Print

p-Value: Effect Size and Sample Size Influence

If you flip a coin 100 times and observe 50 heads, you wouldn't doubt that the coin is fair. But you might be skeptical if you observe 40 or 60 heads. You'd be more skeptical if you observe 37 or 63 heads, and you'd be highly skeptical if you observe 15 or 85 heads. As the difference between the number of heads and tails increases, you have more evidence that the coin is not fair.

Statisticians refer to the difference between the observed statistic and the hypothesized value as the effect size. The null hypothesis of a fair coin suggests 50% heads and 50% tails. If the coin was actually weighted to give 55% heads, the effect size would be 5%.

A p-value measures the probability of observing a value as extreme as the one observed or more extreme, assuming the null hypothesis is true. Suppose you flip the coin 100 times, and you observe 55 heads and 45 tails. The difference of 10 is associated with a p-value of .3682. p-values this large are often seen in experiments with a fair coin.

As the difference between heads and tails gets larger (for example, 20, 26, or as high as 70), the corresponding p-values get smaller. You would rarely see a small p-value (for example, less than .0001) with a fair coin. In the case of 15 heads and 85 tails, you have evidence that the coin is not fair (p-value < .0001). So the p-value is used to determine statistical significance. It helps you assess whether you should reject the null hypothesis.

A p-value is not only affected by the effect size (in this case, the observed proportion of heads). It's also affected by the sample size (in this case, the number of coin flips). For a fair coin, you'd expect 50% of the flips to be heads. What if you get 40% heads instead of the 50% you expect? Is it a fair coin? Let's say that you flip the coin 10 times and observe 40%, or 4 heads. What if you flip the coin 400 times and you observe 40% heads?

The evidence becomes stronger and the p-values become smaller as the number of trials increases. As you saw when we talked about confidence intervals, the variability around the mean estimate gets smaller as the sample size gets larger. For larger sample sizes, you can measure means more precisely. Therefore, 40% heads out of 400 flips makes you more confident that this was not just a chance difference, when compared to 40% heads out of only 10 flips (p-value of 0.0569 versus a p-value of .0120).

The smaller p-values reflect this confidence. The p-value here, less than .0001 for 160 heads and 240 tails, assesses the probablility that this difference from 50% occurred purely by chance. Remember, as you saw earlier, you'd rarely see a p-value less than .0001 with a fair coin.

Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression

Close