Tutorial 1: Board - Solutions

January 17, 2018

Question 1.3.33, Page 34

Consider the following dataset of home sale amounts in 1000s of $:

590	815	575	608	350
1285	408	540	555	679

(a) Calculate and interpret the sample mean and median.

We first sort the data set. We have an even number of elements in our data set so the median will be the average of the two middle values, 575 and 590. $\widetilde{x} = 582.8$ (1000 $s).

$$\overline{x} = \frac{350 + 408 + ... + 815 + 1285}{10} = 640.5 \text{ (1000 \$s)}$$

We conclude that the sale price of homes is centred at 582.8 (1000 $s). The average sale price of a home is 640.5 (1000 $s). We can also see that $\widetilde{x} < \overline{x}$ so the data is right skewed.

(b) Suppose the 6th observation had been 985 rather than 1285. How would the mean and median change?

Note that the median does not change! (recall what it means to be robust). The new mean is 610.5 (1000 $s). Our mean here compared to (a) is smaller by 30 units. Remember that our units is in (1000 $s) so this is actually a change of $30000 which may or may not be a big deal to some people.

(c) Calculate a 20% trimmed mean by first trimming the two smallest and two largest observations.

We delete the values 350, 408, 815, 1285. Our 20% trimmed mean is 591.17 (1000 $s).

(d) Calculate a 15% trimmed mean.

$\alpha = 0.15$.
R Method: $\alpha n = (0.15)(10) = 1.5$, rounding down to the greatest integer gives us 1. Thus we delete the single highest and lowest observations. This is essentially giving us the same result as a 10% trimmed mean. Our 15% trimmed mean is 596.25 (1000 $s). If you don't believe that it's the same as the 10% trimmed mean in R, perform the following:

```
strength <- c(590,815,575,608,350,1285,408,540,555,679)
mean(strength, trim = 0.15)
mean(strength, trim = 0.10)
```

Question 1.S.78, Page 50

Consider a sample $x_1, x_2, ..., x_n$ and suppose that the values of $\overline{x}$, s^2, and s have been calculated.

(a) Let $y_i = x_i - \overline{x}$ for $i = 1, 2, ..., n$. How do the values of s^2 and s for the y_is compare to the corresponding values for the x_is? Explain.

Claim:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

Proof:

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - \overline{x}) \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} \\
&= n\overline{x} - n\overline{x} \\
&= 0
\end{aligned}$$

Now we can find $\overline{y}$ and ${s_y}^2$.

$$\begin{aligned}
\overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}) \\
&= \frac{1}{n} \cdot 0 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
{s_y}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(y_i)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= {s_x}^2
\end{aligned}$$

Since ${s_x}^2 = {s_y}^2$, it follows that $s_x = s_y$.

(b) Let $z_i = (x_i - \overline{x})/s$ for $i = 1, 2, ..., n$. What are the values of the sample variance and sample standard deviation for the z_is?

Let $z_i = (x_i - \overline{x})/s_x$. Then:

$$\begin{aligned}
\overline{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s_x}\right) \\
&= \frac{1}{n \cdot s_x}\sum_{i=1}^{n}(x_i - \overline{x}) \\
&= \frac{1}{n \cdot s_x} \cdot 0 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
{s_z}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(z_i)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s_x}\right)^2 \\
&= \frac{1}{{s_x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= \frac{{s_x}^2}{{s_x}^2} \\
&= 1
\end{aligned}$$

The sample variance is 1 and the sample standard deviation is also 1.

Bonus: Show that the sample variance, s^2, can be calculated using either

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \quad \text{or} \quad \frac{1}{n-1}\left(\sum_{i=1}^{n}{x_i}^2 - n\overline{x}^2\right).$$

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left({x_i}^2 - 2x_i\overline{x} + \overline{x}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}{x_i}^2 - 2\sum_{i=1}^{n}x_i\overline{x} + \sum_{i=1}^{n}\overline{x}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}{x_i}^2 - 2n\overline{x}\,\overline{x} + n\overline{x}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}{x_i}^2 - 2n\overline{x}^2 + n\overline{x}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}{x_i}^2 - n\overline{x}^2\right)
\end{aligned}$$