OpenIntro online supplement

This material is an online resource of *OpenIntro Statistics*, a textbook available as a free PDF at openintro.org and for under $10 as a paperback on amazon.com. This document is licensed to you under a Creative Commons license, and you are welcome to share it with others. For additional details on the license this document is under, see www.openintro.org/rights.php.
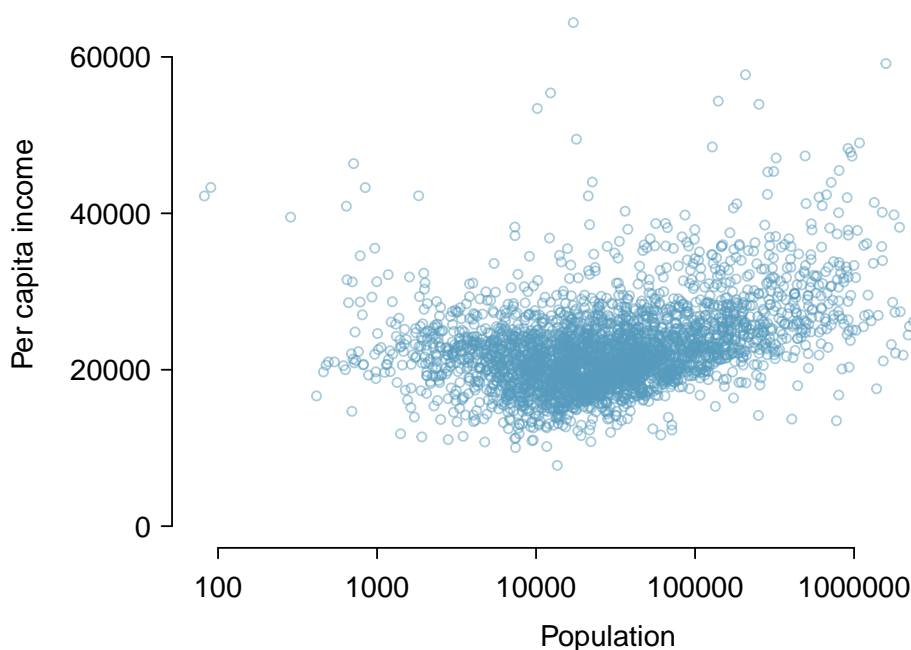
Figure 0.1: Per capita income against population size in 3,143 US counties.

Weighted mean

The weighted mean is the same as the mean, except that it is influenced more by some observations than others. We assign weights to observations as a sort of way of describing its relative importance. In many applications, there are natural choices for weights. For example, in the `county` data set, population is a natural weighting factor.

Given the number of people in each county of the United States and also the average income per person in each county, how should we compute the average income per person for residents in the United States?

An initial guess might ignore the population data. This would yield a mean income of about \$22,505. However, this is not the mean income of the people in the United States, which is actually several thousand dollars more. To see why, consider Figure 1, which shows the income and population of all 3,143 counties. Notice first that the scale on the horizontal axis for population shows a special scale (called the log-scale). Many of the biggest counties have higher incomes than the smaller counties. This is why our original mean calculation gives a wrong answer.

Let's think again about the regular problem: we want the mean income per person in the United States. We need exactly two numbers to complete this calculation: the total income of all people in the US and the number of people in the US:

$$\frac{\text{Total income of all people in the US}}{\text{Number of people in the US}}$$

We're not given these numbers, but the data provided is sufficient to calculate each of the numbers. Starting with the denominator (the bottom), just add up the population in all the counties: 308,745,538.

The numerator (the top of the fraction) is a little tougher, but not much. If we had the total income of all people each county, then we could add up all those values to get the total income. To get the income for any given county, we can multiple the average income of the county with the number of people in the county. This just relies on a rearrangement of the equation defining the average income for the county:

$$\text{Ave income} = \frac{\text{County income}}{\text{Num. people in county}}$$

$$\Downarrow$$

$$(\text{Ave income})(\text{Num. people in county}) = \text{County income}$$

Now that we can get the county incomes, we can get the income for the entire United States: about \$8.4447 trillion (\$8,444,700 million). Finally, we can calculate the average income in the United States:

$$\frac{\text{Total income of all people in the US}}{\text{Number of people in the US}} = \frac{\$8,444,700,000,000}{308,745,538} = \$27,348$$

Recall that the simple mean was \$22,505.

Let's put the entire calculation into a single expression. We'll use w_1 to represent the population of the first county, w_2 to represent the population of the second county, and so on. The label x_1 will represent the average income of county 1, x_2 for the average income of county 2, and so on. Then the mean weighted by population can be written as

$$\frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_{3143} x_{3143}}{w_1 + w_2 + w_3 + \cdots + w_{3143}}$$

This equation represents the **weighted mean** of income, where the weights are given by the population values.

Weighted mean

The weighed mean of observations x_1, x_2, ..., x_n using weights w_1, w_2, ..., w_n is given by

$$\text{weighted mean of } x_i\text{s} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

If the second equation with the Σs isn't math notation you feel comfortable with, just consider the first definition. The second expression is just a fancy way of making the equation compact.

TIP: Simple mean is a special case of the weighted mean

The simple mean is a weighted mean where all the weights are 1:

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1 \times x_1 + 1 \times x_2 + 1 \times x_3 + \cdots + 1 \times x_n}{1 + 1 + 1 + \cdots + 1}$$

The ideas of weighted means are important to some methods that are usually encountered in a second or later course in statistics. We don't touch on the topic again in *OpenIntro Statistics*.