Homework 2 Solutions
Math 150

Enrique Treviño

2.2 Associations. Plots (1) and (3) show a positive association. (1) is linear and (3) is not. Plot (2) shows no association. Plot (4) shows a linear negative association.
2.6 Sleeping in college. The sample mean is 6.25 hours per night. The population mean is 5.5 hours per night.
2.8 Medians and IQRs.

(a) The median is the same. The IQR is close to the same.

(b) The median and IQR are 1 more in (2) than in (1).

(c) Median is 5 more in (2) than in (1). The IQR is the same in (2) as in (1).

(d) The median is ten times larger in (2) and the IQR is ten times larger as well.

2.9 Means and SDs.

(a) The mean and standard deviation are a littler bigger in (2).

(b) The mean is a little smaller in (2) but the standard deviation is bigger in (2).

(c) The mean in (2) is 20 more than the mean in (1). The standard deviation is the same in (2) as in (1).

(d) Both have the same mean, but the standard deviation is larger in (2) than in (1).

2.10 Mix-and-match.

(a) Box plot (2). It is kind of symmetric but it has a few outliers to the right and some to the left. It has a peak at the median around 60.

(b) Box plot (3). It is symmetric and uniform. All values are almost equally likely. The histogram is spread out. It has a high standard deviation.

(c) Box plot (1). It is right skewed with many outliers to the right.

2.12 Median vs. Mean. The median is in the [80, 85] bin, so a good approximation is 82.5.

2.16 Distributions and appropriate statistics, Part II.

(a) It is Right skewed because the difference between Q_3 and the median is much larger than between the median and Q_1. The median represents a typical observation better than the mean. IQR represents the variability better than the standard deviation because of how skewed to the right it is.

(b) It is symmetric because the median is the average of Q_1 and Q_3 and we don't have too many outliers. The mean and the standard deviation represent a typical observation and the variability better, respectively.

(c) Right skewed since the peak is at 0 and the right tail goes far to the right. The median and IQR represent a typical observation and the variability better, respectively.

(d) Right skewed because some executives earn much more, but no one earns much less than the typical employee. The median and IQR represent a typical observation and the variability better, respectively.

2.22 Views on immigration.

(a) $372/910 \approx 0.4088 = 40.88\%$.

(b) $278/910 \approx 0.3055 = 30.55\%$.

(c) $57/910 \approx 0.0626 = 6.26\%$.

(d) $57/372 \approx 0.1532 = 15.32\%$ for conservatives, $120/363 \approx 0.3306 = 33.06\%$ for moderates, and $101/175 \approx 0.5771 = 57.71\%$ for liberals.

(e) They don't appear to be independent. It seems that being more liberal is associated to being in favor of the citizenship option for immigrants.

2.23 Views on DREAM Act. They don't appear to be independent, because liberals seem to be more in favor of supporting than conservatives or moderates.

2.24 Raise taxes. They do not appear to be independent. Considerably more democrats seem to be in favor of raising taxes on the rich over the proportion of Republicans in favor of raising taxes on the rich.

2.26 Heart transplants.

(a) It doesn't seem independent. The bar of survivors is bigger in the treatment group.

(b) The transplant group survives more days on average. The median is considerably larger in the treatment group, the first quartile of the treatment group is about the same as the whisker in the control group. So that means even the ones that don't last long in the treatment group would be considered outliers in the control group!

(c) $30/34 \approx 0.882 = 88.2\%$ of the control group died. $45/69 \approx 0.652 = 65.2\%$ for the treatment group.

(d) i. One claim is that more patients in the treatment group survive, that is, if p_A is the proportion of people that die without treatment (in some fixed amount of time) and p_B is the proportion of people that die with a heart transplant, then $p_A > p_B$. The null hypothesis (or independence model) is where we assume $p_A = p_B$ and the alternative hypothesis (or alternative model) is where $p_A > p_B$.

ii. The answers to the fill in the blanks in order: "28", "75", "69", "34", "0", "-0.23 or lower".

iii. Only 2 out of 100 are -.23 or lower, so it suggests that if the independence model is true, then only 2% of the time do we get data like this. Therefore this is strong evidence that the treatment helps prevent death.

2.27 Make-up Exam.

(a) Decrease.

(b)

$$\frac{24 \times 74 + 64}{25} = 73.6.$$

(c) It increases. Now that the mean changed a little, the sum of the squares of the deviations is a little bigger. Since $|64 - 73.6| > 8.9$, this deviation also contributes more than before, so the standard deviation increases slightly. If you want the precise amount, the calculation is below:

First note that

$$8.9 = \sqrt{\sum_{i=1}^{24} \frac{(x_i - 74)^2}{23}}.$$

Therefore

$$\sum_{i=1}^{24} (x_i - 74)^2 = 23 \times (8.9)^2.$$

Now, we can find the new standard deviation:

$$\begin{aligned} s &= \sqrt{\frac{\sum_{i=1}^{24}(x_i - 73.6)^2 + (64 - 73.6)^2}{24}} \\ &= \sqrt{\frac{\sum_{i=1}^{24}((x_i - 74) - .4)^2 + (9.6)^2}{24}} \\ &= \sqrt{\frac{\sum_{i=1}^{24}(x_i - 74)^2 - 2\sum_{i=1}^{24}(x_i - 74)(.4) + 24 \times (.4)^2 + (9.6)^2}{24}}. \end{aligned}$$

Since 74 is the mean of the first 24 grades, then $\sum_{i=1}^{24}(x_i - 74) = 0$. Therefore

$$\begin{aligned} s &= \sqrt{\frac{23 \times (8.9)^2 + 24 \times (.4)^2 + (9.6)^2}{24}} \\ &= \sqrt{\frac{1917.83}{24}} = \sqrt{79.9096} \approx 8.94. \end{aligned}$$

The increase is very slight, when rounding to one decimal, the standard deviation stays the same.

2.33 Stats scores.