JOJ Ophthalmology
ISSN: 2473-5477

Juniper Publishers
key to the Researchers

Perspective
Volume 7 Issue 5 - July 2019
DOI: 10.19080/JOJO.2019.07.555724

JOJ Ophthalmol

Biostatistics 101: Perspective on Statistics in Ophthalmology

Matthew Hirabayashi[1], Dale Smith[2] and Jella An[1,3]*

[1]*University of Missouri Columbia School of Medicine, USA*

[2]*Olivet Nazarene University, USA*

[3]*Department of Ophthalmology, University of Missouri School of Medicine, USA*

Submission: June 22, 2019; **Published:** July 02, 2019

***Corresponding author:** Jella An, Department of Ophthalmology, Mason Eye Institute, University of Missouri School of Medicine, Columbia, Missouri, USA

Keywords: Biostatistics; Researchers; Statistics; Interpret; Collaboration; Clinician; Progression; Medicine; Interpretation; Translation;

Perspective

Biostatistics has trended towards more complex modeling that allows researchers to analyze data in new and more interesting ways, but these models often require a graduate degree in statistics to implement and interpret. The collaboration between clinician and biostatistician is truly remarkable and allows for the progression of medicine with the most innovative and advanced analytic techniques available. Despite this transferring data across departments for analysis unfortunately has limitations. Other than the cost and time, it's possible for the clinician's goals for a project to become lost in the exchange. Additionally, many clinicians don't have biostatistics training so the details of what specific tests and manipulations were performed and the nuances of interpretation can also become lost in translation.

The benefits of a physician running his/her own statistics for a project include reduced cost and faster turnaround. Additionally, the deeper understanding of the medicine allows for the selection of a statistical test that fits their goals for the project. This gives results that the physician understands and can better interpret, draw conclusions, and plan future studies or refine their own practice (or medical guidelines).

In cases of large-scale longitudinal studies with many between and within subject variables, these high-level statistics have allowed for new types of analyses. Some studies and funding sources require their use. We believe there is still an important place for rudimentary ("old-fashioned") statistics both in understanding and practice. Clinician researchers can perform them due to their simplicity. Most of the statistics we discuss are nearly universally used for at least describing baseline patient characteristics and understanding their appropriate use and interpretation will allow for a more critical eye when reading the work of other researchers.

Herein, a glaucoma specialist and a biostatistician present an algorithm for assigning the appropriate statistical test given study design and parameters and discuss their basics and applications (Figure 1). This represents one possible thought process and we do not explore high level; complex models here and instead focus on simple, classical statistical tests. We hope this may prove helpful for students hoping to learn basic statistics and clinicians hoping to conduct their own basic analyses in cases of simple studies like retrospective chart reviews. This is certainly not comprehensive and much of the heavy statistical jargon necessary for truly understanding the mathematics behind these tests is omitted and this can serve as a starting point for independent learning. Always check assumptions and ensure that every statistical test will provide the appropriate answer to the desired question.

Contingency Tables: Chi-Square and Fisher's Exact

These tests are appropriate for situations where the data can be presented as a contingency table (mutually exclusive categorical groups and observations). These test the null hypothesis that there is no relationship between two nominal variables. An example of this would be determining if one surgical procedure resulted in a higher proportion of successes than another. Type of surgical procedure would be one categorical variable and success would be the other. The true assumption made by Chi-Square analysis is that expected cell counts exceed 5, rather than actual counts. However, if actual cell counts exceed 5, in most cases this assumption will be met. They also require post-hoc testing by looking at adjusted residuals for each cell.

Cells with standardized residuals more than ± 2 are considered to have contributed to a significant Chi-square test due to observed values being different than expected.

Fisher's exact test tests the exact P-value for 2 X 2 contingency tables and works well for small sample sizes when cell counts are <5. Appropriate phrasing of results from the example above could include: "no association was found between surgical type and success." Good form also suggests reporting proportions with the sample sizes for clarity, e.g. 50% (50/100).

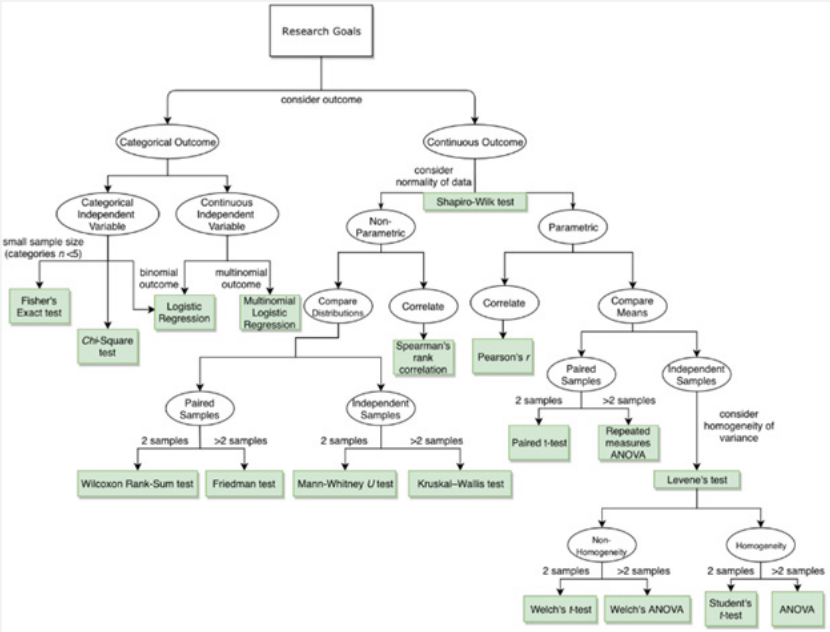

Figure 1: Flow-chart for determining the appropriate statistical test based on circumstance.

Logistic Regression

Logistic regression also tests the null hypothesis that no relationship exists between the X and Y variable. The major difference from Chi-square/Fisher's exact is that this model assesses the predictive value of continuous or categorical variables on a categorical outcome (either binomial or multinomial). This is great for looking for predictive factors of success, for example: are patients who respond well to a certain medication more likely to respond well to a certain surgery? Although linearity of relationships between independent and dependent variables, or residuals, is no longer assumed with logistic or ordinal logistic regression, assumptions regarding lack of multicollinearity and linearity of the relationship between logit of the outcome and each predictor are assumed. Further, sample size limitations also exist. Although comprehensive assessment of sample size and power are beyond the scope of this report, a general rule of thumb that can often be used as a starting point is to take ten times the number of covariates and divide by proportion of cases or positive outcomes. Using the example from Chi-square/Fisher's exact, appropriate phrasing of results would be: "response to medication did not significantly predict response to surgery."

Normality: Shapiro-Wilk

In cases where the outcome is continuous, for example trying to determine IOP or medication reduction, characterizing the data is the first step. Shapiro-Wilk can determine if the data are normally distributed. This is not only important for what tests are appropriate for the data set but also for how the data should be presented. Mean ± standard deviation (SD) is an appropriate way to report central tendency from a data set that has normally distributed values.

Mean ± SD wouldn't make sense for nonparametric (not normally distributed) data though since the mean can be influenced by skew and since there is no bell curve shape to the data there is no true SD. In these cases, median often makes more sense to report along with interquartile range (IQR). IQR is the range in which the middle 50% of the values lie. It is a great measure of variability for non-normal data. Since many people are so used to mean, non-parametric central tendency can also be reported as: mean ± SD [median (IQR)]. Naturally though, it's important to explain to the reader how the data is reported.

Correlation: Spearman's Rank Correlation and Pearson's

For cases to discover correlation between continuous independent variable and a continuous dependent variable (e.g. number of visits to an ophthalmologist and IOP) either the Spearman's rank correlation for normally distributed values or Pearson's r are appropriate test. These test the null hypothesis that there is no effect or relationship between the

How to cite this article: Matthew H, Dale Smith, Jella A. Biostatistics 101: Perspective on Statistics in Ophthalmology. JOJ Ophthalmol. 2019; 7(5): 555724. DOI: 10.19080/JOJO.2019.07.555724

groups. Again, the normality of the data is described with a Shapiro-Wilk. These produce rho values that correspond to the strength of the correlation. Values closer to 1 represent a strong positive correlation and values closer to -1 represent a strong negative correlation between variables. A rho near 0 represents no correlation. Most statistics programs will also provide the P-value, or this can be calculated online. Appropriate phrasing for findings using the example above would be: "number of visits to an ophthalmologist had no significant relationship with IOP."

Comparing Central Tendency of Normally Distributed Data

To compare means of normally distributed data (e.g. the postoperative IOP of one procedure vs. another) then the classic t-test might be appropriate. When comparing two means of two data sets that are normally distributed and independent, it's important to consider the homogeneity of variance since this is an assumption of the Student's t-test. Luckily, some software packages will automatically calculate this when running t-tests. If this assumption is met, then the Student's t-test is appropriate for comparing means. If this assumption is violated, then the Welch's t-test can correct for this. These both test the assumption that there is no difference between means.

When comparing more than two means from normally distributed data, the Analysis of Variance (ANOVA) is appropriate. This also tests the null hypothesis that there is no difference between means. A significant finding here though merits post-hoc testing. This is usually done by individual t-tests. Due to the number of t-tests that this can result in, methods like the Bonferroni correction can account for the multiple comparisons. This involves "resetting" the threshold for considering a P value significant by dividing α (usually .05) by the number of hypotheses tested. P-values lower than this new cut off may be considered significant.

Appropriate phrasing for Student's/Welch's t-test and ANOVA would be: "there was no statistically significant difference between mean IOP between the procedures." Another consideration for comparing means is if the groups are not independent. An example would be testing a value from a population at one time point and then testing the same population again (e.g. IOP before and after a procedure). For comparing two means, a paired t-test is appropriate. This tests the null hypothesis that the mean of the group is the same at both time points. For comparing more than two means, a repeated measures ANOVA is appropriate. These are both sensitive to attrition and only count cases where the same individual is represented at both time points and for that reason repeated measures ANOVA has fallen out of fashion for the more complex models that may not have this limitation. Appropriate phrasing for these paired tests would be: "there was no statistically significant difference preoperative vs. postoperative IOP."

Comparing Central Tendency of Non-Normally Distributed Data

If the research question involves comparing numbers that are not normally distributed and seeing if the distributions are different (e.g. medications a patient is taken after one procedure vs. another), the t-tests and ANOVA variants unfortunately are out of the question. To compare distributions of independent data a Mann-Whitney U test is appropriate. This does not have the assumption that the groups have normally distributed values and is considered the "non-parametric t-test alternative." One could consider it a test that compares medians but that's not quite accurate. The null hypothesis it tests is that the distributions are the same. Appropriate phrasing for this would be: "the difference in number of medications patients took was not statistically significant between procedures. "Like the paired t-test, this also has a paired variant for evaluating distributions at two different time points, the Wilcoxon Rank-Sum test. Appropriate phrasing for this would be: "the difference in number of medications patients took postoperatively was not statistically significant from preoperative status." For comparing more than two distributions there is a non-parametric alternative: the Kruskal-Wallace test. There's even a repeated measures ANOVA alternative: The Friedman test. Assumptions still remain for these types of tests as well, for example we still assume that the shape of the distributions are similar.

Conclusion

Hopefully we have provided at least a starting point for students, clinicians, and beginning statisticians to learn more about the most appropriate ways to report and analyze data. The tests we have presented in our algorithm have a wide range of uses from describing baseline patient characteristics (i.e. seeing if significant baseline differences exist between groups before comparing them) to comparisons for outcomes in cases of simple retrospective chart reviews. Understanding these basic tests allows for better communication with biostatisticians as well as better comprehension of the results they produce. We will leave the generalized estimating equations to the professionals.

How to cite this article: Matthew H, Dale Smith, Jella A. Biostatistics 101: Perspective on Statistics in Ophthalmology. JOJ Ophthalmol. 2019; 7(5): 555724. DOI: 10.19080/JOJO.2019.07.555724

DOI: 10.19080/JOJO.2019.07.555724

How to cite this article: Matthew H, Dale Smith, Jella A. Biostatistics 101: Perspective on Statistics in Ophthalmology. JOJ Ophthalmol. 2019; 7(5): 555724. DOI: 10.19080/JOJO.2019.07.555724