How to Present Statistical Results in Journal Articles

Dr. Lianjun Zhang (张连军)

State University of New York
College of Environmental Science and Forestry
(SUNY-ESF), Syracuse, NY, USA

Outline of Presentation

- Importance of appropriately reporting statistical results
- Preliminary concepts of statistics
- What should be included in each section of your manuscripts / articles
- Some notes

Importance of Appropriately Reporting Statistical Results

One of the last steps in your research is to communicate your research findings with the scientific community. Your manuscripts must be logical and rational both scientifically and statistically.

Importance of Appropriately Reporting Statistical Results

Every manuscript has to pass a peer-review process, which is an unbiased, independent, and critical assessment of the manuscript by experts.

- To determine if a manuscript reports “good science”. No journal wants to publish unimportant, poorly executed, or flawed studies. No scientist wants to read them.
- To allow the experts and colleagues in the same study field to criticize and comment on the quality of research and manuscript before it is published.
Importance of Appropriately Reporting Statistical Results

Deficiencies in reporting statistical results


They found that “the occurrence of statistical errors in Chinese medical journals was as high as 89%, with an incidence of 49% for publications sponsored by foundations at the national level and 80% at others.” And “The number of papers that use statistical tests has considerably increased among Chinese authors, but use of inappropriate statistical methods remains high, especially in the presentation of results as well as in fundamental statistics.”

Importance of Appropriately Reporting Statistical Results

Deficiencies in reporting statistical results


“Current deficiencies in presentation include: (i) failure to specify the assumptions on which an analysis is based; (ii) failure to provide sufficient quantitative information; (iii) failure to match the statistical analysis to the experimental situation; and (iv) failure to consider whether the assumptions on which an analysis has been based have been violated to an extent that would invalidate the results.”

Importance of Appropriately Reporting Statistical Results

Deficiencies in reporting statistical results

International Committee of Medical Journal Editors “Uniform Requirements for Manuscripts Submitted to Biomedical Journals” Updated October 2008 - Guidelines

• “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.
• When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals).
• Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size.
• References to the design of the study and statistical methods should be to standard works when possible (with pages stated).
• Define statistical terms, abbreviations, and most symbols. Specify the computer software used.”
Statistics is the science of collecting, organizing, and interpreting data. The objective is to make statistical inferences on a population based on the information contained in a sample selected from the population.

Key Ideas of Statistics:
- Statistical inference (sample → population)
- Sampling variability (uncertainty)
- Quantifying uncertainty (due to variability)

Procedure for a hypothesis testing
- State the null ($H_0$) and alternative ($H_a$) hypothesis
- Specify a significance level $\alpha$
- Specify the test statistic under $H_0$
- Specify the decision rule for rejecting $H_0$
- Compute the test statistic based on the sample data
- Draw conclusions

Type I and II errors
- Type I error: reject $H_0$ when it is true
  \[ P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha \]
- Type II error: fail to reject $H_0$ when it is false
  \[ P(\text{Type II error}) = P(\text{Fail to reject } H_0 \mid H_0 \text{ is false}) = \beta \]
  \[ P(\text{Reject } H_0 \mid H_0 \text{ is false}) = 1 - \beta \text{ = Power of the Test} \]
Preliminary Concepts of Statistics

- **p-value** associated with a hypothesis test
- p-value is defined as the probability of observing the test statistic based on the sample data, assuming the null hypothesis is true.
- Commonly, if p-value < $\alpha = 0.05$, we reject $H_0$. Otherwise, we fail to reject $H_0$.

What Should be Included and Reported

Abstract (usually 200-250 or 300 words)
- Problem to be investigated and purpose of research (Why)
- Important materials and methods used (How)
- Major results, interpretation, and implication (What)

What Should be Included and Reported

- Title (Title, Authors and Authors' affiliation)
- Abstract
- Introduction
- Materials and Methods
- Results
- Discussion
- Conclusion
- Acknowledgement
- References Cited
- Tables
- Figure captions and Figures

What Should be Included and Reported

Introduction
- State the nature and scope of research problem to be addressed
- Review previous research relevant to the topic and point out the “gap” in the literature
- Justify the motivation and significance of your study
- State the objectives or hypotheses of your study
What Should be Included and Reported

Materials and Methods (General Requirements)

- Subjects used and their pre-experiment handling and care
- Field site description including physical and biological features, and exact location
- Sample preparation techniques and experimental designs
- Data collection protocol and measurements of response variables
- Statistical data analysis techniques
- Equipment and computer software used

What Should be Included and Reported

Materials and Methods (Statistical Details)

- Type of experimental design (CR, Block, Split-plot, etc.)
- Treatment levels or factorial treatment combinations
- Experimental units and replications
- Nature of repeated measures (e.g., length of time intervals between measurements and total number of intervals)
- Specify the significance level used in the study (e.g., significance level \( \alpha = 0.05 \) or “We used an alpha level of 0.05 for all statistical tests”)

What Should be Included and Reported

Materials and Methods (Statistical Details)

- Sampling frames for survey data (e.g., simple random, stratified, systematic, etc.)
- Observed relationships between variables for regression studies (i.e., linear, quadratic, exponential, nonlinear, etc.) and scatterplots between variables
- Correlations between variables (Note: correlation represents the strength of linear association)
- Units of the variables
- Measurement precision (e.g., to the nearest 0.1 m)

What Should be Included and Reported

Materials and Methods (Statistical Details)

- Descriptive statistics for each variable should include: \( N, Mean, Median, SD, Minimum, Maximum \)
- For summary statistics, report one digit more than was present in the raw data (e.g., Age is recorded as the whole year, mean = 54.3 year)
- Measures of precision (e.g., SD, SE, etc.) is usually reported one more digit than the means (e.g., mean = 54.3 year; SD = 5.45 year)
- If using the notation \( A \pm B \), state clearly what the B is (i.e., SD or SE)
- It is preferred to use: The mean initial weight of animals in the study was 34.7 kg (SD = 2.61, \( N = 48 \))
### What Should be Included and Reported

#### Materials and Methods (Statistical Details)

- Provide reasons if a transformation was performed on a variable.
- Justify your choice of sample size and state the full details of a formal sample size calculation.
- Describe the nature of missing data (e.g., non-respondents, dropouts, etc.) and how you have handled missing data.

#### Results (General Requirements)

- Present **KEY** results of the experiment **without interpreting their meaning**.
- Answer the research questions logically and relate them with the methods.
- Summarize statistical analysis results with text, tables, and figures.
- Tables and figures should be self-explanatory.

#### Results (Statistical Details)

- Provide **ANOVA** table that includes the source of variation, degrees of freedom, sum of squares, appropriate error terms, test statistic, and p-value.
- State the statistical model and nature of the factors (e.g., fixed or random, nested or crossed).
- For factorial experiment, interpret the interaction term first, followed by the main factors.
- If appropriate, state the multiple comparison test used, and give reasons for the chosen test.
- Present the statistical power of the test for all non-significant results.

### What Should be Included and Reported

#### Results (Statistical Details)

- Report **R^2** and adjusted **R^2** and sample size.
- Provide the estimates of regression coefficients, standard errors and statistical test for the coefficients.
- Explain model selection criteria for the final model.
- Interpret the form and meaning of the regression model.
- Provide standard error estimates for model predictions.
- Provide checking on model assumptions and diagnostics.
- Plot raw data and fitted equation on the same graph to demonstrate the fit of regression model.
What Should be Included and Reported

Results (Statistical Details)
- For reporting statistical tests, give full details of a result in the following way: *state the test name, followed by a colon, then the test statistic (degrees of freedom), and the p-value associated with the test*
- For example: (1) t-test: $t(49) = 2.10, p-value = 0.041$; (2) ANOVA: $F(2,12) = 5.6, p-value = 0.019$; and (3) Chi-squared: $\chi^2(22) = 19.34, p-value = 0.62$
- State the hypothesis test is an one-tail or two-tail test
- Report both point estimation (e.g., mean, treatment difference) and confidence interval

What Should be Included and Reported

Results (Statistical Details)
- Test statistics and p-values should be rounded to 2 significant decimal places.
- Report the exact p-value, e.g., p-value = 0.032 rather than p-value < 0.05, p-value = 0.16 rather than p-value > 0.05.
- Do not use * for significance at $\alpha = 0.05$, ** for significance at $\alpha = 0.01$, and NS for non-significance.
- If the exact p-value is less than 0.001, say 0.0000014, use p-value < 0.001.

What Should be Included and Reported

Results (Examples)
Reporting a single sample t-test ($H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$):
Students taking statistics courses reported studying more hours for exams (Mean = 121, SD = 14.2) than did the SUNY-ESF students in general ($\mu_0 = 115$), $t(29) = 2.31, p-value = 0.014$.
What Should be Included and Reported

Results (Examples)

Reporting a significant t-test for independent groups
(H0: \( \mu_1 = \mu_2 \) vs. Ha: \( \mu_1 \neq \mu_2 \)):

Over a two-day period, the participants ate significantly less food in the experimental group (Mean = 4.67, SD = 1.55) than did those in the control group (Mean = 8.00, SD = 2.00), \( t(18) = -4.16 \), p-value = 0.00059.

What Should be Included and Reported

Results (Examples)

Reporting a significant F test for a one-way ANOVA:

An analysis of variance showed that the effect of noise was significant, \( F(3,27) = 5.94 \), p-value = 0.003. Post hoc analyses using the LSD criterion for significance indicated that the average number of errors was significantly lower in the white noise condition (Mean = 12.4, SD = 2.26) than in the other two noise conditions (traffic and industrial) combined (Mean = 13.62, SD = 5.56).

What Should be Included and Reported

Results (Examples)

Reporting the results of a \( \chi^2 \) test of independence:

A chi-square test of independence was performed to examine the relation between religion and college interest. The relation between these variables was significant, \( \chi^2(2, N = 170) = 14.14 \), p-value < 0.01. Catholic teens were less likely to show an interest in attending college than were Protestant teens.
What Should be Included and Reported

Results (Examples)

Reporting the results of a $\chi^2$ test of goodness-of-fit:
A chi-square test of goodness-of-fit was performed to determine whether the three sodas were equally preferred. Preference for the three sodas was not equally distributed in the population, $\chi^2(2, N = 55) = 4.53$, p-value < 0.05.

What Should be Included and Reported

Discussion

• Present the principles, relationships and generalizations shown by the results
• Interpret the results and relate to previous research
• Explain the reasons for confirming or denying the research hypotheses
• Explain unexpected findings and possible sources of errors
• Mention possible applications and implications of the study
• Provide suggestions for improvement and future research

What Should be Included and Reported

Conclusion

• Summarize the major results
• Interpret the significance of your study and article

Some Notes

Can We Accept a Null Hypothesis?

• Technically no.
• A commonly held misconception is the failing to reject $H_0$ suggests that $H_0$ is true.
• Rejecting $H_0$ means there is a small chance that the obtained result would have occurred if $H_0$ were in fact true.
• Failing to reject $H_0$ suggests that either $H_0$ is true (reflecting a correct decision) or $H_0$ is false but we do not have enough statistical power to reject $H_0$. Thus, failing to reject $H_0$ is inconclusive!
Some Notes

Use Confidence Intervals

• Failure to reject $H_0$ should not suggest that we cannot draw meaningful conclusions or research is without value
• We cannot accept $H_0$, but confidence intervals can tell us whether or not the differences between treatments would likely be meaningful
• Confidence intervals provide information on the magnitude of the effect size and how estimates would vary in other samples/studies

Some Notes

How to Increase the Power of Statistical Tests

• Ideally, the power of test is 0.80, but typically 0.40 – 0.60.
• Low statistical power ($< 0.50$) indicates there is a high probability that the research design was not sensitive enough to detect the treatment effects
• Increasing the power of test by increasing (1) sample size, (2) significance level ($\alpha$) (use 0.10 rather than 0.05), and (3) effect size of treatments.

Some Notes

Statistical Significance vs. Practical Significance:

• Statistical significance is a statistical calculation that reflects how far a given association exceeds those which would be expected by chance, while practical significance refers to the potential for the research findings to make a real and measurable difference in practice.
• A non-significant result does not necessarily prove null hypothesis.
• A “statistically significant” finding is one in which researchers are 95% confident that an association exists. However, statistically significant finding does not necessarily prove a cause-effect association or that the results are practically significant.

Statistics Courses Offered at SUNY-ESF

APM 391 – Introductory Statistics – 3 credits. By Dr. D. Kiernan
FOR 323 – Forest Biometrics – 3 credits. By Dr. L. Zhang
APM 510 – Introductory Statistics – 3 credits. By Dr. E. Bevilacqua
APM 620 – Experimental Design – 3 credits. By Dr. S. Stehman
APM 625 – Introduction to Sampling Techniques – 3 credits. By Dr. S. Stehman
APM 630 – Regression Analysis – 3 credits. By Dr. L. Zhang
APM 635 – Multivariate Statistical Methods – 3 credits. By Dr. L. Zhang
APM 645 – Nonparametric Statistics and Categorical Data Analysis – 3 credits. By Dr. L. Zhang
Thank you!