STA 6126 Practice Problems (Old Exams)

Chapter 2

QB.1 A researcher is collecting data regarding auctions on Ebay. Are the following variables best described as **I**nterval, **N**ominal, or **O**rdinal scale?

p.1.a. Product Category (Clothing, Electronice, Auto Parts, Software) _____________

p.1.b. Number of bids made during the auction _______________

p.1.c. Size of item (Small, Medium, Large, X-Large) _________________

QB.2. **True or False:** Sampling error is avoided when we take random samples. _____

QB.3. A questionaire is sent out to academic social scientists who are members of a national association. For each of the following elements of the survey, are the responses best described as **I**nterval, **N**ominal, or **O**rdinal scale?

p.3.a. Gender ____________________

p.3.b. Annual Salary _________________________

p.3.c. Attitude toward faculty union (Strongly oppose, oppose, neutral, favor, strongly favor) _______

QB.4. What sampling plan was used in the previous question, under the following scenarios.

p.4.a. Random samples of scientists were taken from each classification of highest degree offered by college (Associates, Bachelors, Masters, Doctorate)

p.4.b. The phone directory is obtained, and from a random starting name, every 200th member is sampled.

QB.5. **True or False:** Sampling error refers to when the subjects give incorrect responses to survey questions.

QB.6. A questionaire is sent out to academic social scientists who are members of a national association. For each of the following elements of the survey, are the responses best described as **I**nterval, **N**ominal, or **O**rdinal scale?

p.6.a. Age ____________________

p.6.b. Academic Institution Type (Public, Private Non-Religious, Private Religious) _____

p.6.c. Attitude toward faculty union (Strongly oppose, oppose, neutral, favor, strongly favor) _________

QB.7. **True or False:** Sampling error refers to the fact that statistics based on samples don't equate to the corresponding population parameter.

QB.8. A questionaire is sent out to academic social scientists who are members of a national association. For each of the following elements of the survey, are the responses best described as **I**nterval, **N**ominal, or **O**rdinal scale?

p.8.a. Gender ___________________

p.8.b. Annual Salary ________________________

p.8.c. Attitude toward faculty union (Strongly oppose, oppose, neutral, favor, strongly favor) ________

QB.9. What sampling plan was used in the previous question, under the following scenarios.

p.9.a Random samples of scientists were taken from each classification of highest degree offered by college (Associates, Bachelors, Masters, Doctorate)

p.9.b. The phone directory is obtained, and from a random starting name, every 200th member is sampled.

QB.10. **True or False:** Sampling error refers to when the subjects give incorrect responses to survey questions.

QB.11. A person's attitude toward prohibition of alcohol (favor, neutral, oppose) is best described as nominal scale

QB.12. The number of children living in a household is a discrete random variable

QB.13. A sample is conducted at a small private university that has 4 colleges (Arts/Sciences, Education, Agriculture, and Health Sciences). Random samples of 25 students are selected from the each college. This is best described as a cluster sample.

QB.14. A random sample of 250 likely voters in a town is selected, and 42% of the voters favor a referendum to allow a new store to be built in the town. When the actual vote takes place on election day, 44% of all voters vote to allow the new store to be built. The fact that 42% of the sample, and 44% of the population favor the store being built is an example of sampling error.

QB.15. A researcher is interested in subjects' reactions to a new type of snack food. She has them rate the likelihood of purchasing the snack food after tasting it. The choices are: Very Unlikely, Unlikely, Likely, and Very Likely. This response is best described as Ordinal.

QB.16. Sampling error refers to the fact that numeric measures based on random samples do not equal the same numeric measure based on the entire population.

Chapter 3

QC.1. **True or False:** For a distribution that is skewed right, the median will be larger than the mean. ______

QC.2. A study considered the level of stability of postwar peace settlements for a sample of 150 disputes. The following partial table gives frequency counts of the stability levels.

Level of Stability	Frequency
Low	77
Moderate	28
High	??

p.2.a. How many disputes were classified as having a high level of stability?
p.2.b. What's the median category?
p.2.c. What proportion of the disputes were classified as having a moderate level of stability?
p.2.d Can we obtain the mean stability level? **Yes / No**

QC.3. Cotton boll lengths were measured on specimens at an archaeological site. The lengths recorded were 25.0, 19.0, 19.0, 21.0 mm for the 4 specimens.

p.3.a. Compute the sample mean and median.

p.3.b. Compute the sample standard deviation.

QC.4. A sample of 1000 college students was obtained, and each was asked the amount of time they spent last month on cell phone.

Amount of Time on cell phone	Frequency
None	200
1-10 minutes	60
11-30 minutes	100
31-60 minutes	280
61-120 minutes	230
>120 minutes	

p.4.a. How many spent over 120 minutes on the phone last month?

p.4.b. What's the median category? The modal category?

QC.5. A manager observes the amount of time spent online in a week among a sample of 5 employees, observing (in hours): 0, 2, 1, 4, 13.

p.5.a. Compute the sample mean and median.

p.5.b. Compute the sample standard deviation.

QC.6. **True or False:** A survey reported by a news agency says 45% of a random sample of adults in a city favor a supercenter being built. This number is an estimate of the population mean. ______

QC.7. A sample of 500 college students was obtained, and each was asked the amount of time they spent last month on cell phone.

Amount of Time on cell phone	Frequency
None	120
1-10 minutes	60
11-30 minutes	80
31-60 minutes	130
61-120 minutes	80
>120 minutes	

p.7.a. How many spent over 120 minutes on the phone last month?

p.7.b. What's the median category?

p.7.c. What proportion of the students that did not use a cell phone last month?

p.7.d. Can we obtain the mean amount of time? **Yes / No**

QC.8. A hurricane researcher takes a random sample of 6 hurricane seasons over a 200 year span, and determines the number of hurricanes each of the seasons. She observes 4,6,7,2,3,14 hurricanes for the 6 seasons.

p.8.a. Compute the sample mean and median.

p.8.b. Compute the sample standard deviation.

QC.9. A sample of 1000 college students was obtained, and each was asked the amount of time they spent last month on cell phone.

Amount of Time on cell phone	Frequency
None	200
1-10 minutes	60
11-30 minutes	100
31-60 minutes	280
61-120 minutes	230
>120 minutes	

p.9.a. How many spent over 120 minutes on the phone last month?

p.9.b. What's the median category? The modal category?

p.9.c. A manager observes the amount of time spent online in a week among a sample of 5 employees, observing (in hours): 0, 2, 1, 4, 13.

p.9.d. Compute the sample mean and median.

p.9.e. Compute the sample standard deviation.

QC.10. A researcher studying the size of family run farms in her state finds that the mean size is 800 acres, while the median size is 400 acres. The distribution of family farm sizes would be skewed to the right.

QC.11. A psychologist is interested in whether people tend to react positively to a particular stimulus. She conducts an experiment, observing whether sample subjects react in a positive manner. Her parameter of interest is the population proportion (π) that would react that way (among all people).

QC.12. The sample standard deviation can be affected by a small number of extreme outliers, while the inter-quartile range will not be.

QC.13. The following table gives the frequency of students by grade (lowest to highest) in a large undergraduate marketing class 2000 students. (0.0='F', ..., 4.0='A')

Grade	Frequency
0.0 (F)	100
1.0 (D)	300
2.0(C)	600
3.0 (B)	700
4.0 (A)	300

p.13.a. Give the median score.
p.13.b. Give the inter-quartile range.
p.13.c. Give the mean score (Hint: Turn the frequencies into probabilities and use $E(Y) = \Sigma y^* p(y)$)

QC.14. A clerk in a records office records the number of requests for records for a sample of 5 days. The number of requests are: 0, 2, 5, 1, 12. Compute the mean and standard deviation.

QC.15. If a distribution of measurements is skewed to the right, the median will be larger than the mean.

QC.16. The following (partial) table gives the frequency of reviews for the book *Jurassic Park* on Amazon.com based on a total of 1000 reviews. Note that a rating of 5 is best and 1 is worst.

Rating	Frequency
5*	730
4*	190
3*	40
2*	10
1*	

p.16.a. Give the median score.
p.16.b. Give the inter-quartile range.

QC.17. The number of people visiting a particular website on a sample of n=5 days were: 10, 12, 20, 16, and 22. (Show your work for all parts)

p.17.a. Compute the sample mean and median.

p.17.b. Compute the sample standard deviation.

Chapter 4

QD.1. **True or False:** For large samples, estimators of means and proportions have sampling distributions that are approximately normal. ______

QD.2. Exam scores on a standardized test are normally distributed with a mean of 830 and a standard deviation of 80. The middle 95% of students taking this exam scored between _________ and _____________

QD.3. Give the approximate sampling distribution for each of the following estimators (State the shape, mean, standard deviation, and a sketch). For each case, give the margin of error (with 95% Confidence).

p.3.a. Sample mean of a sample of n=64 students exam scores, when the population mean and standard deviation of individual scores are 1000 and 160, respectively.

p.3.b. Sample proportion of immigrants in a sample of n=900, and 20% of population are immigrants

QD.4. Profit margins among retail establishments are approximately normally distributed with a mean of 3% and a standard deviation of 4% (negative profits mean the firm lost money).

p.4.a. What proportion of firms make over 10% profit?

p.4.b. What proportion of firms lose money (have negative profits)?

p.4.c. The most successful 2.5% (.025) of firms have profits above what level?

QD.5. Give the approximate sampling distribution for each of the following estimators:

p.5.a. Sample mean of a sample of n=100, when the population mean and standard deviation of individual mesurements are 80 and 20, respectively.

p.5.b. Sample proportion of males in a sample of n=400, and 60% of population is female

QD.6. The true length of listed 10' boards cut at a mill is normally distributed with a mean of 125 inches, and a standard deviation of 2.0 inches.

p.6.a. What proportion of boards are shorter than the advertised 10' (120 inches).

p.6.b. The longest 10% of all boards exceed what length?

QD.7. Give the approximate sampling distribution for each of the following estimators (State the shape, mean, standard deviation, and a sketch):

p.7.a. Sample mean of a sample of n=400, when the population mean and standard deviation of individual mesurements are 500 and 100, respectively.

p.7.b. Sample proportion of females in a sample of n=100, and 50% of population is female

QD.8. Profit margins among retail establishments are approximately normally distributed with a mean of 3% and a standard deviation of 4% (negative profits mean the firm lost money).

p.8.a. What proportion of firms make over 10% profit?

p.8.b. What proportion of firms lose money (have negative profits)?

p.8.c. The most successful 2.5% (.025) of firms have profits above what level?

QD.9. Give the approximate sampling distribution for each of the following estimators:

p.9.a. Sample mean of a sample of n=100, when the population mean and standard deviation of individual mesurements are 80 and 20, respectively.

p.9.b. Sample proportion of males in a sample of n=400, and 60% of population is female

QD.10. Scores on a standardized exam are approximately normally distributed with a mean of 600 and standard deviation 100. What is the probability a randomly selected student scored over 750?

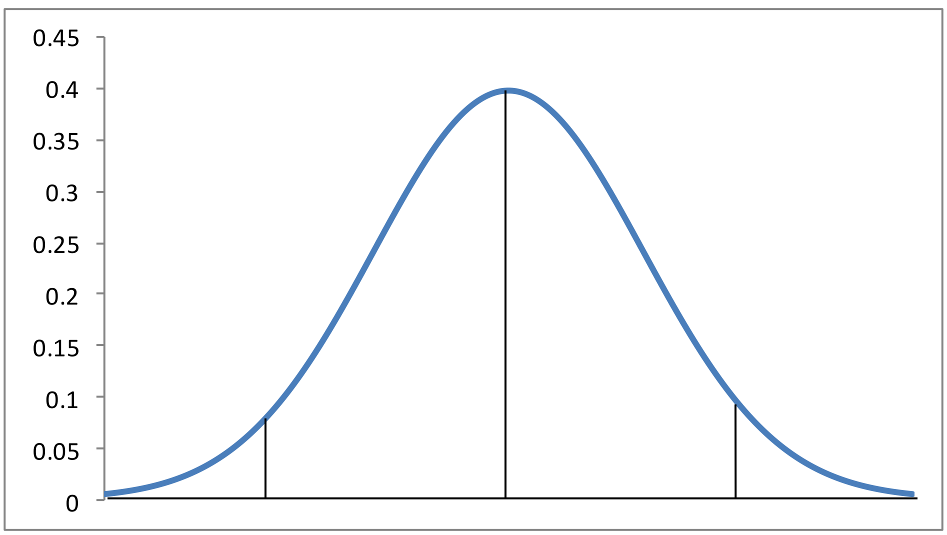
QD.11. The standard error of the sample mean based on samples of size n=100, will be twice as large as the standard error of the sample mean based on samples of size n=400.

QD.12. Scores on a standardized exam are approximately normally distributed with a mean of 1000 and standard deviation 200.

p.12.a. What is the probability a randomly selected student scored over 1250?

p.12.b. Above what score did the top 5% of scores exceed? (That is, what is the 95th-percentile of the distribution)

QD.13. In a population of consumers, 20% (0.20 as a proportion) buys a particular brand of tissue. The sampling distribution of the sample proportion, based on samples of size n=100 would be as shown below. Identify the points a,b,c on the graph, where the area between a and c is 0.95.



a b c

a = ________________ b = _______________ c = ______________________

Chapter 5

QE.1. **True or False:** All else being equal, the width of a 95% Confidence Interval for a population mean increases as the sample size increases. _______________

QE.2. Francis Galton made many anthropological measurements on people in the late 19th Century. He found that among adults (by gender), many of these were normally distributed. He found that among females, arm spans had a mean of 63 inches and a standard deviation of 3 inches.

p.2.a. What proportion of females had arm spans exceeding 65 inches?.

p.2.b. The longest 5% of female arm spans exceeded what length?

p.2.b. What is the probability that the sample mean of a random sample of n=16 women exceeds 65 inches?

QE.3. An economist is interested in estimating the population mean gas price among all gas stations in the U.S. She selects a random sample of 256 gas stations, and observes a sample mean price for regular unleaded of $2.60, and a standard deviation of $0.32.

p.3.a. Give a 95% Confidence Interval for the population mean price among all gas stations in the U.S.

p.3.b. Is it likely that the underlying distribution of prices is normal, if it's known that no stations sell gas for less than $2.30? **Yes / No**

p.3.c. Does your result from a) depend on the distribution being normal? **Yes / No**

QE.4. A sample survey asked 900 registered voters whether they were concerned about identity theft, and 513 said they were very concerned. Give a 95% Confidence Interval for the proportion of all registered voters who are very concerned about identity theft.

QE.5. Among a random sample of 64 families in a rural town, the sample mean and standard deviation of living children are 3.8 and 4.0, respectively. Give a 95% Confidence Interval for the population mean of living children among all families in the town.

QE.6. Among a random sample of 144 households in a rural town, the sample mean and standard deviation of residents are 4.2 and 3.0, respectively.

p.6.a. Give a 95% Confidence Interval for the population mean number of residents among all households in the town.

p.6.b. Is it likely that the underlying distribution of residents is normal? **Yes / No** p.6.c. Does your result from a) depend on the distribution being normal? **Yes / No**

QE.7. A drug manufacturer wishes to estimate the adverse event rate among patients who take a new drug. A sample of 400 patients are given the drug, and 30 suffer the adverse event.

p.7.a. Give a 95% Confidence Interval for the proportion suffering an adverse event among the population of all patients who could be given the drug.

p.7.b. Based on your interval, is the firm justified in claiming that less than 10% of population will suffer the adverse event ($p < .10$)? **Yes / No**

QE.8. Among a random sample of 64 families in a rural town, the sample mean and standard deviation of living children are 3.8 and 4.0, respectively. Give a 95% Confidence Interval for the population mean of living children among all families in the town.

QE.9. All else being equal, as the sample size increases, the margin of error for a sample mean increases.

QE.10. A researcher reports a 95% Confidence Interval for the mean household income of juveniles being charged with misdemeanors as ($18,000 to $28,000). This means that 95% of all such juveniles come from households with incomes in this range.

QE.11. A journalist is interested in studying the complexities of articles in a particular periodic journal. She takes a random sample of n=36 articles from the journal and measures an index of complexity for each article. The mean and standard deviation for the sample were 14.5 and 3.0, respectively. Compute a 95% Confidence Interval for the mean index of all articles published in the journal.

QE.12. All else being the same, the width of a 95% Confidence Interval for a population mean will be wider than the width of a 99% Confidence Interval.

QE.13. A news agency reports that based on a random sample of likely voters, 38% favor Candidate A in the upcoming election, with a margin of error of ± 3.5% (with 95% confidence). We can be 95% confident that the proportion of all likely voters for Candidate A is between 34.5% and 41.5%.

QE.14. A confidence interval (based on a random sample of n=400 students) for the population mean number of times students use Google per week is reported as (8.0 , 12.0). This means that 95% of students use Google between 8 and 12 times per week.

QE.15. A random sample of n=100 students at a college is obtained and each is asked how many hours they work per week at a paid job. Of those sampled, the mean and standard deviation were 8 hours and 12 hours, respectively. Obtain a 95% Confidence Interval for the mean time per week among all students at the college.

QE.16. We wish to estimate the average size of farms in North Central Florida to within 50 acres with 95% confidence. A small preliminary study suggests the standard deviation is approximately 250 acres. How large a sample is needed?

QE.17. Based on responses of 1600 subjects in General Social Surveys in the mid-1980s, a 95% Confidence Interval for the mean number of close friends equals (5.6, 7.2). For each statement, circle the best response: **Correct** or **Incorrect**.

p.17.a. If random samples of 1600 were repeatedly selected and 95% Confidence intervals computed based on each sample, then in the long run 95% of the intervals would contain μ. **Correct** or **Incorrect**?

p.17.b. We can be 95% confident that $\overline{Y}$ is between 5.6 and 7.2. **Correct** or **Incorrect**?

p.17.c. 95% of respondents had between 5.6 and 7.2 close friends. **Correct** or **Incorrect**?

p.17.d. We can be 95% confident that μ is between 5.6 and 7.2. **Correct** or **Incorrect**?

QE.18. A pharmaceutical scientist is interested in estimating the adverse event rate (π) of patients taking a psychotropic drug. She samples 100 patients at random and gives them the drug, 20 report an adverse event.

p.18.a. Give a point estimate for the adverse event rate

p.18.b. Give a 95% confidence interval for the true adverse event rate in the population

p.18.c. Using your point estimate as a "starting value" in the standard error equation, how many subjects would be needed to estimate the true adverse event rate within 0.01 with probability 0.95?

QE.19. A researcher reports a 95% confidence interval for the mean distance traveled daily to work for residents in a large city is 12 to 16 miles. The best statement to describe her result is:

a) If she repeatedly took samples of this size, her sample mean would fall in this range 95% of the time
b) 95% of the people in the city drive between 12 to 16 miles to work
c) If we repeatedly took samples of this size and constructed a confidence interval in this form, in 95% of the samples the interval would contain the true population mean
d) The population mean will be in this range 95% of the time

QE.20. All else being the same, the effect of increasing a sample by a factor of 4, has what effect on the width of a 95% confidence interval?

a) Increases by a factor of 4
b) Decreases by a factor of 2
c) Decreases by a factor of 4
d) Increases by a factor of 2

QE.21. You would like to estimate the proportion of adults in a city who favor the building of a new superstore within 0.02 (2%). How large of a random sample would you need to be 95% confident that your estimate will be within 0.02 of the proportion of all adults in the city who favor the building of the superstore.

QE.22. A random sample of 16 households in a rural village in an underdeveloped country yields a mean weekly income of $15 and a standard deviation of $4. Assuming incomes are approximately normally distributed, give a 95% confidence interval for the mean income of all households in the village.

QE.22. Two graduate students obtain individual random samples of members of the Graduate Student Union at their large university. They ask each sampled student whether they favor a candidate for the president of their union. Satchel samples 100 students, of which 45 favor the candidate. Lee samples 200 students, of which 110 favor the candidate. Whose 95% confidence interval for the true proportion of all graduate students will be narrower (more precise)? Why? Assume that the university has over 10,000 graduate students so that the sample sizes are negligible relative to the population.

a) Satchel's because of a lower proportion
b) Satchel's because of a lower sample size
c) Lee's because of a higher proportion
d) Lee's because of a higher sample size
e) They will be the same, since .45(1-.55) = .55(1-.45)

QE.23. You would like to estimate the mean farm size among farms in a developing country within 15 acres with probability 0.95. A preliminary study suggests a standard deviation of 100 acres. How large a sample of farms is required?

QE.24. A cable news broadcast reports that 45% of registered voters feel a certain way about an issue. Further they report a margin of error of ±4.5%. What size sample was their survey based on?

QE.25. All else being equal, as the sample size increases by a multiple of 4, the width of a 95% Confidence Interval for μ will decrease by one-half. (T/F)

Chapter 6

QF.1. The *P*-value for testing H_0: $\mu=100$ vs H_A: $\mu\neq100$ is 0.001 and $\overline{Y} < 100$. This indicates strong evidence that (select the correct answer):

a) $\mu=100$ b) $\mu>100$ c) $\mu<100$

QF.2. A large-sample 95% confidence interval for $\mu_1-\mu_2$ is (-2.0, 7.0). The *P*-value for testing H_0: $\mu_1-\mu_2 = 0$ vs H_A: $\mu_1-\mu_2 \neq 0$ based on the same sample data is (select the correct answer):

a) above 0.05 b) below –0.05 c) below 0.05 d) Need more information

QF.3. Researchers reporting a Z-test to test the alterntative hypothesis that the proportion of males favoring a candidate is larger than the proportion of females favoring the candidate (H_A:π_M-π_F > 0) give a *P*-value of 0.0643. Their computed test statistic was:

a) 0.0643 b) 1.52 c) 1.85 d) 0.9357 e) We need more information

QF.4. For each of the following problems, give the null and alternative hypotheses (be very careful in notation with respect to symbols and subscripts)

p.4.a. Researchers believe that when faced with the question: "Do you have a favorable attitude toward President Bush"?, men are more likely to respond "Yes" than women.

p.4.b. A researcher wishes to demonstrate that the mean time to read a paragraph differs, depending on the color of the paper that the paragraph is printed on (White vs Red).

QF.5. For a large-sample test of H_0: $\mu = 0$ vs H_A: $\mu < 0$, the z-test statistic is 0.40. Give the P-value:

a) .3446
b) .6892
c) .6554
d) .8446

QF.6. A family researcher is interested in testing a claim that a majority of couples living together are not married. Letting π be the proportion of all households with two adults co-habitating who are not married, the null and alternative hypotheses are: H_0: $\pi = 0.5$ and H_A: $\pi > 0.5$. She samples 500 couples and finds that 265 of them are not married.

p.6.a. Compute the appropriate test statistic:

p.6.b. Compute the P-value:

p.6.c. If the test is being conducted at the α=0.05 significance level, will we conclude that a majority of couples living together are not married?

i) Yes   ii) No   iii) Need more information

QF.7. A Type II Error occurs when a hypothesis test results in rejecting the null hypothesis when in fact it is true. (T/F)

QF.8. An archaeologist wishes to test whether the true mean diameter of certain items found at an excavation site differ from 2". She randomly samples 15 of the items at a site and measures them. She will conclude that the true mean differs from 2 at the 0.05 significance level if the absolute value of her t-statistic is greater than or equal to 1.761. (T/F)

QF.9. If the test statistic and P-value for testing H_0: μ=100 against H_a: $\mu \neq 100$ are t=3.29 and P=0.001, respectively, then we have strong evidence that μ>100. (T/F)

QF.10. A researcher wishes to test whether the recidivism (re-arrest) rate for juveniles from low income households exceeds 50% (0.5). The test statistic for testing H_0: $\pi = 0.5$ against H_a: $\pi > 0.5$ is $z_{obs} = 2.36$ based on a large sample of such youths. Report the P-value for this test.

Chapter 7

QG.1. For each of the following problems, state whether the samples are best described as independent or dependent (paired) samples and whether the responses are numeric or nominal for each subject.

p.1.a. A study is conducted to determine whether two training methods differ for employees learning a new computer system. Among new employees who have no experience with the system, half receive training method 1, the others receive training method 2, and a skills test based on the number of correctly completed of 20 tasks is given after training.

p.1.b. A wine company has a sample of subjects taste their wine twice. Once when the bottle is said to be \$15, once when it's said to be \$8. Under each condition, they are asked if they like the taste of the wine. (Yes/No). Subjects are not aware the wine is the same under the two conditions.

QG.2. A social scientist wants to determine whether average clothing expenditures differ among two cultural groups. They randomly sample 12 individuals from each group, obtaining annual clothing expenditures from each (incomes are similar among all surveyed individuals, but they are NOT paired). The following table gives the sample statistics (assume population standard deviations are equal):

	Grp 1	Grp 2
n	12	12
Mean	5000	3000
Std Dev	1000	800

p.2.a. Give the test statistic for testing H_0: μ_1-μ_2 = 0 vs H_A: μ_1-$\mu_2 \neq 0$

p.2.b. Sketch the *P*-value, and clearly identify the distribution and values that correspond to the P-value, based on your answer to Part a).

p.2.c. For your test and test statisic in Part a), you will conclude (based on α=0.05):

$\mu_1 > \mu_2$ if the test statistic falls in the range: _______________

$\mu_1 < \mu_2$ if the test statistic falls in the range: _______________

μ_1 and μ_2 are not significantly different if test statistic is in range:_____________

QG.3. The following output gives SPSS output from a **paired difference t-test**, to determine whether Botox reduces intensity of migraine scores after treatment, compared to before treatment. Data for each of n=29 subjects is the difference between pre- and post-treatment migraine intensity scores.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean

Pair 1	PREINT	8.862	29	.9533	.1770
	POSTINT	3.03	29	3.669	.681

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	PREINT - POSTINT	5.828	3.9827	.7396	4.313	7.343	7.880	28	.000

We wish to test H_0: $\mu_D = 0$ vs H_A: $\mu_D \neq 0$, give the following quantities/conclusions:

$\overline{D}$ = ________________

s_D = _________________

$s_D/\sqrt{n}$ = ________________

t_{obs} = _________________

P-value: _____________________

95% CI for μ_D: __________________

Conclusion (α=0.05): i) Botox has no effect on migraine intensity.
ii) Botox increases migraine intensity
iii) Botox decreases migraine intensity

QG.4. A movie producer wants to determine whether a movie preview changes movie fans intentions to view the producer's new film. Movie fans are exposed to a brief description of movie: title, director, stars. They are then asked whether they believe they will view the film. After having been exposed to other questions/stimuli, the subjects see a preview of the movie, and are again asked viewing intentions. The following table gives the results.

Before Preview \ AfterPreview	Will Go	Won't Go
Will go	100	40
Will not go	20	200

Use the McNemar Z-test to determine whether the probability of a viewer will say they will go to the movie is different before or after they see the preview, based on α=0.05 significance level. (H_0: $\pi_B = \pi_A$ vs H_A: $\pi_B \neq \pi_A$)

p.1.a. Test Statistic:

p.1.b. P-value:

p.1.c. Conclusion (Circle one):

i. Cannot conclude proportions saying they will go differ
ii. Conclude more likely to say they will go before seeing preview
iii. Conclude more likely to say they will go after seeing preview

QG.5. Independent random samples of 500 males and 500 females were obtained, and each asked whether they believe in afterlife. Of the males, 340 replied Yes, while 420 females replied Yes. Labelling π_M the population proportion of all adult males who believe in the afterlife, and π_F the propulation proportion of all adult females believing in the afterlife, compute a 95% Confidence Interval for π_M-π_F.

QG.6. A test that the proportions of males and females favoring a change in a city ordinance are the same results in a *P*-value of 0.003. The 95% confidence interval for the difference ($\pi_M - \pi_F$) will:

f) Contain 0
g) Be entirely positive if the sample proportion of males is higher
h) Be entirely negative if the sample proportion of males is higher
i) Be unknown as to whether it includes 0

QG.7. A study is being conducted to determine whether a new formulation of whitening mouthwash works better than a current formulation. Subjects are randomly assigned to the two formulations, and after 12 weeks, a measure of whiteness of teeth is made (higher scores are representative of whiter teeth). The appropriate null and alternative hypoitheses are:

a) H_0: $\mu_{NEW} > \mu_{OLD}$ H_A: $\mu_{NEW} = \mu_{OLD}$
b) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} < \mu_{OLD}$
c) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} > \mu_{OLD}$
d) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} \neq \mu_{OLD}$

QG.8. A study is conducted to observe whether a method of teaching drawing improves people's drawings. A sample of 5 individuals is obtained, and each is asked to draw an image before and after the training session. Their drawings are rated by a team of evaluators who are unaware of whether the drawings were done before or after the session (their rating scale ranges from 0 to 25). The scores are given below:

Student	**Before**	**After**
Alex	14	19
Chris	8	14
Sandy	13	17
Sam	17	23
Dana	14	18

a) Obtain the mean and standard deviation of the differences (Before-After)
b) Compute a 95% Confidence Interval for the true mean difference.

QG.9. A clinical trial comparing two antidepressants has 1000 patients receiving each drug (independent samples). Of those receiving Drug A, 640 show a distinct level of improvement. Of those receiving Drug B, 560 show level of improvement, Test whether the improvement rates of the two drugs differ by completing the following parts:

p.9.a. Null Hypothesis:

p.9.b. Alternative Hypothesis:

p.9.c. Test Statistic:

p.9.d. *P*-value:

p.9.e. If you use an α=0.05 significance level, what will you conclude?

i. Drug A has higher improvement rate
ii. Drug B has higher improvement rate
iii. Cannot conclude the improvement rares differ

QG.10. A study was conducted to determine the effect of priests administering a test of religious attitudes versus teachers. Independent random samples of children in parochial schools were administered the tests (100 by priests, 100 by teachers). The following table gives the results. Test whether there is a priest (versus teacher) effect on scores. That is, test whether the true population mean is higher for priests.

	Priests	**Teachers**
Mean	20.0	18.5
Std Dev	5.0	6.0
Sample Size	100	100

p.10.a. Null Hypothesis:

p.10.b. Alternative Hypotheis:

p.10.c. Test Statistic:

p.10.d. P-value:

QG.11. A research article conducts a significance test to determine whether the mean food expenditures differ among families in two societies. Their 95% Confidence Interval for $\mu_2 - \mu_1$ is (32 , 48). Which of these statements best describes the P-value for testing H_0: $\mu_2 - \mu_1 = 0$ vs H_A $\mu_2 - \mu_1 \neq 0$:

a) P-value ≥ 0.05
b) P-value ≥ 0.95
c) P-Value < 0.05
d) We need more information to obtain the P-value

QG.12. A study is being conducted to determine whether a new formulation of whitening mouthwash differs from a current formulation (researchers leave the possibility that it may be better, but may be worse). Subjects are randomly assigned to the two formulations, and after 12 weeks, a measure of whiteness of teeth is made (higher scores are representative of whiter teeth). The appropriate null and alternative hypotheses are:

a) H_0: $\mu_{NEW} \neq \mu_{OLD}$ H_A: $\mu_{NEW} = \mu_{OLD}$
b) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} < \mu_{OLD}$

c) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} > \mu_{OLD}$
d) H_0: $\mu_{NEW} = \mu_{OLD}$ H_A: $\mu_{NEW} \neq \mu_{OLD}$

QG.13. A study compares members of two political parties in terms of the amount of time watching national news. Random samples of 100 voters are taken from each party, and for party A, the mean and standard deviation are 42 minutes per week and 68 minutes per week, respectively. For party B, they are 56 minutes and 74 minutes respectively.

p.13.a Give a single value that "best" estimates $\mu_A - \mu_B$:

p.13.b. Give the estimated standard error of the answer in part a).

p.13.c. Give a 95% Confidence Interval for $\mu_A - \mu_B$

QG.14. A theory says that people generally view risk differently when it is stated as a gain versus a loss. A priori, the researchers don't have a direction in mind for $\pi_{POS}-\pi_{NEG}$ where π_{POS} is the true proportion of people favoring the gamble (Choice A) in the following situation:

CHOICE A: Win \$1000 with probability ½, Win \$0 with probability ½

CHOICE B: Win \$500 with certainty

and π_{NEG} is the true proportion favoring the gamble (Choice A') when it is framed as a loss:

CHOICE A': Lose \$1000 with probability ½, Lose \$0 with probability ½

CHOICE B': Lose \$500 with certainty

p.14.a. Give the appropriate null and alternative hypotheses

p.14.b. Two samples of 100 people each are obtained. Of the sample exposed to Choices A and B (positive frame), 72 selected Choice A. Of the sample exposed to Choices A' and B' (negative frame), 38 selected Choice A'. Compute the test statistic.

p.14.c. Complete the decision rule, based on an α=0.05 significance level: Reject H_0 if the test statistic…

QG.15. When testing H_0: $\mu_2 - \mu_1 = 0$ versus H_A: $\mu_2 - \mu_1 \neq 0$ if the sample means are equal, the P-value is 0. (T/F)

QG.16. A 95% Confidence Interval for $\mu_2 - \mu_1$ contains 0, the P-value for testing H_0: $\mu_2 - \mu_1 = 0$ versus H_A: $\mu_2 - \mu_1 \neq 0$ will be greater than .05. (T/F)

QG.17. All else being equal, as we increase the sample sizes, the width of a 95% confidence interval for $\pi_2 - \pi_1$ will increase. (T/F)

QG.18. Researchers report a t-statistic based on the paired difference t-test to be -2.40. Their sample is based on 16 pairs. They are justified in claiming the P-value for testing H_0: μ_D=0 against H_a: μ_D≠0 is less than 0.01. (T/F)

QG.19. Researchers report a 95% confidence interval for the difference between 2 population means to be (4.5, 7.8). The P-value based on a significance test will be greater than 0.05. (T/F)

QG.20. McNemar's test is appropriate when we have paired measurements on a qualitative variable with 2 levels. (T/F)

QG.21. Two movie reviewers are being compared with respect to their rates of giving positive reviews. A sample of 200 movies that each reviewer has rated is obtained, giving the following cross-tabulation. Give the test statistic and P-value for testing H_0: $\pi_A=\pi_B$ against H_A: $\pi_A \neq \pi_B$ where π_A and π_B are the true proportions of positive reviews for reviewers A and B.

Reviewer A\B	**Positive**	**Not Positive**
Positive	**60**	**30**
Not Positive	**40**	**70**

QG.22. A study is conducted to determine whether there is an effect of a new pricing policy on the average number of visitors to a state park. The wildlife officer randomly sampled 12 days prior to the price increase, and 12 days after the price increase and determined the number of visitors from the "sign-in sheets." She computes the following statistics from the records (she finds no evidence against the variances being equal for the 2 periods).

$$\overline{y}_{Before} = 370 \quad \overline{y}_{After} = 350 \quad s_p\sqrt{\frac{1}{12}+\frac{1}{12}} = 15.0$$

p.22.a. Obtain a 95% confidence interval for the population mean difference between the 2 time periods (μ_{Before} - μ_{After})

p.22.b. Based on your confidence interval, would the P-value for testing H_0: $\mu_{Before} - \mu_{After} = 0$ Against H_a: $\mu_{Before} - \mu_{After} \neq 0$ be **Larger** or **Smaller** than 0.05?

QG.23. A study was conducted to measure the effect of a new training program for new employees at a large company. A sample of 16 new employees were selected and given a test regarding ethics in the workplace before and after a 1-day training session on ethics was given. For each employee, the difference between the scores (After – Before) was obtained. The mean and standard deviation of these differences were 12.0 and 6.0, respectively.
Test to determine whether the training course is effective in increasing true mean scores:

p.23.a. Null Hypothesis: μ_D 0 Alternative Hypothesis: μ_D 0

p.23.b. Test Statistic:

p.23.c. Rejection Region:

p.23.d. Do we conclude that the training course is effective in increasing true mean scores at the α=0.05 significance level? **Yes** or **No**

QG.24. A study considered the effect of executive succession circumstances in small companies. Of 19 companies with planned successions (e.g. long-planned retirement of chief executive), 12 had undiminished profitability and 7 had diminished profitability. Of 17 companies without planned successions (e.g. untimely death or legal problems of chief executive) 3 had undiminished profitability and 14 had diminished profitability. Which method of analysis would be most appropriate for this study.

a) McNemar's Test
b) Linear Regression
c) Gamma / Kendall's Tau
d) Fisher's Exact Test

QG.25. A researcher is interested in comparing married men's and women's attitudes toward same-sex marriage. A random sample of 1000 married couples is obtained. Of these couples, both the male and female opposed it for 200 of the couples, both the male and female favored it for 300 of the couples. The male favored and the female opposed it for 260 of the couples. For the remaining couples the male opposed and the female favored.

p.25.a. Give the test statistic for testing whether the proportions of males and females favoring same sex marriage are the same (H_0) or differ (H_A)

p.25.b. What do we conclude at the α=0.05 significance level?

- Do not conclude the proportions differ
- Conclude a higher proportion of males favor same-sex marriage
- Conclude a higher proportion of females favor same-sex marriage

Chapter 8

QH.1. A chi-square test is being conducted to compare 3 groups' distributions across a response variable with 4 categories. We will conclude there is an association at the α=0.05 significance level exceeds 12.59 (T/F)

QH.2. A content analysis was conducted to compare the reporting of 3 news organizations. Random samples of 50 stories were obtained from each organization, and each was classified by their attitude toward the current government: Positive, Neutral, or Negative. The table of observed frequencies is given below.

fo	Positive	Neutral	Negative	Total
A	30	10	10	50
B	10	30	10	50
C	10	10	30	50
Total	50	50	50	150

p.2.a. Obtain the expected count for each cell under the null hypothesis of no association between news organization and attitude toward the current government. Hint: It is the same for each of the 9 cells.

p.2.b. The chi-square statistic is X_{obs}^2 = 16. Can we conclude that there is an association between news organization and attitude toward the current government at the 0.05 level? Why?

QH.3. Interchanging two rows in a contingency table will not have an effect on gamma. **True or False.**

QH.4. Interchanging two rows in a contingency table will not have an effect on the chi-squared statistic. **True or False.**

QH.5.When conducting a test for independence, a cell in the table has a large positive adjusted residual. This is:

a) Consistent with the null hypothesis of independence
b) The observed cell count is higher than expected if the variables were independent
c) The observed cell count is higher than expected if the variables were independent

QH.6. A test for independence is conducted where the nominal explanatory variable has 4 levels and the nominal response variables has 3 levels. If we conduct the chi-square test at the α=0.05 significance level, we will conclude the variables are not independent (dependent) if the test statistic:

a) < 12.59 b) < 21.03 c) > 12.59 d) > 21.03 e) < 16.92 f) > 16.92

QH.7. A researcher is interested in the distribution of college professors party affiliations. She samples professors at public universities from each of the following colleges: medicine, law, engineering, and liberal arts/sciences. The following table gives the conditional distributions (percentages) for party affiliations within college.

College\Party	Independent	Republican	Democrat	Total % (cases)
Medicine	5%	50%	45%	100% (n=200)
Law	10%	40%	50%	100% (n=150)
Engineering	2%	58%	40%	100% (n=100)
Arts/Sciences	15%	30%	55%	100% (n=200)
Marginal Distribution (%)	8.77%	42.77%	48.46%	100% (n=650)

p.7.a. The researcher wishes to test whether the distribution of party affiliation is independent of college (H_0) or that it is dependent of college (H_A).

p.7.b. Give the observed cell count (f_o) for independent party affiliation among medicine professors.

p.7.c. Give the expected cell count under H_0 (f_e) for independent party affiliation among medicine professors.

p.7.d. Give the contribution to the chi-square statistic for the cell corresponding to independent party affiliation among medicine professors.

QH.8. A sample of Gator fans are classified on the frequency they listen to Sports talk radio (never, occasionally, frequently) and their attitude toward coach Ron Zook (unfavorable, neutral, favorable). The following SPSS output gives gives the following results for the measure of association gamma.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	-.432	.052	-7.803	.000
N of Valid Cases		450			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

p.8.a. Give a 95% Confidence interval for Gamma based on the population of all Gator fans

p.8.b. How would you describe the results? **Circle one**.

i) The more people listen to talk radio, the less favorable they are to Zook.

ii) The more people listen to talk radio, the more favorable they are.

iii) People's attitudes toward Zook are independent of how much they listen to talk radio.

QH.9. You are conducting a literature view in your field of study. The author of an article has conducted a chi-squared test for independence (Nominal variables). He has 2 groups he's comparing and 3 possible outcomes. He reports a chi-square statistic of 5.50 and a P-value of 0.025. What do you conclude?

a) He has understated P-value b) He has overstated P-value c) Cannot determine

QH.10. Interchanging the rows (levels of the independent variable) in a contingency table will:

a) Effect the chi-square statistic, but not the numbers of concordant/discordant pairs
b) Effect the numbers of concordant/discordant pairs, but not the chi-square statistic
c) Effect both the chi-square statistic and the numbers of concordant/discordant pairs
d) Effect neither the chi-square statistic or numbers of concordant/discordant pairs

QH.11. Two researchers conduct a chi-square test with contingency tables of the exact same dimensions (say r=2 and c=3). Jack sampled 100 individuals from each of the 2 populations and Jill sampled 200 individuals from each population. They obtained the exact same conditional distributions for each of the two populations (although the conditional distribution for population 1 is not identical to the conditional distribution for population 2).

a) Their chi-square statistics will be the same
b) Jack's chi-square statistic will be twice as large as Jill's
c) Jill's chi-square statistic will be twice as large as Jack's
d) Jill's chi-square statistic will be 4 times as large as Jack's
e) Jack's chi-square statistic will be 4 times Jack's chi-square statistic will be twice as large as Jill's as large as Jill's

QH.12. A study compared St. John's Wort (SJW), Sertraline, and placebo in patients with major depressive disorder. Patients were assigned at random to one of the three treatments and were classified as having any response or no response. The contingency table is given below.

Trt \ Outcome	**Any Response**	**No Response**	**Total**
SJW	**43**	**70**	**113**
Sertraline	**53**	**56**	**109**
Placebo	**50**	**66**	**116**
Total	**146**	**192**	**338**

p.12.a. Give the conditional distributions for each treatment and overall.

Trt \ Outcome	Any Response	No Response	Total
SJW			100%
Sertraline			100%
Placebo			100%
Total			100%

p.12.b. Give the expected count for Any Response among SJW patients under the hypothesis of no association between response and treatment.

QH.13. A study considered the association between governors' strength and their control over state agencies. State administrators were classified by the strength of their state governor (very weak, weak, moderately strong, strong) and were asked who had control over their agency (legislature, about equal between legislature and governor, governor). The following table contains the measures gamma and Kendall's tau based on this sample.

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.177	.028	6.250	.000
	Gamma	.255	.040	6.250	.000
N of Valid Cases		865			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

p.13.a. Obtain a 95% confidence interval for the true (population) value of Kendall's tau.

QH.14. A study considered the occurrence of upholstery in households in Philadelphia during 4 time periods. The following table gives a cross-tabulation of period by upholstery for samples of households within the periods.

PERIOD * UPHOLSTE Crosstabulation

Count

	UPHOLSTE		
	No	Yes	Total
1	61	19	80

PERIOD	2	68	14	82
	3	64	18	82
	4	53	27	80
Total		246	78	324

p.14.a. Give the expected number of Yes in Period 1 under the hypothesis that upholstery occurrence is independent of period.

p.14.b. Give the contribution for that cell to the chi-square statistic.

QH.15. A study is conducted to determine whether there is an association between perceived price and quality assessments. A sample of 150 wine tasters was obtained and 50 were told it was Low price, 50 told it was Medium price, and 50 told it was High price. Each taster rated the wine on a 3-point scale (1=Lowest, 3=Highest). Note that actually everyone was tasting the same wine. The following table gives the results.

Price\Quality	**Low Quality**	**Med Quality**	**Hi Quality**
Low price	**28**	**15**	**7**
Medium Price	**14**	**24**	**12**
High Price	**9**	**17**	**24**

- Give the numbers of concordant and discordant pairs and the estimate of gamma.
- The estimated standard error of the estimated gamma is 0.092. Give a 95% Confidence Interval for the population-based value of gamma. Can you conclude that there is a positive association between perceived quality ratings and price?

Chapter 9

QI.1. Interchanging the explanatory and response variables will not have an effect on the correlation coefficient, *r*. **True or False.**

QI.2. A researcher reports that her regression model "explains" 50% of the variation in her dependent variable. **Which one of the following statements must be true**.

a) b = 0.50

b) r = 0.25

c) $\hat{\sigma} = 50$

d) $r^2 = 0.50$

QI.3. A regression model is fit relating % party vote (Y) to party unlikeness score (X) in votes in the U.S. Congress. The following SPSS output gives the results of the simple linear regression model being fit. **Note: High values of Y mean the two parties are more polarized (majority of one party voted one way, majority of other party voted other way). The party unlikeness score is scaled so that the more the parties differ, the higher X will be.**

Model: $E(Y) = \alpha + \beta X$ Conditional Standard Deviation: σ

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.955[a]	.913	.911	4.1967

a. Predictors: (Constant), UNLIKE

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10139.207	1	10139.207	575.682	.000[a]
	Residual	968.688	55	17.613		
	Total	11107.895	56			

a. Predictors: (Constant), UNLIKE

b. Dependent Variable: PRTYVOTE

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	10.215	2.131		4.793	.000	5.944	14.486
	UNLIKE	1.045	.044	.955	23.993	.000	.958	1.132

a. Dependent Variable: PRTYVOTE

p.3.a. Give the fitted equation: ____________________

p.3.b. Give the estimate of σ ____________________

p.3.c Give the proportion of variation in Y "explained" by X _____________

p.3.d Give a 95% confidence interval for β _______________

p.3.e. Give the coefficient of correlation _____________

p.3.f. Give the P-value for testing H_0: $\beta=0$ vs H_A: $\beta \neq 0$ ________________

QI.4. A researcher is interested in relating dose of a cough lozenge to clarity of speech among people suffering from a cold. Doses are varied along the range of X=0,2,4,6 mg. The following quantities are reported by research assistant (Y=clarity of speech measurement) based on a spreadsheet analysis of the data:

$$\sum(X-\overline{X})^2 = 100 \qquad \sum(X-\overline{X})(Y-\overline{Y}) = 700 \qquad \sum(Y-\overline{Y})^2 = 10000$$

p.4.a. Compute the estimate of increase in average clarity per unit increase in dose.

p.4.b. Compute the correlation coefficient.

QI.5. A linear regression model is fit, relating the general merchandise sales (in millions) in cities to their population (in 1000s). A model is fit and the estimated regression equation is 1.0+0.80x. As x increases by 1 unit (1000 people), general merchandise sales increase by approximately?

a) \$0.80 b) \$8000 c) \$800,000 d) \$1.80 e) \$800

QI.6. The slope of the least squares prediction equation and the Pearson correlation coefficient are similar in the sense that (Circle any that are True):

- They do not depend on the units of measurement
- They both must fall between -1 and +1
- They both have the same sign
- They both have the same *t* staistic value for testing H_0: Independence

QI.7. One can interpret $r = 0.40$ as (Circle any that are True)

- A 16% reduction in (total squared) error occurs when using *X* to predict *Y* as opposed to using $\overline{Y}$ to predict *Y*
- A 40% reduction in (total squared) error occurs when using *X* to predict *Y* as opposed to using $\overline{Y}$ to predict *Y*
- 16% of the time $\hat{Y} = Y$
- *Y* changes on average, by an estimated 0.40 units, for a one-unit increase in *X*
- When *X* is used to predict *Y*, the “typical” residual is 0.40.

QI.8. You find the following partial ANOVA table in the grad student computer lab in your department. Unfortunately, someone spilled coffee on it, and some values are unreadable. Complete the table from a simple linear regression model.

SOURCE	DF	SUM SQUARES	MEAN SQUARE	F
REGRESSION	1			
RESIDUAL		1200.0		
TOTAL	24	2000.0		

QI.9. An anthropologist is interested in the (linear) relationship between body length and head volume among remains found at an excavation sight. She does not feel that changes in one variable "cause" changes in the other, but does believe that there will be a positive linear relationship between the two. Her appropriate null and alternative hypotheses would be (Choose one):

a) H_0: b = 0 H_A: b > 0
b) H_0: r = 0 H_A: r > 0
c) H_0: $\beta = 0$ H_A: $\beta > 0$
d) H_0: $\rho = 0$ H_A: $\rho > 0$

QI.10. The conditional standard deviation, σ, represents the variation in *Y* values at the same level of *X* in the simple linear regression model.

TRUE **FALSE**

QI.11. The following SPSS output gives results of a simple linear regression relating annual food expenditures (*Y*) to household size (*X*) for a sample of households in an urban area of a country.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.973(a)	.947	.947	.68816

a Predictors: (Constant), hhsize

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	837.562	1	837.562	1768.644	.000[a]
	Residual	46.409	98	.474		
	Total	883.971	99			

a. Predictors: (Constant), hhsize
b. Dependent Variable: foodcost

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.407	.154		9.118	.000
	hhsize	1.886	.045	.973	42.055	.000

a. Dependent Variable: foodcost

Fill in the following values:

Total Sum of Squares $TSS = \sum\left(Y - \overline{Y}\right)^2$ ____________________________

Estimated Slope of Regression line $b = \dfrac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$ ______________________________

Correlation Coefficient $r = \dfrac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 (Y - \overline{Y})^2}}$ _________________________________

Estimated conditional standard deviation $\hat{\sigma} = \sqrt{\dfrac{\sum\left(Y - \hat{Y}\right)^2}{n-2}}$ ______________________________

Estimated standard error of slope of regression line $\hat{\sigma}_b = \dfrac{\hat{\sigma}}{\sqrt{\sum (X - \overline{X})^2}}$ _________________________

QI.12. A researcher fits a simple linear regression model, relating condominium price (in $1000s) to the number of bedrooms among condos of similar age in a community. He fits the analysis on a random sample of *n* = 24 condos of similar age in his town. The results of the analysis are given below:

$\hat{Y} = 80.0 + 40.0X \quad \sum\left(Y - \hat{Y}\right)^2 = 198 \quad \sum\left(X - \overline{X}\right)^2 = 16$

p.12.a. Give the fitted (predicted) price for a condo with 2 bedrooms.

p.12.a. Give a 95% Confidence Interval for the change in average condo price as the number of rooms increases by 1.

QI.13. A researcher publishes the results of a correlation analysis between two variables. The correlation coefficient based on the sample data is *r* = 0.25, based on a sample of *n* = 20 cases. Test whether we can conclude there is an association between the two variables at the 0.05 significance level. $H_0: \rho = 0$ vs $H_A: \rho \neq 0$:

p.13.a. Test Statistic:

p.13.b. Decision Rule: Reject H_0 if the absolute value of the test statistic is _________________________________

p.13.c. P-value (Circle one): **< 0.05** **> 0.05**

QI.14. A simple linear regression model is fit on a computer spreadsheet and the following quantities are obtained:

$\sum(X-\overline{X})(Y-\overline{Y})=-2400 \quad \sum(X-\overline{X})^2=10000 \quad \sum(Y-\overline{Y})^2=625$

Compute the estimated slope of the regression line (*b*) and the estimated correlation coefficient (*r*):

QI.15. For each of the following problems based on a simple linear regression model, give the elements of a test of:

H_0: *Y* is independent of *X* ($\beta = 0$) versus H_A: *Y* is not independent of *X* ($\beta \neq 0$)

The estimated regression coefficient, its standard error, and sample size are: $b=25.0 \quad \hat{\sigma}_b=10.0 \quad n=12$

Test Statistic:

Decision Rule (α = 0.05): Reject H_0 if absolute value of test statistic is greater than or equal to ___________

Conclusion: **Reject H_0** **Fail to Reject H_0**

The analysis of variance yields the following sums of squares (*n* = 26): $SSE=1000 \quad SSR=600$

Test Statistic:

Decision Rule: Reject H_0 if test statistic is greater than or equal to ________________________

P-value: **> 0.05** **< 0.05**

QI.16. The correlation coefficient (r) is a more appropriate measure to report than the slope of the regression line (b) when describing the association between two variables when there is not a clear independent and dependent variable. (T/F)

QI.17. The slope of the regression line and the correlation coefficient both must lie between -1 and +1. (T/F)

QI.18. Authors of a report state that the coefficient of correlation for their analysis is r=0.5. This means that using X to predict Y reduces prediction error by 50% as opposed to not using X for the predictions. (T/F)

QI.19. A regression equation is fit, relating weekly food expenditures (Y) to number of household members. The prediction equation is Y-hat = 25+60X. The estimated increase in mean weekly food expenditures increases by $60 for each extra household member. (T/F)

QI.20. For a regression relating salary (Y) to work experience (X), the Total sum of squares (around the sample mean Y-bar) is 2000 and the Error sum of squares (around the fitted line Y-hat) is 500. The coefficient of determination r^2 is 0.25.

QI.21. A study relating the font of the print and the time to read a 1000 word newspaper article found a negative linear association (as font increased, time to complete the reading tended to decrease) with a sample correlation of $r = -0.25$. The study was based on a sample of n=24 subjects. Test whether we can conclude that the population correlation coefficient, ρ differs from 0 at the α=0.05 significance level.

- Null Hypothesis ____________ Alternative Hypothesis _____________
- Test Statistic:
- Reject H_0 if the test statistic falls in the range(s) _____________________
- Conclusion
  - i. Conclude that there is a positive association in the population
  - ii. Cannot Conclude there is an association in the population
  - iii. Conclude that there is a negative association in the population.

QI.22. The following computer output gives the results from a regression of criminal rate (Y, criminals per 100,000) to ale/pub rate (X, Pubs per 100,000) for a sample of English towns in 1850. We fit the model: $Y = \alpha + \beta X + \varepsilon$.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.463[a]	.214	.194	37.19272

a. Predictors: (Constant), alepubrt

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression		1			.003[a]
	Residual	52565.326		1383.298		
	Total	66897.600	39			

a. Predictors: (Constant), alepubrt

b. Dependent Variable: crmnlrt

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	109.340	14.755		7.410	.000	79.469	139.211
	alepubrt	.116	.036	.463	3.219	.003	.043	.189

a. Dependent Variable: crmnlrt

p.22.a. Complete the Analysis of Variance Table.

p.22.b. Complete the following parts.

$n =$ ____________________

$\hat{Y} =$ _________________________

$\sum\left(Y-\overline{Y}\right)^2 =$ ________________

$\sum\left(Y-\hat{Y}\right)^2 =$ ___________________

Correlation Coefficient ________________

Proportion of Variation Explained __________

p.22.c. Give the elements of the **t-test** in determining whether there is an association between rates of ale/pubs and criminal rate. H_0: $\beta = 0$ versus H_A: $\beta \neq 0$

Test Statistic ________________ P-value ________________________

p.22.d. Give the elements of the **F-test** in determining whether there is an association between rates of ale/pubs and criminal rate. H_0: $\beta = 0$ versus H_A: $\beta \neq 0$

Test Statistic ________________ P-value ________________________ p.22.e. Based on these tests, what can we conclude at the 0.05 significance level?

- Conclude there is a positive association ($\beta > 0$)
- Cannot conclude there is an association (Do not reject that $\beta = 0$)
- Conclude there is a negative association ($\beta < 0$)

Chapter 10

QJ.1. Simpson's Paradox refers to the situation where overall the association between X and Y is one direction, but when we control for a factor Z, the X-Y association is in the opposite direction for each level of Z. (T/F)

QJ.2. A study is conducted to measure the association between gender and exercise activity in adults. The following table gives the results overall, as well as separately for senior citizens and non-senior citizens. Exercise activity is classified as Low versus High. The samples are random samples from each gender/age group.

Overall	Low	High	Total
Female	100	100	200
Male	125	75	200
Total	225	175	400
Seniors	**Low**	**High**	**Total**
Female	25	50	75
Male	50	25	75
Total	75	75	150
Non-Seniors	**Low**	**High**	**Total**
Female	75	50	125
Male	75	50	125
Total	150	100	250

- Overall, is there an association between gender and exercise activity? Give the Chi-square statistic, and note that the critical value is 3.84 (for α=0.05).
- The Chi-square statistics for Seniors and Non-Seniors are 16.67 and 0, respectively. This is an example of (circle all that apply):

  i. Simpson's paradox

  ii. Spurious Association

  iii. Statistical Interaction

Chapter 11

QK.1. A multiple regression model is fit, relating salary (Y) to the following predictor variables: experience (X_1, in years), accounts in charge of (X_2) and gender (X_3=1 if female, 0 if male). The following ANOVA table and output gives the results for fitting the model. Conduct all tests at the 0.05 significance level:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

ANOVA

	df	*SS*	*MS*	*F*	*P-value*
Regression	**3**	**2470.4**	**823.5**	**76.9**	**.0000**
Residual	**21**	**224.7**	**10.7**		
Total	**24**	**2695.1**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**39.58**	**1.89**	**21.00**	**0.0000**
experience	**3.61**	**0.36**	**10.04**	**0.0000**
accounts	**-0.28**	**0.36**	**-0.79**	**0.4389**
gender	**-3.92**	**1.48**	**-2.65**	**0.0149**

p.1.a. Test whether salary is associated with any of the predictor variables:

H_0: $\beta_1=\beta_2=\beta_3=0$ H_A: Not all $\beta_i = 0$ (i=1,2,3)

Test Statistic ___________________________

Reject H_0 if the test statistic falls in the range(s) __________________________

P-value _______________________________

Conclude (Circle One) Reject H_0 Fail to Reject H_0

p.1.b. Set-up the predicted value (all numbers, no symbols) for a male employee with 4 years of experience and 2 accounts.

p.1.c. The following tables give the results for the full model, as well as a reduced model, containing only expereience.

Test H_0: $\beta_2 = \beta_3 = 0$ vs H_A: β_2 and/or $\beta_3 \neq 0$

Complete Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

ANOVA

	df	*SS*	*MS*	*F*	*P-value*
Regression	**3**	**2470.4**	**823.5**	**76.9**	**.0000**
Residual	**21**	**224.7**	**10.7**		
Total	**24**	**2695.1**			

Reduced Model: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

	df	*SS*	*MS*	*F*	*P-value*
Regression	**1**	**2394.9**	**2394.9**	**183.5**	**0.0000**
Residual	**23**	**300.2**	**13.1**		
Total	**24**	**2695.1**			

Test Statistic:

Rejection Region:

Conclude (Circle one): Reject H_0 Fail to Reject