

Name: _____

Lab Section/TA name: _____

Topics to be considered for final exam:

- Module 6: Hypothesis Testing
- Module 7: Simple Linear Regression
- Big picture ideas:
 - The difference between a population and sample
 - The difference between a parameter and a statistic
 - How are confidence intervals and hypothesis tests similar and different
 - Notation involving probabilities on a normal distribution and t distribution
 - Notation that has been used throughout the semester, including:
 - * $\mu, \bar{x}, \sigma, s, n, z, t$
 - * $H_0, H_A, r, R^2, \beta_0, b_0, b_1, \beta_1, \epsilon, \rho, r$
 - Definitions of key words that have been used throughout the semester, including:
 - * mean, standard deviation, variance, confounding variables, percentile, distribution, sampling distribution, standard error, confidence intervals, margin of error
 - * null and alternative hypothesis, p-value, Type I and Type II errors, correlation, response variable, predictor variable, intercept, slope, R squared

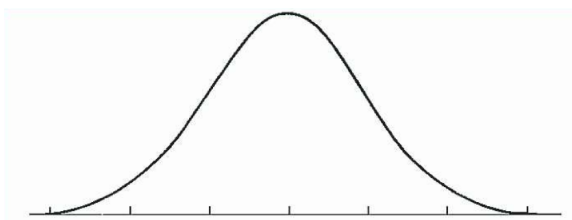
This practice exam will focus on modules 6 and 7 but you should review the big picture ideas on your own and be prepared for a couple of questions on those topics.

1. You recently got a complaint from a customer who said they were on hold with customer service for too long. You want to investigate the average time a caller spends on hold. The head of customer service claims that the average wait time is only 5 minutes, however you suspect it might be longer than that.

(a) State the null and alternative hypothesis being tested.

(b) You call in at 36 random times and the average time on hold is 6.1 minutes with a standard deviation of 3 minutes. Give the formula for the t-statistic and calculate the numerical result.

(c) Use the t-distribution provided to label the axes, your t-statistic, and the p-value desired.

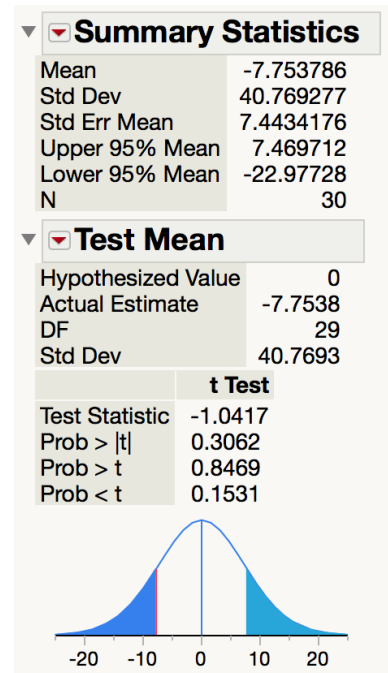


(d) If you know that $P(t < 2.2) = .9827$ when degrees of freedom is 35, find the p-value desired.

(e) Explain what the p-value represents in words and in the context of this problem.

(f) Based on your hypothesis test what would you conclude? Do people spend on average more than 5 minutes on hold?

2. In September 2018 Nike launched a new ad campaign that caused some controversy. You want to see if tweeting in support of Nike would change your number of twitter followers. You randomly select 30 people who have between 200 and 1000 twitter followers and who tweeted about Nike. You record the number of followers they had before they tweeted in support of Nike and after they tweeted. The JMP output below shows the JMP output for the differences (Before – After) in Twitter followers.



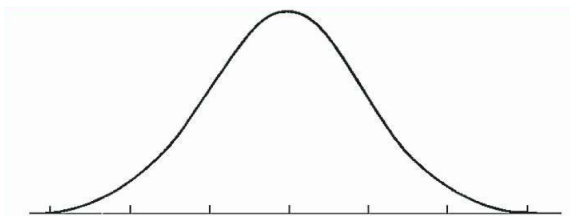
- (a) Is this data paired or not paired?

- (b) State the null and alternative hypothesis being tested.

- (c) Give the formula for the t-statistic and give the numerical result.

- (d) Give the p-value for this hypothesis test.

- (e) Use the t-distribution provided to label the axes, your t-statistic, and the p-value.



(f) In the sample, did people on average tend to gain or lose followers after tweeting in support of Nike?

(g) Based on the hypothesis test what should you conclude about the difference in means? Does tweeting in support of Nike change your number of twitter followers significantly?

(h) Based on the p-value of your test, would you expect the confidence interval for the difference in means to include zero? Explain why or why not.

3. Put the following t-statistics in order from smallest p-value to largest p-value, assuming the same degrees of freedom for all and a two sided test.

$t = 3.2$

$t = -.5$

$t=1.4$

$t = -1.3$

smallest p-value _____ largest p-value

4. Fill in each graph with it's correct correlation. Not all values will be used.

$r=-1.4$

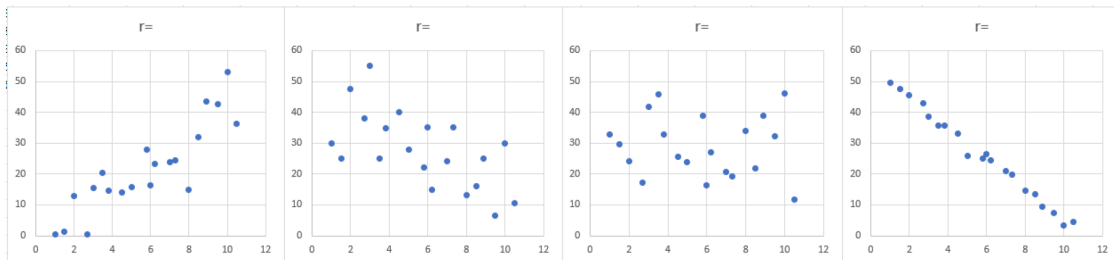
$r=-.95$

$r=-.6$

$r=0$

$r=.75$

$r=1.1$



5. The next two questions will use the following information. Insurance companies use personal information to predict the medical costs you will need covered during a year in order to set policy prices and make decisions. The following variables are measured for a sample size of 1338 people.

- Charges: Individual medical costs billed by health insurance
- Age : age of insurance policy holder (between 18 and 65)

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	3165.885	937.1495	3.38	0.0008*
age	257.72262	22.50239	11.45	<.0001*

- (a) Give the theoretical model you are estimating.
- (b) Based on the JMP output give the line of best fit.
- (c) Interpret the slope for age in context of this problem.
- (d) Perform a hypothesis test to check if the slope of the line is equal to zero.

$H_0 :$

$H_A :$

$t =$

$p - value =$

conclusion:

- (e) Give the confidence interval for the slope of age.

95% confidence interval for _____ : (_____ , _____)

- (f) Predict the amount in charges to insurance when the person's age is 60 years old.

(g) There is a 60 year old in the data set whose charges were \$12147. What is the residual error?

6. Following up on the previous question, suppose that we create a 95% confidence interval for ρ , the (true) population correlation coefficient, and it is (0.2494, 0.3470).

(a) What is the sample correlation coefficient?

(b) If we are testing $H_0 : \rho = 0$ vs. $H_A : \rho \neq 0$, which of the following conclusion(s) is appropriate?

- i. Reject H_0
- ii. Reject H_A
- iii. Fail to Reject H_0
- iv. Fail to Reject H_A
- v. Accept H_0
- vi. Accept H_A

(c) Which of the following are true?

- i. As age increases, charges tends to increase
- ii. As age decreases, charges tends to increase
- iii. As age increases, charges tends to decrease
- iv. As age decreases, charges tends to decrease

7. Explain what extrapolation is and why it is something we should avoid.