STAT 210: Final Exam: Name: <u>SOLUTION</u>

1. Answer the following True/False questions.

a.	T	(F)	Extreme observations (i.e. outliers) affect the median more than the mean.
b.	(T)	F	Boxplots can be used to identify outliers in a set of data.
c.	T	(F)	A data point with a Z-Score above 1 or below -1 would be considered a statistical outlier.
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 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.
f.	(T)	F	A p-value cannot be less than 0 nor greater than 1.

The investigation here centers around whether or not the percentage of alumni who donate money is different for private and public schools. Information from colleges and universities in Minnesota, Wisconsin, and Iowa are being analyzed here. The summary statistics for each group are provided here.

2. Consider the summaries above for the Private and Public schools.
    a. The mean and median are larger for Private because this group has more observations in it (N = 29 vs. N = 21).
    (b.) The mean and median are larger for Private which means that alumni from Private institutions tend to donate more often alumni from Public institutions.
    c. Both a. and b.

3. Consider the summaries above for the Private and Public schools.
    a. The standard deviation for Private is larger; thus, the percentage of alumni who donate from Private institutions will have a larger average distance to the middle than the percentage of alumni who donate from Public institutions.
    b. The range for Private is larger; thus, the percentage of alumni who donate from Private institutions will have a larger average distance to the middle than the percentage of alumni who donate from Public institutions.
    c. Both a. and b.

4. Consider the following margin-of-error computations for each group.

    Margin of Error = 2 * Standard Error of Mean

    where standard error of mean = standard deviation / sqrt(sample size)

    - Margin of Error for Private = 2 * 2.54 = 5.08
    - Margin of Error for Public = 2 * 1.82 = 3.64

    a. The margin-of-error for Public is smaller because the number of observations in this group is smaller.
    b. The margin-of-error for Private is larger because the variation in this group is larger.
    c. Both a. and b.

5. Consider the following mean ± margin-of-error values for each group which approximates the 95% confidence intervals provided above.

    Mean ± MOE for Private = 30.17 ± 5.08,   25.09 up to 35.25
    Mean ± MOE for Public = 16.57 ± 3.64,   12.93 up to 20.21

    a. These formulas and their corresponding intervals allow us to compare the true alumni donation rate for private and public institutions in any state.
    b. These formulas and their corresponding intervals allow us to compare the dollar amount that each private school alum donates to the dollar amount that each public school alum donates for institutions in Minnesota, Wisconsin, and Iowa.
    c. These formulas and their corresponding intervals are necessary to compare the average alumni donation rate from the private schools in our sample to the average alumni donation rate from the public schools in our sample (i.e. comparing the 30.17 to the 16.57).
    d. None of the above.

6. Consider the following data on the percentage of homicide victims for ages 18-24 between MN and WI for the years 1976 to 2000 from the U.S. Department of Justice web site.

Difference = MN – WI

	MN	WI	Difference
1	21.6	19.5	2.1
2	17.3	19.2	-1.9
3	22.4	23.7	-1.3
4	24.2	21.2	3
5	27.4	30.8	-3.4
6	13.6	19	-5.4
7	23.2	15.8	7.4
8	13.8	21.1	-7.3
9	13.5	19.8	-6.3
10	18.5	21.5	-3
11	19.2	27.7	-8.5
12	19.2	15.6	3.6
13	17.6	23.4	-5.8
14	14.2	26.6	-12.4
15	10	25.9	-15.9
16	25	26.8	-1.8
17	16.5	20	-3.5
18	24.4	22.6	1.8
19	30.4	29.9	0.5
20	24.4	28	-3.6
21	17.2	24.9	-7.7
22	28.1	27.7	0.4
23	29.1	.	.
24	25.4	31.3	-5.9
25	28.7	27.3	1.4

Difference

Moments

Mean	-3.0625
Std Dev	5.294198
Std Err Mean	1.0806736
Upper 95% Mean	-0.826956
Lower 95% Mean	-5.298044
N	24

Complete Step 4 and circle the best response for Steps 3, 5, and 6.

Step 0:   Question of Interest: Are there differences in the average percentage of homicide victims for ages 18-24 between MN and WI? If so, what are these differences?
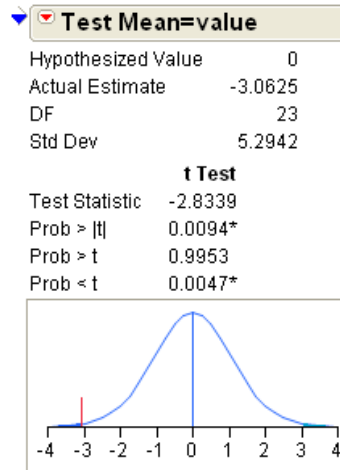
Step 1:   Set up the null and alternative hypothesis
$$H_O : \mu_{Difference} = 0$$
$$H_A : \mu_{Difference} \neq 0$$

Step 2:   Error rate =5%, which gives a confidence rate of 95%

Step 3:   Output from hypothesis test.

Test Mean=value

Hypothesized Value	0
Actual Estimate	-3.0625
DF	23
Std Dev	5.2942

t Test

Test Statistic	-2.8339
Prob > |t|	0.0094*
Prob > t	0.9953
Prob < t	0.0047*

Recall that a test statistic is a type of "Z-Score" that is used to measure whether or not the average difference is extreme against the no difference situation, i.e. 0. The test statistic value here is -2.84. Which is the following is true regarding this value?

a.) The value of this test statistic lends support to our research question.
b. The value of this test statistic does not lend support to our research question.
c. The test statistics cannot be used to show support one way or the other for our research question.

Step 4: Determine the appropriate p-value and make the statistical decision for this test.
The Decision Rule: If p-value is less than error rate, then the data tends to support the alternative hypothesis.

Appropriate P-value: `0.0094 (use the two-tailed as testing for difference between states)`
Decision: `data supports the research question as the p-value is less than 0.05`

Step 5: Circle the most correct conclusion for this test.

a. In any given year, we are 95% certain that MN will have a lower homicide rate than WI for the 18-24 age group.
b.) In any given year, we are 95% certain that there is a difference in the homicide rates between MN and WI for 18-24 age group. **[This conclusion should mention, "on average" as well.]**
c. In any given year, we are 95% certain that WI will have a lower homicide rate than MN for the 18 – 24 age group.
d. In any given year, we are 95% certain that there is no difference in the homicide rate between MN and WI for the 18-24 age group.
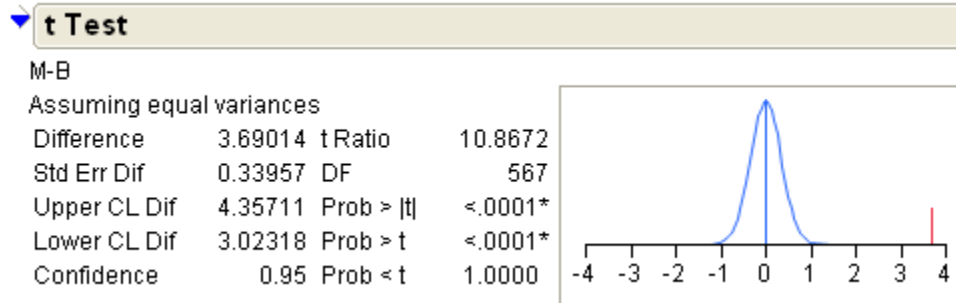
Step 6: I have computed a 95% confidence interval for the difference in the homicide rates between MN and WI. Again, Difference = MN – WI.

Moments	
Mean	-3.0625
Std Dev	5.294198
Std Err Mean	1.0806736
Upper 95% Mean	-0.826956
Lower 95% Mean	-5.298044
N	24

Circle the most correct interpretation for this interval.

e.) In any given year, we are 95% certain that the homicide rate in MN will be about 1 to 5 lower than in WI. **[Again, should mention, "on average" here as well.]**
a. In any given year, we are 95% certain that the homicide rate in WI will be about 1 to 5 lower than in MN.
b. These rates are negative and you cannot have a negative homicide rate, so something is wrong with this interval.
c. In any given year, we are 95% certain that the homicide rate between MN and WI is different because this interval does not contain 0.

Consider the breast cancer dataset from University of WI – Madison that we looked at in class. There were two groups in this dataset (Malignant and Benign). Of interest here is the study of the Texture measurement of the cell. You can see from the JMP output that the confidence interval for the true difference in the average texture measurements between the malignant cells and the benign cells goes from 3.02 up to 4.36. Identify whether the following statements about this interval are correct.

t Test

M-B
Assuming equal variances

Difference	3.69014	t Ratio	10.8672
Std Err Dif	0.33957	DF	567
Upper CL Dif	4.35711	Prob > |t|	<.0001*
Lower CL Dif	3.02318	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000

7. Answer the following True/False statements.

a. We are 95% certain that the true difference in the average texture measurements is between 3.02 and 4.36.
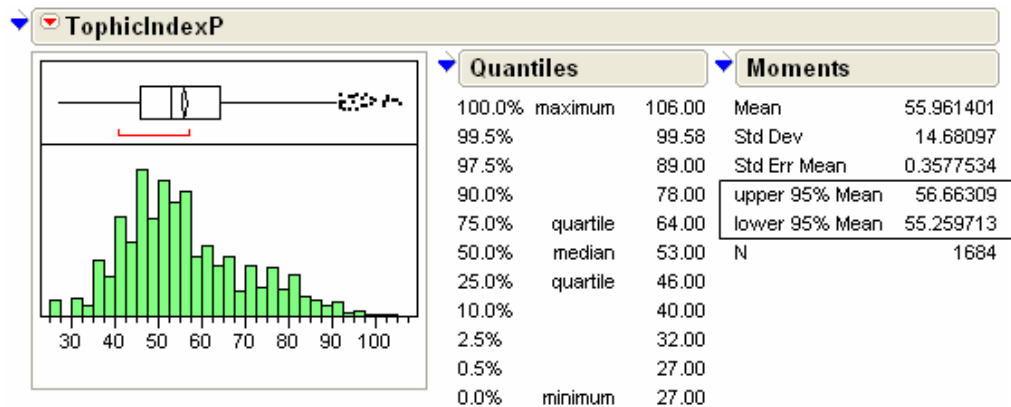
    (TRUE)   FALSE

b. This interval provides evidence that there is a statistical difference in the average texture measurements between malignant and benign cells.

    (TRUE)   FALSE

c. The observed difference in the averages (i.e. 3.69) is guaranteed to be in our 95% confidence interval (i.e. 3.02 up to 4.36).

    (TRUE)   FALSE

A lake assessment study was conducted across the state of Minnesota. The variable being considered here is TrophicIndexP.

TophicIndexP

Quantiles

100.0%	maximum	106.00
99.5%		99.58
97.5%		89.00
90.0%		78.00
75.0%	quartile	64.00
50.0%	median	53.00
25.0%	quartile	46.00
10.0%		40.00
2.5%		32.00
0.5%		27.00
0.0%	minimum	27.00

Moments

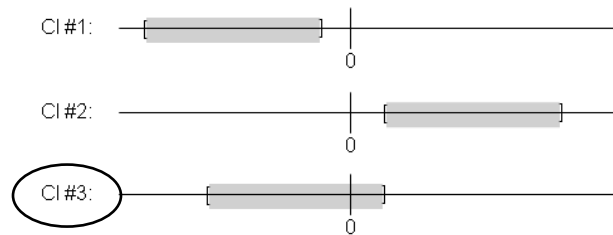|	
Mean	55.961401
Std Dev	14.68097
Std Err Mean	0.3577534
upper 95% Mean	56.66309
lower 95% Mean	55.259713
N	1684

Note that the 95% confidence interval for the true mean TrophicIndexP goes from 55.26 up to 56.66.

8. Which of the following is most correct for testing whether or not the true mean TrophicIndexP is different from 50?

   a. Less than .05, since the 95% confidence interval does not include zero.
   b. Less than .05, since the 95% confidence interval does not include 50.
   c. Greater than .05, since the 95% confidence interval does not include 50.
   d. It is impossible to tell.

9. Suppose that both males and females were asked the question "What is the fastest you have ever driven a car (in mph)?" The sample mean for females was 90.5 mph, and the sample mean for males was 96.1 mph. The p-value for testing for a difference in means was .0944. Identify (by circling the label) which of the following 95% confidence intervals is most likely correct for the true difference in the means (i.e. Average$_{Female}$ – Average$_{Male}$).

10. A researcher is designing a research study. She is hoping to show that the results of an experiment are statistically significant. What type of *p*-value would she want to obtain?

   a. A large *p*-value.
   b. A small *p*-value.
   c. The magnitude of a *p*-value has no impact on statistical significance.