**Exercise 6**

**NAME_____, NAME_____, NAME_____**

**OVERVIEW**

Netflix has recruited our consulting firm (class) to refine an algorithm that they use to target customers based on viewer information. This algorithm performs many functions concerned with customer experience and engagement, but a key aspect centers on updating Netflix's content catalog. Customers are more likely to retain a Netflix account when content suggestions are populated with recommendations that match their specific tastes.

**Objective 1**: provide Netflix with a list of viewer characteristics that could help them better target audiences who enjoy comedic material.

**Objective 2**: conceptualize and operationalize concepts and construct a survey to administer to a hypothetical simple random sample of Netflix consumer population.

**Objective 3**: summarize the survey results in a descriptive statistics table, which includes appropriate statistics for each variable based on its level of measurement.

**Objective 4**: determine how well the sample mean for age estimates the Netflix consumer population.

**INSTRUCTIONS**

Within your Netflix group divide up into teams of two, or three if needed. Each group will receive a bag with 22 numbers, which reflect the ages of all the observations in our Netflix survey data.

**(1)** One person in your team of two sample five ages at random, and the other person record each age. After recording five randomly sampled ages, replace them and pass the bag to the next team.

$Age_1$: _____, $Age_2$: _____, $Age_3$: _____, $Age_4$: _____, $Age_5$: _____

**Note: use a calculator or Excell to solve below**

**(2)** Compute the mean ($\bar{X}$) of the five ages ($n = 5$) you randomly sampled: $\bar{X} = \frac{\sum X_i}{n}$

$\sum X_i$ = $Age_1$: _____ + $Age_2$: _____ + $Age_3$: _____ + $Age_4$: _____ + $Age_5$: _____ = _____

$\bar{X} = \frac{\quad}{5} =$ _____

**(3)** Compute the standard deviation ($sd$) of your sample mean ($\bar{X}$): $sd = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$

$\sum(X_i - \bar{X})^2$ = (Age$_1$: _____ - $\bar{X}$: _____)$^2$ + (Age$_2$: _____ - $\bar{X}$: _____)$^2$ + (Age$_3$: _____ - $\bar{X}$: _____)$^2$ + (Age$_4$: _____ - $\bar{X}$: _____)$^2$ + (Age$_5$: _____ - $\bar{X}$: _____)$^2$ = _____

$\frac{\sum(X_i - \bar{X})^2}{5-1}$= _____

$sd = \sqrt{\frac{\sum(X_i - \bar{X})^2}{4}}$ =_____

**(4)** Compute the standard error ($SE$) of your sample mean: $SE = \frac{sd}{\sqrt{n}}$

$SE = \frac{sd}{\sqrt{5}}$=_____

**(5)** Compute the margin of error ($MoE$) for your sample mean: $MoE = t \times SE$

**Note**: set alpha ($\alpha$) at 0.05, so $t = 2.776$

$MoE = 2.776 \times SE$ =_____

**(6)** Construct the confidence interval ($CI$) by $\pm$ the $MoE$ to/from your sample mean: $CI = \bar{X} \pm MoE$

Lower bound $CI = \bar{X} - MoE$ =_____

Upper bound $CI = \bar{X} + MoE$ =_____

**(7)** Draw a vertical line to represent your mean and a horizontal line to represent your CI

10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Mean Age with 95% CI

**(8)** Download and save RScript "Ex_6_CIs" located in this week's Canvas module, which replicates steps 1 through 6 above. Make modifications to this RScript based on your sample data and save the RScript frequently along the way.

**(9)** Compare results across teams amongst group members….

(a1) What was the smallest mean _____

(a2) What was the greatest mean _____

(a3) What was the smallest sd _____

(a4) What was the greatest sd _____

(a5) Were there any differences between samples with the largest vs smallest sd? For example, in terms of range and/or mean.

(b1) What was the smallest SE _____

(b2) What was the greatest SE _____

(b3) Were these the same teams with the smallest and greatest sd, respectively? Why should that be the case?

(c1) What was the smallest MoE _____

(c2) What was the greatest MoE _____

(c3) Were these the same teams with the smallest and greatest SE, respectively? Why should that be the case?

(d1) What was the lowest bound 95% CI _____

(d2) What was the highest bound 95% CI _____

(d3) Did any team have a mean that fell outside of these lowest and highest bounds? Why should that not be the case?

**(10)** Repeat step 9 as a class.