Statistical Thinking for the 21st Century

Copyright 2020 Russell A. Poldrack

Chapter 10

Quantifying effects

10.2 Effect sizes

> "Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude – **not just, does a treatment affect people, but how much does it affect them.**" Gene Glass (REF)

As the above quote illustrates, we want to move beyond statistical significance to think about practical significance. To discuss practical significance, we need a standard way to describe the size of an effect in terms of the actual data, which we refer to as an *effect size*. In this section we will introduce one way that effect size can be calculated.

An effect size is a standardized measurement that compares the size of some statistical effect to a reference quantity, such as the variability of the statistic. In some fields of science and engineering, this idea is referred to as a "signal to noise ratio." There are many different ways that the effect size can be quantified, which depend on the nature of the data.

10.2.1 Cohen's *d*

One of the most common measures of effect size is known as Cohen's *d*, named after the statistician Jacob Cohen (who is most famous for his 1994 paper titled "The Earth Is Round (p < .05)"). Cohen's *d* is used to quantify the difference between two means, in terms of their standard deviation.

The formula for Cohen's *d* is shown below.

Table 10.1: Interpretation of Cohen's *d*

d	Interpretation
0.000 - 0.199	neglibible
0.200 - 0.499	small
0.500 - 0.799	medium
0.800 - 1.399	large
≥ 1.400	very large

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the means of the two groups, and s is the pooled standard deviation (which is a combination of the standard deviations for the two samples, weighted by their sample sizes):

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where n_1 and n_2 are the sample sizes and $s^2{}_1$ and $s^2{}_2$ are the standard deviations for the two groups respectively.

Table 10.1 provides a commonly used scale for interpreting the size of an effect in terms of Cohen's *d*.

It can be useful to look at some commonly understood effects to help understand these interpretations. For example, the effect size for gender differences in height is very large by reference to our table above. We can also see this by looking at the distributions of male and female heights in our sample.

Figure 10.2 shows that the two distributions are quite well separated, though still overlapping, highlighting the fact that even when there is a large effect size for the difference between two groups, there will be individuals from each group that are more like the other group.

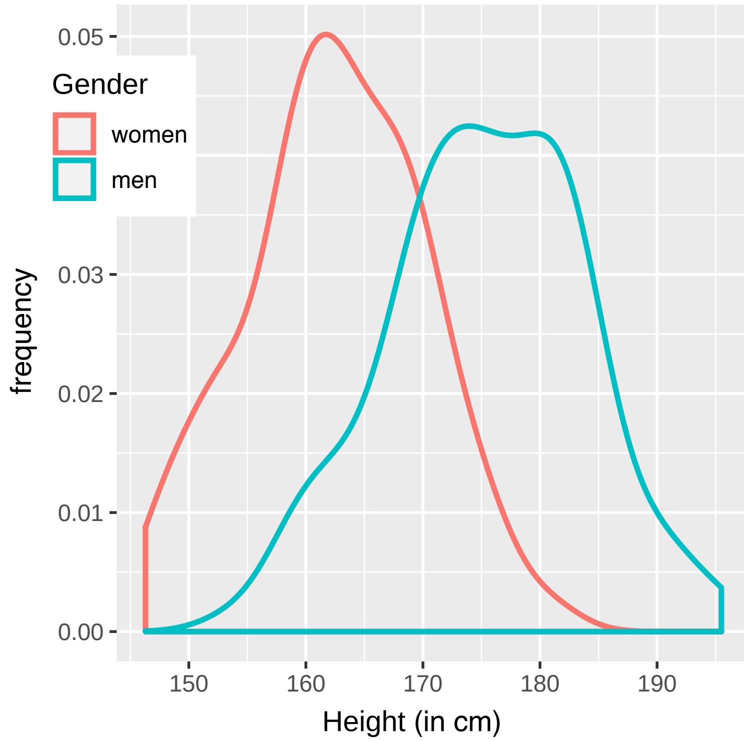

Figure 10.2: Smoothed histogram plots for men and women heights in the NHANES dataset, showing clearly distinct but also clearly overlapping distributions.

It is also worth noting that we rarely encounter effects of this magnitude in science, in part because they are such obvious effects that we don't need scientific research to find them. In fact, very large reported effects in scientific research often reflect the use of questionable research practices rather than truly huge effects in nature.

It is also worth noting that even for such a huge effect as we see between women's and men's heights, the two distributions still overlap - there will be some women who are taller than the average man, and vice versa.

For most interesting scientific effects, the degree of overlap is much greater, so we shouldn't immediately jump to strong conclusions about different populations based on even a large effect size.