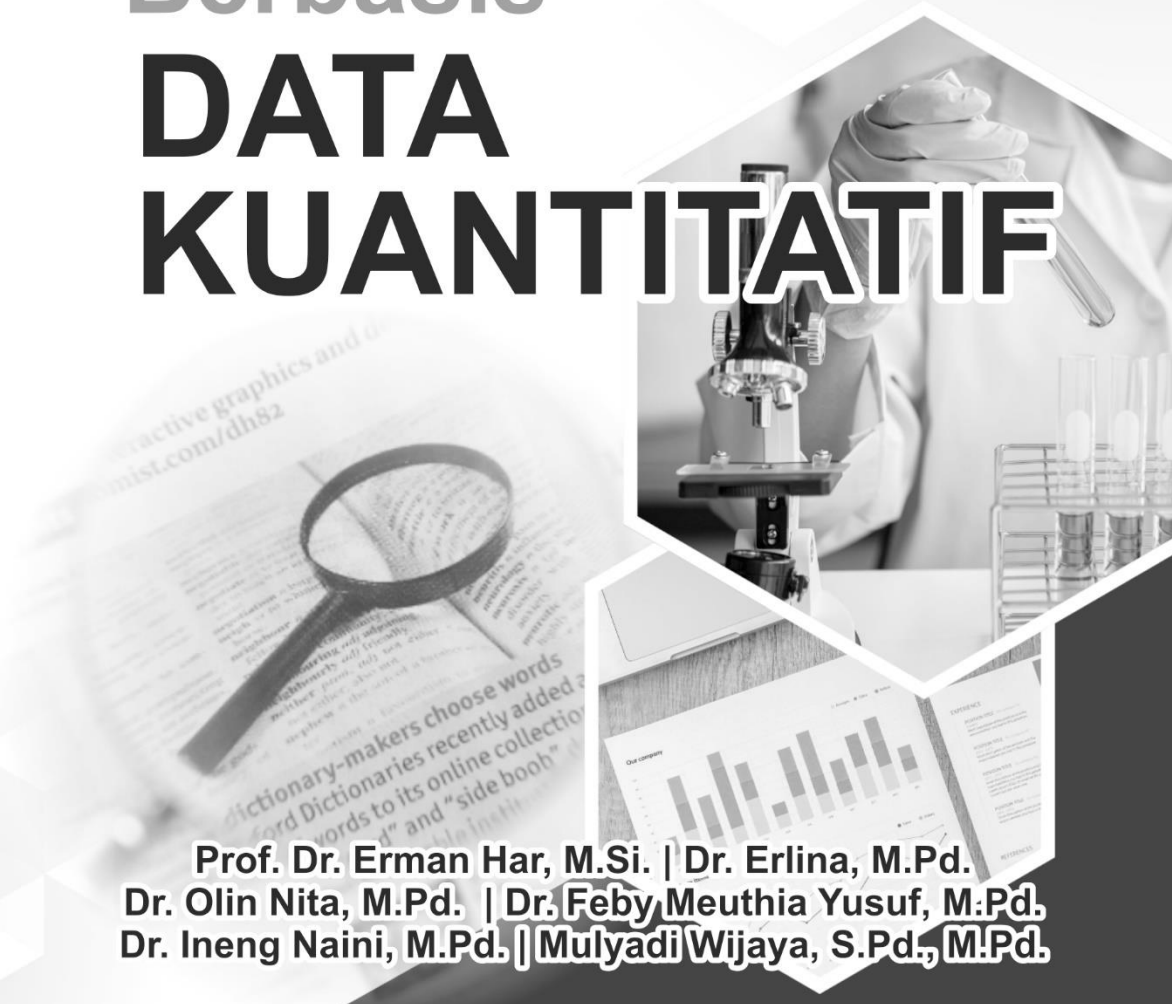
AKURASI RISET Berbasis DATA KUANTITATIF



Prof. Dr. Erman Har, M.Si. | Dr. Erlina, M.Pd.
Dr. Olin Nita, M.Pd. | Dr. Feby Meuthia Yusuf, M.Pd.
Dr. Ineng Naini, M.Pd. | Mulyadi Wijaya, S.Pd., M.Pd.

AKURASI RISET BERBASIS DATA KUANTITATIF

Penulis	:	Prof. Dr. Erman Har, M.Si. Dr. Erlina, M.Pd. Dr. Olin Nita, M.Pd. Dr. Feby Meuthia Yusuf, M.Pd. Dr.Ineng Naini, M.Pd. Mulyadi Wijaya, S.Pd., M.Pd.
Editor	:	Dra. Gusmawetti, M.Si. Prof. Dr. Ristapawa Indra. M.Pd.
Desain Cover	:	Muzammil Akbar
Ilustrasi	:	Gemini

Ukuran: 15.5 x 23 cm; Hal: ix + 189 (198)
Cetakan I, Desember 2025
ISBN 978-634-7244-87-1

Penerbit
Insight Mediatama
Anggota IKAPI No. 338/JTI/2022
Watesnegoro No. 4 (61385) Mojokerto
Whatsapp 087762245559
www.insightmediatama.co.id

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala limpahan rahmat dan karunia-Nya sehingga buku berjudul "AKURASI RISET BERBASIS DATA KUANTITATIF" ini dapat tersusun dengan baik. Buku ini hadir sebagai jawaban atas kebutuhan para peneliti, mahasiswa, dosen, dan praktisi pendidikan maupun sosial yang membutuhkan panduan metodologis komprehensif dalam melakukan penelitian kuantitatif secara akurat, sistematis, dan berbasis data. Di tengah perkembangan teknologi digital, maraknya data besar, serta meningkatnya tuntutan *evidence-based research*, kemampuan melakukan penelitian kuantitatif yang benar menjadi kompetensi ilmiah yang tidak dapat diabaikan.

Buku ini diawali dengan kajian landasan filosofis penelitian kuantitatif, termasuk hakikat penelitian, paradigma positivistik, objektivitas data, logika deduktif, validitas–reliabilitas, serta etika ilmiah. Pendekatan ini mengembalikan riset kuantitatif pada akar epistemologisnya, sehingga pembaca dapat memahami bahwa ketepatan penelitian tidak hanya bertumpu pada statistik, tetapi juga pada integritas dan nilai ilmiah.

Bagian berikutnya membahas proses perumusan masalah penelitian, identifikasi variabel, konseptualisasi dan operasionalisasi konstruk, penyusunan hipotesis, hingga perancangan model konseptual. Tahapan ini sangat menentukan arah dan kualitas riset karena kerangka konseptual yang tepat akan menghasilkan analisis yang lebih akurat dan dapat dipertanggungjawabkan.

Pembahasan selanjutnya memperkenalkan ragam desain penelitian kuantitatif, mulai dari eksperimen, quasi-eksperimen, hingga desain non-eksperimental seperti survei dan studi

korelasional. Selain itu, pembaca dipandu memahami validitas internal dan eksternal serta strategi untuk meningkatkan akurasi desain agar inferensi kausal dapat dicapai dengan lebih kuat.

Pada bagian tentang populasi, sampel, dan teknik sampling, buku ini mengulas konsep populasi target dan terjangkau, sampling frame, teknik probability dan non-probability, perhitungan ukuran sampel, serta kesalahan sampling yang harus diantisipasi. Topik ini menjadi sangat penting karena kualitas data dan kekuatan generalisasi penelitian sangat bergantung pada ketepatan proses sampling.

Bagian instrumen penelitian menyajikan panduan pengembangan instrumen, validitas isi–kriteria–konstruk, analisis faktor, validasi ahli, serta isu kesalahan pengukuran. Pendekatan yang digunakan menekankan bahwa kualitas instrumen merupakan fondasi dari akurasi seluruh proses penelitian.

Pembahasan berikutnya menguraikan reliabilitas dan analisis statistik dasar seperti Cronbach Alpha, split-half, test-retest, statistik deskriptif, normalitas, homogenitas, linearitas, serta analisis hubungan. Materi ini memberikan dasar kuat sebelum pembaca memasuki ranah analisis statistik inferensial.

Selanjutnya dijelaskan teknik analisis inferensial seperti uji *t*, ANOVA, MANOVA, MANCOVA, regresi, korelasi, path analysis, hingga pemodelan struktural (SEM). Bagian ini membantu pembaca memahami bagaimana menguji hipotesis dan menarik kesimpulan berdasarkan data.

Buku ini juga memuat pembahasan analisis kuantitatif lanjutan, termasuk PLS-SEM, analisis mediasi–moderasi, analisis cluster dan diskriminan, machine learning untuk riset sosial, hingga meta-analisis. Pendekatan modern ini memberikan wawasan mengenai metode analitik yang semakin relevan di era digital dan data besar.

Selain itu, disajikan pula panduan penggunaan software statistik seperti SPSS, AMOS, SmartPLS, JASP & R, serta pengantar big data analytics. Penguasaan perangkat lunak menjadi keterampilan penting bagi peneliti agar mampu mengolah dan menafsirkan data secara tepat.

Sebagai penutup, buku ini memberikan panduan pelaporan hasil penelitian, mulai dari struktur penyusunan laporan, interpretasi data secara benar, penyajian tabel dan grafik, pengenalan kesalahan umum dalam pelaporan, hingga penggunaan standar APA sebagai acuan internasional dalam penulisan ilmiah.

Harapan penulis, buku ini dapat menjadi rujukan utama bagi siapa pun yang ingin meningkatkan kemampuan riset kuantitatif, memperkuat akurasi data, serta menghasilkan penelitian yang berintegritas tinggi. Semoga kehadiran buku ini memberi kontribusi nyata dalam pengembangan ilmu pengetahuan dan praktik penelitian berbasis data.

DAFTAR ISI

BAB 1
LANDASAN FILOSOFIS DAN PARADIGMA RISET KUANTITATIF

1.1 Hakikat Penelitian Kuantitatif

Hakikat penelitian kuantitatif tidak dapat dilepaskan dari pijakan filsafat ilmu, khususnya positivisme dan post-positivisme, yang memberi fondasi ontologis, epistemologis, dan metodologis bagi praktik penelitian modern. Menurut Kerlinger (2000), penelitian kuantitatif merupakan "a systematic, controlled, empirical, and critical investigation of natural phenomena guided by theory and hypotheses." Rumusan ini memperlihatkan empat pilar: sistematis, terkontrol, empiris, dan kritikal, yang menjadi struktur dasar pendekatan kuantitatif. Dari perspektif ontologi, pendekatan ini berangkat dari asumsi bahwa realitas bersifat objektif, stabil, dan dapat diukur. Asumsi tersebut sejalan dengan gagasan positivistik Auguste Comte tentang dunia sosial sebagai entitas yang tunduk pada hukum-hukum yang dapat diamati secara ilmiah.

Creswell dan Creswell (2018) menjelaskan bahwa penelitian kuantitatif memfokuskan diri pada pengukuran variabel, pengujian hubungan, dan analisis statistik untuk menghasilkan generalisasi yang dapat dipertanggungjawabkan secara ilmiah. Ini berarti kuantifikasi bukan sekadar prosedur teknis, melainkan representasi epistemologis dari keyakinan bahwa fenomena sosial dapat dipahami melalui struktur numerik. Neuman (2014) menambahkan bahwa penelitian kuantitatif menuntut tingkat presisi tinggi karena data numerik

merepresentasikan fenomena sosial yang kompleks melalui proses reduksi terukur. Dengan demikian, esensi penelitian kuantitatif sesungguhnya adalah usaha menjembatani realitas empiris dan abstraksi teoretis melalui perangkat pengukuran yang valid dan reliabel.

Dari sudut filsafat ilmu, pendekatan kuantitatif tumbuh dari tradisi positivistik, namun dalam penelitian modern dipengaruhi pula oleh post-positivisme. Popper (2002) memberikan landasan kritis melalui konsep falsifikasi, yakni bahwa teori ilmiah harus memungkinkan pengujian yang berpotensi menyalahkannya. Prinsip falsifikasionisme Popper menggeser cara pandang positivistik yang terlalu menekankan verifikasi menuju paradigma ilmiah yang lebih kritis, reflektif, dan terbuka terhadap revisi. Perspektif ini memengaruhi penelitian kuantitatif kontemporer yang tidak lagi memandang data sebagai representasi sempurna realitas, tetapi sebagai indikator probabilistik yang harus diuji ketepatan dan konsistensinya.

Gall, Gall, dan Borg (2019) menekankan bahwa penelitian kuantitatif bertumpu pada proses deduksi dari teori menuju hipotesis yang kemudian diuji melalui pengumpulan data. Dengan demikian, penelitian kuantitatif bukan sekadar mengolah angka, tetapi mengikuti logika ilmiah yang bersifat deduktif. Fraenkel, Wallen, dan Hyun (2021) menguraikan bahwa hubungan antara teori dan data bersifat siklik: teori menghasilkan hipotesis; hipotesis diuji melalui data; hasilnya memperkuat atau melemahkan teori. Siklus inilah yang menjadikan penelitian kuantitatif selalu berorientasi pada pembuktian hubungan kausal maupun korelasional.

Babbie (2021) menambahkan bahwa penelitian kuantitatif mengambil posisi epistemologis bahwa peneliti harus menjaga jarak objektif terhadap fenomena yang diteliti. Objektivitas bukan

sekadar sikap mental, tetapi mekanisme metodologis seperti kontrol variabel, desain eksperimen, prosedur sampling yang representatif, instrumen terstandar, serta pengujian validitas dan reliabilitas. Cronbach memandang reliabilitas sebagai prasyarat dasar karena tanpa konsistensi pengukuran, tidak mungkin melakukan inferensi yang kredibel. Campbell dan Stanley (1963) memperkuat hal ini melalui kerangka validitas internal dan validitas eksternal sebagai standar kualitas desain eksperimen dan quasi-eksperimen.

Dalam era digital, hakikat penelitian kuantitatif berkembang seiring hadirnya big data, machine learning, AI-assisted analytics, serta reproducible research. McMillan dan Schumacher (2014) menekankan bahwa evidence-based inquiry dalam pendidikan menuntut data yang diperoleh melalui prosedur ilmiah yang ketat dan dapat direplikasi. Johnson dan Christensen (2020) menyebut integrasi statistik komputasional sebagai karakter baru penelitian kuantitatif modern, yang tetap berakar pada logika ilmiah tetapi diperluas melalui teknologi digital yang memungkinkan analisis berskala besar. Namun demikian, Neuman (2014) mengingatkan bahwa data besar tetap membutuhkan landasan epistemologis: data tidak pernah netral, melainkan hasil konstruksi metodologis.

Penelitian kuantitatif memiliki karakteristik fundamental yang membedakannya dari pendekatan lain. Pertama, pengukuran variabel: variabel harus didefinisikan secara operasional sehingga dapat diukur dengan instrumen yang memenuhi standar validitas dan reliabilitas. Kedua, struktur desain penelitian: desain eksperimen, survei, dan korelasional harus mengikuti prinsip logika ilmiah agar hubungan antar-variabel dapat diinterpretasi secara tepat. Ketiga, generalizability: menurut Punch (2014), data kuantitatif berorientasi pada

generalisasi temuan dari sampel ke populasi, sehingga representativitas menjadi syarat epistemologis. Keempat, analisis statistik sebagai alat inferensi: statistik bukan tujuan penelitian, melainkan instrumen untuk menguji hipotesis dan memperkirakan kebenaran teoretis berdasarkan data empiris.

Selain itu, hakikat penelitian kuantitatif tidak dapat dipahami tanpa memperhatikan aspek etika penelitian. APA (2020) menegaskan pentingnya integritas ilmiah, kejujuran dalam analisis data, penghindaran manipulasi angka, serta transparansi dalam pelaporan hasil. Etika bukan aspek tambahan, melainkan fondasi epistemologis karena keandalan ilmu bergantung pada kredibilitas prosesnya. Dalam konteks penelitian pendidikan, etika menjadi semakin penting karena menyangkut kerahasiaan siswa, guru, serta institusi pendidikan yang menjadi unit analisis.

Secara filosofis, penelitian kuantitatif menempatkan akurasi, objektivitas, dan generalisasi sebagai nilai ilmiah utama. Namun perkembangan riset modern mendorong penyesuaian paradigma. Post-positivisme menegaskan bahwa pengetahuan ilmiah bersifat tentatif, probabilistik, dan dapat direvisi. Karena itu, penelitian kuantitatif masa kini lebih menekankan estimasi probabilistik, confidence intervals, effect size, serta reproducible workflows, bukan hanya signifikansi statistik. Perubahan ini mencerminkan pergeseran dari positivisme klasik menuju paradigma ilmiah yang lebih reflektif dan berbasis ketidakpastian (uncertainty-aware science).

Hakikat penelitian kuantitatif memberikan implikasi penting bagi metodologi penelitian pendidikan. Peneliti harus memahami bahwa angka hanyalah representasi fenomena, bukan fenomena itu sendiri; bahwa teori harus diuji secara kritikal, bukan diterima secara dogmatis; dan bahwa akurasi statistik harus disertai integritas etika. Dalam era big data, logika

ilmiah kuantitatif semakin relevan untuk mengembangkan evidence-based decision making dalam pendidikan. Namun tantangan metodologis semakin kompleks sehingga peneliti perlu memiliki literasi statistik, literasi digital, dan literasi etik yang kuat untuk memastikan bahwa penelitian benar-benar berkontribusi pada peningkatan mutu ilmu maupun praktik pendidikan.

1.2 Paradigma Positivistik dan Objektivitas Data

Paradigma positivistik menempati posisi fundamental dalam evolusi penelitian kuantitatif. Menurut Babbie (2021), positivisme merupakan keyakinan filosofis bahwa realitas sosial dapat dipahami melalui metode ilmiah yang sama dengan ilmu alam, yakni observasi terkontrol, pengukuran objektif, dan analisis sistematis. Asal-usulnya dapat ditelusuri pada Auguste Comte yang memandang ilmu sebagai sarana untuk menemukan hukum-hukum universal yang mengatur gejala sosial. Kerlinger (2000) menegaskan bahwa positivisme memberi landasan epistemologis bagi penelitian kuantitatif: dunia sosial diperlakukan sebagai objek yang dapat diobservasi secara empiris melalui prosedur terstandar, bebas nilai, dan mengikuti hukum kausalitas.

Cohen, Manion, dan Morrison (2018) menjelaskan bahwa paradigma ini dibangun di atas tiga asumsi utama: (1) realitas bersifat objektif dan eksis secara independen dari peneliti; (2) pengetahuan diperoleh melalui pengukuran dan observasi yang dapat diverifikasi; dan (3) hukum-hukum yang mengatur fenomena sosial dapat dipahami melalui desain penelitian terkontrol dan analisis statistik. Akibatnya, penelitian kuantitatif memandang objektivitas sebagai syarat epistemologis: data harus diperoleh melalui instrumen yang terstandar, bukan interpretasi subjektif peneliti.

Namun positivisme tidak berkembang tanpa kritik. Popper (2002) mengkritik tradisi verifikasionisme positivistik yang menganggap bahwa akumulasi bukti empiris cukup untuk membenarkan teori. Ia menegaskan bahwa teori ilmiah harus terbuka terhadap pengujian kritis melalui falsifikasi. Dengan demikian, objektivitas ilmiah bukan semata hasil pengamatan, tetapi produk dari *critical scrutiny* dan prosedur metodologis yang memungkinkan klaim ilmiah ditolak jika tidak sesuai dengan data. Pandangan ini memengaruhi perkembangan post-positivisme, yaitu pandangan bahwa realitas memang objektif, tetapi pemahaman manusia terhadapnya selalu tidak sempurna dan probabilistik.

Creswell dan Creswell (2018) menekankan bahwa penelitian kuantitatif modern lebih dekat dengan post-positivisme, bukan positivisme murni. Dalam pendekatan post-positivistik, objektivitas bukan berarti peneliti benar-benar bebas nilai, melainkan berusaha meminimalkan bias melalui prosedur ilmiah yang ketat: penggunaan instrumen reliabel, kontrol variabel, desain eksperimen yang valid, prosedur pengkodean data, dan pelaporan transparan. Hal ini selaras dengan gagasan Neuman (2014) bahwa objektivitas dalam riset sosial adalah upaya metodologis, bukan kondisi mutlak.

Gall, Gall, dan Borg (2019) menunjukkan bagaimana positivisme mempengaruhi desain penelitian kuantitatif, terutama dalam penyusunan hipotesis, pengukuran variabel, dan pencarian hubungan kausal. Logika ilmiah positivistik memandang bahwa jika variabel dapat didefinisikan secara operasional dan diukur secara konsisten, maka hubungan antar-variabel dapat dianalisis secara statistik untuk menghasilkan inferensi ilmiah. Di sinilah objektivitas data berperan sebagai

dasar dalam menguji teori, mengidentifikasi pola, dan melakukan generalisasi ke populasi yang lebih luas.

Dalam praktik penelitian, objektivitas tidak hanya berkaitan dengan instrumen, tetapi juga dengan proses pengumpulan data. Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa representativitas sampel adalah komponen esensial objektivitas, karena generalisasi hanya dapat dilakukan jika sampel mencerminkan karakteristik populasi. Punch (2014) menambahkan bahwa objektivitas menuntut prosedur sampling yang bebas bias, misalnya melalui teknik random sampling atau stratified sampling yang dirancang untuk memastikan keberagaman data.

Objektivitas data dalam penelitian kuantitatif juga sangat bergantung pada konsep validitas dan reliabilitas. Menurut Cronbach, reliabilitas merupakan indikator konsistensi pengukuran, sehingga sebuah instrumen yang reliabel dapat menghasilkan data yang stabil meskipun diuji ulang pada kondisi yang sama. Campbell dan Stanley (1963) mengembangkan kerangka validitas internal dan eksternal sebagai kriteria objektivitas dalam desain eksperimen. Validitas internal memastikan bahwa hubungan antar-variabel benar-benar disebabkan oleh manipulasi eksperimen, bukan variabel pengganggu. Validitas eksternal memastikan hasil dapat digeneralisasikan ke konteks lain. Dengan demikian, objektivitas bukan sekadar kondisi ideal, tetapi hasil dari rancangan metodologis yang matang.

Perkembangan teknologi digital membawa tantangan baru terhadap konsep objektivitas data. McMillan dan Schumacher (2014) menegaskan bahwa dalam era evidence-based education, keputusan pendidikan harus didasarkan pada data valid dan dapat direplikasi. Namun big data menghadapkan peneliti pada

keberlimpahan informasi yang sering kali diperoleh tanpa kontrol sampling, sehingga menimbulkan pertanyaan epistemologis mengenai representativitas. Johnson dan Christensen (2020) menekankan bahwa meskipun big data dapat meningkatkan kekuatan analisis, objektivitas tetap bergantung pada kualitas algoritma, definisi variabel, serta prosedur pemrosesan data. Artinya, data besar tidak otomatis objektif; objektivitasnya dibentuk melalui desain metodologis dan refleksi epistemologis.

Neuman (2014) mengingatkan bahwa data digital membawa potensi bias baru, seperti *algorithmic bias* dan *measurement error* berbasis teknologi. Karena itu, penelitian kuantitatif modern menekankan prinsip reproducible research, yakni bahwa proses analisis harus dapat diulang oleh peneliti lain dengan hasil yang konsisten. Prinsip ini selaras dengan tuntutan objektivitas positivistik, namun diformulasikan ulang dalam konteks teknologi digital. Transparansi kode analisis, dokumentasi dataset, serta pelaporan prosedur statistik menjadi bagian integral dari objektivitas ilmiah di era modern.

Selain aspek metodologis, objektivitas perlu ditopang oleh etika penelitian. APA (2020) menegaskan bahwa objektivitas mengandung aspek moral: peneliti harus jujur, tidak memanipulasi data, tidak melakukan *p-hacking*, tidak menyembunyikan hasil yang tidak signifikan, serta menjaga integritas proses ilmiah. Dengan demikian, objektivitas bukan hanya persoalan teknis, tetapi juga karakter etis seorang ilmuwan.

Paradigma positivistik telah membentuk fondasi penelitian kuantitatif, namun perkembangan post-positivisme dan teknologi digital menuntut reinterpretasi konsep objektivitas. Penelitian kuantitatif masa kini perlu memadukan prinsip positivistik tentang observasi empiris dengan kesadaran post-positivistik

bahwa pengukuran manusia selalu mengandung ketidaksempurnaan. Objektivitas tidak lagi dipahami sebagai kondisi mutlak, melainkan *aspirasi metodologis*, yakni upaya terus-menerus untuk mengurangi bias melalui desain penelitian yang ketat, prosedur transparan, algoritma yang etis, dan pelaporan yang dapat direplikasi. Bagi penelitian pendidikan, paradigma ini menegaskan bahwa data harus menjadi dasar pengambilan keputusan, namun interpretasinya harus dilakukan dengan kewaspadaan epistemologis agar tidak terjebak pada determinisme angka semata.

1.3 Logika Deduktif dalam Penelitian

Logika deduktif merupakan fondasi epistemologis yang mengarahkan desain penelitian kuantitatif. Menurut Kerlinger (2000), pendekatan kuantitatif selalu bergerak dari teori ke data: teori dijadikan kerangka berpikir untuk menurunkan proposisi, merumuskan hipotesis, dan mengujinya dengan bukti empiris. Pandangan ini sejalan dengan tradisi positivistik yang menempatkan teori sebagai perangkat rasional untuk menjelaskan gejala, sekaligus prinsip metodologis untuk mengarahkan prosedur penelitian. Dalam kerangka deduktif, penelitian menjadi proses sistematis yang menguji konsistensi antara apa yang diprediksi oleh teori dan apa yang terjadi dalam kenyataan.

Creswell dan Creswell (2018) menjelaskan bahwa dalam penelitian kuantitatif, hipotesis adalah “statements that predict relationships between variables based on existing theory.” Dengan demikian, hipotesis tidak muncul secara intuitif, tetapi merupakan hasil logika deduktif dari teori. Gall, Gall, dan Borg (2019) menegaskan bahwa validitas penelitian kuantitatif bergantung pada ketepatan proses deduksi: jika teori dirumuskan

secara jelas, maka hipotesis yang diturunkannya pun dapat diuji dengan instrumen pengukuran yang tepat. Deduksi menjadi alat penghubung antara abstraksi teoritis dan realitas empiris.

Dalam perspektif filsafat ilmu, logika deduktif memperoleh penguatan melalui pemikiran Popper (2002) tentang falsifikasi. Popper menolak verifikasionisme dan menegaskan bahwa teori ilmiah harus menghasilkan prediksi yang dapat diuji dan—jika meleset—dapat disalahkan (falsified). Ini menunjukkan bahwa deduksi bukan sekadar menurunkan pernyataan, tetapi merumuskan "risiko epistemologis," yaitu peluang bagi teori untuk diuji secara ketat. Dengan demikian, penelitian deduktif tidak bertujuan membuktikan teori, tetapi menguji ketahanannya terhadap data. Proses deduktif dalam penelitian kuantitatif modern harus memahami konsep ini: hipotesis deduktif tidak hanya diuji untuk mencari dukungan data, tetapi juga untuk mencari ketidaksesuaian yang dapat memperbaiki teori.

Babbie (2021) menunjukkan bahwa logika deduktif memerlukan konsistensi internal antara teori, konsep, variabel, dan indikator. Setiap konsep abstrak harus didefinisikan secara operasional agar dapat diukur secara empiris. Neuman (2014) menekankan bahwa definisi operasional merupakan langkah epistemologis penting karena mengubah konsep teoritis menjadi entitas yang dapat diobservasi. Proses ini memastikan bahwa hubungan deduktif tidak berhenti pada level abstraksi, tetapi terhubung dengan realitas empiris melalui pengukuran kuantitatif.

Fraenkel, Wallen, dan Hyun (2021) menggarisbawahi bahwa logika deduktif dalam penelitian pendidikan memiliki implikasi khusus karena variabel-variabel yang diteliti (motivasi, kemampuan, prestasi, persepsi) sering kali bersifat laten dan tidak dapat diukur secara langsung. Oleh karena itu, instrumen

pengukuran harus dirancang melalui validasi teoretis dan empiris yang ketat. Reliabilitas Cronbach dan validitas Campbell & Stanley menjadi prasyarat metodologis agar data empiris dapat menilai kebenaran deduksi teoretis secara akurat.

Punch (2014) menjelaskan bahwa kekuatan logika deduktif terletak pada kemampuannya menghasilkan *predictive power*. Sebuah teori yang kuat akan menghasilkan prediksi yang jelas, sehingga hipotesis yang diturunkannya dapat diuji melalui desain eksperimen, korelasional, atau survei. Dalam desain eksperimen, misalnya, deduksi menentukan variabel bebas yang dimanipulasi, variabel terikat yang diukur, dan variabel kontrol yang dijaga. Tanpa kerangka deduktif, penelitian akan kehilangan arah dan berisiko menjadi proses pengumpulan data tanpa struktur ilmiah.

Dalam konteks big data dan statistik modern, logika deduktif tetap relevan meskipun metode analisis berkembang. Johnson dan Christensen (2020) menegaskan bahwa meskipun machine learning dan eksplorasi data bersifat induktif, interpretasi ilmiah tetap membutuhkan deduksi dari teori. Tanpa landasan deduktif, analisis statistik mudah terjebak dalam korelasi semu (*spurious correlations*) yang tidak memiliki landasan teoretis. Oleh karena itu, penelitian modern menekankan integrasi antara deduksi teoretis dan analisis berbasis data, sehingga penemuan statistik dapat dipahami dalam kerangka ilmiah yang koheren.

Logika deduktif memainkan peran penting dalam struktur pelaporan penelitian. Menurut McMillan dan Schumacher (2014), laporan penelitian kuantitatif selalu mengikuti alur deduktif: mulai dari teori, rumusan masalah, hipotesis, metode, analisis data, dan kesimpulan yang mengarah kembali pada teori. Struktur ini mencerminkan cara kerja deduksi sebagai kerangka berpikir ilmiah. Standar APA (2020) mendukung hal ini melalui

struktur pelaporan yang sistematis dan konsisten, memastikan bahwa hasil empiris selalu dihubungkan dengan teori yang mendasarinya.

Di era reproducible research, logika deduktif memperoleh peran baru sebagai mekanisme transparansi ilmiah. Data, kode analisis, dan prosedur penelitian harus dicatat secara jelas sehingga peneliti lain dapat mereplikasi pengujian hipotesis yang sama. Neuman (2014) menekankan bahwa salah satu fungsi deduksi adalah memastikan bahwa proses ilmiah dapat diikuti dan diverifikasi oleh komunitas ilmiah. Transparansi tidak hanya mendukung objektivitas data, tetapi juga memperkuat validitas pengujian deduktif.

Meski demikian, logika deduktif tidak bebas dari kritik. Tokoh-tokoh interpretivis berargumen bahwa fenomena sosial terlalu kompleks untuk dijelaskan melalui struktur deduksi yang linear. Namun penelitian kuantitatif menanggapi kritik tersebut melalui pendekatan post-positivistik yang mengakui probabilisme: deduksi tidak menghasilkan kepastian mutlak, tetapi estimasi probabilitas. Hal ini tercermin dalam penggunaan *effect size*, *confidence intervals*, dan model statistik probabilistik. Dengan demikian, logika deduktif tetap menjadi kerangka utama, tetapi diintegrasikan dengan pemahaman bahwa pengetahuan ilmiah bersifat tentatif dan revisibel.

Logika deduktif merupakan tulang punggung penelitian kuantitatif karena memberikan arah, kontrol, dan konsistensi ilmiah. Melalui deduksi, teori diuji secara kritis dan sistematis sehingga menghasilkan pengetahuan yang dapat dipertanggungjawabkan. Namun dalam penelitian modern, deduksi harus dipahami secara dinamis: teori tidak hanya mengarahkan penelitian, tetapi juga diuji melalui ketidakpastian, probabilitas, dan kemungkinan falsifikasi. Bagi penelitian

pendidikan, logika deduktif mengajarkan pentingnya kerangka teoretis, kejelasan konsep, dan pengukuran yang valid agar penelitian benar-benar berkontribusi pada pengembangan kebijakan, praktik pembelajaran, dan inovasi pedagogis berbasis bukti.

1.4 Validitas, Reliabilitas, dan Replikasi

Validitas, reliabilitas, dan replikasi merupakan tiga pilar epistemologis yang memastikan bahwa penelitian kuantitatif benar-benar menghasilkan pengetahuan ilmiah yang akurat dan dapat dipercaya. Dalam pandangan Kerlinger (2000), kualitas penelitian kuantitatif terletak pada akurasi dalam mendeskripsikan dan memprediksi fenomena perilaku, sehingga ketepatan pengukuran menjadi syarat mutlak. Validitas menilai apakah instrumen benar-benar mengukur konstruk yang seharusnya diukur, reliabilitas menilai kestabilan hasil pengukuran, dan replikasi menilai ketahanan temuan ketika diuji ulang oleh peneliti lain. Ketiganya tidak dapat dipisahkan karena secara filosofis merupakan fondasi objektivitas dan keilmuan dalam tradisi positivistik maupun post-positivistik.

Validitas dipahami sebagai ketepatan representasi realitas melalui data. Creswell dan Creswell (2018) menegaskan bahwa validitas bukanlah sifat instrumen semata, tetapi kualitas interpretasi yang dibuat berdasarkan skor pengukuran. Fraenkel, Wallen, dan Hyun (2021) menambahkan bahwa validitas merupakan proses evaluasi terhadap kesesuaian antara teori, definisi operasional, indikator, dan hasil empiris. Kerangka klasik validitas banyak dipengaruhi oleh Campbell dan Stanley (1963) yang memperkenalkan dua dimensi utama: validitas internal dan validitas eksternal. Validitas internal mengacu pada sejauh mana perubahan dalam variabel terikat dapat diatribusikan secara sah

kepada variabel bebas, bukan faktor lain atau variabel pengganggu. Desain eksperimen yang kuat, randomisasi, dan kontrol variabel menjadi mekanisme utama untuk memperkuat validitas internal ini. Sebaliknya, validitas eksternal berorientasi pada kemampuan generalisasi temuan ke populasi dan konteks yang lebih luas. Punch (2014) menegaskan bahwa generalisasi ilmiah baru dapat dipertanggungjawabkan apabila sampel bersifat representatif dan kondisi penelitian tidak terlalu artifisial sehingga tidak relevan dengan konteks dunia nyata.

Validitas konstruk, seperti dijelaskan Neuman (2014), menjadi dimensi yang sangat penting dalam penelitian pendidikan dan perilaku, mengingat banyak variabel tidak tampak secara langsung, seperti motivasi, sikap, atau kepuasan belajar. Konstruk tersebut hanya dapat diakses melalui indikator empiris yang dihasilkan dari proses definisi operasional. Dengan demikian, validitas konstruk merupakan ujian langsung terhadap kualitas hubungan antara teori dan empiris, sehingga penelitian kuantitatif yang tidak memiliki validitas konstruk yang kuat dapat menghasilkan kesimpulan yang menyesatkan. Di era big data, tantangan validitas menjadi semakin kompleks. Johnson dan Christensen (2020) mengingatkan bahwa data besar sering dihasilkan melalui proses otomatis yang tidak dirancang dengan prinsip-prinsip pengukuran ilmiah sehingga meskipun volume data sangat besar, validitas konstruk atau validitas eksternal tidak otomatis meningkat. McMillan dan Schumacher (2014) juga menekankan bahwa data digital rentan terhadap bias algoritmik, keterbatasan representativitas pengguna internet, serta distorsi yang dihasilkan sistem digital. Oleh karena itu, validitas tetap merupakan proses kritis yang tidak dapat diabaikan hanya karena kuantitas data berlimpah.

Reliabilitas merupakan dimensi kedua yang menjadi syarat dasar bagi konsistensi pengukuran. Dalam pandangan Cronbach, reliabilitas menggambarkan sejauh mana instrumen dapat menghasilkan hasil yang stabil dan konsisten. Tanpa konsistensi, tidak mungkin sebuah instrumen menghasilkan inferensi ilmiah yang valid. Fraenkel, Wallen, dan Hyun (2021) menjelaskan bahwa reliabilitas mencakup beberapa bentuk, seperti test-retest reliability yang mengukur kestabilan hasil dari waktu ke waktu, internal consistency yang mengukur koherensi antar-butir instrumen (misalnya melalui Cronbach's alpha), serta inter-rater reliability yang memastikan bahwa penilai berbeda menghasilkan skor yang relatif sama. Babbie (2021) menggarisbawahi bahwa reliabilitas bukan hanya persoalan teknis pengukuran, tetapi fondasi epistemologis yang menentukan apakah data dapat diandalkan untuk menggambarkan realitas. Jika reliabilitas rendah, data menjadi "noise" dan hubungan antar-variabel tidak dapat ditafsirkan secara sah.

Dalam penelitian pendidikan, reliabilitas menghadapi tantangan tersendiri karena banyak konstruk bersifat laten dan dipengaruhi konteks. Oleh karena itu, desain instrumen harus disertai pilot study, analisis butir, dan pengujian ulang untuk memastikan konsistensi. Gall, Gall, dan Borg (2019) menekankan bahwa reliabilitas merupakan proses berkelanjutan, bukan sekadar angka koefisien. Di era digital, reliabilitas algoritma juga menjadi isu penting. Neuman (2014) mengingatkan bahwa sistem digital dapat tampak konsisten tetapi justru menghasilkan konsistensi yang bias apabila algoritma dibangun dengan asumsi yang keliru. Dengan demikian, reliabilitas modern harus diperluas tidak hanya pada instrumen tradisional tetapi juga pada perangkat komputasional, sistem pengolahan data, dan metode statistik otomatis.

Replikasi, sebagai pilar ketiga, merupakan ujian tertinggi dalam ilmu pengetahuan. Kerlinger (2000) menyebut replikasi sebagai "the supreme test of scientific truth" karena hanya melalui pengulangan penelitian oleh peneliti lain temuan dapat memperoleh status ilmiah. Creswell dan Creswell (2018) menekankan bahwa replikasi tidak mungkin terjadi apabila metode penelitian tidak dilaporkan secara transparan. Neuman (2014) membedakan replikasi menjadi tiga jenis: replikasi langsung (mengulangi penelitian dengan kondisi identik), replikasi sistematis (mengulangi penelitian dengan variasi kondisi tertentu), dan replikasi konseptual (menguji teori yang sama dengan pendekatan berbeda). Dalam penelitian pendidikan, replikasi sangat penting untuk memastikan bahwa intervensi atau kebijakan tidak hanya efektif dalam satu konteks, tetapi juga dalam konteks lain. McMillan dan Schumacher (2014) menyatakan bahwa replikasi merupakan penopang utama evidence-based practice karena efektivitas suatu praktik hanya dapat dipercaya apabila ditemukan secara konsisten pada berbagai studi.

Di era reproducible research, replikasi memperoleh posisi baru yang semakin krusial. Data mentah, kode analisis, dokumentasi prosedur, dan pelaporan statistik harus disediakan agar peneliti lain dapat memverifikasi dan mengulangi analisis. Hal ini semakin penting dalam penelitian berbasis big data dan machine learning yang proses analisisnya sering kali kompleks dan tidak transparan. APA (2020) menegaskan bahwa transparansi analisis, kejujuran pelaporan, dan integritas dalam penyajian data merupakan bentuk etika ilmiah yang memungkinkan replikasi dan menjaga kepercayaan publik terhadap ilmu pengetahuan.

Ketika dipandang dari sudut epistemologis, validitas, reliabilitas, dan replikasi merupakan rangkaian konsep yang saling terkait. Reliabilitas memberikan konsistensi, validitas memberikan ketepatan representasi realitas, dan replikasi memberikan ketahanan terhadap kritik dan pengujian ulang oleh komunitas ilmiah. Punch (2014) menegaskan bahwa ketiganya membentuk fondasi sains yang bersifat kumulatif, kritis, dan berkembang melalui koreksi diri. Pandangan ini sejalan dengan semangat post-positivistik bahwa ilmu tidak menghasilkan kepastian absolut, tetapi estimasi probabilistik yang terus disempurnakan melalui proses falsifikasi Popperian dan pengujian berulang.

Secara metodologis, prinsip validitas, reliabilitas, dan replikasi membawa implikasi penting bagi penelitian pendidikan. Instrumen harus dirancang secara teoretis dan empiris, proses pengumpulan data harus dicatat secara rinci, analisis statistik harus transparan, dan temuan harus dilaporkan dengan integritas. Dalam era big data, peneliti harus mewaspadai bias algoritmik dan memastikan bahwa proses analisis dapat direplikasi. Tanpa tiga pilar ini, penelitian kuantitatif akan kehilangan legitimasi ilmiahnya dan tidak dapat berkontribusi secara signifikan pada evidence-based education maupun kebijakan pendidikan yang akuntabel.

1.5 Etika Ilmiah dalam Penelitian

Etika ilmiah dalam penelitian kuantitatif merupakan fondasi moral dan epistemologis yang memastikan bahwa proses pencarian pengetahuan tidak hanya akurat secara metodologis, tetapi juga benar secara moral. Pada tingkat paling dasar, etika penelitian berangkat dari keyakinan bahwa ilmu harus dibangun di atas kejujuran, tanggung jawab, dan penghormatan terhadap

martabat manusia. Kerlinger (2000) menegaskan bahwa ilmu merupakan "a public enterprise" yang menuntut integritas karena pengetahuan ilmiah hanya bernilai jika dapat dipercaya oleh komunitas akademik. Persoalan etika bukan sekadar aturan administratif, tetapi merupakan inti dari hakikat keilmuan itu sendiri. Popper (2002) mengingatkan bahwa falsifikasi ilmiah hanya mungkin terjadi dalam masyarakat ilmiah yang jujur, terbuka pada kritik, dan menghargai transparansi. Dengan demikian, pelanggaran etika bukan hanya kesalahan moral, tetapi juga ancaman langsung terhadap proses produksi pengetahuan ilmiah.

Creswell dan Creswell (2018) menjelaskan bahwa etika harus mengarahkan setiap tahap penelitian, mulai dari perumusan masalah, desain metode, pengumpulan data, analisis, hingga pelaporan hasil. Dalam penelitian pendidikan, etika menjadi semakin penting karena subjek penelitian biasanya adalah siswa, guru, atau institusi pendidikan yang rentan. Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa peneliti wajib menghormati prinsip perlindungan terhadap peserta, termasuk hak atas privasi, kerahasiaan data, dan kebebasan untuk berpartisipasi tanpa tekanan. Dalam kerangka ini, persetujuan sadar (*informed consent*) bukan hanya formalitas, tetapi tindakan etis untuk memastikan bahwa peserta memahami tujuan penelitian, risiko yang mungkin muncul, dan hak mereka untuk menolak. Babbie (2021) mengingatkan bahwa tanpa informed consent, penelitian berpotensi memosisikan peserta sebagai objek, bukan subjek otonom yang memiliki hak atas dirinya.

Aspek etika juga berkaitan erat dengan validitas ilmiah. Neuman (2014) menyebut bahwa manipulasi data, fabrikasi, atau penghilangan temuan yang tidak sesuai hipotesis (selective

reporting) merupakan bentuk pelanggaran etika yang meruntuhkan validitas sekaligus merusak integritas ilmu. Ilmuwan tidak boleh menjadikan statistik sebagai alat manipulasi untuk mencapai hasil yang diinginkan. Oleh karena itu, penggunaan prosedur seperti *p-hacking*, *data dredging*, atau mengubah hipotesis setelah melihat hasil data (HARKing: Hypothesizing After Results are Known) merupakan pelanggaran serius terhadap etika ilmiah. APA (2020) secara eksplisit mewajibkan peneliti untuk melaporkan metode, data, dan hasil secara transparan, termasuk temuan yang tidak signifikan atau yang bertentangan dengan harapan. Kejujuran pelaporan merupakan bentuk tanggung jawab ilmiah karena kesalahan interpretasi atau pelaporan dapat mengakibatkan kesimpulan kebijakan atau praktik pendidikan yang keliru.

Etika ilmiah juga mencakup perlindungan kerahasiaan data. Penelitian kuantitatif sering melibatkan pengumpulan data sensitif, seperti nilai akademik, identitas demografis, atau persepsi peserta didik. Menurut McMillan dan Schumacher (2014), kerahasiaan bukan hanya upaya administratif, tetapi komitmen moral untuk menjaga kepercayaan peserta. Data harus dianonimkan, disimpan dengan aman, dan digunakan hanya untuk tujuan ilmiah. Dalam era digital, tantangan etika menjadi lebih kompleks. Johnson dan Christensen (2020) mengingatkan bahwa data digital dapat dengan mudah dibagikan, disalin, atau dimanfaatkan secara tidak bertanggung jawab, sehingga peneliti harus memastikan bahwa seluruh proses pengelolaan data mematuhi prinsip keamanan informasi, perlindungan privasi, dan kepatuhan terhadap peraturan. Penelitian big data menuntut refleksi etis tambahan karena data yang terkumpul sering kali tidak diperoleh melalui persetujuan individu, misalnya data dari media sosial atau sistem daring. Hal ini membuka perdebatan

epistemologis dan moral tentang batas-batas penggunaan data publik dan tanggung jawab peneliti dalam memastikan bahwa analisis data tidak merugikan kelompok tertentu.

Gall, Gall, dan Borg (2019) menyoroti bahwa pelanggaran etika dalam penelitian pendidikan dapat berdampak langsung pada proses pembelajaran dan kebijakan sekolah. Misalnya, penyebutan identitas sekolah dalam hasil penelitian tanpa izin dapat memengaruhi reputasi institusi. Demikian pula, analisis statistik yang salah atau dilebih-lebihkan dapat berujung pada implementasi kebijakan pendidikan yang tidak berdasarkan bukti sahih. Karena itu, etika ilmiah bukan sekadar menjaga peserta penelitian, tetapi juga menjaga masyarakat dari konsekuensi buruk akibat kesalahan interpretasi ilmiah. Dalam perspektif yang lebih luas, ilmu tanpa etika kehilangan arah dan dapat menjadi alat pembenaran yang menyesatkan.

Reproducible research memberikan dimensi baru dalam etika ilmiah. Dalam ekosistem sains modern, transparansi merupakan nilai etis sekaligus teknis. Peneliti harus memberikan akses pada prosedur analisis, kode komputasi, dokumentasi data, dan proses statistika sehingga peneliti lain dapat menguji ulang hasil penelitian. Neuman (2014) menegaskan bahwa keterbukaan metodologis merupakan syarat agar komunitas ilmiah dapat melakukan kritik, replikasi, dan verifikasi temuan. Tanpa transparansi, replikasi mustahil dilakukan, dan ilmu kehilangan kemampuan koreksi diri. APA (2020) juga menegaskan bahwa peneliti wajib menjaga integritas dalam analisis data digital dengan melaporkan semua prosedur secara jujur. Dalam konteks big data dan AI, keterbukaan algoritma dan model analisis menjadi isu etika baru. Peneliti harus menjelaskan bagaimana data diproses, bagaimana model memprediksi hasil, dan apakah algoritma yang digunakan berpotensi bias kepada kelompok

tertentu. Dengan demikian, etika ilmiah sekarang mencakup tanggung jawab untuk memastikan bahwa kecerdasan buatan digunakan dengan adil, akurat, dan tidak merugikan.

Secara epistemologis, etika ilmiah tidak dapat dipisahkan dari validitas dan reliabilitas. Jika validitas memastikan ketepatan representasi realitas, dan reliabilitas memastikan konsistensi pengukuran, maka etika memastikan bahwa proses empiris dilakukan dengan integritas moral. Punch (2014) menegaskan bahwa kualitas ilmiah yang sejati hanya mungkin terjadi ketika peneliti menggabungkan ketepatan metodologis dengan komitmen etis. Etika menjaga agar ilmu tidak terjatuh menjadi sekadar produksi angka, tetapi tetap berpegang pada tujuan utama ilmu: mencari kebenaran demi kebaikan bersama.

Dalam penelitian pendidikan, implikasi metodologis etika ilmiah sangat luas. Peneliti harus menjaga keamanan data siswa, memastikan bahwa intervensi pembelajaran tidak merugikan peserta didik, melaporkan hasil dengan jujur, dan menghormati hak setiap subjek penelitian. Selain itu, peneliti harus mencermati isu-isu baru seperti bias algoritmik, keamanan data digital, dan keberlanjutan penggunaan teknologi dalam analisis statistik. Pada akhirnya, etika ilmiah menjadi penyeimbang yang memastikan bahwa penelitian kuantitatif tidak hanya menghasilkan temuan yang valid dan reliabel, tetapi juga memberikan manfaat nyata bagi perkembangan ilmu dan kemaslahatan manusia. Tanpa etika, seluruh bangunan metodologi kehilangan legitimasi ilmiahnya.

Daftar Rujukan

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). APA.

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education* (10th ed.). Cengage.

Babbie, E. (2021). *The practice of social research* (15th ed.). Cengage.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2021). *How to design and evaluate research in education* (10th ed.). McGraw-Hill.

Gall, M. D., Gall, J. P., & Borg, W. R. (2019). *Educational research: An introduction* (12th ed.). Pearson.

Johnson, B., & Christensen, L. (2020). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). SAGE.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Harcourt.

McMillan, J. H., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry* (7th ed.). Pearson.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson.

Popper, K. (2002). *The logic of scientific discovery*. Routledge. *(Karya asli diterbitkan 1959)*

Punch, K. F. (2014). *Introduction to social research: Quantitative and qualitative approaches* (3rd ed.). SAGE.

Trochim, W. M. (2020). *Research methods: The essential knowledge base* (3rd ed.). Cengage.

BAB 2
PERUMUSAN MASALAH, VARIABEL, DAN HIPOTESIS

2.1 Karakteristik Masalah Penelitian

Masalah penelitian merupakan titik awal sekaligus fondasi konseptual bagi seluruh proses penelitian kuantitatif. Kerlinger (2000) menegaskan bahwa masalah penelitian adalah "a discrepancy between what is known and what needs to be known," sehingga penelitian selalu dimulai dari adanya kesenjangan pengetahuan yang perlu dijelaskan melalui prosedur ilmiah. Dalam konteks pendidikan, sosial, dan perilaku, masalah penelitian tidak hanya muncul dari fenomena empiris, tetapi juga dari ketidakkonsistenan teori, kurangnya bukti empiris, kelemahan dalam praktik kebijakan, atau perubahan sosial yang menuntut pemahaman baru. Creswell (2014) menekankan bahwa masalah penelitian kuantitatif harus dapat dijelaskan melalui hubungan antarvariabel, karena orientasi penelitian kuantitatif adalah menguji prediksi atau estimasi pengaruh berdasarkan teori.

Karakteristik utama masalah penelitian kuantitatif adalah kejelasan, keterukuran, dan relevansi. Menurut Cohen, Manion, dan Morrison (2018), masalah penelitian harus memiliki batasan konseptual yang jelas agar dapat diterjemahkan ke dalam variabel-variabel yang dapat diukur secara empiris. Masalah yang kabur atau bersifat terlalu luas akan menyulitkan proses identifikasi variabel dan perumusan hipotesis. Gall, Gall, dan Borg (2019) menyatakan bahwa masalah penelitian yang baik harus

memiliki *researchability*, yaitu dapat dijawab melalui metode ilmiah, bukan melalui opini atau spekulasi. Ini berarti masalah harus dapat diterjemahkan ke dalam pengukuran yang objektif, prosedur pengumpulan data yang sistematis, dan analisis statistik yang relevan.

Neuman (2014) menambahkan bahwa masalah penelitian kuantitatif harus berbasis teori, karena teori menyediakan struktur logis yang memungkinkan peneliti merumuskan hubungan antarvariabel secara jelas. Tanpa fondasi teoretis, masalah penelitian tidak memiliki arah, tidak dapat diuji secara deduktif, dan tidak dapat menghasilkan kontribusi ilmiah yang berarti. Sekaran dan Bougie (2020) menggarisbawahi pentingnya menyusun masalah penelitian berdasarkan *symptom versus problem distinction*. Dalam banyak kasus, gejala empiris seperti rendahnya prestasi belajar, tingginya turnover pegawai, atau rendahnya motivasi tidak boleh langsung dipahami sebagai masalah utama. Peneliti harus mengidentifikasi akar masalah melalui kajian literatur dan analisis teoritik agar penelitian tidak terjebak pada diagnosis yang keliru.

Dalam tradisi kuantitatif, masalah penelitian umumnya diarahkan untuk menguji hubungan antarvariabel, baik berbentuk pengaruh (causal-effect), asosiasi (correlation), maupun prediksi (regression). Trochim dan Donnelly (2006) menyebut bahwa suatu masalah penelitian yang baik harus memungkinkan peneliti mengidentifikasi variabel independen, variabel dependen, serta variabel mediator atau moderator yang relevan berdasarkan kajian teori. Persoalan metodologis ini sangat penting karena salah identifikasi variabel dapat menghasilkan kesimpulan statistik yang bias atau keliru. Field (2018) memperingatkan bahwa penelitian kuantitatif yang tidak mendasarkan masalahnya pada struktur variabel yang benar

berisiko tinggi menghasilkan *model misspecification*, yang berdampak langsung pada validitas inferensi.

Relevansi masalah penelitian merupakan karakteristik penting lainnya. Menurut McMillan dan Schumacher (2014), masalah penelitian pendidikan harus memberikan kontribusi terhadap pemahaman, kebijakan, atau praktik pembelajaran. Dalam konteks evidence-based education, masalah penelitian harus terkait langsung dengan kebutuhan praktis seperti efektivitas metode pembelajaran, faktor penentu prestasi belajar, kualitas proses asesmen, atau dinamika psikologis peserta didik. Masalah penelitian yang baik bukan hanya menarik dari sudut akademik, tetapi juga memiliki manfaat praktis yang dapat diimplementasikan dalam pengambilan keputusan berbasis data (data-driven decision making).

Perkembangan teknologi digital dan big data mengubah secara signifikan dinamika perumusan masalah penelitian. Data yang sangat besar, cepat terakumulasi, dan beragam dapat menimbulkan ilusi bahwa masalah penelitian dapat ditemukan hanya dari eksplorasi data. Hair, Black, Babin, dan Anderson (2019) memperingatkan bahwa meskipun big data membuka peluang analisis kompleks, masalah penelitian tetap harus dirumuskan berdasarkan teori. Jika tidak, penelitian akan terjebak pada korelasi acak, pola kebetulan, atau *spurious relationships* yang tidak memiliki makna teoretis. Kline (2016) menegaskan bahwa model konseptual yang baik harus mendahului analisis statistik, bukan sebaliknya. Hal ini menunjukkan bahwa teori tetap menjadi basis utama dalam merumuskan masalah, meskipun teknologi analisis semakin canggih.

Karakteristik penting lainnya adalah *feasibility* atau keterlaksanaan. Sekaran (2020) menyebut bahwa masalah

penelitian harus mempertimbangkan ketersediaan sumber daya, kemampuan peneliti mengakses data, waktu, serta keterbatasan etis. Masalah yang terlalu ambisius atau tidak dapat diukur akan menghasilkan penelitian yang tidak dapat diselesaikan atau tidak sah secara empiris. Sugiyono (2017) bahkan menekankan bahwa peneliti pemula sering gagal bukan karena tidak mampu mengolah data, tetapi karena merumuskan masalah yang tidak feasible.

Dalam konteks analisis statistik, masalah penelitian harus memungkinkan identifikasi variabel yang dapat dioperasionalkan. Pedhazur (1997) mengingatkan bahwa analisis regresi, korelasi, atau SEM hanya akan bermakna jika masalah penelitian menetapkan hubungan antarvariabel secara jelas. Karena itu, karakteristik masalah penelitian yang baik harus kompatibel dengan analisis statistik yang direncanakan. Miles dan Shevlin (2001) menegaskan bahwa kesalahan dalam merumuskan masalah dapat berujung pada kesalahan model, kesalahan pengujian hipotesis, dan interpretasi yang tidak valid.

Dalam penelitian pendidikan, masalah penelitian sering muncul dari disparitas hasil belajar, ketidakselarasan kurikulum, perubahan perilaku siswa, penggunaan teknologi, maupun faktor psikologis seperti motivasi, kecemasan, atau self-regulated learning. Namun, tidak semua fenomena dapat langsung dijadikan masalah penelitian. Creswell (2014) menganjurkan penggunaan *problem mapping*, yaitu menyusun peta isu dari fenomena hingga akar masalah berdasarkan literatur. Pendekatan ini memastikan bahwa masalah yang dirumuskan bukan hanya gejala, tetapi benar-benar representasi dari gap keilmuan.

Secara filosofis, masalah penelitian adalah pintu masuk ke dalam logika ilmiah. Jika masalah dirumuskan secara tepat, proses identifikasi variabel, operasionalisasi konsep, dan

perumusan hipotesis dapat berjalan dengan konsisten. Sebaliknya, kesalahan pada tahap ini akan menghasilkan efek domino pada seluruh tahapan penelitian. Menurut Trochim (2006), masalah penelitian yang salah arah akan menghasilkan hipotesis yang salah, variabel yang tidak relevan, instrumen yang tidak valid, dan analisis statistik yang tidak bermakna.

Refleksi metodologis terhadap karakteristik masalah penelitian menegaskan bahwa ketepatan perumusan masalah adalah fondasi utama validitas penelitian kuantitatif. Masalah yang kabur menghasilkan variabel yang kabur; variabel yang kabur menghasilkan hipotesis yang tidak logis; dan hipotesis yang tidak logis menghasilkan analisis yang tidak dapat dipertanggungjawabkan. Dalam era big data, tantangan semakin besar karena peneliti memiliki akses pada data yang melimpah, tetapi tanpa kejelasan masalah, data tersebut hanya menjadi deretan angka tanpa nilai ilmiah. Oleh karena itu, masalah penelitian harus dirumuskan secara teoretis, empiris, feasible, dan relevan agar dapat menghasilkan pengetahuan yang valid, reliabel, dan berguna bagi perkembangan pendidikan maupun praktik profesional.

2.2 Identifikasi Variabel

Identifikasi variabel merupakan tahap kritis dalam penelitian kuantitatif karena seluruh proses analisis—mulai dari perumusan hipotesis, desain instrumen, pemilihan teknik statistik, hingga interpretasi temuan—bergantung pada ketepatan peneliti dalam menentukan variabel yang relevan. Kerlinger (2000) menyatakan bahwa variabel adalah “constructs that take on different values,” sehingga suatu fenomena hanya dapat diteliti secara kuantitatif apabila peneliti mampu mengidentifikasi aspek mana dari fenomena tersebut yang

berubah, dapat diukur, dan memiliki dasar teoretis untuk dijelaskan hubungannya dengan variabel lain. Menurut Creswell (2014), variabel harus muncul dari kajian teori, bukan sekadar intuisi, karena teori memberikan landasan logis untuk menghubungkan variabel independen dengan variabel dependen, termasuk peran variabel moderator dan mediator dalam memperkuat atau menjembatani pengaruh.

Dalam penelitian pendidikan dan perilaku, variabel biasanya berkaitan dengan konstruk-konstruk abstrak seperti motivasi, minat, kecemasan akademik, prestasi belajar, kepuasan, atau efektivitas pembelajaran. Cohen, Manion, dan Morrison (2018) menekankan bahwa identifikasi variabel membutuhkan pemahaman mendalam tentang konsep yang diteliti agar tidak terjadi kesalahan kategorisasi, misalnya menganggap variabel mediator sebagai moderator atau sebaliknya. Kesalahan identifikasi ini berdampak serius terhadap model konseptual, teknik analisis statistik, dan interpretasi hasil. Hair, Black, Babin, dan Anderson (2019) menjelaskan bahwa dalam analisis multivariat modern, jenis variabel menentukan alat analisis yang digunakan, seperti regresi, ANOVA, SEM, atau analisis jalur. Karena itu, peneliti wajib memahami struktur hubungan antarvariabel sebelum menentukan teknik statistiknya.

Menurut Gall, Gall, dan Borg (2019), proses identifikasi variabel harus dimulai dengan membaca literatur untuk mengidentifikasi konstruk utama dan faktor-faktor yang berhubungan. Setelah itu, peneliti harus menganalisis apakah variabel tersebut berfungsi sebagai variabel bebas (independent variable), variabel terikat (dependent variable), variabel moderator yang memengaruhi kekuatan hubungan, atau variabel mediator yang menjelaskan mekanisme hubungan. Trochim dan Donnelly (2006) menekankan bahwa model ilmiah yang baik

harus menjelaskan alasan teoretis mengapa satu variabel diperlakukan sebagai penyebab, sedangkan variabel lain diperlakukan sebagai akibat. Tanpa justifikasi teoretis, identifikasi variabel bersifat spekulatif dan tidak dapat diuji secara deduktif.

Sekaran dan Bougie (2020) menambahkan bahwa variabel harus diidentifikasi berdasarkan *theoretical reasoning* dan *empirical evidence* yang jelas. Dalam penelitian bisnis dan sosial, peneliti cenderung memiliki banyak variabel potensial, tetapi tidak semuanya relevan. Pemilihan variabel harus mempertimbangkan relevansi teoritis, kontribusi ilmiah, kemudahan pengukuran, serta keterhubungan logis antarvariabel. Neuman (2014) memperingatkan bahwa peneliti sering terjebak dalam kesalahan *conceptual overload* ketika memasukkan terlalu banyak variabel hanya karena ditemukan dalam literatur, padahal tidak semua variabel tersebut memiliki hubungan kuat dengan masalah penelitian utama.

Identifikasi variabel juga memerlukan pemahaman tentang level pengukuran (nominal, ordinal, interval, dan rasio). Field (2018) menegaskan bahwa tingkat pengukuran menentukan jenis statistik yang dapat digunakan. Misalnya, variabel nominal seperti jenis kelamin atau kategori sekolah tidak dapat dianalisis menggunakan statistik parametrik tanpa prosedur transformasi tertentu. Dalam penelitian dengan big data, Hair et al. (2019) menekankan bahwa variabel dapat berasal dari data digital seperti log aktivitas, klik sistem, waktu akses, atau pola interaksi. Namun, variabel digital tersebut tetap membutuhkan definisi teoretis agar memiliki makna ilmiah dan tidak menjadi sekadar angka tanpa konteks.

Dalam penelitian pendidikan, identifikasi variabel sangat penting untuk memastikan kesesuaian antara tujuan penelitian dan model analisis. Misalnya, penelitian tentang efektivitas

model pembelajaran akan memasukkan variabel independen seperti metode pembelajaran, variabel dependen seperti hasil belajar, dan variabel moderator seperti gaya belajar atau motivasi siswa. Jika hubungan tersebut tidak diidentifikasi dengan benar, hipotesis akan menjadi lemah dan analisis statistik tidak dapat memberikan jawaban yang relevan.

Pada tingkat epistemologis, identifikasi variabel merupakan proses menghubungkan dunia abstrak teoritis dengan dunia empiris yang dapat diukur. Tanpa akurasi dalam menentukan variabel, penelitian kehilangan fondasi ilmiahnya. Kesalahan identifikasi variabel akan berdampak domino terhadap kesalahan operasionalisasi, instrumen yang tidak tepat, analisis yang tidak valid, dan kesimpulan yang menyesatkan. Oleh karena itu, identifikasi variabel harus dilakukan secara sistematis, berbasis teori, dan disesuaikan dengan tuntutan evidence-based research serta analisis statistik modern.

2.3 Konseptualisasi dan Operasionalisasi Variabel

Konseptualisasi dan operasionalisasi variabel merupakan tahap fundamental dalam penelitian kuantitatif karena menentukan sejauh mana konstruk teoretis dapat diterjemahkan ke dalam indikator empiris yang dapat diukur secara akurat. Kerlinger (2000) menyatakan bahwa konsep merupakan abstraksi dari fenomena, sedangkan variabel merupakan bentuk terukur dari konsep tersebut. Dengan demikian, konseptualisasi adalah proses mendefinisikan makna teoretis suatu konstruk, sementara operasionalisasi adalah proses memecah konstruk tersebut menjadi indikator-indikator terukur. Creswell (2014) menegaskan bahwa tanpa konseptualisasi yang jelas, operasionalisasi tidak akan memiliki dasar ilmiah, dan instrumen yang dihasilkan akan kehilangan validitas. Dalam penelitian pendidikan dan sosial,

konseptualisasi sangat penting karena banyak konstruk seperti motivasi, kepuasan, kecemasan belajar, resiliensi, atau self-regulated learning tidak dapat diobservasi langsung, melainkan direpresentasikan melalui perilaku, persepsi, atau respons yang dapat diukur.

Menurut Cohen, Manion, dan Morrison (2018), konseptualisasi harus diawali dengan kajian literatur yang mendalam untuk memahami berbagai definisi konstruk, model teoretisnya, serta dimensi-dimensi yang membangunnya. Misalnya, motivasi belajar pada siswa dapat dipandang dari perspektif behavioristik, kognitif, atau humanistik, yang masing-masing menghasilkan indikator penelitian yang berbeda. Neuman (2014) menekankan bahwa kesalahan konseptualisasi terjadi ketika peneliti memilih definisi konstruk secara sembarangan atau memadukan indikator dari teori yang tidak kompatibel. Karena itu, peneliti harus memilih pendekatan teoretis yang paling relevan untuk menjelaskan fenomena yang diteliti dan memastikan bahwa konsep tersebut bersifat konsisten secara epistemologis.

Setelah definisi teoretis ditetapkan, operasionalisasi variabel dilakukan dengan merumuskan indikator-indikator terukur. Sekaran dan Bougie (2020) menyatakan bahwa operasionalisasi bertujuan menerjemahkan konstruk abstrak ke dalam item yang dapat diamati, diklasifikasikan, atau diukur melalui skala. Proses ini harus mempertimbangkan tiga aspek: kejelasan indikator, kesesuaian indikator dengan konsep, dan kelayakan pengukuran. Misalnya, konstruk “kepuasan belajar” dapat dioperasionalkan melalui indikator seperti persepsi terhadap metode pembelajaran, kualitas interaksi dengan dosen, kejelasan materi, dan kenyamanan lingkungan belajar. DeVellis (2017) menekankan bahwa indikator harus disusun berdasarkan

prinsip pengembangan skala agar reliabilitas dan validitasnya dapat dijamin, terutama dalam penelitian yang menggunakan skala Likert atau instrumen psikometrik.

Hair, Black, Babin, dan Anderson (2019) menjelaskan bahwa operasionalisasi variabel memiliki implikasi langsung terhadap teknik analisis statistik. Dalam analisis multivariat seperti SEM, indikator harus mencerminkan konstruk secara reflektif atau formatif sesuai model teoretisnya. Kesalahan dalam operasionalisasi dapat menyebabkan model statistik tidak fit atau menghasilkan interpretasi yang keliru. Field (2018) menambahkan bahwa kualitas analisis regresi atau korelasi sangat bergantung pada keakuratan indikator, karena indikator yang tidak tepat akan menghasilkan error measurement yang tinggi, sehingga mengurangi validitas inferensi statistik.

Dalam konteks penelitian berbasis big data, operasionalisasi variabel menghadapi tantangan baru. Data digital seperti clickstream, waktu akses, pola log pembelajaran, atau aktivitas media sosial sering kali sudah tersedia dalam jumlah besar, tetapi tidak selalu mencerminkan konstruk teoretis. Trochim dan Donnelly (2006) mengingatkan bahwa data kuantitatif tidak otomatis mewakili konsep; peneliti harus tetap menafsirkan data digital melalui kerangka teori. Misalnya, tingginya frekuensi akses LMS tidak otomatis menjadi indikator motivasi belajar kecuali terdapat kerangka teoretis yang menghubungkannya dengan perilaku belajar. Oleh karena itu, integrasi teori tetap menjadi kunci utama operasionalisasi, bahkan ketika peneliti menggunakan sumber data otomatis dan algoritmik.

Operasionalisasi juga harus mempertimbangkan level pengukuran variabel, mulai dari nominal, ordinal, interval, hingga rasio. Pemilihan level pengukuran memengaruhi jenis analisis

statistik yang dapat digunakan. Misalnya, variabel nominal seperti jenis kelamin hanya dapat dianalisis melalui statistik non-parametrik, sedangkan variabel interval dapat dianalisis melalui ANOVA atau regresi linear. Menurut McMillan dan Schumacher (2014), peneliti harus berhati-hati dalam mengasumsikan bahwa skala Likert bersifat interval, padahal secara teknis bersifat ordinal. Meskipun dalam praktik penelitian sosial sering diperlakukan sebagai data interval untuk memudahkan analisis, peneliti tetap harus memahami implikasi metodologisnya.

Kesalahan konseptualisasi dan operasionalisasi dapat menghasilkan instrumen yang tidak valid, model statistik yang tidak tepat, dan kesimpulan yang menyesatkan. Sebagai contoh, jika peneliti mengukur "kecemasan belajar" hanya melalui indikator perilaku eksternal tanpa mempertimbangkan aspek kognitif dan afektif, maka konstruk tersebut menjadi tidak lengkap dan tidak mewakili realitas teoretis yang seharusnya. Demikian pula, jika peneliti mengoperasionalkan variabel mediator atau moderator tanpa pemahaman teoretis yang kuat, analisis statistik seperti regresi moderasi atau mediasi akan memberikan hasil yang tidak bermakna.

Secara metodologis, konseptualisasi dan operasionalisasi variabel merupakan jembatan yang menghubungkan teori dan data. Ketepatan dalam tahap ini menentukan kualitas seluruh proses penelitian. Refleksi metodologis menunjukkan bahwa ketika konstruk teoretis didefinisikan secara akurat dan dioperasionalkan dengan indikator yang relevan, penelitian kuantitatif mampu menghasilkan bukti yang kuat dan dapat diandalkan untuk pengambilan keputusan berbasis data. Sebaliknya, kesalahan pada tahap ini menyebabkan hilangnya validitas penelitian, meskipun analisis statistik dilakukan dengan teknik yang canggih. Oleh karena itu, peneliti harus memastikan

bahwa konseptualisasi bersifat teoretis, operasionalisasi bersifat empiris, dan keduanya selaras sehingga penelitian dapat memberikan kontribusi ilmiah yang signifikan.

2.4 Rumusan Hipotesis

Perumusan hipotesis merupakan tahap sentral dalam penelitian kuantitatif karena hipotesis berfungsi sebagai jembatan logis antara teori dan data empiris. Kerlinger (2000) menyatakan bahwa hipotesis adalah "a conjectural statement of the relation between two or more variables," sehingga hakikat hipotesis bukan sekadar dugaan, tetapi pernyataan yang berakar pada teori dan harus dapat diuji melalui prosedur statistik. Creswell (2014) menambahkan bahwa hipotesis kuantitatif harus terumus secara spesifik, mencerminkan prediksi teoretis, serta menyatakan hubungan antarvariabel secara jelas, apakah hubungan tersebut bersifat linier, positif, negatif, langsung, tidak langsung, moderatif, atau mediatif. Dalam penelitian pendidikan dan sosial, hipotesis membantu peneliti memfokuskan analisis pada hubungan yang paling penting, memastikan bahwa pengumpulan data dan pengukuran variabel dilakukan secara sistematis dan sesuai dengan desain penelitian.

Menurut Cohen, Manion, dan Morrison (2018), kekuatan hipotesis terletak pada kemampuannya merangkum model teoretis ke dalam bentuk pernyataan yang dapat diuji secara empiris. Oleh karena itu, hipotesis tidak boleh dirumuskan tanpa dasar teori yang kuat. Neuman (2014) mengingatkan bahwa hipotesis yang baik harus didasarkan pada literatur yang relevan dan hasil penelitian terdahulu, sehingga prediksi yang dibuat merupakan refleksi dari akumulasi pengetahuan ilmiah, bukan opini subjektif. Sekaran dan Bougie (2020) bahkan menekankan prinsip *theoretical justification*, yakni bahwa setiap hipotesis

harus menunjukkan alasan teoretis mengapa variabel independen diprediksi memengaruhi variabel dependen, serta apakah terdapat mediator atau moderator yang memperkuat atau menjelaskan hubungan tersebut.

Hipotesis dalam penelitian kuantitatif umumnya diklasifikasikan ke dalam beberapa jenis. Pertama, hipotesis deskriptif yang memprediksi nilai atau karakteristik tertentu dari suatu variabel, meskipun jenis ini jarang digunakan karena penelitian kuantitatif lebih berorientasi pada hubungan antarvariabel. Kedua, hipotesis komparatif yang memprediksi perbedaan antara dua kelompok atau lebih, seperti perbedaan hasil belajar antara model pembelajaran A dan B. Ketiga, hipotesis asosiatif atau kausal yang memprediksi hubungan antarvariabel, seperti pengaruh motivasi terhadap prestasi belajar. Field (2018) menjelaskan bahwa jenis hipotesis yang dipilih akan menentukan jenis analisis statistik yang digunakan, misalnya analisis regresi, ANOVA, korelasi, uji-t, atau SEM. Karena itu, peneliti harus merumuskan hipotesis sesuai dengan struktur variabel dan model analisis yang direncanakan.

Hair, Black, Babin, dan Anderson (2019) menekankan bahwa hipotesis modern juga harus mempertimbangkan kompleksitas hubungan antarvariabel, terutama dalam penelitian multivariat. Misalnya, hubungan antara literasi digital dan prestasi akademik dapat dimediasi oleh self-regulated learning, atau hubungan antara metode pembelajaran dan keterlibatan siswa dapat dimoderasi oleh gaya belajar. Dalam konteks SEM, hipotesis harus menggambarkan arah pengaruh (direct effect), pengaruh tidak langsung (indirect effect), dan pengaruh total (total effect). Kline (2016) menyatakan bahwa hipotesis yang tidak mempertimbangkan struktur teoretis seperti mediator atau

moderator berpotensi menghasilkan model penelitian yang miskin atau gagal menjelaskan fenomena secara holistik.

Trochim dan Donnelly (2006) menegaskan bahwa hipotesis harus memenuhi tiga kriteria utama: dapat diuji (*testable*), dapat dibantah (*falsifiable*), dan dapat diobservasi (*empirically grounded*). Hal ini sejalan dengan pemikiran Popper bahwa klaim ilmiah harus dapat diuji dan berpotensi ditolak. Hipotesis yang terlalu abstrak, tidak terukur, atau tidak memiliki indikator empiris yang jelas tidak memenuhi syarat sebagai hipotesis ilmiah. Misalnya, pernyataan "motivasi memengaruhi keberhasilan hidup" tidak dapat diuji dalam konteks penelitian kuantitatif tanpa definisi operasional yang spesifik.

Miles dan Shevlin (2001) menekankan bahwa dalam analisis regresi dan korelasi, hipotesis harus menentukan arah hubungan (directionality). Hipotesis nondireksional—misalnya "terdapat hubungan antara X dan Y"—kurang informatif dan tidak mencerminkan prediksi teoretis yang kuat. Sebaliknya, hipotesis direksional seperti "motivasi belajar berpengaruh positif terhadap prestasi belajar" menunjukkan prediksi yang lebih jelas dan dapat diuji secara statistik.

Dalam era big data, perumusan hipotesis menghadapi tantangan baru. Data yang sangat besar dapat menggoda peneliti untuk mengabaikan teori dan menggantinya dengan eksplorasi data semata. Pedhazur (1997) memperingatkan bahwa analisis data tanpa hipotesis teoretis hanya menghasilkan korelasi dangkal yang tidak memiliki makna ilmiah. Bahkan, algoritma machine learning dapat menemukan ribuan pola yang tidak memiliki dasar teoritis sehingga menimbulkan risiko *spurious findings*. Karena itu, meskipun teknologi analitik semakin canggih, hipotesis tetap diperlukan sebagai panduan intelektual

untuk menjaga penelitian tetap berada dalam kerangka teoretis yang konsisten.

Kesalahan dalam merumuskan hipotesis dapat berdampak serius pada seluruh penelitian. Hipotesis yang terlalu luas akan menghasilkan variabel yang kabur; hipotesis yang tidak didukung teori akan menghasilkan analisis yang tidak valid; hipotesis yang tidak dapat diuji akan menghasilkan kesimpulan yang tidak bermakna. McMillan dan Schumacher (2014) mengingatkan bahwa banyak penelitian pendidikan gagal bukan karena kesalahan statistik, tetapi karena hipotesis tidak dirumuskan berdasarkan analisis teoretis yang memadai. Oleh karena itu, perumusan hipotesis harus dipandang sebagai proses intelektual yang membutuhkan ketelitian, argumentasi, dan integrasi teori.

Secara metodologis, hipotesis adalah fondasi pengujian ilmiah. Ketepatan hipotesis menentukan kualitas desain penelitian, instrumen, model analisis, dan kesimpulan. Dalam konteks evidence-based research, hipotesis yang baik memungkinkan penelitian memberikan kontribusi nyata bagi pemecahan masalah pendidikan. Refleksi metodologis menunjukkan bahwa hipotesis bukan sekadar pernyataan formal, tetapi inti dari logika deduktif yang memastikan bahwa penelitian bergerak dari teori menuju bukti dengan cara yang sistematis, terukur, dan konsisten.

2.5 Model Konseptual

Model konseptual merupakan representasi visual dan logis dari hubungan antarvariabel yang telah diidentifikasi dan didefinisikan dalam penelitian. Model ini berfungsi sebagai peta teoretis yang menunjukkan bagaimana variabel independen, dependen, moderator, mediator, dan kontrol saling berhubungan dalam suatu kerangka penelitian. Kerlinger (2000) menekankan

bahwa model konseptual adalah jantung dari proses deduksi ilmiah karena ia merangkum asumsi teoretis ke dalam struktur yang memungkinkan diuji secara empiris. Dengan demikian, model konseptual tidak hanya menggambarkan hubungan antarvariabel, tetapi juga mengekspresikan mekanisme teoretis yang menghubungkan fenomena yang diteliti dengan prediksi hipotesis.

Menurut Creswell (2014), model konseptual harus dibangun berdasarkan teori yang telah ada dan temuan penelitian terdahulu. Hal ini memastikan bahwa hubungan antarvariabel tidak bersifat spekulatif, tetapi didukung oleh landasan ilmiah yang kuat. Cohen, Manion, dan Morrison (2018) menegaskan bahwa model konseptual membantu peneliti mengorganisasi pemikiran, mengurangi ambiguitas konseptual, serta memberikan arah metodologis pada langkah operasionalisasi, desain instrumen, dan teknik analisis statistik. Tanpa model konseptual yang jelas, penelitian berisiko menghasilkan data yang tidak relevan, rancu, atau tidak dapat dianalisis secara sistematis.

Neuman (2014) menjelaskan bahwa membangun model konseptual melibatkan proses identifikasi hubungan logis antarvariabel berdasarkan teori, termasuk mengklarifikasi apakah hubungan tersebut bersifat kausal, korelasional, langsung, tidak langsung, atau dipengaruhi oleh faktor moderasi. Hair, Black, Babin, dan Anderson (2019) menambahkan bahwa model konseptual harus mencerminkan struktur hubungan yang dapat diuji melalui analisis multivariat, seperti regresi, path analysis, atau structural equation modeling (SEM). Dalam konteks SEM, Kline (2016) menegaskan bahwa model konseptual menjadi blueprint analisis karena menentukan bagaimana konstruk laten diukur dan bagaimana hubungan antarvariabel diuji melalui

koefisien lintasan (path coefficients). Jika model konseptual tidak dirumuskan secara tepat, analisis statistik akan gagal menunjukkan hubungan teoretis yang sesungguhnya.

Dalam penelitian pendidikan, model konseptual digunakan untuk menjelaskan dinamika pembelajaran, perilaku siswa, efektivitas metode mengajar, serta hubungan antara faktor psikologis dan prestasi akademik. Misalnya, pada studi mengenai pengaruh literasi digital terhadap prestasi belajar dengan mediasi self-regulated learning, model konseptual akan menampilkan literasi digital sebagai variabel independen, prestasi belajar sebagai variabel dependen, dan self-regulated learning sebagai mediator yang menjembatani hubungan tersebut. Dengan demikian, model konseptual memberikan gambaran tentang proses bagaimana literasi digital dapat memengaruhi prestasi melalui kemampuan regulasi diri. Model ini tidak hanya mengarahkan analisis statistik, tetapi juga membantu menjelaskan mekanisme perilaku yang terlibat dalam proses pembelajaran.

Sekaran dan Bougie (2020) menekankan bahwa model konseptual juga berperan dalam pengambilan keputusan praktis karena ia membantu pemangku kebijakan melihat struktur hubungan faktor-faktor kunci yang memengaruhi suatu fenomena. Dalam penelitian organisasi pendidikan, misalnya, model konseptual dapat menunjukkan hubungan antara kompetensi guru, penggunaan teknologi pembelajaran, motivasi siswa, dan hasil belajar. Dengan memahami hubungan ini secara sistematis, sekolah dan lembaga pendidikan dapat merumuskan intervensi yang lebih efektif dan berbasis data.

Dalam konteks era big data, pembangunan model konseptual tetap relevan meskipun peneliti memiliki akses ke volume data yang sangat besar. Trochim dan Donnelly (2006)

mengingatkan bahwa data tanpa teori hanya menghasilkan korelasi yang tidak bermakna. Big data dapat mengungkap pola, tetapi tanpa model konseptual, pola tersebut tidak dapat dipahami secara ilmiah. Bahkan, analisis machine learning yang bersifat eksploratif memerlukan model konseptual untuk menjelaskan mengapa pola terjadi dan bagaimana pola tersebut berhubungan dengan konstruk teoretis. Oleh karena itu, meskipun big data menawarkan kekuatan prediktif, model konseptual tetap diperlukan untuk memberikan landasan interpretasi dan menjamin bahwa temuan analitik tidak bersifat semu (spurious).

Pentingnya model konseptual juga terlihat dalam hubungannya dengan validitas dan reliabilitas. Field (2018) menunjukkan bahwa model konseptual yang buruk mengarah pada pemilihan indikator yang tidak tepat, sehingga mengurangi validitas konstruk dan meningkatkan error measurement. Hal ini berdampak langsung pada ketepatan analisis statistik dan interpretasi hasil. Sebaliknya, model konseptual yang baik memperkuat struktur operasionalisasi variabel dan menjamin bahwa instrumen penelitian mengukur aspek penting dari konstruk yang relevan secara teoretis.

Kesalahan dalam membangun model konseptual memiliki konsekuensi metodologis yang signifikan. Jika hubungan antarvariabel tidak didasarkan pada teori atau tidak logis secara causal ordering, analisis statistik seperti regresi atau SEM akan menghasilkan model yang tidak fit atau koefisien yang tidak stabil. Pedhazur (1997) menegaskan bahwa kesalahan dalam menentukan arah hubungan (directionality) akan menyesatkan proses inferensi sehingga penelitian gagal menjawab pertanyaan penelitian secara sah. Model yang memasukkan variabel moderator sebagai mediator atau sebaliknya juga akan

menghasilkan interpretasi yang keliru dan berpotensi menghasilkan implikasi kebijakan yang tidak tepat.

Secara epistemologis, model konseptual berperan sebagai mekanisme integrasi antara teori dan data. Model ini tidak hanya menunjukkan apa yang ingin diuji, tetapi juga bagaimana teori bekerja ketika diterjemahkan ke dalam bentuk empiris. Dalam konteks evidence-based education, model konseptual memungkinkan peneliti merumuskan intervensi dan rekomendasi kebijakan yang lebih tepat sasaran karena hubungan antarvariabel dijelaskan secara sistematis. Dengan demikian, model konseptual memainkan peran strategis dalam menjamin bahwa penelitian kuantitatif tidak hanya menghasilkan angka, tetapi juga menghasilkan pengetahuan yang bermakna, dapat diuji ulang, dan dapat digunakan untuk meningkatkan kualitas pendidikan.

Refleksi metodologis menunjukkan bahwa model konseptual adalah kerangka intelektual yang membimbing seluruh proses penelitian. Ketika model dikembangkan dengan baik, penelitian berjalan secara terarah dan analisis statistik dapat memberikan jawaban yang valid. Sebaliknya, model yang lemah mengakibatkan penelitian kehilangan fokus, menghasilkan data yang tidak relevan, dan memberikan kesimpulan yang tidak dapat dipertanggungjawabkan. Oleh karena itu, pengembangan model konseptual merupakan tahap esensial yang menuntut ketelitian teoretis, ketepatan logis, dan kejelasan metodologis.

Daftar Rujukan

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

Gall, M. D., Gall, J. P., & Borg, W. R. (2019). *Educational research: An introduction* (12th ed.). Pearson.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Kerlinger, F. N. (2000). *Foundations of behavioral research* (4th ed.). Harcourt.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

McMillan, J. H., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry* (7th ed.). Pearson.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. SAGE.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. Harcourt Brace.

Sekaran, U., & Bougie, R. (2020). *Research methods for business: A skill-building approach* (8th ed.). Wiley.

Sugiyono. (2017). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Trochim, W. M. K., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Atomic Dog.

BAB 3
DESAIN PENELITIAN KUANTITATIF

3.1 Desain Eksperimental

Desain eksperimental merupakan bentuk desain penelitian yang paling kuat dalam menguji hubungan sebab-akibat karena memungkinkan peneliti memanipulasi variabel independen secara langsung dan mengontrol variabel lain yang berpotensi memengaruhi hasil. Campbell dan Stanley (1963) menyatakan bahwa eksperimen sejati (*true experiment*) memiliki tiga karakteristik pokok: adanya manipulasi variabel bebas, pengukuran variabel terikat, dan yang paling penting adalah randomisasi dalam penugasan partisipan. Randomisasi dipandang sebagai mekanisme dasar untuk menghilangkan bias seleksi dan menyamakan karakteristik kelompok secara statistik sehingga setiap perbedaan hasil dapat diatribusikan pada perlakuan. Konsep ini diperkuat oleh Shadish, Cook, dan Campbell (2002) yang menegaskan bahwa desain eksperimental merupakan standar emas (*gold standard*) dalam inferensi kausal, karena tidak hanya menjamin kontrol terhadap ancaman validitas internal tetapi juga menyediakan struktur logis bagi generalisasi temuan apabila desain dilakukan secara ketat.

Creswell (2014) menjelaskan bahwa eksperimen sejati digunakan untuk menguji efektivitas suatu intervensi, misalnya metode pembelajaran, program pelatihan, strategi bimbingan, atau kebijakan pendidikan tertentu. Pada konteks pendidikan dan psikologi, eksperimen biasanya membandingkan dua atau lebih kelompok, misalnya kelompok eksperimen yang menerima

intervensi tertentu dan kelompok kontrol yang tidak menerima intervensi. Johnson dan Christensen (2020) menyebut bahwa eksperimen memungkinkan peneliti menguji *causal directionality*, yaitu apakah perubahan pada variabel X menyebabkan perubahan pada variabel Y, bukan sekadar terkait. Inilah yang membedakan eksperimen dari desain korelasional atau non-eksperimental yang hanya dapat mengidentifikasi hubungan, tetapi tidak dapat memberikan kepastian kausal.

Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa eksperimen sejati menuntut kontrol yang ketat terhadap variabel perancu melalui randomisasi, penyetaraan awal (*matching*), penggunaan instrumen reliabel, dan keseragaman prosedur perlakuan. Di bidang pendidikan, desain pretest-posttest control group (randomized) merupakan salah satu bentuk yang paling banyak digunakan karena memungkinkan peneliti mengevaluasi perbedaan perubahan sebelum dan sesudah perlakuan sekaligus membandingkannya antar kelompok. Gall, Gall, dan Borg (2019) menambahkan bahwa eksperimen juga dapat dilakukan pada tingkat kelas atau sekolah dalam bentuk *cluster randomization*, meskipun hal ini menimbulkan tantangan tambahan dalam mengontrol varians antar kelompok.

Kerlinger (2000) menggarisbawahi bahwa eksperimen bukan hanya manipulasi teknis, tetapi juga implementasi logika ilmiah yang kuat. Ia menekankan bahwa perlakuan harus didasarkan pada teori yang jelas sehingga eksperimen dapat menguji hipotesis deduktif yang bersumber dari model konseptual. Tanpa fondasi teoretis, eksperimen berpotensi menghasilkan data yang tidak bermakna meskipun dilaksanakan secara teknis sempurna. Dalam pendidikan, penting untuk memastikan bahwa intervensi benar-benar berakar pada prinsip

pedagogis atau teori perilaku sehingga hasil penelitian dapat diinterpretasikan dengan tepat.

Namun, eksperimen juga tidak bebas dari ancaman validitas internal dan eksternal. Cohen, Manion, dan Morrison (2018) mengingatkan bahwa meskipun randomisasi mengurangi bias seleksi, ancaman seperti *history*, *maturation*, *testing effect*, *instrumentation*, dan *statistical regression* masih dapat terjadi. Oleh karena itu, desain eksperimen harus mencakup strategi mitigasi, misalnya penggunaan instrumen yang stabil, kontrol terhadap kondisi lingkungan, dan penjadwalan yang konsisten. Shaughnessy, Zechmeister, dan Zechmeister (2015) menekankan pentingnya "standardization of procedure" untuk memastikan bahwa perbedaan hasil antar kelompok benar-benar berasal dari perlakuan, bukan perbedaan perlakuan yang tidak konsisten.

Dalam perkembangan penelitian kontemporer, eksperimen menghadapi tantangan baru, terutama dalam konteks digital dan big data. Dengan sistem pembelajaran daring, platform e-learning, dan aplikasi edukasi digital, eksperimen dapat dilakukan secara otomatis melalui *A/B testing*, yaitu memberikan dua versi perlakuan berbeda kepada kelompok pengguna. Field (2018) mengungkapkan bahwa pendekatan ini memperluas peluang eksperimen karena data dapat dikumpulkan secara cepat, real-time, dan dalam jumlah besar. Namun, eksperimen digital tidak selalu memenuhi syarat desain eksperimental tradisional karena randomisasi pengguna platform seringkali tidak sepenuhnya terkendali. Selain itu, variabel perancu seperti preferensi pengguna atau latar belakang digital literasi sulit dikontrol secara sempurna. Maka dari itu, peneliti modern harus menggabungkan prinsip eksperimental klasik dengan teknik analitik big data seperti propensity score matching atau algoritma machine learning untuk memperkuat inferensi kausal.

Maxwell dan Delaney (2004) menekankan bahwa analisis statistik berperan penting dalam eksperimen karena membantu menilai efek perlakuan dengan lebih presisi. Dalam eksperimen multikelompok atau multilevel, analisis seperti ANOVA, ANCOVA, regresi hierarkis, dan *linear mixed models* sering digunakan untuk mengontrol variabel pengganggu dan meningkatkan sensitivitas deteksi efek. Pedhazur (1997) mengingatkan bahwa penggunaan kovariat harus didasarkan pada teori, bukan sekadar menambah variabel secara sembarangan untuk meningkatkan signifikansi statistik. Oleh karena itu, eksperimen yang baik harus dibangun atas fondasi teoretis, desain metodologis yang ketat, serta analisis statistik yang tepat.

Di bidang pendidikan, eksperimen memiliki keterbatasan praktis dan etis. Tidak semua intervensi dapat diberikan secara acak karena sekolah sering memiliki kebijakan tertentu atau pertimbangan etis tentang pemerataan perlakuan. Tuckman dan Harper (2012) menyatakan bahwa eksperimen dalam situasi alami sering menghadapi resistensi administratif atau keterbatasan sumber daya. Meski demikian, keberhasilan eksperimen dalam pendidikan bergantung pada kolaborasi peneliti dengan guru, sekolah, dan pemangku kepentingan agar perlakuan dapat diterapkan secara terstandar.

Secara epistemologis, desain eksperimental memperkuat kemampuan penelitian untuk menghasilkan inferensi kausal yang kuat. Ketika eksperimen dilakukan dengan kontrol yang memadai, peneliti dapat menyimpulkan bahwa perubahan pada variabel dependen disebabkan oleh perlakuan, bukan faktor lain. Namun refleksi metodologis menunjukkan bahwa kekuatan desain eksperimental bukan tanpa batas. Desain yang buruk, randomisasi yang gagal, atau perlakuan yang tidak konsisten dapat menghilangkan kekuatan kausalitas, meskipun analisis

statistik dilakukan dengan benar. Oleh karena itu, eksperimen harus dipahami tidak hanya sebagai prosedur teknis tetapi juga sebagai praktik ilmiah yang membutuhkan ketelitian desain, integritas implementasi, dan ketepatan interpretasi.

3.2 Quasi-Eksperimental

Desain quasi-eksperimental muncul sebagai alternatif ketika peneliti tidak dapat melakukan randomisasi penuh terhadap partisipan atau kelompok, sebuah kondisi yang sangat umum dalam penelitian pendidikan dan sosial. Campbell dan Stanley (1963) memperkenalkan istilah ini untuk menggambarkan desain yang tetap memanipulasi variabel bebas tetapi tidak mampu mengontrol seluruh ancaman validitas internal sebagaimana true experiment. Dalam konteks sekolah, kelas, atau institusi pendidikan, peneliti sering tidak dapat memindahkan siswa secara acak ke kelompok perlakuan dan kontrol karena adanya batasan administratif, etis, atau praktis. Oleh karena itu, quasi-eksperimen tetap mempertahankan prinsip inti eksperimen—yaitu pemberian perlakuan—namun tanpa random assignment, sehingga peneliti harus menggunakan strategi lain untuk meminimalkan bias.

Shadish, Cook, dan Campbell (2002) menegaskan bahwa kekuatan quasi-eksperiment bukan terletak pada randomisasi, tetapi pada desain alternatif dan kontrol statistik yang mampu memperkuat inferensi kausal. Salah satu bentuk yang paling umum adalah desain *non-equivalent control group*, di mana kelompok eksperimen dan kontrol sudah terbentuk secara alami sebelum penelitian dimulai. Creswell (2014) menjelaskan bahwa desain ini tetap memungkinkan peneliti membandingkan efek perlakuan, meskipun adanya perbedaan karakteristik awal antar kelompok harus diperhitungkan dan dikontrol. Penelitian

pendidikan sering menggunakan desain ini, misalnya ketika dua kelas berbeda dibandingkan dalam efektivitas suatu model pembelajaran.

Fraenkel, Wallen, dan Hyun (2021) menyatakan bahwa kelemahan utama quasi-eksperimen adalah ancaman validitas internal, khususnya *selection bias*. Tanpa randomisasi, tidak ada jaminan bahwa kelompok eksperimen dan kontrol sama sebelum perlakuan. Untuk mengatasi hal ini, peneliti dapat menggunakan *pretest* sebagai alat untuk membandingkan kondisi awal, serta melakukan *matching* berdasarkan karakteristik tertentu seperti nilai awal, kemampuan kognitif, atau variabel demografis. Johnson dan Christensen (2020) juga merekomendasikan penggunaan statistik kovariat seperti ANCOVA untuk mengontrol perbedaan awal tersebut, sehingga estimasi dampak perlakuan menjadi lebih akurat.

Cohen, Manion, dan Morrison (2018) menyoroti bahwa quasi-eksperimen sangat berguna dalam situasi naturalistik di mana eksperimen penuh tidak mungkin dilakukan. Dalam pendidikan, misalnya, jika sekolah ingin menguji program intervensi literasi tetapi tidak dapat merandom siswa, peneliti tetap dapat merancang quasi-eksperimen yang valid dengan mengontrol variabel perancu melalui desain berulang seperti *time-series design*. Desain ini memungkinkan peneliti mengukur variabel dependen pada banyak titik waktu sebelum dan sesudah perlakuan, sehingga perubahan dapat diamati secara lebih jelas. Gall, Gall, dan Borg (2019) menganggap desain *interrupted time-series* sebagai salah satu bentuk quasi-eksperimen yang paling kuat karena struktur data yang berulang dapat meminimalkan ancaman *history* dan *maturation*.

Dalam era big data dan analitik digital, quasi-eksperimen mendapatkan relevansi baru. Field (2018) menjelaskan bahwa

banyak platform digital, seperti sistem pembelajaran daring, menghasilkan data besar yang dapat dimanfaatkan untuk evaluasi kebijakan atau intervensi tanpa randomisasi formal. Teknik seperti *propensity score matching* (PSM) dan *regression discontinuity design* (RDD) sering digunakan untuk meniru kondisi randomisasi dalam data observasional. Shadish et al. (2002) menganggap pendekatan ini sebagai bagian dari quasi-eksperimen modern yang memungkinkan peneliti mendekati inferensi kausal meskipun tidak memiliki kendali penuh terhadap proses alokasi perlakuan.

Quasi-eksperiment tetap memiliki keterbatasan penting. Tanpa randomisasi, ancaman seperti *attrition*, *differential selection*, dan *instrumentation* lebih sulit dikontrol. Namun, desain yang baik dapat meminimalkan kelemahan tersebut. Tuckman dan Harper (2012) menekankan bahwa dokumentasi proses perlakuan, keseragaman implementasi, dan kontrol lingkungan penelitian merupakan langkah penting untuk mempertahankan integritas kausal. Dalam pendidikan, keberhasilan quasi-eksperimen sangat bergantung pada kerja sama guru, konsistensi perlakuan, dan kejelasan prosedur penelitian.

Secara metodologis, quasi-eksperimen menempati posisi strategis antara eksperimen sejati dan penelitian non-eksperimental. Ia memberikan peluang untuk menguji hubungan sebab-akibat dalam konteks yang realistis, meskipun tidak sekuat true experiment dalam hal kontrol. Refleksi metodologis menunjukkan bahwa quasi-eksperimen memberikan kontribusi penting bagi penelitian pendidikan karena ia memungkinkan pengujian intervensi dalam kondisi alami tanpa melanggar batasan etis atau administratif. Ketika dirancang dengan hati-hati, quasi-eksperimen dapat menghasilkan bukti ilmiah yang kuat dan

relevan, serta memberikan dasar bagi pengambilan keputusan berbasis data dalam praktik pendidikan.

3.3 Non-Eksperimental

Desain non-eksperimental digunakan ketika peneliti tidak dapat melakukan manipulasi variabel bebas ataupun memberikan perlakuan tertentu terhadap partisipan. Desain ini banyak digunakan dalam penelitian pendidikan, sosial, dan psikologi karena peneliti sering kali harus mengamati fenomena sebagaimana adanya tanpa intervensi. Creswell (2014) menyatakan bahwa desain non-eksperimental berfokus pada pengukuran variabel sebagaimana muncul secara alami, sehingga hubungan yang diteliti bersifat korelasional, komparatif pasif, atau prediktif. Cohen, Manion, dan Morrison (2018) menegaskan bahwa desain non-eksperimental sangat penting ketika manipulasi variabel tidak mungkin dilakukan karena alasan etis, praktis, atau logis, misalnya ketika meneliti pengaruh tingkat kecemasan terhadap prestasi belajar atau hubungan antara latar belakang sosial-ekonomi dan kemampuan literasi.

Menurut Johnson dan Christensen (2020), penelitian non-eksperimental mencakup beberapa jenis utama, seperti studi survei, penelitian korelasional, studi *ex post facto*, dan penelitian prediktif. Survei digunakan untuk mengumpulkan informasi dari sampel besar mengenai sikap, persepsi, motivasi, atau pengalaman peserta didik. Dalam pendidikan, survei sering diterapkan untuk memperoleh gambaran umum tentang efektivitas pembelajaran, kepuasan mahasiswa, atau budaya sekolah. Sementara itu, penelitian korelasional bertujuan mengidentifikasi hubungan antarvariabel tanpa menyimpulkan sebab-akibat. Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa meskipun korelasi tidak menunjukkan kausalitas, analisis

korelasional dapat menghasilkan model prediksi yang kuat, terutama jika didukung oleh teori yang jelas.

Gall, Gall, dan Borg (2019) menyebut studi *ex post facto* sebagai bentuk non-eksperimental yang mencoba menemukan hubungan kausal setelah peristiwa terjadi, misalnya meneliti dampak pola asuh atau pengalaman belajar sebelumnya terhadap hasil akademik siswa. Meskipun tidak ada manipulasi, peneliti tetap dapat menilai pola hubungan menggunakan analisis statistik seperti regresi, ANOVA non-eksperimental, atau model mediasi. Namun, Trochim (2006) memperingatkan bahwa desain ini rawan terhadap ancaman validitas internal, terutama *selection bias*, sehingga peneliti harus mengontrol variabel perancu melalui matching, kovariat statistik, atau stratifikasi sampel.

Perkembangan era digital memberikan peluang baru bagi penelitian non-eksperimental. Field (2018) menjelaskan bahwa big data dari Learning Management System (LMS), aplikasi pendidikan, dan platform digital memungkinkan peneliti menganalisis pola perilaku tanpa intervensi langsung. Analisis berbasis log data, *learning analytics*, atau data perilaku online dapat memberikan wawasan prediktif mengenai keterlibatan belajar atau risiko putus studi. Namun, Shaughnessy et al. (2015) mengingatkan bahwa besarnya volume data tidak menjamin validitas; peneliti tetap harus mengacu pada teori ketika menentukan variabel yang akan dianalisis untuk menghindari korelasi semu.

Keterbatasan utama penelitian non-eksperimental adalah ketidakmampuannya memastikan hubungan sebab-akibat secara definitif karena tidak ada manipulasi dan kontrol perlakuan. Namun, desain ini tetap memiliki nilai ilmiah yang signifikan karena dapat mengungkap pola hubungan, kecenderungan, dan

prediksi yang berguna dalam praktik pendidikan. Dalam banyak kasus, seperti pengaruh motivasi belajar atau faktor keluarga terhadap prestasi, penelitian non-eksperimental adalah satu-satunya pendekatan yang etis dan praktis. Refleksi metodologis menunjukkan bahwa kekuatan desain non-eksperimental terletak pada kualitas teori, ketepatan definisi variabel, dan penggunaan teknik statistik yang relevan. Ketika ketiga komponen tersebut terpenuhi, desain non-eksperimental dapat menghasilkan bukti empiris yang kuat dan relevan untuk pengambilan keputusan pendidikan berbasis data.

3.4 Validitas Internal & Eksternal

Validitas internal dan eksternal merupakan dua dimensi esensial dalam desain penelitian kuantitatif karena keduanya menentukan sejauh mana hasil penelitian dapat dipercaya dan digeneralisasikan. Campbell dan Stanley (1963) mendefinisikan validitas internal sebagai tingkat keyakinan bahwa perubahan pada variabel dependen benar-benar disebabkan oleh variabel independen, bukan oleh faktor lain. Dalam eksperimen sejati, randomisasi menjadi mekanisme utama untuk mengurangi ancaman validitas internal seperti *selection bias*, *history*, *maturation*, *instrumentation*, *testing effect*, dan *statistical regression*. Shadish, Cook, dan Campbell (2002) memperluas konsep ini dengan menegaskan bahwa validitas internal adalah dasar inferensi kausal; tanpa validitas internal yang kuat, klaim sebab-akibat tidak dapat dipertahankan meskipun analisis statistik menunjukkan hubungan signifikan.

Dalam penelitian pendidikan, ancaman terhadap validitas internal sangat sering terjadi karena lingkungan pembelajaran sulit dikontrol sepenuhnya. Fraenkel, Wallen, dan Hyun (2021) menyebut bahwa variabel perancu seperti perbedaan

kemampuan awal siswa, motivasi, dukungan keluarga, atau gaya mengajar guru dapat memengaruhi hasil. Oleh karena itu, strategi penguatan validitas internal mencakup penggunaan pretest, *matching*, kontrol statistik seperti ANCOVA, penggunaan instrumen reliabel, serta prosedur perlakuan yang distandardisasi. Johnson dan Christensen (2020) menekankan bahwa dokumentasi proses penelitian dan konsistensi implementasi perlakuan merupakan aspek penting untuk memastikan penelitian berjalan sebagaimana dirancang.

Validitas eksternal, di sisi lain, merujuk pada sejauh mana hasil penelitian dapat digeneralisasikan ke populasi lain, konteks lain, atau waktu lain. Creswell (2014) menegaskan bahwa validitas eksternal bergantung pada representativitas sampel, karakteristik setting penelitian, dan kesamaan kondisi antara penelitian dan dunia nyata. Campbell dan Stanley (1963) mengidentifikasi ancaman seperti *reactive arrangements*, efek Hawthorne, dan *interaction of* selection *and treatment* sebagai faktor yang dapat mengurangi kemampuan generalisasi. Dalam penelitian pendidikan, misalnya, perlakuan yang efektif di sekolah dengan fasilitas lengkap tidak selalu memberikan hasil yang sama pada sekolah dengan sumber daya terbatas.

Gall, Gall, dan Borg (2019) menyatakan bahwa validitas eksternal sering kali ditingkatkan melalui replikasi penelitian di berbagai konteks. Ketika hasil eksperimen atau quasi-eksperimen konsisten di berbagai sekolah atau kelompok siswa, tingkat keyakinan terhadap generalisasi meningkat. Selain itu, penggunaan desain sampel bertingkat (*multistage sampling*) atau *cluster sampling* dapat meningkatkan representativitas sampel dalam penelitian pendidikan yang melibatkan banyak kelas atau sekolah. Cohen et al. (2018) juga menekankan bahwa peneliti harus berhati-hati dalam menarik kesimpulan generalisasi,

terutama jika sampel tidak acak atau jika kondisi penelitian terlalu artifisial.

Perkembangan era digital dan big data memberikan dimensi baru pada pembahasan validitas. Trochim (2006) mencatat bahwa data digital besar sering kali memperkuat validitas eksternal karena mencerminkan perilaku alami pengguna dalam konteks real time. Namun, validitas internal bisa menurun karena data observasional tidak mengontrol faktor perancu. Oleh karena itu, pendekatan seperti *propensity score matching*, *instrumental variables*, atau *regression discontinuity design* semakin penting untuk meningkatkan validitas internal dalam penelitian big data. Field (2018) menegaskan bahwa meskipun data besar menawarkan kekuatan prediktif, inferensi kausal tetap membutuhkan struktur desain yang kuat.

Maxwell dan Delaney (2004) menekankan keseimbangan antara kedua jenis validitas. Desain yang sangat ketat untuk meningkatkan validitas internal sering kali mengurangi validitas eksternal karena kondisi penelitian menjadi tidak alami. Sebaliknya, penelitian naturalistik yang meningkatkan validitas eksternal sering kali mengorbankan kontrol yang diperlukan untuk inferensi kausal. Johnson dan Christensen (2020) menyarankan bahwa peneliti pendidikan harus memilih strategi desain yang sesuai dengan tujuan penelitian: jika tujuan utama adalah uji efektivitas yang ketat, validitas internal menjadi prioritas; jika tujuan adalah penerapan luas, validitas eksternal lebih ditekankan.

Secara metodologis, validitas internal dan eksternal merupakan fondasi yang menentukan kekuatan bukti ilmiah. Refleksi kritis menunjukkan bahwa keduanya tidak boleh dipahami sebagai dimensi yang saling bertentangan, melainkan sebagai dua aspek yang harus diseimbangkan. Penelitian yang

kuat adalah penelitian yang mampu mengontrol ancaman internal sekaligus relevan dengan situasi nyata. Dengan demikian, validitas bukan hanya persoalan teknis tetapi juga keputusan epistemologis tentang bagaimana penelitian dapat berkontribusi pada pemahaman ilmiah dan praktik pendidikan berbasis bukti.

3.5 Strategi Meningkatkan Akurasi Desain

Akurasi desain penelitian merupakan komponen esensial yang menentukan kekuatan inferensi kausal dan kualitas keseluruhan hasil penelitian kuantitatif. Campbell dan Stanley (1963) menegaskan bahwa akurasi desain sangat dipengaruhi oleh kemampuan peneliti mengontrol ancaman validitas internal, eksternal, dan konstruk secara simultan. Untuk memperkuat akurasi desain, peneliti harus mengidentifikasi potensi bias sejak tahap perencanaan serta menerapkan strategi mitigasi yang sistematis. Dalam eksperimen sejati, randomisasi memegang peran sentral sebagai strategi untuk menyamakan karakteristik partisipan secara statistik sehingga setiap perbedaan hasil dapat dikaitkan dengan perlakuan. Shadish, Cook, dan Campbell (2002) menambahkan bahwa randomisasi bukan sekadar teknik penugasan acak, tetapi sebuah prinsip metodologis untuk menghilangkan bias seleksi dan memastikan bahwa estimasi kausal bersifat tidak bias (*unbiased estimator*).

Namun dalam praktik penelitian pendidikan, randomisasi sering sulit dilakukan karena keterbatasan administratif, sehingga strategi lain diperlukan. Creswell (2014) menyarankan penggunaan pretest sebagai alat untuk mengukur kesetaraan awal antar kelompok, sehingga perbedaan setelah perlakuan dapat dianalisis dengan lebih akurat. Teknik statistik seperti ANCOVA atau regresi juga dapat digunakan untuk mengontrol variabel perancu yang tidak dapat dihilangkan melalui desain.

Cohen, Manion, dan Morrison (2018) menekankan bahwa penyetaraan kelompok melalui *matching* atau *propensity score matching* dapat meningkatkan akurasi desain quasi-eksperimental terutama ketika randomisasi tidak mungkin dilakukan. Strategi ini memungkinkan peneliti menciptakan kelompok yang secara statistik serupa meskipun tidak dilakukan penugasan acak.

Fraenkel, Wallen, dan Hyun (2021) menyoroti pentingnya *standardization of procedures* sebagai strategi untuk memastikan akurasi desain. Perlakuan, instruksi, waktu penelitian, serta kondisi lingkungan harus dijaga agar sama antar kelompok. Dalam pendidikan, misalnya, guru yang menerapkan perlakuan harus mengikuti panduan yang jelas sehingga perlakuan tidak bias akibat variasi implementasi. Gall, Gall, dan Borg (2019) menyebut bahwa akurasi desain dapat ditingkatkan melalui *treatment fidelity*, yaitu memastikan bahwa intervensi diterapkan sesuai dengan rancangan aslinya. Pengawasan implementasi, pelatihan fasilitator, dan dokumentasi proses merupakan elemen penting untuk menjamin konsistensi perlakuan.

Di sisi lain, akurasi desain juga memerlukan strategi untuk meningkatkan validitas eksternal. Johnson dan Christensen (2020) menyarankan penggunaan sampel yang representatif, desain penelitian lintas konteks, serta replikasi dalam berbagai setting pendidikan. Dalam penelitian sekolah, teknik pengambilan sampel bertingkat (multistage sampling) atau cluster sampling dapat meningkatkan relevansi hasil penelitian terhadap populasi yang lebih luas. Maxwell dan Delaney (2004) menekankan bahwa generalisasi temuan tidak hanya bergantung pada sampel, tetapi juga kesesuaian konteks penelitian dengan lingkungan nyata di mana hasil penelitian akan diterapkan.

Era big data menghadirkan peluang baru untuk meningkatkan akurasi desain, tetapi juga membawa tantangan. Trochim (2006) menjelaskan bahwa penggunaan data digital dapat memperkuat validitas eksternal karena mencerminkan perilaku alami dalam konteks real time. Namun data observasional besar rentan terhadap bias internal karena tidak ada manipulasi perlakuan. Field (2018) menyarankan penggunaan model statistik lanjutan seperti *instrumental variables*, *regression discontinuity design* (RDD), atau *difference-in-differences* (DiD) untuk meningkatkan akurasi inferensi kausal dalam desain non-eksperimental. Pendekatan ini membantu mengurangi bias perancu ketika peneliti tidak memiliki kendali penuh terhadap kondisi penelitian.

Selain itu, validitas konstruk juga merupakan bagian dari akurasi desain yang tidak dapat diabaikan. Penggunaan instrumen yang reliabel dan valid membantu memastikan bahwa variabel yang diukur benar-benar mencerminkan konstruk teoretis. Pengembangan skala, analisis faktor, dan uji reliabilitas internal menjadi aspek yang sangat penting. Kline (2016) menegaskan bahwa instrumen yang buruk akan menghasilkan error measurement yang tinggi, yang pada gilirannya melemahkan akurasi model statistik apa pun, bahkan dalam desain eksperimental yang kuat sekalipun.

Secara metodologis, akurasi desain adalah hasil dari kombinasi strategi desain, kontrol prosedural, teknik statistik, dan ketepatan teoretis. Refleksi kritis menunjukkan bahwa akurasi desain tidak terjadi secara kebetulan; ia harus dibangun dengan kesadaran epistemologis sejak tahap awal penelitian. Ketika akurasi desain tercapai, penelitian kuantitatif tidak hanya menghasilkan data yang presisi, tetapi juga inferensi kausal yang

kuat, generalisasi yang relevan, dan kontribusi nyata bagi praktik pendidikan berbasis bukti.

Daftar Rujukan

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2021). *How to design and evaluate research in education* (10th ed.). McGraw-Hill.

Gall, M. D., Gall, J. P., & Borg, W. R. (2019). *Educational research: An introduction* (12th ed.). Pearson.

Johnson, B., & Christensen, L. (2020). *Educational research:* Quantitative*, qualitative, and mixed approaches* (7th ed.). SAGE.

Kerlinger, F. N. (2000). *Foundations of behavioral research* (4th ed.). Harcourt.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). Routledge.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. Harcourt.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2015). *Research methods in psychology* (10th ed.). McGraw-Hill.

Trochim, W. M. K. (2006). *The research methods knowledge base*. Atomic Dog.

Tuckman, B. W., & Harper, B. E. (2012). *Conducting educational research* (7th ed.). Rowman & Littlefield.

BAB 4
POPULASI, SAMPEL, DAN TEKNIK SAMPLING

4.1 Populasi & Sampling Frame

Populasi merupakan konsep fundamental dalam metodologi penelitian kuantitatif karena menjadi dasar bagi proses generalisasi hasil penelitian. Menurut Cochran (1977), populasi adalah keseluruhan elemen yang memiliki karakteristik tertentu dan menjadi perhatian peneliti untuk diteliti atau disimpulkan. Dalam konteks penelitian pendidikan dan sosial, populasi dapat berupa siswa, guru, kepala sekolah, orang tua, masyarakat, atau bahkan dokumen dan aktivitas digital. Creswell (2014) menegaskan bahwa definisi populasi harus jelas, terukur, dan relevan dengan tujuan penelitian sehingga peneliti dapat menentukan batasan siapa atau apa yang termasuk dalam wilayah inferensi penelitian. Kesalahan mendasar dalam tahap ini, seperti populasi yang terlalu luas atau tidak terdefinisi dengan baik, akan memengaruhi seluruh tahapan berikutnya, termasuk teknik sampling dan validitas eksternal.

Dalam literatur metodologi, dikenal dua jenis populasi: populasi target dan populasi terjangkau. Fraenkel, Wallen, dan Hyun (2021) menjelaskan bahwa populasi target adalah keseluruhan unit yang menjadi sasaran generalisasi penelitian, sementara populasi terjangkau adalah bagian dari populasi target yang dapat diakses peneliti secara realistis. Misalnya, populasi target penelitian adalah seluruh siswa sekolah menengah di Indonesia, tetapi populasi terjangkaunya mungkin hanya siswa di Kota Malang yang dapat dijangkau peneliti. Neuman (2014)

menegaskan bahwa perbedaan ini sangat penting karena gagal membedakan populasi target dan terjangkau dapat menyebabkan bias generalisasi, terutama jika populasi terjangkau tidak representatif terhadap populasi target.

Selanjutnya, salah satu konsep paling krusial dalam proses sampling adalah sampling frame. Lohr (2009) mendefinisikan sampling frame sebagai daftar atau representasi operasional dari elemen-elemen populasi yang dapat digunakan peneliti untuk memilih sampel. Dalam penelitian tradisional, sampling frame dapat berupa daftar siswa, database pegawai, daftar kelas, atau arsip administrasi. Kish (1965) menyebut sampling frame sebagai prasyarat bagi proses sampling probabilistik karena tanpa daftar yang lengkap dan akurat, sampel tidak dapat dipilih secara benar. Fowler (2014) menambahkan bahwa kualitas sampling frame sangat menentukan tingkat bias sampling; semakin lengkap dan bebas duplikasi, semakin besar peluang menghasilkan sampel representatif.

Dalam penelitian pendidikan modern, sampling frame sering diperoleh dari sistem informasi sekolah atau database administrasi. Namun, tantangan sering muncul ketika data tidak lengkap, tidak terbarui, atau memiliki inkonsistensi. Hal ini dapat menyebabkan *coverage error*, yaitu perbedaan antara populasi aktual dengan populasi yang tercantum dalam sampling frame. Babbie (2021) menegaskan bahwa coverage error merupakan salah satu sumber bias yang paling sering diabaikan oleh peneliti, terutama dalam survei skala besar.

Era digital membawa perubahan signifikan dalam konsep sampling frame. Dillman, Smyth, dan Christian (2014) menjelaskan bahwa penelitian berbasis internet atau survei daring menghadapi tantangan karena sampling frame digital jarang lengkap. Misalnya, penelitian menggunakan kuesioner

Google Form atau survei media sosial tidak memiliki sampling frame formal karena peneliti tidak mengetahui siapa saja yang memiliki akses atau peluang untuk berpartisipasi. Groves et al. (2009) menganggap ini sebagai tantangan metodologis besar karena sampel yang muncul cenderung *self-selected*, sehingga tidak dapat digeneralisasikan ke populasi yang lebih luas. Untuk mengatasi masalah ini, teknik seperti *panel sampling* online, frame integrasi, atau penggunaan big data yang terstruktur seperti database LMS (Learning Management System) dapat meningkatkan representativitas sampling frame digital.

Sekaran dan Bougie (2020) menekankan bahwa definisi populasi dan sampling frame juga harus disesuaikan dengan tujuan penelitian. Untuk penelitian eksperimen, populasi sering kali lebih sempit karena peneliti membutuhkan kondisi homogen untuk meningkatkan kontrol. Sebaliknya, untuk penelitian survei atau evaluasi program, populasi biasanya lebih luas dan memerlukan sampling frame yang komprehensif agar generalisasinya sah. Dalam penelitian perilaku, Gall, Gall, dan Borg (2019) menegaskan bahwa sampling frame dapat mencakup unit perilaku seperti aktivitas pembelajaran atau interaksi digital, bukan hanya individu. Hal ini semakin relevan pada penelitian big data di mana populasi dapat berupa log aktivitas siswa, rekaman klik LMS, atau data transaksi akademik.

Konsep evaluasi kualitas sampling frame mencakup empat aspek utama: *completeness*, *accuracy*, *up-to-dateness*, dan *uniqueness*. Thompson (2012) menjelaskan bahwa sampling frame harus lengkap (memuat semua elemen populasi), akurat (tidak memuat data salah), mutakhir (tidak kedaluwarsa), dan bebas duplikasi. Jika salah satu aspek ini tidak terpenuhi, risiko *sampling bias* meningkat. Misalnya, sampling frame siswa tahun 2022 tidak dapat digunakan untuk penelitian tahun 2024 tanpa

pembaruan, karena data perpindahan, kelulusan, atau mutasi dapat mengubah struktur populasi secara signifikan.

Dalam konteks big data, sampling frame tidak selalu eksplisit tetapi dapat dibentuk melalui algoritma penyaringan data. Namun, Neuman (2014) mengingatkan bahwa meskipun big data memiliki ukuran besar, hal ini tidak menjamin bebas bias karena tidak semua anggota populasi memiliki peluang yang sama untuk terekam dalam dataset. Hal ini disebut *coverage bias digital*, salah satu sumber kesalahan terbesar dalam penelitian era modern.

Secara metodologis, populasi dan sampling frame merupakan fondasi dari seluruh proses generalisasi hasil penelitian. Refleksi kritis menunjukkan bahwa ketidakjelasan populasi atau sampling frame akan melemahkan validitas eksternal dan mengurangi tingkat kepercayaan terhadap hasil penelitian, meskipun analisis statistik dilakukan dengan benar. Oleh karena itu, peneliti harus menyusun definisi populasi yang tegas, mengidentifikasi sampling frame akurat, serta mengantisipasi bias dalam pemilihan sampel agar penelitian kuantitatif dapat menghasilkan bukti yang valid, reliabel, dan dapat digeneralisasikan secara sah.

4.2 Probability Sampling

Probability sampling merupakan teknik pengambilan sampel yang memberikan peluang sama bagi setiap elemen populasi untuk terpilih sebagai anggota sampel. Menurut Cochran (1977), probability sampling memungkinkan peneliti melakukan estimasi statistik yang tidak bias dan menghitung kesalahan sampling secara akurat, sehingga hasil penelitian dapat digeneralisasikan ke populasi target. Kish (1965) bahkan menyebut probability sampling sebagai fondasi inferensi ilmiah

dalam survei karena pemilihan sampel dilakukan secara acak dan mengikuti prinsip peluang. Dalam penelitian pendidikan dan sosial, probability sampling digunakan ketika peneliti ingin memperoleh representativitas tinggi, seperti pada survei nasional, evaluasi program, atau penelitian korelasional berskala besar.

Salah satu bentuk probability sampling yang paling sederhana adalah simple random sampling, yaitu setiap elemen populasi memiliki peluang yang sama untuk dipilih tanpa memperhatikan struktur populasi. Creswell (2014) menjelaskan bahwa teknik ini sangat efektif ketika populasi homogen atau relatif seragam. Namun, ketika populasi besar dan heterogen, simple random sampling menjadi kurang efisien. Lohr (2009) menyarankan penggunaan stratified random sampling, yaitu teknik di mana populasi dibagi ke dalam strata atau kelompok kecil berdasarkan karakteristik tertentu seperti jenis kelamin, wilayah, atau tingkat sekolah. Teknik ini meningkatkan presisi estimasi karena variasi dalam strata dapat dikendalikan dengan lebih baik.

Pada penelitian pendidikan, cluster sampling sering digunakan karena unit penelitian biasanya berupa kelas, sekolah, atau daerah. Gall, Gall, dan Borg (2019) menjelaskan bahwa cluster sampling lebih efisien secara logistik, terutama ketika populasi tersebar secara geografis. Namun, teknik ini memiliki kelemahan yaitu meningkatnya *sampling error* akibat korelasi antar individu dalam satu klaster (*intra-cluster correlation*). Lohr (2009) menegaskan perlunya desain berlapis atau penyesuaian analisis untuk mengatasi kelemahan ini, misalnya dengan *design effect*.

Selain itu, systematic sampling merupakan teknik probability sampling yang digunakan ketika sampling frame

tersusun secara berurutan. Fowler (2014) menyatakan bahwa systematic sampling dapat menghasilkan sampel representatif dengan cepat, selama urutan dalam sampling frame tidak mengikuti pola tertentu yang dapat menimbulkan bias.

Dalam era digital, probability sampling menghadapi tantangan baru. Survei online jarang memiliki sampling frame lengkap, sehingga peluang elemen populasi untuk berpartisipasi tidak dapat dihitung secara jelas. Groves et al. (2009) mengingatkan bahwa survei daring sering kali tidak memenuhi prinsip probability sampling karena keterbatasan akses internet atau bias partisipasi. Untuk mengatasi hal ini, beberapa peneliti mengembangkan probability-based online panels, yaitu panel partisipan yang direkrut menggunakan teknik random sampling tradisional lalu diberi akses untuk survei daring. Thompson (2012) menyebut pendekatan ini sebagai bentuk hybrid yang menggabungkan sampling klasik dan teknologi digital.

Representativitas tetap menjadi tujuan utama probability sampling. Sekaran dan Bougie (2020) menekankan bahwa akurasi generalisasi sangat bergantung pada desain probability sampling yang tepat, bukan sekadar ukuran sampel yang besar. Yamane (1967) mendukung hal ini melalui rumus ukuran sampel yang mempertimbangkan proporsi populasi dan margin of error. Dengan menggunakan probability sampling, peneliti dapat menghitung tingkat presisi, interval kepercayaan, dan kemungkinan kesalahan dengan cara yang sistematis.

Secara metodologis, probability sampling merupakan strategi penting untuk meningkatkan validitas eksternal. Refleksi kritis menunjukkan bahwa meskipun teknik ini lebih kompleks dan memerlukan sampling frame yang akurat, probability sampling memberikan dasar ilmiah yang kuat untuk generalisasi hasil penelitian. Dalam konteks pendidikan, penggunaan

probability sampling memastikan bahwa kesimpulan yang diambil tidak bersifat bias terhadap kelompok tertentu dan dapat diterapkan pada populasi yang lebih luas. Karena itu, probability sampling tetap menjadi pilar utama dalam desain penelitian kuantitatif modern.

4.3 Non-Probability Sampling

Non-probability sampling adalah teknik pengambilan sampel di mana elemen populasi tidak memiliki peluang yang sama untuk terpilih, dan probabilitas pemilihan sampel tidak dapat dihitung. Menurut Neuman (2014), teknik ini banyak digunakan dalam penelitian sosial dan pendidikan ketika sampling frame tidak tersedia, populasi sulit dijangkau, atau ketika tujuan penelitian lebih bersifat eksploratif daripada generalisasi statistik. Karena pemilihannya tidak acak, non-probability sampling memiliki keterbatasan dalam representativitas; namun dalam praktiknya, ia sering menjadi pilihan realistis terutama ketika sumber daya terbatas atau populasi bersifat khusus.

Salah satu teknik paling umum adalah purposive sampling, yaitu pemilihan sampel berdasarkan pertimbangan peneliti terhadap karakteristik tertentu yang relevan. Creswell (2014) menyatakan bahwa purposive sampling sangat berguna ketika peneliti membutuhkan responden dengan pengalaman atau karakteristik spesifik, misalnya guru berpengalaman dalam model pembelajaran tertentu, siswa berprestasi tinggi, atau pengguna aktif suatu aplikasi pendidikan. Teknik ini relevan dalam evaluasi program pendidikan, meskipun keterbatasan generalisasi harus diakui secara metodologis.

Jenis lain adalah convenience sampling, yaitu pemilihan sampel berdasarkan kemudahan akses. Babbie (2021) mencatat

bahwa convenience sampling sering digunakan pada survei mahasiswa atau pengguna media sosial karena partisipan mudah dijangkau. Namun, teknik ini memiliki tingkat bias tinggi karena partisipan tidak mewakili populasi secara proporsional. Fowler (2014) menyebut convenience sampling sebagai teknik yang harus digunakan dengan kehati-hatian, terutama ketika peneliti berniat menarik kesimpulan inferensial.

Snowball sampling digunakan ketika populasi sulit diidentifikasi atau tersembunyi, seperti kelompok tertentu dalam penelitian sosial. Neuman (2014) menjelaskan bahwa snowball sampling dilakukan dengan meminta responden awal merekomendasikan responden berikutnya. Meskipun teknik ini berguna untuk menemukan partisipan yang tidak terdaftar dalam sampling frame, bias jaringan sangat mungkin terjadi karena responden cenderung merekomendasikan orang yang mirip dengan mereka.

Dalam konteks pendidikan, quota sampling menjadi alternatif ketika peneliti ingin memastikan distribusi sampel sesuai proporsi tertentu tanpa randomisasi. Sekaran dan Bougie (2020) menjelaskan bahwa quota sampling memungkinkan peneliti mengumpulkan data dengan cepat sambil tetap mempertahankan proporsi kelompok, seperti jenis kelamin, tingkatan sekolah, atau program studi. Namun, karena pemilihan sampel dalam setiap kuota tidak acak, bias tetap tidak dapat dihindari.

Era digital memperluas cakupan non-probability sampling. Survei daring, polling media sosial, dan platform e-learning pada umumnya menggunakan self-selection sampling, di mana partisipan secara sukarela memilih untuk ikut serta. Groves et al. (2009) menyebut bahwa teknik ini rawan *volunteer bias*, karena responden yang berminat mengisi survei biasanya memiliki

karakteristik berbeda dengan populasi umum. Meskipun demikian, pada penelitian eksploratif atau analitik prediktif berbasis big data, teknik ini banyak digunakan karena volume datanya besar dan cepat diperoleh.

Lohr (2009) menegaskan bahwa keterbatasan utama non-probability sampling bukan sekadar ketiadaan peluang acak, tetapi risiko bias sistematis yang sulit diukur. Oleh karena itu, analisis hasil penelitian harus menyertakan peringatan tentang keterbatasan generalisasi. Gall, Gall, dan Borg (2019) menyarankan bahwa ketika non-probability sampling digunakan, peneliti harus memperkuat validitas melalui triangulasi, replikasi studi, atau penggunaan analisis statistik yang sesuai seperti *post-stratification*.

Secara metodologis, non-probability sampling penting dalam penelitian pendidikan karena sering kali kondisi lapangan tidak memungkinkan penerapan probability sampling. Refleksi kritis menunjukkan bahwa meskipun teknik ini lebih fleksibel dan praktis, peneliti harus berhati-hati dalam menarik inferensi dan harus secara eksplisit membatasi klaim generalisasi. Dengan pemilihan teknik yang tepat, analisis yang hati-hati, dan transparansi metodologis, non-probability sampling tetap dapat memberikan kontribusi signifikan, terutama pada penelitian eksploratif, evaluasi program, dan studi perilaku dalam konteks digital.

4.4 Penentuan Ukuran Sampel

Penentuan ukuran sampel merupakan langkah strategis dalam penelitian kuantitatif karena berkaitan langsung dengan tingkat presisi estimasi, kekuatan statistik, dan kemampuan generalisasi temuan. Yamane (1967) menekankan bahwa ukuran sampel harus mencerminkan keseimbangan antara kebutuhan

akurasi dan keterbatasan sumber daya. Dalam studi survei, analisis regresi, maupun eksperimen pendidikan, ukuran sampel yang terlalu kecil dapat meningkatkan kesalahan sampling dan menurunkan kekuatan analitik, sementara ukuran sampel yang terlalu besar dapat menghabiskan sumber daya tanpa meningkatkan manfaat ilmiah secara signifikan. Oleh karena itu, penentuan ukuran sampel tidak dapat dilakukan secara arbitrer, melainkan harus mengikuti prinsip-prinsip metodologis yang terukur.

Cochran (1977) memperkenalkan beberapa rumus untuk menentukan ukuran sampel, yang paling dasar adalah rumus sampel untuk populasi besar:

$$n = \frac{Z^2 P(1-P)}{e^2}$$

di mana *Z* adalah nilai z untuk tingkat kepercayaan tertentu, *P* adalah proporsi populasi yang diperkirakan memiliki karakteristik tertentu, dan *e* adalah margin of error yang diinginkan. Rumus ini banyak digunakan dalam penelitian pendidikan skala besar, terutama ketika peneliti ingin mengestimasi proporsi, seperti persentase siswa yang memahami materi tertentu atau proporsi guru yang menguasai strategi pembelajaran tertentu. Ketika populasi terbatas, Cochran menyarankan koreksi ukuran sampel dengan rumus populasi terbatas (*finite population correction*) untuk meningkatkan akurasi estimasi.

Salah satu pendekatan praktis yang banyak digunakan adalah rumus Yamane (1967):

$$n = \frac{N}{1 + Ne^2}$$

yang memudahkan peneliti menentukan ukuran sampel dengan cepat ketika populasi diketahui dan tanpa perlu mengestimasi proporsi. Rumus ini menjadi acuan umum dalam penelitian sosial karena sederhana namun cukup akurat untuk survei pendidikan berskala kecil hingga menengah. Lohr (2009) menekankan bahwa rumus-rumus tersebut bekerja secara optimal hanya ketika sampling frame lengkap dan teknik probability sampling digunakan; jika tidak, ukuran sampel harus ditambah untuk mengimbangi potensi bias.

Dalam penelitian pendidikan yang menggunakan teknik *cluster sampling*, ukuran sampel harus mempertimbangkan *design effect* karena individu dalam satu klaster cenderung lebih mirip dibanding antar klaster. Kish (1965) menjelaskan bahwa *design effect* memperbesar ukuran sampel yang diperlukan, sehingga penelitian yang menggunakan sampel kelas atau sekolah sering memerlukan sampel lebih besar dibanding simple random sampling. Thompson (2012) mengingatkan bahwa peneliti harus menghitung efek kelompok (intraclass correlation coefficient, ICC) untuk menentukan jumlah klaster dan ukuran per klaster yang optimal, terutama dalam penelitian sekolah berjenjang.

Selain pendekatan rumus, Creswell (2014) menyatakan bahwa ukuran sampel juga harus memperhatikan jenis analisis statistik yang digunakan. Misalnya, regresi multipel membutuhkan minimal 10–20 sampel per variabel independen, sementara analisis SEM membutuhkan ukuran sampel lebih besar (150–400) untuk menghasilkan model yang stabil. Sekaran dan Bougie (2020) menambahkan bahwa penelitian eksploratori dapat menggunakan sampel lebih kecil, tetapi penelitian konfirmatori harus memiliki ukuran sampel memadai untuk meningkatkan reliabilitas dan validitas hasil.

Dalam konteks big data, penentuan ukuran sampel berubah secara paradigmatik. Groves et al. (2009) menjelaskan bahwa dalam dataset digital yang sangat besar, isu utama bukan lagi ukuran sampel, tetapi representativitas dan bias. Walaupun dataset besar meningkatkan presisi estimasi, ia tidak otomatis memperbaiki bias jika data tidak mencerminkan seluruh populasi. Karena itu, peneliti harus memastikan bahwa data digital mencakup berbagai kelompok secara proporsional atau menggunakan teknik koreksi seperti *post-stratification* dan *weighting adjustment*.

Gall, Gall, dan Borg (2019) menyarankan bahwa dalam penelitian pendidikan, ukuran sampel harus mempertimbangkan ketersediaan partisipan, etika penelitian, dan konteks implementasi. Sebagai contoh, eksperimen kelas sering menghadapi keterbatasan jumlah siswa dan tidak mungkin memenuhi persyaratan statistik ideal. Dalam situasi tersebut, peneliti harus menggabungkan strategi desain dan analisis—misalnya, menggunakan repeated measures untuk meningkatkan kekuatan statistik tanpa menambah sampel.

Refleksi metodologis menunjukkan bahwa penentuan ukuran sampel adalah proses ilmiah yang membutuhkan pertimbangan matematis, teoretis, dan praktis. Ukuran sampel yang tepat meningkatkan validitas eksternal dan internal serta memberikan dasar kuat bagi inferensi statistik. Sebaliknya, ukuran sampel yang tidak memadai menyebabkan hasil penelitian menjadi tidak stabil, bias, dan sulit digeneralisasikan. Oleh karena itu, peneliti harus menggunakan kombinasi rumus, teori, dan pertimbangan kontekstual agar desain penelitian menghasilkan temuan yang akurat dan kredibel.

4.5 Kesalahan Sampling

Kesalahan sampling merupakan salah satu aspek penting yang memengaruhi kualitas data dalam penelitian kuantitatif. Cochran (1977) mendefinisikan kesalahan sampling (*sampling error*) sebagai perbedaan antara nilai statistik sampel dan parameter populasi yang terjadi karena hanya sebagian elemen populasi yang diambil sebagai sampel. Semakin besar ukuran sampel dan semakin tepat teknik sampling probabilistik yang digunakan, semakin kecil kemungkinan terjadinya kesalahan sampling. Kesalahan ini bersifat matematis dan dapat dihitung melalui margin of error serta interval kepercayaan. Oleh karena itu, sampling error merupakan bagian alami dari penelitian survei dan dapat diminimalkan dengan desain yang baik.

Namun, para metodolog seperti Groves et al. (2009) menekankan bahwa kesalahan sampling hanyalah salah satu jenis kesalahan dalam survei. Kesalahan lain yang tidak kalah penting adalah non-sampling error, yaitu kesalahan yang terjadi bukan karena penggunaan sampel, tetapi karena cacat proses pengumpulan data, pengukuran, atau bias partisipasi. Lohr (2009) menjelaskan bahwa non-sampling error sering kali lebih berbahaya daripada sampling error karena sifatnya sistematis dan sulit dideteksi melalui statistik sederhana. Non-sampling error mencakup bias nonresponse, kesalahan pengukuran, kesalahan pencatatan, kesalahan frame, dan bias seleksi. Dalam penelitian pendidikan, misalnya, bias dapat terjadi ketika siswa dengan kemampuan rendah enggan mengisi survei, sehingga sampel tidak mewakili populasi secara akurat.

Dillman, Smyth, dan Christian (2014) menguraikan bahwa kesalahan frame (*coverage error*) terjadi ketika sampling frame tidak mencakup seluruh elemen populasi atau mengandung data yang salah. Misalnya, daftar siswa yang tidak mutakhir atau data

guru yang belum diperbarui dapat menyebabkan sebagian populasi tidak memiliki peluang untuk terpilih. Fowler (2014) menilai kesalahan frame sebagai salah satu sumber bias yang paling sering diabaikan, terutama dalam penelitian sekolah atau survei institusional. Untuk mengurangi kesalahan ini, peneliti harus memastikan bahwa sampling frame lengkap, akurat, dan diperbarui sebelum proses sampling dilakukan.

Kesalahan nonresponse (*nonresponse error*) merupakan jenis kesalahan lain yang sering muncul dalam survei pendidikan maupun survei daring. Groves et al. (2009) menjelaskan bahwa nonresponse error terjadi ketika sebagian individu dalam sampel tidak memberikan respons, dan karakteristik mereka berbeda secara sistematis dari mereka yang merespons. Hal ini akan menghasilkan data yang bias meskipun teknik sampling sudah benar. Dalam survei online, tingkat respons rendah dan bias partisipasi menjadi tantangan besar. Strategi seperti *follow-up reminders*, insentif, desain kuesioner yang menarik, dan penggunaan multi-mode (online–offline) dapat meningkatkan tingkat respons.

Kesalahan pengukuran (*measurement error*) terjadi ketika instrumen tidak mengukur apa yang seharusnya diukur karena pertanyaan ambigu, respons tidak jujur, atau kondisi pengisian tidak kondusif. Neuman (2014) mengingatkan bahwa kesalahan ini sering muncul pada survei pendidikan yang menggunakan skala sikap atau kuesioner motivasi. Untuk mengurangi kesalahan pengukuran, instrumen harus diuji validitas dan reliabilitasnya sebelum digunakan secara luas.

Dalam era big data, kesalahan sampling dan non-sampling memperoleh bentuk baru. Groves et al. (2009) menyatakan bahwa meskipun dataset digital besar, data tersebut tidak selalu representatif. *Big data bias* muncul ketika hanya sebagian

kelompok yang terekam dalam sistem digital, misalnya hanya siswa aktif menggunakan LMS yang memiliki data lengkap, sementara siswa dengan keterbatasan akses tidak terdokumentasi. Hal ini menciptakan *coverage error* digital yang jauh lebih besar daripada survei tradisional. Selain itu, algoritma seleksi data dalam platform digital dapat menghasilkan bias otomatis yang tidak disadari peneliti.

Gall, Gall, dan Borg (2019) menekankan pentingnya mengidentifikasi dan mengontrol semua bentuk kesalahan sampling dalam penelitian pendidikan. Peneliti tidak hanya harus menghitung margin of error tetapi juga memahami sumber bias dan mengambil langkah pencegahan sejak awal, seperti memperbaiki sampling frame, menggunakan teknik randomisasi, meningkatkan desain instrumen, dan menerapkan strategi pengendalian kualitas data.

Refleksi metodologis menunjukkan bahwa kualitas penelitian tidak hanya ditentukan oleh analisis statistik, tetapi juga oleh kemampuan peneliti mengelola kesalahan sampling dan non-sampling. Ketika kesalahan ini diminimalkan melalui desain yang baik, validitas eksternal meningkat dan hasil penelitian menjadi lebih dapat dipercaya. Dengan demikian, pemahaman mendalam tentang kesalahan sampling merupakan syarat utama untuk menghasilkan penelitian kuantitatif yang kredibel dan relevan.

Daftar Rujukan

Babbie, E. (2021). *The practice of social research* (15th ed.). Cengage.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Fowler, F. J. (2014). *Survey research methods* (5th ed.). SAGE.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2021). *How to* design *and evaluate research in education* (10th ed.). McGraw-Hill.

Gall, M. D., Gall, J. P., & Borg, W. R. (2019). *Educational research: An introduction* (12th ed.). Pearson.

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Kish, L. (1965). *Survey sampling*. Wiley.

Lohr, S. L. (2009). *Sampling: Design and analysis* (2nd ed.). Cengage.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson.

Sekaran, U., & Bougie, R. (2020). *Research methods for business: A skill-building approach* (8th ed.). Wiley.

Thompson, S. K. (2012). *Sampling* (3rd ed.). Wiley.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77–100.

Yamane, T. (1967). *Statistics: An introductory analysis* (2nd ed.). Harper & Row.

BAB 5

INSTRUMEN PENELITIAN DAN UJI VALIDITAS

5.1 Pengembangan Instrumen

Pengembangan instrumen merupakan tahap fundamental dalam penelitian kuantitatif karena menjadi penentu utama kualitas data dan ketepatan inferensi ilmiah yang dihasilkan. Instrumen penelitian berfungsi sebagai alat untuk mengoperasionalkan konstruk teoretis yang bersifat laten ke dalam bentuk skor numerik yang dapat dianalisis secara statistik. Menurut Anastasi dan Urbina (1997), instrumen pengukuran dalam penelitian pendidikan dan psikologi harus mampu merepresentasikan atribut yang diukur secara akurat, konsisten, dan bebas dari bias sistematis. Dengan demikian, pengembangan instrumen tidak dapat dilakukan secara intuitif atau pragmatis, melainkan harus mengikuti prosedur metodologis yang sistematis dan berlandaskan teori psikometri.

Langkah awal dalam pengembangan instrumen adalah pendefinisian konstruk secara konseptual. Konstruk merupakan variabel laten yang tidak dapat diamati secara langsung, seperti motivasi belajar, sikap, persepsi, atau self-efficacy. Menurut DeVellis (2017), kegagalan mendefinisikan konstruk secara jelas akan menyebabkan instrumen kehilangan fokus konseptual dan berisiko mengukur atribut yang berbeda dari yang dimaksudkan. Oleh karena itu, definisi konstruk harus disusun melalui kajian literatur yang mendalam, mencakup teori klasik dan temuan empiris terkini. Cohen, Manion, dan Morrison (2018)

menegaskan bahwa definisi konstruk yang baik harus memiliki batasan konseptual yang jelas agar dapat diturunkan menjadi indikator yang terukur secara empiris.

Setelah konstruk didefinisikan, tahap berikutnya adalah penurunan dimensi dan indikator konstruk. Dimensi berfungsi menjelaskan aspek-aspek utama dari konstruk, sedangkan indikator merepresentasikan perilaku atau atribut yang dapat diukur. Nunnally dan Bernstein (1994) menjelaskan bahwa setiap konstruk seharusnya direpresentasikan oleh sejumlah indikator yang mencerminkan keseluruhan domain atribut yang diukur. Reduksi konstruk yang berlebihan akan menyebabkan instrumen kehilangan validitas isi, sementara indikator yang terlalu luas dapat menurunkan konsistensi internal. Oleh karena itu, peneliti harus menjaga keseimbangan antara kelengkapan cakupan konstruk dan kejelasan operasional indikator.

Tahap selanjutnya adalah penyusunan butir instrumen (item writing). Item merupakan pernyataan atau pertanyaan yang secara langsung direspons oleh responden. Menurut DeVellis (2017), kualitas instrumen sangat ditentukan oleh kualitas item, sehingga penyusunan item harus memperhatikan prinsip kejelasan bahasa, kesederhanaan struktur kalimat, dan ketepatan makna. Clark dan Watson (1995) menegaskan bahwa item yang baik harus fokus pada satu aspek konstruk, tidak ambigu, serta tidak mengandung pernyataan ganda (double-barreled). Selain itu, item sebaiknya ditulis dengan bahasa yang sesuai dengan tingkat pemahaman responden agar tidak menimbulkan kesalahan pengukuran akibat interpretasi yang keliru.

Dalam praktik pengembangan instrumen, peneliti dianjurkan untuk menyusun jumlah item awal yang relatif lebih banyak daripada jumlah item yang akan digunakan dalam instrumen final. Pendekatan ini sejalan dengan prinsip seleksi

item dalam psikometri. Menurut Field (2018), proses analisis empiris seperti korelasi item–total dan analisis faktor akan mengeliminasi item-item yang tidak berkontribusi secara signifikan terhadap konstruk. Dengan demikian, penghapusan item bukan merupakan kegagalan desain instrumen, melainkan bagian dari proses penyempurnaan berbasis data empiris.

Tahap berikutnya adalah penentuan format respons dan sistem skoring. Skala Likert merupakan format respons yang paling banyak digunakan dalam penelitian pendidikan, psikologi, dan sosial. Nunnally dan Bernstein (1994) menyatakan bahwa pemilihan jumlah kategori respons harus mempertimbangkan keseimbangan antara sensitivitas pengukuran dan kemampuan diskriminasi responden. Skala dengan kategori yang terlalu sedikit dapat menghasilkan varians skor yang rendah, sedangkan skala dengan kategori yang terlalu banyak berpotensi membebani responden secara kognitif. Oleh karena itu, pemilihan format respons harus disesuaikan dengan karakteristik populasi dan tujuan pengukuran.

Instrumen yang telah disusun kemudian perlu melalui penelaahan awal secara teoretis dan linguistik sebelum diuji secara empiris. Menurut Trochim (2006), tahap ini merupakan bentuk validasi konseptual awal yang bertujuan memastikan kesesuaian item dengan indikator konstruk serta meminimalkan kesalahan sistematis sebelum pengumpulan data lapangan. Penelaahan ini juga mencakup konsistensi terminologi, kejelasan redaksi, dan kesesuaian konteks budaya. Instrumen yang lolos tahap ini memiliki dasar teoretis yang lebih kuat untuk memasuki tahap pengujian validitas dan reliabilitas.

Dalam kerangka evidence-based research, pengembangan instrumen tidak dapat dilepaskan dari tuntutan validitas modern. Menurut Messick (1995), validitas bukan sekadar karakteristik

instrumen, melainkan kualitas inferensi yang dibuat berdasarkan skor yang dihasilkan. Oleh karena itu, setiap keputusan dalam pengembangan instrumen—mulai dari definisi konstruk hingga pemilihan format respons—harus diarahkan untuk mendukung interpretasi skor yang sahih dan bermakna. Kesalahan pada tahap awal pengembangan instrumen akan berdampak sistemik terhadap seluruh proses analisis dan penarikan kesimpulan.

Perkembangan perangkat lunak statistik seperti SPSS, Amos, LISREL, dan Mplus telah memperkuat pendekatan pengembangan instrumen berbasis analisis empiris. Menurut Hair et al. (2019), pengembangan instrumen modern menuntut integrasi antara teori pengukuran klasik, analisis faktor eksploratori, dan pendekatan konfirmatori sebagai bentuk validitas konstruk lanjutan. Dengan demikian, pengembangan instrumen tidak lagi dipahami sebagai tahap awal yang sederhana, melainkan sebagai proses metodologis berlapis yang menentukan kualitas inferensi ilmiah dalam penelitian pendidikan, psikologi, dan sosial.

Secara reflektif, pengembangan instrumen merupakan fondasi utama penelitian kuantitatif yang bermutu. Instrumen yang disusun secara sistematis, berbasis teori, dan diuji secara empiris akan menghasilkan data yang reliabel dan valid. Sebaliknya, instrumen yang dikembangkan tanpa landasan psikometrik yang kuat berpotensi menghasilkan kesimpulan yang bias dan tidak dapat dipertanggungjawabkan secara ilmiah. Oleh karena itu, pengembangan instrumen harus diposisikan sebagai proses ilmiah yang kritis dalam menjaga integritas dan akurasi penelitian berbasis bukti.

5.2 Validitas Isi, Kriteria, dan Konstruk

Validitas merupakan konsep sentral dalam pengukuran kuantitatif karena menentukan sejauh mana instrumen benar-benar mengukur konstruk yang dimaksud. Dalam psikometri modern, validitas tidak lagi dipahami sebagai sifat tunggal instrumen, melainkan sebagai kualitas inferensi yang dibuat berdasarkan skor hasil pengukuran. Menurut Messick (1995), validitas adalah konsep terpadu (*unitary concept*) yang mencakup bukti teoretis dan empiris untuk mendukung interpretasi skor. Dengan demikian, pembahasan validitas harus melampaui uji statistik semata dan mencakup rasional konseptual, struktur konstruk, serta kegunaan hasil pengukuran dalam konteks penelitian pendidikan, psikologi, dan sosial.

Jenis validitas yang paling awal diuji dalam pengembangan instrumen adalah validitas isi (*content validity*). Validitas isi merujuk pada sejauh mana butir-butir instrumen merepresentasikan keseluruhan domain konstruk yang diukur. Menurut Nunnally dan Bernstein (1994), validitas isi bersifat teoretis dan tidak dapat diuji sepenuhnya dengan statistik inferensial, melainkan melalui penilaian sistematis oleh pakar. DeVellis (2017) menegaskan bahwa validitas isi yang lemah akan menyebabkan instrumen kehilangan relevansi konseptual, meskipun secara statistik tampak reliabel. Oleh karena itu, validitas isi merupakan fondasi awal sebelum instrumen diuji secara empiris.

Dalam praktik penelitian modern, validitas isi sering diukur menggunakan Content Validity Index (CVI). Polit dan Beck menjelaskan bahwa CVI memungkinkan peneliti mengkuantifikasi penilaian ahli terhadap relevansi item. CVI terdiri atas indeks tingkat item (Item-*level Content Validity Index* atau I-CVI) dan indeks tingkat skala (*Scale-level Content Validity Index* atau S-CVI).

I-CVI dihitung berdasarkan proporsi ahli yang menilai item sebagai relevan, sedangkan S-CVI menggambarkan rata-rata atau proporsi keseluruhan item yang memenuhi kriteria validitas isi. Nilai I-CVI ≥ 0,78 umumnya dianggap memadai apabila jumlah ahli minimal tiga orang. Pendekatan ini sejalan dengan prinsip *evidence-based measurement* karena menggabungkan pertimbangan teoretis dan kuantifikasi sistematis.

Selain validitas isi, validitas kriteria (criterion-related validity) juga menjadi aspek penting dalam pengembangan instrumen. Validitas kriteria merujuk pada sejauh mana skor instrumen berkorelasi dengan ukuran eksternal yang dianggap sebagai kriteria yang relevan. Menurut Anastasi dan Urbina (1997), validitas kriteria menunjukkan kegunaan praktis instrumen dalam memprediksi atau merefleksikan perilaku aktual. Validitas kriteria dibedakan menjadi validitas konkuren dan validitas prediktif. Validitas konkuren diperoleh ketika skor instrumen berkorelasi dengan ukuran kriteria yang diukur pada waktu yang sama, sedangkan validitas prediktif berkaitan dengan kemampuan instrumen memprediksi kinerja atau perilaku di masa depan.

Dalam penelitian pendidikan, validitas kriteria sering diuji melalui korelasi skor instrumen dengan nilai akademik, hasil asesmen standar, atau indikator kinerja lain yang relevan. Cohen, Manion, dan Morrison (2018) menegaskan bahwa pemilihan kriteria harus mempertimbangkan kesesuaian konseptual dan kualitas psikometrik ukuran pembanding. Korelasi yang tinggi dengan kriteria yang tidak relevan justru dapat menimbulkan kesalahan interpretasi. Oleh karena itu, validitas kriteria harus dipahami sebagai bukti pendukung, bukan satu-satunya dasar klaim validitas instrumen.

Jenis validitas yang paling komprehensif dalam psikometri modern adalah validitas konstruk (construct validity). Validitas konstruk merujuk pada sejauh mana instrumen benar-benar mengukur konstruk teoretis yang mendasarinya. Menurut Cronbach dan Meehl, validitas konstruk dicapai melalui akumulasi bukti dari berbagai sumber, termasuk struktur internal instrumen, hubungan antar konstruk, serta konsistensi dengan teori yang mendasarinya. Messick (1995) memperluas pandangan ini dengan menekankan bahwa validitas konstruk mencakup implikasi sosial dan konsekuensi penggunaan skor.

Salah satu pendekatan utama untuk menguji validitas konstruk adalah analisis struktur internal instrumen. Dalam konteks ini, korelasi item–total dan analisis faktor memainkan peran penting. Menurut Field (2018), korelasi item–total digunakan untuk menilai sejauh mana setiap item berkontribusi terhadap keseluruhan skala. Item dengan korelasi rendah menunjukkan ketidaksesuaian dengan konstruk dan perlu dipertimbangkan untuk direvisi atau dihapus. Namun, korelasi item–total hanya memberikan gambaran awal dan tidak cukup untuk menjelaskan struktur konstruk secara menyeluruh.

Analisis faktor eksploratori (EFA) dan analisis faktor konfirmatori (CFA) merupakan pendekatan lanjutan dalam pengujian validitas konstruk. Menurut Hair et al. (2019), EFA digunakan untuk mengidentifikasi struktur laten instrumen ketika hubungan antar item belum sepenuhnya diketahui, sedangkan CFA digunakan untuk menguji kesesuaian model faktor yang telah ditentukan berdasarkan teori. Kline (2016) menegaskan bahwa CFA memberikan bukti validitas konstruk yang lebih kuat karena memungkinkan pengujian model pengukuran secara eksplisit melalui indeks kesesuaian (*goodness-of-fit indices*). Dengan

demikian, EFA dan CFA dipahami sebagai pendekatan komplementer dalam membangun validitas konstruk.

Perkembangan teknologi statistik dan perangkat lunak seperti SPSS, Amos, LISREL, dan Mplus telah memudahkan peneliti dalam menguji validitas konstruk secara lebih komprehensif. Tabachnick dan Fidell (2019) menekankan bahwa penggunaan analisis multivariat harus disertai pemahaman teoretis yang kuat agar hasil analisis tidak disalahartikan. Validitas konstruk bukan sekadar persoalan statistik, melainkan kesesuaian antara data empiris dan kerangka teoretis yang mendasarinya.

Secara reflektif, validitas isi, kriteria, dan konstruk merupakan pilar utama dalam pengembangan instrumen penelitian. Ketiga jenis validitas ini saling melengkapi dan tidak dapat dipisahkan satu sama lain. Instrumen yang memiliki validitas isi yang baik tetapi gagal menunjukkan validitas konstruk akan menghasilkan inferensi yang lemah. Sebaliknya, instrumen dengan struktur faktor yang baik tetapi domain konstruk yang tidak lengkap juga bermasalah secara konseptual. Oleh karena itu, pengujian validitas harus dipahami sebagai proses berkelanjutan yang bertujuan menjaga akurasi, relevansi, dan integritas inferensi ilmiah dalam penelitian berbasis bukti.

5.3 Analisis Faktor (Exploratory Factor Analysis / EFA)

Analisis Faktor Eksploratori (Exploratory Factor Analysis/EFA) merupakan teknik statistik multivariat yang digunakan untuk mengidentifikasi struktur laten dari seperangkat item dalam suatu instrumen pengukuran. EFA bertujuan mengungkap dimensi atau faktor yang mendasari hubungan antar item tanpa menetapkan model struktur sebelumnya. Menurut Hair et al. (2019), EFA sangat relevan dalam tahap awal pengembangan instrumen ketika struktur konstruk belum

sepenuhnya terkonseptualisasi secara empiris. Dalam konteks penelitian pendidikan, psikologi, dan sosial, EFA berfungsi sebagai alat utama untuk memberikan bukti validitas konstruk melalui analisis struktur internal instrumen.

Sebelum melakukan EFA, peneliti harus memastikan bahwa data memenuhi asumsi kelayakan analisis faktor. Salah satu indikator utama adalah ukuran sampel yang memadai. Menurut Tabachnick dan Fidell (2019), ukuran sampel minimal yang disarankan adalah 5–10 responden per item, meskipun Hair et al. (2019) menekankan bahwa kualitas korelasi antar item lebih penting daripada sekadar jumlah sampel. Selain ukuran sampel, kelayakan EFA juga diuji melalui Kaiser–Meyer–Olkin Measure of Sampling Adequacy (KMO) dan Bartlett's Test of Sphericity. Field (2018) menjelaskan bahwa nilai KMO di atas 0,60 menunjukkan data layak dianalisis secara faktor, sedangkan Bartlett's Test yang signifikan menandakan bahwa matriks korelasi antar item tidak bersifat identitas.

Setelah asumsi terpenuhi, tahap berikutnya adalah pemilihan metode ekstraksi faktor. Metode ekstraksi bertujuan mengidentifikasi faktor laten yang menjelaskan varians bersama (*shared variance*) antar item. Menurut Hair et al. (2019), metode yang paling umum digunakan dalam EFA adalah *Principal Axis Factoring* (PAF) dan *Maximum Likelihood* (ML). PAF lebih disarankan ketika data tidak sepenuhnya berdistribusi normal, sedangkan ML memungkinkan pengujian statistik lanjutan dan cocok digunakan ketika asumsi normalitas terpenuhi. Field (2018) menegaskan bahwa penggunaan *Principal Component Analysis* (PCA) sebaiknya dibedakan secara konseptual dari EFA karena PCA bertujuan reduksi data, bukan identifikasi konstruk laten.

Tahap selanjutnya adalah penentuan jumlah faktor yang diekstraksi. Penentuan jumlah faktor merupakan keputusan kritis

karena akan memengaruhi interpretasi konstruk. Menurut Kaiser's criterion, faktor dengan nilai eigenvalue lebih besar dari satu dianggap layak dipertahankan. Namun, Hair et al. (2019) memperingatkan bahwa kriteria ini sering kali menghasilkan jumlah faktor yang berlebihan. Oleh karena itu, peneliti disarankan menggunakan pendekatan tambahan seperti *scree plot* dan *parallel analysis*. Field (2018) menekankan bahwa *parallel analysis* merupakan metode yang lebih akurat karena membandingkan eigenvalue empiris dengan eigenvalue data acak, sehingga mengurangi risiko overfitting struktur faktor.

Setelah jumlah faktor ditentukan, peneliti melakukan rotasi faktor untuk memperoleh struktur faktor yang lebih mudah diinterpretasikan. Rotasi bertujuan memaksimalkan *factor loading* item pada satu faktor dan meminimalkan *cross-loading* pada faktor lain. Menurut Tabachnick dan Fidell (2019), rotasi faktor dapat bersifat ortogonal (misalnya Varimax) atau oblique (misalnya Oblimin dan Promax). Rotasi ortogonal mengasumsikan bahwa faktor tidak berkorelasi, sedangkan rotasi oblique mengizinkan korelasi antar faktor. Dalam penelitian psikologi dan pendidikan, di mana konstruk sering kali saling berkaitan, Hair et al. (2019) menyarankan penggunaan rotasi oblique karena lebih realistis secara teoretis.

Interpretasi hasil EFA dilakukan dengan menganalisis nilai factor loading setiap item. Factor loading menunjukkan kekuatan hubungan antara item dan faktor laten. Menurut Hair et al. (2019), nilai loading minimal 0,40 umumnya dianggap memadai, meskipun standar ini dapat disesuaikan dengan ukuran sampel dan tujuan penelitian. Item dengan *cross-loading* tinggi atau loading rendah perlu dipertimbangkan untuk direvisi atau dihapus. Field (2018) menegaskan bahwa keputusan penghapusan item tidak boleh semata-mata berbasis statistik,

tetapi harus mempertimbangkan relevansi teoretis item terhadap konstruk.

EFA juga berperan penting dalam penyempurnaan instrumen sebelum dilakukan pengujian lanjutan. Item-item yang tidak sesuai dengan struktur faktor dapat dieliminasi untuk meningkatkan homogenitas skala dan kejelasan konstruk. Menurut DeVellis (2017), proses ini merupakan bagian integral dari pengembangan skala yang bertujuan mencapai keseimbangan antara validitas konstruk dan reliabilitas internal. Instrumen yang lolos tahap EFA diharapkan memiliki struktur faktor yang stabil dan koheren, sehingga layak diuji lebih lanjut menggunakan analisis faktor konfirmatori (CFA).

Perbedaan antara EFA dan CFA perlu dipahami secara metodologis. EFA bersifat eksploratif dan digunakan ketika struktur faktor belum diketahui secara pasti, sedangkan CFA bersifat konfirmatori dan digunakan untuk menguji kesesuaian model faktor yang telah ditetapkan berdasarkan teori. Menurut Kline (2016), CFA memberikan bukti validitas konstruk yang lebih kuat karena memungkinkan pengujian hubungan laten secara eksplisit serta evaluasi *model fit* melalui berbagai indeks kesesuaian. Oleh karena itu, EFA sering diposisikan sebagai tahap awal dalam rangkaian pengujian validitas konstruk yang berkelanjutan.

Dalam era *big data* dan kemajuan perangkat lunak statistik seperti SPSS, Amos, LISREL, dan Mplus, penerapan EFA menjadi lebih mudah secara teknis namun menuntut ketelitian interpretatif yang lebih tinggi. Tabachnick dan Fidell (2019) menekankan bahwa hasil EFA harus ditafsirkan secara kritis dan tidak dilepaskan dari kerangka teoretis yang mendasarinya. EFA yang dilakukan tanpa pemahaman konseptual berisiko

menghasilkan struktur faktor yang bersifat artifisial dan tidak bermakna secara substantif.

Secara reflektif, analisis faktor eksploratori merupakan alat penting dalam menjaga validitas konstruk instrumen penelitian. EFA membantu peneliti memahami struktur laten data, menyempurnakan item, dan membangun dasar empiris bagi pengujian lanjutan. Namun, EFA bukan tujuan akhir, melainkan bagian dari proses pengembangan instrumen yang berkelanjutan. Ketepatan penggunaan EFA akan menentukan kualitas pengukuran dan kekuatan inferensi ilmiah dalam penelitian pendidikan, psikologi, dan sosial.

5.4 Uji Coba dan Validasi Ahli

Uji coba instrumen dan validasi ahli merupakan tahap krusial dalam pengembangan instrumen penelitian karena berfungsi menjembatani antara rancangan teoretis dan kinerja empiris instrumen di lapangan. Tahap ini memastikan bahwa instrumen tidak hanya valid secara konseptual, tetapi juga dapat dipahami, digunakan, dan direspons secara tepat oleh populasi sasaran. Menurut DeVellis (2017), uji coba dan validasi ahli merupakan proses iteratif yang memungkinkan peneliti mengidentifikasi kelemahan item sebelum instrumen digunakan pada penelitian utama. Dengan demikian, tahap ini berperan strategis dalam meminimalkan kesalahan pengukuran dan meningkatkan kualitas inferensi ilmiah.

Validasi ahli (*expert judgment*) biasanya dilakukan sebelum uji coba lapangan secara luas. Validasi ini bertujuan menilai kesesuaian item dengan konstruk, kejelasan redaksi, relevansi indikator, serta potensi bias konten. Menurut Anastasi dan Urbina (1997), validasi ahli merupakan bentuk evaluasi substantif yang tidak dapat digantikan oleh analisis statistik, karena berfokus

pada rasional teoretis dan makna psikologis item. Para ahli yang dilibatkan idealnya memiliki kompetensi pada bidang substansi konstruk, metodologi penelitian, dan psikometri, sehingga penilaian yang diberikan bersifat komprehensif dan kredibel.

Dalam praktik penelitian kuantitatif modern, validasi ahli sering dikombinasikan dengan pendekatan kuantitatif melalui Content Validity Index (CVI). Menurut Polit dan Beck, CVI memungkinkan peneliti mengukur tingkat kesepakatan antar ahli mengenai relevansi item terhadap konstruk. I-CVI menunjukkan proporsi ahli yang menilai item sebagai relevan, sedangkan S-CVI mencerminkan tingkat validitas isi skala secara keseluruhan. DeVellis (2017) menegaskan bahwa penggunaan CVI memperkuat objektivitas validasi ahli karena mengurangi subjektivitas penilaian individual. Item dengan nilai I-CVI rendah perlu direvisi atau dieliminasi sebelum instrumen dilanjutkan ke tahap uji coba empiris.

Setelah melalui validasi ahli, instrumen memasuki tahap uji coba awal (pilot study). Uji coba ini dilakukan pada sampel yang memiliki karakteristik serupa dengan populasi penelitian utama, meskipun dengan ukuran sampel yang lebih kecil. Menurut Cohen, Manion, dan Morrison (2018), tujuan utama pilot study bukan untuk menguji hipotesis, melainkan untuk mengevaluasi kelayakan instrumen secara praktis dan empiris. Uji coba memungkinkan peneliti mengidentifikasi masalah teknis seperti item yang sulit dipahami, instruksi yang ambigu, waktu pengisian yang terlalu lama, serta respons ekstrem atau tidak konsisten.

Dalam konteks psikometri, uji coba awal juga berfungsi untuk memperoleh data awal yang digunakan dalam analisis reliabilitas dan validitas empiris. Nunnally dan Bernstein (1994) menjelaskan bahwa data pilot dapat digunakan untuk menghitung reliabilitas internal, seperti koefisien Cronbach’s

Alpha, serta korelasi item–total. Item dengan korelasi rendah atau kontribusi negatif terhadap reliabilitas skala perlu ditinjau ulang. Namun demikian, keputusan penghapusan item tidak boleh semata-mata berbasis angka statistik, melainkan harus mempertimbangkan relevansi teoretis dan hasil validasi ahli sebelumnya.

Uji coba instrumen juga memberikan kesempatan untuk menilai respons kognitif responden terhadap item. Menurut Tourangeau, proses respons mencakup pemahaman item, penarikan informasi dari memori, pembentukan penilaian, dan pemilihan respons. Apabila item tidak dipahami sesuai maksud peneliti, maka skor yang dihasilkan tidak lagi merefleksikan konstruk yang diukur. Oleh karena itu, beberapa peneliti merekomendasikan penggunaan teknik tambahan seperti *cognitive interviewing* atau diskusi terbatas dengan responden pilot untuk memperoleh umpan balik kualitatif mengenai kejelasan item.

Setelah uji coba dilakukan, peneliti melakukan revisi instrumen berdasarkan temuan validasi ahli dan hasil analisis pilot. Revisi dapat berupa perbaikan redaksi item, penghapusan item yang tidak valid atau tidak reliabel, penambahan item untuk memperkuat dimensi tertentu, serta penyempurnaan instruksi dan format respons. DeVellis (2017) menekankan bahwa revisi instrumen merupakan bagian alami dari pengembangan skala dan tidak boleh dianggap sebagai kelemahan desain awal. Instrumen yang direvisi secara sistematis justru menunjukkan proses ilmiah yang reflektif dan berbasis bukti.

Tahap uji coba dan validasi ahli juga memiliki implikasi penting terhadap validitas konstruk instrumen. Item yang lolos validasi ahli dan menunjukkan kinerja empiris yang baik dalam uji coba memberikan bukti awal bahwa instrumen mengukur

konstruk sesuai dengan definisi teoretisnya. Menurut Messick (1995), bukti validitas konstruk diperoleh melalui akumulasi berbagai sumber, termasuk rasional teoretis, struktur internal, dan konsistensi respons. Oleh karena itu, uji coba dan validasi ahli tidak berdiri sendiri, melainkan terintegrasi dengan analisis faktor dan uji reliabilitas dalam keseluruhan proses pengembangan instrumen.

Dalam era digital dan *evidence-based research*, uji coba instrumen semakin dimudahkan oleh platform survei daring dan perangkat lunak statistik. Namun, kemudahan teknis ini tidak mengurangi tuntutan ketelitian metodologis. Field (2018) mengingatkan bahwa data pilot yang dikumpulkan secara daring tetap harus dievaluasi secara kritis, terutama terkait kualitas respons, missing data, dan pola jawaban ekstrem. Peneliti harus memastikan bahwa data yang digunakan dalam uji coba benar-benar mencerminkan respons autentik responden, bukan artefak teknis atau kesalahan desain instrumen.

Secara reflektif, uji coba dan validasi ahli merupakan tahap penentu dalam menjaga kualitas instrumen penelitian. Tahap ini memastikan bahwa instrumen tidak hanya valid secara teoretis, tetapi juga fungsional secara empiris. Instrumen yang melewati uji coba dan validasi ahli dengan baik memiliki peluang lebih besar untuk menghasilkan data yang reliabel, valid, dan bermakna. Sebaliknya, pengabaian tahap ini berpotensi menghasilkan instrumen yang cacat secara konseptual maupun teknis, sehingga melemahkan keseluruhan proses penelitian. Oleh karena itu, uji coba dan validasi ahli harus diposisikan sebagai investasi metodologis yang esensial dalam penelitian kuantitatif berbasis bukti.

5.5 Kesalahan Pengukuran

Kesalahan pengukuran (*measurement error*) merupakan isu fundamental dalam penelitian kuantitatif karena memengaruhi ketepatan skor, kekuatan analisis statistik, dan validitas inferensi ilmiah. Setiap skor hasil pengukuran pada dasarnya merupakan kombinasi antara skor sejati (*true score*) dan kesalahan pengukuran. Menurut Nunnally dan Bernstein (1994), tidak ada instrumen pengukuran yang sepenuhnya bebas dari kesalahan; yang dapat dilakukan peneliti adalah meminimalkan dan mengendalikan kesalahan tersebut agar interpretasi hasil penelitian tetap sahih. Oleh karena itu, pemahaman mengenai sumber dan implikasi kesalahan pengukuran menjadi prasyarat penting dalam pengembangan instrumen penelitian pendidikan, psikologi, dan sosial.

Konsep kesalahan pengukuran secara klasik dijelaskan dalam Classical Test Theory (CTT). Dalam CTT, skor observasi (*observed score*) dipandang sebagai hasil penjumlahan antara skor sejati dan kesalahan pengukuran, sebagaimana dirumuskan oleh Crocker dan Algina (2006). Skor sejati merepresentasikan nilai sebenarnya dari atribut yang diukur, sedangkan kesalahan pengukuran mencakup seluruh variasi skor yang tidak berkaitan dengan konstruk. Variasi ini dapat bersifat acak (*random error*) maupun sistematis (*systematic error*). Kesalahan acak cenderung tidak terprediksi dan berubah-ubah antar pengukuran, sedangkan kesalahan sistematis bersumber dari bias instrumen, prosedur, atau karakteristik responden.

Sumber kesalahan pengukuran dapat berasal dari berbagai aspek. Anastasi dan Urbina (1997) mengelompokkan sumber kesalahan ke dalam faktor instrumen, faktor responden, faktor administrasi, dan faktor situasional. Faktor instrumen mencakup item yang ambigu, redaksi yang tidak jelas, atau format respons

yang tidak sesuai. Faktor responden berkaitan dengan kondisi psikologis, motivasi, kelelahan, atau kecenderungan memberikan jawaban sosial yang diharapkan (*social desirability bias*). Faktor administrasi meliputi instruksi yang tidak konsisten, waktu pengisian yang tidak memadai, atau gangguan lingkungan saat pengumpulan data. Sementara itu, faktor situasional mencakup konteks sosial dan budaya yang memengaruhi cara responden menafsirkan item.

Kesalahan pengukuran memiliki implikasi langsung terhadap reliabilitas instrumen. Menurut Nunnally dan Bernstein (1994), reliabilitas mencerminkan proporsi varians skor yang disebabkan oleh skor sejati relatif terhadap varians total. Semakin besar kesalahan pengukuran, semakin rendah reliabilitas instrumen. Koefisien reliabilitas seperti Cronbach's Alpha digunakan untuk mengestimasi konsistensi internal instrumen, namun nilai reliabilitas yang tinggi tidak secara otomatis menjamin validitas. DeVellis (2017) menegaskan bahwa reliabilitas merupakan syarat perlu tetapi tidak cukup untuk validitas; instrumen dapat reliabel tetapi mengukur konstruk yang salah secara konsisten.

Kesalahan pengukuran juga berdampak signifikan terhadap validitas inferensi statistik. Menurut Trochim (2006), kesalahan pengukuran yang tinggi akan melemahkan korelasi antar variabel, menurunkan kekuatan uji statistik (*statistical power*), dan meningkatkan risiko kesimpulan yang keliru. Dalam penelitian korelasional atau regresi, kesalahan pengukuran dapat menyebabkan *attenuation*, yaitu hubungan antar variabel tampak lebih lemah daripada hubungan sebenarnya. Akibatnya, peneliti berpotensi menolak hipotesis yang sebenarnya benar (*Type II error*). Oleh karena itu, pengendalian kesalahan

pengukuran merupakan aspek krusial dalam menjaga akurasi hasil penelitian.

Dalam konteks validitas konstruk, Messick (1995) menekankan bahwa kesalahan pengukuran harus dipertimbangkan sebagai bagian dari evaluasi konsekuensi penggunaan skor. Skor yang dihasilkan dari instrumen dengan tingkat kesalahan tinggi berpotensi menghasilkan keputusan yang tidak adil atau tidak akurat, terutama ketika instrumen digunakan untuk tujuan evaluatif atau selektif. Oleh karena itu, analisis kesalahan pengukuran tidak hanya bersifat teknis, tetapi juga memiliki dimensi etis dan praktis dalam penelitian dan praktik pendidikan.

Pendekatan statistik lanjutan memberikan alternatif untuk memahami dan mengendalikan kesalahan pengukuran secara lebih eksplisit. Dalam kerangka Structural Equation Modeling (SEM), kesalahan pengukuran dimodelkan secara langsung melalui *error variance* pada setiap indikator. Menurut Kline (2016), SEM memungkinkan peneliti memisahkan varians sejati dan varians kesalahan, sehingga estimasi hubungan antar konstruk menjadi lebih akurat. Pendekatan ini merupakan pengembangan dari CTT dan sering digunakan dalam analisis faktor konfirmatori (CFA) sebagai bentuk validitas konstruk lanjutan.

Selain SEM, Item Response Theory (IRT) juga menawarkan perspektif berbeda terhadap kesalahan pengukuran. Embretson dan Reise (2000) menjelaskan bahwa dalam IRT, kesalahan pengukuran tidak dianggap konstan di seluruh rentang skor, melainkan bervariasi tergantung pada tingkat kemampuan responden. IRT memungkinkan estimasi *standard error of measurement* pada setiap level kemampuan, sehingga memberikan informasi yang lebih rinci mengenai presisi

pengukuran. Meskipun lebih kompleks, pendekatan ini semakin relevan dalam penelitian berbasis data besar dan asesmen berbantuan komputer.

Upaya meminimalkan kesalahan pengukuran harus dilakukan sejak tahap awal pengembangan instrumen. Menurut DeVellis (2017), strategi utama meliputi pendefinisian konstruk yang jelas, penyusunan item yang berkualitas, validasi ahli, uji coba instrumen, serta analisis empiris yang ketat. Selain itu, konsistensi prosedur pengumpulan data, pelatihan enumerator, dan pemanfaatan teknologi survei yang andal juga berkontribusi dalam mengurangi kesalahan non-sampling. Peneliti harus menyadari bahwa pengendalian kesalahan pengukuran merupakan proses berkelanjutan, bukan satu tahap terpisah.

Secara reflektif, kesalahan pengukuran merupakan kenyataan inheren dalam penelitian kuantitatif yang tidak dapat dihilangkan sepenuhnya. Namun, pemahaman teoretis dan metodologis yang memadai memungkinkan peneliti mengelola dan meminimalkan dampaknya. Instrumen yang dikembangkan dengan mempertimbangkan sumber dan implikasi kesalahan pengukuran akan menghasilkan data yang lebih presisi dan inferensi yang lebih dapat dipercaya. Oleh karena itu, analisis kesalahan pengukuran harus diposisikan sebagai bagian integral dari evaluasi kualitas instrumen dan sebagai fondasi bagi penelitian berbasis bukti yang akurat dan bertanggung jawab.

Daftar Rujukan

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Cengage Learning.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

Trochim, W. M. (2006). *The research methods knowledge base* (2nd ed.). Atomic Dog Publishing.

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 45–79). Elsevier.

BAB 6

RELIABILITAS DAN ANALISIS STATISTIK DASAR

6.1 Konsep Reliabilitas

Reliabilitas merupakan konsep fundamental dalam penelitian kuantitatif karena berkaitan langsung dengan tingkat konsistensi dan keandalan hasil pengukuran. Instrumen yang reliabel akan menghasilkan skor yang relatif stabil dan konsisten apabila digunakan berulang kali pada kondisi yang sebanding. Dalam konteks penelitian pendidikan, sosial, dan perilaku, reliabilitas menjadi prasyarat utama sebelum data dianalisis lebih lanjut menggunakan teknik statistik inferensial. Menurut Gravetter dan Wallnau (2016), reliabilitas menunjukkan sejauh mana suatu alat ukur terbebas dari kesalahan pengukuran acak (*random error*). Tanpa reliabilitas yang memadai, hasil analisis statistik kehilangan makna karena varians skor lebih banyak dipengaruhi oleh kesalahan daripada konstruk yang diukur.

Konsep reliabilitas berakar kuat pada Classical Test Theory (CTT) yang memandang skor observasi sebagai kombinasi antara skor sejati (*true score*) dan kesalahan pengukuran (*error score*). Menurut Nunnally dan Bernstein, yang kemudian dijelaskan kembali oleh Cohen (1988), semakin kecil proporsi kesalahan pengukuran dalam skor total, semakin tinggi reliabilitas instrumen tersebut. Dengan demikian, reliabilitas tidak pernah bernilai sempurna, melainkan berada pada suatu rentang probabilistik yang mencerminkan tingkat presisi pengukuran. Prinsip ini menegaskan bahwa reliabilitas bukanlah sifat absolut

instrumen, melainkan karakteristik skor dalam konteks pengukuran tertentu.

Dalam praktik penelitian, reliabilitas sering disalahartikan sebagai jaminan kebenaran pengukuran. Howell (2012) menegaskan bahwa reliabilitas tidak identik dengan validitas. Instrumen dapat menunjukkan reliabilitas tinggi namun tetap tidak valid apabila secara konsisten mengukur konstruk yang salah. Namun demikian, reliabilitas tetap menjadi syarat perlu bagi validitas, karena instrumen yang tidak reliabel hampir pasti tidak valid. Oleh karena itu, evaluasi reliabilitas harus dipandang sebagai langkah awal yang krusial sebelum membahas validitas konstruk dan analisis hubungan antarvariabel.

Jenis reliabilitas dapat diklasifikasikan berdasarkan sumber konsistensi yang diukur. Menurut Hinkle, Wiersma, dan Jurs (2003), reliabilitas dapat mencakup reliabilitas konsistensi internal, reliabilitas stabilitas waktu, dan reliabilitas ekuivalensi bentuk. Reliabilitas konsistensi internal berkaitan dengan sejauh mana item-item dalam suatu instrumen mengukur konstruk yang sama. Reliabilitas stabilitas waktu berkaitan dengan konsistensi skor dari waktu ke waktu, sedangkan reliabilitas ekuivalensi bentuk berkaitan dengan kesetaraan dua versi instrumen. Masing-masing jenis reliabilitas memiliki implikasi metodologis yang berbeda dan harus dipilih sesuai dengan tujuan penelitian.

Dalam penelitian pendidikan dan sosial, reliabilitas konsistensi internal merupakan bentuk yang paling sering digunakan, terutama pada instrumen berbentuk skala sikap, persepsi, atau kuesioner psikologis. Menurut George dan Mallery (2020), reliabilitas konsistensi internal memberikan informasi penting tentang homogenitas item dalam suatu skala. Skala yang baik diharapkan memiliki item-item yang saling berkorelasi secara moderat karena mengukur aspek yang sama, namun tidak

bersifat redundan. Korelasi item yang terlalu tinggi justru dapat mengindikasikan pengulangan makna yang tidak efisien secara pengukuran.

Nilai reliabilitas biasanya dinyatakan dalam bentuk koefisien yang berkisar antara 0 hingga 1. Cohen (1988) menjelaskan bahwa interpretasi koefisien reliabilitas harus mempertimbangkan konteks penelitian dan tujuan penggunaan instrumen. Dalam penelitian eksploratif, nilai reliabilitas sekitar 0,60 masih dapat diterima, sedangkan dalam penelitian konfirmatori atau evaluatif, nilai reliabilitas minimal 0,70 atau bahkan 0,80 sering dijadikan standar. Namun demikian, angka-angka ini bukanlah aturan mutlak, melainkan pedoman umum yang harus dikombinasikan dengan pertimbangan substantif.

Reliabilitas juga memiliki implikasi langsung terhadap daya uji statistik (*statistical power*). Menurut Cohen (1988), rendahnya reliabilitas akan menyebabkan pelemahan hubungan antarvariabel (*attenuation*), sehingga koefisien korelasi atau efek perlakuan tampak lebih kecil daripada kondisi sebenarnya. Akibatnya, peneliti berisiko gagal mendeteksi efek yang sesungguhnya ada (*Type II error*). Oleh karena itu, peningkatan reliabilitas instrumen bukan hanya persoalan kualitas pengukuran, tetapi juga strategi untuk meningkatkan sensitivitas analisis statistik.

Dalam konteks penggunaan perangkat lunak statistik modern seperti SPSS, Jamovi, JASP, dan R, pengujian reliabilitas menjadi lebih mudah secara teknis. Namun, Field (2018) mengingatkan bahwa kemudahan perhitungan koefisien reliabilitas tidak boleh mengaburkan pemahaman konseptual peneliti. Reliabilitas harus ditafsirkan secara kritis dengan mempertimbangkan struktur instrumen, karakteristik sampel, dan tujuan analisis. Reliabilitas yang dihitung pada satu sampel

tidak otomatis berlaku pada sampel lain, sehingga pelaporan reliabilitas harus selalu dikaitkan dengan konteks data yang digunakan.

Secara metodologis, reliabilitas juga berkaitan erat dengan kesalahan pengukuran dan asumsi statistik dalam analisis lanjutan. Hair et al. (2019) menegaskan bahwa analisis multivariat yang kompleks, seperti regresi atau analisis faktor, sangat sensitif terhadap kualitas pengukuran awal. Instrumen dengan reliabilitas rendah akan menghasilkan estimasi parameter yang bias dan menurunkan kepercayaan terhadap hasil analisis. Oleh karena itu, reliabilitas harus diposisikan sebagai fondasi analisis statistik, bukan sekadar prosedur administratif dalam pelaporan hasil penelitian.

Secara reflektif, reliabilitas merupakan penjaga awal integritas penelitian kuantitatif. Ia memastikan bahwa variasi skor yang dianalisis benar-benar merefleksikan perbedaan individu atau fenomena yang diteliti, bukan fluktuasi acak akibat kesalahan pengukuran. Peneliti yang mengabaikan reliabilitas berisiko membangun kesimpulan di atas data yang rapuh secara metodologis. Sebaliknya, perhatian serius terhadap reliabilitas akan memperkuat validitas inferensi dan meningkatkan kredibilitas hasil penelitian pendidikan, sosial, dan perilaku.

6.2 Cronbach Alpha, Split-Half, dan Test–Retest

Pengujian reliabilitas instrumen dalam penelitian kuantitatif umumnya dilakukan melalui beberapa pendekatan statistik yang dirancang untuk mengestimasi tingkat konsistensi skor. Tiga teknik yang paling banyak digunakan adalah Cronbach's Alpha, Split-Half Reliability, dan Test–Retest Reliability. Ketiga pendekatan ini memiliki dasar teoretis yang sama dalam Classical Test Theory (CTT), namun berbeda dalam cara mengestimasi

proporsi varians skor sejati terhadap varians total. Menurut Howell (2012), pemilihan teknik reliabilitas harus disesuaikan dengan karakteristik instrumen, tujuan pengukuran, dan desain penelitian.

Pendekatan yang paling populer dalam penelitian pendidikan dan sosial adalah Cronbach's Alpha (α). Koefisien ini digunakan untuk mengestimasi reliabilitas konsistensi internal, yaitu sejauh mana item-item dalam suatu skala saling berkorelasi dan secara bersama-sama mengukur konstruk yang sama. Menurut George dan Mallery (2020), Cronbach's Alpha sangat sesuai digunakan pada instrumen berbentuk skala Likert dengan lebih dari dua kategori respons. Secara matematis, α merefleksikan rasio varians sejati terhadap varians total skor, sehingga semakin tinggi nilai α, semakin kecil kontribusi kesalahan pengukuran acak.

Interpretasi nilai Cronbach's Alpha harus dilakukan secara hati-hati. Cohen (1988) menjelaskan bahwa nilai α sekitar 0,70 umumnya dianggap memadai untuk penelitian eksploratif, sedangkan nilai di atas 0,80 lebih diharapkan dalam penelitian konfirmatori atau evaluatif. Namun, nilai α yang sangat tinggi (misalnya di atas 0,95) justru dapat mengindikasikan redundansi item, yaitu adanya item-item yang terlalu mirip secara substansi. Field (2018) menegaskan bahwa tujuan pengujian reliabilitas bukan sekadar mengejar nilai α setinggi mungkin, melainkan memastikan keseimbangan antara homogenitas dan cakupan konstruk.

Selain melihat nilai α secara keseluruhan, peneliti juga perlu memperhatikan korelasi item–total dan informasi *"Cronbach's Alpha if Item Deleted"*. Menurut Gravetter dan Wallnau (2016), item dengan korelasi item–total yang rendah menunjukkan bahwa item tersebut tidak berkontribusi secara optimal terhadap

konstruk yang diukur. Dalam praktik analisis menggunakan SPSS, Jamovi, atau JASP, item dengan korelasi di bawah 0,30 sering dipertimbangkan untuk direvisi atau dihapus. Namun, keputusan ini tidak boleh semata-mata berbasis angka statistik, melainkan harus mempertimbangkan relevansi teoretis item tersebut.

Pendekatan kedua adalah Split-Half Reliability, yang mengukur reliabilitas dengan membagi instrumen menjadi dua bagian yang setara, kemudian menghitung korelasi antara skor kedua bagian tersebut. Menurut Hinkle, Wiersma, dan Jurs (2003), teknik ini bertujuan mengestimasi konsistensi internal instrumen tanpa harus mengadministrasikan tes lebih dari satu kali. Namun, reliabilitas split-half sangat bergantung pada cara pembagian item. Pembagian yang tidak seimbang dapat menghasilkan estimasi reliabilitas yang bias.

Untuk mengatasi keterbatasan tersebut, digunakan Spearman–Brown Prophecy Formula yang berfungsi menyesuaikan koefisien korelasi split-half agar merefleksikan reliabilitas instrumen secara keseluruhan. Menurut Howell (2012), rumus Spearman–Brown memungkinkan peneliti memprediksi perubahan reliabilitas apabila jumlah item ditambah atau dikurangi. Dalam konteks pengembangan instrumen, formula ini sangat berguna untuk mengevaluasi apakah penambahan item tertentu berpotensi meningkatkan reliabilitas skala secara signifikan.

Pendekatan ketiga adalah Test–Retest Reliability, yang mengukur stabilitas skor dari waktu ke waktu. Teknik ini dilakukan dengan mengadministrasikan instrumen yang sama kepada responden yang sama dalam dua waktu yang berbeda, kemudian menghitung korelasi antara skor pada waktu pertama dan kedua. Menurut Gravetter dan Wallnau (2016), reliabilitas test–retest sangat relevan untuk konstruk yang relatif stabil, seperti

kemampuan kognitif atau sikap dasar. Sebaliknya, untuk konstruk yang mudah berubah, seperti emosi sesaat atau motivasi situasional, nilai reliabilitas test–retest yang rendah tidak selalu menunjukkan kelemahan instrumen.

Penentuan interval waktu antara dua pengukuran merupakan aspek krusial dalam reliabilitas test–retest. Howell (2012) menegaskan bahwa interval yang terlalu pendek berisiko menghasilkan efek ingatan (*memory effect*), sedangkan interval yang terlalu panjang berpotensi mencerminkan perubahan konstruk yang sebenarnya. Oleh karena itu, peneliti harus menyesuaikan interval waktu dengan sifat konstruk yang diukur serta konteks penelitian lapangan.

Dalam praktik analisis menggunakan perangkat lunak statistik modern, ketiga teknik reliabilitas ini dapat dihitung dengan relatif mudah. SPSS menyediakan fasilitas *Reliability Analysis* untuk Cronbach's Alpha dan Split-Half, sementara Jamovi dan JASP menawarkan antarmuka yang lebih visual dan intuitif. Namun, Field (2018) menekankan bahwa kemudahan teknis tidak boleh menggantikan pemahaman konseptual. Peneliti harus mampu menjelaskan makna koefisien reliabilitas, implikasinya terhadap kualitas data, serta keterbatasan setiap pendekatan yang digunakan.

Reliabilitas juga memiliki implikasi langsung terhadap analisis statistik lanjutan. Hair et al. (2019) menjelaskan bahwa reliabilitas yang rendah akan menyebabkan *attenuation* pada korelasi dan koefisien regresi, sehingga hubungan antarvariabel tampak lebih lemah dari kondisi sebenarnya. Dalam konteks analisis multivariat, reliabilitas yang buruk dapat menghasilkan estimasi parameter yang bias dan menurunkan validitas model. Oleh karena itu, pengujian reliabilitas harus dilakukan sebelum

analisis inferensial, bukan sebagai pelengkap laporan hasil penelitian.

Secara reflektif, Cronbach's Alpha, Split-Half, dan Test–Retest bukanlah teknik yang saling menggantikan, melainkan saling melengkapi. Cronbach's Alpha memberikan gambaran konsistensi internal, Split-Half menawarkan estimasi alternatif berbasis pembagian item, dan Test–Retest mengukur stabilitas temporal. Kombinasi ketiganya, apabila memungkinkan, akan memberikan gambaran reliabilitas instrumen yang lebih komprehensif. Peneliti yang memahami perbedaan dan keterbatasan masing-masing teknik akan mampu memilih pendekatan reliabilitas yang paling sesuai dengan tujuan dan konteks penelitiannya.

6.3 Statistik Deskriptif

Statistik deskriptif merupakan tahap awal yang sangat penting dalam analisis data kuantitatif karena berfungsi menggambarkan karakteristik dasar data sebelum dilakukan pengujian inferensial. Statistik deskriptif membantu peneliti memahami pola distribusi, kecenderungan pusat, dan tingkat variasi data yang dikumpulkan. Menurut Gravetter dan Wallnau (2016), statistik deskriptif tidak bertujuan untuk menguji hipotesis, melainkan menyajikan ringkasan data secara sistematis agar peneliti dapat melakukan interpretasi awal yang bermakna. Dalam penelitian pendidikan, sosial, dan perilaku, statistik deskriptif menjadi fondasi untuk menentukan kelayakan data terhadap asumsi analisis statistik lanjutan.

Ukuran pemusatan data (*measures of central tendency*) merupakan komponen utama statistik deskriptif yang mencakup mean, median, dan modus. Mean atau rata-rata aritmetika merupakan ukuran pemusatan yang paling sering digunakan

karena mempertimbangkan seluruh nilai dalam data. Menurut Howell (2012), mean sangat sensitif terhadap nilai ekstrem (*outliers*), sehingga interpretasinya harus dilakukan dengan hati-hati, terutama pada data yang berdistribusi tidak normal. Dalam penelitian pendidikan, mean sering digunakan untuk menggambarkan tingkat pencapaian belajar, sikap, atau persepsi responden terhadap suatu variabel.

Median merupakan nilai tengah dalam distribusi data yang telah diurutkan, sedangkan modus adalah nilai yang paling sering muncul. Gravetter dan Wallnau (2016) menjelaskan bahwa median lebih stabil dibandingkan mean ketika data mengandung nilai ekstrem atau distribusi yang menceng. Oleh karena itu, pada data ordinal atau data interval yang tidak berdistribusi normal, median sering menjadi ukuran pemusatan yang lebih representatif. Modus, meskipun jarang digunakan dalam analisis lanjutan, berguna untuk mengidentifikasi kecenderungan kategori dominan, terutama pada data kategorikal atau skala nominal.

Selain ukuran pemusatan, statistik deskriptif juga mencakup ukuran dispersi yang menggambarkan seberapa besar variasi data. Ukuran dispersi yang paling umum digunakan adalah standar deviasi (SD) dan varian. Menurut Field (2018), standar deviasi menunjukkan sejauh mana skor individu menyimpang dari mean. Nilai SD yang kecil menunjukkan bahwa data relatif homogen, sedangkan nilai SD yang besar mengindikasikan keragaman respons yang tinggi. Dalam penelitian sosial, standar deviasi sering digunakan untuk menilai heterogenitas sikap atau persepsi responden terhadap suatu fenomena.

Pemahaman terhadap dispersi data sangat penting karena berkaitan dengan interpretasi hasil analisis inferensial. Cohen (1988) menegaskan bahwa varians yang terlalu kecil dapat

membatasi kemampuan peneliti mendeteksi perbedaan atau hubungan antarvariabel, sedangkan varians yang terlalu besar dapat mengindikasikan adanya subkelompok yang tidak terkontrol dalam sampel. Oleh karena itu, analisis deskriptif harus dilakukan sebelum pengujian hipotesis untuk mengidentifikasi potensi masalah dalam data.

Statistik deskriptif juga mencakup ukuran bentuk distribusi, yaitu skewness dan kurtosis. Skewness menunjukkan tingkat kemencengan distribusi data, apakah condong ke kiri atau ke kanan, sedangkan kurtosis menggambarkan tingkat keruncingan distribusi. Menurut George dan Mallery (2020), nilai skewness dan kurtosis antara −2 hingga +2 umumnya dianggap masih dapat diterima untuk asumsi normalitas dalam analisis parametris. Namun, Field (2018) menekankan bahwa interpretasi skewness dan kurtosis harus disesuaikan dengan ukuran sampel dan konteks penelitian, karena pada sampel besar, penyimpangan kecil dari normalitas sering kali tidak berdampak signifikan.

Dalam praktik penelitian, statistik deskriptif biasanya disajikan dalam bentuk tabel dan visualisasi seperti histogram, boxplot, atau diagram batang. Visualisasi data membantu peneliti mengenali pola distribusi, nilai ekstrem, dan potensi kesalahan entri data. Menurut Dancey dan Reidy (2017), visualisasi data merupakan langkah penting dalam *data screening* yang sering diabaikan oleh peneliti pemula. Dengan melihat grafik distribusi, peneliti dapat segera mengidentifikasi apakah data mendekati distribusi normal atau menunjukkan pola yang tidak lazim.

Penggunaan perangkat lunak statistik modern seperti SPSS, Jamovi, JASP, dan R sangat memudahkan penyajian statistik deskriptif. George dan Mallery (2020) menjelaskan bahwa SPSS menyediakan menu *Descriptive Statistics* yang memungkinkan peneliti memperoleh mean, median, standar deviasi, skewness,

dan kurtosis secara simultan. Jamovi dan JASP menawarkan antarmuka berbasis menu yang lebih intuitif serta visualisasi langsung, sehingga memudahkan interpretasi bagi peneliti pendidikan dan sosial. Namun demikian, kemudahan teknis ini tidak menghilangkan tanggung jawab peneliti untuk memahami makna statistik yang dihasilkan.

Statistik deskriptif juga berfungsi sebagai dasar untuk menentukan teknik analisis inferensial yang tepat. Menurut Sheskin (2011), keputusan penggunaan uji parametris atau nonparametris harus didasarkan pada karakteristik data yang terungkap melalui analisis deskriptif. Data yang sangat menceng, memiliki banyak nilai ekstrem, atau berasal dari skala ordinal mungkin lebih sesuai dianalisis menggunakan teknik nonparametris. Sebaliknya, data yang mendekati distribusi normal dengan varians yang relatif homogen lebih cocok untuk analisis parametris.

Secara reflektif, statistik deskriptif bukan sekadar tahap awal yang bersifat teknis, melainkan proses analitis yang menentukan arah dan kualitas analisis selanjutnya. Statistik deskriptif memungkinkan peneliti memahami data secara menyeluruh, mengidentifikasi potensi masalah, dan membuat keputusan metodologis yang tepat. Tanpa analisis deskriptif yang memadai, pengujian inferensial berisiko menghasilkan kesimpulan yang bias atau tidak valid. Oleh karena itu, statistik deskriptif harus diposisikan sebagai fondasi penting dalam analisis data kuantitatif yang bertanggung jawab dan berbasis bukti.

6.4 Normalitas, Homogenitas, dan Linearitas

Uji asumsi statistik merupakan tahapan penting dalam analisis data kuantitatif karena menentukan kelayakan

penggunaan teknik analisis parametris. Analisis parametris seperti uji *t*, ANOVA, regresi, dan korelasi Pearson mensyaratkan terpenuhinya beberapa asumsi dasar, antara lain normalitas distribusi data, homogenitas varians, dan linearitas hubungan antarvariabel. Menurut Field (2018), pengabaian terhadap asumsi-asumsi ini dapat menghasilkan kesimpulan statistik yang bias dan menurunkan validitas inferensi penelitian. Oleh karena itu, pengujian asumsi statistik harus dilakukan secara sistematis sebelum peneliti melanjutkan ke analisis inferensial.

Asumsi pertama yang paling sering diuji adalah normalitas distribusi data. Normalitas mengacu pada sejauh mana distribusi skor mendekati distribusi normal teoretis. Menurut Gravetter dan Wallnau (2016), banyak prosedur statistik parametris dikembangkan dengan asumsi bahwa data berasal dari populasi yang berdistribusi normal. Meskipun beberapa teknik cukup robust terhadap pelanggaran normalitas, pemeriksaan awal tetap diperlukan untuk memastikan kesesuaian metode analisis. Normalitas dapat diuji melalui pendekatan visual seperti histogram dan Q–Q plot, maupun melalui uji statistik formal.

Uji statistik normalitas yang umum digunakan adalah Kolmogorov–Smirnov (K–S) dan Shapiro–Wilk. Menurut Sheskin (2011), uji Kolmogorov–Smirnov lebih sesuai digunakan pada sampel besar, sedangkan uji Shapiro–Wilk direkomendasikan untuk sampel kecil hingga menengah karena memiliki daya uji yang lebih tinggi. Interpretasi kedua uji ini didasarkan pada nilai signifikansi (*p-value*). Apabila nilai *p* lebih besar dari 0,05, maka data dianggap berdistribusi normal. Namun, Field (2018) mengingatkan bahwa pada sampel besar, uji normalitas sering kali terlalu sensitif sehingga penyimpangan kecil dari normalitas dapat menghasilkan nilai *p* yang signifikan. Oleh karena itu,

peneliti perlu mengombinasikan uji statistik dengan pemeriksaan visual dan ukuran skewness–kurtosis.

Asumsi kedua yang penting adalah homogenitas varians, yaitu kesamaan varians antar kelompok yang dibandingkan. Asumsi ini sangat relevan dalam analisis komparatif seperti uji *t* dan ANOVA. Menurut Howell (2012), pelanggaran homogenitas varians dapat menyebabkan kesalahan estimasi nilai statistik uji dan meningkatkan risiko kesimpulan yang tidak akurat. Uji yang paling umum digunakan untuk menguji homogenitas varians adalah Levene's Test. Dalam uji ini, nilai signifikansi di atas 0,05 menunjukkan bahwa varians antar kelompok dapat dianggap homogen.

Interpretasi hasil uji Levene harus dilakukan dengan mempertimbangkan konteks penelitian. Field (2018) menjelaskan bahwa ketika asumsi homogenitas varians dilanggar, peneliti tidak harus menghentikan analisis, tetapi dapat menggunakan prosedur alternatif yang lebih robust, seperti uji *t* Welch atau ANOVA Welch. Dengan demikian, uji homogenitas berfungsi sebagai dasar pengambilan keputusan metodologis, bukan sebagai penghalang absolut dalam analisis data.

Asumsi ketiga yang sering diabaikan namun sangat penting adalah linearitas hubungan antarvariabel. Asumsi linearitas menyatakan bahwa hubungan antara variabel independen dan dependen harus bersifat linear agar analisis korelasi Pearson dan regresi linear memberikan hasil yang valid. Menurut Cohen (1988), hubungan yang kuat namun non-linear dapat menghasilkan koefisien korelasi yang rendah atau menyesatkan apabila dianalisis dengan teknik linear. Oleh karena itu, pemeriksaan linearitas menjadi prasyarat penting dalam analisis hubungan.

Linearitas dapat diuji melalui pendekatan visual seperti scatterplot, maupun melalui uji statistik menggunakan tabel ANOVA untuk uji linearitas. George dan Mallery (2020) menjelaskan bahwa dalam SPSS, uji linearitas dapat dilakukan melalui menu *Compare Means* dengan memeriksa signifikansi komponen linear dan deviasi dari linearitas. Apabila komponen linear signifikan dan deviasi dari linearitas tidak signifikan, maka asumsi linearitas terpenuhi. Pendekatan ini membantu peneliti memastikan bahwa hubungan antarvariabel sesuai dengan asumsi model yang digunakan.

Pengujian asumsi normalitas, homogenitas, dan linearitas harus dipahami sebagai bagian dari proses *data screening* yang bertujuan meningkatkan kualitas analisis statistik. Menurut Hair et al. (2019), data screening yang baik memungkinkan peneliti mengidentifikasi masalah sejak dini dan memilih teknik analisis yang paling sesuai dengan karakteristik data. Dalam penelitian pendidikan dan sosial, data lapangan sering kali tidak ideal, sehingga fleksibilitas metodologis menjadi sangat penting.

Penggunaan perangkat lunak statistik modern seperti SPSS, Jamovi, JASP, dan R memudahkan pengujian asumsi statistik ini. George dan Mallery (2020) menekankan bahwa meskipun perangkat lunak menyediakan hasil uji secara otomatis, peneliti tetap harus memahami makna statistik dan implikasi metodologis dari hasil tersebut. Pengambilan keputusan tidak boleh hanya didasarkan pada nilai *p*, tetapi harus mempertimbangkan konteks penelitian, ukuran sampel, dan tujuan analisis.

Secara reflektif, pengujian asumsi statistik merupakan bentuk kehati-hatian metodologis dalam penelitian kuantitatif. Normalitas, homogenitas, dan linearitas bukan sekadar persyaratan teknis, melainkan fondasi bagi validitas inferensi statistik. Peneliti yang memahami dan menerapkan pengujian

asumsi secara tepat akan mampu memilih teknik analisis yang sesuai dan menghindari kesimpulan yang menyesatkan. Sebaliknya, pengabaian terhadap asumsi-asumsi ini dapat melemahkan kredibilitas hasil penelitian, meskipun analisis statistik dilakukan dengan perangkat lunak yang canggih.

6.5 Analisis Hubungan

Analisis hubungan merupakan teknik statistik yang digunakan untuk mengetahui derajat keterkaitan antara dua atau lebih variabel tanpa maksud menarik kesimpulan kausal. Dalam penelitian pendidikan, sosial, dan perilaku, analisis hubungan sering digunakan untuk mengeksplorasi keterkaitan antara sikap, persepsi, motivasi, prestasi, atau variabel psikologis lainnya. Menurut Cohen (1988), analisis hubungan berperan penting dalam memahami struktur keterkaitan antarvariabel serta dalam menentukan besarnya efek (*effect size*) yang mendasari fenomena empiris. Oleh karena itu, analisis hubungan tidak hanya berfokus pada signifikansi statistik, tetapi juga pada makna substantif dari koefisien hubungan yang diperoleh.

Teknik analisis hubungan yang paling umum digunakan adalah korelasi Pearson Product–Moment. Korelasi Pearson digunakan untuk mengukur kekuatan dan arah hubungan linear antara dua variabel yang diukur pada skala interval atau rasio dan memenuhi asumsi normalitas serta linearitas. Menurut Howell (2012), koefisien korelasi Pearson (r) berkisar antara −1 hingga +1, di mana nilai positif menunjukkan hubungan searah dan nilai negatif menunjukkan hubungan berlawanan arah. Nilai r mendekati nol menunjukkan hubungan yang lemah atau tidak ada hubungan linear.

Interpretasi nilai korelasi harus mempertimbangkan konteks penelitian dan ukuran sampel. Cohen (1988)

mengemukakan pedoman umum untuk menginterpretasikan kekuatan korelasi, yaitu r ≈ 0,10 sebagai hubungan kecil, r ≈ 0,30 sebagai hubungan sedang, dan r ≥ 0,50 sebagai hubungan besar. Namun, Cohen juga menekankan bahwa pedoman ini bersifat konvensional dan harus disesuaikan dengan bidang kajian. Dalam penelitian pendidikan, misalnya, korelasi sebesar 0,30 dapat memiliki implikasi praktis yang signifikan, terutama jika variabel yang dikaji berkaitan dengan hasil belajar atau kesejahteraan peserta didik.

Apabila asumsi normalitas atau linearitas tidak terpenuhi, peneliti dapat menggunakan korelasi Spearman Rank-Order (ρ) sebagai alternatif nonparametris. Menurut Field (2018), korelasi Spearman digunakan ketika data berskala ordinal atau ketika distribusi data menyimpang secara signifikan dari normalitas. Korelasi Spearman mengukur kekuatan hubungan monotonic, bukan hubungan linear, sehingga lebih fleksibel dalam menghadapi data lapangan yang tidak ideal. Meskipun demikian, interpretasi nilai koefisien Spearman tetap mengikuti prinsip yang sama dengan korelasi Pearson dalam hal arah dan kekuatan hubungan.

Selain korelasi sederhana, penelitian kuantitatif sering memerlukan korelasi parsial untuk mengendalikan pengaruh variabel lain. Korelasi parsial mengukur hubungan antara dua variabel dengan mengontrol satu atau lebih variabel kovariat. Menurut Howell (2012), teknik ini sangat berguna dalam penelitian sosial dan pendidikan yang melibatkan banyak variabel saling terkait. Sebagai contoh, hubungan antara motivasi belajar dan prestasi akademik dapat dianalisis dengan mengontrol pengaruh kemampuan awal atau latar belakang sosial ekonomi. Dengan demikian, korelasi parsial memberikan gambaran hubungan yang lebih murni antarvariabel yang dikaji.

Analisis hubungan juga tidak dapat dilepaskan dari konsep effect size. Cohen (1988) menegaskan bahwa signifikansi statistik (*p-value*) hanya menunjukkan kemungkinan bahwa hubungan yang diamati tidak terjadi secara kebetulan, tetapi tidak memberikan informasi tentang besarnya efek hubungan tersebut. Koefisien korelasi itu sendiri merupakan ukuran effect size yang menunjukkan proporsi varians bersama antara dua variabel. Nilai r^2, yang merupakan kuadrat dari koefisien korelasi, menunjukkan persentase varians satu variabel yang dapat dijelaskan oleh variabel lain. Sebagai contoh, korelasi sebesar 0,40 menunjukkan bahwa sekitar 16% varians satu variabel dapat dijelaskan oleh variabel lainnya.

Dalam praktik analisis menggunakan perangkat lunak statistik seperti SPSS, Jamovi, JASP, dan R, analisis korelasi dapat dilakukan dengan relatif mudah. George dan Mallery (2020) menjelaskan bahwa SPSS menyediakan menu *Bivariate Correlations* yang memungkinkan peneliti menghitung korelasi Pearson, Spearman, dan Kendall secara simultan. Jamovi dan JASP menawarkan tampilan visual yang memudahkan interpretasi output, termasuk penyajian nilai r, *p-value*, dan interval kepercayaan. Namun, Field (2018) mengingatkan bahwa kemudahan teknis tidak boleh mengurangi ketelitian interpretasi, terutama terkait asumsi statistik dan konteks substantif.

Penting untuk ditekankan bahwa korelasi tidak menunjukkan hubungan sebab–akibat. Menurut Sheskin (2011), korelasi hanya menggambarkan keterkaitan statistik antara variabel, tanpa memberikan informasi tentang arah kausalitas. Hubungan yang tinggi antara dua variabel dapat disebabkan oleh variabel ketiga yang tidak teramati atau oleh hubungan yang bersifat timbal balik. Oleh karena itu, interpretasi hasil korelasi

harus dilakukan secara hati-hati dan tidak melampaui desain penelitian yang digunakan.

Analisis hubungan juga memiliki keterkaitan erat dengan reliabilitas instrumen. Hair et al. (2019) menjelaskan bahwa reliabilitas yang rendah akan melemahkan koefisien korelasi yang diperoleh, suatu fenomena yang dikenal sebagai *attenuation*. Oleh karena itu, interpretasi korelasi harus selalu mempertimbangkan kualitas pengukuran variabel yang terlibat. Korelasi yang rendah tidak selalu menunjukkan tidak adanya hubungan substantif, melainkan dapat mencerminkan keterbatasan reliabilitas instrumen.

Secara reflektif, analisis hubungan merupakan alat penting dalam penelitian kuantitatif untuk memahami keterkaitan antarvariabel dan mengestimasi besarnya efek empiris. Pemahaman yang tepat tentang jenis korelasi, asumsi statistik, dan interpretasi effect size akan membantu peneliti menghasilkan kesimpulan yang lebih bermakna dan bertanggung jawab. Analisis hubungan yang dilakukan secara cermat dan kontekstual dapat memberikan dasar yang kuat bagi penelitian lanjutan yang bersifat kausal maupun eksperimental.

Daftar Rujukan

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Dancey, C. P., & Reidy, J. (2017). *Statistics without maths for psychology* (7th ed.). Pearson Education.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (10th ed.). Cengage Learning.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Houghton Mifflin.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). McGraw-Hill.

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). CRC Press.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

BAB 7

ANALISIS STATISTIK INFERENSIAL

7.1 Uji t dan ANOVA

Uji *t* dan Analysis of Variance (ANOVA) merupakan teknik statistik inferensial yang paling mendasar dan luas digunakan dalam penelitian kuantitatif, khususnya pada bidang pendidikan, psikologi, manajemen, dan ilmu sosial. Kedua teknik ini dirancang untuk menguji perbedaan rata-rata antar kelompok dan menjadi fondasi bagi analisis inferensial yang lebih kompleks. Menurut Field (2018), uji *t* dan ANOVA merupakan pintu masuk utama bagi peneliti dalam memahami logika pengujian hipotesis berbasis perbedaan kelompok, sekaligus melatih ketelitian dalam memeriksa asumsi statistik yang menyertainya.

Uji *t* digunakan untuk membandingkan rata-rata dua kelompok. Dalam praktik penelitian, terdapat dua jenis uji *t* yang paling umum, yaitu uji *t* independen dan uji *t* berpasangan. Uji *t* independen digunakan ketika dua kelompok yang dibandingkan bersifat terpisah dan tidak saling bergantung, misalnya perbandingan hasil belajar antara siswa yang mengikuti metode pembelajaran daring dan luring. Menurut Howell (2012), uji *t* independen menguji hipotesis nol bahwa tidak terdapat perbedaan mean antara dua populasi yang diasumsikan memiliki distribusi normal dan varians yang homogen.

Sebaliknya, uji *t* berpasangan (paired-samples t-test) digunakan ketika dua pengukuran berasal dari subjek yang sama atau pasangan yang setara, misalnya skor pretest dan posttest siswa setelah diberikan perlakuan tertentu. Menurut Field (2018),

uji *t* berpasangan lebih sensitif dibandingkan uji *t* independen karena mengendalikan variasi individual yang tidak relevan dengan perlakuan. Oleh karena itu, dalam desain eksperimen pendidikan, uji *t* berpasangan sering menjadi pilihan utama untuk mengevaluasi efektivitas intervensi pembelajaran.

Baik uji *t* independen maupun berpasangan mensyaratkan beberapa asumsi statistik, antara lain normalitas distribusi skor dan, khusus untuk uji *t* independen, homogenitas varians antar kelompok. Stevens (2012) menegaskan bahwa pelanggaran asumsi ini dapat memengaruhi tingkat kesalahan tipe I dan tipe II. Namun demikian, uji *t* relatif robust terhadap pelanggaran normalitas ringan, terutama pada ukuran sampel yang seimbang dan cukup besar. Apabila asumsi homogenitas varians tidak terpenuhi, peneliti dapat menggunakan alternatif seperti uji *t* Welch yang lebih adaptif terhadap perbedaan varians.

Ketika jumlah kelompok yang dibandingkan lebih dari dua, penggunaan uji *t* menjadi tidak efisien dan meningkatkan risiko kesalahan tipe I akibat pengujian berulang. Dalam konteks inilah ANOVA digunakan sebagai solusi metodologis. Menurut Field (2018), ANOVA menguji apakah terdapat perbedaan rata-rata yang signifikan di antara tiga atau lebih kelompok dengan membandingkan varians antar kelompok (*between-group variance*) dan varians dalam kelompok (*within-group variance*). Rasio kedua varians ini dinyatakan dalam statistik *F*, yang menjadi dasar pengambilan keputusan inferensial.

ANOVA satu arah (one-way ANOVA) digunakan ketika peneliti menguji pengaruh satu variabel bebas kategorikal terhadap satu variabel terikat kontinu. Contoh klasik dalam penelitian pendidikan adalah perbandingan hasil belajar siswa berdasarkan tiga metode pembelajaran yang berbeda. Menurut Gravetter dan Wallnau (2016), ANOVA satu arah menguji

hipotesis nol bahwa seluruh kelompok memiliki mean yang sama. Jika hasil uji *F* signifikan, maka setidaknya terdapat satu pasang kelompok yang memiliki perbedaan mean yang signifikan.

Namun, ANOVA tidak secara langsung menunjukkan kelompok mana yang berbeda secara signifikan. Oleh karena itu, analisis dilanjutkan dengan uji lanjut (post hoc test) seperti Tukey, Bonferroni, atau Scheffé. Stevens (2012) menjelaskan bahwa pemilihan uji post hoc harus mempertimbangkan jumlah kelompok, ukuran sampel, dan asumsi homogenitas varians. Uji Tukey, misalnya, cocok digunakan ketika ukuran sampel relatif seimbang dan asumsi homogenitas terpenuhi, sedangkan Bonferroni lebih konservatif dan sering digunakan ketika jumlah perbandingan cukup banyak.

Selain ANOVA satu arah, terdapat pula ANOVA dua arah (two-way ANOVA) yang memungkinkan peneliti menguji pengaruh dua variabel bebas sekaligus serta interaksi di antara keduanya. Menurut Field (2018), keunggulan ANOVA dua arah terletak pada kemampuannya mengidentifikasi efek interaksi, yaitu kondisi ketika pengaruh satu variabel bebas bergantung pada tingkat variabel bebas lainnya. Dalam penelitian pendidikan, misalnya, efektivitas metode pembelajaran dapat berbeda tergantung pada tingkat motivasi belajar siswa. Analisis interaksi semacam ini memberikan wawasan yang lebih kaya dibandingkan analisis efek utama saja.

ANOVA juga memiliki asumsi statistik yang serupa dengan uji *t*, yaitu normalitas, homogenitas varians, dan independensi pengamatan. Tabachnick dan Fidell (2019) menegaskan bahwa pemeriksaan asumsi ini merupakan bagian integral dari analisis ANOVA. Apabila asumsi tidak terpenuhi, peneliti dapat mempertimbangkan transformasi data atau menggunakan alternatif nonparametris seperti uji Kruskal–Wallis. Namun,

keputusan ini harus didasarkan pada pertimbangan metodologis, bukan sekadar preferensi teknis.

Dalam praktik analisis menggunakan perangkat lunak statistik seperti SPSS, Jamovi, dan JASP, uji *t* dan ANOVA dapat dilakukan dengan relatif mudah melalui menu analisis yang tersedia. Field (2018) menekankan bahwa meskipun perangkat lunak menghasilkan output statistik secara otomatis, peneliti tetap harus mampu menafsirkan nilai *t*, *F*, derajat kebebasan, dan *p-value* secara kritis. Selain itu, pelaporan hasil analisis sebaiknya dilengkapi dengan ukuran efek (*effect size*) seperti Cohen's *d* untuk uji *t* dan eta squared (η^2) atau partial eta squared untuk ANOVA, guna memberikan informasi tentang signifikansi praktis temuan penelitian.

Secara reflektif, uji *t* dan ANOVA bukan sekadar alat statistik, melainkan kerangka logis untuk memahami perbedaan empiris antar kelompok dalam penelitian kuantitatif. Pemilihan yang tepat antara uji *t* dan ANOVA, pemeriksaan asumsi yang cermat, serta interpretasi hasil yang kontekstual akan meningkatkan validitas dan kekuatan temuan penelitian. Oleh karena itu, penguasaan uji *t* dan ANOVA menjadi kompetensi dasar yang harus dimiliki peneliti sebelum melangkah ke analisis inferensial multivariat yang lebih kompleks.

7.2 MANOVA dan MANCOVA

Multivariate Analysis of Variance (MANOVA) dan Multivariate Analysis of Covariance (MANCOVA) merupakan pengembangan dari ANOVA yang dirancang untuk menangani situasi penelitian dengan lebih dari satu variabel dependen yang saling berkorelasi. Dalam penelitian pendidikan, psikologi, manajemen, dan ilmu sosial, fenomena yang dikaji sering kali bersifat multidimensional sehingga tidak cukup

direpresentasikan oleh satu indikator tunggal. Menurut Hair et al. (2019), MANOVA memungkinkan peneliti menguji perbedaan kelompok secara simultan pada beberapa variabel dependen, sehingga memberikan gambaran yang lebih holistik dan mengurangi risiko kesalahan tipe I akibat pengujian berulang.

Secara konseptual, MANOVA menguji apakah vektor mean dari beberapa variabel dependen berbeda secara signifikan antar kelompok yang ditentukan oleh variabel bebas kategorikal. Berbeda dengan ANOVA yang membandingkan mean tunggal, MANOVA mempertimbangkan struktur kovarians antar variabel dependen. Stevens (2012) menegaskan bahwa MANOVA lebih tepat digunakan ketika variabel dependen memiliki korelasi moderat, karena korelasi tersebut justru menjadi kekuatan analisis multivariat. Apabila variabel dependen tidak berkorelasi sama sekali, MANOVA tidak memberikan keuntungan substantif dibandingkan ANOVA terpisah.

Asumsi statistik MANOVA mencakup beberapa persyaratan penting, antara lain normalitas multivariat, homogenitas matriks kovarians, independensi pengamatan, serta tidak adanya multikolinearitas ekstrem antar variabel dependen. Menurut Tabachnick dan Fidell (2019), normalitas multivariat jarang terpenuhi secara sempurna dalam data lapangan, namun MANOVA relatif robust terhadap pelanggaran ringan, terutama ketika ukuran sampel seimbang antar kelompok. Asumsi homogenitas matriks kovarians biasanya diuji menggunakan Box's M Test, di mana nilai signifikansi yang tidak signifikan menunjukkan bahwa asumsi terpenuhi.

Dalam output MANOVA, terdapat beberapa statistik uji multivariat yang dapat digunakan untuk pengambilan keputusan, seperti Wilks' Lambda, Pillai's Trace, Hotelling's Trace, dan Roy's Largest Root. Menurut Field (2018), Wilks' Lambda merupakan

statistik yang paling sering dilaporkan karena memiliki interpretasi yang relatif intuitif, yaitu proporsi varians total yang tidak dijelaskan oleh perbedaan kelompok. Nilai Wilks' Lambda yang kecil dan signifikan menunjukkan adanya perbedaan multivariat yang signifikan antar kelompok. Namun, Pillai's Trace sering direkomendasikan ketika asumsi homogenitas kovarians dilanggar karena lebih robust terhadap pelanggaran asumsi.

Setelah MANOVA menunjukkan hasil signifikan, analisis dilanjutkan dengan ANOVA univariat pada masing-masing variabel dependen serta uji lanjut (*post hoc*) jika diperlukan. Hair et al. (2019) menegaskan bahwa interpretasi hasil MANOVA harus dilakukan secara berjenjang, dimulai dari efek multivariat, diikuti oleh efek univariat, dan diakhiri dengan interpretasi substantif berdasarkan konteks penelitian. Pendekatan ini mencegah peneliti menarik kesimpulan parsial yang tidak didukung oleh hasil multivariat secara keseluruhan.

Sementara itu, MANCOVA merupakan pengembangan MANOVA yang memasukkan satu atau lebih variabel kovariat kontinu untuk mengendalikan pengaruh faktor luar yang tidak menjadi fokus utama penelitian. Menurut Stevens (2012), MANCOVA memungkinkan peneliti mengevaluasi perbedaan kelompok pada beberapa variabel dependen setelah mengontrol variabel pengganggu, sehingga estimasi efek perlakuan menjadi lebih akurat. Dalam penelitian pendidikan, misalnya, MANCOVA dapat digunakan untuk membandingkan hasil belajar dan motivasi siswa berdasarkan metode pembelajaran dengan mengontrol kemampuan awal atau latar belakang sosial ekonomi.

Asumsi MANCOVA mencakup seluruh asumsi MANOVA, ditambah dengan asumsi linearitas hubungan antara kovariat dan variabel dependen, serta homogenitas kemiringan regresi (homogeneity of regression slopes). Menurut Field (2018),

pelanggaran asumsi homogenitas kemiringan regresi dapat menyebabkan interpretasi hasil MANCOVA menjadi tidak valid karena hubungan antara kovariat dan variabel dependen berbeda antar kelompok. Oleh karena itu, pengujian asumsi ini merupakan langkah penting sebelum melanjutkan analisis MANCOVA.

Dalam praktik penggunaan perangkat lunak statistik seperti SPSS dan Jamovi, analisis MANOVA dan MANCOVA dapat dilakukan melalui menu *General Linear Model*. SPSS secara otomatis menyediakan statistik uji multivariat, uji asumsi, serta ANOVA lanjutan. Namun, Kline (2016) mengingatkan bahwa peneliti harus berhati-hati dalam menafsirkan output yang kompleks dan memastikan bahwa keputusan inferensial didasarkan pada pemahaman konseptual, bukan sekadar signifikansi statistik.

Dari perspektif metodologis, MANOVA dan MANCOVA memiliki keunggulan utama dalam meningkatkan efisiensi analisis dan mempertahankan tingkat kesalahan tipe I. Namun, analisis ini juga memiliki keterbatasan, terutama dalam hal interpretasi yang lebih kompleks dan tuntutan asumsi yang lebih ketat dibandingkan ANOVA. Oleh karena itu, pemilihan MANOVA atau MANCOVA harus didasarkan pada rasional teoretis yang kuat, bukan sekadar keinginan menggunakan teknik yang lebih canggih.

Secara reflektif, MANOVA dan MANCOVA merupakan alat inferensial yang sangat berguna untuk penelitian multivariat yang kompleks. Ketika digunakan secara tepat, teknik ini memungkinkan peneliti menangkap dinamika multidimensional fenomena sosial dan pendidikan secara lebih komprehensif. Pemahaman yang mendalam tentang asumsi, logika statistik, dan interpretasi hasil menjadi kunci untuk memastikan bahwa

temuan penelitian memiliki validitas dan relevansi substantif yang tinggi.

7.3 Regresi

Analisis regresi merupakan teknik statistik inferensial yang digunakan untuk memodelkan hubungan fungsional antara satu atau lebih variabel independen dengan satu variabel dependen. Dalam penelitian pendidikan, psikologi, manajemen, dan sosial, regresi digunakan untuk memprediksi nilai variabel terikat, menjelaskan kontribusi relatif prediktor, serta menguji hipotesis hubungan antarvariabel secara kuantitatif. Menurut Pedhazur (1997), regresi bukan sekadar alat prediksi, tetapi kerangka analitis yang memungkinkan peneliti menguji teori secara empiris melalui estimasi parameter yang terukur dan dapat diinterpretasikan.

Bentuk paling sederhana dari regresi adalah regresi linear sederhana, yang melibatkan satu variabel independen dan satu variabel dependen. Model ini dinyatakan dalam persamaan linear yang menggambarkan hubungan antara prediktor dan kriteria. Menurut Howell (2012), koefisien regresi dalam model ini menunjukkan perubahan rata-rata variabel dependen akibat satu satuan perubahan variabel independen. Dalam penelitian pendidikan, regresi linear sederhana sering digunakan untuk memprediksi prestasi belajar berdasarkan satu faktor, seperti motivasi atau waktu belajar.

Ketika penelitian melibatkan lebih dari satu variabel independen, digunakan regresi linear berganda. Regresi berganda memungkinkan peneliti mengevaluasi kontribusi unik setiap prediktor terhadap variabel dependen dengan mengendalikan pengaruh prediktor lain. Pedhazur (1997) menegaskan bahwa kekuatan utama regresi berganda terletak

pada kemampuannya memisahkan pengaruh variabel yang saling berkorelasi, sehingga interpretasi hasil menjadi lebih akurat. Dalam konteks penelitian sosial, regresi berganda sering digunakan untuk memodelkan fenomena kompleks yang dipengaruhi oleh berbagai faktor sekaligus.

Interpretasi hasil regresi berganda melibatkan beberapa parameter kunci, antara lain koefisien regresi tidak terstandar (B), koefisien terstandar (β), nilai R^2, dan uji signifikansi model. Menurut Field (2018), koefisien B menunjukkan perubahan absolut variabel dependen, sedangkan β memungkinkan perbandingan kontribusi relatif antar prediktor karena telah dinyatakan dalam satuan standar. Nilai R^2 menunjukkan proporsi varians variabel dependen yang dapat dijelaskan oleh seluruh prediktor dalam model. Sebagai contoh, nilai R^2 sebesar 0,40 menunjukkan bahwa 40% variasi skor dependen dapat dijelaskan oleh model regresi.

Analisis regresi memiliki sejumlah asumsi statistik yang harus dipenuhi agar hasil estimasi parameter valid dan dapat diinterpretasikan secara tepat. Asumsi utama regresi meliputi linearitas hubungan, normalitas residual, homoskedastisitas, independensi error, dan tidak adanya multikolinearitas antar prediktor. Menurut Tabachnick dan Fidell (2019), pelanggaran terhadap asumsi-asumsi ini dapat menyebabkan bias estimasi, kesalahan standar yang tidak akurat, serta kesimpulan inferensial yang menyesatkan. Oleh karena itu, pemeriksaan asumsi regresi merupakan bagian integral dari analisis.

Multikolinearitas menjadi perhatian khusus dalam regresi berganda karena berkaitan dengan korelasi tinggi antar variabel independen. Hair et al. (2019) menjelaskan bahwa multikolinearitas dapat meningkatkan varians koefisien regresi sehingga menurunkan stabilitas dan interpretabilitas model.

Indikator umum multikolinearitas meliputi nilai Variance Inflation Factor (VIF) dan Tolerance. Nilai VIF di atas 10 sering dianggap sebagai indikasi masalah serius, meskipun beberapa peneliti menggunakan batas yang lebih konservatif. Peneliti perlu mempertimbangkan penghapusan atau penggabungan prediktor yang sangat berkorelasi untuk meningkatkan kualitas model.

Selain regresi linear, dalam praktik penelitian juga dikenal berbagai bentuk regresi lanjutan seperti regresi hierarkis dan regresi stepwise. Regresi hierarkis memungkinkan peneliti memasukkan prediktor secara bertahap berdasarkan pertimbangan teoretis, sehingga dapat mengevaluasi peningkatan R^2 pada setiap tahap. Menurut Field (2018), pendekatan ini sangat berguna untuk menguji model teoretis yang bersifat bertingkat. Sebaliknya, regresi stepwise lebih bersifat eksploratif dan sering dikritik karena terlalu bergantung pada kriteria statistik semata tanpa dasar teoretis yang kuat.

Penggunaan perangkat lunak statistik seperti SPSS, Jamovi, JASP, dan R sangat memudahkan analisis regresi. SPSS menyediakan menu *Linear Regression* yang menampilkan output lengkap, termasuk koefisien, uji asumsi, dan diagnostik residual. Namun, Kline (2016) mengingatkan bahwa interpretasi hasil regresi harus selalu dikaitkan dengan teori dan desain penelitian. Regresi yang signifikan secara statistik tidak otomatis menunjukkan hubungan kausal, terutama dalam desain survei non-eksperimental.

Dalam konteks inferensi kausal, regresi memiliki keterbatasan yang perlu disadari. Stevens (2012) menegaskan bahwa regresi hanya mengestimasi hubungan prediktif, bukan sebab–akibat. Hubungan yang ditemukan dapat dipengaruhi oleh variabel perancu yang tidak dimasukkan ke dalam model. Oleh karena itu, hasil regresi harus ditafsirkan secara hati-hati dan,

apabila memungkinkan, dilengkapi dengan desain penelitian yang lebih kuat seperti eksperimen atau analisis longitudinal.

Secara reflektif, analisis regresi merupakan jembatan antara analisis hubungan sederhana dan pemodelan struktural yang lebih kompleks. Regresi memungkinkan peneliti menguji kontribusi relatif variabel, mengevaluasi model teoretis, dan menghasilkan prediksi berbasis data empiris. Penguasaan konsep, asumsi, dan interpretasi regresi menjadi kompetensi esensial bagi peneliti kuantitatif sebelum melangkah ke analisis lanjutan seperti *path analysis* dan Structural Equation Modeling.

7.4 Korelasi dan Path Analysis

Korelasi dan *path analysis* merupakan dua pendekatan statistik yang berperan penting dalam memahami struktur hubungan antarvariabel dalam penelitian kuantitatif. Korelasi berfungsi untuk mengidentifikasi derajat keterkaitan antarvariabel, sedangkan *path analysis* memperluas analisis tersebut dengan memodelkan hubungan kausal hipotetik berdasarkan teori. Dalam penelitian pendidikan, psikologi, manajemen, dan ilmu sosial, kedua teknik ini sering digunakan secara komplementer untuk mengevaluasi model konseptual yang melibatkan beberapa variabel saling terkait. Menurut Pedhazur (1997), korelasi dan *path analysis* merupakan fondasi penting bagi pengembangan pemodelan struktural yang lebih kompleks.

Analisis korelasi, sebagaimana telah dibahas pada bab sebelumnya, bertujuan mengukur kekuatan dan arah hubungan antarvariabel. Korelasi Pearson digunakan untuk data interval atau rasio yang memenuhi asumsi normalitas dan linearitas, sedangkan korelasi Spearman digunakan ketika data berskala ordinal atau tidak memenuhi asumsi parametris. Menurut Howell

(2012), korelasi memberikan informasi awal yang penting mengenai potensi hubungan antarvariabel, namun tidak cukup untuk menjelaskan mekanisme hubungan tersebut. Oleh karena itu, korelasi sering digunakan sebagai langkah awal sebelum membangun model analisis yang lebih kompleks.

Keterbatasan utama korelasi terletak pada ketidakmampuannya membedakan antara hubungan langsung dan hubungan yang dimediasi oleh variabel lain. Cohen (1988) menegaskan bahwa korelasi yang signifikan tidak memberikan informasi tentang arah pengaruh atau struktur hubungan kausal. Dua variabel dapat berkorelasi karena hubungan langsung, hubungan tidak langsung melalui variabel ketiga, atau karena keduanya dipengaruhi oleh faktor lain yang tidak terukur. Kesadaran akan keterbatasan ini mendorong penggunaan *path analysis* sebagai pendekatan analitis yang lebih kaya.

Path analysis merupakan pengembangan dari regresi berganda yang digunakan untuk menguji model hubungan kausal hipotetik antarvariabel yang telah ditentukan berdasarkan teori. Menurut Pedhazur (1997), *path analysis* memungkinkan peneliti memisahkan efek langsung, efek tidak langsung, dan efek total dari suatu variabel terhadap variabel lainnya. Efek langsung merepresentasikan pengaruh langsung satu variabel terhadap variabel lain, sedangkan efek tidak langsung mencerminkan pengaruh yang dimediasi oleh satu atau lebih variabel perantara (*mediator*).

Dalam *path analysis*, hubungan antarvariabel direpresentasikan melalui diagram jalur (*path diagram*) yang menunjukkan arah pengaruh sesuai dengan asumsi teoretis. Setiap jalur diestimasi menggunakan koefisien regresi terstandar (β) yang memungkinkan perbandingan kekuatan pengaruh antarjalur. Menurut Stevens (2012), interpretasi koefisien jalur

harus selalu dikaitkan dengan rasional teoretis dan desain penelitian, karena *path analysis* tidak secara otomatis membuktikan kausalitas, melainkan menguji konsistensi data dengan model kausal yang diajukan.

Asumsi statistik *path analysis* pada dasarnya sama dengan asumsi regresi linear berganda, yaitu linearitas hubungan, normalitas residual, homoskedastisitas, dan tidak adanya multikolinearitas yang ekstrem. Hair et al. (2019) menegaskan bahwa pelanggaran asumsi-asumsi ini dapat memengaruhi stabilitas dan interpretabilitas koefisien jalur. Oleh karena itu, sebelum melakukan *path analysis*, peneliti perlu melakukan analisis korelasi dan regresi pendahuluan untuk memastikan kelayakan data.

Salah satu keunggulan *path analysis* adalah kemampuannya menjelaskan mekanisme hubungan yang kompleks dalam penelitian sosial dan pendidikan. Sebagai contoh, dalam penelitian pendidikan, pengaruh status sosial ekonomi terhadap prestasi belajar dapat dimediasi oleh motivasi dan dukungan belajar di rumah. Melalui *path analysis*, peneliti dapat menguji apakah pengaruh status sosial ekonomi terhadap prestasi bersifat langsung atau sebagian besar bekerja melalui mediator tertentu. Menurut Hox dan Bechger (2007), pendekatan ini memberikan wawasan yang lebih mendalam dibandingkan analisis korelasi atau regresi sederhana.

Namun demikian, *path analysis* memiliki keterbatasan penting yang perlu diperhatikan. Kline (2016) menekankan bahwa *path analysis* hanya dapat diterapkan pada variabel yang terukur secara langsung (observed variables) dan tidak memperhitungkan kesalahan pengukuran secara eksplisit. Akibatnya, estimasi koefisien jalur dapat terdistorsi apabila reliabilitas instrumen rendah. Keterbatasan ini menjadi salah satu

alasan berkembangnya Structural Equation Modeling (SEM), yang memungkinkan penggabungan *path analysis* dengan analisis faktor untuk mengakomodasi variabel laten dan kesalahan pengukuran.

Dalam praktik analisis, *path* analysis dapat dilakukan menggunakan perangkat lunak statistik seperti SPSS (melalui regresi bertahap), Amos, LISREL, dan SmartPLS. Byrne (2016) menjelaskan bahwa Amos menyediakan antarmuka visual yang memudahkan peneliti menggambar diagram jalur dan mengestimasi koefisien secara langsung. Namun, terlepas dari perangkat lunak yang digunakan, peneliti harus memastikan bahwa model jalur yang diuji didasarkan pada teori yang kuat, bukan sekadar eksplorasi data.

Secara metodologis, perbedaan utama antara korelasi dan *path analysis* terletak pada tingkat kompleksitas dan kedalaman interpretasi. Korelasi bersifat deskriptif-inferensial dan terbatas pada hubungan dua variabel, sedangkan *path analysis* bersifat konfirmatori dan memungkinkan pengujian struktur hubungan yang lebih kompleks. Menurut Hair et al. (2019), pemilihan antara korelasi, regresi, dan *path analysis* harus disesuaikan dengan tujuan penelitian, kompleksitas model teoretis, dan kualitas data yang tersedia.

Secara reflektif, korelasi dan *path analysis* merupakan tahapan penting dalam perjalanan analisis inferensial menuju pemodelan struktural yang lebih komprehensif. Korelasi memberikan gambaran awal keterkaitan antarvariabel, sementara *path analysis* memungkinkan peneliti menguji mekanisme hubungan yang lebih mendalam. Pemahaman yang tepat terhadap kedua pendekatan ini akan membantu peneliti membangun model empiris yang lebih kuat, transparan, dan

bermakna sebelum melangkah ke analisis Structural Equation Modeling.

7.5 Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) merupakan pendekatan statistik multivariat lanjutan yang mengintegrasikan analisis faktor dan analisis jalur dalam satu kerangka pemodelan yang komprehensif. SEM memungkinkan peneliti menguji hubungan kausal hipotetik antarvariabel laten maupun variabel teramati secara simultan. Dalam penelitian pendidikan, psikologi, manajemen, dan ilmu sosial, SEM digunakan untuk menguji model teoretis yang kompleks, mengakomodasi kesalahan pengukuran, serta mengevaluasi kesesuaian model dengan data empiris. Menurut Kline (2016), SEM bukan sekadar teknik statistik, melainkan pendekatan konfirmatori untuk menguji teori secara kuantitatif.

Keunggulan utama SEM terletak pada kemampuannya memisahkan model pengukuran (measurement model) dan model struktural (structural model). Model pengukuran menjelaskan hubungan antara variabel laten dan indikator-indikatornya, sedangkan model struktural menjelaskan hubungan kausal antar variabel laten. Hair et al. (2019) menegaskan bahwa pemisahan ini memungkinkan peneliti mengevaluasi kualitas pengukuran dan hubungan antar konstruk secara terpisah, sehingga interpretasi hasil menjadi lebih akurat dan transparan.

Tahap awal dalam SEM umumnya dimulai dengan Confirmatory Factor Analysis (CFA). CFA digunakan untuk menguji apakah indikator-indikator yang dirancang benar-benar merefleksikan konstruk laten sesuai dengan teori. Berbeda dengan Exploratory Factor Analysis (EFA) yang bersifat eksploratif, CFA bersifat konfirmatori karena struktur faktor telah ditentukan

sebelumnya. Menurut Byrne (2016), CFA memberikan bukti validitas konstruk melalui pengujian *factor loading*, *construct reliability*, dan *average variance extracted*. Indikator dengan *loading* rendah menunjukkan kontribusi yang lemah terhadap konstruk laten dan perlu dipertimbangkan untuk direvisi atau dihapus.

Setelah model pengukuran menunjukkan hasil yang memadai, analisis dilanjutkan ke model struktural. Model struktural menguji hubungan kausal antar konstruk laten sebagaimana dirumuskan dalam hipotesis penelitian. Koefisien jalur dalam model struktural diinterpretasikan sebagai pengaruh langsung antar konstruk, serupa dengan koefisien regresi terstandar. Menurut Pedhazur (1997), keunggulan SEM dibandingkan regresi terpisah terletak pada kemampuannya mengestimasi seluruh hubungan secara simultan, sehingga mengurangi bias estimasi dan meningkatkan konsistensi model.

Evaluasi hasil SEM sangat bergantung pada goodness of fit indices, yaitu ukuran statistik yang menunjukkan sejauh mana model teoretis sesuai dengan data empiris. Indeks kesesuaian model yang umum dilaporkan meliputi Chi-square (χ^2), Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), dan Standardized Root Mean Square Residual (SRMR). Menurut Kline (2016), tidak ada satu indeks tunggal yang dapat dijadikan patokan absolut; evaluasi model harus mempertimbangkan beberapa indeks secara simultan.

Chi-square digunakan untuk menguji perbedaan antara matriks kovarians model dan data, namun indeks ini sangat sensitif terhadap ukuran sampel. Oleh karena itu, peneliti sering mengombinasikannya dengan indeks lain. Hair et al. (2019) menyarankan nilai CFI dan TLI ≥ 0,90 atau 0,95 sebagai indikator

kecocokan model yang baik, nilai RMSEA ≤ 0,08 atau 0,06, serta nilai SRMR ≤ 0,08. Interpretasi indeks-indeks ini harus dilakukan secara kontekstual dan tidak bersifat mekanistik.

Apabila model awal tidak menunjukkan kesesuaian yang memadai, peneliti dapat melakukan modifikasi model dengan mengacu pada *modification indices*. Namun, Byrne (2016) menegaskan bahwa modifikasi model harus didasarkan pada rasional teoretis yang kuat, bukan semata-mata untuk meningkatkan nilai *fit*. Modifikasi yang tidak didukung teori berisiko menghasilkan model yang overfitting dan sulit direplikasi pada sampel lain. Oleh karena itu, transparansi dan kehati-hatian sangat diperlukan dalam proses modifikasi model SEM.

SEM juga memiliki sejumlah asumsi statistik yang perlu diperhatikan, antara lain ukuran sampel yang memadai, normalitas multivariat, serta spesifikasi model yang benar. Stevens (2012) menyarankan ukuran sampel minimal 5–10 kali jumlah parameter yang diestimasi, meskipun kebutuhan sampel dapat bervariasi tergantung kompleksitas model dan metode estimasi. Pelanggaran asumsi normalitas dapat diatasi dengan metode estimasi alternatif seperti *robust maximum likelihood* atau *bootstrap*.

Dalam praktik penelitian, SEM dapat diimplementasikan menggunakan berbagai perangkat lunak statistik seperti SPSS Amos, LISREL, EQS, SmartPLS, dan paket *lavaan* dalam R. Byrne (2016) menjelaskan bahwa Amos populer karena antarmuka grafisnya yang intuitif, sedangkan LISREL dan EQS menawarkan fleksibilitas tinggi dalam spesifikasi model. Rosseel (2012) menunjukkan bahwa *lavaan* menjadi alternatif open-source yang kuat dan semakin banyak digunakan dalam penelitian akademik. Pemilihan perangkat lunak harus disesuaikan dengan tujuan penelitian, jenis data, dan kompetensi peneliti.

Dari perspektif metodologis, SEM menawarkan kerangka analisis yang sangat kuat, namun juga menuntut pemahaman teoretis dan statistik yang mendalam. Hox dan Bechger (2007) menekankan bahwa SEM bukan alat eksplorasi data, melainkan pendekatan konfirmatori yang bergantung pada teori yang jelas dan spesifikasi model yang tepat. Kesalahan dalam spesifikasi model dapat menghasilkan kesimpulan yang keliru meskipun indeks *fit* tampak memadai.

Secara reflektif, SEM merepresentasikan puncak analisis statistik inferensial dalam penelitian kuantitatif. Dengan kemampuannya mengintegrasikan pengukuran dan struktur kausal, SEM memungkinkan peneliti menguji teori secara lebih komprehensif dan akurat. Namun, kekuatan ini harus diimbangi dengan kehati-hatian metodologis, kejelasan teori, dan pelaporan yang transparan. Pemilihan SEM sebagai teknik analisis inferensial harus didasarkan pada kebutuhan penelitian dan kesiapan data, sehingga temuan yang dihasilkan benar-benar memiliki validitas dan kekuatan inferensial yang tinggi.

Daftar Rujukan

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (10th ed.). Cengage Learning.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hox, J. J., & Bechger, T. M. (2007). An introduction to structural equation modeling. *Family Science Review, 11*(4), 354–373.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Cengage Learning.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th ed.). Routledge.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. In J. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (pp. 607–634). Wiley.

BAB 8

ANALISIS LANJUTAN (MODERN KUANTITATIF)

8.1 Partial Least Squares Structural Equation Modeling (PLS-SEM)

Partial Least Squares Structural Equation Modeling (PLS-SEM) merupakan pendekatan pemodelan struktural berbasis varians yang semakin populer dalam penelitian kuantitatif modern, khususnya pada bidang pendidikan, psikologi, manajemen, bisnis, dan ilmu sosial. Berbeda dengan covariance-based SEM (CB-SEM) yang berorientasi pada pengujian kesesuaian model teoretis, PLS-SEM lebih menekankan pada kemampuan prediktif dan eksplorasi hubungan antarvariabel. Menurut Hair, Hult, Ringle, dan Sarstedt (2022), PLS-SEM sangat sesuai digunakan ketika tujuan penelitian adalah prediksi, pengembangan teori, atau eksplorasi model struktural yang kompleks dengan keterbatasan ukuran sampel dan asumsi distribusi data.

Perbedaan mendasar antara PLS-SEM dan CB-SEM terletak pada filosofi statistik yang mendasarinya. CB-SEM bertujuan meminimalkan perbedaan antara matriks kovarians empiris dan kovarians model, sehingga menuntut asumsi normalitas multivariat dan ukuran sampel besar. Sebaliknya, PLS-SEM memaksimalkan varians terjelaskan (*explained variance*) dari variabel endogen melalui pendekatan iteratif berbasis regresi. Hair et al. (2022) menegaskan bahwa PLS-SEM lebih fleksibel

terhadap pelanggaran asumsi statistik dan cocok untuk penelitian lapangan yang bersifat dinamis dan kompleks.

PLS-SEM terdiri atas dua komponen utama, yaitu model pengukuran (outer model) dan model struktural (inner model). Model pengukuran menjelaskan hubungan antara konstruk laten dan indikator-indikatornya, sedangkan model struktural menjelaskan hubungan kausal antar konstruk laten. Dalam PLS-SEM, konstruk laten dapat dimodelkan sebagai reflektif atau formatif, suatu fleksibilitas yang jarang tersedia dalam CB-SEM. Sarstedt, Ringle, dan Hair (2017) menekankan bahwa pemilihan tipe konstruk harus didasarkan pada rasional teoretis, bukan pertimbangan statistik semata.

Evaluasi model pengukuran reflektif dalam PLS-SEM mencakup pengujian reliabilitas indikator, reliabilitas konstruk, validitas konvergen, dan validitas diskriminan. Reliabilitas indikator dievaluasi melalui nilai *outer loading*, sedangkan reliabilitas konstruk dinilai menggunakan *Composite Reliability* dan Cronbach's Alpha. Menurut Hair et al. (2022), nilai *Composite Reliability* antara 0,70 hingga 0,95 dianggap memadai. Validitas konvergen dievaluasi melalui *Average Variance Extracted (AVE)*, dengan nilai minimal 0,50 sebagai indikator bahwa konstruk menjelaskan lebih dari setengah varians indikatornya.

Validitas diskriminan dalam PLS-SEM dievaluasi menggunakan beberapa pendekatan, termasuk kriteria Fornell–Larcker dan Heterotrait–Monotrait Ratio (HTMT). Sarstedt et al. (2017) menunjukkan bahwa HTMT merupakan pendekatan yang lebih sensitif dan direkomendasikan dalam praktik penelitian modern. Nilai HTMT di bawah 0,85 atau 0,90 menunjukkan bahwa konstruk laten dapat dibedakan secara empiris. Evaluasi ini sangat penting dalam penelitian sosial dan pendidikan yang melibatkan konstruk abstrak dan saling berkaitan.

Setelah model pengukuran memenuhi kriteria, analisis dilanjutkan ke model struktural. Evaluasi model struktural dalam PLS-SEM mencakup pengujian multikolinearitas antar konstruk, signifikansi dan kekuatan koefisien jalur, nilai R^2, *effect size* (f^2), serta *predictive relevance* (Q^2). Menurut Hair et al. (2022), nilai R^2 digunakan untuk menilai kemampuan prediktif model, dengan nilai 0,25, 0,50, dan 0,75 masing-masing menunjukkan kekuatan prediksi lemah, sedang, dan kuat. Koefisien jalur diinterpretasikan serupa dengan koefisien regresi terstandar.

Pengujian signifikansi dalam PLS-SEM dilakukan menggunakan bootstrapping, suatu prosedur resampling nonparametris yang tidak bergantung pada asumsi normalitas. Hair et al. (2022) menjelaskan bahwa bootstrapping memungkinkan estimasi *standard error*, nilai *t*, dan interval kepercayaan untuk setiap koefisien jalur. Pendekatan ini sangat relevan dalam penelitian sosial dan pendidikan yang sering menghadapi distribusi data tidak normal dan ukuran sampel terbatas.

PLS-SEM juga sangat cocok digunakan dalam konteks predictive analytics, yang menjadi ciri utama riset kuantitatif modern. Shmueli et al. (2016) menegaskan bahwa pergeseran dari statistik konfirmatori menuju pemodelan prediktif menuntut pendekatan analisis yang berorientasi pada akurasi prediksi, bukan sekadar kesesuaian model. Dalam konteks ini, PLS-SEM memberikan keunggulan karena fokus pada varians terjelaskan dan kemampuan memprediksi nilai variabel endogen pada data baru.

Dalam praktik penelitian, PLS-SEM banyak diimplementasikan menggunakan perangkat lunak seperti SmartPLS, WarpPLS, serta paket *plspm* dan *seminr* dalam R. SmartPLS menjadi pilihan populer karena antarmuka grafisnya

yang intuitif dan dukungan analisis lanjutan seperti *importance–performance map analysis (IPMA)*. Namun, Hair et al. (2022) mengingatkan bahwa kemudahan perangkat lunak tidak boleh menggantikan pemahaman teoretis dan metodologis peneliti dalam menyusun dan mengevaluasi model.

Meskipun memiliki banyak keunggulan, PLS-SEM juga memiliki keterbatasan. Sarstedt et al. (2017) menegaskan bahwa PLS-SEM kurang tepat digunakan untuk pengujian teori yang mapan dan pengujian *model fit* global secara ketat, yang lebih sesuai ditangani oleh CB-SEM. Oleh karena itu, pemilihan PLS-SEM harus didasarkan pada tujuan penelitian, tahap pengembangan teori, dan karakteristik data, bukan sekadar preferensi metode.

Secara reflektif, PLS-SEM merepresentasikan pergeseran paradigma dalam analisis kuantitatif modern menuju pendekatan yang lebih prediktif, fleksibel, dan kontekstual. Dalam penelitian pendidikan, psikologi, dan ilmu sosial, PLS-SEM membuka peluang untuk menguji model kompleks yang sulit ditangani oleh pendekatan tradisional. Dengan pemahaman teoretis yang kuat dan penerapan metodologis yang tepat, PLS-SEM dapat menjadi alat analisis yang sangat efektif dalam menjawab tantangan penelitian kontemporer yang semakin multidimensional dan berbasis data.

8.2 Mediasi dan Moderasi

Analisis mediasi dan moderasi merupakan pendekatan kuantitatif lanjutan yang digunakan untuk memahami mekanisme dan kondisi hubungan antarvariabel dalam penelitian sosial, pendidikan, psikologi, dan manajemen. Berbeda dengan analisis hubungan sederhana yang hanya menguji apakah dua variabel berkaitan, mediasi dan moderasi memungkinkan peneliti

menjawab pertanyaan *bagaimana* dan *kapan* suatu hubungan terjadi. Menurut Hayes (2018), kedua pendekatan ini menjadi semakin penting dalam riset kontemporer karena fenomena sosial jarang bersifat linier dan tunggal, melainkan dipengaruhi oleh proses perantara dan kondisi kontekstual tertentu.

Mediasi terjadi ketika pengaruh variabel independen terhadap variabel dependen disalurkan melalui satu atau lebih variabel perantara (*mediator*). Model mediasi klasik diperkenalkan oleh Baron dan Kenny (1986), yang mengemukakan bahwa suatu variabel dapat dianggap sebagai mediator apabila memenuhi empat kondisi, yaitu: (1) variabel independen berpengaruh signifikan terhadap variabel dependen, (2) variabel independen berpengaruh signifikan terhadap mediator, (3) mediator berpengaruh signifikan terhadap variabel dependen ketika variabel independen dikendalikan, dan (4) pengaruh variabel independen terhadap variabel dependen berkurang setelah mediator dimasukkan ke dalam model. Meskipun berpengaruh besar dalam literatur awal, pendekatan ini kemudian dikritik karena terlalu bergantung pada signifikansi statistik dan kurang sensitif terhadap efek tidak langsung.

Perkembangan metodologis selanjutnya menekankan pentingnya pengujian efek tidak langsung (indirect effect) secara langsung, terutama melalui teknik bootstrapping. Hayes (2018) menegaskan bahwa signifikansi mediasi seharusnya ditentukan berdasarkan interval kepercayaan dari efek tidak langsung, bukan sekadar pengujian jalur individual. Bootstrapping merupakan prosedur resampling nonparametris yang memungkinkan estimasi distribusi efek tidak langsung tanpa mengasumsikan normalitas. Apabila interval kepercayaan bootstrap tidak mencakup nilai nol, maka efek mediasi dianggap signifikan.

Pendekatan ini kini menjadi standar emas dalam analisis mediasi modern.

Dalam penelitian pendidikan, mediasi sering digunakan untuk menjelaskan mekanisme pengaruh variabel kontekstual terhadap hasil belajar. Sebagai contoh, pengaruh status sosial ekonomi terhadap prestasi akademik dapat dimediasi oleh motivasi belajar atau dukungan orang tua. Dengan analisis mediasi, peneliti dapat mengidentifikasi titik intervensi yang lebih spesifik dan relevan secara praktis. Menurut Hayes (2018), kekuatan analisis mediasi terletak pada kemampuannya menghubungkan teori dengan implikasi kebijakan dan praktik.

Berbeda dengan mediasi, moderasi terjadi ketika kekuatan atau arah hubungan antara variabel independen dan dependen bergantung pada tingkat variabel lain yang disebut moderator. Baron dan Kenny (1986) mendefinisikan moderator sebagai variabel yang memengaruhi hubungan antarvariabel utama, bukan menyalurkan pengaruh tersebut. Dalam penelitian sosial dan pendidikan, moderator sering berupa karakteristik individu atau konteks, seperti jenis kelamin, tingkat motivasi, atau lingkungan belajar. Analisis moderasi membantu peneliti memahami kondisi di mana suatu hubungan menjadi lebih kuat, lebih lemah, atau bahkan berubah arah.

Secara statistik, analisis moderasi biasanya dilakukan melalui interaksi antara variabel independen dan moderator dalam model regresi. Menurut Cohen (1988), koefisien interaksi yang signifikan menunjukkan adanya efek moderasi. Interpretasi hasil moderasi tidak cukup hanya berdasarkan signifikansi koefisien, tetapi juga memerlukan visualisasi seperti *simple slopes analysis* untuk memahami pola hubungan pada tingkat moderator yang berbeda. Hayes (2018) menekankan bahwa

interpretasi visual sangat penting untuk menghindari kesalahan pemahaman terhadap efek interaksi yang kompleks.

Perkembangan terbaru dalam analisis mediasi dan moderasi mengarah pada conditional process modeling, yang mengintegrasikan mediasi dan moderasi dalam satu kerangka analisis. Pendekatan ini memungkinkan peneliti menguji apakah efek mediasi berbeda pada tingkat moderator tertentu. Hayes (2018) mengembangkan berbagai model *PROCESS* yang memfasilitasi analisis ini secara sistematis dan mudah diimplementasikan menggunakan perangkat lunak statistik seperti SPSS, Jamovi, dan R. Model-model ini sangat berguna dalam penelitian kontemporer yang menuntut pemahaman hubungan bersyarat yang kompleks.

Dalam praktik analisis, penggunaan perangkat lunak modern sangat memudahkan penerapan mediasi dan moderasi. *PROCESS Macro* karya Hayes memungkinkan peneliti melakukan analisis mediasi, moderasi, dan *conditional process* tanpa harus membangun model secara manual. Namun, Hayes (2018) mengingatkan bahwa kemudahan teknis ini harus diimbangi dengan pemahaman konseptual yang kuat. Peneliti harus mampu menjelaskan makna efek langsung, tidak langsung, dan total, serta keterbatasan desain penelitian yang digunakan.

Analisis mediasi dan moderasi juga memiliki implikasi metodologis yang penting terkait inferensi kausal. Meskipun teknik ini sering digunakan untuk menguji model kausal hipotetik, Stevens (2012) menegaskan bahwa kesimpulan kausal tetap bergantung pada desain penelitian dan asumsi yang mendasarinya. Dalam desain survei potong lintang (*cross-sectional*), misalnya, interpretasi kausal harus dilakukan secara hati-hati dan didukung oleh teori yang kuat.

Secara reflektif, analisis mediasi dan moderasi merepresentasikan pergeseran penting dalam riset kuantitatif modern dari sekadar pengujian hubungan langsung menuju pemahaman mekanisme dan kondisi hubungan tersebut. Dengan menggabungkan pendekatan klasik dan teknik modern seperti bootstrapping dan *conditional process modeling*, peneliti dapat menghasilkan temuan yang lebih kaya, kontekstual, dan relevan. Analisis ini tidak hanya meningkatkan ketepatan inferensi statistik, tetapi juga memperkuat kontribusi teoretis dan praktis penelitian sosial, pendidikan, dan perilaku.

8.3 Analisis Cluster dan Diskriminan

Analisis cluster dan analisis diskriminan merupakan teknik multivariat yang digunakan untuk memahami struktur kelompok dalam data dan membedakan karakteristik antar kelompok tersebut. Kedua teknik ini memiliki orientasi yang berbeda namun saling melengkapi. Analisis cluster bersifat eksploratif dan tidak memerlukan informasi awal mengenai keanggotaan kelompok, sedangkan analisis diskriminan bersifat konfirmatori karena digunakan untuk menguji atau memprediksi keanggotaan kelompok yang telah ditentukan. Menurut Everitt (2011), kombinasi kedua teknik ini sangat relevan dalam riset sosial modern yang bertujuan memetakan heterogenitas populasi dan mengembangkan segmentasi berbasis data.

Analisis cluster bertujuan mengelompokkan objek atau individu ke dalam kelompok-kelompok yang relatif homogen di dalam kelompok dan heterogen antar kelompok. Dalam penelitian pendidikan dan sosial, analisis cluster sering digunakan untuk mengidentifikasi profil siswa, pola perilaku belajar, segmentasi konsumen pendidikan, atau tipologi organisasi. Tabachnick dan Fidell (2019) menegaskan bahwa analisis cluster

membantu peneliti memahami struktur laten dalam data tanpa harus menetapkan asumsi awal tentang jumlah atau karakteristik kelompok.

Terdapat dua pendekatan utama dalam analisis cluster, yaitu hierarchical clustering dan non-hierarchical clustering, khususnya K-means clustering. Hierarchical clustering membangun struktur kelompok secara bertahap, baik dengan pendekatan aglomeratif (menggabungkan objek dari yang paling mirip) maupun divisif (memisahkan objek dari kelompok besar). Hasil analisis ini sering divisualisasikan dalam bentuk *dendrogram*, yang membantu peneliti menentukan jumlah cluster yang paling masuk akal secara substantif. Menurut Everitt (2011), hierarchical clustering sangat berguna pada tahap eksplorasi awal, terutama ketika jumlah cluster yang optimal belum diketahui.

Sebaliknya, K-means clustering merupakan pendekatan non-hierarkis yang mengelompokkan data ke dalam jumlah cluster yang telah ditentukan sebelumnya. Metode ini bekerja dengan meminimalkan variasi dalam cluster dan memaksimalkan variasi antar cluster. Menurut Hair et al. (2019), K-means sangat efisien untuk dataset berukuran besar dan sering digunakan dalam riset manajemen dan bisnis untuk segmentasi pasar. Namun, kelemahan utama K-means terletak pada ketergantungannya terhadap penentuan jumlah cluster awal, sehingga peneliti harus menggunakan pertimbangan teoretis dan empiris secara hati-hati.

Interpretasi hasil analisis cluster memerlukan evaluasi terhadap profil cluster, yaitu karakteristik variabel yang membedakan satu cluster dari cluster lainnya. Dalam penelitian pendidikan, misalnya, cluster siswa dapat dibedakan berdasarkan tingkat motivasi, prestasi, dan strategi belajar. Shmueli et al.

(2016) menegaskan bahwa analisis cluster berkontribusi penting dalam pendekatan *predictive analytics* karena memungkinkan peneliti dan praktisi mengembangkan intervensi yang lebih terarah berdasarkan profil kelompok yang teridentifikasi.

Berbeda dengan analisis cluster, analisis diskriminan digunakan ketika kelompok telah diketahui sebelumnya dan peneliti ingin mengidentifikasi variabel yang paling efektif dalam membedakan kelompok tersebut. Analisis diskriminan linear (Linear Discriminant Analysis/LDA) membangun satu atau lebih fungsi diskriminan yang merupakan kombinasi linear dari variabel prediktor. Menurut Tabachnick dan Fidell (2019), tujuan utama analisis diskriminan adalah memaksimalkan perbedaan antar kelompok relatif terhadap variasi dalam kelompok.

Asumsi statistik analisis diskriminan meliputi normalitas multivariat, homogenitas matriks kovarians, dan tidak adanya multikolinearitas ekstrem antar prediktor. Hair et al. (2019) menjelaskan bahwa meskipun pelanggaran asumsi dapat memengaruhi akurasi klasifikasi, analisis diskriminan relatif robust terhadap pelanggaran moderat, terutama pada ukuran sampel yang seimbang. Evaluasi hasil analisis diskriminan biasanya dilakukan melalui tingkat akurasi klasifikasi dan *cross-validation* untuk menilai kemampuan prediksi model.

Dalam praktik penelitian, analisis cluster dan diskriminan sering digunakan secara berurutan. Pertama, analisis cluster digunakan untuk mengidentifikasi kelompok-kelompok laten dalam data. Selanjutnya, analisis diskriminan digunakan untuk memvalidasi dan menjelaskan perbedaan antar kelompok tersebut. Pendekatan ini memberikan kombinasi eksplorasi dan konfirmasi yang kuat, terutama dalam penelitian sosial dan pendidikan yang melibatkan populasi heterogen.

Penggunaan perangkat lunak statistik seperti SPSS, Jamovi, JASP, dan R sangat memudahkan penerapan kedua teknik ini. SPSS menyediakan menu *Hierarchical Cluster*, *K-means Cluster*, dan Discriminant *Analysis*, sementara R menawarkan fleksibilitas tinggi melalui berbagai paket statistik. Namun, Tabachnick dan Fidell (2019) mengingatkan bahwa peneliti harus lebih menekankan interpretasi substantif daripada sekadar output numerik, karena nilai statistik tanpa konteks teoretis memiliki makna yang terbatas.

Secara reflektif, analisis cluster dan diskriminan memainkan peran strategis dalam riset kuantitatif modern yang berorientasi pada segmentasi, klasifikasi, dan pengambilan keputusan berbasis data. Analisis cluster membuka wawasan tentang struktur laten populasi, sedangkan analisis diskriminan memperkuat pemahaman tentang faktor pembeda antar kelompok. Dengan pemahaman metodologis yang kuat dan interpretasi yang kontekstual, kedua teknik ini dapat memberikan kontribusi signifikan terhadap pengembangan teori dan praktik dalam ilmu sosial, pendidikan, dan manajemen.

8.4 Machine Learning untuk Riset Sosial

Machine learning (ML) merupakan pendekatan analisis data modern yang berfokus pada pengembangan model prediktif berbasis algoritma yang mampu belajar dari data tanpa harus diprogram secara eksplisit untuk setiap aturan keputusan. Dalam konteks riset sosial, pendidikan, psikologi, dan manajemen, machine learning menjadi semakin relevan seiring meningkatnya ketersediaan data besar (*big data*), data digital, dan kebutuhan akan prediksi yang akurat. Shmueli et al. (2016) menegaskan bahwa machine learning menandai pergeseran paradigma dari statistik inferensial tradisional yang berorientasi pada pengujian

hipotesis menuju predictive analytics yang menekankan akurasi prediksi dan generalisasi model.

Perbedaan mendasar antara pendekatan statistik klasik dan machine learning terletak pada tujuan analisis. Statistik klasik berfokus pada inferensi, estimasi parameter, dan pengujian signifikansi untuk mendukung atau menolak hipotesis teoretis. Sebaliknya, machine learning berfokus pada kinerja prediktif model terhadap data baru yang belum pernah diamati. Menurut Kuhn dan Johnson (2013), dalam machine learning, akurasi prediksi dan kemampuan generalisasi lebih diprioritaskan daripada interpretasi koefisien model. Meskipun demikian, kedua pendekatan ini tidak bersifat saling meniadakan, melainkan dapat saling melengkapi dalam riset sosial modern.

Machine learning secara umum dibagi menjadi dua kategori utama, yaitu supervised learning dan unsupervised learning. Supervised learning digunakan ketika data memiliki variabel target yang jelas, seperti kategori atau nilai numerik yang ingin diprediksi. Contoh penerapan supervised learning dalam riset pendidikan adalah memprediksi kelulusan mahasiswa berdasarkan data akademik dan demografis. Algoritma yang umum digunakan dalam supervised learning meliputi regresi logistik, *decision tree*, *random forest*, *support vector machine* (SVM), dan *gradient boosting*. Breiman (2001) menunjukkan bahwa *random forest* memiliki keunggulan dalam menangani data berdimensi tinggi dan interaksi nonlinier tanpa memerlukan asumsi distribusi yang ketat.

Sebaliknya, unsupervised learning digunakan ketika data tidak memiliki label atau variabel target yang jelas. Tujuan utamanya adalah menemukan pola, struktur, atau pengelompokan laten dalam data. Analisis cluster yang telah dibahas pada subbab sebelumnya dapat dipandang sebagai

bentuk awal dari unsupervised learning. Menurut James et al. (2021), teknik unsupervised learning seperti *K-means*, *hierarchical clustering*, dan *principal component analysis* (PCA) sangat berguna untuk eksplorasi data sosial yang kompleks dan heterogen. Dalam riset sosial digital, unsupervised learning sering digunakan untuk mengidentifikasi pola perilaku pengguna, segmentasi populasi, atau topik laten dalam teks.

Salah satu kekuatan utama machine learning adalah kemampuannya menangani hubungan nonlinier dan interaksi kompleks antarvariabel yang sulit ditangkap oleh model statistik linear. Bishop (2006) menjelaskan bahwa algoritma seperti SVM dan *neural networks* mampu memodelkan fungsi nonlinier melalui transformasi ruang fitur. Dalam konteks riset sosial, kemampuan ini sangat penting karena fenomena sosial jarang mengikuti pola linier sederhana. Namun demikian, kompleksitas model machine learning juga membawa tantangan dalam hal interpretabilitas dan transparansi.

Isu interpretabilitas model menjadi perhatian utama dalam penerapan machine learning untuk riset sosial. Shmueli et al. (2016) menegaskan bahwa model dengan akurasi tinggi namun sulit dijelaskan dapat menimbulkan dilema etis dan metodologis, terutama ketika hasil analisis digunakan untuk pengambilan keputusan kebijakan atau evaluasi pendidikan. Oleh karena itu, pendekatan *explainable AI* (XAI) mulai banyak dikembangkan untuk menjembatani kebutuhan prediksi dan interpretasi. Teknik seperti *feature importance*, *partial dependence plots*, dan *SHAP values* membantu peneliti memahami kontribusi relatif variabel dalam model prediktif.

Evaluasi kinerja model machine learning berbeda dari evaluasi model statistik tradisional. Dalam machine learning, data biasanya dibagi menjadi data latih (training set) dan data uji

(testing set) untuk menilai kemampuan generalisasi model. Menurut Kuhn dan Johnson (2013), metrik evaluasi yang umum digunakan meliputi akurasi, *precision*, *recall*, *F1-score*, dan *area under the curve (AUC)* untuk klasifikasi, serta *mean squared error (MSE)* atau *root mean squared error (RMSE)* untuk regresi. Pendekatan ini menekankan validasi eksternal model, bukan sekadar signifikansi statistik internal.

Dalam praktik riset sosial, machine learning sering diimplementasikan menggunakan perangkat lunak dan bahasa pemrograman seperti R dan Python. Paket seperti *caret*, *tidymodels*, dan *scikit-learn* menyediakan kerangka kerja yang komprehensif untuk pengembangan, evaluasi, dan validasi model machine learning. Meskipun demikian, Kuhn dan Johnson (2013) mengingatkan bahwa penggunaan machine learning memerlukan kehati-hatian dalam pemilihan fitur, pengendalian *overfitting*, dan validasi silang (*cross-validation*). Tanpa prosedur ini, model dengan kinerja tinggi pada data latih dapat gagal ketika diterapkan pada data baru.

Dalam konteks metodologis, machine learning tidak dimaksudkan untuk menggantikan statistik inferensial, melainkan memperluas cakupan analisis kuantitatif. Penggabungan machine learning dan pendekatan statistik klasik memungkinkan peneliti menguji teori sekaligus membangun model prediktif yang kuat. Pendekatan hibrida ini semakin relevan dalam riset sosial kontemporer yang menuntut akurasi, relevansi, dan skalabilitas analisis.

Secara reflektif, machine learning membuka peluang besar bagi riset sosial untuk menjawab pertanyaan penelitian yang sebelumnya sulit dijangkau dengan metode konvensional. Dengan kemampuannya menangani data besar, kompleks, dan nonlinier, machine learning memungkinkan analisis yang lebih

presisi dan adaptif terhadap dinamika sosial modern. Namun, pemanfaatannya harus disertai dengan kesadaran metodologis, etika penelitian, dan integrasi teori yang kuat agar temuan yang dihasilkan tidak hanya akurat secara prediktif, tetapi juga bermakna secara ilmiah dan sosial.

8.5 Meta-Analisis

Meta-analisis merupakan pendekatan kuantitatif lanjutan yang digunakan untuk mensintesis hasil dari sejumlah penelitian empiris yang mengkaji topik atau hubungan yang sama. Berbeda dengan tinjauan pustaka naratif yang bersifat deskriptif, meta-analisis menggunakan teknik statistik untuk menggabungkan ukuran efek (*effect size*) dari berbagai studi sehingga menghasilkan estimasi yang lebih presisi dan generalizable. Menurut Borenstein, Hedges, Higgins, dan Rothstein (2009), meta-analisis memainkan peran penting dalam pengembangan ilmu pengetahuan karena mampu mereduksi inkonsistensi temuan antar penelitian dan meningkatkan kekuatan inferensial melalui akumulasi bukti empiris.

Konsep sentral dalam meta-analisis adalah effect size, yaitu ukuran kuantitatif yang merepresentasikan besarnya pengaruh atau hubungan yang diteliti secara independen dari ukuran sampel. Hedges dan Olkin (1985) menjelaskan bahwa penggunaan effect size memungkinkan perbandingan hasil antar studi yang menggunakan desain, instrumen, dan ukuran sampel yang berbeda. Effect size yang umum digunakan dalam penelitian sosial dan pendidikan meliputi *Cohen's d* untuk perbedaan rata-rata, *Hedges' g* sebagai koreksi bias sampel kecil, koefisien korelasi (*r*), serta *odds ratio* dalam penelitian kategorikal. Standarisasi ini menjadi kunci utama dalam penggabungan hasil penelitian yang heterogen.

Proses meta-analisis umumnya dimulai dengan penentuan kriteria inklusi dan eksklusi studi. Kriteria ini harus ditetapkan secara transparan dan sistematis untuk meminimalkan bias seleksi. Menurut Borenstein et al. (2009), kejelasan kriteria inklusi menentukan validitas eksternal hasil meta-analisis karena memengaruhi ruang lingkup inferensi yang dapat ditarik. Dalam penelitian pendidikan, misalnya, peneliti perlu menentukan apakah meta-analisis mencakup studi eksperimental saja atau juga studi kuasi-eksperimental dan korelasional.

Setelah studi yang relevan dikumpulkan, langkah berikutnya adalah ekstraksi dan pengkodean data. Data yang diekstraksi mencakup informasi tentang desain penelitian, karakteristik sampel, variabel yang diukur, serta statistik yang diperlukan untuk menghitung effect size. Borenstein et al. (2009) menekankan pentingnya reliabilitas pengkodean, yang sering dilakukan melalui pengkodean ganda (*double coding*) untuk mengurangi kesalahan dan subjektivitas. Tahap ini sangat krusial karena kesalahan pengkodean dapat berdampak langsung pada estimasi effect size gabungan.

Salah satu isu metodologis utama dalam meta-analisis adalah heterogenitas, yaitu variasi effect size antar studi yang melebihi variasi yang diharapkan secara kebetulan. Heterogenitas mencerminkan perbedaan substantif dalam desain, konteks, atau karakteristik sampel penelitian. Menurut Higgins dan Thompson, heterogenitas biasanya diukur menggunakan statistik Q dan I^2. Nilai I^2 menunjukkan persentase variasi total yang disebabkan oleh heterogenitas nyata, bukan oleh kesalahan sampling. Nilai I^2 yang tinggi mengindikasikan bahwa hasil studi sangat bervariasi dan memerlukan analisis lanjutan.

Pemilihan model meta-analisis sangat bergantung pada tingkat heterogenitas yang terdeteksi. Dua model utama yang

digunakan adalah fixed-effect model dan random-effects model. Fixed-effect model mengasumsikan bahwa seluruh studi mengestimasi satu effect size populasi yang sama, sehingga perbedaan antar studi hanya disebabkan oleh kesalahan sampling. Sebaliknya, random-effects model mengasumsikan bahwa effect size bervariasi antar studi dan mengikuti distribusi tertentu. Borenstein et al. (2009) menegaskan bahwa random-effects model lebih realistis dan umum digunakan dalam penelitian sosial dan pendidikan yang melibatkan konteks dan populasi yang beragam.

Selain mengestimasi effect size gabungan, meta-analisis juga memungkinkan eksplorasi faktor-faktor yang memoderasi variasi hasil penelitian melalui meta-regression atau analisis subkelompok. Pendekatan ini membantu peneliti memahami kondisi di mana suatu efek menjadi lebih kuat atau lebih lemah. Sebagai contoh, dalam meta-analisis efektivitas metode pembelajaran, variabel seperti tingkat pendidikan, durasi intervensi, atau karakteristik peserta didik dapat berfungsi sebagai moderator. Menurut Borenstein et al. (2009), analisis moderator meningkatkan nilai teoretis dan praktis meta-analisis karena menghubungkan temuan empiris dengan konteks implementasi.

Isu penting lainnya dalam meta-analisis adalah publication bias, yaitu kecenderungan studi dengan hasil signifikan lebih mungkin dipublikasikan dibandingkan studi dengan hasil non-signifikan. Bias ini dapat menyebabkan estimasi effect size gabungan menjadi terlalu besar. Hedges dan Olkin (1985) menekankan bahwa pengabaian publication bias dapat merusak validitas kesimpulan meta-analisis. Beberapa teknik digunakan untuk mendeteksi bias publikasi, antara lain *funnel plot*, *Egger's regression test*, dan metode *trim and fill*. Pendekatan-pendekatan

ini membantu peneliti menilai sejauh mana hasil meta-analisis dipengaruhi oleh bias seleksi publikasi.

Dalam praktik penelitian modern, meta-analisis didukung oleh berbagai perangkat lunak seperti Comprehensive Meta-Analysis (CMA), paket *meta* dan *metafor* dalam R, serta modul khusus dalam SPSS. Borenstein et al. (2009) menjelaskan bahwa perangkat lunak ini memudahkan perhitungan effect size, pengujian heterogenitas, dan visualisasi hasil melalui *forest plot*. Namun, peneliti tetap dituntut memahami konsep dan asumsi di balik analisis yang dilakukan agar tidak terjebak pada penggunaan teknis semata.

Secara reflektif, meta-analisis merepresentasikan puncak integrasi bukti empiris dalam riset kuantitatif modern. Dengan menggabungkan hasil berbagai studi secara sistematis dan statistik, meta-analisis memberikan dasar yang kuat bagi pengambilan keputusan berbasis bukti, pengembangan teori, dan rekomendasi kebijakan. Namun, kekuatan meta-analisis sangat bergantung pada kualitas studi primer, ketelitian metodologis, dan transparansi proses analisis. Oleh karena itu, meta-analisis harus dipandang bukan sekadar teknik statistik lanjutan, melainkan pendekatan ilmiah yang menuntut integritas metodologis dan pemahaman teoritis yang mendalam.

Daftar Rujukan

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Everitt, B. S. (2011). *Cluster analysis* (5th ed.). Wiley.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific Software International.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). *Partial least squares structural equation modeling*. Springer.

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance

of models. *Statistical Science, 31*(3), 332–349. https://doi.org/10.1214/16-STS597

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

BAB 9

SOFTWARE STATISTIK

9.1 IBM SPSS Statistics

IBM SPSS Statistics merupakan salah satu perangkat lunak statistik paling luas digunakan dalam penelitian kuantitatif, khususnya pada bidang pendidikan, psikologi, ilmu sosial, dan manajemen. Popularitas SPSS tidak terlepas dari kemampuannya menjembatani kompleksitas analisis statistik dengan antarmuka yang relatif mudah digunakan. Menurut Field (2018), SPSS dirancang untuk memfasilitasi peneliti dalam melakukan analisis statistik mulai dari tahap eksplorasi data, pengujian asumsi, hingga analisis inferensial dan multivariat, tanpa menuntut kemampuan pemrograman tingkat lanjut.

Secara filosofis, SPSS berakar kuat pada paradigma statistik klasik yang menekankan inferensi berbasis probabilitas, pengujian hipotesis, dan estimasi parameter. George dan Mallery (2020) menegaskan bahwa SPSS sangat cocok digunakan dalam penelitian yang mengikuti desain kuantitatif konvensional, seperti survei, eksperimen, dan kuasi-eksperimen. Dalam konteks ini, SPSS berfungsi sebagai alat untuk menerjemahkan desain penelitian ke dalam prosedur analisis statistik yang sistematis dan terstandar.

Salah satu keunggulan utama SPSS terletak pada kemampuannya menangani analisis statistik dasar hingga lanjutan secara terintegrasi. Untuk tahap awal analisis data, SPSS menyediakan fasilitas statistik deskriptif yang komprehensif, termasuk perhitungan mean, median, modus, standar deviasi, skewness, dan kurtosis. Field (2018) menekankan bahwa

eksplorasi data awal merupakan tahap krusial dalam *data* workflow karena memungkinkan peneliti mendeteksi kesalahan input, outlier, dan pola distribusi sebelum melakukan analisis inferensial.

Dalam konteks pengujian asumsi statistik, SPSS menawarkan berbagai uji yang esensial dalam analisis parametrik, seperti uji normalitas (Kolmogorov–Smirnov dan Shapiro–Wilk), uji homogenitas varians (Levene), dan uji linearitas. Menurut George dan Mallery (2020), kemudahan akses terhadap uji asumsi ini menjadikan SPSS sangat relevan dalam penelitian pendidikan dan sosial, di mana pelanggaran asumsi sering kali menjadi tantangan metodologis utama.

SPSS juga dikenal luas sebagai perangkat lunak utama untuk analisis inferensial klasik, termasuk uji *t*, ANOVA satu dan dua arah, ANCOVA, serta korelasi Pearson dan Spearman. Field (2018) menunjukkan bahwa SPSS memungkinkan peneliti tidak hanya menjalankan analisis tersebut, tetapi juga menyajikan output yang mendukung interpretasi hasil, seperti *effect size*, interval kepercayaan, dan grafik pendukung. Dengan demikian, SPSS berkontribusi langsung terhadap peningkatan kualitas pelaporan hasil penelitian.

Pada tingkat yang lebih lanjut, SPSS menyediakan fasilitas analisis multivariat, seperti MANOVA, analisis faktor eksploratori (EFA), analisis diskriminan, dan regresi berganda. George dan Mallery (2020) menegaskan bahwa meskipun SPSS bukan perangkat khusus untuk pemodelan struktural, kemampuannya dalam analisis multivariat menjadikannya fondasi penting sebelum peneliti beralih ke perangkat lunak lanjutan seperti AMOS atau SmartPLS. Dalam praktik penelitian, SPSS sering digunakan untuk praproses data dan eksplorasi struktur laten sebelum analisis SEM dilakukan.

Dalam konteks transformasi digital penelitian, SPSS juga beradaptasi dengan menyediakan fitur syntax-based analysis, yang memungkinkan analisis dilakukan secara lebih transparan dan replikatif. Menurut Field (2018), penggunaan syntax membantu peneliti mendokumentasikan setiap langkah analisis, sehingga meningkatkan reprodusibilitas penelitian. Meskipun banyak pengguna mengandalkan menu grafis, penggunaan syntax menjadi jembatan penting antara SPSS dan praktik *open science* yang menekankan transparansi analisis.

Namun demikian, SPSS memiliki sejumlah keterbatasan yang perlu dipertimbangkan secara kritis. Salah satu keterbatasan utama adalah sifatnya yang proprietary, sehingga memerlukan lisensi berbayar yang tidak selalu terjangkau bagi semua institusi. Selain itu, SPSS relatif kurang fleksibel dalam menangani analisis berbasis *big data* dan algoritma machine learning dibandingkan dengan bahasa pemrograman seperti R dan Python. Shmueli et al. (2019) menegaskan bahwa kebutuhan analitik modern yang berskala besar dan prediktif menuntut perangkat lunak yang lebih terbuka dan skalabel.

Meskipun demikian, dalam *workflow* penelitian kuantitatif modern, SPSS tetap memiliki posisi strategis sebagai perangkat analisis dasar–lanjutan yang stabil, terstandar, dan mudah diakses. SPSS sering menjadi titik awal pembelajaran statistik bagi mahasiswa dan peneliti pemula, sekaligus alat yang andal bagi peneliti berpengalaman dalam melakukan analisis inferensial konvensional. Integrasi SPSS dengan perangkat lunak lain, seperti AMOS dan R, memungkinkan peneliti membangun alur analisis yang komprehensif dan adaptif terhadap kompleksitas penelitian.

Sebagai refleksi metodologis, SPSS tidak seharusnya dipandang sekadar sebagai alat teknis, melainkan sebagai bagian dari ekosistem metodologi kuantitatif yang lebih luas. Pemilihan

SPSS harus didasarkan pada kesesuaian antara desain penelitian, jenis data, dan tujuan analisis. Dengan pemahaman yang tepat, SPSS dapat berfungsi sebagai fondasi kuat dalam menghasilkan analisis yang akurat, replikatif, dan bermakna dalam penelitian kuantitatif di era digital.

9.2 AMOS (Analysis of Moment Structures)

AMOS (Analysis of Moment Structures) merupakan perangkat lunak statistik yang dirancang khusus untuk analisis Structural Equation Modeling (SEM) berbasis kovarian. AMOS banyak digunakan dalam penelitian pendidikan, psikologi, manajemen, dan ilmu sosial yang bertujuan menguji model teoretis secara konfirmatori. Menurut Byrne (2016), AMOS sangat efektif untuk menguji hubungan laten yang kompleks karena mengintegrasikan analisis faktor konfirmatori (*Confirmatory Factor Analysis/CFA*) dan model struktural dalam satu kerangka analisis yang koheren.

Secara filosofis, AMOS berakar pada paradigma confirmatory modeling, yang menekankan pengujian kesesuaian model teoretis dengan data empiris. Berbeda dengan pendekatan eksploratif atau prediktif, SEM berbasis kovarian bertujuan meminimalkan perbedaan antara matriks kovarians yang diobservasi dan matriks kovarians yang diprediksi oleh model. Byrne (2016) menegaskan bahwa pendekatan ini sangat relevan ketika peneliti memiliki teori yang mapan dan hipotesis struktural yang jelas sebelum analisis dilakukan.

Salah satu fungsi utama AMOS adalah Confirmatory Factor Analysis (CFA), yang digunakan untuk menguji validitas konstruk dan struktur faktor instrumen pengukuran. CFA memungkinkan peneliti mengonfirmasi apakah indikator-indikator yang digunakan benar-benar merefleksikan konstruk laten yang

diasumsikan secara teoretis. Menurut Byrne (2016), CFA merupakan tahap krusial dalam SEM karena kualitas model struktural sangat bergantung pada validitas model pengukuran. Dalam praktik penelitian pendidikan dan psikologi, CFA sering digunakan untuk menguji instrumen sikap, motivasi, atau kompetensi yang bersifat laten.

AMOS menyediakan berbagai indeks kesesuaian model (*goodness of fit indices*) yang menjadi dasar evaluasi model SEM. Indeks-indeks tersebut meliputi *Chi-square*, *Comparative Fit Index (CFI)*, *Tucker–Lewis Index (TLI)*, *Root Mean Square Error of Approximation (RMSEA)*, dan *Standardized Root Mean Square Residual (SRMR)*. Byrne (2016) menekankan bahwa tidak ada satu indeks pun yang dapat digunakan secara tunggal untuk menilai kelayakan model, sehingga interpretasi harus didasarkan pada kombinasi beberapa indikator kesesuaian model.

Selain CFA, AMOS digunakan untuk membangun model struktural yang menggambarkan hubungan kausal antar konstruk laten. Model struktural memungkinkan peneliti menguji hipotesis tentang pengaruh langsung dan tidak langsung antar variabel, termasuk efek mediasi. Dalam konteks ini, AMOS menyediakan estimasi koefisien jalur (*path coefficients*) beserta nilai signifikansi statistiknya. Menurut Byrne (2016), interpretasi koefisien jalur dalam SEM harus mempertimbangkan konteks teoretis dan kesesuaian keseluruhan model, bukan hanya signifikansi individual.

Dari sisi teknis, AMOS menawarkan antarmuka grafis yang intuitif, memungkinkan peneliti membangun model SEM melalui *drag-and-drop*. Fitur ini menjadikan AMOS relatif mudah diakses oleh peneliti yang tidak memiliki latar belakang pemrograman. Gaskin dan Lim (2016) menambahkan bahwa plugin AMOS dapat memperluas kemampuan analisis, seperti perhitungan *composite*

reliability, *average variance extracted (AVE)*, dan visualisasi hasil yang lebih informatif. Kemudahan ini menjadikan AMOS populer di kalangan peneliti akademik dan mahasiswa pascasarjana.

Namun demikian, penggunaan AMOS menuntut pemenuhan asumsi statistik yang relatif ketat, termasuk normalitas multivariat, ukuran sampel yang memadai, dan spesifikasi model yang benar. Byrne (2016) menegaskan bahwa pelanggaran asumsi dapat berdampak signifikan terhadap estimasi parameter dan kesesuaian model. Oleh karena itu, AMOS sering digunakan setelah tahap eksplorasi data dan pengujian asumsi dilakukan menggunakan perangkat lunak seperti SPSS.

Dalam *workflow* penelitian kuantitatif modern, AMOS sering dibandingkan dengan PLS-SEM yang diimplementasikan melalui SmartPLS. Perbedaan utama antara AMOS dan SmartPLS terletak pada tujuan analisis dan pendekatan statistik. AMOS lebih sesuai untuk pengujian teori yang mapan dan evaluasi kesesuaian model secara ketat, sedangkan PLS-SEM lebih berorientasi pada prediksi dan eksplorasi model dengan asumsi yang lebih fleksibel. Hair et al. (2022) menegaskan bahwa pemilihan antara AMOS dan SmartPLS harus didasarkan pada tujuan penelitian, tahap pengembangan teori, dan karakteristik data.

Keterbatasan AMOS juga perlu dicermati secara kritis. Selain sifatnya yang berlisensi dan berbiaya relatif tinggi, AMOS kurang fleksibel dalam menangani model prediktif berskala besar dan data *non-normal* yang kompleks. Dalam konteks *big data* dan *machine learning*, AMOS tidak dirancang untuk analisis prediktif berbasis algoritma. Shmueli et al. (2019) menekankan bahwa kebutuhan analitik modern yang berorientasi prediksi

memerlukan pendekatan dan perangkat lunak yang berbeda dari SEM berbasis kovarian.

Meskipun demikian, AMOS tetap memiliki peran strategis dalam penelitian kuantitatif yang menekankan validasi teori dan konstruk. Dalam disiplin ilmu sosial dan pendidikan, di mana pengukuran konstruk laten menjadi isu sentral, AMOS menyediakan kerangka analisis yang kuat dan terstandar. Integrasi AMOS dengan SPSS memungkinkan peneliti membangun alur analisis yang sistematis, mulai dari eksplorasi data hingga pengujian model struktural yang kompleks.

Sebagai refleksi metodologis, AMOS sebaiknya dipandang sebagai alat konfirmatori yang melengkapi, bukan menggantikan, pendekatan analisis lainnya. Pemahaman yang mendalam tentang teori SEM, asumsi statistik, dan interpretasi model sangat menentukan kualitas hasil analisis. Dengan penggunaan yang tepat, AMOS dapat meningkatkan ketepatan inferensi, validitas konstruk, dan kontribusi teoretis penelitian kuantitatif di era digital.

9.3 SmartPLS

SmartPLS merupakan perangkat lunak statistik yang dikembangkan untuk mengimplementasikan Partial Least Squares Structural Equation Modeling (PLS-SEM), yaitu pendekatan pemodelan struktural berbasis varians yang berorientasi pada prediksi dan pengembangan teori. Dalam beberapa dekade terakhir, SmartPLS menjadi salah satu perangkat lunak yang paling banyak digunakan dalam penelitian manajemen, bisnis, pendidikan, dan ilmu sosial yang melibatkan model struktural kompleks. Menurut Hair, Hult, Ringle, dan Sarstedt (2022), SmartPLS sangat sesuai digunakan ketika tujuan utama penelitian adalah memaksimalkan varians terjelaskan

(*explained variance*) dan mengevaluasi kemampuan prediktif model, bukan sekadar menguji kesesuaian model teoretis.

Secara filosofis, PLS-SEM yang diimplementasikan dalam SmartPLS berangkat dari paradigma predictive modeling, berbeda dari SEM berbasis kovarian yang bersifat konfirmatori. PLS-SEM tidak berfokus pada reproduksi matriks kovarians, melainkan pada estimasi skor variabel laten yang mampu memprediksi variabel endogen secara optimal. Hair et al. (2022) menegaskan bahwa pendekatan ini sangat relevan dalam riset sosial kontemporer yang menghadapi keterbatasan ukuran sampel, distribusi data yang tidak normal, dan model dengan kompleksitas tinggi.

SmartPLS menyediakan lingkungan grafis yang intuitif untuk membangun model pengukuran (outer model) dan model struktural (inner model) secara simultan. Model pengukuran dalam SmartPLS dapat berupa konstruk reflektif maupun formatif, suatu fleksibilitas yang menjadi keunggulan utama PLS-SEM. Menurut Sarstedt, Ringle, dan Hair (2017), kemampuan memodelkan konstruk formatif secara eksplisit sangat penting dalam penelitian manajemen dan pendidikan, di mana indikator sering kali membentuk konstruk secara kausal, bukan sekadar merefleksikannya.

Evaluasi model pengukuran reflektif dalam SmartPLS mencakup pengujian reliabilitas indikator (*outer loadings*), reliabilitas konstruk (*Composite Reliability*), validitas konvergen (*Average Variance Extracted/AVE*), serta validitas diskriminan menggunakan kriteria Fornell–Larcker dan HTMT (Heterotrait–Monotrait Ratio). Hair et al. (2022) menekankan bahwa HTMT merupakan pendekatan yang lebih sensitif dan direkomendasikan dalam praktik PLS-SEM modern. Sementara itu, evaluasi konstruk

formatif menitikberatkan pada pengujian multikolinearitas dan signifikansi bobot indikator (*outer weights*).

Pada tahap model struktural, SmartPLS memungkinkan peneliti mengevaluasi hubungan antar konstruk laten melalui koefisien jalur, nilai R^2, *effect size* (f^2), serta *predictive relevance* (Q^2). Salah satu ciri khas SmartPLS adalah penggunaan bootstrapping untuk menguji signifikansi parameter tanpa bergantung pada asumsi normalitas. Hair et al. (2022) menjelaskan bahwa prosedur bootstrapping memungkinkan estimasi *standard error*, nilai *t*, dan interval kepercayaan, sehingga meningkatkan ketahanan inferensi statistik dalam kondisi data yang kompleks.

SmartPLS juga unggul dalam mendukung analisis mediasi dan moderasi secara langsung dalam kerangka PLS-SEM. Peneliti dapat menguji efek langsung, tidak langsung, dan total dalam satu model terintegrasi. Selain itu, SmartPLS menyediakan fitur Importance–Performance Map Analysis (IPMA) yang memungkinkan evaluasi simultan antara pentingnya suatu konstruk dan kinerjanya. Menurut Hair et al. (2022), IPMA sangat berguna dalam penelitian terapan karena menghubungkan hasil statistik dengan implikasi manajerial atau kebijakan.

Dalam *workflow* penelitian kuantitatif modern, SmartPLS sering diposisikan sebagai alternatif atau pelengkap AMOS. Perbandingan antara SmartPLS dan AMOS mencerminkan perbedaan orientasi metodologis yang mendasar. AMOS lebih tepat digunakan untuk pengujian teori yang mapan dan evaluasi *model fit* global, sedangkan SmartPLS lebih sesuai untuk pengembangan teori, eksplorasi hubungan kompleks, dan analisis prediktif. Hair et al. (2022) menegaskan bahwa pemilihan perangkat lunak harus disesuaikan dengan tujuan penelitian, bukan preferensi teknis semata.

Dari sisi keterbatasan, SmartPLS tidak menyediakan indeks *goodness of fit* global yang seketat SEM berbasis kovarian. Meskipun beberapa ukuran kesesuaian telah dikembangkan, Sarstedt et al. (2017) mengingatkan bahwa PLS-SEM tidak dimaksudkan untuk menggantikan CB-SEM dalam pengujian teori yang telah matang. Oleh karena itu, penggunaan SmartPLS harus disertai pemahaman yang jelas tentang tujuan analisis dan keterbatasan inferensinya.

SmartPLS juga berperan penting dalam menjembatani analisis struktural dengan pendekatan predictive analytics dan machine learning. Shmueli et al. (2019) menegaskan bahwa orientasi prediktif PLS-SEM sejalan dengan kebutuhan analitik modern yang menekankan akurasi prediksi dan relevansi praktis. Dalam konteks ini, SmartPLS dapat diposisikan sebagai bagian dari ekosistem analitik yang lebih luas, berdampingan dengan R dan Python dalam menangani data yang semakin kompleks.

Sebagai refleksi metodologis, SmartPLS tidak sekadar alat teknis, melainkan representasi dari pergeseran paradigma penelitian kuantitatif menuju pendekatan yang lebih fleksibel, prediktif, dan kontekstual. Dengan kemampuan menangani model kompleks, asumsi yang lebih longgar, dan fokus pada varians terjelaskan, SmartPLS menjadi pilihan strategis bagi peneliti yang ingin menggabungkan teori dan eksplorasi data. Namun, kualitas analisis tetap bergantung pada ketepatan desain penelitian, kekuatan teori, dan kehati-hatian dalam interpretasi hasil.

9.4 JASP dan R

Perkembangan penelitian kuantitatif di era digital ditandai oleh pergeseran signifikan menuju penggunaan perangkat lunak open-source yang menekankan transparansi, replikasi, dan

fleksibilitas analisis. Dalam konteks ini, JASP dan R menempati posisi strategis sebagai alternatif terhadap perangkat lunak statistik berlisensi. Keduanya tidak hanya menawarkan kemampuan analisis yang setara dengan perangkat lunak komersial, tetapi juga mendukung praktik reproducible research yang semakin menjadi tuntutan utama dalam komunitas ilmiah internasional. Love et al. (2019) menegaskan bahwa adopsi perangkat lunak open-source merupakan langkah penting untuk meningkatkan integritas dan keterbukaan penelitian kuantitatif.

JASP (Jeffreys's Amazing Statistics Program) dikembangkan sebagai perangkat lunak statistik berbasis antarmuka grafis yang mudah digunakan, khususnya bagi pengguna yang terbiasa dengan SPSS. Secara filosofis, JASP menggabungkan kemudahan penggunaan dengan pendekatan Bayesian statistics, suatu ciri khas yang membedakannya dari banyak perangkat lunak statistik lain. Menurut Love et al. (2019), JASP dirancang untuk menjembatani kesenjangan antara statistik klasik dan Bayesian, sehingga memungkinkan peneliti memilih pendekatan analisis yang paling sesuai dengan tujuan dan filosofi penelitiannya.

Dalam praktik penelitian, JASP menyediakan berbagai analisis statistik dasar dan lanjutan, termasuk statistik deskriptif, uji *t*, ANOVA, regresi, korelasi, analisis faktor, hingga model Bayesian yang kompleks. Salah satu keunggulan utama JASP adalah kemampuannya menyajikan output analisis secara otomatis dalam format yang siap dilaporkan, lengkap dengan nilai *effect size* dan interval kepercayaan. Love et al. (2019) menekankan bahwa fitur ini sangat membantu peneliti dan mahasiswa dalam meningkatkan kualitas pelaporan hasil penelitian secara konsisten dan transparan.

Selain kemudahan teknis, JASP juga mendukung reproducibility melalui sistem *analysis workflow* yang

terdokumentasi. Setiap perubahan pada data atau pengaturan analisis akan langsung tercermin dalam output, sehingga meminimalkan kesalahan pelaporan. Pendekatan ini sejalan dengan prinsip *open science* yang menuntut keterlacakan setiap langkah analisis. Dalam konteks pendidikan tinggi, JASP sering digunakan sebagai alat pengajaran statistik karena memudahkan mahasiswa memahami hubungan antara konsep statistik dan hasil analisis.

Berbeda dengan JASP yang berorientasi pada antarmuka grafis, R merupakan bahasa pemrograman dan lingkungan komputasi statistik yang menawarkan fleksibilitas dan kekuatan analitik yang sangat tinggi. R Core Team (2022) menjelaskan bahwa R dikembangkan sebagai lingkungan terpadu untuk manipulasi data, perhitungan statistik, dan visualisasi grafis. Dengan ribuan paket yang tersedia, R mampu menangani hampir seluruh spektrum analisis kuantitatif, mulai dari statistik dasar hingga *machine learning* dan *big data analytics*.

Secara metodologis, R merepresentasikan paradigma script-based analysis, di mana setiap langkah analisis ditulis dalam bentuk kode. Pendekatan ini memungkinkan dokumentasi yang lengkap dan eksplisit dari seluruh proses analisis, sehingga meningkatkan reprodusibilitas dan transparansi penelitian. Kuhn dan Johnson (2013) menegaskan bahwa penggunaan R mendorong peneliti untuk berpikir secara sistematis tentang alur analisis, mulai dari praproses data hingga interpretasi hasil.

Dalam konteks riset sosial dan pendidikan, R sering digunakan untuk analisis statistik lanjutan seperti regresi multilevel, SEM, PLS-SEM, *machine learning*, dan meta-analisis. Paket-paket seperti *lavaan*, *caret*, *tidymodels*, dan *metafor* menyediakan kerangka kerja yang komprehensif untuk analisis kompleks. James et al. (2021) menekankan bahwa fleksibilitas R

memungkinkan integrasi antara analisis inferensial dan prediktif, suatu kebutuhan utama dalam penelitian kuantitatif modern.

Salah satu kekuatan utama R adalah kemampuannya berintegrasi dengan prinsip reproducible research melalui penggunaan *R Markdown* dan *Quarto*. Dengan pendekatan ini, kode analisis, output statistik, dan narasi laporan dapat digabungkan dalam satu dokumen yang dinamis. Shmueli et al. (2019) menegaskan bahwa praktik ini tidak hanya meningkatkan efisiensi penelitian, tetapi juga memperkuat kredibilitas ilmiah melalui transparansi penuh terhadap proses analisis.

Meskipun menawarkan fleksibilitas tinggi, penggunaan R juga memiliki tantangan tersendiri, terutama terkait kurva pembelajaran yang relatif curam bagi pengguna pemula. Berbeda dengan JASP dan SPSS yang berbasis menu, R menuntut pemahaman dasar pemrograman dan logika analisis. Namun, Kuhn dan Johnson (2013) menegaskan bahwa investasi waktu untuk mempelajari R akan terbayar melalui peningkatan kemampuan analitik dan kontrol penuh terhadap proses penelitian.

Dalam *workflow* penelitian kuantitatif modern, JASP dan R sering digunakan secara komplementer. JASP cocok digunakan pada tahap eksplorasi awal dan pengajaran statistik, sedangkan R digunakan untuk analisis lanjutan, replikasi, dan pengembangan model yang lebih kompleks. Integrasi keduanya mencerminkan pergeseran menuju ekosistem analitik yang terbuka, fleksibel, dan berorientasi pada kualitas ilmiah.

Sebagai refleksi metodologis, adopsi JASP dan R menandai transformasi penting dalam praktik penelitian kuantitatif menuju keterbukaan, akuntabilitas, dan inovasi metodologis. Dengan memanfaatkan kekuatan open-source analytics, peneliti dapat meningkatkan akurasi analisis, memfasilitasi replikasi, dan

berkontribusi pada pengembangan ilmu pengetahuan yang lebih transparan dan kolaboratif di era digital.

9.5 Big Data Analytics

Big Data Analytics merupakan pendekatan analisis data yang dirancang untuk menangani data dengan karakteristik volume, velocity, dan variety yang jauh melampaui kapasitas metode statistik dan perangkat lunak konvensional. Dalam konteks penelitian sosial dan pendidikan modern, big data tidak lagi terbatas pada data survei berskala kecil, tetapi mencakup jejak digital (*digital traces*), data pembelajaran daring, media sosial, log sistem informasi pendidikan, dan data administratif berskala nasional. Provost dan Fawcett (2013) menegaskan bahwa big data analytics menandai pergeseran mendasar dari analisis berbasis sampel menuju pemanfaatan data populasi yang masif dan dinamis.

Secara filosofis, big data analytics berakar pada paradigma data-driven science, di mana pola dan prediksi sering kali dieksplorasi langsung dari data tanpa ketergantungan penuh pada hipotesis awal. Pendekatan ini berbeda dari statistik inferensial klasik yang menempatkan teori sebagai titik awal analisis. Shmueli et al. (2019) menegaskan bahwa dalam big data analytics, tujuan utama sering kali adalah prediksi, *pattern discovery*, dan pengambilan keputusan berbasis data, bukan sekadar pengujian signifikansi statistik. Meskipun demikian, pendekatan ini tidak meniadakan peran teori, melainkan menuntut integrasi yang lebih dinamis antara teori dan eksplorasi data.

Dari sisi infrastruktur, big data analytics sangat bergantung pada teknologi komputasi terdistribusi seperti Hadoop dan Apache Spark. Hadoop menyediakan kerangka kerja

penyimpanan dan pemrosesan data terdistribusi melalui *Hadoop Distributed File System (HDFS)*, sedangkan Spark menawarkan kecepatan pemrosesan yang lebih tinggi melalui pemrosesan *in-memory*. Menurut Provost dan Fawcett (2013), teknologi ini memungkinkan analisis data berskala besar yang sebelumnya tidak mungkin dilakukan dengan perangkat lunak statistik tradisional seperti SPSS atau AMOS.

Dalam praktik penelitian sosial dan pendidikan, big data analytics sering diimplementasikan melalui bahasa pemrograman Python dan R, yang menyediakan ekosistem pustaka (*libraries*) untuk pemrosesan data besar dan machine learning. Python, dengan pustaka seperti *pandas*, *NumPy*, *scikit-learn*, dan *TensorFlow*, menjadi pilihan utama dalam pengembangan model prediktif dan analisis perilaku berbasis data digital. James et al. (2021) menegaskan bahwa fleksibilitas Python memungkinkan integrasi antara analisis statistik, machine learning, dan visualisasi data dalam satu alur kerja yang kohesif.

Salah satu ciri utama big data analytics adalah pemanfaatan machine learning dan artificial intelligence untuk mengekstraksi pola kompleks dari data. Algoritma seperti *random forest*, *gradient boosting*, *neural networks*, dan *deep learning* memungkinkan analisis hubungan nonlinier dan interaksi tingkat tinggi yang sulit ditangkap oleh model statistik linear. Breiman (2001) menunjukkan bahwa pendekatan *ensemble learning* seperti random forest memiliki keunggulan dalam akurasi prediksi dan ketahanan terhadap *overfitting*, menjadikannya sangat relevan dalam analisis data sosial berskala besar.

Selain aspek teknis, big data analytics juga menuntut perhatian serius terhadap isu metodologis dan etis. Dalam penelitian sosial dan pendidikan, penggunaan data digital sering kali melibatkan isu privasi, keamanan data, dan bias algoritmik.

Shmueli et al. (2019) menegaskan bahwa akurasi prediksi yang tinggi tidak selalu menjamin keadilan atau validitas substantif hasil analisis. Oleh karena itu, peneliti dituntut untuk mengintegrasikan prinsip etika penelitian, transparansi algoritma, dan interpretabilitas model dalam praktik big data analytics.

Big data analytics juga mendorong perubahan signifikan dalam workflow penelitian kuantitatif. Proses analisis tidak lagi bersifat linear, melainkan iteratif dan kolaboratif, melibatkan tahap eksplorasi data, pengembangan model, evaluasi prediktif, dan penyempurnaan berkelanjutan. Dalam konteks ini, komputasi awan (*cloud computing*) memainkan peran penting dengan menyediakan sumber daya komputasi yang skalabel dan fleksibel. Platform cloud memungkinkan peneliti mengakses kapasitas pemrosesan tinggi tanpa investasi infrastruktur fisik yang besar, sehingga memperluas akses terhadap analisis big data.

Dalam bidang pendidikan, big data analytics membuka peluang baru melalui *learning analytics* dan *educational data mining*, yang memungkinkan pemantauan proses belajar secara real-time dan pengembangan intervensi berbasis data. Dalam penelitian sosial, analisis media sosial dan data digital lainnya memungkinkan pemahaman fenomena sosial secara lebih dinamis dan kontekstual. Provost dan Fawcett (2013) menekankan bahwa nilai utama big data analytics terletak pada kemampuannya menghasilkan wawasan yang relevan dan dapat ditindaklanjuti (*actionable insights*), bukan sekadar volume data yang besar.

Meskipun menawarkan potensi besar, big data analytics tidak dapat menggantikan sepenuhnya metode statistik klasik dan inferensial. Sebaliknya, pendekatan ini harus dipandang

sebagai pelengkap yang memperluas cakupan analisis kuantitatif. Integrasi antara big data analytics, statistik inferensial, dan teori substantif menjadi kunci dalam menghasilkan penelitian yang tidak hanya akurat secara teknis, tetapi juga bermakna secara ilmiah dan sosial.

Sebagai refleksi metodologis penutup BAB 9, big data analytics menandai transformasi fundamental dalam praktik penelitian kuantitatif di era digital. Pemanfaatan teknologi komputasi terdistribusi, machine learning, dan cloud computing memungkinkan analisis data yang lebih luas, cepat, dan prediktif. Namun, keberhasilan big data analytics sangat bergantung pada kemampuan peneliti untuk mengintegrasikan teknologi dengan teori, metodologi yang kuat, dan kesadaran etis. Dengan pendekatan yang seimbang, big data analytics dapat menjadi pilar utama dalam pengembangan penelitian sosial dan pendidikan yang relevan, akurat, dan berorientasi masa depan.

Daftar Rujukan

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Gaskin, J., & Lim, J. (2016). *AMOS plugin*. Gaskination's StatWiki.

George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference* (16th ed.). Routledge.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Love, J., Selker, R., Marsman, M., Jamil, T., Verhagen, J., Ly, A., ... Wagenmakers, E.-J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software, 88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). *Data mining for business analytics: Concepts, techniques, and applications in R* (2nd ed.). Springer.

IBM Corp. (2021). *IBM SPSS Statistics documentation*. IBM Corporation.

BAB 10

PELAPORAN HASIL PENELITIAN

10.1 Struktur Laporan Penelitian

Struktur laporan penelitian merupakan kerangka sistematis yang digunakan untuk menyajikan proses dan hasil penelitian secara logis, transparan, dan dapat dipertanggungjawabkan secara ilmiah. Dalam penelitian kuantitatif di bidang pendidikan, sosial, dan manajemen, struktur laporan tidak hanya berfungsi sebagai format administratif, tetapi juga sebagai instrumen epistemologis yang memastikan keterhubungan antara rumusan masalah, metode, analisis data, dan kesimpulan. American Psychological Association (APA, 2020) menegaskan bahwa struktur laporan yang baku memungkinkan pembaca menilai kualitas metodologis penelitian serta mereplikasi prosedur analisis yang dilakukan.

Secara umum, laporan penelitian kuantitatif mengikuti struktur formal yang terdiri atas judul, abstrak, pendahuluan, metode, hasil, diskusi, dan kesimpulan, disertai daftar rujukan dan lampiran jika diperlukan. Creswell (2014) menyatakan bahwa struktur ini mencerminkan alur logika penelitian ilmiah, mulai dari perumusan masalah hingga interpretasi temuan. Konsistensi dalam mengikuti struktur tersebut menjadi prasyarat utama agar laporan penelitian dapat diterima dalam forum akademik internasional, termasuk jurnal bereputasi dan laporan tesis atau disertasi.

Bagian judul merupakan representasi ringkas dari fokus penelitian dan harus mencerminkan variabel utama, subjek

penelitian, serta konteks kajian. APA (2020) menekankan bahwa judul yang baik bersifat informatif, spesifik, dan bebas dari bias interpretatif. Dalam penelitian kuantitatif, judul sebaiknya menghindari klaim kausal yang berlebihan kecuali didukung oleh desain eksperimen yang kuat. Judul yang jelas memudahkan pembaca mengidentifikasi relevansi penelitian sejak awal.

Abstrak berfungsi sebagai ringkasan komprehensif dari seluruh laporan penelitian. Menurut APA (2020), abstrak penelitian kuantitatif idealnya mencakup tujuan penelitian, desain dan metode, karakteristik sampel, teknik analisis, serta temuan utama. Creswell (2014) menegaskan bahwa abstrak bukan sekadar ringkasan deskriptif, melainkan representasi akurat dari isi penelitian yang memungkinkan pembaca menilai relevansi dan kontribusi penelitian tanpa membaca keseluruhan laporan.

Bagian pendahuluan merupakan fondasi teoretis dan konseptual penelitian. Neuman (2014) menyatakan bahwa pendahuluan harus menguraikan latar belakang masalah, kesenjangan penelitian, tujuan penelitian, dan hipotesis atau pertanyaan penelitian secara sistematis. Dalam laporan kuantitatif, pendahuluan juga harus mengaitkan teori dengan variabel yang diteliti serta menjelaskan rasional metodologis pemilihan pendekatan analisis. Pendahuluan yang kuat menunjukkan pemahaman peneliti terhadap konteks ilmiah dan relevansi praktis penelitian.

Bagian metode penelitian memiliki peran sentral dalam menjamin transparansi dan reprodusibilitas penelitian. APA (2020) menekankan bahwa metode harus dijelaskan secara rinci, mencakup desain penelitian, partisipan atau sampel, instrumen pengukuran, prosedur pengumpulan data, dan teknik analisis statistik. Fraenkel, Wallen, dan Hyun (2021) menegaskan bahwa

kejelasan metode memungkinkan pembaca menilai validitas internal dan eksternal penelitian, serta mereplikasi studi dalam konteks yang berbeda.

Bagian hasil penelitian menyajikan temuan empiris secara objektif tanpa interpretasi yang berlebihan. Field (2018) menekankan bahwa pelaporan hasil harus berfokus pada statistik yang relevan, seperti nilai mean, standar deviasi, koefisien hubungan, nilai signifikansi, *effect size*, dan interval kepercayaan. Penyajian hasil sebaiknya didukung oleh tabel dan grafik yang mengikuti standar APA agar informasi dapat dipahami secara akurat oleh pembaca. Dalam bagian ini, peneliti harus menghindari diskusi implikasi atau spekulasi teoretis.

Bagian diskusi merupakan ruang utama untuk interpretasi dan integrasi temuan penelitian dengan teori dan penelitian sebelumnya. Gall, Gall, dan Borg (2019) menekankan bahwa diskusi yang baik tidak sekadar mengulang hasil, tetapi menjelaskan makna temuan, implikasi teoretis dan praktis, serta keterbatasan penelitian. Dalam penelitian kuantitatif, diskusi juga harus menafsirkan signifikansi statistik secara substantif, bukan hanya formal, dengan mempertimbangkan *effect size* dan konteks penelitian.

Bagian kesimpulan merangkum temuan utama dan kontribusi penelitian secara ringkas dan jelas. Sekaran dan Bougie (2020) menyatakan bahwa kesimpulan harus menjawab tujuan penelitian dan memberikan rekomendasi yang proporsional dengan bukti empiris yang tersedia. Dalam laporan akademik, kesimpulan tidak boleh memperkenalkan temuan baru, melainkan mensintesis hasil dan diskusi secara integratif.

Selain bagian utama, daftar rujukan dan lampiran juga merupakan komponen penting dalam struktur laporan penelitian. APA (2020) menegaskan bahwa setiap sumber yang

dirujuk dalam teks harus dicantumkan secara lengkap dan konsisten dalam daftar rujukan. Lampiran digunakan untuk menyajikan informasi tambahan seperti instrumen penelitian, tabel analisis tambahan, atau output statistik yang terlalu rinci untuk dimasukkan dalam tubuh utama laporan.

Sebagai refleksi metodologis, struktur laporan penelitian bukanlah sekadar format teknis, melainkan kerangka epistemologis yang menjamin koherensi, transparansi, dan kredibilitas penelitian kuantitatif. Dengan mengikuti struktur yang baku dan berstandar internasional, peneliti tidak hanya memudahkan pembaca memahami penelitian, tetapi juga berkontribusi pada praktik ilmiah yang etis, replikatif, dan bermutu tinggi.

10.2 Interpretasi Data Penelitian

Interpretasi data merupakan tahap krusial dalam pelaporan hasil penelitian karena menjadi jembatan antara temuan statistik dan makna substantif penelitian. Dalam penelitian kuantitatif di bidang pendidikan, sosial, dan manajemen, interpretasi data tidak sekadar melaporkan angka atau hasil uji statistik, melainkan menjelaskan arti temuan tersebut dalam konteks teori, tujuan penelitian, dan implikasi praktis. Creswell (2014) menegaskan bahwa interpretasi data yang baik harus didasarkan pada pemahaman metodologis yang kuat serta kesadaran terhadap keterbatasan desain dan analisis yang digunakan.

Salah satu kesalahan umum dalam interpretasi data adalah penyamaan antara signifikansi statistik dan signifikansi substantif. Field (2018) menjelaskan bahwa nilai *p* hanya menunjukkan probabilitas hasil yang diperoleh jika hipotesis nol benar, bukan ukuran kekuatan atau pentingnya suatu efek. Oleh karena itu, interpretasi hasil penelitian tidak boleh berhenti pada pernyataan

“signifikan” atau “tidak signifikan”, melainkan harus mempertimbangkan besaran efek (*effect size*) dan konteks penelitian. Pendekatan ini sejalan dengan anjuran APA (2020) yang menekankan pentingnya pelaporan *effect size* dan interval kepercayaan.

Dalam interpretasi statistik deskriptif, peneliti harus menjelaskan makna nilai mean, median, standar deviasi, dan distribusi data secara substantif. Neuman (2014) menyatakan bahwa statistik deskriptif berfungsi memberikan gambaran awal tentang karakteristik data dan konteks empiris penelitian. Misalnya, perbedaan rata-rata skor antara dua kelompok harus diinterpretasikan dengan mempertimbangkan variasi data dan relevansi perbedaan tersebut terhadap fenomena yang diteliti, bukan hanya perbedaan numerik semata.

Interpretasi statistik inferensial menuntut kehati-hatian yang lebih tinggi karena berkaitan dengan generalisasi hasil penelitian. Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa peneliti harus memahami asumsi yang mendasari setiap uji statistik sebelum menarik kesimpulan. Pelanggaran asumsi seperti normalitas atau homogenitas varians dapat memengaruhi validitas inferensi. Oleh karena itu, interpretasi hasil uji *t*, ANOVA, regresi, atau SEM harus selalu dikaitkan dengan hasil pengujian asumsi yang relevan.

Penggunaan effect size menjadi elemen penting dalam interpretasi data modern. Cohen (1988) menegaskan bahwa *effect size* memberikan informasi tentang kekuatan hubungan atau perbedaan yang diamati, terlepas dari ukuran sampel. Dalam penelitian pendidikan, misalnya, perbedaan yang secara statistik signifikan tetapi memiliki *effect size* kecil mungkin tidak memiliki implikasi praktis yang berarti. Oleh karena itu,

interpretasi data harus menyeimbangkan antara signifikansi statistik dan signifikansi praktis.

Selain *effect size*, interval kepercayaan (confidence interval) juga berperan penting dalam interpretasi data. APA (2020) merekomendasikan pelaporan interval kepercayaan karena memberikan informasi tentang ketidakpastian estimasi dan rentang nilai yang mungkin dari parameter populasi. Creswell (2014) menegaskan bahwa interpretasi interval kepercayaan membantu pembaca memahami stabilitas dan presisi temuan penelitian, sehingga meningkatkan kualitas inferensi ilmiah.

Dalam penelitian kuantitatif lanjutan, interpretasi data juga mencakup pemahaman terhadap hubungan kompleks seperti mediasi, moderasi, dan model struktural. Neuman (2014) menekankan bahwa peneliti harus berhati-hati dalam menafsirkan hubungan kausal, terutama dalam desain non-eksperimental. Interpretasi mediasi, misalnya, harus menekankan mekanisme hubungan antarvariabel, bukan sekadar keberadaan jalur tidak langsung yang signifikan. Demikian pula, interpretasi moderasi harus menjelaskan kondisi di mana hubungan antarvariabel menjadi lebih kuat atau lebih lemah.

Aspek etika juga menjadi bagian integral dari interpretasi data. APA (2020) menegaskan bahwa peneliti harus menghindari *overinterpretation*, *cherry-picking* hasil, atau pelaporan selektif yang dapat menyesatkan pembaca. Kejujuran ilmiah menuntut peneliti untuk melaporkan temuan yang tidak mendukung hipotesis awal dan mendiskusikan kemungkinan alternatif penjelasan. Dalam konteks *reproducible research*, transparansi interpretasi menjadi syarat utama untuk menjaga integritas ilmiah.

Interpretasi data yang baik juga harus mempertimbangkan keterbatasan penelitian. Gall et al. (2019) menekankan bahwa

pengakuan terhadap keterbatasan bukanlah kelemahan, melainkan bagian dari praktik ilmiah yang bertanggung jawab. Keterbatasan terkait ukuran sampel, instrumen pengukuran, atau konteks penelitian harus dijelaskan secara eksplisit agar pembaca dapat menilai ruang lingkup generalisasi temuan.

Sebagai refleksi metodologis, interpretasi data merupakan proses sintesis yang menuntut keseimbangan antara ketelitian statistik dan pemahaman substantif. Peneliti dituntut untuk tidak hanya "membaca" output statistik, tetapi juga "menerjemahkan" maknanya dalam bahasa ilmiah yang jelas dan bertanggung jawab. Dengan interpretasi data yang tepat, hasil penelitian kuantitatif dapat memberikan kontribusi yang bermakna bagi pengembangan teori, praktik, dan kebijakan di bidang pendidikan, sosial, dan manajemen.

10.3 Penyajian Tabel dan Grafik

Penyajian tabel dan grafik merupakan komponen penting dalam pelaporan hasil penelitian kuantitatif karena berfungsi sebagai sarana komunikasi visual yang memperjelas temuan statistik. Dalam penelitian pendidikan, sosial, dan manajemen, tabel dan grafik tidak hanya menyajikan data, tetapi juga membantu pembaca memahami pola, perbandingan, dan hubungan antarvariabel secara cepat dan akurat. American Psychological Association (APA, 2020) menegaskan bahwa visualisasi data yang baik harus akurat, jelas, konsisten, dan mendukung narasi ilmiah, bukan menggantikannya.

Secara metodologis, tabel dan grafik memiliki fungsi yang berbeda namun saling melengkapi. Tabel digunakan untuk menyajikan data numerik secara rinci dan presisi, seperti statistik deskriptif, koefisien regresi, atau hasil uji inferensial. Grafik, di sisi lain, digunakan untuk menyoroti pola, tren, dan perbandingan

antar kelompok yang sulit ditangkap melalui angka semata. Weissgerber et al. (2015) menyatakan bahwa pemilihan antara tabel dan grafik harus didasarkan pada tujuan komunikasi, bukan preferensi estetika semata.

Dalam standar APA 7, penyajian tabel mengikuti kaidah format yang ketat. Setiap tabel harus diberi nomor urut sesuai kemunculannya dalam teks dan judul tabel ditempatkan di atas tabel dengan format miring (*italic*). APA (2020) menekankan bahwa judul tabel harus ringkas namun informatif, sehingga pembaca dapat memahami isi tabel tanpa harus merujuk kembali ke teks utama. Kolom dan baris tabel harus diberi label yang jelas, serta menggunakan satuan pengukuran yang konsisten.

Selain format, isi tabel harus dipilih secara selektif. Field (2018) menegaskan bahwa tabel sebaiknya hanya memuat informasi yang relevan dengan pertanyaan penelitian dan analisis yang dilaporkan. Menyajikan terlalu banyak tabel atau data mentah yang tidak diperlukan dapat membebani pembaca dan mengaburkan pesan utama penelitian. Oleh karena itu, peneliti harus menyeimbangkan antara kelengkapan informasi dan kejelasan komunikasi.

Grafik dalam laporan penelitian kuantitatif juga harus mengikuti prinsip dan standar yang jelas. APA (2020) merekomendasikan penggunaan grafik yang sederhana, dengan elemen visual yang minimal namun informatif. Jenis grafik yang umum digunakan meliputi grafik batang untuk perbandingan kategori, grafik garis untuk menunjukkan tren, diagram pencar (scatterplot) untuk menggambarkan hubungan antarvariabel, dan boxplot untuk menunjukkan distribusi data. Pemilihan jenis grafik harus disesuaikan dengan sifat data dan tujuan analisis.

Prinsip utama dalam visualisasi data adalah kejujuran visual. Tufte (2001) menekankan bahwa grafik tidak boleh

memanipulasi persepsi pembaca melalui skala yang menyesatkan, pemotongan sumbu yang tidak proporsional, atau penggunaan warna yang berlebihan. Grafik manipulatif dapat memperbesar atau memperkecil perbedaan data secara artifisial, sehingga melanggar etika pelaporan ilmiah. Oleh karena itu, peneliti harus memastikan bahwa skala dan proporsi grafik mencerminkan data secara akurat.

Selain aspek kejujuran, keterbacaan juga menjadi kriteria penting dalam penyajian tabel dan grafik. Weissgerber et al. (2015) menekankan bahwa ukuran huruf, kontras warna, dan tata letak harus disesuaikan agar grafik dapat dibaca dengan mudah oleh berbagai audiens. Dalam konteks publikasi ilmiah, grafik sering dicetak dalam hitam-putih, sehingga penggunaan warna harus mempertimbangkan kemungkinan kehilangan informasi ketika warna tidak tersedia.

Integrasi tabel dan grafik dengan narasi teks juga merupakan aspek metodologis yang krusial. APA (2020) menegaskan bahwa setiap tabel dan grafik harus dirujuk dan dijelaskan dalam teks, bukan dibiarkan "berbicara sendiri". Peneliti harus menyoroti temuan utama yang ditampilkan dalam tabel atau grafik dan mengaitkannya dengan tujuan penelitian. Namun, peneliti juga harus menghindari pengulangan data numerik secara verbatim dalam teks, karena informasi tersebut sudah tersedia dalam visualisasi.

Dalam praktik pelaporan hasil penelitian kuantitatif modern, visualisasi data juga berkaitan dengan prinsip reproducible research. Field (2018) menyatakan bahwa tabel dan grafik sebaiknya dihasilkan langsung dari data dan analisis statistik, bukan dimodifikasi secara manual. Pendekatan ini mengurangi risiko kesalahan dan meningkatkan transparansi proses analisis. Penggunaan perangkat lunak statistik dan bahasa

pemrograman seperti SPSS, R, atau Python memungkinkan pembuatan visualisasi yang konsisten dan dapat direplikasi.

Kesalahan umum dalam penyajian tabel dan grafik meliputi penggunaan grafik yang tidak sesuai dengan jenis data, penyajian informasi berlebihan, serta ketidaksesuaian antara teks dan visualisasi. Weissgerber et al. (2015) menegaskan bahwa grafik batang sering disalahgunakan untuk data kontinu, padahal grafik titik atau boxplot lebih informatif dalam menunjukkan distribusi data. Kesadaran terhadap kesalahan-kesalahan ini penting untuk meningkatkan kualitas komunikasi ilmiah.

Sebagai refleksi metodologis, penyajian tabel dan grafik bukan sekadar aspek estetika dalam laporan penelitian, melainkan bagian integral dari proses interpretasi dan komunikasi ilmiah. Tabel dan grafik yang disusun sesuai standar APA dan prinsip visualisasi data yang baik akan memperkuat kredibilitas penelitian, memudahkan pemahaman pembaca, dan mendukung transparansi analisis. Dengan demikian, kemampuan menyajikan data secara visual merupakan kompetensi esensial bagi peneliti kuantitatif di era akademik dan digital saat ini.

10.4 Kesalahan Pelaporan Hasil Penelitian

Kesalahan pelaporan merupakan persoalan serius dalam penelitian kuantitatif karena dapat mengaburkan makna temuan, menyesatkan pembaca, dan merusak kredibilitas ilmiah. Dalam konteks pendidikan, sosial, dan manajemen, kesalahan pelaporan tidak selalu muncul dalam bentuk manipulasi data yang disengaja, tetapi sering kali berupa kekeliruan metodologis dan interpretatif yang bersumber dari pemahaman statistik yang tidak utuh. American Psychological Association (APA, 2020) menegaskan bahwa pelaporan hasil penelitian harus menjunjung

tinggi prinsip akurasi, kejujuran, dan transparansi untuk menjaga integritas ilmu pengetahuan.

Salah satu kesalahan pelaporan yang paling umum adalah misinterpretasi nilai p (*p-value*). Banyak laporan penelitian menyamakan nilai *p* dengan ukuran kekuatan efek atau kepastian kebenaran hipotesis. Field (2018) menjelaskan bahwa *p-value* hanya menunjukkan probabilitas memperoleh hasil setidaknya sebesar yang diamati jika hipotesis nol benar, bukan probabilitas bahwa hipotesis penelitian benar. Kesalahan ini sering memicu klaim berlebihan terhadap hasil penelitian, terutama ketika nilai *p* berada sedikit di bawah batas signifikansi konvensional (misalnya $p < 0,05$).

Kesalahan lain yang sering terjadi adalah inflasi signifikansi statistik, yaitu kecenderungan melaporkan hasil signifikan tanpa mempertimbangkan konteks analisis secara menyeluruh. Neuman (2014) menyatakan bahwa praktik seperti melakukan banyak uji statistik tanpa koreksi (*multiple testing*) dapat meningkatkan peluang kesalahan tipe I. Dalam laporan penelitian, inflasi signifikansi sering terlihat ketika peneliti menyoroti hasil signifikan sambil mengabaikan hasil non-signifikan yang relevan dengan tujuan penelitian. Praktik ini tidak hanya menyesatkan pembaca, tetapi juga bertentangan dengan prinsip kejujuran ilmiah.

Kesalahan pelaporan juga muncul dalam bentuk pengabaian ukuran efek (*effect size*) dan interval kepercayaan. APA (2020) secara eksplisit merekomendasikan pelaporan *effect size* untuk memberikan konteks substantif terhadap signifikansi statistik. Namun, dalam banyak laporan penelitian kuantitatif, hasil analisis hanya dilaporkan dalam bentuk nilai *p* tanpa penjelasan tentang besaran dan makna praktis efek yang ditemukan. Gall et al. (2019) menegaskan bahwa pelaporan yang

demikian mengurangi nilai informatif penelitian dan menyulitkan pembaca dalam menilai relevansi temuan.

Kesalahan berikutnya adalah pelaporan grafik dan tabel yang menyesatkan. Tufte (2001) mengingatkan bahwa visualisasi data dapat digunakan secara tidak tepat untuk memperbesar atau memperkecil perbedaan data melalui manipulasi skala, pemotongan sumbu, atau penggunaan warna yang berlebihan. Grafik manipulatif dapat menciptakan persepsi yang tidak sesuai dengan data sebenarnya, sehingga melanggar etika pelaporan ilmiah. Oleh karena itu, peneliti harus memastikan bahwa setiap tabel dan grafik disajikan secara proporsional dan sesuai standar.

Dalam konteks interpretasi, overgeneralization merupakan kesalahan pelaporan yang sering terjadi. Fraenkel, Wallen, dan Hyun (2021) menekankan bahwa kesimpulan penelitian harus dibatasi oleh desain dan karakteristik sampel yang digunakan. Menarik kesimpulan yang melampaui ruang lingkup data, misalnya menggeneralisasi hasil studi lokal ke populasi nasional tanpa dasar metodologis yang kuat, merupakan bentuk kesalahan pelaporan yang serius. Kesalahan ini sering terjadi ketika peneliti tidak secara eksplisit menyatakan keterbatasan penelitian.

Kesalahan pelaporan juga berkaitan dengan pelanggaran etika penelitian, seperti plagiarisme, fabrikasi data, atau pelaporan selektif. APA (2020) menegaskan bahwa plagiarisme, baik disengaja maupun tidak, merusak integritas ilmiah dan kepercayaan publik terhadap penelitian. Selain itu, praktik *selective reporting* atau *p-hacking*—memilih dan melaporkan hanya hasil yang mendukung hipotesis—merupakan pelanggaran etika yang dapat mengarah pada bias sistematis dalam literatur ilmiah.

Dalam era *reproducible research*, kesalahan pelaporan juga mencakup kurangnya transparansi analitik. Field (2018)

menekankan bahwa peneliti seharusnya melaporkan prosedur analisis secara rinci, termasuk uji asumsi, keputusan eksklusi data, dan parameter analisis yang digunakan. Ketidakjelasan dalam pelaporan prosedur analitik menyulitkan replikasi penelitian dan menurunkan kredibilitas temuan. Transparansi bukan hanya tuntutan metodologis, tetapi juga kewajiban etis dalam praktik ilmiah modern.

Upaya meminimalkan kesalahan pelaporan memerlukan kombinasi antara pemahaman metodologis, kepatuhan terhadap standar penulisan, dan kesadaran etis. Neuman (2014) menegaskan bahwa pelatihan metodologi dan statistik yang memadai merupakan langkah preventif utama untuk mengurangi kesalahan pelaporan. Selain itu, proses *peer review* dan penggunaan panduan pelaporan seperti APA 7 membantu memastikan kualitas dan konsistensi laporan penelitian.

Sebagai refleksi metodologis, kesalahan pelaporan bukanlah sekadar persoalan teknis, melainkan isu fundamental yang menyangkut integritas dan kredibilitas ilmu pengetahuan. Peneliti kuantitatif dituntut untuk melaporkan hasil penelitian secara jujur, proporsional, dan transparan, dengan menghindari klaim berlebihan dan praktik yang menyesatkan. Dengan kesadaran kritis terhadap potensi kesalahan pelaporan, penelitian kuantitatif dapat memberikan kontribusi ilmiah yang lebih bermakna, dapat dipercaya, dan berkelanjutan.

10.5 Standar APA (American Psychological Association)

Standar APA (American Psychological Association) merupakan pedoman internasional yang paling luas digunakan dalam penulisan dan pelaporan penelitian ilmiah, khususnya di bidang pendidikan, psikologi, ilmu sosial, dan manajemen. Edisi ketujuh dari *Publication Manual of the American Psychological*

Association menekankan konsistensi format, kejelasan komunikasi, dan etika akademik sebagai fondasi utama praktik ilmiah yang kredibel. Menurut APA (2020), kepatuhan terhadap standar penulisan bukan sekadar persoalan teknis, melainkan bagian integral dari integritas dan transparansi penelitian.

Dalam konteks pelaporan penelitian kuantitatif, standar APA mengatur secara rinci struktur naskah, mulai dari judul hingga daftar rujukan. Judul penelitian harus ringkas, informatif, dan mencerminkan variabel serta konteks penelitian tanpa klaim berlebihan. APA (2020) menegaskan bahwa kejelasan judul membantu pembaca memahami fokus penelitian dan mengurangi risiko misinterpretasi sejak awal. Selain itu, penggunaan *running head*, nomor halaman, dan format margin menjadi bagian dari konsistensi visual naskah.

Standar APA juga memberikan panduan khusus mengenai penulisan abstrak. Abstrak harus bersifat informatif, mencakup tujuan, metode, hasil utama, dan implikasi penelitian. Dalam penelitian kuantitatif, abstrak sebaiknya menyertakan informasi statistik utama, seperti ukuran sampel dan temuan kunci, tanpa detail teknis yang berlebihan. Creswell (2014) menekankan bahwa abstrak berfungsi sebagai “etalase” penelitian, sehingga harus disusun secara akurat dan representatif.

Dalam sitasi dalam teks, APA 7 mengatur penggunaan sitasi naratif dan parentetik secara konsisten. Setiap ide, data, atau kutipan yang berasal dari sumber lain harus dirujuk secara jelas untuk menghindari plagiarisme. Neuman (2014) menegaskan bahwa sitasi bukan hanya kewajiban akademik, tetapi juga mekanisme untuk menunjukkan posisi penelitian dalam peta keilmuan yang lebih luas. Standar APA membantu peneliti menjaga keseimbangan antara penggunaan sumber dan kontribusi orisinal.

Bagian daftar rujukan dalam standar APA dirancang untuk memastikan keterlacakan sumber secara akurat. Setiap rujukan harus ditulis dengan format yang konsisten, mencakup nama penulis, tahun publikasi, judul, dan informasi penerbitan. APA (2020) menekankan bahwa ketepatan format referensi bukan sekadar formalitas, melainkan prasyarat bagi reproducibility dan verifikasi ilmiah. Kesalahan dalam daftar rujukan dapat menghambat pembaca dalam menelusuri sumber dan menurunkan kredibilitas laporan penelitian.

Standar APA juga mengatur pelaporan statistik secara rinci. Field (2018) menjelaskan bahwa APA 7 mendorong pelaporan statistik yang lengkap dan informatif, termasuk nilai *p*, *effect size*, dan interval kepercayaan. Pelaporan statistik harus disajikan secara konsisten dan disertai penjelasan substantif yang relevan. Pendekatan ini sejalan dengan upaya meningkatkan transparansi dan akurasi interpretasi hasil penelitian kuantitatif.

Aspek penting lainnya dalam standar APA adalah bahasa dan gaya penulisan. APA (2020) menekankan penggunaan bahasa yang jelas, objektif, dan bebas dari bias. Dalam penelitian sosial dan pendidikan, peneliti harus menghindari istilah yang diskriminatif atau ambigu, serta menggunakan istilah teknis secara konsisten. Gaya penulisan yang baik tidak hanya memudahkan pemahaman pembaca, tetapi juga mencerminkan sikap ilmiah yang profesional.

Standar APA juga memberikan perhatian khusus pada etika akademik, termasuk isu plagiarisme, duplikasi publikasi, dan konflik kepentingan. Fraenkel, Wallen, dan Hyun (2021) menegaskan bahwa kepatuhan terhadap standar etika merupakan prasyarat utama kepercayaan publik terhadap penelitian ilmiah. Dalam konteks *reproducible research*, APA mendorong peneliti untuk melaporkan prosedur dan data secara

transparan agar penelitian dapat direplikasi dan diverifikasi oleh peneliti lain.

Dalam praktik penelitian modern, standar APA juga mendukung prinsip reproducibility dan open science. Pelaporan metode dan analisis secara rinci memungkinkan peneliti lain mengulangi atau mengembangkan penelitian yang sama. Sekaran dan Bougie (2020) menekankan bahwa standar pelaporan yang jelas dan konsisten memperkuat akumulasi pengetahuan ilmiah dan mencegah kesalahpahaman metodologis.

Sebagai refleksi metodologis penutup BAB 10, standar APA tidak boleh dipandang sebagai aturan kaku yang membatasi kreativitas ilmiah, melainkan sebagai kerangka normatif yang memastikan kualitas, kejelasan, dan integritas pelaporan penelitian. Dengan mematuhi standar APA 7, peneliti kuantitatif dapat meningkatkan kredibilitas temuan, memfasilitasi komunikasi ilmiah lintas disiplin, dan berkontribusi pada pengembangan ilmu pengetahuan yang bertanggung jawab dan berkelanjutan.

Daftar Pustaka

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2021). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Gall, M. D., Gall, J. P., & Borg, W. R. (2019). *Educational research: An introduction* (12th ed.). Pearson Education.

Levitt, H. M. (2019). Reporting quantitative results in psychology: Guidance based on APA style. *Training and Education in Professional Psychology, 13*(2), 89–97. https://doi.org/10.1037/tep0000224

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson Education.

Sekaran, U., & Bougie, R. (2020). *Research methods for business: A skill-building approach* (8th ed.). Wiley.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar graphs: Improving data visualization. *PLoS Biology, 13*(6), e1002128. https://doi.org/10.1371/journal.pbio.1002128