Statistics

After reading this article, you should be able to use statistics to make your argument as effective as possible.

Why Do Statistics Matter?

Numbers and statistics are powerful pieces of evidence that can effectively strengthen any argument. As simple and straightforward as these numbers promise to be, statistics can create more problems than they solve if they aren't used correctly.

The average reader does not know how to properly evaluate and interpret the statistics s/he reads. Many people think that statistics can speak for themselves, but numbers are as ambiguous as words and need just as much explanation.

Using statistics is similar to using direct quotes; they are expected to do all the work or they are not presented as a piece of evidence that requires interpretation. If your statistics are confusing or misleading, then the reader has to guess what you mean. The only way to avoid off-the-wall interpretations is to be as clear as possible.

Reading Statistics

Numbers are powerful. This makes statistics persuasive pieces of evidence, but they can also be intimidating. Too often, we don't wonder about the truthfulness behind a set of data, yet being an effective reader means thinking critically and asking questions. Below are a set of hard questions to ask of the numbers you find.

1. Does Your Evidence Come From Reliable Sources?

This is an important question with any evidence you use in a paper. There are many ways that statistics can be played with or misrepresented to produce a desired outcome, so it's important to take your statistics from reliable sources (See our handout on evaluating print sources). While reliable sources aren't perfect, they're probably less likely to use deceptive practices. Try to question the truthfulness of statistics instead of trusting that they are accurate.

2. What Is the Data's Background?

Unfortunately, data and statistics don't fall from heaven fully formed; they're always the product of research, so it's important to know where they come from. For example, if the statistics come from a survey or poll, ask:

- **Who asked the questions?**
- **What questions were asked?**
- **Who interpreted the data?**
- **What issue prompted the survey/poll?**
- **What policy/procedure potentially hinges on the results of the poll?**
- **Who stands to gain from particular interpretations of the data?**

These questions will help you discover possible biases or weaknesses in a data set. The goal of this exercise is not to find "pure, objective" data, but to make any biases explicit in order to accurately interpret the evidence.

3. Are All Data Reported?

No, not in most cases. It's helpful to ask whether all data have been presented in context. This is much more complicated when you consider the bigger issue (whether the text or source presents enough evidence for you to draw your own conclusion). A reliable source should not exclude data that contradicts or weakens its argument.

Here's an example using the news: In winter, newscasters might warn you to stay off treacherous roads because 25 accidents have already been reported in a single day. While this number sounds high, some studies have found that the number of accidents actually goes down on days with severe weather. One possible explanation for this is that the number of accidents is less than on an "average" day because there are fewer people on the road, given the dangerous conditions. **The critical lesson here is that even when the general interpretation is "accurate," the data may not actually be evidence for the particular interpretation.** This means you have no way to verify if the interpretation is in fact correct.

There's generally a comparison implied in the use of statistics. If you're trying to make a valid comparison without having all the facts, look to another source (or sources) to find all the data you need.

4. Have the Data Been Interpreted Correctly?

When an author gives you his or her statistics, interpret them yourself. It's useful to read and understand an author's argument, but remember that it's an interpretation, not the final word on the matter. Sometimes authors can use perfectly good statistics and come up with perfectly bad interpretations. Here are two common mistakes to watch out for:

- **Confusing correlation with causation.** Just because two things **vary** together does not mean that one of them is **causing** the other. It could be nothing more than a coincidence, or both could be caused by a third factor (This is called **spurious**).
    - **Example:** according to a study, the more firefighters were sent to put out a fire, the more damage the fire did. Yikes! Before we start shutting down fire stations, though, it might be useful to consider alternative explanations. This seemingly contradictory finding can be explained by pointing to a third factor that causes both: the size of the fire. The lesson here? **Correlation does not equal causation.** It's important to show that two variables co-vary and also think about the causal mechanism.
- **Ignoring the margin of error.** When survey results are reported, they frequently include a margin of error. You might see this written as "a margin of error of plus or minus 5 percentage points." Surveys are normally generated from samples of a larger population, so they're never exact. If I say that the number of UTEP students who find it difficult to use statistics in their writing is 60%, plus or minus 4%, that means (assuming the normal confidence interval of 95%) that with 95% certainty we can say that the actual number is between 56% and 64%.
    - If I introduce this handout to UTEP students and a new poll finds that only 56% (plus or minus 3%) are having difficulty with statistics, I could go to the Writing Center director and ask for a raise, since I have made a significant contribution to students' writing skills. However, he would point out that a) this may be a spurious relationship and b) the actual change is not significant because it falls within the margin of error for the original results. The lesson

here? **Margins of error matter**, so you cannot just compare simple percentages.

Keep in mind that the source you are looking at may not be the original source of your data. If an essay quotes a number of statistics to support its argument, that author is likely using someone else's data. It's helpful to consider your source and the author's sources as well.

Writing statistics

As you write with statistics, remember your own experience as a reader. Be as clear and straightforward as you can be with your numbers. Avoid changing the numbers a little bit to help your argument because your reader will want to know the answers to the questions that we discussed above. Below is a list of common pitfalls in the world of statistics, along with suggestions for avoiding them.

1. The Mistake of the "Average" Writer

Nobody wants to be average. Similarly, nobody wants to see the word "average" in a piece of writing because nobody knows exactly what it means. There are three different definitions of "average" in statistics, so your reader has a 33.3% chance of correctly guessing which one you mean.

For the following definitions, please refer to this set of numbers: **5, 5, 5, 8, 12, 14, 21, 33, 38.**

- **Mean** (arithmetic mean): This is the weighted average—a total of all numbers included divided by the quantity of numbers represented. Thus the mean of the above set of numbers is 5+5+5+8+12+14+21+33+38, all divided by 9, which

equals 15.644444444444 (That's a lot of numbers after the decimal! Precision is a good thing, but too much of it does not necessarily make your argument stronger. Consider the reasonable amount of precision based on your input and round accordingly. In this case, 15.6 should do the trick.)

- **Median:** Depending on whether you have an odd or even set of numbers, the median is either a) the number midway through an odd set of numbers or b) a value halfway between the two middle numbers in an even set. For the above set (an odd set of 9 numbers), the median is 12. (5, 5, 5, 8 < 12 < 14, 21, 33, 38)
- **Mode:** The mode is the number or value that occurs most frequently in a series. If two or more values occur with the same frequency, then take the mean of the values. For our set, the mode would be 5, since it occurs 3 times, whereas all other numbers occur only once.

Numbers can vary considerably, as can their significance; therefore, you should always inform your reader which average you're using. Otherwise, you'll confuse your readers (and likely receive a bad grade).

2. Match Your Facts and Questions

Make sure that your statistics apply to the point/argument you're making. For example, imagine your professor hands back the midterm and the grades are distributed as follows:

Grade	# Received
100	4

98 5

95 2

63 4

58 6

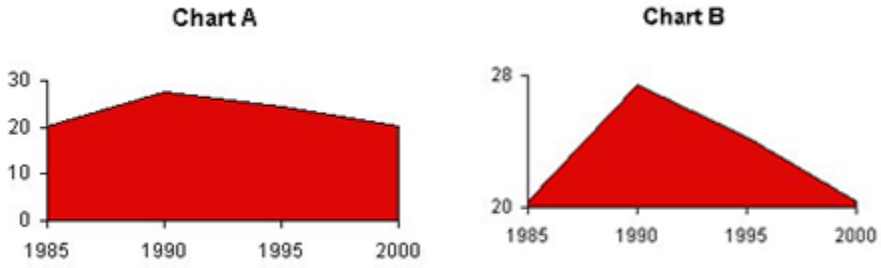
The professor felt that the test must have been too easy, because the average (**median**) grade was a 95. When a colleague asked how the midterm grades came out, the professor answered (knowing that her classes were gaining a reputation for being "too easy") that the average (**mean**) grade was an 80. When your parents ask why you did poorly on the midterm, you answer, "Don't worry about my 63. It is not as bad as it sounds. The average (**mode**) grade was a 58."

As you can see from the examples above, selecting the appropriate facts or statistics will help your argument immensely. The best way to maintain precision is to specify which of the three forms of "average" you are using.

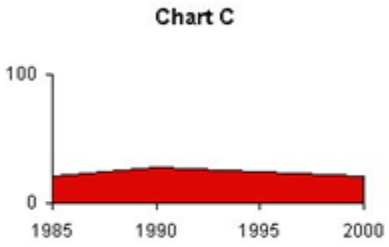
3. Show the Entire Picture

Sometimes, you may misrepresent your evidence by accident. Other times, however, misrepresentation may be slightly less innocent. This can be achieved by presenting visual aids (charts/graphs) in numerous ways; either the range can be shortened (to cut out data points which do not fit, e.g., starting a time series too late or ending it too soon), or the scale can be manipulated so that small changes look big and vice versa. In your own writing, be sure not to

shape the representation so that it "best supports" your argument, or adjust the proportions vertically or horizontally.

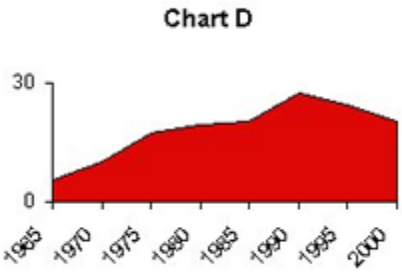


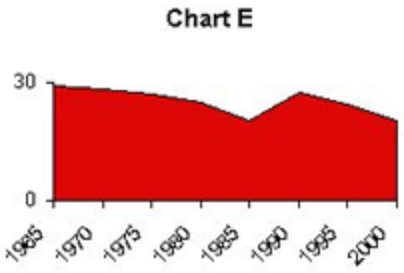
Charts A, B, and C all use the same data points, but the stories they tell are quite different. Chart A shows a mild increase, followed by a slow decline. Chart B, on the other hand, reveals a steep jump, with a sharp drop-off immediately following. Chart C demonstrates that there was virtually no change over time. These variations are a product of changing the scale of the chart. One way to alleviate this problem is to supplement the chart by using the actual numbers in your text.



Another point of concern can be seen in Charts D and E. Both use the same data as charts A, B, and C for the years 1985-2000, but additional time points, using two hypothetical sets of data, have been added back to 1965. Given the different trends leading up to 1985, consider how

the significance of recent events can change. In Chart D, the downward trend from 1990 to 2000 is going against a long-term upward trend, whereas in Chart E, it is merely the continuation of a larger downward trend after a brief upward turn.





One of the difficulties with visual aids is that there's no rule about how much to include or what to exclude. Be sure to present your visual aids so that your readers can draw their own conclusions from the facts and verify your assertions. If what you have cut out could affect the reader's interpretation of your data, then consider keeping it in.

4. Give Bases of All Percentages

Percentages are always derived from a specific base, so they're meaningless until they're associated with a base. If I tell you that you'll be 23% more persuasive as a writer after reading this handout, it's not a meaningful assertion because you have no idea what it's based on (23% more persuasive than what?).

- **For example:** Suppose we have two cities, Springfield and Shelbyville. In Springfield, the murder rate has gone up 75%, while in Shelbyville, the rate has only increased by 10%. Which city is having a bigger murder problem? Well, that's obvious, right? It has to be Springfield.

- **In order to know which city has a worse problem, we have to look at the actual numbers.** If I told you that Springfield had 4 murders last year and 7 this year, and Shelbyville had 30 murders last year and 33 murders this year, would you change your answer?
- **Unfortunately, we still don't have all the facts.** We have to compare the two based on equivalent standards, examining the per capita rate (often given in rates per 100,000 people per year). If Springfield has 700 residents while Shelbyville has 3.3 million, then Springfield has a murder rate of 1,000 per 100,000 people, and Shelbyville's rate is merely 1 per 100,000. Gadzooks! The residents of Springfield are dropping like flies.

Percentages are really no different from any other form of statistics: they gain their meaning only through their context. Percentages should be presented in context so that readers can draw their own conclusions. If your statistics really do support your point, then you should have no problem with revealing the larger context that frames them.

Important Questions to Ask (and Answer) About Statistics

- **Is the question relevant?**
- **Do the data come from reliable sources?**
- **Margin of error/confidence interval—when is a change really a change?**
- **Are all data reported, or just the best/worst?**
- **Are the data presented in context?**
- **Have the data been interpreted correctly?**
- **Does the author confuse correlation with causation?**

Conclusion

Numbers are powerful; therefore, they come with great responsibility. Now you can use this knowledge to manipulate your numbers to your advantage, or you can use statistics to make accurate and fair arguments. The choice is yours.

This information originally appeared in a handout from the University of North Carolina at Chapel Hill Writing Center.