Statistical Significance 101

Educational Briefing

What is statistical significance?

- The result of a study is considered statistically significant when it is highly unlikely to have occurred by chance or due to sampling error (the sample being different from the general population).
- In statistics terminology, a result is statistically significant when the probability (measured by a **p-value**) of obtaining that result versus the **null hypothesis** (the result you would expect if an intervention had no impact) is below a certain **significance threshold**. In most studies, this threshold is .05, meaning a researcher will say a result is statistically significant if there is less than a 5% chance that the result occurred by accident.
- P-values are calculated through a statistical method which takes into account how many data points are included (the **sample size**) and how much variation there is in these data points (the **standard deviation**).

What steps are required to determine if a result is statistically significant?

1. Researchers first determine the study's null hypothesis or expected results. For example, perhaps a researcher is testing whether men or women are more likely to have a heart attack and thinks that women are more likely. The null hypothesis would be that men and women have equal risk of a heart attack, as this is the 'normal' or 'expected' result. Therefore, if the null hypothesis was true, you would expect the same incidence of heart attack in men and women.
2. Then researchers conduct the study and track the outcomes they observe.
3. They then compare the observed results of the study to the expected results to determine if the null hypothesis was true. This is done through a **chi-squared** calculation and results in a p-value (the probability of the observed result occurring). *(Note: There are other methods of doing this p-value calculation depending on the variables being studied and the study design).*
4. To determine if the result was significant, researchers determine if this p-value is greater or smaller than the level of significance they decided upon at the start of the study (conventionally .05). If the p-value was larger, they accept the null hypothesis (meaning the results could have occurred by chance). If it is lower, they reject the null hypothesis and claim that the results are statistically significant.

What is an example of this significance calculation?

- Researchers want to figure out if running for 30 minutes a day has an impact on one's risk of developing diabetes. To begin, they state the null hypothesis or normal result–that those who run 30 minutes per day have the same risk for diabetes as those who don't run 30 minutes a day.
- They then do a study where the intervention group runs 30 minutes daily while the control group remains sedentary, and track the incidence of diabetes in both groups over the next 3 years.
- Upon looking at the results, they find that 25% of those in the intervention group (the runners) developed diabetes while 32% of those in the control group developed diabetes.
- To calculate if this difference is significant, they do a chi-squared calculation to see if the expected result (that both groups should have a 32% risk of developing diabetes, as seen in the control group) with the observed result (that the running group had a 7% lower chance.)
- After doing this calculation, the result is a p-value of .12 (or a 12% chance of the outcome occurring independent of the intervention). Since this is above their significance threshold (.05), they accept the null hypothesis and claim that they cannot say that running has a statistically significant impact on diabetes risk.

How is statistical significance used in evidence-based medicine?

- Statistical significance is usually considered the basis of determining if an intervention has an impact on a result and therefore if it should be adopted as a practice in medicine. However, this does not mean that studies that don't find significance are not useful–rather knowing that an intervention *isn't* always tied to a certain outcome is also useful in informing clinical practice or health interventions.
- Generally, however, studies are about three times more likely to be published if they find a statistically significant result, which can lead to a "publication bias" or "file drawer problem" whereby results that are not statistically significant, but important because they are *not* significant, are not published and therefore often left out of systematic reviews or meta-analyses.

September 2018

Source: "Publication bias in the social sciences: Unlocking the file drawer", *Science,* 2014*;* Advisory Board insight and analysis.