Using Statistics in Research Papers

East Los Angeles College Library ● elac.edu/library

What is a statistic?

Simply put, a statistic is a number that describes a group. There are two main categories of statistics: descriptive and inferential.

- Descriptive statistics describe only the group to which they belong. For example, "This season, our local hockey team scored a mean of 2.3 goals per game."
- Inferential statistics can be extrapolated to draw conclusions about a larger group than the one they describe. For example, "Our research suggests that 33% of American women prefer green cars." In this case, researchers have, of course, not asked every woman in America her color preference; they have polled a smaller, representative group.

Key Terms

Data point: one particular number or item from a data set
Data set: a group of information
Distribution: describes how the points in a data set are clustered, for example, whether they're spread apart equally or clustered toward the middle or edges; two data sets may have the same mean or median but have wildly different distributions
Mean: the sum of all numbers in a set divided by the number of data points in the set; this is the number most commonly referred to as the average
Median: the middle data point in a set when all points are arranged from least to greatest
Mode: the number that occurs most frequently in a data set
Null hypothesis: the hypothesis that there is no significant difference between specified populations or no correlation among variables and any observed difference is due to sampling or experimental error; generally, this is the opposite of the hypothesis the researcher wants to prove
P-Value: an indicator of statistical significance, the p-value indicates the probability of obtaining an effect at least as extreme as the one in the sample data, assuming the truth of the null hypothesis; a low P-value suggests that it's unlikely the results would have occurred if the null hypothesis were true
Population: all of the members contained within a group; in inferential statistics, this is the group you want your results to generalize to
Sample: all of the members contained within the group actually studied; a subset of the population. The more representative the sample is of the population in size and demographics, the more accurate any inferences drawn from collected data will be.
Standard Deviation: a measure of variability; the average distance that individual data points are from the mean. Data sets with distributions clustered around the mean (for example, a bell curve) have lower standard deviations than data sets that are more spread out with several points at either extreme.
T-test: a common statistical test used to compare two groups; typically, the difference of two means divided by a measure of variability taking into account the number of units in the sample

Inferential Statistics Basics

It is important to understand how the sample being studied was drawn from the population. Inferential statistics are most accurate when the sample is drawn at random from the population; given a large enough sample, drawing at random ensures a fair and representative sample of a population.

Be wary of statistics drawn from small sample sizes. Professional statisticians can and often do interpret results correctly from small sample sizes, but novice statisticians often incorrectly interpret results.

Even with a large sample size, a mean isn't meaningful without a measure of variability. If you're told that most people recovered from a cold in an average of 5 days, it's possible that everyone recovered in 4-6 days or, alternatively, that half of people recovered within 1 day and the other half took 10 days.

A low p-value is a good sign. This indicates that it's unlikely the results were obtained by pure chance.

Writing with Statistics

DO:

- Know your source. Authors with an agenda may knowingly misinterpret results. Use peer-reviewed journal articles when possible.
- Any time you use a statistic, credit the source. If you calculated the statistic yourself but used data obtained by others, be sure to cite the source of the original data.
- When using other researchers' statistics, evaluate and interpret them yourself rather than assuming the interpretation given is the only correct one. Sometimes authors use good statistics but reach bad conclusions.
- Consider your audience. If they might not understand the statistics you're using, be sure to explain your procedures in detail. It's better to include too much information than too little.
- In general, you should always 'translate' your statistics into some understandable form for your reader. Say what you mean plainly first, then provide the statistical evidence afterward.
- Present as much information as needed so that your reader can make their own interpretation of your data. Give your reader enough information that they can reconstruct your argument from your statistics. Without sufficient information, your reader may suspect that you are being deceptive and intentionally leaving out details that contradict your argument, which may damage your credibility.
- Provide all of the context needed to give your reader the full picture.
- When possible, include a measure of variability (usually a standard deviation).
- Remember that "average" can be used to refer to mean, median, or mode. Be clear about which measurement you are presenting.
- If you have a lot of statistics to report, consider presenting them using graphs or tables. You can then highlight the statistics of greatest interest in your writing.
- Be clear about which population(s) your statistic is meant to generalize to.

DON'T:

- Don't confuse correlation with causation**.** Two events occurring together doesn't always indicate that one is caused by the other. This could be a coincidence, or both could be caused by a different factor. For example, it's often true that the more firefighters are sent to a fire, the more damage the fire does. The firefighters aren't making the fires worse; instead, both the number of firefighters deployed and the damage caused are the result of a third variable: the size of the fire.
- Never calculate or use a statistical procedure you don't fully understand.
- Never attempt to interpret the results of a statistical procedure you don't fully understand.
- Don't feel the need to include information that is irrelevant to your argument. If the median, for example, isn't a meaningful representation of your data set for the purposes of your argument, it's okay to leave that out.
- Don't include every data point from your set. One purpose of statistics is to condense large amounts of information into more manageable pieces. Presenting the entire data set defeats this purpose.
- Don't imply that your sample generalizes to everyone if your statistic was calculated from a specific population (e.g. only college students).
- Don't use percentages without providing the base numbers, and don't manipulate the scale of a chart to exclude or distort relevant data.

Get More Help

In person or by phone:

- **ELAC Math Tutoring Center (G5-LL0099):** (323) 415-4191
- **ELAC Reading and Writing Center (E3-220):** (323) 415-4147

Online:

- **Purdue OWL:** http://bit.ly/owlstats
- **The Writing Center at UNC Chapel Hill:** http://bit.ly/writingstatistics

Adapted from the Online Writing Lab at Purdue University: http://owl.english.purdue.edu