Lew

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400 Review Materials for the FINAL (ANSWERS WILL BE POSTED ON 12/5/2004)

Exam coverage: Chapters 1, 2, 5, 6, 7, 8, 9 (basic concepts only), 12, 13, 14, 15 (possible extra credit), 16, 18, 19, 20, 21 (basic concepts only), 23

OFFICE HOURS 10th and FINALS WEEK

M 12/6 11am-noon & 2pm-4pm W 12/8 8am-9:30am & 3:15pm-4:15pm F 12/10 11am-12:30pm Sunday 12/12 2:30pm-4:00pm Room To Be Announced

CONCEPTS COVERED

Variable, Categorical Variables, Quantitative Variables (discrete & continuous and they have units), Ordinal variables, frequency table, relative frequency table, distribution, modes, unimodal, bimodal, multimodal, uniform, symmetric, tails, skewed right AKA positive and left AKA negative skewness) outliers, mean AKA average, median, range, interquartile range, quartiles (25th, 50th, and 75th percentiles, Q1, Q2 AKA Median AKA 50th percentile, Q3), percentiles, five number summary, variance, standard deviation, z-scores AKA standard scores AKA standardized values, rescaling, normal, parameters, standard normal distribution, 68-95-99.7 rule, normal percentiles, random, simulation, outcome, trial, population, sample, sample surveys, bias, randomizing, sample size, census, population parameter, sample statistic, representative, simple random sample, sampling frame, sampling variability, stratified random sampling, cluster sample, multistage samples, voluntary response bias, convenience samples, undercoverage, nonresponse bias, response bias, wording of the question, observational study, retrospective study, prospective study, experiment, random assignment, factor, response, subjects, levels, treatment, control, replication, statistically significant, control group, single blind, double blind, placebo, placebo effect, confounding. Independence, Probability, Mutually Exclusive, Probability distribution, Sampling Distribution, Sampling Distribution for a mean and for a proportion, standard deviation, Central Limit Theorem, Standard Error, Confidence Interval, Margin of Error, Alternative hypothesis, Null Hypothesis, Z-test, one sided alternative, p-value. Sample size. t hypothesis test, t confidence interval, t-table Correlation. Scatter diagram or scatter plot. Regression Line. Predicted value. Residuals. Slope. Intercept. Extrapolation. Outliers.

Suggested Extra Problems From Your Textbook:

pp. 104-112: 25, 28, 29, 37 pp. 199-208: 2, 11a-d only, 13, 14a-b & d, 39 pp. 262-267 :1, 3, 5, 23, 25, 31, 34 pp. 329-334: 1, 3, 5, 7, 11, 13, 17, 19 pp. 424-428 : 5, 9, 13, 21, 23, 29 pp. 447-451: 13, 14, 15, 16, 21, 27

Final Details

The final is worth 200 points spread over 4 questions. The breakdown is approximate since I have not written the final yet, but this is what I am thinking about as I assemble the review materials:

Chapters	7 & 8: 60 points	(Correlation & Regression)
Chapters	1, 2, 5, 12-13: 40 points	(Descriptive Statistics, Surveys, Observational Studies, Experiments)
Chapters	14, 16 40 points	(Expected Value & Standard Deviation)
Chapters	6, 18, 19, 20, 23: 60 poin	nts (Z and t distribution & applications)

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

Department regulations forbid me from revealing final grades via e-mail or phone. 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 a lot of extra practice. Best wishes.

Lew

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

1. A study of 150 first year college students, selected at random after their first full year of college, gives the following results for high school ranking (RANK) first year GPA (grade point average) and high school SAT score:

Average GPA =3.19;Standard deviation GPA= 0.49Average RANK=2.04;Standard deviation RANK = 1.13Average SAT =1291;Standard deviation SAT = 163Correlation coefficient for GPA and RANK= -.33Correlation coefficient for GPA and SAT= .24Correlation coefficient for RANK and SAT= .42

Assume the SAT scores had a minimum of 910 and a maximum of 1580 and are normally distributed. RANK has a minimum of 0 and a maximum of 6 and is not normally distributed. GPA has a minimum of 1.56 and a maximum of 3.98 and it is normally distributed.

(a) Of the 3 correlation coefficients given to you above, please identify which pair has the strongest correlation and which has the weakest correlation.

(b) A student is interested in regressing first year college GPA on SAT. 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.

(d) Peter and Paul are USC undergraduates who were selected in this study. Peter had an SAT score of 1040 and Paul had an SAT score of 820, what are their predicted GPAs?

Stat 10

2. 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 as large as or larger than their year-end 2001 bonuses?

b) A simple random sample of 100 lawyers has an average year-end bonus at the 15th percentile, what is the actual value of that average?

c) A simple random sample of 100 lawyers has an average year-end bonus of -7.5%, what is the chance of getting a sample average of -7.5% or higher?

d) What is the chance that a lawyer, selected at random, will have a year-end bonus of -7.5% or higher?

e) What percentage of lawyers have year end bonuses between -1% and +5%?

3. The IQ scores of adult humans (age 18 and over) are approximately normally distributed with a mean of 100 and a standard deviation of 15. The highest IQ of a currently living adult as reported by the Guiness Book of World Records belongs to Marilyn vos Savant who scored a 186 (nearly six standard deviations above average). The maximum IQ score is 200, values estimated above it are deemed unreliable. The lowest score on record is 40.

(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 size 256 is drawn from the adult human population. What is the chance that the sample average will exceed 101?

(c) How large of a sample would a researcher need to select to insure that he or she is within plus or minus 1 IQ point of the population mean IQ with 99% confidence?

(d) A simple random sample of 256 USC students is drawn from the adult human population. The sample average is 97 and the sample standard deviation is 30. Please test the hypothesis that USC students have lower 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.

4. 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. Assume overseas returns and US returns are both normally distributed.

Lew

Year		Overseas	U.S.	
		% Return	% Return	
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 -.3239. Please describe the relationship between U.S. and overseas returns in words, using r to make your description more precise.

(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.

(c) In 1993, the return on U.S. stocks was 10.1%, what was the predicted return on overseas stocks for that year? Suppose I told you that the actual return on Overseas Stocks that year was 32.9%? Why are they so different?

Stat 10

Lew

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

5. 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.

6. 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= 0.4Min. GPA = 2.5 Max. GPA = 4.3Average MCAT:10;Standard deviation= 1.1Min MCAT=7.0 Max MCAT=13Correlation coefficient = 0.65

a) Suppose the percentile of one student's GPA is at the 90th percentile. Predict the student's percentile on the MCAT. The scatter diagram is football shaped and the MCAT and the GPA are normal.

b) Please construct the regression equation for MCAT on GPA. Please interpret the values of the slope and intercept in plain English.

c) Your wealthy cousin, a graduate of USC, is really, really dumb. He told you his GPA is a whopping 1.9. He's now thinking of going to medical school. What is his predicted MCAT score?

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

7. 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.

(a) 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.

(b) What proportion of pregnancies have durations as long as or longer than 300 days?

(c) Suppose a researcher is only interested in studying the proportion of pregnancies that have durations as long as or longer than 300 days. How large of a sample would he or she have to select in order to properly invoke the Central Limit Theorem to create confidence intervals or test hypotheses about these 300+ day pregnancies?

8. Here are two statistics on all persons who consider themselves computer programmers in 1999:

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

Which one of these numbers is the mean salary from computer programming and which one is the median salary from computer programming in 1999? Assume the samples were of good quality.

The mean is _____

The median is_____

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

9.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?

10. Los Angeles International Airport handles an average of 6,000 international passengers an hour. Suppose 80% can pass through primary security, but the rest are detained for interrogation by the FBI. And suppose the FBI can handle 1,500 passengers an hour without unreasonable delays for travelers and extra costs to the airlines (due to missed flights and connections).

a.Over break, it is expected that as many as 8,000 international passengers will arrive per hour. When that occurs, what is the expected proportion of passengers who will be detained?

b. Referring to part a, find the approximate chance that less than 2,000 out of 8000 international passengers will be detained?

c. Suppose the FBI decides to randomly sample passengers in order to speed up the screening process. What is the chance that a simple random sample of 100 will have between 22 and 28 passengers detained by the FBI?

d. Certain ethnic/racial groups appear to be detained at much higher rates than others. Suppose a human rights organization sends 64 persons who appear to be of middle eastern origin through the airport and 21 are detained for interrogation. Please test the hypothesis that persons of middle eastern origin are detained in higher proportions than the typical traveler. 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. You may treat the 64 as if it were a simple random sample and it is of reasonable size.

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

11. 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", 175 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.

Lew

(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 true market share is REALLY 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?

Lew

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

12. A student issued the following command for analysis variable FAMILYSEI

. summarize familysei, detail

familysei				
	Percentiles	Smallest		
1%	0	0		
5%	22.5	0		
10%	28.2	0	Obs	1428
25%	37.3	0	Sum of Wgt.	1428
50%	63.5		Mean	68.9631
		Largest	Std. Dev.	39.68948
75%	92.3	175		
90%	129.5	180.6	Variance	1575.255
95%	145.7	194.4	Skewness	.6213121
99%	166.9	194.4	Kurtosis	2.792152

Please answer the following questions based on the Stata results for variable FAMILYSEI. You may round the numbers given above to one or two decimal places. For example, 39.68948 can be rounded to 39.7 or 39.69

A. Using the Stata results above, please calculate the range, the interquartile range, and list the values of the quartiles (i.e. Q1, Q2 and Q3) (10 points)

 B. Is the distribution for this variable skewed? (circle one): YES NO (1 point) Please justify your response in the space below. If you think it is skewed, please indicate the direction (left or right skewed) of the skewness. (4 points)

C. Suppose we could treat the 1,428 persons as a population, from it, a simple random sample of size 51 is drawn, assume it is large enough to overcome the non-normality in the population. The mean familysei for the sample is 70. Please test the hypothesis that something is wrong with the sample because it is biased towards higher values. Please state a null and alternative hypothesis, perform the appropriate test and use an alpha=.05 as your decision rule. Please state clearly whether you reject or do not reject the null and give a brief verbal (simple English) conclusion.

Lew

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

13. There are 20,000 restaurants in the County of Los Angeles, 50% of them received a letter grade of "A" during inspections, 40% received either a B or a C grade and 10% failed their inspections. Restaurant grades are not normally distributed. My financial adviser, the Oracle, has hired you as a temporary personal assistant. Your job is to schedule his next 16 dinners (Oracle never eats at home). Unfortunately, you didn't know about the rating system and you never eat out because you don't have the money. So you listened to your best friend and picked 25 restaurants at random with replacement from an internet database of the 20,000 restaurants in Los Angeles. The Oracle will give you +3 points if you choose "A" restaurants, +1.25 points if you choose "B" or "C" restaurants, and -20 points if you choose a restaurant with a failing grade. Treat your restaurant selections as if they were a simple random sample of restaurants.

A. Please onstruct the probability distribution for this problem

B. What is the expected value for the probability distribution? What is the expected value for the mean restaurant scores for a sample of 25 restaurants selected at random with replacement (please assume that the sample is large enough to use Z)?

C. What is the standard deviation for this distribution?

D. What is the standard deviation for the average score of a sample of 25 restaurants?

E. To convert your temporary job into a permanent job, you must have accumulated an average of at least +1 points from the Oracle after picking 25 restaurants for him. What's your chance of getting an average of at least +1 points after picking 25 restaurants? If it is not possible to calculate the chance, please write "not possible" below and explain why.

14. You know that every UCLA student will definitely get a job after graduation. The only uncertainty is the salary. Suppose this is what you know about the job prospects of UCLA students after graduation:

There is a 35% chance that the salary will be \$20,000 per year; a 45% chance that it will be \$90,000 per year; and a 20% chance that it will be \$40,000 per year.

- a. Construct the probability distribution for the salary of UCLA students
- b. Find the expected value and the standard deviation.

c. Suppose a random sample of 36 UCLA students who have graduated is drawn. What is their expected average salary? What is the standard deviation of the salary for the sample of 36? What is the chance that the sample average will exceed \$57,000

d. Suppose a random sample of 49 UCLA students who have graduated is drawn. What is their expected average salary? What is the standard error of the salary? What is the chance that the sample average will exceed \$57,000

Stat 10

15. An investment advisor advertises that he is "da very best" and proclaims that his clients will "retire in style" if they hand over all of their savings to him. However, after an audit, the Securities and Exchange Commission (SEC) learned that only 6.9% of his clients had percentage returns greater than or equal to 10%. Additionally, 25% of his clients reported percentage returns as low as or lower than -8%. One client had lost all of her savings (a total loss of 100%) but another client was at the maximum with returns of +91%. Noting the variation in returns, the SEC assumes that his clients' percentage returns are normally distributed (and you should too) (Note: outliers can occur in normal distributions).

Using the information above please answer the following:

A. Please calculate the mean and standard deviation of all of his client's percentage returns.

B. Please list and clearly identify the values of the five number summary for the percentage returns of all of his clients. Please show the numeric computations when appropriate.

C. Hoping to maintain his claim of "retiring in style" the advisor believes that by changing his investment decisions for his clients, it will increase the percentage of clients with percentage returns in excess of 10% to 24% instead of the 6.9% reported by the SEC. If the standard deviation (calculated in part A, we will assume it is correct) remains the same, what new mean would he need to achieve his goal?

D. Suppose the SEC performed a poor audit and in fact, the advisor has client percentage returns that are normally distributed with mean 2.5% and standard deviation 6%. Please calculate the probability that at least one of three randomly selected returns are \geq -4%. Please assume independence.

Stat 10 Lew FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

16. Here is a correlation table:

	insurance	unemploy	poverty	income	incarcerat	io violen	ce
insurance	+ 1.0000						
unemploy	0.4872	1.0000					
poverty	0.1106	0.5222	1.0000				
income	0.4115	-0.1345	-0.7099	1.0000			
incarceratio	0.4473	0.4389	0.4756	-0.1276	1.0000		
violence	0.5284	0.4381	0.3883	-0.0014	0.8144	1.0000	

Please identify the strongest positive correlation, the strongest negative correlation and the weakest correlation in the table.

17. The UCLA housing office wants to estimate the mean monthly rent for one bedroom apartments in the 90024 zip code. A random sample of size 36 is selected from the zip code area and the sample mean is found to be \$1,100. The sample standard deviation is \$200.

- a. Please construct a 95% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.
- b. Please construct a 99% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.

c. What is the minimum sample size needed to generate a 95% confidence interval which has a margin of error of \pm \$50?

18. Mark ONE of the columns

True	False	Statement		
		You can determine whether a distribution is right-skewed if you know its standard deviation and		
		its mean		
		Standard Deviations are either zero or are positive values		
		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		
		Observational studies differ from randomized controlled experiments in that the researchers do		
		not assign the subjects to treatment or control groups		
		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)		
		Quantitative (or numerical) variables can be discrete or continuous		
		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%.		

19. The Los Angeles Department of Transportation fines Taxicab companies that take too long to respond to customers. Each year, the Department of Transportation tests each taxicab company by making 100 independent calls to that company (you can treat this as a random sample of size 100) and uses this sample to check the proportion (or percentage) of calls that are answered in a timely manner where within 15 minutes is defined as timely.

Suppose that a taxi company's true proportion of calls answered in a timely manner is .81 (or 81%). That is, in the course of thousands and thousands of calls, the taxicab company would be able to successfully answer 81% of the calls within fifteen minutes.

- (a) What is the chance that a random sample of 100 calls made to this taxi company will only show 77% were answered in a timely manner?
- (b) The Los Angeles Department of Transportation doesn't know that this company has a true proportion of calls answered in a time manner of .81 or 81%. Instead, it makes 100 independent calls to the company (treat it as a random sample) and finds that only 75% were answered in a timely manner. Please construct a 98% confidence interval for the proportion of calls answered in a timely manner.

20. A bank wonders whether omitting the annual credit card fee for customers who charge at least \$2400 in a year would increase the amount charged on the credit card. The bank makes this offer to a simple random sample of 25 of its existing credit card customers. It then compares how much these customers charge this year with the amount they charged last year. The mean percentage change is +12% with a standard deviation of 6%. The amounts are nearly normally distributed.

(a) Is there a statistically significant increase in the mean amount charged? Please state a null and alternative hypothesis, perform the appropriate test, state a p-value and use an alpha=.05 as your decision rule. Give you conclusion in plain English please – do you reject or not reject the null and what does this mean?