Statistical Significance

A finding or an observation is said to be **statistically significant** if it is unlikely to have occurred by chance. That is, if we *expect* to get a certain sample result and don't, it could be because of sampling variability (in repeated sampling from the same population, we will get different sample results even though the population value is fixed), or it could be because the sample came from a different population than we thought. If the result is far enough from expected that we think something other than chance is operating, then the result is statistically significant.

example: Todd claims that he can throw a football 50 yards. If he throws the ball 50 times and averages 49.5 yards, we have no reason to doubt his claim. If he only averages 30 yards, the finding is *statistically significant* in that he is unlikely to have a sample average this low if his claim was true.

In the above example, most people would agree that 49.5 was consistent with Todd's claim (that is, it was a likely average if the true value is 50) and that 30 is inconsistent with the claim (it is *statistically significant*). It's a bit more complicated to decide where between 30 and 49.5 the cutoff is between "likely" and "unlikely."

There are some general agreements about how unlikely a finding needs to be in order to be significant. Typical significance levels, symbolized by the Greek letter a, are probabilities of 0.1, 0.5, and 0.01. If a finding has a lower probability of occurring than the significance level, then the finding is statistically significant.

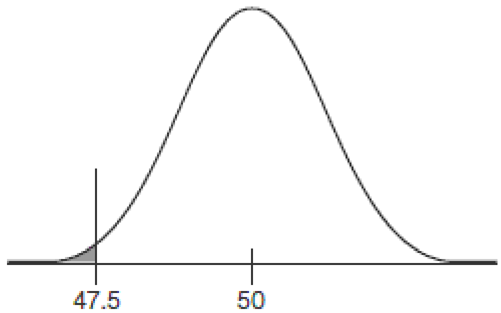
example: The school statistics teacher determined that the probability that Todd would only average 30 yards per throw if he really could throw 50 yards is 0.002. This value is so low that it seems unlikely to have occurred by chance, and so we say that the finding is significant. It is lower than any of the commonly accepted significance levels.

P-Value

We said that a finding is statistically significant, or significant, if it is unlikely to have occurred by chance. *P*-value is what tells us just how unlikely a finding actually is. The **P-value** is the probability of getting a finding (statistic) as extreme, or more extreme, as we obtained by chance alone. This requires that we have some expectation about what we ought to get. In other words, the *P*-value is the probability of getting a finding at least as far removed from expected as we got. A decision about significance can then be made by comparing the obtained P-value with a stated value of α

example: Suppose it turns out that Todd's 50 throws are approximately normally distributed with mean 47.5 yards and standard deviation 8 yards. His claim is that he can average 50 yards per throw. What is the probability of getting a finding this far below expected by chance alone (that is, what is the P-value) if his true average is 50 yards (assume the population standard deviation is 8 yards)? Is this finding significant at $\alpha = 0.05$? At $\alpha = 0.01$?

solution: We are assuming the population is normally distributed with mean 50 and standard deviation 8. The situation is pictured below:



$$P(\bar{x} < 47.5) = P\left(z < \frac{47.5 - 50}{8/\sqrt{50}} = -2.21 \right) = 0.014.$$

This is the *P*-value: it's the probability of getting a finding as far below 50 as we did by chance alone. This finding is significant at the 0.05 level but not (quite) at the 0.01 level.