Aligning Research Objectives with Appropriate Statistical Tools: A Decision Framework with Applied Examples from African Research Contexts

Malcolm Shambana[1], Ephraim Kaang'andu Belemu, PhD[2],

Business Studies,Africa Research University,Lusaka,Zambia

DOI: https://doi.org/10.47772/IJRISS.2026.101000579

Received: 07 October 2026; Accepted: 12 October 2026; Published: 11 November 2026

ABSTRACT

The credibility of empirical research rests on the correct alignment between research objectives, the nature of the data collected, and the statistical procedures used to analyse those data. A recurring weakness in graduate and early-career scholarship, particularly within African universities, is the selection of statistical tools on the basis of familiarity or software convenience rather than the measurement properties of variables and the logic of the research design. Drawing on an integrative review of established and recent methodological literature, this article examines the theoretical basis for statistical tool selection and proposes a four-step decision framework that links research objectives, levels of measurement, distributional assumptions and design structure to appropriate parametric, non-parametric and advanced procedures. The framework is presented as a decision matrix and a flowchart and is positioned against published test-selection guides and recent work on model selection and reproducible research. Nine worked examples drawn from social science, health, agricultural and business research across several African contexts illustrate correct application; they use hypothetical data constructed for teaching purposes. The article identifies eight recurrent errors in practice and concludes with practical recommendations for researchers, supervisors and institutions on specifying the statistical analysis plan before, rather than after, data collection. Its contribution is a practical, objective-led reference tool for proposal development, supervision and examination.

Keywords: Statistical Tool Selection; Research Methodology; Decision Framework; Decision Flowchart; Parametric Tests; Non-Parametric Tests; Levels Of Measurement; Reproducible Research; Postgraduate Supervision; African Academic Research

INTRODUCTION

Statistical analysis is not a mechanical add-on to research; it is the analytical bridge through which raw observations are translated into defensible knowledge claims. The validity of any empirical conclusion depends jointly on the quality of the research design, the properties of the data collected, and the appropriateness of the statistical procedure applied to those data (Field, 2018). When these three elements are misaligned, the resulting inferences are compromised, however carefully the data were collected and however sophisticated the software used to process them.

Correct alignment between research objectives and statistical tools matters because every statistical test carries a set of assumptions and a specific inferential logic. A test correctly applied answers the question the researcher is actually asking; a test incorrectly applied may produce an answer that looks plausible but does not correspond to the underlying research question at all (Pallant, 2020). Misapplication commonly shows itself as an inflated Type I error rate, where a researcher concludes that a relationship or difference exists when it does not, or an elevated Type II error rate, where a genuine effect is missed because the chosen test lacked the power or the appropriate structure to detect it (Cohen, 1988; Gujarati and Porter, 2009).

The consequences of poor tool selection extend beyond individual dissertations or journal submissions. Editors and reviewers increasingly scrutinise the analytic sections of manuscripts, and reviews of the published health

research literature have long documented that a substantial proportion of errors arise not from fabricated data but from statistical misapplication: parametric tests run on ordinal data, correlation coefficients interpreted as evidence of causation, or regression models estimated without verifying the assumptions on which their coefficients depend (Altman, 1994; Strasak et al., 2007). There is little reason to believe that applied social science, agricultural and business research are immune to the same problems. For early-career researchers in African universities, where access to dedicated biostatistics or econometrics support can be limited, the risk of such errors is compounded by reliance on default software routines and by supervisory guidance that may not itself have kept pace with developments in applied statistics.

This article responds to that gap. It sets out, in accessible terms, the logic that should govern the choice of a statistical tool, presents a decision framework that links research objectives and data characteristics to appropriate techniques, and illustrates the framework with worked examples grounded in African research contexts across the social sciences, health, agriculture and business. The article is intended primarily for postgraduate students, early-career researchers and supervisors, but it is equally relevant to examiners and journal reviewers, who must judge whether a reported analysis is fit for purpose.

Statement of the Problem

A pattern frequently encountered in the supervision and examination of postgraduate research in many African universities is the selection of statistical tools on the basis of convenience, familiarity or perceived prestige rather than the measurement properties of the data and the logic of the research design. This runs contrary to the long-established principle that the analytic technique should follow from the research objective and design (Kothari, 2004). Several manifestations of this pattern are common.

First, researchers frequently apply parametric procedures such as the independent-samples t-test or one-way ANOVA to data that are ordinal in nature, most notably single-item Likert-scale responses, without first establishing that the assumptions of interval-level measurement, normality and homogeneity of variance are reasonably satisfied (Pallant, 2020). Second, correlation coefficients are routinely computed and interpreted as though they demonstrate causal relationships, even where the research objective explicitly frames a cause-and-effect question that correlation alone cannot answer (Gujarati and Porter, 2009). Third, researchers default to whichever tool is most familiar from a single statistics course or most easily executed in a particular software package, irrespective of whether that tool matches the data type or the stated hypothesis. Fourth, assumption testing (normality, homogeneity of variance, multicollinearity, independence of observations) is frequently omitted altogether, or conducted only after a significant result has already been reported, which reverses the correct sequence of analysis.

These practices are not confined to any single discipline. In social science research, ordinal attitudinal data are commonly analysed with parametric comparisons of means. In health research, small and skewed clinical samples are analysed with tests that assume large-sample normality. In agricultural research, count data such as pest incidence or disease occurrence are modelled with ordinary least squares regression rather than with Poisson or negative binomial models designed for count outcomes. In business research, panel data spanning multiple firms and years are analysed with pooled cross-sectional regression that ignores firm-level heterogeneity. The consequence in each case is the same: statistical output that appears authoritative but rests on a foundation that does not match the data or the question being asked.

Objectives and Approach

This article is guided by three specific objectives:

- To classify statistical tools according to data type (nominal, ordinal, interval/ratio) and research objective (description, comparison, association, prediction and classification).
- To demonstrate correct statistical tool selection through applied, discipline-specific examples drawn from African research contexts, including the justification for each choice and the interpretation of representative output.

- To provide a decision framework and practical recommendations that researchers and supervisors can apply at the design stage of a study, before data collection begins, so that the analytic plan is determined by the research question rather than retrofitted to the data that happen to be available.

The article adopts a conceptual, integrative-review approach. Rather than reporting new primary data, it synthesises established methodological texts in statistics, econometrics and research design into a structured decision framework, presents that framework as a decision matrix and a flowchart, and demonstrates it through worked examples. Empirical validation of the framework is beyond the scope of this article and will be reported in a separate study. The worked examples are set in realistic African research contexts, but the scenarios and all numerical outputs are hypothetical, constructed solely for illustration, and should not be cited as empirical findings. The scope is limited to the frequentist techniques most commonly used in postgraduate research; Bayesian methods, machine-learning approaches and qualitative analysis fall outside the remit of the article.

LITERATURE REVIEW

The theoretical basis for statistical tool selection rests on the classical taxonomy of levels of measurement first articulated by Stevens (1946) and subsequently embedded in mainstream statistics teaching (Field, 2018). Nominal data classify observations into unordered categories, such as sex, marital status or district of residence. Ordinal data classify observations into ordered categories in which the intervals between categories are not assumed to be equal, such as Likert-scale ratings of agreement or satisfaction. Interval and ratio data have equal and meaningful intervals between values, and ratio data also have a true zero point, as in income, yield in kilograms or reaction time. The level of measurement of the dependent variable, together with that of any predictor variables, is the first and most fundamental determinant of which statistical procedures are admissible (Field, 2018).

The treatment of Likert-type data deserves particular care, because it is the point at which this taxonomy is most often contested. A single Likert item is, strictly, ordinal. However, a substantial methodological literature argues that composite scores formed by summing or averaging several Likert items with acceptable internal consistency can, in many applications, be treated as approximately interval, and that parametric tests are reasonably robust to this practice where samples are adequate and distributions are not severely skewed (Norman, 2010). The defensible position is therefore neither to prohibit parametric analysis of Likert data outright nor to apply it uncritically, but to state explicitly whether the variable is a single item or a multi-item scale, to justify the measurement assumption adopted, and, where doubt remains, to confirm that a non-parametric analysis leads to the same substantive conclusion.

Pallant (2020) emphasises that parametric procedures (t-tests, analysis of variance, Pearson correlation and ordinary least squares regression) share a common set of distributional assumptions: the dependent variable should be measured at the interval or ratio level, should be approximately normally distributed within groups or overall, and, where groups are compared, should have approximately equal variances across those groups (homogeneity of variance, commonly assessed with Levene's test). Observations are further assumed to be independent of one another unless the design explicitly models dependence, as in repeated-measures or panel structures. Where these assumptions are violated and cannot be remedied through transformation, non-parametric alternatives, which rely on ranks rather than raw scores and make substantially weaker distributional assumptions, become the statistically defensible choice (Pallant, 2020).

Gujarati and Porter (2009) and Wooldridge (2010) extend this logic into the regression framework, distinguishing between the level of measurement of the outcome variable and the functional form of the model required to estimate it consistently. Ordinary least squares regression is appropriate for continuous, approximately normally distributed outcomes conditional on the predictors; it is not appropriate for binary, count or strictly ordinal outcomes, for which it can produce predicted values outside the logically possible range and standard errors that are not valid for hypothesis testing. Hosmer, Lemeshow and Sturdivant (2013) provide an authoritative treatment of logistic regression as the appropriate tool where the outcome is binary, multinomial or ordinal. They note that the logit transformation constrains predicted probabilities to the [0,1]

interval and that model fit should be assessed through likelihood-based statistics and calibration measures, such as the Hosmer–Lemeshow goodness-of-fit test, rather than the R-squared statistics conventionally reported for linear regression.

Kothari (2004) places these technical considerations within the broader logic of research design, arguing that the statistical technique is not an isolated decision made after data collection but a direct consequence of the research objective, the hypotheses formulated and the sampling design adopted. A study whose objective is to compare two independent groups implies a different analytic pathway from one whose objective is to establish a predictive relationship among several variables, and the choice of tool should be settled at the proposal stage, before instruments are finalised, so that the data collected are actually capable of answering the question posed. Taken together, the classical literature establishes four criteria that should jointly determine the choice of statistical tool: the research objective, the level of measurement of the variables involved, the distributional properties of the data, and the structural features of the research design, including independence, sample size and any repeated or clustered observations (Field, 2018; Pallant, 2020; Gujarati and Porter, 2009; Wooldridge, 2010; Hosmer, Lemeshow and Sturdivant, 2013; Kothari, 2004).

Published test-selection guides already translate this logic into decision aids. Nayak and Hazra (2011), for example, present branching schemes that sort the choice of test by whether the question concerns a difference between groups or an association between variables, whether the groups are paired or independent, and whether the data are numerical or categorical. Suner et al. (2015) reviewed the decision trees then available and found that each covered only a limited set of tests, so that a researcher first had to choose the right tree before choosing the right test; they combined the trees into StatXFinder, a web-based tool that leads users through 44 yes–no questions to 85 parametric and non-parametric tests. The framework proposed in this article shares the branching logic of these guides but differs in two respects. It is organised by research objective first, and it extends beyond hypothesis tests to the regression, count, panel, multilevel, time-series and latent-variable models that dominate applied social science, health, agricultural and business research. It also attaches to each choice the assumptions that must be checked before the result can be trusted.

Recent methodological work also cautions against applying such decision rules mechanically. Rochon, Gondan and Kieser (2012) showed that the widely taught two-stage procedure, in which a preliminary Shapiro–Wilk test decides between the t-test and the Mann–Whitney U test, itself distorts the Type I error rate of the main analysis. Fagerland (2012) demonstrated that in large samples the t-test remains valid even for heavily skewed data, and that the Wilcoxon–Mann–Whitney test answers a different question from a comparison of means, so that switching to a rank test in a large study can answer the wrong question. Delacre, Lakens and Leys (2017) argue that Welch's t-test should be the default for two independent groups, because it controls the Type I error rate when variances differ and loses little when they are equal. A parallel debate concerns count data: Ver Hoef and Boveng (2007) show that quasi-Poisson and negative binomial regression weight large and small counts differently, because they assume different relationships between the variance and the mean, and can therefore give materially different estimates. The common lesson is that assumption checking is a matter of judgement, informed by sample size, plots and knowledge of the design, rather than a single pre-test p-value.

Where several models are defensible, the literature on model selection offers explicit criteria for choosing among them. Burnham and Anderson (2004) set out the information-theoretic basis of the Akaike and Bayesian information criteria and argue for multimodel inference, in which a small set of candidate models is specified in advance and compared, rather than one model being chosen informally and then treated as if it were known to be true. Evidence from many-analysts studies shows why this matters. When Silberzahn et al. (2018) gave 29 research teams the same data set and the same research question, their estimates varied widely, and neither the analysts' expertise nor their prior beliefs explained the variation. Breznau et al. (2022) repeated the exercise in social science with 73 teams testing one hypothesis on the same survey data and found results ranging from substantial negative to substantial positive effects, with most of the variation unexplained by the analytical decisions they could identify. A decision framework can narrow the set of defensible techniques, but it cannot remove every choice; the remaining choices must be made visible.

This is the concern of the reproducible-research movement. Simmons, Nelson and Simonsohn (2011) showed through simulation that undisclosed flexibility in data collection and analysis can raise the actual false-positive rate far above the nominal 5%. Munafò et al. (2017) and Nosek et al. (2018) propose remedies that include specifying the research questions and analysis plan before the outcome data are seen. On the reporting side, the SAMPL guidelines ask authors to state which test was used for each analysis and to confirm that the data met that test's assumptions (Lang and Altman, 2015), while the American Statistical Association has urged researchers to move beyond treating $p < 0.05$ as a verdict (Wasserstein, Schirm and Lazar, 2019). The FAIR principles (Wilkinson et al., 2016) extend the same logic to the data and code themselves, which should be findable, accessible, interoperable and reusable. Read together, this literature supports the central claim of this article: the choice of statistical tool should be made at the design stage, recorded in advance and reported transparently.

Framework and Decision Criteria

The framework proposed here turns the literature reviewed above into a sequential decision process. The researcher first identifies the type of research objective being pursued: description, comparison, association, prediction or classification. Second, the level of measurement of the dependent (outcome) variable is established. Third, the distributional assumptions relevant to that variable, together with the structural features of the design (independence of observations, repeated or clustered measurement, and sample size), are assessed; for continuous outcomes the most common checks are normality and homogeneity of variance. Fourth, the appropriate statistical tool is selected at the intersection of these considerations. Table 1 summarises this decision matrix for the tools discussed in this article, and Figure 1 presents it as a flowchart.

Table 1: Decision matrix linking research objective, data type, distributional assumptions and design features to appropriate statistical tools

Research Objective	Data Type (DV)	Distribution / Design	Appropriate Tool(s)
Describe a single group	Nominal	n/a	Frequencies, percentages, mode
Describe a single group	Ordinal	n/a	Median, IQR, frequencies
Describe a single group	Interval/Ratio	Normal	Mean, standard deviation
Describe a single group	Interval/Ratio	Non-normal / skewed	Median, IQR
Compare two independent groups	Interval/Ratio	Normal, equal variances	Independent-samples t-test
Compare two independent groups	Interval/Ratio	Normal, unequal variances	Welch's t-test
Compare two independent groups	Ordinal or non-normal interval	Non-normal / small n	Mann–Whitney U test
Compare two related/paired groups	Interval/Ratio	Normal	Paired-samples t-test
Compare two related/paired groups	Ordinal or non-normal	Non-normal	Wilcoxon signed-rank test
Compare 3+ independent groups	Interval/Ratio	Normal, equal variances	One-way ANOVA (with post-hoc)
Compare 3+ independent groups	Interval/Ratio	Normal, unequal variances	Welch's ANOVA
Compare 3+ independent groups	Ordinal or non-normal	Non-normal	Kruskal–Wallis H test
Compare 3+ related groups (repeated measures)	Interval/Ratio	Normal, sphericity	Repeated-measures ANOVA
Compare 3+ related groups (repeated measures)	Ordinal or non-normal	Non-normal	Friedman test

Research Objective	Data Type (DV)	Distribution / Design	Appropriate Tool(s)
Compare groups controlling for a covariate	Interval/Ratio	Normal, linear covariate relation	ANCOVA
Compare groups on multiple DVs simultaneously	Interval/Ratio (multiple)	Multivariate normal	MANOVA
Test association between two categorical variables	Nominal/Ordinal	Adequate expected cell counts	Chi-square test of independence (Fisher's exact test for small expected counts)
Test association between two continuous variables	Interval/Ratio	Normal, linear relationship	Pearson correlation
Test association between two ordinal variables	Ordinal	Monotonic relationship	Spearman's rho / Kendall's tau
Predict a continuous outcome from predictors	Interval/Ratio	Normal residuals, linearity	Simple / multiple linear regression
Predict a binary outcome	Nominal (binary)	Independent observations; adequate events per predictor	Binary logistic regression
Predict a multi-category outcome	Nominal (>2 categories)	Independence of irrelevant alternatives	Multinomial logistic regression
Predict an ordered categorical outcome	Ordinal	Proportional odds	Ordinal logistic regression
Predict a count outcome	Ratio (counts)	Poisson / overdispersed	Poisson or negative binomial regression
Analyse repeated observations over time/firms	Interval/Ratio, panel	Fixed or random effects	Panel data (fixed/random effects) models
Analyse nested or clustered data (e.g., pupils within schools)	Interval/Ratio, binary or count	Dependence within clusters	Multilevel (mixed-effects) models
Model long-run and short-run time-series relationships	Interval/Ratio, time series	Mixed I(0)/I(1); no I(2) variables	ARDL / time-series models
Test complex latent-variable relationships	Interval/Ratio (indicators)	Approx. normal, adequate sample	Structural Equation Modelling (SEM)

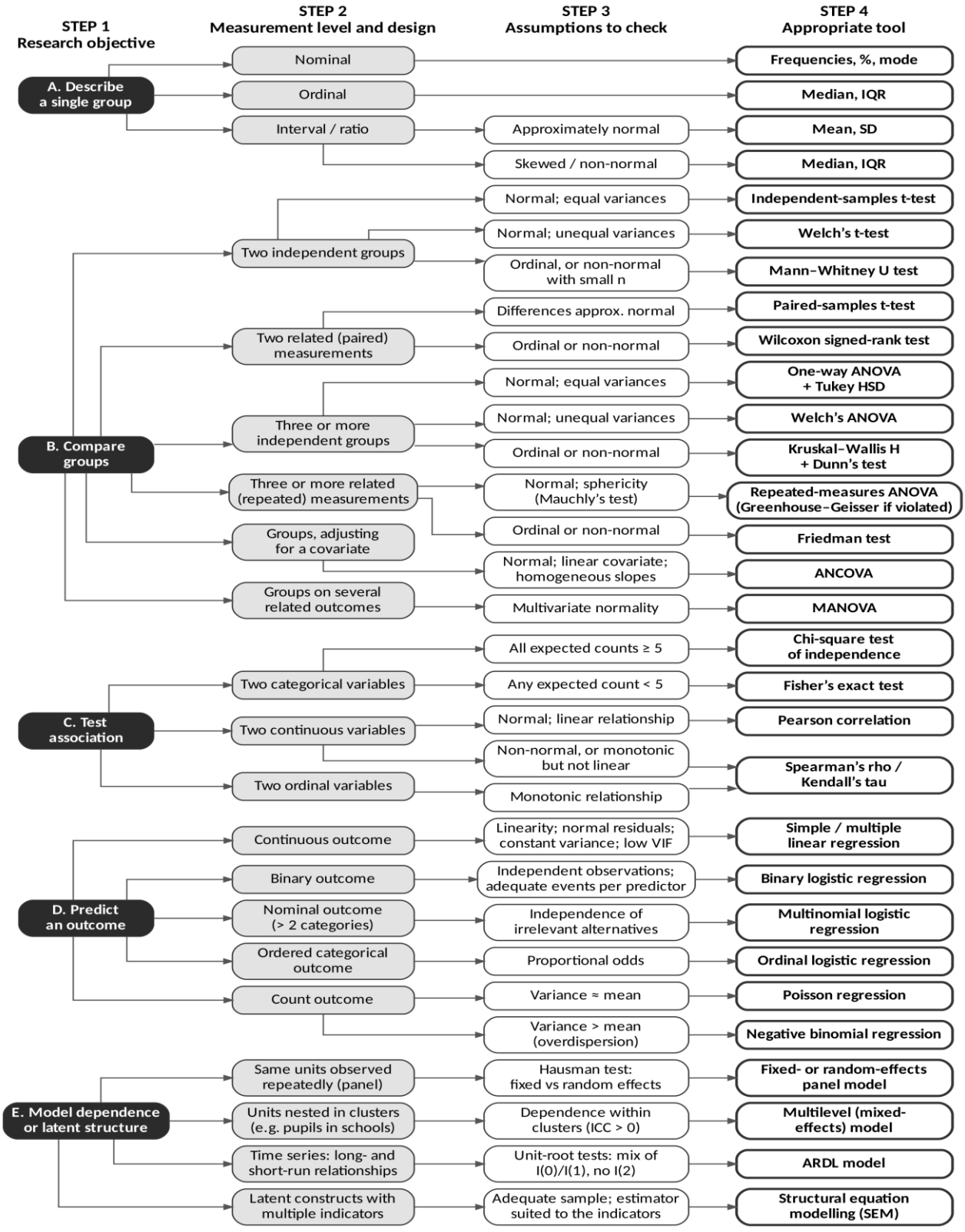
Source: Authors' compilation based on Field (2018), Pallant (2020), Hosmer, Lemeshow and Sturdivant (2013) and Wooldridge (2010).

This matrix is deliberately organised by objective rather than by tool, reflecting the central argument of this article: the statistical technique should be derived from the question being asked and the nature of the data available to answer it, not selected first and imposed on the data regardless of fit. The matrix is a guide to defensible defaults rather than an exhaustive algorithm; where a design combines several features, such as clustered count data, the researcher should consult specialised sources and, where possible, a statistician.

Figure 1 presents the same decision logic as a flowchart, so that it can be used at the proposal stage without reading across the full matrix. The researcher enters at Step 1 by identifying the research objective, then follows the branch that matches the measurement level and design of the outcome (Step 2) and the assumptions that the data must satisfy (Step 3), arriving at the appropriate tool (Step 4).

Branch E takes precedence whenever observations are repeated, clustered or ordered in time, even if the objective is a comparison or a prediction, because standard tests that assume independent observations understate uncertainty in these designs.

Figure 1: Four-step decision flowchart for selecting an appropriate statistical tool



Note: Step 3 names the conditions that should hold before the Step 4 tool is used; they call for judgement based on plots, sample size and design, not a single pre-test p-value (Rochon, Gondan and Kieser, 2012). Branch E applies whenever observations are repeated, clustered or time-ordered. IQR = interquartile range; SD = standard deviation; VIF = variance inflation factor; ICC = intraclass correlation coefficient; I(0), I(1), I(2) = order of integration. *Source:* Authors' construction based on Table 1.

Choosing between defensible alternatives

The flowchart identifies a defensible default, but at several points more than one technique can be justified, for example Student's or Welch's t-test for two groups, a t-test or a rank test for a moderately skewed outcome, or Poisson or negative binomial regression for counts. Four rules settle most of these cases. First, choose the tool whose null hypothesis matches the research question: a t-test compares means, whereas the Mann–Whitney U test compares whole distributions, and the two are not interchangeable in large samples (Fagerland, 2012). Second, where the alternatives differ mainly in robustness, prefer the one that remains valid if the most plausible assumption fails; on this basis Welch's t-test is the safer default for two independent groups (Delacre, Lakens and Leys, 2017), and negative binomial regression is preferred to Poisson regression when the variance of the counts exceeds the mean. Third, where competing models remain defensible, specify the candidate models in advance and compare them with explicit criteria, such as likelihood-ratio tests or information criteria, rather than choosing among them after seeing which gives the most favourable result (Burnham and Anderson, 2004). Fourth, name one analysis as primary in the analysis plan and report the others as sensitivity analyses; agreement across them strengthens the conclusion, and disagreement is itself a finding that should be reported (Silberzahn et al., 2018).

Discussion of Statistical Tools

Parametric Tools

Parametric tools are so called because they estimate, or make assumptions about, population parameters such as the mean and variance, and they are the most powerful class of tests when their assumptions are satisfied (Field, 2018). The independent-samples t-test compares the means of a continuous outcome across two independent groups and requires approximate normality within each group; where variances differ, Welch's version of the test should be used. The paired-samples t-test performs the corresponding comparison for two related measurements taken from the same subjects, such as pre-test and post-test scores. One-way ANOVA extends the logic of the t-test to three or more independent groups, testing whether at least one group mean differs from the others; a statistically significant omnibus result is typically followed by post-hoc pairwise comparisons (for example, Tukey's HSD) to identify which specific groups differ. Two-way ANOVA examines the separate and interactive effects of two categorical factors on a continuous outcome, while repeated-measures ANOVA is used when the same subjects are measured on three or more occasions. It requires the additional assumption of sphericity, assessed with Mauchly's test and corrected, where violated, using the Greenhouse–Geisser adjustment.

ANCOVA (analysis of covariance) combines the logic of ANOVA with regression, allowing the researcher to compare group means on an outcome while statistically controlling for one or more continuous covariates, for example comparing post-intervention scores across groups while controlling for the pre-intervention baseline. MANOVA (multivariate analysis of variance) is used when the researcher has two or more related continuous dependent variables and wishes to test whether groups differ on the combination of these outcomes, which avoids the inflated Type I error risk of running several separate ANOVAs (Field, 2018).

Pearson correlation quantifies the strength and direction of a linear association between two continuous, approximately normally distributed variables, producing a coefficient (r) between −1 and +1. A significant correlation establishes association, not causation, a distinction that is frequently blurred in applied writing (Gujarati and Porter, 2009). Simple linear regression extends this logic to predict a continuous outcome from a single predictor, while multiple linear regression accommodates several predictors simultaneously, allowing the researcher to estimate the unique contribution of each predictor while holding the others constant. Both are subject to diagnostic checks for linearity, homoscedasticity, normality of residuals and the absence of severe multicollinearity (Gujarati and Porter, 2009; Wooldridge, 2010).

Logistic regression is the appropriate parametric tool when the outcome variable is categorical rather than continuous. Binary logistic regression models the probability of one of two outcomes (for example, adoption versus non-adoption of an agricultural technology); multinomial logistic regression extends this to outcomes with more than two unordered categories; and ordinal logistic regression is used when the outcome has more

than two categories with a natural order, such as a three-point rating of food security status, and relies on the proportional odds assumption (Hosmer, Lemeshow and Sturdivant, 2013).

Non-Parametric Tools

Non-parametric tools make substantially weaker assumptions about the distribution of the underlying data and are appropriate when outcome variables are ordinal, when samples are small, or when continuous data depart markedly from normality and cannot be adequately transformed (Pallant, 2020). Despite a common misconception that non-parametric tests are a lesser or fallback option, in these circumstances they are the more statistically valid choice and should be presented as equally legitimate rather than as a compromise. Their principal cost is a modest loss of statistical power when parametric assumptions are in fact satisfied, which is why the choice should rest on the properties of the data rather than on habit in either direction.

The chi-square test of independence examines whether two categorical (nominal or ordinal) variables are associated, comparing the observed cell frequencies in a contingency table with the frequencies expected under independence; the chi-square goodness-of-fit test instead compares the observed frequency distribution of a single variable with a hypothesised distribution. The Mann–Whitney U test is the non-parametric counterpart of the independent-samples t-test, comparing the rank distributions of two independent groups, and is appropriate for ordinal outcomes or for continuous outcomes that violate normality. The Wilcoxon signed-rank test plays the equivalent role for two related or paired measurements. The Kruskal–Wallis H test extends the Mann–Whitney logic to three or more independent groups, as the non-parametric counterpart of one-way ANOVA, while the Friedman test plays the same role for three or more related measurements, corresponding to repeated-measures ANOVA.

Spearman's rank-order correlation (rho) and Kendall's tau both quantify the strength of association between two ordinal variables, or between continuous variables whose relationship is monotonic but not necessarily linear, by working with ranks rather than raw values. Kendall's tau is generally preferred to Spearman's rho for small samples or where there are many tied ranks, as it tends to give a more accurate estimate under those conditions (Field, 2018).

Advanced and Specialised Tools

Beyond the standard parametric and non-parametric toolkit, several specialised techniques are increasingly relevant to African social science, health, agricultural and business research. Count models are appropriate when the outcome is a count of events, such as the number of antenatal visits, pest infestations per plot or customer complaints per branch, rather than a continuous measurement: Poisson regression in general, and negative binomial regression where the variance of the counts substantially exceeds the mean (overdispersion) (Wooldridge, 2010; Cameron and Trivedi, 2013). Panel data models, particularly fixed-effects and random-effects regression, are appropriate when the same units (individuals, firms, districts) are observed repeatedly over time, because they allow the researcher to control for unobserved, time-invariant differences between units; the Hausman test is conventionally used to decide between a fixed- and a random-effects specification (Wooldridge, 2010).

ARDL (autoregressive distributed lag) and related time-series models are appropriate when the research objective concerns the long-run and short-run relationships among variables observed over time, particularly where the variables are integrated of different orders (a mixture of I(0) and I(1), provided that no variable is I(2)), so that a conventional cointegration test such as the Johansen procedure would not apply (Pesaran, Shin and Smith, 2001). Structural equation modelling (SEM) is appropriate where the research objective involves testing relationships among latent constructs measured by multiple indicators, such as a hypothesised model linking service quality, customer satisfaction and loyalty in business research; it combines elements of factor analysis and multiple regression within a single simultaneous framework (Kline, 2016). Multilevel (mixed-effects) models are appropriate where observations are nested within higher-level units, such as patients within clinics or pupils within schools, because they partition variance between levels and correct the standard errors that single-level models would understate (Field, 2018).

Applied Examples

This section presents nine worked examples across the major tool categories, each grounded in a plausible African research scenario, to demonstrate the practical application of the decision framework set out in Section 5. All scenarios, sample sizes and statistical outputs are hypothetical and were constructed for illustration only; they are not drawn from real studies and should not be cited as empirical evidence. Each example follows a common structure (context, problem, objective, hypothesis, tool, justification and output) so that readers can trace the reasoning from research question to analytic choice. Hypotheses are stated as alternative hypotheses about population parameters; statistical significance is a property of a test result, not of a hypothesis.

Independent-Samples t-Test

Context: Kenya (Health)

Problem Statement: A study seeks to determine whether a community health-worker nutrition counselling intervention improves maternal haemoglobin levels among pregnant women in a rural sub-county.

Research Objective: To compare mean haemoglobin concentration (g/dL) at 36 weeks' gestation between women who received the intervention and those who received standard care.

Hypothesis: H1: Mean haemoglobin concentration differs between the intervention and control groups.

Statistical Tool Applied: Independent-samples t-test.

Justification: The outcome (haemoglobin concentration, g/dL) is a continuous, ratio-scaled variable, two independent groups are compared, and inspection of the data together with Shapiro–Wilk and Levene's tests indicated approximate normality within each group and homogeneity of variance. Had the outcome instead been the number of antenatal visits attended, a count model (Section 6.3) would have been more appropriate.

Illustrative Output and Interpretation: $t(148) = 2.86$, $p = 0.005$, Cohen's $d = 0.47$; mean difference = 0.6 g/dL, 95% CI [0.19, 1.01]. Because $p < 0.05$, the null hypothesis of no difference is rejected; the intervention group recorded a significantly higher mean haemoglobin concentration than the control group, with a moderate effect size.

One-Way ANOVA with Tukey's HSD Post-Hoc Test

Context: Nigeria (Agriculture)

Problem Statement: Agricultural extension officers wish to compare maize yield across four fertiliser treatment regimes applied to demonstration plots.

Research Objective: To compare mean maize yield (kg/ha) across four fertiliser treatments.

Hypothesis: H1: At least one fertiliser treatment produces a mean yield that differs from the others.

Statistical Tool Applied: One-way ANOVA with Tukey's HSD post-hoc test.

Justification: The outcome (yield) is continuous and ratio-scaled, there are more than two independent groups (treatments), and normality and homogeneity of variance were checked and found to be reasonably satisfied across treatment plots, meeting the assumptions of ANOVA.

Illustrative Output and Interpretation: $F(3, 76) = 5.92$, $p = 0.001$, $\eta^2 = 0.19$, a large effect by conventional benchmarks (Cohen, 1988). The significant omnibus result indicates a treatment effect; Tukey's post-hoc comparisons show that Treatment 3 produced a significantly higher mean yield than Treatments 1 and 2 ($p < 0.05$), while the other pairwise differences were not significant.

Chi-Square Test of Independence

Context: Zambia (Social Science)

Problem Statement: A researcher examines whether household food security status is associated with the sex of the household head, using survey data from two districts.

Research Objective: To determine whether an association exists between the sex of the household head (male/female) and food security status (secure/insecure).

Hypothesis: H1: Food security status is associated with the sex of the household head.

Statistical Tool Applied: Chi-square test of independence.

Justification: Both variables are nominal, and the chi-square test of independence is the correct tool for testing association between two categorical variables. All expected cell frequencies exceeded the conventional minimum of 5; had they not, Fisher's exact test would have been preferred.

Illustrative Output and Interpretation: $\chi^2(1, N = 240) = 6.41$, $p = 0.011$, Cramér's $V = 0.16$ (equivalent to the phi coefficient for a 2×2 table). The association is statistically significant: female-headed households in the sample show a higher proportion of food insecurity than male-headed households, though the effect size indicates a small relationship.

Kruskal–Wallis H Test

Context: Ghana (Business)

Problem Statement: A study investigates whether employee ratings of supervisor support, measured on a single five-point Likert item, differ across three departments of a manufacturing firm.

Research Objective: To compare employee-rated supervisor support across three departments.

Hypothesis: H1: The distribution of supervisor support ratings differs across the three departments.

Statistical Tool Applied: Kruskal–Wallis H test.

Justification: Supervisor support was measured on a single ordinal Likert item. Treating these responses as interval data and applying ANOVA would require a measurement assumption that a single item does not support, so the non-parametric Kruskal–Wallis test, which compares rank distributions across three or more independent groups, is the appropriate choice. Had support been measured by a validated multi-item scale, a one-way ANOVA on the composite score could also have been defended (see Section 4).

Illustrative Output and Interpretation: $H(2) = 8.74$, $N = 150$, $p = 0.013$, $\varepsilon^2 = 0.06$. Because $p < 0.05$, the null hypothesis of equal distributions is rejected; Dunn's pairwise comparisons with Bonferroni adjustment indicate that the Production department reported significantly lower supervisor support than the Sales and Administration departments.

Multiple Linear Regression

Context: Uganda (Agriculture/Business)

Problem Statement: A cooperative wants to know whether farm size and access to credit predict the quantity of coffee (kg) sold by smallholder farmers.

Research Objective: To determine the extent to which farm size and access to credit predict the quantity of coffee sold.

Hypothesis: H1: Farm size and access to credit are positively associated with the quantity of coffee sold.

Statistical Tool Applied: Multiple linear regression.

Justification: The outcome (quantity sold, kg) is continuous, two predictors are of interest simultaneously (credit access is entered as a binary indicator), and residual diagnostics confirmed linearity, homoscedasticity and approximately normal residuals, with acceptable variance inflation factors, satisfying the assumptions of ordinary least squares.

Illustrative Output and Interpretation: $R^2 = 0.34$, adjusted $R^2 = 0.33$, $F(2, 117) = 29.85$, $p < 0.001$. Standardised coefficients show that farm size ($\beta = 0.41$, $p < 0.001$) and credit access ($\beta = 0.22$, $p = 0.008$) are both significant positive predictors; together the model explains 34% of the variance in the quantity of coffee sold. Because the design is cross-sectional, these coefficients indicate association rather than causal effect.

Binary Logistic Regression

Context: Tanzania (Health)

Problem Statement: A hospital records department wants to identify predictors of whether patients default on tuberculosis treatment (default versus completed).

Research Objective: To identify factors predicting treatment default among TB patients.

Hypothesis: H1: Distance to the health facility and patient age are associated with the likelihood of treatment default.

Statistical Tool Applied: Binary logistic regression.

Justification: The outcome is binary (default/completed), which violates the continuous-outcome assumption of linear regression; logistic regression correctly models the log-odds of default as a function of the predictors and constrains predicted probabilities to the valid [0,1] range.

Illustrative Output and Interpretation: Model $\chi^2(2) = 18.63$, $p < 0.001$; distance (per additional 5 km): OR = 1.34, 95% CI [1.10, 1.63], $p = 0.004$; age (per year): OR = 0.97, 95% CI [0.94, 0.999], $p = 0.041$; Hosmer–Lemeshow $\chi^2(8) = 6.12$, $p = 0.63$, indicating adequate calibration. Patients living farther from the facility have significantly higher odds of default (34% higher for every additional 5 km), while older patients have slightly lower odds (approximately 3% lower per year of age), controlling for the other predictor.

Negative Binomial Regression

Context: Malawi (Agriculture)

Problem Statement: A plant pathology study records the number of armyworm-infested maize plants per plot across three agro-ecological zones over a growing season.

Research Objective: To model the effect of agro-ecological zone on the count of infested plants per plot.

Hypothesis: H1: The expected count of infested plants differs across agro-ecological zones.

Statistical Tool Applied: Negative binomial regression.

Justification: The outcome is a count variable with evidence of overdispersion (variance substantially exceeding the mean, confirmed by a dispersion statistic), which violates the equidispersion assumption of Poisson regression. Negative binomial regression relaxes this assumption while retaining the count-model framework, making it the statistically defensible choice over ordinary least squares or standard Poisson regression (Cameron and Trivedi, 2013; Ver Hoef and Boveng, 2007).

Illustrative Output and Interpretation: Likelihood ratio $\chi^2(2) = 14.27$, $p < 0.001$; incidence rate ratio (IRR) for the high-risk zone = 1.89, 95% CI [1.26, 2.83], $p = 0.002$. Plots in the high-risk zone are expected to have, on average, 89% more infested plants than plots in the reference zone.

Fixed-Effects Panel Regression

Context: Ethiopia (Business/Economics)

Problem Statement: A microfinance institution tracks loan repayment performance for the same 60 branches over five consecutive years and wants to determine the effect of branch-level training investment on repayment rates while accounting for unobserved branch characteristics.

Research Objective: To estimate the effect of training investment on the loan repayment rate across branches over time.

Hypothesis: H1: Training investment has a positive effect on the repayment rate after controlling for unobserved branch heterogeneity.

Statistical Tool Applied: Fixed-effects panel regression.

Justification: The data have a panel structure (the same branches observed repeatedly over time). A pooled OLS regression would ignore unobserved, time-invariant branch characteristics that may be correlated with training investment and so bias the estimate, whereas the fixed-effects estimator controls for this heterogeneity. A Hausman test confirmed that fixed effects were preferred to random effects for this data set, and standard errors were clustered by branch to allow for serial correlation within branches.

Illustrative Output and Interpretation: Within $R^2 = 0.28$; coefficient on training investment = 0.62 (cluster-robust SE = 0.20), $p = 0.003$. After controlling for branch fixed effects, a one-unit increase in training investment (ETB 100,000 per branch per year) is associated with a significant 0.62-percentage-point increase in the repayment rate.

Structural Equation Modelling (SEM)

Context: South Africa (Business)

Problem Statement: A retail research team wishes to test a hypothesised model in which perceived service quality influences customer satisfaction, which in turn influences customer loyalty, using multi-item Likert scales for each latent construct.

Research Objective: To test the structural relationships among service quality, satisfaction and loyalty as latent constructs.

Hypotheses: H1: Service quality has a positive effect on satisfaction. H2: Satisfaction has a positive effect on loyalty.

Statistical Tool Applied: Structural equation modelling (SEM).

Justification: Each construct (service quality, satisfaction, loyalty) is latent and measured by multiple observed indicator items rather than a single variable. SEM allows simultaneous estimation of the measurement model (relating indicators to latent constructs) and the structural model (relating constructs to one another), which neither simple regression nor correlation can accommodate. Because the indicators are ordinal Likert items, an estimator suited to ordinal or non-normal indicators (such as WLSMV or robust maximum likelihood) should be used, and the measurement model should be validated (factor loadings, composite reliability and average variance extracted) before the structural paths are interpreted (Kline, 2016).

Illustrative Output and Interpretation: Model fit (N = 356): $\chi^2(87) = 164.2$, $p < 0.001$; CFI = 0.96, RMSEA = 0.05, SRMR = 0.04, indicating acceptable fit against commonly used benchmarks (Hu and Bentler, 1999); a significant χ^2 is expected in samples of this size and is not by itself evidence of poor fit. Standardised path coefficients: service quality to satisfaction, $\beta = 0.58$, $p < 0.001$; satisfaction to loyalty, $\beta = 0.64$, $p < 0.001$. Both hypothesised structural paths are supported.

Common Errors in Practice

Several recurring errors account for a large share of the statistical misapplication encountered in African postgraduate and early-career research. Recognising them explicitly is a necessary step towards correcting the underlying pattern described in Section 2.

- Applying parametric tests or Pearson correlation to single-item Likert-scale (ordinal) data without stating and justifying the measurement assumption. Because a single item does not guarantee equal intervals between response categories, Spearman's rho, Kendall's tau or a rank-based comparison is the default choice unless the measure is a multi-item composite that can be justified as approximately interval (Pallant, 2020; Norman, 2010). Where a parametric analysis is used, a non-parametric check should confirm that the substantive conclusion is unchanged. The aim is to state and check the measurement assumption, not to prohibit parametric tests.

- Running independent-samples t-tests or ANOVA without first examining normality and homogeneity of variance. Normality should be judged from histograms and Q–Q plots as well as formal tests such as Shapiro–Wilk or Kolmogorov–Smirnov, which become oversensitive in large samples, and a single pre-test should not decide the analysis mechanically (Rochon, Gondan and Kieser, 2012). Where normality fails and cannot be corrected by transformation, the Mann–Whitney U or Kruskal–Wallis test should be used; where only the equal-variance assumption fails, Welch's t-test or Welch's ANOVA should be used (Field, 2018; Delacre, Lakens and Leys, 2017).

- Using linear regression for binary, multinomial or ordinal outcomes. This produces predicted values that can fall outside the [0,1] probability range and standard errors that are invalid for inference; logistic regression in its binary, multinomial or ordinal form should be used instead (Hosmer, Lemeshow and Sturdivant, 2013).

- Interpreting a statistically significant correlation or regression coefficient as evidence of causation, particularly in cross-sectional survey designs that cannot rule out reverse causality or confounding (Gujarati and Porter, 2009).

- Treating panel, clustered or repeated-measures data as independent cross-sectional observations. This violates the independence assumption underlying standard OLS, t-test and ANOVA standard errors and understates the true uncertainty of the estimates (Wooldridge, 2010).

- Applying Poisson regression to count data without checking for overdispersion, which leads to underestimated standard errors and overstated statistical significance; where overdispersion is present, a negative binomial model should be fitted and compared (Cameron and Trivedi, 2013; Ver Hoef and Boveng, 2007).

- Reporting p-values without effect sizes or confidence intervals, and treating $p < 0.05$ as a binary verdict on the importance of a finding rather than as one piece of evidence to be weighed alongside the magnitude and precision of the estimate (Wasserstein and Lazar, 2016; Wasserstein, Schirm and Lazar, 2019).

- Selecting a statistical technique only after data collection is complete, rather than at the proposal stage. This frequently results in data that are structurally incapable of answering the stated research question, whichever test is subsequently applied (Kothari, 2004).

RECOMMENDATIONS

The following recommendations are directed at researchers, supervisors and research training programmes seeking to reduce the incidence of statistical misapplication.

- Determine the statistical analysis plan at the proposal stage, deriving it directly from the stated research objectives and hypotheses rather than after the data have been collected.
- Record the analysis plan, including the primary analysis and any planned sensitivity analyses, before outcome data are examined, and share data and analysis code where ethics and consent permit (Nosek et al., 2018; Wilkinson et al., 2016).
- At the proposal stage, use an a priori power analysis to determine the sample size needed to detect a substantively meaningful effect with the chosen statistical technique (Cohen, 1988).
- Classify each variable explicitly by level of measurement (nominal, ordinal, interval or ratio) before selecting an analytic technique, and record this classification in the methodology chapter.
- Test, rather than assume, the statistical assumptions relevant to the chosen technique (normality, homogeneity of variance, linearity, independence, absence of severe multicollinearity and, for count models, dispersion) and report the results of these diagnostic checks transparently (Lang and Altman, 2015).
- Report effect sizes and confidence intervals alongside p-values for every principal test, and interpret findings in terms of their practical as well as their statistical significance (Wasserstein and Lazar, 2016).
- Treat non-parametric techniques as equally legitimate methodological choices where the properties of the data warrant them, rather than as inferior substitutes for parametric tests.
- Where the research design involves repeated or clustered observations (panel data, repeated measures, nested sampling), select a technique that explicitly models that dependence rather than defaulting to methods that assume independent observations.
- Supervisors and research methods courses should build the decision framework in Table 1 and Figure 1 into proposal review, requiring students to justify tool selection with reference to objective, data type and distributional assumptions before proceeding to data collection.
- Institutions should strengthen access to statistical and biostatistical consultation at the proposal stage, when correction is inexpensive, rather than at the point of thesis examination or journal review, when correction may require re-analysis or re-collection of data.

CONCLUSION

The alignment between research objectives, data properties and statistical tools is not a peripheral technical detail but a foundational determinant of research validity. This article has argued that the recurring pattern of statistical misapplication in African postgraduate and early-career research, such as parametric tests applied to ordinal data, correlational evidence interpreted causally and default reliance on familiar procedures irrespective of data type, stems from a failure to treat tool selection as a design decision governed by clear criteria. The decision matrix, flowchart and worked examples presented here show that a small set of principles, drawn from the research objective, the level of measurement of the data and the distributional and structural assumptions of the design, is sufficient to guide correct tool selection across the social sciences, health, agriculture and business. Embedding these principles in proposal development and supervisory practice, rather than applying them retrospectively at the point of analysis, offers the most direct and sustainable route to improving the methodological rigour and credibility of African academic research.

The framework has limitations. It is designed as a guide to defensible defaults for the techniques most commonly used in postgraduate research, not as an exhaustive algorithm, and it does not address Bayesian inference, machine-learning methods, survival analysis or the integration of quantitative and qualitative strands in mixed-methods designs. Its worked examples rely on hypothetical data, and the framework has not yet been validated empirically. Validation is planned for a separate study, for example through simulation of common analytic errors and an audit of the alignment between stated objectives and analytic techniques in a sample of theses examined at African universities. Future research could also evaluate whether the framework's use in research methods teaching improves the quality of research proposals.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest: The authors declare no conflict of interest.

Data availability: No new data were created or analysed in this study. All numerical results reported in Section 7 are hypothetical and are presented for illustration only.

REFERENCES

1. Altman, D. G. (1994) 'The scandal of poor medical research', BMJ, 308(6924), pp. 283–284.
2. Breznau, N., Rinke, E. M., Wuttke, A. et al. (2022) 'Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty', Proceedings of the National Academy of Sciences, 119(44), e2203150119.
3. Burnham, K. P. and Anderson, D. R. (2004) 'Multimodel inference: Understanding AIC and BIC in model selection', Sociological Methods & Research, 33(2), pp. 261–304.
4. Cameron, A. C. and Trivedi, P. K. (2013) Regression Analysis of Count Data. 2nd edn. Cambridge: Cambridge University Press.
5. Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
6. Delacre, M., Lakens, D. and Leys, C. (2017) 'Why psychologists should by default use Welch's t-test instead of Student's t-test', International Review of Social Psychology, 30(1), pp. 92–101.
7. Fagerland, M. W. (2012) 't-tests, non-parametric tests, and large studies—a paradox of statistical practice?', BMC Medical Research Methodology, 12, article 78.
8. Field, A. (2018) Discovering Statistics Using IBM SPSS Statistics. 5th edn. London: SAGE Publications.
9. Gujarati, D. N. and Porter, D. C. (2009) Basic Econometrics. 5th edn. New York: McGraw-Hill/Irwin.
10. Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013) Applied Logistic Regression. 3rd edn. Hoboken, NJ: John Wiley & Sons.
11. Hu, L. and Bentler, P. M. (1999) 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives', Structural Equation Modeling: A Multidisciplinary Journal, 6(1), pp. 1–55.
12. Kline, R. B. (2016) Principles and Practice of Structural Equation Modeling. 4th edn. New York: Guilford Press.
13. Kothari, C. R. (2004) Research Methodology: Methods and Techniques. 2nd edn. New Delhi: New Age International Publishers.
14. Lang, T. A. and Altman, D. G. (2015) 'Basic statistical reporting for articles published in biomedical journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines', International Journal of Nursing Studies, 52(1), pp. 5–9.
15. Munafò, M. R., Nosek, B. A., Bishop, D. V. M. et al. (2017) 'A manifesto for reproducible science', Nature Human Behaviour, 1, article 0021.
16. Nayak, B. K. and Hazra, A. (2011) 'How to choose the right statistical test?', Indian Journal of Ophthalmology, 59(2), pp. 85–86.

17. Norman, G. (2010) 'Likert scales, levels of measurement and the "laws" of statistics', Advances in Health Sciences Education, 15(5), pp. 625–632.
18. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. and Mellor, D. T. (2018) 'The preregistration revolution', Proceedings of the National Academy of Sciences, 115(11), pp. 2600–2606.
19. Pallant, J. (2020) SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS. 7th edn. London: Routledge.
20. Pesaran, M. H., Shin, Y. and Smith, R. J. (2001) 'Bounds testing approaches to the analysis of level relationships', Journal of Applied Econometrics, 16(3), pp. 289–326.
21. Rochon, J., Gondan, M. and Kieser, M. (2012) 'To test or not to test: Preliminary assessment of normality when comparing two independent samples', BMC Medical Research Methodology, 12, article 81.
22. Silberzahn, R., Uhlmann, E. L., Martin, D. P. et al. (2018) 'Many analysts, one data set: Making transparent how variations in analytic choices affect results', Advances in Methods and Practices in Psychological Science, 1(3), pp. 337–356.
23. Simmons, J. P., Nelson, L. D. and Simonsohn, U. (2011) 'False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant', Psychological Science, 22(11), pp. 1359–1366.
24. Stevens, S. S. (1946) 'On the theory of scales of measurement', Science, 103(2684), pp. 677–680.
25. Strasak, A. M., Zaman, Q., Pfeiffer, K. P., Göbel, G. and Ulmer, H. (2007) 'Statistical errors in medical research – a review of common pitfalls', Swiss Medical Weekly, 137(3–4), pp. 44–49.
26. Suner, A., Karakülah, G., Koşaner, Ö. and Dicle, O. (2015) 'StatXFinder: A web-based self-directed tool that provides appropriate statistical test selection for biomedical researchers in their scientific studies', SpringerPlus, 4, article 633.
27. Ver Hoef, J. M. and Boveng, P. L. (2007) 'Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?', Ecology, 88(11), pp. 2766–2772.
28. Wasserstein, R. L. and Lazar, N. A. (2016) 'The ASA statement on p-values: Context, process, and purpose', The American Statistician, 70(2), pp. 129–133.
29. Wasserstein, R. L., Schirm, A. L. and Lazar, N. A. (2019) 'Moving to a world beyond "$p < 0.05$"', The American Statistician, 73(sup1), pp. 1–19.
30. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. et al. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data, 3, article 160018.
31. Wooldridge, J. M. (2010) Econometric Analysis of Cross Section and Panel Data. 2nd edn. Cambridge, MA: MIT Press.