Stat 10 Lew FINAL WILL BE HELD IN THE LECTURE HALL Review Materials for the FINAL

Exam coverage: Chapters 1, 2, 3.1-3.4, 4, 5, 8, 9, 10.1, 10.3, 12.1, 16.4, 17, 18, 19, 20.1-20.3, 21.1-21.3, 23.1-23.3, 26.1-26.4

FINAL IS 3/20/03 11:30am-2:30pm

Suggested Extra Problems From Your Textbook:

Chapter 8

Set A: #2 #6 Set B: #1, #2 Set D: #1

Chapter 9

Set A: #2 - #6 Set C: #1

Chapter 10.1 and 10.3

Set A: #1 - #4 Set C: #1, #2

Chapter 12.1

Set Â: #1, #2

The are some additional Problems for Review on following pages if you need more problems: pages 265, 428, 565, 567

Final Details

The final is worth 120 points spread over 6 questions or so (approximately the size of your progress tests combined). The breakdown is approximate since I have not written the final yet:

Chapter 1-2, 19	10 points
Chapter 3-4	10 points
Chapter 5	15 points
Chapter 17, 18, 20.1-20.3, 23.1-23.2	25 points
Chapter 21.1-21.3, 23.3	15 points
Chapter 26.1-26.4	25 points
Chapter 8, 9, 10.1, 10.3, 12.1	20 points

Bring a calculator, writing instruments, identification, and a note sheet (both sides allowed). You are not allowed to eat during the final, you may bring something to drink (non-alcoholic, please, remember some of your classmates are under 21) however.

I am not allowed to reveal final grades via e-mail or phone. If you want to know yours as soon as possible, leave a grade card with me. Otherwise, the grades will be posted on URSA in a timely manner.

What follows are practice problems, the final is not this long, it's just extra practice. Good luck

FINALS WEEK OFFICE HOURS have not been scheduled yet. They will most likely fall on Tuesday March 18th in the late afternoon (after 3pm) and on Wednesday March 19th in the morning.

1. Mr. Joe Potato works very hard during the workweek (Monday through Friday), but likes to watch a lot of television on weekends. The number of minutes of television viewing for Joe, on each of 60 consecutive days, was recorded. For this data set of 60 values, which of the following would be true? (Choose one)

- a. The mean of this data set would smaller than the median.
- b. The data set would be skewed left.
- c. The data set would be skewed right.
- d. Both a and b are true

Stat 10

e. Both a and c are true

2. The GPA (on a scale of 1-4) of a sample of students at UCLA has sample mean 2.9106 and sample standard deviation 0.256, with 95% confidence interval (2.488,3.333). Which of the following is correct?

- a. If we sample many times, the proportion of intervals computed that cover the true mean is about 95%.
- b. There's a 95% chance that the true mean falls between 2.488 and 3.333.
- c. The estimate is within 95% of the true mean.
- d. The population has a normal distribution with mean 2.9106 and standard deviation 0.256.

Lew

- e. None of the above are correct
- 3. Which of the following is always true?
 - a. 95% of the data are within 2 standard deviations of the mean.
 - b. The distribution of a variable is always bell-shaped.
 - c. Histograms always have means and medians.
 - d. Histograms whose means are greater than their medians are always left skewed
 - e. None of the above are correct

4. At least 68% of the values in a data set fall within 1 standard deviation of the mean. TRUE or FALSE.

5. If the smallest value in a data set is removed, it would cause the standard deviation to decrease. TRUE or FALSE.

- 6. Which of the random variables listed below are continuous?
- a) The time it takes for a tow truck to arrive.
- b) The number of buttons on a shirt.
- c) The distance a long jumper jumps in a competition.
- d) All of the above.
- e) Only (a) and (b)
- f) Only (b) and (c)
- g) Only (a) and (c)
- 7. If the correlation between body weight and annual income were high and positive, we could conclude that: a. high incomes cause people to eat more food.
 - b. low incomes cause people to eat less food.
 - c. high income people tend to spend more of their income on food than low income people, on average.
 - d. high income people tend to be heavier than low income people, on average.
 - e. high incomes cause people to gain weight.
- 8. The correlation coefficient provides:
 - a. a measure of the extent to which changes in one variable cause changes in another variable.
 - b. a measure of the strength of the linear association between two categorical variables.
 - c. a measure of the strength of the association (not necessarily linear) between two categorical variables.
 - d. a measure of the strength of the linear association between two quantitative variables.
 - e. a measure of the strength of the linear association between a quantitative variable and a categorical variable.

- 9. In the histogram below, which of the following is true? (choose 1)
- a) The proportion of scores greater than 15 is 0.46.
- b) The proportion of scores between 15 and 35 is 0.65.
- c) The proportion of scores less than 55 is 0.05
- d) Both a and b are true.
- e) Both b and c are true.
- f) Both a and c are true.





11. The standard deviation is a common measure of variability that displays the average distance of scores from the mean. TRUE OR FALSE?

12. Two confidence intervals are calculated for a proportion p: a 90% and a 99% confidence interval. Each confidence interval is based on the same random sample. Which one of the following statements is true?

A. The 99% confidence interval would be narrower.

B. The 90% confidence interval would be wider.

C. The 99% confidence interval would be wider.

D. It is not possible to determine which is wider and which is narrower, based on the information given.

13. Which of the following statements is most appropriate for a hypothesis test?

A. When the p-value is small, say less than .05, we can reject the null hypothesis or equivalently, accept the alternative hypothesis.

B. When the p-value is not small, for example if the p-value is greater than .10, we cannot reject the null hypothesis.

C. It is almost always correct to say "I accept the alternative hypothesis".

D. It is almost always correct to say "I accept the null hypothesis".

True	False	Question
		The sample standard deviation of a data set must be zero or larger.
		If X is a continuous random variable which is normally distributed with a mean of 100 and a
		standard deviation of 15 then the probability that X>115 is 0.5.
		In order to have a valid probability distribution, the sum of the probabilities must equal to 1 or
		100% and the probabilities themselves cannot be negative or greater than 100%.
		The height of a randomly selected UCLA football player is a quantitative variable.
		If the mean of a variable is less than the median of that variable, it is correct to say that the
		distribution is right skewed.
		The idea behind statistical inference is to find distribution(s) of statistic(s).

14. Mark ONE of the columns

Stat 10

1. The IQ scores of adult humans (age 18 and over) is approximately normal with a mean of 100 and a standard deviation of 15.

(a) How low is the lowest 5% of all IQ scores (that is, at or below what IQ score is the lowest 5%) How high is the highest 10% of IQ scores (that is, at or above what IQ Score is the highest 10%)?

(b) A simple random sample of 225 college students is drawn from the adult human population. The sample average is 103 and the sample standard deviation is 30. Please test the hypothesis that college students have higher IQ scores than the average human. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

2. Investors ask about the relationship between returns on investments (the money you make by investing your money) in the United States and on investments overseas. Below is a table of total returns on investments on U.S. and overseas stocks over a 10 year period.

	Year	Overseas	U.S.
		% Return	% Return
	1987	24.6	5.1
	1988	28.5	16.8
	1989	10.6	31.5
	1990	-23	-3.1
	1991	12.8	30.4
	1992	-12.1	7.6
	1993	32.9	10.1
	1994	6.2	1.3
	1995	11.2	37.6
	1996	6.4	23
Average	1991.5000	9.8100	16.0300
Standard Deviation	2.7386	15.6493	12.6810

(a) Suppose the correlation, r, of the U.S. and overseas returns is .32. Using this information describe the relationship between U.S. and overseas returns in plain English.

(b) Find the regression line of overseas returns on U.S. returns. Please interpret the values of the slope and of the intercept of this line.

(continued from above)

(c) In 1993, the return on U.S. stocks was 10.1%, what was the predicted return on overseas stocks. Is the predicted return the same as the actual return? If it is the same, please explain why this is so. If it is different, please explain why they are different.

3. You got a job working for a marketing company and your supervisor is planning a sample survey of households in Los Angeles. Your supervisor instructs you to contact households by random-digit dialing phone numbers. Your supervisor knows from past experience that about 70% of the households you contact in this manner will respond.

(a) If you randomly dial 1500 telephone numbers, what are the mean and standard error of the number of households who respond?

(b) Find the probability that you will get at least 1000 responses.

4. You are planning to perform a significance test of

 H_0 : mean = 0

Versus

H₁: mean < 0

What values of Z would lead you to reject H_0 at the 1% level of significance? Then answer this question: True or False and explain why. A significance test that is significant at the 1% level of significance must always be significant at the 5% level of significance.

5. An investigator looks up the rainfall in a certain city on January 15 for the past 70 years. She finds the average rainfall on that day to be 0.30 inches and the SD to be about 0.14 inches. She then concludes that the interval from 0.25 to 0.35 inches is a 99.7% confidence interval for the average rainfall on January 15 in the city. Is this conclusion justified? Why or why not?

6. The speed of light is measured 2,500 times by a new process. The average of these 2,500 measurements is 299,774 kilometers per second, with an SD of 14 kilometers per second.

a) Find an approximate 95% confidence interval for the speed of light. (You may assume normality, with no bias.)

b. Now the investigators determine the speed of light again by taking the average of another 2,500 measurements. The average is now 299,781 kilometers per second. Is this a surprising result? Again, you may assume normality with no bias. Why or why not?

7. In government data, a household consists of all occupants of a dwelling unit. Choose an American household at random and count the number of people it contains. Here is the assignment of proportions for your outcome:

Number of persons	1	2	3	4	5	6	7
Box Proportions	0.25	0.32	???	???	0.07	0.03	0.01

The proportions of 3 people in a household is the same as that of 4 people. These proportions are marked with a ???.

a) Find the proportions of households that contain 3 people.

b) Pretend the table above is a box model. What is the box average?

c) 100 families are going to be drawn at random from the "box" and will become a part of a new study on poverty. What is the expected number of people in the study?

8. Suppose that 47% of all adult women think they did not get enough time for themselves. An opinion poll interviews 1025 randomly chosen women and records the sample proportion that doesn't feel they get enough time for themselves. This statistic will vary from sample to sample if the poll is repeated. The sampling distribution is approximately normal with mean 0.47 and standard error about 0.016.

a) The truth about the population is 0.47. In what range will the middle 95% of all sample results fall for samples of size 1,025?

b) What is the chance that a new poll of size 1,025 gets a sample in which fewer than 45.4% say they do not get enough time for themselves?

9.A study of many families gave the following results:

average height of father=68 inches, SD=3 inches average height of daughter=63 inches, SD=2.5 inches r=0.6

Using the regression method, estimate the height of a daughter whose father is 62 inches tall

Lew

FINAL WILL BE HELD IN THE LECTURE HALL

10. Does salt cause high blood pressure? One large study was done at 52 centers in 32 counties. Each center recruited 200 subjects in 8 age- and sex- groups. Salt intake was measured, as well as blood pressure and several possible confounding variables. After adjusting for age, sex, and the confounding variables, 25 of the centers found a positive association between diastolic pressure and salt intake; 27 found a negative association. Do the data support the theory that salt causes high blood pressure? Answer yes or no, and explain briefly.

11. A study on pre-meds, selected at random, gives the following results for the medical college admissions test (MCAT) and undergraduate GPA (grade point average):

Average GPA: 3.3;Standard deviation= 0.4Average MCAT:10;Standard deviation= 1.1Correlation coefficient = 0.65

Suppose the percentile rank of one student's GPA is 90%. Predict the student's percentile rank on the MCAT. The scatter diagram is football shaped and the MCAT and the GPA are normal.

12. The pregnancy duration of human females (age 18 and over) is approximately normal with a mean of 266 days and a standard deviation of 16 days. It is believed that older pregnant women have longer pregnancy durations. A simple random sample of 121 older pregnant women is drawn from the population of all pregnant women. The average pregnancy duration for the sample is 267 days and the sample standard deviation is 35. Please test the hypothesis that older women have longer pregnancy durations than the average woman. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

13. The following comes from a recent article in The Wall Street Journal :

Generation X-ers Aren't Relying On the Survival of Social Security

BY JOHN SIMONS

According to the most recent Wall Street Journal/NBC News poll, only 39% of X-ers believe that Social Security will still be able to provide benefits when they retire. That compares to recent surveys of all Americans which show that 45% think so.

Assume that 45% figure is a stable, long-run, historical fact about American beliefs about Social Security benefits. Also assume that the survey of Generation X-ers had 121 respondents.

A. Test the hypothesis that the belief that Social Security will still be able to provide retirement benefits has decreased over time. State the null and the alternative, perform a test, and state a p-value. Please use a 5% level of significance as your decision rule. On the basis of your test results, do you think that Generation X-ers are like other Americans in their beliefs about social security or are they different?

B. Suppose in a few years the Wall Street Journal decided to replicate this study (i.e.draw a new sample) on Generation Y-ers (that's you all, I think...). Let's assume that the 39% figure is now the stable, long run fact about belief in Social Security benefits by Americans.

What is the chance that a sample of 100 will have at least 32% of the surveyed Generation Y-ers believing in Social Security?

Stat 10

Lew

FINAL WILL BE HELD IN THE LECTURE HALL

14. Here are two statistics on all persons who consider themselves medical doctors in 1999:

\$820,000 dollars per year \$141,000 dollars per year

Which one of these numbers is the mean salary and which one is the median?

The mean is _____

The median is_____

Explain your choice in the space below. Be brief. This is not a long answer.

15.High Bias and High Variance are both considered undesirable features of certain sample statistics (such as a sample mean for example). You are working with a team on a marketing study, a sample of size 100 is drawn. One of the variables you are interested in is the average time spent on the internet on any day. You plan to construct confidence intervals and perform some unspecified hypothesis tests. Studies always have problems, and today you have your choice: High Bias or High Variance. Which one would you rather deal with and why?

16. A marketing survey interviewed 1000 adults selected at random from the population of all U.S. adults. Of the adults, 529 said they currently own a personal computer. When asked about the manufacturer of their computer, 144 of them said "Dull", 115 of them said "Compact", 275 of them said "some other company" and the rest of them said "I don't know". The mean time of ownership (in months) for the 529 was 12.9 with a standard deviation of 8.7.

(a) A Compact executive saw the survey and is now upset, he believes that the survey was poorly done and argues that Compact's true market share is 25% (i.e. he thinks that 25% of all adults who own computers own a Compact) and cannot be nearly as low as the survey suggests.

Let's help the executive out. Please test the hypothesis that Compact's market share is actually 25%. Use a 5% level of significance as your decision rule. State the null hypothesis, the alternative hypothesis, perform a test, give a p-value, and state your conclusion in plain English: would you reject the null and on the basis of your test result do you also think the survey was poorly done?

(b) Suppose Compact's market share REALLY IS 25%. What is the chance that among 529 computer owners you would get less than 22% of them saying they owned a Compact? What is the chance that you would get between 22% and 28% saying they owned a Compact? What is the chance that you would get at least 28% saying they owned a Compact?

17. A study of 250 first year college students, selected at random after their first full year of college, gives the following results for the course units (NUNITS) first year GPA (grade point average) and high school SAT I score:

Average GPA:3.3;Standard deviation GPA = 0.4Average NUNITS:38;Standard deviation NUNITS:6.1Average SAT I:1010;Standard deviation SAT I = 201Correlation coefficient for GPA and NUNITS=.35Correlation coefficient for GPA and SAT=.65Correlation coefficient for NUNITS and SAT=-.55

Assume SAT scores have a minimum of 400 and a maximum of 1600 and are normally distributed. NUNITS has a minimum of 12 and a maximum of 60 and is not normally distributed. GPA has a minimum of 0.0 and a maximum of 4.3 and it is not normally distributed.

(a) Please interpret the value of the correlation coefficient for SAT and NUNITS in plain English. Discuss the direction and magnitude its value implies.

(b) A student is interested in regressing GPA on SAT I. Using the information at the top of the page, please find the regression equation. Clear identify the slope, intercept, x and y variables.

(c) Please interpret the values of slope and intercept you calculated in part (b) in plain English.

Stat 10

Stat 10

Lew

FINAL WILL BE HELD IN THE LECTURE HALL

18. The amount of money all college students earn in the year after graduation is right skewed with a mean of \$26,600 and a standard deviation of \$5000. Apparently, 20% of all students earn no money in the year after graduation. Administrators at UCLA believe that UCLA students earn more money in the year after graduation than the average college student. A simple random sample of 250 UCLA students is drawn from the population of all recent UCLA graduates. The average for the sample was \$27,300 and the sample standard deviation is \$9,500. The sample also revealed that 15% of UCLA students earned no money in the year after graduation.

Is it possible to test the hypothesis that UCLA students earn more than other college students? If you think it is, please state a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and tell us how to interpret the result -- in plain English) and use a 5% level of significance as your decision rule.

If you think it is not possible to test the hypothesis, please use the space below to explain why this is not possible.

19.A simple lottery gamed called "The Daily 3" is played in California. You pay \$1 to play and you get to choose a 3-digit number. Each digit can have any value from 0 to 9 and you can have repeated values (e.g. 111, 222 etc.) The State of California randomly selects a 3 digit number at the end of each day and pays you \$500 if your number is selected. There are 1,000 3 digit combinations (different 3 digit numbers).

If you were to play every day for one year (365 plays). What is the chance that you will win money? You can treat your plays as if they were a random sample of size 365.

20.Suppose it known that Americans spend an average of 12.2 minutes in the shower each day. If the standard deviation of showering times is 2.3 minutes and times are normally distributed:

a. Find the percentage of Americans who spend at least 15 minutes in the shower

b. Calculate the chance that the average time of a random sample of 144 Americans will be less than 11.9 minutes

c. Find the percentage of Americans who spend between 11 minutes and 15.5 minutes in the shower.

d. What percentage of random samples of size 100 will have a sample average between 12 and 13 minutes?

21. Lawyers frequently receive a "year-end bonus" because law firms are partnerships and the money earned is shared among partners. There are approximately 1,000,000 lawyers in the U.S. and year 2002's "year-end bonus", when calculated as a percentage change of the 2001 "year-end bonus", is normally distributed with a mean year-end bonus of -9% (a decrease, 2002 was a bad year compared to 2001) and a standard deviation of 16%. SHOW YOUR WORK FOR FULL CREDIT.

a) What proportion of lawyers received year-end 2002 bonuses that were larger than their year-end 2001?

- b) A sample of 100 lawyers has an average percentage change at the 15th percentile, what is the value of its average?
- c) How many lawyers have bonuses between -1% and +5%?

d) If it is possible to do this, please calculate the median and interquartile range for the probability histogram of samples of lawyers of size 225. If any of these are not possible to calculate, please identify the ones that are not calculable and please give values for the ones that are calculable.

22. Mark ONE of the columns

True	False	Statement
		You can tell if a distribution is right-skewed if you know its standard deviation and its mean
		Standard Errors are either zero or positive values
		For normal distributions, Chance Error and the Standard Error are equivalent
		Selection Bias is a result of mistakes made by the person or persons who design a study which involves sampling
		Non-response Bias is caused by researchers who fail to write a survey questionnaire properly
		Chance error is the source of difference between statistics and parameters
		Observational studies differ from randomized controlled experiments in that the researchers do
		not assign the subjects to treatment or control
		Association may point to causation, but it is not the same as causation
		An extremely large biased sample (e.g. > 10,000) generates better estimates population
		parameters than a small random sample (e.g. a little over size 100)
		A quantitative variable can be discrete or continuous

YES

YES

23.The Dull Computer Company manufacturers its own computers and delivers them directly to customers who order them via the Internet. Dull's market dominance has arisen from its quick delivery and competitive pricing. The CEO (Chief Executive Officer) of Dull has stated publicly that if customers make unassisted online purchases of their computers (i.e. no interaction with a salesperson), the computers will have an average delivery time of 56 hours from the time of purchase (with a standard deviation of 18 hours). He also noted that about 20% of these computers cost less than \$1300. The CEO refused to give a mean cost of these computers but said the standard deviation was \$600. The cost of these computers is not normally distributed nor is the delivery time.

A consumer research organization decided to test the CEO's delivery time claim and cost claims by purchasing 200 computers from Dull at randomly selected times and days. The 200 purchases were randomly divided into two groups: 72 were purchased by telephone and involved talking to a live salesperson, the remaining 128 were unassisted online purchases. 32 of the 128 computers were delivered in less than 45 hours. The mean cost of the 128 purchases was \$1,661 Please assume that the purchases (i.e. 200, 72, 128) constitute reasonably large samples of good quality.

Using the information above and from what you have learned in this course, can you construct a 80% confidence interval for the population percentage of computers that will be delivered in less than 45 hours? If yes, circle yes below, justify your response and then construct the interval. If no, circle no below and justify your response. (5 points)

NO

(b) Using the information above and from what you have learned in this course, can you construct a 80% confidence interval for the mean cost of all computers involved in online unassisted purchases? If yes, circle yes below, justify your response and then construct the interval. If no, circle no below and justify your response. (5 points)

ΝO