FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

10th WEEK and FINALS WEEK OFFICE HOURS: 12/4 12:30pm-2:00pm, 12/5 4:00pm-5:30pm, 12/11 to be announced and by appointment. E-mail will be accepted until 11:59pm 12/11/02

Review Materials for the FINAL

Exam coverage: Chapters 1, 2, 3.1-3.4, 4.1-4.3, 4.5, 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

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 A: #1, #2
Review: #1, #2, #3

General Problems for Review
Starting on page 263 #4, #9 , #13, #17
Starting on Page 428 #1. #5, #8, #9, #11, #16, #18, #27
Starting on Page 567 #1, #2, #8, #9, #11, #17, #18, #25, #26

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	20 points
Chapter 17, 18, 20.1-20.3, 23.1-23.3	20 points
Chapter 21.1-21.3, 23.1-23.3	20 points
Chapter 26.1-26.4	20 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, check on line using the "gradebook". 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

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

1. Suppose you know that the mean height of US males is 69 inches and that the standard deviation is 4 inches. Also suppose that the shape of the histogram for heights of US males is approximately normal. Which of the following follow from this information?

a) At least 75% of US males have heights between 61 and 77 inches.
b) Approximately 68% of US males have heights between 65 and 73 inches.
c) At most 25% of US males have heights above 77 inches.
d) All of the above.
e) None of the above
f) Only a and b
g) Only b and c
h) Only a and c

2. 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
e. Both a and c are true

3.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.
e. None of the above are correct

4.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

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

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

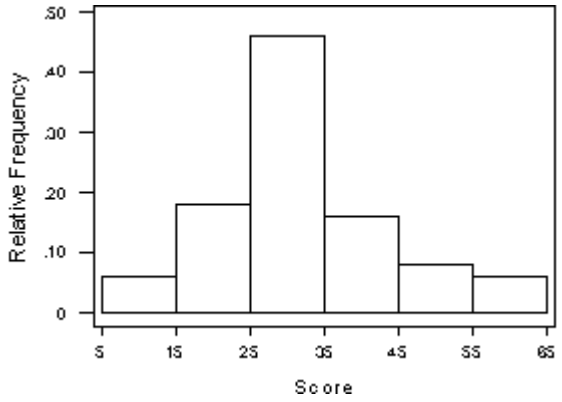
7. Which of the random variables listed below are continuous?

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)

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

8. 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.



9. A correlation coefficient of -.7 is a negative and weaker correlation than +.50. TRUE OR FALSE

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

11. 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.

12. Which of the following statements is not appropriate for an 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 small, for example if the p-value is greater than .10, we cannot reject the null hypothesis.

C. It is almost never correct to say "I accept the null hypothesis".
D. It is almost always correct to say "I accept the null hypothesis".

Mark ONE of the columns

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).

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

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 % Return	U.S. % 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) Find the correlation, r, of the U.S. and overseas returns then 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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

(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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

4. You are planning to perform a significance test of

H_0: mean = 0

Versus

H_1: 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.)

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

b.Now the investigators determine the speed of light once more by the same procedure and get 299,781 kilometers per second. Is this a surprising result? 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 probabilities for your outcome:

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

The probability of finding 3 people in a household is the same as the probability of finding 4 people. These probabilities are marked ??? in the table of the distribution.

Find the probability that a household contains 3 people.

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

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?

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

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 probability 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

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

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 = 0.4
Average MCAT: 10; Standard deviation = 1.1
Correlation 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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

13. 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 = 0.4
Average NUNITS: 38; Standard deviation NUNITS: 6.1
Average SAT I: 1010; Standard deviation SAT I = 201
Correlation coefficient for GPA and NUNITS= .35
Correlation coefficient for GPA and SAT= .65
Correlation 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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

14. 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.

(a) 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.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

(continued from above)

(b) Suppose a hypothesis test results in a p-value of .16 or 16%. Explain (using calculations or a verbal response or both) what this p-value means to your good friend who knows little about statistics and hasn't taken Stat 10 yet.

(c) Answer True or False and justify your response (using calculations are OK, but we are looking for a verbal response here for full credit). Achieving statistical significance at the 5% level (for a one-sided test) is like constructing a 90% confidence interval that doesn't contain the parameter.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

15. A simple lottery game called "The Daily 3" is played in California. You pay $1 to play and choose a 3-digit-number. The state of California chooses a 3 digit number at random and pays you $500 if your number is chosen. There are 1,000 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

16. Psychologist Amos Tversky, now deceased, studied public perceptions of probability. Here is an example of this work:

    (a) Tversky asked randomly selected subjects to choose between 2 public health programs that affect 600 people. Program A has the probability of 50% of saving all 600 and probability 50% that all 600 will die. Program B always guarantee to save exactly 400 out of the 600 people. What is the expected number of people saved by Program A?

    (b) Tversky then offered people a different choice. Again, Program A has the probability of 50% of saving all 600 and probability 50% that all 600 will die. Program C will always lose exactly 200 out of the 600 people. What is the difference between this choice and the choice given in part (a)?

    (c) When surveyed, the majority of his randomly selected subjects choose Program B when given the choices in (a). When given the choices in (b), most people chose Program A. Do the subjects appear to use expected values in their choices? Answer yes or no and please explain your choice in plain English.

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

17. Americans spend an average of 12.2 minutes in the shower. If the standard deviation of the variable is 2.3 minutes and the variable is normally distributed:

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

(b) Find the chance that the mean time of a random sample of 144 Americans who shower will be less than 11.9 minutes

(c) Find the percentage of Americans who spend between 11 minutes and 15.5 minutes in the showe

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

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 groups
		Association may point to causation, but association is not the same as causation
		An extremely large biased sample (e.g. > 1000) generates better estimates of the population parameters than a small random sample (e.g. a little over size 100)
		A quantitative variable can be discrete or continuous