From the Expert: Statistical Practice

From the Expert: Statistical Practice

Communicating with Statistics

Statistics do not end with calculations. The results and conclusions need to communicate back to the audiences. The audiences can be very diverse such as your instructors, co-workers, your boss, professional researchers or lay people. Understanding your audiences, and choosing the appropriated writing or presentations can be very challenge for the successful communications. In general practices, conclusions should be stated first, then follow with the supporting statistics. This is known as BULF or Bottom Line Up Front. Level of the precision number should correspond to the purposes such as rounding up the number using 2 digits decimal instead of 8 digits decimal. For very small or very large numbers, scientific notations may be more appropriated. Do not forget to specify the source of data, the collection methods, and details about data sampling. Whenever the percentage is used, the base should be included to avoid perception misunderstanding such as 10% increase in the house prices of 200,000 and 20,000,000 are big differences. The general audiences may not familiar or have little understandings in statistical analysis, so it is acceptable to re-iterate the same point in different ways.

General guidelines to communicate with different audiences:

For general public:

Writing articles for magazines or newspapers should avoid all technical jargons when conveying the results. The language should be simple or everyday language without too many details in methodology. However, the results and their practical implications should be highlighted.

For professional journal:

Typically, certain formats and styles need to be followed so it is helpful to be familiar with the journal you plan to submit. Choosing the appropriate journal for the research topic should be done carefully for the chance to get published.

For the workplace:

This can be very challenging since there are mixed audiences and the stakes can be high since the analysis results may affect the company directions. In writing, two types of reports should be prepared: an executive summary and a detailed report. An executive summary is usually for upper level managements, who have neither time nor technical background. Therefore, the summary should be a short finding with plain language. In addition, a set of choices and/or recommendations must also be included for decision makers. The detailed report, on the other hand, includes necessary finding details in longer documents. For presentation, two level presentations: an executive summary and

more detailed presentation can be prepared alongside. This means that the presentation is an executive version with additional detail as side notes or footnotes. Technical handouts can be given to technical or interested audiences. Finally, be prepared to explain why particular statistical tests are chosen and the meaning of the results.

Misuse of statistics

Two major reasons of misusing statistics are ignorance and intention. The ignorance occurs when tests, and methodologies are not appropriate for data analysis, especially, for complex procedures such as multivariate analysis. The intentional misuse happens when there are attempts to hide, or mislead the results, for example, ignore the inference assumptions to make the results valid, and in-correct scale on graphs.

Common Problems Errors in Statistical Analysis

- **Misleading graphs** – scales on a graph can distort the data perception. Different scales affect the appearance in a graph. For instance, a plot with bar graph comparisons. A smaller scale makes the differences between bar graphs appear to be larger. Whereas, a bigger scale makes the differences appears to be smaller or minimal.

- **Drawing the wrong conclusion from a confidence interval** - For instance, 95% of confidence interval is (a,b) means that 95 out of 100 cases, the confidence interval will contain the true population mean NOT the probability of the population mean being in the interval (a,b) is 95%.

- **Misinterpreting the results in a hypothesis test** - For instance,
  Mr. A is innocent
  : Mr. A is guilty

  If the null hypothesis is rejected, then Mr. A is guilty. If the null hypothesis is not rejected, it simply means that is not enough evidence to prove that Mr. A is guilty. It does not necessarily means that Mr. A is innocent.

- **Be careful of using coefficient of Determination**
  Typically, indicates the fit of the model. When new independent variables are added to the model, the value of increases automatically even these independent variables do not contribute to additional explanation power to the regression.

- **Normality Distribution**
  Many testing techniques assume that data is normally distributed. This assumption should be verified before the test. Otherwise, the conclusions may be not accurate (e.g. underestimate the risk)

- **Correlation does not imply causality**
  Correlation indicates the tendency of two variables to move in the same or opposite directions. It does not necessary mean one causes the other. Correlation does not imply causality. The classic example is from William Stanley Jevons, an 19th century British economist, who utilized statistical methods in business cycles measurement. He observed that sunspots went through a cycle that lasted about 11 years, and the business cycles last for less than 11 years. Therefore, he concluded his study that sunspots were responsible for the business cycle. With the reason, sunspots have affects on the weather patterns, and as a result, these impact the economy)

- **Drawing conclusions when the assumptions are not satisfied**
  In regression, for example, the assumptions need to be valid, or else, this can lead to the wrong conclusion (e.g. found that the statistics is significant even it is not significant)

- **Including correlated variables in multiple linear regression**
  This is also known as multicollinearity, where two or more independent variables exhibit highly correlated with each one another. Correlated variables can cause high standard errors. Thus, the statistical test is significant even it is not significant (e.g. variables would be removed from the model instead of remaining in the model)

- **Too much confidence in forecasts**

Errors such as bias error (e.g. forecast more or less than actual values) and random error (e.g. unpredicted events such as earthquakes, drop in oil prices) can impact the forecasts accuracy.

- Using wrong distribution
  Conclusions that are drawn from the wrong distributions are less accurate. For instance, use normal distribution instead of uniform distribution.

Other guidelines:

- Check for the consistency of the hypothesis and the results. Has the hypothesis been changed after the results are known?
- Did the investigator try to manipulate the results using a number of tests to make the test significant?
- The simple test that is suitable for your purpose should be employed first before the more complicated one.
- To avoid publication bias, non-significant test results should also be reported in the study. This is also known as file drawer problem, where only significant test results get published but the not significant test stays in the drawer or unpublished.

Ultimately, statistics plays a major role in many aspects of our life and society on economy, medicine, marketing, manufacturing, scientific research, etc. The impact of statistics can be very broad and high stakes. For instance, statistics in medical science or biomedical research could affect personal health, live, and death. The American Statistical Association creates general guidelines for statistics practitioners to communicate with ethical decisions here: http://www.amstat.org/committees/ethics/index.html

References:

American Statistical Association (2015), Ethical Guidelines for Statistical Practice,
Retrieved from: http://www.amstat.org/committees/ethics/index.html

Anderson, A. (2013). Business Statistics for Dummies. Wiley.

Boslaugh, S. (2013) Statistics in a Nutshell. O'Reilly Media, Inc.,

Smith, K.M. (2015) Common Mistakes in Using Statistics: Spotting Them and Avoiding them. Retrieved from: https://www.ma.utexas.edu/users/mks/statmistakes/Resources.html

Reflect in ePortfolio Download Print

Activity Details

Task: View this topic