STAT 110: Exam #3   Name:_____

1. Answer the following True/False questions.

a.	T	F	Suppose the scores on an exam range from 0 to 100. A score of 50 must be equal to the 50th percentile.
b.	T	F	The average is measured on the same scale as the data. That is, if the data is in dollars, then the average is expressed in dollars as well.
c.	T	F	The standard deviation is measured on the same scale as the data. That is, if the data is in dollars, then the standard deviation is expressed in dollars as well.
d.	T	F	The range is the most widely accepted method of measuring spread for a set of data by statisticians.
e.	T	F	The standard deviation uses all the data points in its calculation.
f.	T	F	Z-scores are used to identify outliers in a set of data.
g.	T	F	If a data point has a Z-Score below -2 or above +2, then it is considered an outlier.
h.	T	F	A p-value cannot be less than 0 nor greater than 1.
i.	T	F	A t-test can be used to expand the scope of inference beyond the data you have collected.

I live in Rushford and drive to Winona each day to get to work. It takes about 30 minutes to get to work, but the actual time varies due to: when I leave the house, conditions of the road, road construction, whether or not I have to wait for train, availability of parking spots, etc.

2. In terms of my driving, which of the following would influence **only** the center (or location) of the distribution for "time it takes to get to work".

    a. Drive faster every day or drive slower every day.
    b. Drive faster on some days and drive slower on other days.
    c. Changing your speed will not change the center of this distribution.

3. In terms of my driving, which of the following would influence **only** the variability (or spread) of the distribution for "time it takes to get to work".

    a. Drive faster every day or drive slower every day.
    b. Drive faster on some days and drive slower on other days.
    c. Changing your speed will not change the variability of this distribution.

Consider location across Minnesota for which snowfall amounts have been collected. For this problem, you can assume that the further north the location, the more snow that location gets.

Original Data	New Data #1	New Data #2
↑N		
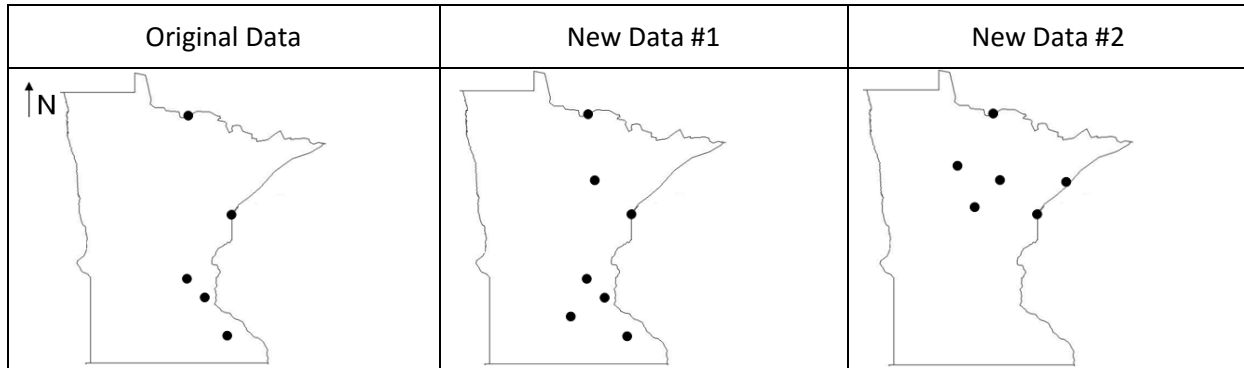
4. Answer the following True/False questions.

a.	T	F	We would expect that New Data #2 would have the largest average snowfall amount.
b.	T	F	We would expect that the Original Data would have the smallest standard deviation because it has the fewest number of locations.
c.	T	F	We would expect that the average of the New Data #1 would be about the same as the average from the original data.
d.	T	F	We would expect the standard deviation to be about the same between the Original Data and New Data#1.
e.	T	F	We would expect that the standard deviation for New Data #2 to be smaller than the standard deviation from New Data #1.

5. A test is given to 100 students, and the scores were calculated. After grading the test, the instructor realized that the 10 students with the highest scores did exceptionally well and decides to award these 10 students by giving them an additional 5 points. Which of the following statements is correct?

    a. The average of the new scores will be smaller than the average of the old scores.
    b. The average of the new scores will be larger than the average of the old scores.
    c. The average will not change.
    d. It is impossible to tell.

6. A test is given to 100 students, and scores were calculated. After grading the test, the instructor realized that the 10 students with the highest scores did exceptionally well. She decides to award these 10 students by giving them an additional 5 points. Which of the following statements is correct?

    a. The median of the new scores will be smaller than the median of the old scores.
    b. The median of the new scores will be larger than the median of the old scores.
    c. The median will not change.
    d. It is impossible to tell.

Consider the Vitamin Intake dataset that we discussed in class. Of interest in the investigation is the intake of Iron, an important mineral for us to consume. We will be making comparisons across two groups – Females and Males here. The units of measurement for this investigation is mg/day.
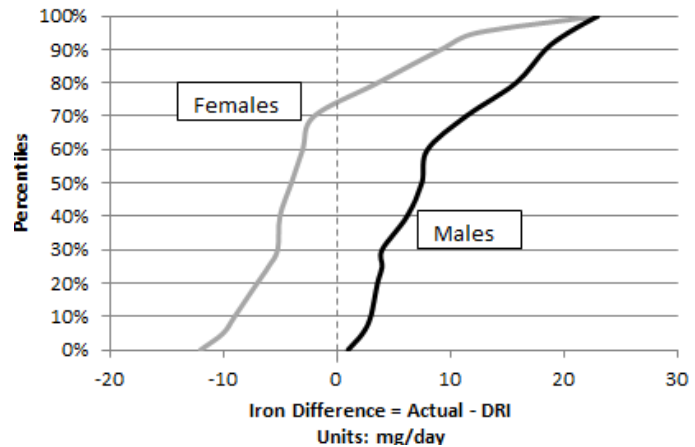
Iron Difference = Actual - DRI

Excel output for making comparisons.

PivotTable Output

Row Labels	Average of Iron Difference	StdDev of Iron Difference	Count of Iron Difference
F	-1.68	7.62	57
M	9.21	6.80	14
Grand Total	0.46	8.61	71

CDF Plot

7. Use the above output to answer the following True/False questions.

a.	T	F	The intake level for iron cannot go below zero, thus the average for females should be +1.68 instead of -1.68.
b.	T	F	A comparison of these averages suggests that, for the students in this study, males are taking in more iron relative to their daily recommended levels than females.
c.	T	F	A comparison of these averages suggests that, for all students at WSU, males are taking in more iron relative to their daily recommended levels than females.
e.	T	F	The CDF plot suggests that a majority of the females are taking in less iron than their daily recommended level.
f.	T	F	The number of females is different than the number of males in this study; thus, fair comparisons cannot be made between the two groups.

8. Suppose the research question for this problem is two-sided. Which of the following would be the appropriate t-test analysis in Excel. The following information is from the Excel help file and defines the last argument for the =TTEST() function in Excel.

IF TYPE EQUALS	THIS TEST IS PERFORMED
1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)

   a. =TTEST(FemaleData, MaleData, 2, 1)
   b. =TTEST(FemaleData, MaleData, 2, 2)

9. Consider the following p-value output from the t-test completed in Excel.

P-Value for T-Test
6.37E-06

   Identify the most appropriate conclusion for this problem.
   a. We have enough statistical evidence to say the data supports the research question (p-value = 6.37E-06).
   b. We have enough statistical evidence to say that, on average, a difference exists in the intake levels of iron between females and males (p-value = 6.37E-06).
   c. We have enough statistical evidence to say that, on average, a difference exists between the actual intake of iron and the daily recommended intake of iron (p-value = 6.37E-06).
   d. We have enough statistical evidence to say that, on average, a difference exists in the actual intake level of iron relative to their daily recommended intake level between females and males (p-value = 6.37E-06).

10. Consider the following 95% confidence interval for this investigation.

0   6.5   10.9   15.3

Avg. Diff for Males – Avg. Diff for Females
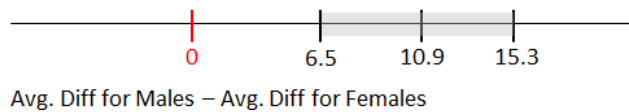
   Answer the following True/False questions.

a.	T	F	This interval seems to suggest a difference, on average, between Females and Males.
b.	T	F	This interval suggests both Females and Males are taking in more iron, on average, then their daily recommended levels.
c.	T	F	The margin of error for this interval is about 8.8.

11. The data below come from the 2005 Youth Risk Behavior Survey of American high schools students, which was conducted by a branch of the United States government. This study investigated the potential relationship between smoking and body mass index (BMI). Body mass index is one possible measurement of one's fitness level. Generally speaking, the smaller the number the more fit someone is.

Smoking Status	Average of BMI
Non-Smoker	21.5
Smoker	24.5

Research Question: "Does the average BMI differ between American high school students who smoke and those who do not smoke?"

a. Smoking status is one of the variables of interest in this investigation and defines the two groups to be compared in this study. Is this variable categorical, or numeric?

    Numerical   Categorical

b. The second variable being considered in this investigation is the subject's BMI. Is this variable categorical, or numeric?

    Numerical   Categorical

c. Consider the research question stated above. Which of the following test should be used to compute a p-value to determine whether not this question is supported statistically? Circle only one answer.
    i.   A binomial test
    ii.   A t-test
    iii.   A chi-square test

d. Conducting a statistical test will allow us to expand our scope-of-interest and permit us to make decisions beyond the data we've collected in our study. Identify which of the following would be considered within the scope of inference for this investigation.

Within Scope of Inference? (Place X is one column)		Description
Yes	No	
		The individuals who took part in this study
		2005 American high school students who smoke
		2005 American high school students who do not smoke
		Anybody in the world who was a high school student in 2005
		American high school students who graduated in 1970 – a time when the perception of smoking was very different than 2005.

Consider the following set of plots we discussed in class. In each plot, females are the open circles and have the dotted trend line.
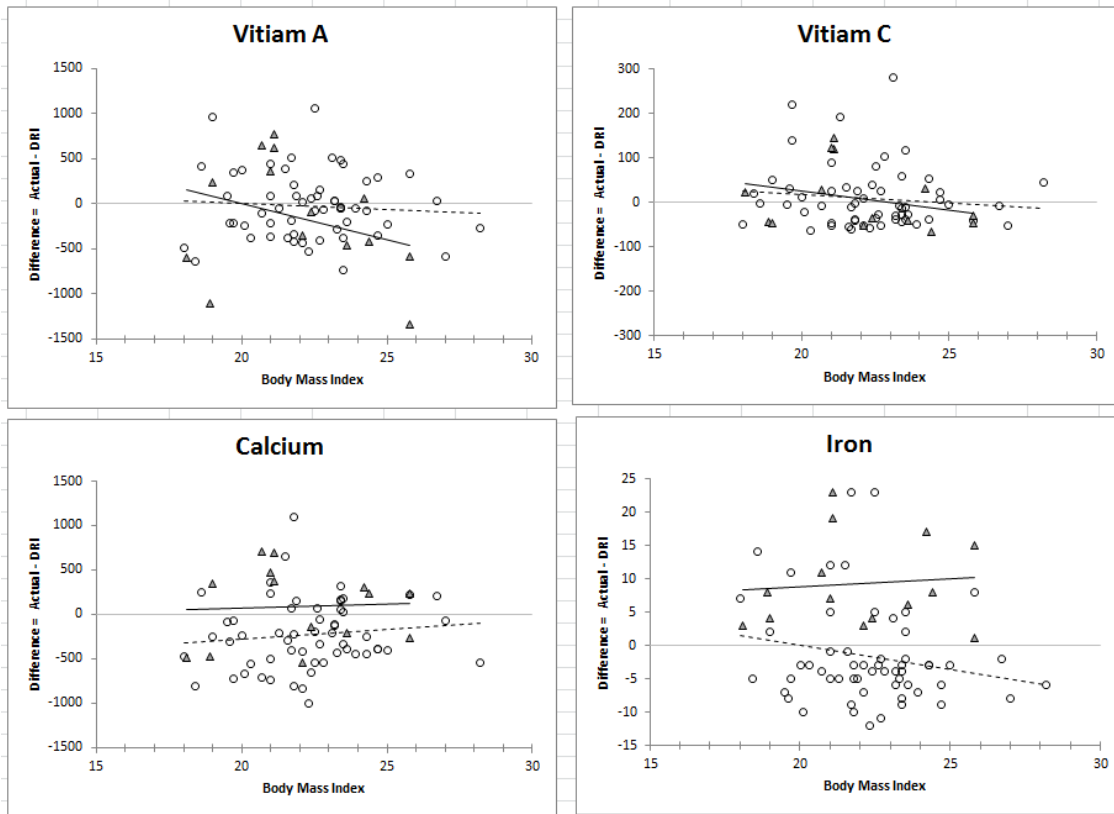
12. Answer the following True / False questions.

a.	T	F	Females tended to take in an appropriate amount of Vitamin A.
b.	T	F	Males tended to take in an appropriate amount of Iron.
c.	T	F	Females tended to be closer to their daily recommended intake for Vitamin A compared to Males.
d.	T	F	If BMI has no impact on intake, then the trend line should be flat.
e.	T	F	The effect of Gender is about the same regardless of BMI levels in regards to an appropriate amount of Calcium intake.
f.	T	F	An interaction is said to exist between Gender and BMI in regards to an appropriate amount of Iron intake.