ENG 342: Advanced Engineering Math II

Quiz #4

November 8, 2016

Problem 1 [3 pts, 0.5 pts each]

Indicate whether each of the following statements is true or false. If it is false, write a sentence explaining why.

(i) Randomization in sampling is important because it guarantees that any sample will accurately represent the population.

False. Randomization doesn't guarantee *every* sample will accurately represent the population. It guarantees that any sample has no systematic tendency to deviate from the population. Differences may exist, but when they do, it will be due to random variation.

(ii) Randomly choosing 100 people to poll from each of the 50 states would constitute a simple random sample of the United States population.

False. In an SRS, every person in the US would have an equal chance of being selected for the sample. Choosing 100 people state-by-state is stratified sampling, where every person in each state has an equal chance, *i.e.*, people who live in states with more people have less chance of being selected.

(iii) If we are worried about outliers, we may use the median and interquartile range as measures of center and spread as opposed to the mean and standard deviation.

True. These measures are not sensitive to outliers.

(iv) The variance has the same unit of measurement as the sample.

False. The variance is in squared units; the standard deviation is in the same units as the sample.

(v) The three quartiles of a dataset can be taken as the 25th, 50th, and 75th percentiles.

True: Q_1, Q_2, and Q_3 are these percentiles.

(vi) The mean and median of a dataset will never be equivalent.

False: They will be equivalent when the distribution is symmetric.

Problem 2 [4 pts, 1 pt each]

Consider the following dataset of 10 samples:

16 1 17 18 23 19 50 18 13 25

(i) [Computer Permitted] What are the mean and standard deviation?

```
##MATLAB code:
x = [16 1 17 18 23 19 50 18 13 25];
mean(x)
std(x)
```

The mean is $\bar{X} = 20$, and the standard deviation $s_X = 12.37$.

(ii) [Computer Permitted] What are the median and 11% trimmed mean?

```
##MATLAB code:
median(x)
trimmean(x,11)
```

The median is 18, and the 11% trimmed mean is 18.63 (ignoring the largest and smallest value).

Dataset from smallest to largest:

1 13 16 17 18 18 19 23 25 50

(iii) [Computer Permitted] What are the quartiles and interquartile range?

```
##MATLAB code:
prctile(x,25)  #Q1
prctile(x,50)  #Q2
prctile(x,75)  #Q3
iqr(x)
```

The quartiles are $Q_1 = 16$, $Q_2 = 18$, and $Q_3 = 23$. The interquartile range is $23 - 16 = 7$.

(iv) Does the dataset have any outliers? Explain.

The smallest and largest values – 1 and 50 – can be considered left and right outliers, since they are rather far from the median. They also qualify as outliers in the quartile definition since $50 > 23 + 1.5 \times \text{IQR}$ and $1 < 16 - 1.5 \times \text{IQR}$.

Problem 3 [3 pts, 1 each]

In circuits lab, we measure the voltage (V) across a resistor 70 times, and come up with an average of 0.4V and a standard deviation of 0.05V.

(i) We determine that there was a systematic error in our sampling process which caused us to estimate too high by 0.1V each time we took a measurement. How would we adjust the mean and standard deviation without recalculating?

If each measurement was too high by 0.1, then the sample mean should be adjusted to $0.4 - 0.1 = 0.3$ V. The standard deviation will not change.

(ii) We want to report measurements in millivolts (mV) rather than volts (1V = 1000mV). How would we adjust the mean and standard deviation without recalculating?

The mean and standard deviation will scale by the same factor: 300 mV and 50 mV.

(iii) We take another 30 measurements, and come up with an average of 0.2V and a standard deviation of 0.01V (without bias) for this new set of 30. What is the mean of the combined sample of 100 measurements?

We have a mean of 0.3 from 70 samples, and 0.2 from 30 samples. To combine them, we take the weighted average:

$$\frac{0.3 \times 70 + 0.2 \times 30}{70 + 30} = \frac{27}{100} = 0.27 \text{ V}$$

Note that the numerator here is the sum of voltages: $\bar{X} = \sum_i X_i/n$, so $\sum_i X_i = \bar{X}n$. We add these sums from the sample of 30 and 70 and divide by the total.