

STAT 350 Practice Final Exam Solution (Spring 2015)

PART 1: Multiple Choice Questions:

- 1) A study was conducted to compare five different training programs for improving endurance. Forty subjects were randomly divided into five groups of eight subjects in each group. A different training program was assigned to each group. After 2 months, the improvement in endurance was recorded for each subject. A one-way ANOVA is used to compare the five training programs, and the resulting p -value is 0.014. At a significance level of 0.05, what is the appropriate conclusion about mean improvement in endurance?

a) The average amount of improvement appears to be the same for all five training programs.

☒ b) It appears that at least one of the five training programs has a different average amount of improvement.

c) The average amount of improvement appears to be different for each of the five training programs.

d) One training program is significantly better than the other four.

- 2) Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars. He collected their weight (in hundreds of pounds) and the mileage (MPG) for each car. From a scatterplot made with the data, a linear model seemed appropriate. Which of the following descriptions of the value of the slope is the correct description?

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	40.44	6.275	6.445	0
Weight	-0.521	0.164	-3.182	0.0072

a) We cannot interpret the slope because we cannot have a negative weight of a car.

b) We estimate the mileage to decrease by 0.521 miles per gallon when the weight of a car increases by 1 pound.

c) We estimate the mileage to decrease by 52.1 miles per gallon when the weight of a car increases by 100 pounds.

☒ d) We estimate the mileage to decrease by 0.521 miles per gallon when the weight of a car increases by 100 pounds.

- 3) Suppose you are testing the null hypothesis that the slope of the regression line is zero versus the alternative hypothesis that the slope is different than zero. Would a very small P -value (i.e., less than .0001) indicate a strong relationship between the explanatory variable and the response variable?

a) Yes, because the P -value will give the strength of the association between the explanatory variable and the response variable.

b) Yes, because if the P -value is small, then R^2 is large.

☒ c) No, because there could be a lot of scatter about the regression line, indicating a weak relationship between the explanatory variable and the response variable.

d) No, because a large P -value would indicate a strong relationship between the explanatory variable and the response variable.

STAT 350 Practice Final Exam Solution (Spring 2015)

- 4) In a simple linear regression model, the deviations, ϵ_i , are assumed to be _____.
a) $N(0,1)$
☒ b) $N(0,\sigma)$
c) $t(1)$
d) $t(0)$
- 5) Do students tend to improve their SAT Mathematics (SAT-M) scores the second time they take the test? A random sample of four students who took the test twice received the following scores. Assume that the change in SAT-M score (second score – first score) for the population of all students taking the test twice is Normally distributed. Suppose we do not believe that students tend to improve their SAT-M scores the second time they take the test. Based on a 90% confidence interval $(-22.56, 72.56)$, we wish to test $H_0: \mu = 0$ versus $H_a: \mu \neq 0$ at the 5% significance level. Determine which of the following statements is true.
a) We cannot make a decision since the confidence interval is so wide.
b) We cannot make a decision since the confidence level we used to calculate the confidence interval is 90%, and we would need a 95% confidence interval.
c) We reject H_0 since the value 0 falls in the 90% confidence interval.
☒ d) We accept H_0 since the value 0 falls in the 90% confidence interval and would therefore also fall in the 95% confidence interval.
- 6) A study was to be undertaken to determine if a particular training program would improve physical fitness. An SRS of 31 university students was selected to be enrolled in the fitness program. One important measure of fitness is maximum oxygen uptake. The researchers wished to determine if there was evidence that their sample of students differed from the general population of untrained subjects. The measurements made on the subjects coming into this study produced a sample mean of $\bar{x} = 47.4$ with a standard deviation of $s = 5.3$. Suppose the 98% confidence interval were determined to be $(45.2, 49.6)$ an interval. If the level of confidence were changed to 95%, what would happen to the confidence interval and the P -value?
a) The confidence interval would become longer and the P -value would decrease.
b) The confidence interval would become shorter and the P -value would increase.
☒ c) The confidence interval would become shorter but the P -value would not change.
d) The confidence interval would become shorter and the P -value would decrease.
e) The confidence interval would not change and the P -value also would not change.

STAT 350 Practice Final Exam Solution (Spring 2015)

- 7) The t distribution has properties similar to the _____ distribution.
- a) Uniform
 - ☒ b) Normal
 - c) F
 - d) Binomial
 - e) None of the above
- 8) The one-sample z statistic is used instead of the one-sample t statistic when _____.
- a) μ is known
 - ☒ b) σ is known
 - c) μ is unknown
 - d) σ is unknown
- 9) Bob has a severe cold. His roommate takes a garlic tablet every morning and has not had a cold in 2 years. Bob's aunt also has a friend who takes garlic tablets daily and has not had a cold in more than a year. Based on these data, Bob decides to start taking garlic tablets as soon as his cold clears up. What type of study is Bob's decision based on?
- a) An experiment
 - b) An observational study based on available data
 - c) An observational study based on a sample survey
 - ☒ d) Anecdotal evidence
- 10) An agricultural researcher plants 25 plots with a new variety of yellow corn. Assume that the yield per acre for the new variety of yellow corn follows a Normal distribution with an unknown mean of μ and a standard deviation of $\sigma = 10$ bushels per acre. Which of the following would produce a confidence interval with a smaller margin of error than the 90% confidence interval?
- a) Compute a 99% confidence interval rather than a 90% confidence interval, because a higher confidence level will result in a smaller margin of error.
 - b) Plant only five plots rather than 25, because five are easier to manage and control.
 - c) Plant 10 plots rather than 25, because a smaller sample size will result in a smaller margin of error.
 - ☒ d) Plant 100 plots rather than 25, because a larger sample size will result in a smaller margin of error.
- 11) Confidence intervals are useful when trying to estimate _____.
- a) unknown statistics
 - b) known statistics
 - ☒ c) unknown parameters
 - d) known parameters

STAT 350 Practice Final Exam Solution (Spring 2015)

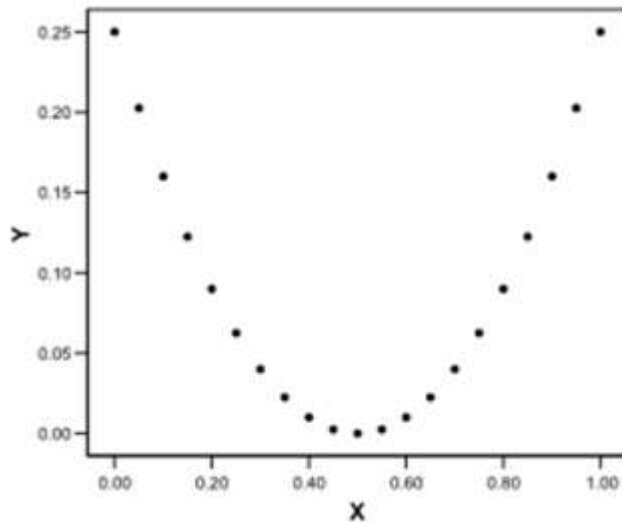
- 12) A test of significance for a null hypothesis has been conducted and the P -value is determined. Which of the following statements about a P -value is TRUE?
- a) The P -value is the probability that the null hypothesis is false.
 - b) The P -value is the probability that the alternative hypothesis is true.
 - c) The P -value is the probability that the null hypothesis is rejected even if that hypothesis is actually true.
 - d) The P -value tells us the strength of the evidence against the null hypothesis. The larger the P -value the stronger the evidence against the null hypothesis.
 - ☒ e) None of the above is a true statement about the P -value.
- 13) When choosing between a one-sided alternative hypothesis and a two-sided alternative hypothesis, you should base the decision on _____.
- a) the data
 - b) the results of the P -value
 - ☒ c) the research question
 - d) All of the above
- 14) The sample standard deviation s is a useful measure of a characteristic of a distribution of data values. Which of the following statements about s is FALSE?
- ☒ a) The standard deviation can never be zero.
 - b) The standard deviation measures the spread of the data around the mean.
 - c) The standard deviation is appropriate as a measure of spread when the mean is chosen as the measure of center.
 - d) The standard deviation is not resistant; a few outliers can make s very large.
- 15) In order to determine if smoking causes cancer, researchers surveyed a large sample of adults. For each adult they recorded whether the person had smoked regularly at any period in their life and whether the person had cancer. They then compared the proportion of cancer cases in those who had smoked regularly at some time in their lives with the proportion of cases in those who had never smoked regularly at any point in their lives. The researchers found a higher proportion of cancer cases among those who had smoked regularly than among those who had never smoked regularly. What type of study is this?
- a) An experiment but not a double-blind experiment
 - b) A double-blind experiment
 - c) A block design
 - ☒ d) An observational study

STAT 350 Practice Final Exam Solution (Spring 2015)

- 16) A researcher is studying the relationship between sugar consumption and weight gain. Twelve volunteers were randomly assigned to one of two groups. The first group of five participants was put on a diet low in sugar and the second group of the remaining seven participants received 10% of their calories from sugar. After 8 weeks, weight gain was recorded from each participant. What type of study is this?
- a) A double-blind experiment
 - b) A matched-pairs experiment
 - ☒ c) An experiment but not a double-blind nor matched pairs experiment
 - d) An observational study
- 17) Do people prefer tap or bottled water? For this study, an investigator pours two water samples into two cups and then marks the type of water on the bottom. Each subject then tastes both water samples and rates the taste on a five-point scale (1 = poor to 5 = excellent.) What type of design is this?
- a) Double-blind experiment
 - ☒ b) Matched-pairs design
 - c) Completely randomized design
 - d) Observational study
- 18) An opinion poll is to be given to a sample of 90 members of a local gym. The members are first divided into men and women, and then a simple random sample of 45 men and a separate simple random sample of 45 women are taken. What is this an example of?
- a) A block design
 - b) A double-blind simple random sample
 - c) A randomized comparative experiment
 - ☒ d) A stratified random sample
- 19) In tests of significance about an unknown parameter, what does the test statistic represent?
- a) The value of the unknown parameter under the null hypothesis
 - b) The value of the unknown parameter under the alternative hypothesis
 - ☒ c) A measure of compatibility between the null hypothesis and the data
 - d) A measure of compatibility between the null and alternative hypotheses
- 20) When trying to explain the relationship between two quantitative variables, it would be best to use a _____.
- ☒ a) scatterplot
 - b) density curve
 - c) boxplot
 - d) histogram

STAT 350 Practice Final Exam Solution (Spring 2015)

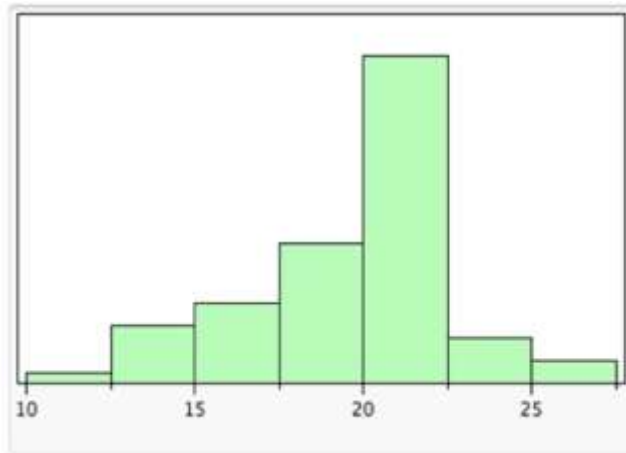
- 21) Consider the following scatterplot of two variables x and y :
What can we conclude from this graph?



- ☒ a. The correlation between x and y is not meaningful.
 - b. The correlation between x and y must be close to 1 because there is nearly a perfect relationship between them.
 - c. The correlation between x and y must be close to -1 because there is nearly a perfect relationship between them, but it is not a straight-line relation.
 - d. There is no relationship between x and y because the correlation is close to 0.
- 22) A researcher reports that on average, the participants in his study lost 10.4 lbs. after two months on his new diet. A friend of yours comments that she tried the diet for two months and lost no weight, so clearly the report must be a fraud. Which of the following statements is correct?
- ☒ a) The report gives only the average. This doesn't imply that all participants in the study lost 10.4 lbs. or even that all participants lost weight. Your friend's experience doesn't necessarily contradict the study results.
 - b) Your friend must not have followed the diet correctly because she did not lose weight.
 - c) Because your friend did not lose weight, the report must not be correct.
 - d) In order for the study to be correct, we must now add your friend's results to those of the study and recomputed the new average.

STAT 350 Practice Final Exam Solution (Spring 2015)

23) In the fuel efficiency study of 2007 compact model automobiles, the following histogram of the distribution of the miles-per-gallon fuel efficiency rating in city driving (MPG-City) for automobiles manufactured in Europe was obtained. From the histogram above, showing the distribution of MPG-City, we can see that



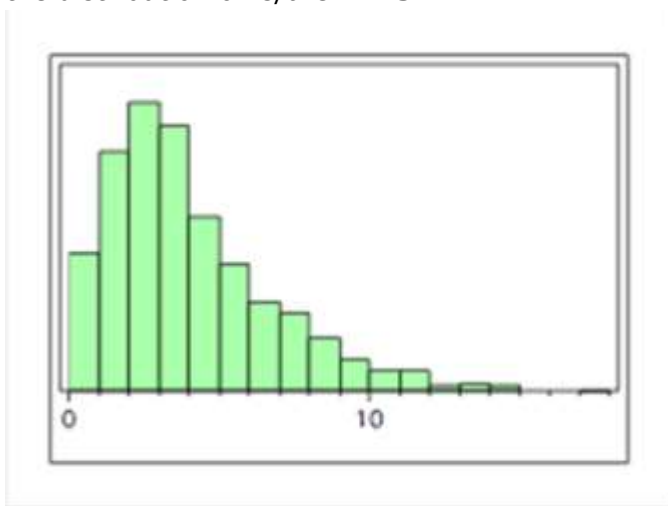
- a) the shape of the distribution is roughly symmetric with one peak.
 - b) the distribution is roughly symmetric with outlier values to the left.
 - ☒ c) the distribution is skewed to the left.
 - d) the distribution is skewed to the right.
- 24) A college newspaper interviews a psychologist about a proposed system for rating the teaching ability of faculty members. The psychologist says, "The evidence indicates that the correlation between a faculty member's research productivity and teaching rating is close to zero." What would be a correct interpretation of this statement?
- a) Good researchers tend to be poor teachers and vice versa.
 - b) Good teachers tend to be poor researchers and vice versa.
 - c) Good research and good teaching go together.
 - ☒ d) Good researchers are just as likely to be good teachers as they are bad teachers. Likewise for poor researchers.
- 25) Suppose that our sample is not representative of the population, however, our coefficient of determination, R^2 , is close to 1. The conclusion is that we have a good linear relationship between our response variable and our explanatory variable in our population.
- ☒ a) False
 - b) True

STAT 350 Practice Final Exam Solution (Spring 2015)

26) Suppose that a random sample of 50 bottles of a particular brand of cough syrup is selected and the alcohol content of each bottle is determined. Suppose that the resulting 95% confidence interval is (7.8, 9.4). We may conclude that

- a) there is a 95% chance that μ is between 7.8 and 9.4.
- b) if the process of selecting a sample of size 50 and then computing the corresponding 95% interval is repeated a very large number of times, approximately 95% of the resulting intervals will include μ .
- c) we can be highly confident that 95% of all bottles of this type of cough syrup have an alcohol content that is between 7.8 and 9.4.
- d) all of the above

27) The following histogram shows the distribution of 1000 sample observations from a population with a mean of $\mu = 4$ and a variance of $\sigma^2 = 8$. Suppose a simple random sample of 100 observations is to be selected from the population and the sample average is calculated. Which of the following statements about the distribution of is/are **FALSE**?



- a. Because the distribution shown in the histogram above is clearly skewed to the right, the shape of the distribution of \bar{X} will also show skewness to the right.
- b. The distribution of \bar{X} will have a mean of 4.
- c. The distribution \bar{X} will be approximately Normal.
- d. Even though the distribution of the population variable appears to be skewed to the right, the distribution of \bar{X} will be approximately symmetric around $\mu = 4$.
- e. The standard deviation of the distribution of \bar{X} will be 0.283.

STAT 350 Practice Final Exam Solution (Spring 2015)

Part 2: Free Response Questions.

2. (25 points) Alice's waiting time (in minutes) for the morning shuttle bus to Purdue can be modeled by a continuous distribution with density function

$$f(x) = 3(4x - x^2)/32 \text{ for } 0 < x < 4.$$

(a) (5 points) Find the probability that Alice's waiting time is more than 2.16 minutes.

Let X = Alice's waiting time in minutes

1 point	$P(X > 2.16)$
1 point	$= \int_{2.16}^4 \frac{3(4x - x^2)}{32} dx$
2 point	$= \frac{3}{32} (2x^2 - \frac{1}{3}x^3) \Big _{2.16}^4$
1 point	$= 0.440128$
	$= 0.44$

(b) (10 points) Find Alice's average waiting time.

2 points	$\mu = \int_0^4 xf(x) dx$
2 points	$\mu = \int_0^4 x \frac{3(4x - x^2)}{32} dx$
4 points	$= \frac{3}{32} (\frac{4x^3}{3} - \frac{x^4}{4}) \Big _0^4$
2 point	$= 2$

STAT 350 Practice Final Exam Solution (Spring 2015)

(c) (10 points) The standard deviation of the waiting time is 0.8 minutes. Next year, Alice will take the shuttle 100 times. What is the probability that her average waiting time will be more than 2.16 minutes?

1pts	\bar{X} is approximately Normal
1pts	with mean 2 and
3pts	standard deviation $0.8/10 = 0.08$
1pts	$P(\bar{X} > 2.16)$
2pts	$= P(Z > \frac{2.16 - 2}{0.08})$
1pts	$= P(Z > 2) = P(Z < -2)$
1pts	$= 0.0228$

If the student does not realize that this is a sampling distribution, please see an instructor.

3. (35 points) The weight gain of women during pregnancy has an important effect on the birth weight of their children. If the weight gain is not adequate, the infant is more likely to be small and will tend to be less healthy. In a study conducted in three countries, weight gains (in kilograms) of women during the third trimester of pregnancy were measured. The results are summarized in the following table:

Country	N	Mean	StDev
Egypt	46	3.7	2.5
Kenya	100	3.1	1.8
Mexico	52	2.9	1.8

Analysis of Variance				
Source	SS	MS	F-Value	P-Value
Country (Model)	17.22	8.61	xxx	0.1321
Error	767.25	xxxx		

(a) (5 points) Is it reasonable to use the assumption of equal standard deviation when analyzing these data? Give a reason for your answer.

1pts	Yes
4pts	$2.5/1.8 < 2$

(b) (5 points) Find the pooled standard deviation. Be sure to state the value of MSE.

1pts	$DFE = 46 + 100 + 52 - 3 = 195$
2pts	$MSE = SSE / DEF = 767.25 / 195 = \mathbf{3.93}$
2pts	$s_p = \sqrt{3.93} = 1.98$

STAT 350 Practice Final Exam Solution (Spring 2015)

(c) (5 points) What are the numerator and denominator degrees of freedom for the F statistic?

F(2, 206)

2pts	DF (Denominator) = DFE = 195
3pts	DF (Numerator) = 3 - 1 = 2

(d) (5 points) Find the value of the F statistic.

2pts	$F = \text{MSM}/\text{MSE}$
3pts	$F = 8.61/3.93 = \mathbf{2.19}$

3. (cont'd)

(d) (10 points) Carry out a significance test to compare the mean birth weights for the three countries at a significance level 0.05. Please follow the four step procedure and make a conclusion in context.

Let μ_1, μ_2, μ_3 be the population weight gains (in kilograms) of women during the third trimester of pregnancy in Egypt, Kenya, and Mexico respectively.

1pt	$H_0: \mu_1 = \mu_2 = \mu_3$
2pts	H_1 : not all of the μ_i are equal
1pt	$F(2, 206) = 2.19$
2pts	P-Value = 0.1321 > 0.05
2pts	Fail to reject H_0
2pts	The population weight gains of women during the third trimester of pregnancy in the three countries are the same.

If using a 5 step process (Findsen):

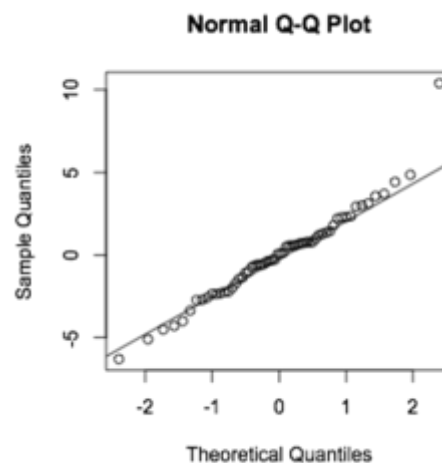
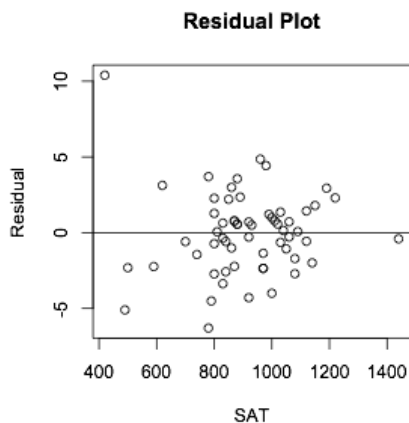
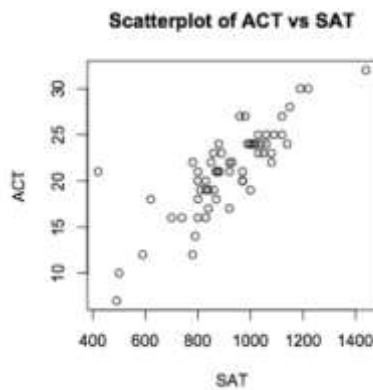
2 pt.	Let μ_1, μ_2, μ_3 be the population weight gains (in kilograms) of women during the third trimester of pregnancy in Egypt, Kenya, and Mexico respectively.
1pt	$H_0: \mu_1 = \mu_2 = \mu_3$
1pts	H_1 : not all of the μ_i are equal
1pt	$F(2, 206) = 2.19$
2pts	P-Value = 0.1321 > 0.05
1pts	Fail to reject H_0
2pts	The population weight gains of women during the third trimester of pregnancy in the three countries are the same.

STAT 350 Practice Final Exam Solution (Spring 2015)

(e) (5 points) Should a multiple comparison test be performed in this situation? Please explain your answer in one sentence.

2pts	No
3pts	We concluded that the population weight gains of women during the third trimester of pregnancy in the three countries are the same.

4. (28 points) The SAT and the ACT are the two major standardized tests that colleges use to evaluate candidates. Most students take just one of these tests. However, some students take both. Consider the scores of 60 students who took both tests. A scatterplot, residual plot, QQplot of the residuals, and regression output are shown below.



STAT 350 Practice Final Exam Solution (Spring 2015)

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.63	1.84	0.882	0.382
SAT	0.02374	0.001983	xxxxx	0.000

Residual standard error (ROOT MSE): $s = 2.744$

Multiple R-squared: 0.667

(a) (4 points) What are the conditions or assumptions for using linear regression analysis?

1	The data must be a SRS .
1	The relationship between x and y must be linear
1	The residuals must be independent and Normally distributed
1	with mean 0 and the same variance

(b) (4 points) Using the regression output, write the equation of the fitted regression line.

$ACT = 1.63 + 0.021374 \cdot SAT$ (2 points each for intercept and slope)

(c) (10 points) Find the 95% confidence interval for the slope parameter. Make sure to interpret your interval. (Assume that the conditions for the inference are satisfied.)

1	$b1 \pm t * SE(b1)$
2	$DF = n - 2 = 60 - 2 = 58$
2	$t = 2.009$ (using $df = 50$ - Findsen) or $t = 2$ ($df = 60$)
2	$0.021374 \pm 2 * \mathbf{0.001983} = 0.021374 \pm 0.003966$ (0.003984)
1	(0.0174, 0.02534)
	(0.017, 0.025)
2	We are 95% confident that the population (true) slope is between 0.017 and 0.025.

(d) (10 points) Is there an association between ACT score and SAT score? If so, does a high score on the SAT test cause a high score on the ACT test? Please explain your answer for both parts.

3	Yes, there is an association between ACT score and SAT score.
2	From the scatter plot and part (c), there is an obvious positive association between ACT score and SAT score.
3	However, a high score on the SAT test does not cause a high score on the ACT test.
2	Strong correlation does not imply causation. Student's knowledge and learning is the lurking variable.