Analyses

Analysis Questionable Research Practices (Wicherts et al, 2016, Table 1)

1. Choosing between different options of dealing with incomplete or missing data on ad hoc grounds.
2. Specifying pre-processing of data (e.g., cleaning, normalization, smoothing, motion correction) in an ad hoc manner.
3. Deciding how to deal with violations of statistical assumptions in an ad hoc manner.
4. Deciding on how to deal with outliers in an ad hoc manner.
5. Selecting the dependent variable out of several alternative measures of the same construct.
6. Trying out different ways to score the chosen primary dependent variable.
7. Selecting another construct as the primary outcome.
8. Selecting independent variables out of a set of manipulated independent variables.
9. Operationalizing manipulated independent variables in different ways (e.g., by discarding or combining levels of factors).
10. Choosing to include different measured variables as covariates, independent variables, mediators, or moderators.
11. Operationalizing non-manipulated independent variables in different ways.
12. Using alternative inclusion and exclusion criteria for selecting participants in analyses.
13. Choosing between different statistical models.
14. Choosing the estimation method, software package, and computation of standard errors (SEs).
15. Choosing inference criteria (e.g., Bayes factors, alpha level, sidedness of the test, corrections for multiple testing).

Guidance:

All of the above practices inappropriately (and arguably unethically) allow researchers to make analysis decisions based on the nature of the data obtained. You can avoid these QRPs by constructing a formal **data analysis plan**that concretely addresses all decisions. For example, if you plan to conduct a moderated multiple regression analysis, you should specify (prior to data collection) what alternative procedure you will use if you violate the assumptions of regression (e.g., high reliability of predictors, multivariate normality of errors, etc.). Likewise, if you plan to use a covariate in your regression, you should specify (prior to data collection), not only what the covariate is, but what alternative procedure you will use if the covariate interacts with another predictor. The principle behind doing so is that the researcher will have a clear record of their analysis intentions prior to data collection so that they can demonstrate researcher flexibility was not used during analyses. Some data analysis plans go so far as to have blank templates for the tables and graphs that will be used in the final thesis. Ideally, the data analysis plan is stored before data collection in a repository such as the Open Science Foundation. As per point 3 above, it is crucial to evaluate the assumptions under your analyses (see Osborne, 2017)

When interpreting data, a common practice is to use *p*-values. If reporting *p*-values, report them exactly and do not round down to meet significance. For example, do not round .054 to .05 (doing so avoids Questionable Research Practice #5). Unfortunately, *p*-values are poorly understood by psychological researchers. Indeed, approximately 80% of psychology professors do not understand the correct interpretation of *p*-values (Haller & Kraus, 2002; Kline, 2009, pp. 120, 125). A correct definition of a *p*-value is available in Kline (2009)—be sure to consult this reference. In addition, there is a long history of criticism of the Null Hypothesis Significance Testing Process (NHSTP) that questions the value

of the practice (e.g., Cohen, 1994; Cumming, 2008). Indeed, although most journals accept *p*-values, some have banned them (see Woolston, 2015).

In addition, the American Statistical Association has issued a statement with a few key points about *p*-values (see below). These points were designed to provide "*principles to improve the conduct and interpretation of quantitative science.*" Context for these points is also available.

"*p*-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone"

"By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis."

"A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result."

"Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold."

The consequence of these truths is that a thesis based exclusively or primarily on *p*-values does not represent good science.

How should student researchers proceed in light of these truths? One promising approach is captured in the quotation below from the Executive Director of the American Statistical Association:

"In the post p<0.05 era, scientific argumentation is not based on whether a p-value is small enough or not. Attention is paid to effect sizes and confidence intervals. Evidence is thought of as being continuous rather than some sort of dichotomy."

Ron Wasserstein,
Executive Director,
American Statistical Association, 2016
(Read the complete interview)

Confidence Intervals

The recommendation of the Executive Director of the American Statistical Associations to interpret data using effect sizes and confidence intervals is consistent with APA task force on statistical significance (see PDF links at bottom of the task force webpage). The APA task force report stated "*Always present effect sizes for primary outcomes*" and "*Interval estimates should be given for any effect sizes involving principal outcomes*" (*p*. 599, Wilkinson, 1999). The 2016 American Statistical Association position goes beyond this by suggesting that confidence intervals and effect sizes should be the primary means of interpretation.

Confidence intervals can be constructed using raw data units (e.g., CI around a mean or mean difference) or around a standardized effect size (e.g., *r* or *d*). A survey of researchers indicated that

researchers frequently fail to understand what is conveyed by a confidence interval (Cummings, Williams & Fidler, 2004; also see this document by Howell). Consequently, it may be helpful to review how to interpret confidence intervals "by eye" (Cumming & Finch, 2005; for standard error whiskers see Cumming, Fidler, & Vaux, 2007). In most cases, if dealing with the difference between means, it's easier to interpret a confidence interval for the difference (e.g., *d*-value with CI), rather than the two means.

In short, a confidence interval can be interpreted as a plausible estimate of the range of population-level effects that could have caused the sample effect (see Cumming & Finch, 2005). Population values closer to the middle of the confidence interval are somewhat more likely than those at the extremes. In using confidence intervals, there is a temptation to use them merely as a proxy for significance testing (i.e., in a dichotomous way). This practice is ill-advised. Nevertheless, there is a tendency to do so, as indicated in the article "Editors can lead researchers to confidence intervals but can't make them think: Statistical reform lessons from medicine" by Fidler et al. (2004). That is, many researchers, when switching to confidence intervals, make the error of trying to use them to provide dichotomous evidence (i.e., reject/fail to reject the null hypothesis) rather than continuous evidence. Continuous evidence requires researchers to think about the full range of the confidence interval when interpreting their findings.

Confidence Interval Walk Away Message

We suggest using confidence intervals as the primary basis for your conclusions. When making scientific or applied conclusions ask yourself: "*are my conclusions consistent with the full range of effect sizes in the confidence interval*?" If not, revise your conclusions.

It can be difficult to know what to focus on when reporting confidence intervals. If an effect is significant, it makes sense to discuss the (absolute magnitude) lower bound of the confidence interval to indicate how small the effect could be. Conversely, it makes sense to discuss the (absolute magnitude) upper bound of the confidence interval to indicate how large the (non-significant) effect could be.

In some instances, the confidence interval may be sufficiently wide that few meaningful conclusions appear possible (e.g., the plausible population effect size ranges from near zero to large). In this event, the primary conclusion may be that a larger sample size is needed in that research domain. We provide example text for reporting confidence intervals in the next section.

Practical Significance

In addition to statistical significance, there is an increasing emphasis on the practical significance of findings (e.g., How many fewer days does a major depressive episode last given a certain treatment versus control?). Here is a great example of how to investigate practical significance in the context of testing for an interaction using regression.

Student Check List 4 of 5: Data Analyses

_____ The student has presented the committee with a **data analysis plan**that addresses each of the above points.

_____ This data analysis plan for confirmatory hypothesis testing must be completed prior to data collection.

_____ The data analysis plan clearly indicates the specific analysis that will be used for each hypothesis.

_____ Where relevant, the data analysis plan clearly indicates how the assumptions will be assessed for each hypothesis test and handled if violated.

_____ The data analysis plan clearly indicates how problematic outliers (i.e., points of influence) will be dealt with.

_____ The data analysis plan clearly indicates a strategy for handling missing data in analyses (e.g., pairwise deletion, listwise deletion, etc).

_____ In the event that covariates are to be included in any analysis, the specific covariates for each hypothesis test are mentioned in the pre-data collection analysis plan.

_____ Confidence intervals will be reported for all tests.

_____ Consider avoiding *p*-values when conducting exploratory analyses.

_____ Consistent with the TriAgency position on data management, uploading analysis scripts, descriptions of variables in the data file (i.e., data code book), and the data to an open access platform (e.g., osf.io) was considered.

Source URL:https://www.uoguelph.ca/psychology/book-page/analyses