Stat 301 **Spring 2019** **Final exam practice problems**

*These questions are all taken from previous semesters' exams. For the exam you may use a calculator and two 4x6 index cards that you fill out yourself. It is **strongly** advised that you attempt to solve these problems on your own before looking at the solutions. Working through practice problems unassisted will help you best determine where to direct your studying before you take the real final exam.*

1. Suppose we have used sample data to create this 95% confidence interval for mean height of adult females:

$$95\%\ CI\ for\ \mu = (62.5,\ 66.7)$$

Identify which null hypothesis would result in which p-value from the list on the right.

a. $H_0: \mu = 63$ $p - value =$ ___________

b. $H_0: \mu = 65$ $p - value =$ ___________

c. $H_0: \mu = 67$ $p - value =$ ___________

Possible p-values
-0.02241
0.02586
0.7035
0.1320
1.0556
64.60

2. Why do we say that a small p-value provides stronger evidence for a scientific hypothesis than a large p-value provides? For full credit, your answer should refer to the logic behind how a p-value is used to provide evidence.

3. We have looked at a lot of hypothesis tests over the semester. The box on the right contains brief descriptions of how test statistics are calculated for various tests. For parts a., b., and c. below, match the given null hypothesis to the appropriate description of how to calculate its test statistic. Writing down the appropriate Roman numeral as your answer is sufficient.

a. $\beta_1 = 0$

b. $\mu_d = 0$

c. $\mu_1 - \mu_2 = 0$

How the test statistic is calculated
i. The sample means and standard deviations for both groups are calculated individually. The difference in sample means is divided by the standard error of the difference in sample means.
ii. The sample mean and standard deviation of the differences between the paired observations is calculated. The average paired difference is divided by the standard error of the average paired difference.
iii. The correlation coefficient is squared, and this number is divided by its compliment.
iv. The estimated value of the slope is divided by its standard error

4. The SAT is a standardized test taken by high school students before applying to college. There has been a lot of research (and controversy) regarding what kind of factors might influence SAT scores. This page and the next refer to data on SAT results from every state in the country. The following variables will be used:

total_sat:	Average combined SAT score (math and verbal)
expenditure:	Average expenditure per student, in thousands of dollars
student_faculty_ratio:	The average number of students per faculty member
percent_taking:	The percentage of students who take the SAT

Results from 4 simple regression models are given below

Model A (response variable: *total_sat*)

RSquare 0.139914

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1081.1194	44.94187	24.06	<.0001*
expenditure	-19.93503	7.368016	-2.71	0.0096*

Model B (response variable: *total_sat*)

RSquare 0.783231

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1050.6406	8.574524	122.53	<.0001*
percent_taking	-2.435282	0.190984	[redacted]	<.0001*

Model C (response variable: *percent_taking*)

RSquare 0.345058

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-31.47782	14.25205	-2.21	0.0323*
expenditure	11.376992	2.336559	4.87	<.0001*

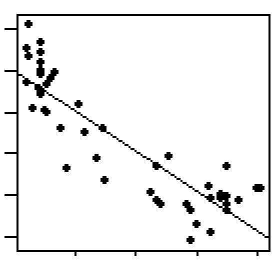
Model D (response variable: *total_sat*)

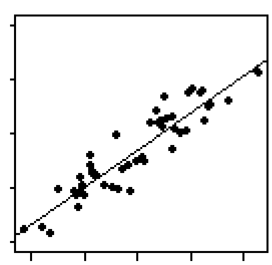
RSquare 0.013019

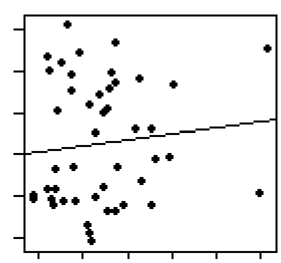
Parameter Estimates

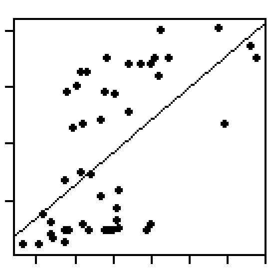
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	900.3193	81.64092	11.03	<.0001*
student_faculty_ratio	3.7102608	4.815762	0.77	0.4451

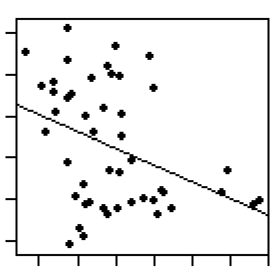
a. Below are 5 scatterplots with labels and axes missing. 4 of these scatterplots correspond to one of the 4 models above, with the predictor variable on the horizontal axis and the response variable on the vertical axis. One scatterplot does not refer to any of the 4 models. Label each plot "A", "B", "C", "D", or "none" on the blank lines provided. (**5 points**)

 ________

 ________

 ________

 ________

 ________

b. On the top left plot above, draw a set of bands that could realistically be 95% confidence bands for the mean.

c. One the top middle plot above, draw a set of bands that could realistically be 95% prediction bands for an individual observation.

d. Write out the formula for the estimated line of best fit for predicting total_sat using expenditure.

e. Create an approximate 95% confidence interval for the slope in model D. *Using this confidence interval*, state whether $H_0: \beta_1 = 0$ can be rejected, and briefly justify your answer.

f. Calculate the test statistic testing $H_0: \beta_1 = 0$ in model B. *Using the p-value*, state whether this null hypothesis should be rejected, and briefly justify your answer.

g. How much does the predicted percentage of students taking the SAT change when per student expenditure increases by $1000?

h. All of the plots of the previous page have lines of best fit superimposed on them. Suppose that, on one of these plots, we were to draw some line other than the "best" fitting line. What mathematical property would this line have that would distinguish it from the "best" line?

i. Suppose we wanted to modify model A so that it could estimate the relationship between expenditure and total_sat in states for which more students take the ACT than the SAT, and in states for which more students take the SAT than the ACT. You could do this with a new indicator variable called "ACT", which equals 1 if more students take the ACT than the SAT, and 0 if more students take the SAT than the ACT.

   i. Write out what this model would look like, generically, if we wanted to allow for both the intercept and the slope for expenditure to be different for ACT = 1 vs. ACT = 0.

   ii. Make up values for the relevant slopes such that the slope of expenditure is -20 in states where more students take the ACT than the SAT, and -18 in states where more students take the SAT than the ACT.

j. So far, we have only considered simple regression models for this data. Here are the results from a multiple regression model:

```
total_sat = 993.13+11.87*expenditure-2.80*percent_taking
```

RSquare	0.815695

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	993.13409	22.15022	44.84	<.0001*
expenditure	11.865419	4.262135	2.78	0.0079*
percent_taking	-2.795154	0.220062	-12.70	<.0001*

Compare these results to the simple regression models relating *total_sat*, expenditure, and *percent_taking*. There is some evidence that the apparent effect of expenditure on *total_sat* is confounded by *percent_taking*. Use the results from this multiple regression model and any relevant simple regression models to explain how *percent_taking* could be confounding the relationship between expenditure and *total_sat*. Your answer should be thorough enough that it is clear you understand what "confounding" means in this context, and that you can identify all the relevant statistical evidence for this confounding.

k. How does the proportion of variability in *total_sat* explained by the model in j. compare to the proportion of variability in *total_sat* explained by the simple regression model in which *percent_taking* is the only predictor?

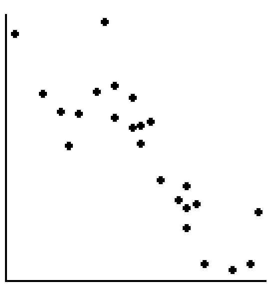
l. Suppose one state in your data set has an average expenditure of $5000 per student, 65% of students take the SAT, and the average combined SAT score in this state is 864. Calculate the residual for this observation, using the model in part j.
(hint: you might want to double check the units that the predictor variables are in before you plug in any numbers)

5. Suppose you conduct a two-tailed, two-sample t-test and get a p-value of 0.023. Which of the following statements are valid conclusions to draw from this test? Select all that apply.

   a. There is a 0.977 probability that the null hypothesis is false.
   b. There is a 0.023 probability that the null hypothesis is true.
   c. There is a 0.023 probability of getting results at least this extreme, assuming the null hypothesis is true.
   d. There is a 0.977 probability of getting results at least this extreme, assuming the alternative hypothesis is true.
   e. There is a 0.023 probability that these results occurred due to chance.
   f. The sample means would be considered significantly different at the $\alpha = 0.05$ level.
   g. The sample means are more than one standard error apart.
   h. If we repeatedly take new random samples from the same population, we expect to reject the null hypothesis 97.7% of the time.
   i. If we use the same data to construct a 95% confidence interval for the difference in means, it will not contain zero.
   j. If we use the same data to construct a 95% margin of error for the difference in means, it will be smaller than the point estimate for the difference in means.

6. Suppose we have a data set with the variables Wealth, Age, School, and Hours. In each scatter plot below, the X (horizontal) axis variable is given but the Y (vertical) axis variable is not. Use the provided correlation matrix to identify which variable is represented on each Y axis. (You don't have to write down the correlation coefficients, just fill in the variable name for each Y axis.)

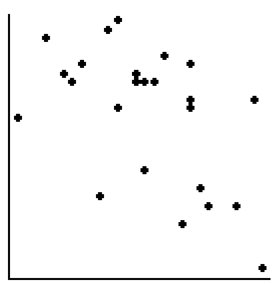
i.



X axis: Age

Y axis: ____________

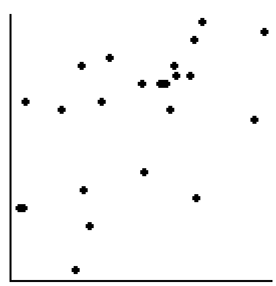
ii.



X axis: Age

Y axis: ______________

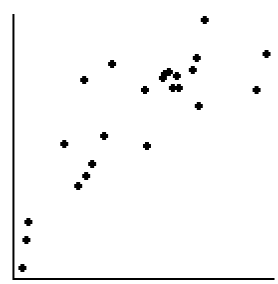
iii.



X axis: Hours

Y axis: ______________

iv.



X axis: Hours

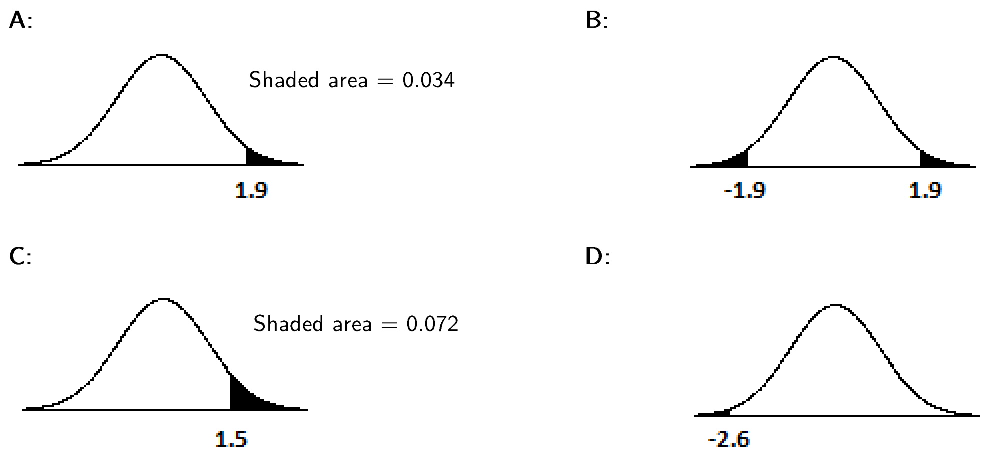
Y axis: ______________

Correlation matrix:

	Wealth	**Age**	**School**	**Hours**
Wealth	1	-0.72	0.78	0.82
Age	-0.72	1	-0.52	-0.85
School	0.78	-0.52	1	0.53
Hours	0.82	-0.85	0.53	1

7. Explain the difference between what standard deviation quantifies and what standard error quantifies.

8. Suppose four different one-sample t-tests are conducted, tests A, B, C, and D. Each test is from a different sample, the sample size for each test is the same, and each test is done using $\alpha = 0.05$. The diagrams below show the sampling distribution of each test statistic under H_0. For each diagram, the shaded area is the p-value, and the value of the test statistic is given. The value of the shaded area is given for diagrams A & C.



Use this information to answer questions a. through f. below: (2 **points each**)

a. Make up a plausible value for the p-value in test **D**?

b. For a right sided test using whatever sample size was used in these tests, make up a plausible value for a test statistic that would result in a p-value of 0.04.

c. Is the sample mean significantly different than the hypothesized mean in test B? Briefly justify your answer.

d. If you were to make a 95% CI for the mean in test A, would it contain the hypothesized mean? Briefly justify your answer.

e. Which test resulted in a test statistic that would be <u>least</u> likely to occur by chance if H_0 were true? Briefly justify your answer.

9. When assessing statistical power, we primarily look at three quantities: effect size, sample size, and power. For each scenario listed below, state whether *required sample size* will increase, decrease, or if there isn't enough information to know:

   a. Desired power increases

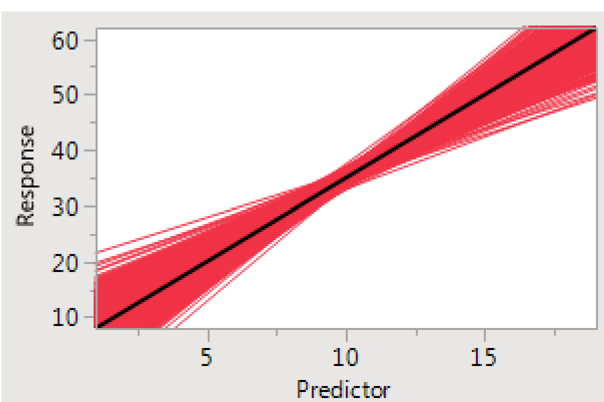
   b. Assumed effect size increases

   c. Desired power decreases *and* assumed effect size decreases

   d. Assumed effect size decreases *and* desired power increases

10. Suppose we are using t-tests to test multiple null hypotheses, and we decide to do a Bonferroni correction. What affect does the Bonferroni correction have on the power of each t-test? Explain your answer.

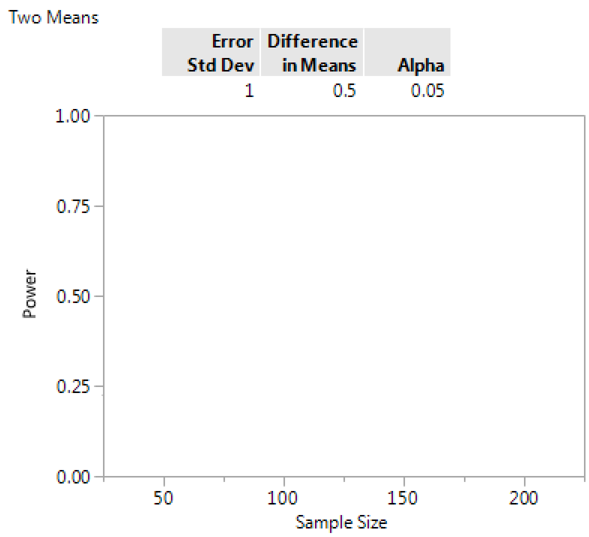
11. The picture below shows the sampling distribution of the line of best fit, which was covered in the notes on inference for regression. Explain what we mean when we call this a "sampling distribution".

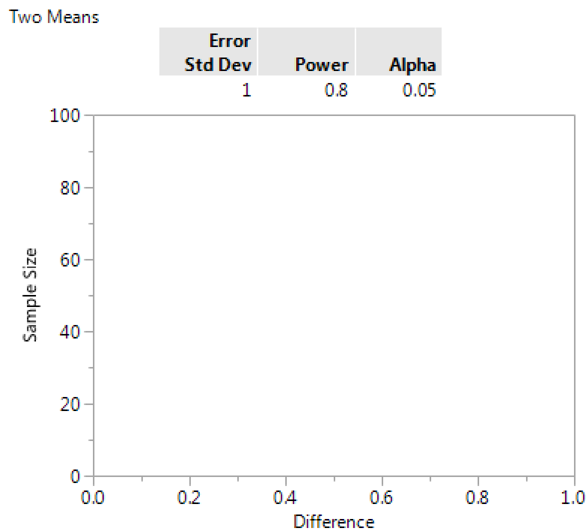


NOTE: the following problems are all on statistical power. They're together at the end because we had originally planned to put them on exam 2.

12. Define "statistical power" in English (i.e. don't write it out using formal mathematical notation).

13. Below are two plots made in JMP, with the lines erased. On each plot, draw a line that shows roughly the relationship between the variables on the axes.





14. Suppose you are preparing to conduct an experiment in which you will test for a difference in means between a control group and a treatment group. You believe that the effect size of this difference is d = 0.8. For this effect size and a sample size of n = 80, you calculate that the power of the test is 0.94.

a. Suppose the sample size could increase to n =100. Make up a plausible value for the new power of the test.

b. Suppose that the sample size decreases to n = 60. Make up an effect size that would result in power staying at 0.94.

c. Suppose you are worried that the effect size is smaller than you initially believed, and you want to know what the consequences of this would be. You calculate that, for an effect size of d = 0.5 and power = 0.8, the required sample size is n =128. Make up a sample size that would achieve power = 0.8 if the effect size were 0.6 rather than 0.5.

d. Suppose that you believe that your test will only have a power of 0.2. Why might it not be worth conducting this test in the first place?