Statistical considerations in reporting cardiovascular research

Citation for published version:

Digital Object Identifier (DOI):
10.1152/ajpheart.00309.2018

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
American Journal of Physiology-Heart and Circulatory Physiology

Publisher Rights Statement:
This is the authors accepted manuscript as accepted by American Journal of Physiology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Statistical Considerations in Reporting Cardiovascular Research

Merry L. Lindsey¹*, Gillian A Gray², Susan K. Wood³, and Douglas Curran-Everett⁴*

¹Mississippi Center for Heart Research, Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216 USA; and Research Service, G.V. (Sonny) Montgomery Veterans Affairs Medical Center, Jackson, MS 39216 USA; ²BHF/University Centre for Cardiovascular Science, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK; ³Department of Pharmacology, Physiology and Neuroscience, University of South Carolina School of Medicine, Columbia, SC; and ⁴Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, Colorado; and Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Denver, CO

*co-chairs and co-corrresponding authors

Address for Correspondence:

Merry L. Lindsey, Ph.D.
Department of Physiology and Biophysics
University of Mississippi Medical Center
2500 North State St., Jackson, MS 39216-4505, USA
(phone) +1 601-815-1329 (fax) +1 601-984-1817
email: mllindsey@umc.edu or

Douglas Curran-Everett, Ph.D.
Division of Biostatistics and Bioinformatics, M222
National Jewish Health 1400 Jackson St., Denver, CO 80206–2761
(phone) 303.398.1209 (fax) 303.270.2131
e-mail: EverettD@NJHealth.org

Running Title: Statistical Guidelines

Word count: 11,121
Abstract

The problem of inadequate statistical reporting is long-standing and widespread in the biomedical literature, including in cardiovascular physiology. Although guidelines for reporting statistics have been available in clinical medicine for some time, there are currently no guidelines specific to cardiovascular physiology. To assess the need for guidelines, we determined the type and frequency of statistical tests and procedures currently used in American Journal of Physiology: Heart and Circulatory Physiology. A PubMed search for articles published in the Journal between January 1, 2017 and October 6, 2017 provided a final sample of 146 articles evaluated for methods used and 38 articles for in depth analysis. The t-test and ANOVA accounted for 71% (212/300) of the statistical tests performed. Of 6 categories of post hoc tests, Bonferroni and Tukey were used in 63% (62/98). There was an overall lack in details provided by authors publishing in the Journal, and we compiled a list of recommended minimum reporting guidelines to aid authors in preparing manuscripts. Following these guidelines could substantially improve the quality of statistical reports and enhance data rigor and reproducibility.

Keywords: statistics, physiology, cardiovascular disease, big data, rigor and reproducibility, meta-research, meta-science
Introduction

Measuring variables of cardiac physiology is a foundation of cardiovascular research, and analyzing physiological measurements involves statistics. With increasing discussion over rigor and reproducibility,(66, 135) the goals of guidelines are to provide best practice information regarding statistical analysis and to recommend how to report statistics for cardiovascular physiology research. Up to 50% of studies are not reproducible, perhaps in part because the statistical analyses cannot be evaluated from the information given.(8) Potential issues with statistics include studies that lack adequate statistical power, use inappropriate statistical tests, fail to confirm test assumptions, fail to account for and explain outlying values or missing data, and do not consider units of analysis. Adequate reporting of statistics will help to determine if any of these issues are applicable.

This article focuses on the statistics used in cardiovascular physiology research. We review the most commonly used tests in AJP Heart publications and summarize current best practices. We provide a checklist for authors to use in designing experiments and writing manuscripts and for reviewers to use in assessing the statistical tests and procedures reported in manuscripts. In addition, the reference section is a resource for those who wish to learn more about the technical aspects of statistical approaches, which are not discussed in detail here.

We focus on statistical use in animal research, which is the majority of research reported in this journal. For statistical guidelines for clinical research, please see the recent Guidelines for the Content of Statistical Analysis Plans in Clinical Trials published by the Journal of the American Medical Association and other resources.(60, 91, 137) Our guidelines add to previous guidelines on statistical use (41, 43, 44) and dovetail with recent efforts by AJP Heart to provide guidelines for articles on antibody use, recording sympathetic nerve activity, animal models of myocardial ischemia and infarction, and cardiac physiology measurements.(19, 76, 115, 116)
Most commonly used statistical tests in AJP Heart publications

We assessed articles published by *AJP Heart* to identify the most commonly used statistical tests and to evaluate current practices in reporting statistics. The search included all 2017 journal articles published in *AJP Heart*, from January 1 to date of search (October 6, 2017).

Articles were identified from PubMed using the search term “[journal] Am J Physiol Heart Circ Physiol”. Of these 254 articles, those concerning corrections, errata, reviews, editorials, and articles in press were excluded, leaving 160 original research articles, of which all were downloaded for evaluation of methods used. Of these downloaded articles, 40 were chosen by formal random selection for an additional, in depth evaluation of the statistics used. Of the 160 articles, 14 were not evaluated because they were false positive selections (8 editorials, 1 historical perspective, and 5 computational or modeling articles that used no statistics), leaving 146 articles evaluated for statistical methods used and 38 articles for the in depth analysis. (1, 4, 5, 7, 9-18, 21-23, 25, 26, 45-59, 61, 63, 64, 67-69, 71-75, 77-80, 82-90, 92-97, 99-101, 103, 104, 108-114, 117-121, 123-126, 128-131, 133, 136, 138-145, 147-159, 162-183, 186-189, 192-199) Three evaluators abstracted the data and performed the analysis (GAG, MLL, and SKW). To assess consistency across evaluators, 20 of the 146 (5 of the 38) were randomly selected and analyzed twice (by GAG and MLL); all had good degree of concordance. Both analyzers identified the same statistical tests and were identical with the in depth evaluations of the details provided by authors.

We identified 6 categories of statistical tests: analysis of variance (ANOVA), chi-square tests, regression, t-tests, other two-sample tests, and other. Of the 300 tests, the t-test and ANOVA accounted for 212 (71%; Table 1). Because the statistics details were grouped, it was difficult to ascertain how many cases there were where multiple t-tests were used when an ANOVA was appropriate. There were only a few cases (<5) where we had suspicions that a t-test had been used instead of ANOVA. Overall, authors appear to understand what tests are appropriate to use or reviewers are requesting corrections during peer review. For the other
test category, the most frequent tests were the Shapiro-Wilk normality test, the Kolmogorov-Smirnov normality test, and the Bland-Altman analysis, accounting for 13 of 36 other tests (36%). Of the 7 post hoc tests used, including Bonferroni, Dunnett, Holm Sidak, Least Significant Difference, Student-Newman-Keuls, Tukey, and other, Bonferroni and Tukey post-tests accounted for 62/98 (63%; Table 2).

Standard error of the mean (SE) was used to report error 82% of the time (n=31 of 38), as opposed to 4 uses of the standard deviation and 3 cases where type of error reported was not identified or other was used (Table 3). All 38 articles named the tests used, and of the 146 articles evaluated, only 1 did not report what test(s) had been used. The statistical software program was reported in 66% of the 38 articles, with GraphPad Prism (https://www.graphpad.com/) and SPSS Software (IBM) accounting for 80%. Actual sample sizes for each group were reported 79% of the time; the remaining articles reported sample sizes as a range (e.g., n=6 to 8 per group). The \( P \) value was reported as \(<0.05\) for in 89%, and different \( P \)-value thresholds (i.e., assigning differences among \( P<0.05, P<0.01, \) and \( P<0.001 \)) were reported in 47% of the articles. The assumption of normality was tested for 21% of the articles, but a power analysis was reported in only 3%. In most cases, information on whether normality testing or power analysis had been completed was not provided. Practices that are not good habits in clinical research, including optional stopping or not following sequential analysis rules, could not be evaluated based on the information provided. Whether there is a proclivity towards collecting data until significance is reached may be an issue for animal and \textit{in vitro} research. Overall, this analysis highlights that while most groups appear to be using statistics appropriately, more detailed instructions are needed on what should be reported.

\textbf{Guidelines for reporting statistics: minimum details needed}

The minimum information we recommend for reporting statistical analyses comes from several sources (Table 4). This advice is in line with published guidelines, including the Animals in Research: Reporting \textit{In Vivo} Experiments (ARRIVE)
guidelines. Having a stand-alone statistical section in the methods may not be the best way to allow rigorous assessment and reproducibility of findings. Instead, incorporating statistical information in individual methods sections and figure and table legends may be more appropriate. Other options include hosting analysis scripts, data, and more detailed information (e.g., degrees of freedom and F-ratio) on repository sites such as FigShare and Open Science Framework (https://cos.io/our-products/osf/).

The use of P value thresholds (e.g., $P<0.05$) reflects both historical, formal statistical theory and practice, and the fact that P values were obtained using tables because of computational limitations. Reporting exact $P$ values rather than threshold values is important for assessing reproducibility; this is particularly true when the P value is in the 0.01 to 0.10 range. For example, a $P$ value of 0.04 in one study is statistically significant, whereas a $P$ value of 0.06 in a replicate study is not. Reproducibility issues would arise if the only information provided were whether the threshold for significance was met. At the same time, reporting exact $P$ values in an attempt to say that one comparison is more significant (has a lower $P$ value) than another comparison, is not appropriate.

The standard deviation should be reported when one replicate measurement is made for each data point. For example, if blood pressure is acquired once for each subject, standard deviation should be reported. The standard error of the mean should be used when multiple measurements are made for each data point. For example, if blood pressure is acquired multiple times for each subject and averaged, standard error of the mean should be reported. Interquartile range is another way to show variability within a group. Confidence intervals provide details on the uncertainty about the true value of the population and keep the interpretation focused on the physiology and not merely on statistical probabilities (or chance) as an explanation for differences.

Using box and whisker plots or similar graph to show individual responses instead of bar graphs is recommended for data visualization. This will allow readers to assess the
variation in individual responses. Showing individual responses may not be practical and may reduce clarity; for example, when using multiple line graphs such as in articles by Brooks et al and Zhang et al. (20, 200) We recommend that the authors select graphs that best represent the data reported.

**Common statistical tests**

Analysis plans should be chosen a priori, and contingency plans set in case there are violations of assumptions of the original tests (see http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf for more details). Flow charts can be used to determine which descriptive statistics and tests may be most appropriate for analyzing a dataset; for example figures in Bernard Rosner’s Fundamentals of Biostatistics. (146) Table 5 provides a list of the common statistical tests with descriptions and assumptions. More details on these concepts can be found in the Exploration in Statistics series published by Advances in Physiology Education. (29-39, 41) Additional resources also provide more details on specifics of individual tests. (185) In addition, we highly recommend that a statistician be consulted as needed. All statistical tests have assumptions, so it is important to determine whether your data met the assumptions of the analysis and whether the results of your statistical analysis are meaningful. There are a number of tests that can be performed to assess analysis quality; for example, test statistics, testing for residuals, and testing for colinearity. While not commonly used in the analysis of cardiovascular physiology, there are additional details that can be reported, including the coefficient of multiple determination, degrees of freedom, and measures of goodness-of-fit.

**Determination of statistical power.** Power analyses should be done during the experimental design process in order to estimate the sample size needed to detect a difference that is scientifically important. (27) Sample sizes that are too large wastes resources, while sample sizes that are too low are subject to false negative results (type II error). There is also a balance between theoretically ideal and practically feasible that needs to be considered when
designing experiments. There are a number of online calculators for power analysis that are easy to use, including http://powerandsamplesize.com/,

The main assumption of the power analysis is that the data involve random sampling. Two other considerations are 1) the power analysis is performed a priori to set a pre-planned sample size and 2) the effect size is the smallest of interest rather than a pre-observed value. A more in-depth discussion of power, including bias that occurs when small sample pilot studies are utilized to estimate the expected effect size in prospective power analysis, is beyond the scope of this article.(2)

**Outlier assessment.** An outlier is defined as a data point that deviates markedly from the other observations in the sample, located on the remote tail end of the true population. Physiologists filter outliers in several ways. Statistical analyses assume the data are free of outliers, and thus every data set should be evaluated for the presence of relevant statistical outliers before analysis to avoid faulty conclusions. If the outlying value was demonstrably incorrectly measured or an error occurred while documenting the data and correction is not possible, the value may be dropped. Determining whether the outlier is physiologically possible is one criteria that can be used to make this assessment. Several tests can be used to statistically detect outliers.(6)

The Dixon test determines whether a value is too small or large compared to its nearest neighbor.(184) The Grubb’s Test determines whether a single outlier is present, whereas the Generalized Extreme Studentized Deviate can detect more than one outlier.(70) Of course, the physiology should be considered into this assessment, and physiological plausibility can be a criteria for inclusion. The truncated outlier filtering method first replaces the maximum and minimum or the sample population prior to computing the exclusion criterion. This results in a more compact criterion for the determination of the outlier.(28)
Whether the outlier should be removed can be decided using the following guidelines. If the outlier does not change the results, it is acceptable to include the outlier. If the outlier affects overall results, the final statistical analysis with and without the outlier should be presented. In the end, whichever statistical method you chose and rationale you use to filter an outlier, it is critical to report this information in the methods and results.

**Missing data.** Even with the most rigorous study designs, missing data or subject dropouts are possible. Although missing data imposes a serious challenge to statistical analysis, there are acceptable strategies to handle such events. Several comprehensive reviews have been written on this topic; Slinker and Glantz review how to handle missing data under conditions of a two-way ANOVA,(161) and He reviews multiple imputation, a common statistical technique for analyzing incomplete data sets.(81)

**Big data analysis.** Analysis of big datasets such as omics datasets are distinct from the traditional statistical approaches discussed in these guidelines and are thus beyond the scope of the present recommendations. Big data analysis requires bioinformatics coupled with statistics for data visualization. Several tools and tests can help provide new perspectives on data, including heat maps, volcano plots, principal component analysis, pathway analysis, and clustering. Statistically, controlling for false discovery rates in evaluating multiple comparisons is particularly important for large transcriptomics or proteomics datasets.(40) Although big data analysis of omics datasets is currently not prevalent in *AJP Heart* articles, they have appeared (57, 122, 164, 171) and more are anticipated.

**Resources and software packages**

Several resources contain more detail on the use and reporting of statistics; for example, Common Statistical Errors and How to Avoid Them.(65) A number of useful decision trees on how to choose an appropriate test are available online: [www.microsiris.com/Statistical%20Decision%20Tree/](http://www.microsiris.com/Statistical%20Decision%20Tree/) and [http://statpages.info/#WhichAnalysis](http://statpages.info/#WhichAnalysis). Commonly used software include GraphPad Prism and SPSS, as well as STATA Software.
(https://www.stata.com/) and SAS Software (https://www.sas.com/en_us/home.html). Of these, GraphPad Prism is user-friendly and great for graph development but limited in performing ANOVA because it can only do a 2x2 analysis and not a larger MANOVA. There are free, valid, point-and-click alternatives such as jamovi (jamovi.org) and JASP (jasp-stats.org), and both programs include effect size estimates and other analysis options. Additionally, R (http://cran.us.r-project.org/) has a virtually endless number of packages or extensions useful for data analysis, including a markdown useful for reducing transcription errors and several advanced data visualization options. Several other online research tools that include statistical analysis and bioinformatics platforms are available. For example, Metaboanalyst (http://www.metaboanalyst.ca/) is an online program originally developed as a comprehensive tool for metabolomics analysis and interpretation that can be used for any dataset; it is not limited to only analyzing metabolomics. Metaboanalyst is a good resource of bioinformatics tools, including heat maps, volcano plots, principal component analysis, and clustering. Enrichr (http://amp.pharm.mssm.edu/Enrichr/) is an online enrichment analysis tool that contains >180,000 annotated gene sets from >100 gene set libraries.(24, 102)

**Conclusions**

This article summarizes current practices in statistical analysis reported in *AJP Heart* articles and identifies the minimum that should be included in manuscripts to allow reviewers and readers to assess data quality. The take-home messages are that statistics should be considered during the experimental design and throughout data analysis, the methods and results sections of the manuscript should describe sufficiently which tests were done for each evaluation, and there are a number of readily available resources to assist you with statistics and data visualization. Improving clarity in statistics will improve rigor and reproducibility of cardiovascular physiology studies.

**Acknowledgements**
The authors acknowledge funding from the National Institutes of Health under Award Numbers GM104357, GM114833, and GM115428, HL051971, HL075360, HL129823, and MH113892; the Biomedical Laboratory Research and Development Service of the Veterans Affairs Office of Research and Development under Award Numbers 5I01BX000505 and 1I21 BX002085; the American Heart Association Award Number 15SDG22430017; the Brain and Behavior Research Foundation Award Number 26809; the British Heart Foundation Award Number TG/18/1/33408; and Research Excellence Award (Cardiovascular Science, University of Edinburgh). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health or the Veterans Administration.
References


Table 1. Statistical procedures used in *AJP Heart* articles published between January 1, 2017 and October 6, 2017.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of variance</td>
<td>40</td>
</tr>
<tr>
<td><em>t</em> tests</td>
<td>31</td>
</tr>
<tr>
<td>Another two-sample test</td>
<td>7</td>
</tr>
<tr>
<td>Regression analyses</td>
<td>9</td>
</tr>
<tr>
<td>Chi-square tests</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 2. Frequency of post-hoc tests used following ANOVA in *AJP Heart* articles published from January 1, 2017 to October 6, 2017.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonferroni</td>
<td>33</td>
</tr>
<tr>
<td>Tukey</td>
<td>31</td>
</tr>
<tr>
<td>Dunnett</td>
<td>12</td>
</tr>
<tr>
<td>Student-Newman-Keuls</td>
<td>8</td>
</tr>
<tr>
<td>Least significant difference</td>
<td>6</td>
</tr>
<tr>
<td>Holm-Šídák</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3. Frequency of reporting details in the 38 *AJP Heart* articles evaluated.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEM reported</td>
<td>82</td>
</tr>
<tr>
<td>Statistical software identified</td>
<td>66</td>
</tr>
<tr>
<td>Sample size listed for each individual group</td>
<td>79</td>
</tr>
<tr>
<td><em>P</em> value reported as threshold (<em>P</em>&lt;0.05 vs exact <em>P</em> value)</td>
<td>89</td>
</tr>
<tr>
<td>Spurious precision</td>
<td>47</td>
</tr>
<tr>
<td>Tests for normality reported</td>
<td>21</td>
</tr>
<tr>
<td>Power analysis reported</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 4. Minimum requirements checklist for reporting statistical analyses. We recommend that the following details be provided in manuscripts to allow the data and the study's reproducibility to be assessed.

**Experimental Design**- define:
- hypothesis tested and purpose of the statistical analysis
- variables, groups, sample sizes (preferably determined by power analysis), sample randomization, significance (alpha) level

**Methods**- provide details on:
- name and version of the statistical software used
- any procedures taken to modify raw data before analysis (e.g., transformation, ratios, combining categories)
- which tests were used for which comparisons, including post-hoc tests for ANOVA, and whether corrections were made for multiple comparisons
- ancillary analyses (assumptions testing, identification and treatment of outliers and missing values)
- data and details of statistical analysis should be available for requests to assess reproducibility on open repository sites

**Results**- report:
- precise $P$ values to 2 (for 1.0 to 0.01) or 3 (for 0.009 to 0.001) decimal places; precision below $P<0.001$ not needed except for genetic associations
- variability reported using standard deviation
- confidence intervals
- data with appropriate scientific precision (e.g., report body weight with no significant digits after the decimal point)
- upload source data into a public repository (e.g., Figshare, [https://figshare.com/](https://figshare.com/)) at submission.

**Table and figure legends**:
- name tests used and sample sizes for each group in figure legends and tables
- provide information on sex of animals used, unless only one sex is stated in the methods
- data visualization- use box and whisker plots or similar instead of column graphs, to show individual responses; consider clarity of information presented
<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics</td>
<td>measures of center (mean- arithmetic average &amp; median- value in the middle and variability (SD, mean or median absolute deviation, &amp; IQR)</td>
<td>may need to be normalized; SD for single measurements, IQR for data not normally distributed</td>
</tr>
<tr>
<td>One sample comparisons</td>
<td>used to evaluate a single group- one-sample t-test (parametric) &amp; one sample chi-square test for variances</td>
<td>variables continuous, data independent, randomly selected; &amp; normally distributed; no outliers</td>
</tr>
<tr>
<td>Two group comparisons-T-test</td>
<td>used to evaluate two groups: • paired t-test (Wilcoxon signed-rank test is the non-parametric version) • unpaired t-test (Mann Whitney U test is the non-parametric version)</td>
<td>all- no outliers • parametric; dependent variable is continuous; subjects paired or dependent; data normally distributed or sample size large enough that central limit theorem is satisfied; homogeneity of variance- if unequal variation, log transform or use Wilcoxon signed-rank test • parametric; dependent variable is continuous; independent variable is categorical; dependent variable normally distributed (or sample size large enough that central limit theorem is satisfied) and randomly selected; observations are independent</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>• association- determines whether observed distribution differs from chance • goodness of fit- determines whether an observed distribution differs from known distribution.</td>
<td>non-parametric; variables are independent; relatively large sample size (minimum expected n&gt;5 for each group; if n&lt;50 for 2x2 table, use Fisher’s exact test)</td>
</tr>
<tr>
<td>Kaplan-Meier</td>
<td>time to event (e.g., survival) analysis; can accommodate censored data; non-parametric log-rank test used to compare distributions</td>
<td>data independent; time intervals uniform &amp; clearly defined; censoring similar between groups</td>
</tr>
<tr>
<td>Regression</td>
<td>predicts value of one variable from a predictor (univariate) or ≥2 predictors (multivariate) • Linear regression- correlation coefficients • Deming regression- line of best fit for</td>
<td>variables are multivariate; little or no multi-collinearity; limited autocorrelation; homogeneity of variance</td>
</tr>
<tr>
<td>Method</td>
<td>Description</td>
<td>Assumptions</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>a two-dimensional dataset</td>
<td>• Logistic regression- odds ratio (with 95% confidence intervals)</td>
<td></td>
</tr>
<tr>
<td>Bland-Altman Plot</td>
<td>• analyzes agreement between two different assays</td>
<td>data independent, randomly selected; &amp; normally distributed</td>
</tr>
<tr>
<td>≥3 group comparisons- ANOVA</td>
<td>test for differences of means among groups</td>
<td>continuous dependent variable; categorical independent variable; independent observations; data</td>
</tr>
<tr>
<td></td>
<td>• one-way- one variable examined</td>
<td>randomly sampled; dependent variables are normally distributed or sample size large enough that Central</td>
</tr>
<tr>
<td></td>
<td>• multi-way- ≥2 variables examined</td>
<td>Limit Theorem is satisfied (use log or arcsin transformation for data not normally distributed);</td>
</tr>
<tr>
<td></td>
<td>• repeated measures- over time, dose range</td>
<td>homogeneity of variance; no outliers</td>
</tr>
<tr>
<td></td>
<td>• Non-parametric: Kruskal-Wallis and Friedman</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-tests evaluate which groups are different- examples:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parametric: Bonferroni, Duncan, Dunnett, false discovery rate, Student-Newman-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keuls, Fisher least significant difference, Sidak, Holm-Sidak, Tukey</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-parametric: Dunns</td>
<td></td>
</tr>
</tbody>
</table>

ANOVA- analysis of variance; IQR- interquartile range