# Meninjau Ulang Praktik Analisis Kuantitatif: Kesalahan Fundamental Peneliti Pemula dan Pendekatan Metodologis Berbasis Bukti Empiris

**Irma Rodiana[1*], Tommy Hendratno[2], Isidora Kadarwati[3], Satra[4], Supandi[5], Teguh Iswanto[6], Nandang Hidayat[7]**

[1,2,3,4,5,6,7]Universitas Pakuan, Indonesia

Corresponding Author's e-mail: *irma.rodiana2025@gmail.com*

ARMADA
JURNAL PENELITIAN MULTIDISIPLIN

e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

https://ejournal.45mataram.ac.id/index.php/armada

Vol. 04, No. 08 Agustus, 2026

Page: 4263-4273

DOI:
https://doi.org/10.55681/armada.v4i8.3518

Article History:

Received: Juni 20, 2026

Revised: Juli 13, 2026

Accepted: Agustus 19, 2026

***Abstract*** *: Quantitative research methodology is a crucial instrument in the development of science due to its emphasis on objectivity and numerical data testing. Unfortunately, a lack of conceptual understanding often traps novice researchers into fatal methodological misconceptions. This article is a Systematic Literature Review aimed at deconstructing 10 major misunderstandings in quantitative data analysis. The limitation to these 10 essential errors is intentionally applied so that the evaluation remains deep, focused, and prevents cognitive overload for novice researchers. Unlike normative views, this article presents empirical evidence of misconceptions occurring in the field ranging from findings that 90% of researchers misinterpret the p-value illusion, 100% of a 15-thesis sample violated validity/reliability assumptions, low empirical scores of students' critical reasoning (77.50), to actual mechanical errors in reading statistical distribution tables. Through this review, each technical error is explained conceptually along with its mitigation. The review concludes that research validity is measured not by the complexity of statistical tools, but by the accuracy of the methodological design, adherence to parametric assumptions, and empirical evidence-based analytical justification.*

**Keywords** : *Data Analysis, Empirical Data, Statistical Literacy, Research Misconceptions, Quantitative Research.*

**Abstrak** : Metodologi penelitian kuantitatif merupakan instrumen krusial dalam pengembangan ilmu pengetahuan karena mengedepankan objektivitas dan pengujian data secara numerik. Sayangnya, rendahnya pemahaman konseptual sering menjebak peneliti pemula ke dalam berbagai miskonsepsi metodologis. Artikel ini merupakan tinjauan literatur sistematis yang bertujuan mendekonstruksi 10 kesalahpahaman utama dalam analisis data kuantitatif. Pembatasan pada 10 kesalahan esensial ini sengaja dilakukan agar evaluasi dapat dibahas secara mendalam, fokus, dan mencegah terjadinya beban kognitif (cognitive overload) bagi peneliti pemula. Berbeda dengan pandangan normatif, artikel ini menyajikan bukti empiris atas miskonsepsi yang terjadi di lapangan mulai dari temuan bahwa 90% peneliti salah mengartikan ilusi p-value, 100% dari sampel 15 skripsi terbukti melanggar uji asumsi validitas/reliabilitas, rendahnya skor empiris penalaran kritis mahasiswa (77,50), hingga kesalahan mekanis nyata dalam pembacaan tabel distribusi statistik. Melalui kajian ini, setiap kesalahan teknis dijelaskan makna konseptualnya

dan diberikan mitigasinya. Hasil tinjauan menyimpulkan bahwa keabsahan penelitian tidak diukur dari kerumitan alat statistik, melainkan dari ketepatan rancangan metodologis, kepatuhan pada asumsi parametrik, dan justifikasi analisis berbasis bukti empiris.

**Kata Kunci:** Analisis Data, Data Empiris, Literasi Statistik, Miskonsepsi Penelitian, Riset Kuantitatif.

## PENDAHULUAN

Penelitian kuantitatif merupakan salah satu pendekatan fundamental dalam pengembangan ilmu sosial dan pendidikan karena memungkinkan pengujian hubungan antarvariabel, pengukuran fenomena secara sistematis, serta penyusunan kesimpulan berdasarkan bukti empiris (Creswell & Creswell, 2018; Mustofa et al., 2024). Pendekatan ini bertumpu pada data numerik dan prosedur analisis yang terstruktur untuk menghasilkan temuan yang objektif dan dapat dipertanggungjawabkan (Damanik et al., 2025; Sugiyono, 2022). Dalam konteks tersebut, literasi statistik menjadi kompetensi penting bagi mahasiswa dan peneliti pemula karena kualitas penelitian kuantitatif tidak hanya ditentukan oleh kemampuan menggunakan perangkat lunak statistik, tetapi juga oleh pemahaman terhadap asumsi, logika inferensial, pemilihan teknik analisis, serta interpretasi hasil secara tepat.

Namun, pembelajaran statistik di pendidikan tinggi masih sering berorientasi pada aspek teknis-operasional, seperti penggunaan perangkat lunak SPSS, tanpa diimbangi pemahaman konseptual dan filosofis yang memadai (Alam et al., 2024; Bezzina & Saunders, 2014; Hardjito, 2025). Kondisi tersebut dapat menyebabkan mahasiswa mampu menjalankan prosedur analisis secara mekanis, tetapi belum tentu memahami alasan pemilihan suatu uji, asumsi yang harus dipenuhi, maupun batas interpretasi hasilnya. Kesenjangan ini berpotensi memunculkan berbagai miskonsepsi statistik yang berdampak pada ketepatan desain penelitian, analisis data, dan validitas kesimpulan yang dihasilkan (Irianto, 2024; Maulana, 2023). Kesalahan dalam penelitian kuantitatif bahkan dapat mencakup persoalan yang luas, mulai dari bias instrumen, penanganan *missing data*, pelaporan hasil, hingga spesifikasi dan pembobotan dalam analisis regresi (Onwuegbuzie & Daniel, 2003).

Salah satu alternatif untuk mengurangi persoalan tersebut adalah menyediakan pembahasan statistik yang tidak berhenti pada petunjuk prosedural, tetapi mengintegrasikan aspek konseptual, metodologis, dan aplikatif. Peneliti pemula membutuhkan kerangka yang mampu menjelaskan bukan hanya bagaimana suatu analisis dilakukan, tetapi juga mengapa prosedur tertentu digunakan, kapan prosedur tersebut tepat diterapkan, serta bagaimana kesalahan interpretasi dapat memengaruhi validitas penelitian. Atas dasar itu, artikel ini menerapkan *purposive scoping* dengan memfokuskan pembahasan pada sepuluh miskonsepsi fundamental yang paling relevan bagi penelitian mahasiswa. Pembatasan tersebut dilakukan untuk menjaga kedalaman analisis dan mengurangi potensi *cognitive overload*, sekaligus memusatkan perhatian pada kesalahan yang secara empiris berkaitan dengan rendahnya kualitas karya ilmiah mahasiswa (Mauliddin, 2017; Sudarto, 2024).

Kajian terdahulu telah membahas berbagai kesalahan statistik, kelemahan literasi metodologis, dan persoalan dalam pembelajaran analisis kuantitatif (Onwuegbuzie & Daniel, 2003; Bezzina & Saunders, 2014; Alam et al., 2024). Namun, kajian tersebut umumnya membahas kesalahan statistik secara luas atau berfokus pada aspek teknis tertentu, sehingga masih terdapat ruang untuk menyusun sintesis yang secara khusus memetakan miskonsepsi fundamental yang paling relevan bagi peneliti pemula dan menghubungkannya dengan implikasi terhadap validitas penelitian. Dengan demikian, *research gap* artikel ini terletak pada masih terbatasnya kajian yang mengintegrasikan bukti empiris mengenai miskonsepsi statistik, penjelasan konseptual mengenai sumber kesalahannya, serta konsekuensi metodologisnya dalam satu kerangka pembahasan yang sistematis dan mudah digunakan oleh mahasiswa.

Kebaruan artikel ini terletak pada penyusunan sepuluh miskonsepsi utama dalam penelitian kuantitatif berdasarkan sintesis bukti empiris dari penelitian terdahulu, disertai penjelasan mengenai dasar konseptual, dampak metodologis, dan arah koreksinya bagi peneliti pemula.

Artikel ini bertujuan mengidentifikasi dan menganalisis sepuluh miskonsepsi fundamental dalam penelitian kuantitatif serta menjelaskan implikasinya terhadap kualitas dan validitas penelitian. Kajian ini penting karena miskonsepsi pada tahap awal analisis dapat memengaruhi seluruh rantai proses penelitian, mulai dari pemilihan teknik statistik hingga penarikan kesimpulan. Oleh karena itu, pemahaman yang lebih tepat terhadap prinsip-prinsip dasar analisis kuantitatif diperlukan untuk memperkuat literasi statistik mahasiswa dan meningkatkan kualitas penelitian akademik di lingkungan pendidikan tinggi.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Systematic Literature Review* (SLR) untuk mengidentifikasi, menelaah, dan mensintesis temuan empiris mengenai miskonsepsi serta kesalahan yang umum terjadi dalam penelitian kuantitatif (Booth et al., 2016; Kitchenham & Brereton, 2021). Literatur yang dianalisis difokuskan pada studi yang membahas kesalahan penggunaan statistik, interpretasi hasil analisis, penentuan sampel, pengujian asumsi, serta kemampuan penalaran statistik dalam penelitian akademik. Sumber-sumber empiris, termasuk telaah terhadap skripsi dan penelitian terdahulu seperti Mauliddin (2017), Pramita et al. (2018), dan Sudarto (2024), digunakan sebagai dasar untuk mengidentifikasi pola kesalahan yang muncul secara berulang. Analisis dilakukan melalui proses ekstraksi, perbandingan, dan pengelompokan temuan dari literatur yang ditelaah. Setiap bentuk kesalahan dibandingkan berdasarkan kesamaan karakteristik, konsekuensi metodologis, dan frekuensi kemunculannya dalam penelitian terdahulu, kemudian disintesis ke dalam sepuluh tipologi miskonsepsi utama. Proses sintesis tidak hanya mengidentifikasi bentuk kesalahan, tetapi juga menelaah dasar konseptual yang menyebabkan miskonsepsi tersebut serta dampaknya terhadap validitas dan ketepatan inferensi penelitian kuantitatif. Pada tahap akhir, setiap miskonsepsi dianalisis dengan mengintegrasikan bukti empiris dan literatur metodologis yang relevan. Hasil sintesis kemudian disajikan dalam tiga komponen, yaitu bentuk miskonsepsi, konsekuensi metodologis, dan rekomendasi mitigasi. Pendekatan ini memungkinkan kajian menghasilkan kerangka yang tidak hanya bersifat deskriptif, tetapi juga aplikatif dalam membantu peneliti pemula mengenali, memahami, dan menghindari kesalahan fundamental dalam penelitian kuantitatif.

## HASIL DAN PEMBAHASAN

### Pemetaan Sepuluh Miskonsepsi Fundamental dalam Penelitian Kuantitatif

Hasil sintesis menunjukkan bahwa persoalan statistik pada peneliti pemula tidak hanya berkaitan dengan kemampuan melakukan perhitungan, tetapi terutama dengan ketepatan mengambil keputusan metodologis dan menginterpretasikan keluaran statistik. Sepuluh miskonsepsi yang teridentifikasi dapat dikelompokkan ke dalam empat domain utama, yaitu: (1) inferensi dan interpretasi statistik; (2) sampling dan representativitas; (3) asumsi, spesifikasi, dan kompleksitas model; serta (4) literasi operasional dan penalaran statistik. Pola tersebut menunjukkan bahwa kesalahan kuantitatif cenderung saling berhubungan. Kesalahan menentukan sampel, misalnya, dapat memengaruhi *statistical power*, sedangkan ketergantungan terhadap *p-value* dapat menyebabkan ukuran efek dan ketidakpastian estimasi terabaikan. Kritik terhadap praktik statistik mekanis semacam ini juga telah lama dikemukakan dalam literatur metodologi (Gigerenzer, 2004; Greenland et al., 2016; Stark & Saltelli, 2018).

**Tabel 1.** Sintesis Sepuluh Miskonsepsi Fundamental dalam Penelitian Kuantitatif

| No. | Miskonsepsi | Inti Permasalahan | Konsekuensi Metodologis | Praktik yang Direkomendasikan |
|---|---|---|---|---|
| 1 | *P-value fallacy* | *P-value* dianggap sebagai probabilitas hipotesis benar/salah atau ukuran besarnya efek | Kesimpulan berlebihan hanya berdasarkan $p < 0{,}05$ | Laporkan *p-value*, ukuran efek, CI, dan interpretasi substantif |
| 2 | Sampel besar selalu lebih baik | Kuantitas sampel dianggap lebih penting | Bias seleksi tetap atau bahkan semakin | Prioritaskan representativitas dan |

| | | | | |
|---|---|---|---|---|
| | | daripada kualitas desain sampling | meyakinkan secara semu | prosedur sampling yang sesuai |
| 3 | Penggunaan Slovin secara universal | Rumus digunakan tanpa memperhatikan tujuan estimasi dan asumsi | Ukuran sampel tidak sesuai dengan desain dan karakter populasi | Tentukan ukuran sampel berdasarkan estimand, variabilitas, desain, dan *power* |
| 4 | Tidak melaporkan *effect size* | Signifikansi statistik dianggap cukup membuktikan pentingnya hasil | Besarnya dampak tidak dapat dinilai secara substantif | Laporkan Cohen's *d*, *r*, OR, $\eta^2$, atau ukuran lain yang relevan beserta CI |
| 5 | Mengabaikan asumsi analisis | Uji parametrik diterapkan secara mekanis tanpa diagnostik model | Standard error, CI, atau pengujian hipotesis berpotensi menyesatkan | Lakukan diagnostik dan gunakan pendekatan robust bila diperlukan |
| 6 | Korelasi dianggap kausalitas | Hubungan statistik dianggap sebagai bukti sebab-akibat | Klaim kausal melampaui kemampuan desain penelitian | Sesuaikan bahasa kesimpulan dengan desain dan strategi identifikasi kausal |
| 7 | Model kompleks pada sampel terbatas | Banyak parameter dipaksakan pada informasi sampel yang tidak memadai | Estimasi tidak stabil dan berisiko *overfitting* | Justifikasi ukuran sampel, sederhanakan model, dan lakukan validasi |
| 8 | Salah membaca distribusi *t*, *F*, dan *r* | Nilai kritis dipilih tanpa memperhatikan *degree of freedom* dan jenis uji | Keputusan hipotesis dapat keliru | Tentukan df dan prosedur pengujian sebelum membaca nilai kritis |
| 9 | Tidak melakukan *power analysis* | Ukuran sampel ditentukan tanpa mempertimbangkan efek yang ingin dideteksi | Risiko *false negative* serta Type S dan Type M meningkat | Lakukan *a priori power analysis* atau simulasi sesuai desain |
| 10 | Menjadi operator perangkat lunak | Output perangkat lunak diterima tanpa evaluasi konseptual | Kesalahan kode, model, dan interpretasi tidak terdeteksi | Integrasikan keterampilan komputasi dengan penalaran statistik kritis |

Tabel 1 menunjukkan bahwa miskonsepsi statistik tidak dapat diperbaiki hanya melalui penambahan kemampuan menggunakan perangkat lunak. Sebaliknya, perbaikannya membutuhkan pemahaman terhadap hubungan antara pertanyaan penelitian, desain, kualitas data, asumsi model, estimasi, dan kekuatan inferensi. Dengan kata lain, kompetensi penelitian kuantitatif lebih tepat dipahami sebagai kompetensi pengambilan keputusan statistik daripada sekadar kompetensi komputasional.

### Miskonsepsi Inferensi Statistik: *P-Value*, Ukuran Efek, dan *Statistical Power*

Miskonsepsi pertama berkaitan dengan interpretasi p-value. Temuan kajian menunjukkan bahwa p-value sering diperlakukan sebagai ukuran probabilitas bahwa hipotesis nol benar, probabilitas hasil terjadi karena kebetulan, atau ukuran kekuatan efek. Ketiga interpretasi tersebut tidak tepat. Secara formal, p-value menunjukkan probabilitas memperoleh hasil yang sama ekstrem atau lebih ekstrem daripada data yang diamati apabila hipotesis nol dan asumsi model statistik yang digunakan diperlakukan sebagai benar. Karena itu, p = 0,001 tidak berarti hanya terdapat peluang 0,1% bahwa hasil disebabkan oleh kebetulan, dan juga tidak menunjukkan bahwa efek yang

ditemukan besar. Wasserstein dan Lazar (2016) serta Greenland et al. (2016) menegaskan bahwa p-value tidak mengukur besarnya efek maupun probabilitas suatu hipotesis benar.

Miskonsepsi tersebut berkaitan langsung dengan minimnya pelaporan effect size. Signifikansi statistik menjawab persoalan yang berbeda dari signifikansi substantif. Sampel yang sangat besar dapat menghasilkan p-value kecil untuk perbedaan yang secara praktis hampir tidak berarti. Sebaliknya, sebuah efek yang substantif dapat gagal memenuhi ambang $p < 0,05$ apabila penelitian memiliki presisi rendah. Sullivan dan Feinn (2012) menegaskan bahwa effect size diperlukan untuk memahami magnitude suatu perbedaan, sedangkan Lakens (2013) menunjukkan bahwa ukuran efek penting untuk interpretasi, perencanaan sampel, dan sintesis meta-analitik.

Dengan demikian, penggunaan Cohen's d, misalnya, sebaiknya tidak berhenti pada pemberian label kecil, sedang, atau besar secara mekanis. Nilai $d = 0,80$ atau 0,85 memang sering dikaitkan dengan kategori "besar" berdasarkan konvensi klasik, tetapi relevansi substantifnya tetap harus ditafsirkan berdasarkan konteks bidang, konsekuensi praktis, serta ukuran efek yang lazim ditemukan dalam literatur terkait. Lakens (2013) secara khusus mengingatkan bahwa effect size lebih informatif ketika dibandingkan dengan efek empiris yang relevan daripada sekadar menggunakan batas kategoris secara kaku.

Temuan berikutnya adalah absennya statistical power dalam perencanaan penelitian. Studi dengan power rendah tidak hanya meningkatkan risiko gagal menemukan efek yang benar-benar ada, tetapi hasil signifikan yang berhasil diperoleh juga berpotensi melebih-lebihkan besarnya efek. Button et al. (2013) menghubungkan penelitian berdaya rendah dengan estimasi efek yang berlebihan dan reproduksibilitas yang rendah, sedangkan Gelman dan Carlin (2014) menunjukkan bahwa kondisi tersebut dapat menghasilkan Type S (sign) error dan Type M (magnitude) error. Implikasinya, jumlah sampel sebaiknya ditentukan secara a priori dengan mempertimbangkan ukuran efek yang secara substantif penting, tingkat signifikansi, desain penelitian, dan power yang ditargetkan. G*Power merupakan salah satu perangkat yang dapat digunakan untuk tujuan tersebut dan menyediakan analisis daya untuk berbagai keluarga uji statistik (Faul et al., 2007, 2009). Target power 0,80 dapat digunakan sebagai konvensi awal, tetapi bukan angka universal yang menggantikan pertimbangan desain dan konsekuensi kesalahan penelitian.

### Miskonsepsi Ukuran Sampel dan Representativitas

Temuan kedua menunjukkan adanya pandangan bahwa semakin besar jumlah responden, semakin tinggi validitas penelitian. Asumsi tersebut mengabaikan sumber kesalahan yang lebih mendasar, yaitu proses pembentukan sampel. Ukuran sampel yang besar menurunkan sampling variance dalam kondisi tertentu, tetapi tidak otomatis menghilangkan selection bias, coverage error, atau non-response bias. Bahkan, sampel berukuran sangat besar dapat menghasilkan interval estimasi yang sangat sempit di sekitar nilai yang tetap bias. Fenomena tersebut ditunjukkan secara empiris oleh Bradley et al. (2021). Dalam estimasi vaksinasi COVID-19 di Amerika Serikat, survei dengan sekitar 250.000 responden per minggu menghasilkan estimasi yang jauh lebih bias dibandingkan panel sekitar 1.000 responden yang menerapkan praktik sampling lebih baik. Dalam analisis tersebut, informasi efektif dari survei yang sangat besar bahkan dapat merosot hingga setara dengan simple random sample yang amat kecil akibat bias seleksi. Temuan ini merupakan ilustrasi kuat dari Big Data Paradox: volume data tidak dapat menggantikan kualitas proses pembentukan data.

Dengan demikian, contoh "150 responden probabilistik lebih baik daripada 5.000 responden convenience" sebaiknya tidak diperlakukan sebagai aturan numerik mutlak. Substansi yang lebih tepat adalah bahwa sampel yang lebih kecil tetapi memiliki mekanisme pemilihan yang lebih sesuai dapat menghasilkan inferensi populasi yang lebih dapat dipertanggungjawabkan dibandingkan sampel besar dengan bias seleksi yang berat. Karena itu, justifikasi sampel harus membahas bukan hanya berapa banyak unit analisis yang dikumpulkan, tetapi juga bagaimana unit tersebut dipilih. Temuan terkait muncul pada penggunaan formula Slovin. Kesederhanaan formula tersebut membuatnya sering digunakan sebagai prosedur universal dalam penelitian mahasiswa. Padahal, Tejada dan Punzalan (2012) menunjukkan bahwa penggunaan formula Slovin memiliki ruang

aplikasi yang jauh lebih terbatas daripada yang umumnya diasumsikan. Formula tersebut berkaitan dengan estimasi proporsi populasi pada kondisi tertentu, termasuk tingkat kepercayaan 95%, dan bekerja optimal ketika proporsi populasi yang belum diketahui berada di sekitar 0,5. Oleh karena itu, formula Slovin tidak seharusnya digunakan secara otomatis untuk setiap survei, terlebih ketika tujuan penelitian, parameter yang diestimasi, struktur populasi, dan desain sampling berbeda.

Pada populasi heterogen, penggunaan stratified sampling dapat relevan ketika lapisan populasi memiliki karakteristik substantif yang perlu direpresentasikan. Namun, stratifikasi bukan "tambahan wajib" untuk memperbaiki Slovin; pemilihan desain sampling seharusnya berasal dari struktur populasi dan tujuan inferensi. Dengan demikian, perencanaan ukuran sampel idealnya bergeser dari logika "memilih satu rumus" menuju logika menetapkan estimand, menentukan desain sampling, menetapkan presisi atau efek minimum yang relevan, kemudian menghitung kebutuhan sampel berdasarkan tujuan tersebut.

**Miskonsepsi Asumsi, Kausalitas, dan Kompleksitas Model**

Miskonsepsi kelima berkaitan dengan uji asumsi statistik. Temuan kajian terdahulu menunjukkan bahwa peneliti pemula sering menjalankan analisis berdasarkan prosedur perangkat lunak tanpa memahami konsekuensi ketika asumsi model tidak terpenuhi (Mauliddin, 2017; Sudarto, 2024). Akan tetapi, penyelesaian terhadap pelanggaran asumsi juga perlu dilakukan secara proporsional. Heteroskedastisitas, misalnya, tidak secara otomatis mengharuskan peneliti meninggalkan regresi linier dan menggunakan analisis nonparametrik. Hayes dan Cai (2007) menunjukkan bahwa pada regresi OLS, heteroskedastisitas terutama dapat menyebabkan estimasi matriks kovarians dan standard error konvensional menjadi tidak tepat, sehingga pengujian signifikansi dan interval kepercayaan dapat menjadi terlalu liberal atau terlalu konservatif. Salah satu pendekatan yang dapat digunakan adalah heteroskedasticity-consistent standard errors. Transformasi data, perubahan spesifikasi model, weighted least squares, atau metode robust lainnya dapat dipertimbangkan berdasarkan penyebab dan tujuan analisis. Dengan demikian, inti kompetensinya bukan "lulus atau gagal uji asumsi", tetapi memahami asumsi mana yang relevan terhadap estimasi tertentu dan bagaimana pelanggarannya memengaruhi inferensi.

Miskonsepsi keenam adalah penyamaan korelasi dengan kausalitas. Koefisien korelasi mengukur pola keterkaitan antarvariabel, tetapi tidak dengan sendirinya menentukan arah kausal maupun mengeliminasi kemungkinan confounding, reverse causation, dan berbagai sumber bias lainnya. Altman dan Krzywinski (2015) menekankan pentingnya membedakan asosiasi, korelasi, dan kausalitas. Karena itu, hasil korelasi Pearson yang signifikan tidak cukup untuk menyimpulkan bahwa perubahan variabel X "menyebabkan" perubahan variabel Y. Implikasinya terhadap penulisan hasil sangat konkret. Penelitian korelasional lebih tepat menggunakan istilah "berhubungan", "berasosiasi", atau "berkorelasi", sedangkan klaim "memengaruhi" atau "menyebabkan" membutuhkan desain dan asumsi identifikasi kausal yang memadai. Contoh hubungan antara konsumsi es krim dan kejadian tenggelam menggambarkan bahwa dua variabel dapat bergerak bersama karena dipengaruhi faktor ketiga, seperti temperatur atau musim, tanpa memiliki hubungan kausal langsung satu sama lain.

Miskonsepsi ketujuh muncul pada penggunaan model multivariat atau SEM yang terlalu kompleks dibandingkan informasi yang tersedia dalam sampel. Wolf et al. (2013), melalui simulasi Monte Carlo, menunjukkan bahwa kebutuhan sampel SEM sangat bergantung pada kompleksitas model, jumlah indikator, besar factor loading, koefisien jalur, dan pola missing data. Kebutuhan sampel dalam simulasi mereka bervariasi secara luas, sehingga penggunaan satu aturan sederhana untuk seluruh SEM tidak dapat dibenarkan. Pada PLS-SEM, Kock dan Hadaya (2018) juga menunjukkan bahwa aturan "10 kali" menghasilkan estimasi kebutuhan sampel yang kurang presisi dan menawarkan pendekatan berbasis inverse square root serta gamma-exponential. Hal penting lainnya adalah bahwa bootstrapping tidak dengan sendirinya "menyembuhkan" model yang terlalu kompleks atau sampel yang secara informasional tidak memadai. Bootstrapping berguna untuk memperkirakan distribusi sampling dan ketidakpastian parameter, tetapi desain yang tidak memadai tetap perlu diatasi melalui justifikasi kebutuhan sampel, penyederhanaan model bila diperlukan, evaluasi stabilitas parameter, serta validasi prediktif atau out-of-sample. Dengan demikian, problem overfitting harus diperlakukan sebagai persoalan keseimbangan antara

fleksibilitas model dan informasi yang tersedia, bukan sebagai masalah yang dapat diselesaikan dengan satu prosedur perangkat lunak.

### Miskonsepsi Operasional dan Krisis Penalaran Statistik

Temuan kedelapan menunjukkan kesalahan dalam membaca tabel distribusi statistik. Mauliddin (2017) menemukan kekeliruan dalam penggunaan nilai kritis pada karya mahasiswa, sementara Pramita et al. (2018) menunjukkan adanya kesulitan mahasiswa dalam memahami tabel distribusi t dan F. Kesalahan semacam ini menunjukkan bahwa mahasiswa dapat mengetahui nama suatu uji statistik tanpa sepenuhnya memahami bagaimana distribusi sampling, tingkat signifikansi, dan degree of freedom menentukan keputusan inferensial. Sebagai contoh, pada independent-samples t-test dengan asumsi varians sama dan masing-masing kelompok memiliki 40 responden, degree of freedom untuk pooled t-test adalah (df=n_1+n_2-2=78), bukan 80. Namun, apabila asumsi kesamaan varians tidak dipertahankan dan digunakan Welch's t-test, derajat kebebasan dihitung dengan pendekatan yang berbeda dan dapat berbentuk non-integer. Contoh ini memperlihatkan mengapa penggunaan tabel statistik tanpa memahami model yang mendasarinya berpotensi menghasilkan keputusan mekanis yang salah.

Miskonsepsi kesepuluh merupakan persoalan yang lebih fundamental, yaitu kecenderungan menjadi operator perangkat lunak statistik. Wahidah dan Atmaja (2022) menunjukkan masih adanya kelemahan pada dimensi penarikan kesimpulan dalam kemampuan berpikir kritis mahasiswa. Temuan tersebut konsisten dengan kritik lebih luas terhadap praktik mindless statistics. Gigerenzer (2004) menjelaskan bagaimana ritual statistik dapat menggantikan penalaran statistik, sementara Stark dan Saltelli (2018) menyebut praktik mekanis semacam ini sebagai cargo-cult statistics, yaitu menjalankan prosedur statistik tanpa memahami logika ilmiah yang mendasarinya. Kondisi tersebut terlihat ketika hasil perangkat lunak diterima sebagai kebenaran tanpa pemeriksaan terhadap kode data, konstruksi instrumen, spesifikasi model, atau teori. Sebagai contoh, koefisien negatif ketika teori memprediksi hubungan positif tidak otomatis berarti terdapat kesalahan reverse coding. Hasil tersebut seharusnya menjadi pemicu pemeriksaan terhadap pengkodean, reliabilitas alat ukur, suppression effect, multikolinearitas, spesifikasi model, karakter sampel, dan kemungkinan bahwa data memang tidak mendukung hipotesis awal. Peneliti kritis tidak memodifikasi hasil agar sesuai dengan teori, tetapi mengevaluasi apakah analisis dan data telah memberikan dasar inferensial yang dapat dipertanggungjawabkan.

## Pembahasan

### Dari Ritual Signifikansi menuju Inferensi Berbasis Besaran dan Ketidakpastian

Secara keseluruhan, sepuluh miskonsepsi menunjukkan bahwa persoalan utama penelitian kuantitatif bukan kurangnya perangkat analisis, melainkan kecenderungan mereduksi inferensi statistik menjadi serangkaian aturan mekanis. Temuan ini sejalan dengan Gigerenzer (2004), yang mengkritik null ritual berupa ketergantungan pada hipotesis nol, batas 0,05, dan keputusan signifikan/tidak signifikan tanpa pertimbangan substantif. Greenland et al. (2016) kemudian mendokumentasikan sejumlah besar kesalahan interpretasi terkait p-value, interval kepercayaan, dan power.

Sintesis penelitian ini memperluas persoalan tersebut ke konteks peneliti pemula dengan menunjukkan bahwa p-value fallacy, absennya effect size, dan ketiadaan power analysis sebenarnya berasal dari akar epistemologis yang sama: statistik diperlakukan sebagai mekanisme menghasilkan keputusan biner, bukan sebagai alat untuk mengukur besaran efek dan ketidakpastian. ASA secara eksplisit menegaskan bahwa kesimpulan ilmiah tidak seharusnya hanya bergantung pada apakah p-value melewati ambang tertentu (Wasserstein & Lazar, 2016).

Karena itu, laporan kuantitatif yang lebih kuat perlu menggeser fokus dari pertanyaan "apakah signifikan?" menuju sekurang-kurangnya tiga pertanyaan: seberapa besar efeknya, seberapa presisi estimasinya, dan apakah efek tersebut bermakna secara substantif? Ukuran efek dan interval kepercayaan membantu menjawab dua pertanyaan pertama, sedangkan konteks teori dan bidang penelitian diperlukan untuk menjawab pertanyaan ketiga. Prinsip ini sejalan dengan Sullivan dan Feinn (2012) serta Lakens (2013), yang menempatkan ukuran efek sebagai bagian esensial pelaporan penelitian.

### Representativitas Lebih Fundamental daripada Sekadar Besarnya Sampel

Temuan mengenai sampel memperlihatkan konflik antara logika kuantitas dan kualitas informasi. Peneliti pemula sering menganggap ukuran sampel sebagai indikator mutu yang berdiri sendiri, padahal kemampuan melakukan generalisasi ditentukan oleh hubungan antara sampel dan populasi target. Bradley et al. (2021) memberikan bukti empiris yang kuat bahwa peningkatan jumlah responden tidak mampu mengatasi bias seleksi secara otomatis. Temuan tersebut memperjelas keterbatasan penggunaan rumus sampel secara ritualistik. Slovin menjadi contoh penting karena kesederhanaannya sering menghilangkan pertanyaan metodologis yang seharusnya datang lebih dahulu: parameter apa yang ingin diestimasi, seberapa heterogen populasi, bagaimana mekanisme sampling, berapa presisi yang diperlukan, dan analisis apa yang akan dilakukan? Kritik Tejada dan Punzalan (2012) menunjukkan bahwa formula tersebut memiliki asumsi dan ruang penggunaan tertentu, sehingga tidak dapat dijadikan formula universal.Dengan demikian, temuan penelitian ini mendukung perubahan orientasi pengajaran metode kuantitatif. Mahasiswa seharusnya tidak pertama-tama ditanya "rumus sampel apa yang digunakan?", tetapi "inferensi apa yang ingin dibuat dan kepada populasi mana hasil tersebut akan digeneralisasikan?". Pergeseran tersebut penting karena ukuran sampel, teknik sampling, *power*, dan desain analisis merupakan keputusan yang saling terkait.

### Diagnostik Model sebagai Proses Penalaran, Bukan Formalitas

Temuan pada asumsi klasik juga menunjukkan adanya kecenderungan menjadikan pengujian asumsi sebagai ritual administratif. Praktik seperti melakukan serangkaian uji normalitas, heteroskedastisitas, atau multikolinearitas hanya untuk memperoleh label "lolos" dapat kehilangan makna apabila peneliti tidak memahami konsekuensi pelanggaran terhadap parameter yang sedang diestimasi. Pada heteroskedastisitas, misalnya, Hayes dan Cai (2007) memperlihatkan bahwa OLS tidak otomatis menjadi tidak berguna; persoalan utamanya terletak pada validitas estimasi *standard error* konvensional dan inferensi turunannya. Hal yang sama berlaku pada model kompleks. Literatur SEM tidak mendukung satu angka minimum sampel yang berlaku universal.

Wolf et al. (2013) menunjukkan bahwa kebutuhan sampel merupakan fungsi dari karakteristik model, sedangkan Kock dan Hadaya (2018) menunjukkan keterbatasan aturan sederhana dalam PLS-SEM. Dengan demikian, pendekatan yang lebih defensibel adalah menghubungkan kebutuhan sampel dengan kompleksitas model, kekuatan efek, kualitas pengukuran, dan tujuan inferensi. Temuan tersebut juga menunjukkan pentingnya membedakan alat koreksi dengan solusi desain. *Bootstrapping*, transformasi data, atau robust standard errors memiliki fungsi metodologis tertentu, tetapi tidak menggantikan desain yang baik. Misalnya, *bootstrapping* tidak dapat mengubah sampel yang sangat tidak representatif menjadi representatif, dan robust standard errors tidak mengoreksi *omitted variable bias*. Pemahaman batas fungsi suatu metode merupakan indikator literasi statistik yang lebih matang daripada sekadar kemampuan menjalankan prosedurnya.

### Literasi Statistik sebagai Kemampuan Menghubungkan Teori, Data, dan Keputusan

Miskonsepsi korelasi-kausalitas, kesalahan membaca distribusi, dan ketergantungan pada perangkat lunak menunjukkan bahwa literasi statistik pada dasarnya merupakan kompetensi penalaran. Altman dan Krzywinski (2015) menegaskan bahwa asosiasi statistik tidak identik dengan kausalitas. Sementara itu, kritik Gigerenzer (2004) dan Stark dan Saltelli (2018) menunjukkan bahwa perangkat statistik yang semakin mudah dioperasikan justru dapat memperbesar praktik ritualistik apabila tidak diikuti pemahaman konseptual. Hasil kajian ini karena itu menempatkan critical thinking sebagai kompetensi penghubung bagi kesepuluh miskonsepsi. Peneliti yang memahami perangkat lunak tetapi tidak mampu mempertanyakan keluaran analisis masih rentan terhadap kesalahan. Sebaliknya, peneliti yang memiliki penalaran statistik akan memeriksa apakah analisis sesuai dengan pertanyaan penelitian, apakah kualitas pengukuran memadai, apakah asumsi substantif masuk akal, apakah ukuran sampel dapat mendukung kompleksitas model, dan apakah bahasa kesimpulan tidak melampaui informasi yang diberikan data.

Dengan demikian, literasi statistik ideal tidak berakhir pada kemampuan menghasilkan tabel SPSS, R, Stata, atau perangkat lainnya. Kompetensi tersebut harus mencapai kemampuan

menjelaskan mengapa metode dipilih, asumsi apa yang digunakan, bagaimana ketidakpastian ditangani, apa yang dapat disimpulkan, dan yang sama pentingnya apa yang tidak dapat disimpulkan dari hasil analisis. Temuan penelitian memiliki implikasi langsung terhadap pembelajaran metodologi penelitian dan statistik di perguruan tinggi. Pertama, pembelajaran perlu bergeser dari orientasi *software-driven* menuju *decision-driven statistical learning*. Mahasiswa tetap perlu menguasai perangkat lunak, tetapi setiap prosedur harus didahului oleh pemahaman terhadap pertanyaan penelitian, jenis variabel, desain, estimand, asumsi, dan konsekuensi inferensial. Latihan statistik dapat dirancang berbasis kasus yang mengharuskan mahasiswa mendeteksi dan memperbaiki miskonsepsi, bukan hanya mereplikasi langkah analisis.

Kedua, perguruan tinggi dapat mengembangkan standar minimum pelaporan penelitian kuantitatif yang mencakup: justifikasi desain dan ukuran sampel; penjelasan mekanisme sampling; diagnostik model yang relevan; pelaporan ukuran efek beserta interval kepercayaan; interpretasi *p-value* secara tepat; serta pembatasan klaim kausal sesuai desain. Pendekatan ini sejalan dengan rekomendasi literatur yang menekankan ukuran efek, presisi estimasi, dan kualitas inferensi dibandingkan ketergantungan eksklusif pada signifikansi statistik (Wasserstein & Lazar, 2016; Sullivan & Feinn, 2012; Lakens, 2013). Ketiga, analisis kebutuhan sampel sebaiknya tidak lagi bergantung pada satu rumus universal. Penelitian eksperimental, korelasional, regresi, SEM, survei populasi, dan desain lainnya membutuhkan pendekatan perencanaan yang berbeda. *Power analysis*, simulasi Monte Carlo, presisi interval kepercayaan, atau metode berbasis desain sampling dapat dipilih sesuai dengan tujuan penelitian. Hal ini akan membantu mahasiswa memahami bahwa keputusan statistik merupakan konsekuensi dari desain ilmiah, bukan sekadar persyaratan administratif penelitian.

Secara lebih luas, sepuluh miskonsepsi yang diidentifikasi dapat digunakan sebagai kerangka diagnostik untuk pengembangan modul literasi statistik bagi mahasiswa. Kerangka tersebut dapat memandu dosen dalam mengajarkan statistik melalui tiga tahap: memahami konsep, mendeteksi miskonsepsi, dan mengambil keputusan metodologis yang tepat. Dengan pendekatan tersebut, kemampuan statistik mahasiswa diharapkan berkembang dari keterampilan prosedural menuju kompetensi analitis dan kritis. Kajian ini secara sengaja menerapkan *purposive scoping* pada sepuluh miskonsepsi fundamental agar setiap persoalan dapat dibahas secara konseptual dan aplikatif dengan kedalaman yang memadai. Fokus tersebut memungkinkan terbentuknya kerangka yang relevan bagi peneliti pemula, tetapi tidak dimaksudkan sebagai inventarisasi seluruh bentuk kesalahan yang mungkin terjadi dalam penelitian kuantitatif. Persoalan lain, seperti *missing data*, *multiple testing*, *measurement error*, *publication bias*, *p-hacking*, pemilihan model berbasis data, dan kesalahan pada analisis longitudinal, dapat dikembangkan dalam kajian lanjutan.

Selain itu, bukti yang disintesis berasal dari literatur dengan desain, disiplin, dan konteks yang beragam. Keragaman tersebut merupakan kekuatan karena memperlihatkan bahwa problem literasi statistik bersifat lintas bidang, tetapi pada saat yang sama membuat kajian ini tidak diarahkan untuk mengestimasi prevalensi kuantitatif setiap miskonsepsi. Penelitian selanjutnya dapat menggunakan *systematic review* atau *meta-research* dengan kriteria inklusi yang lebih terstruktur untuk mengestimasi frekuensi kesalahan statistik pada skripsi, tesis, atau artikel ilmiah berdasarkan disiplin dan jenis analisis. Pengembangan berikutnya juga dapat menguji sepuluh miskonsepsi ini sebagai instrumen diagnostik literasi statistik mahasiswa. Validasi empiris dapat dilakukan melalui pengembangan *scenario-based assessment* yang tidak hanya mengukur kemampuan menghitung, tetapi juga kemampuan memilih metode, mengenali kesalahan, dan menginterpretasikan hasil. Dengan demikian, kajian ini menyediakan landasan konseptual bagi penelitian berikutnya untuk mengembangkan intervensi pembelajaran statistik yang dapat diuji secara eksperimental.

## KESIMPULAN DAN SARAN

Berbagai data dan telaah literatur empiris dalam artikel ini mengonfirmasi bahwa kesalahan dalam penelitian kuantitatif mulai dari pelanggaran asumsi 100% pada skripsi hingga skor pemikiran kritis yang rendah bukanlah sekadar asumsi, melainkan realitas lapangan yang secara masif menggugurkan validitas karya ilmiah peneliti pemula. Keabsahan penelitian sejatinya tidak

ditentukan oleh seberapa mahir seseorang menekan tombol aplikasi statistik, tetapi oleh ketepatan desain instrumen, kepatuhan mutlak pada asumsi parametrik, serta justifikasi pelaporan berbasis *effect size* yang transparan. Perguruan tinggi harus segera mereformasi pendidikan metodologi dari sekadar pendekatan operasional-mekanis menuju pembentukan karakter pemikir kritis (*critical thinker*), agar luaran riset yang dihasilkan sungguh-sungguh empiris, objektif, dan berdampak nyata.

## UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan akademik, masukan metodologis, dan kontribusi selama proses penyusunan penelitian ini. Apresiasi juga disampaikan kepada institusi, dosen, serta rekan sejawat yang turut mendukung pengembangan kajian sehingga penelitian ini dapat diselesaikan dengan baik.

## DAFTAR PUSTAKA

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation. *Nature Methods, 12*(10), 899–900. https://doi.org/10.1038/nmeth.3587

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE Publications.

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., & Flaxman, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature, 600*, 695–700. https://doi.org/10.1038/s41586-021-04198-4

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709–722. https://doi.org/10.3758/BF03192961

Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and Software Technology, 55*(12), 2049–2075. doi:10.1016/j.infsof.2013.07.010.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261. https://doi.org/10.1111/isj.12131

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Mauliddin, M. (2017). Analisis kesalahan penggunaan uji statistik pada skripsi mahasiswa. *Matematika dan Pembelajaran, 5*(2), 141–158. https://doi.org/10.33477/mp.v5i2.231

Mauliddin, M. (2017). Analisis kesalahan penggunaan uji statistik pada skripsi mahasiswa. *Matematika dan Pembelajaran, 5*(2), 141–158. doi:10.33477/mp.v5i2.231.

Pramita, D., Anwar, Y. S., Sirajudin, & Abdillah. (2019). Analisis kesalahan uji statistik pada skripsi mahasiswa Program Studi Pendidikan Matematika FKIP Universitas Muhammadiyah Mataram. *Jurnal Riset Intervensi Pendidikan, 1*(1), 39–44.

Stark, P. B., & Saltelli, A. (2018). Cargo-cult statistics and scientific crisis. *Significance, 15*(4), 40–43. https://doi.org/10.1111/j.1740-9713.2018.01174.x

Sudarto, S. (2024). Analisis kesalahan mahasiswa dalam membuat skripsi. *Jurnal Cakrawala Ilmiah, 3*(9), 2455–2462.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tejada, J. J., & Punzalan, J. R. B. (2012). On the misuse of Slovin's formula. *The Philippine Statistician, 61*(1), 129–136.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237