Calculator and 4 notes sheets (8.5" by 11", both sides) allowed. Include all five steps outlined in class for the hypothesis tests. Tables will be provided. Length may not be comparable to the final exam. I have not checked it.

- **1.** SAT verbal scores are normally distributed with mean 430 and standard deviation 120.
 - a) What is the probability a randomly selected student's score is over 600?
 - b) What is the minimum score a student must attain to be in the top 1% of all SAT verbal scores?
 - c) What is the probability that the sample mean score for 36 randomly selected students is over 600?
- 2. You play a game where you toss a fair die. If an even number shows up, you win that many dollars. If an odd number shows up, you lose 6 dollars. You pay nothing for the initial toss of the die.
 - a) What is the probability distribution of the random variable *x* = net gain for this game?

X		
p(x)		

- b) Calculate the expected value and standard deviation of x.
- c) Estimate your net gain if you play the game 10,000 times.
- **3.** There are 12 light bulbs in a box, of which 5 are burnt out. Find the probability of randomly selecting 2 burnt out bulbs if:
 - a) Each bulb is replaced before the next is picked
 - b) The bulbs are not replaced after being picked
- **4.** Suppose it is known that 80% of children like the movie <u>Finding Nemo</u>. If 9 randomly selected children are interviewed, what is the probability
 - a) Less than 5 like <u>Finding Nemo</u>
 - b) Between 5 and 7, inclusive, like <u>Finding Nemo</u>
 - c) If 63% of all children watch Sesame Street regularly, what is the probability that, in a random sample of 10 children, exactly 7 watch Sesame Street regularly?
- **5.** Air America is considering a new policy of booking as many as 400 persons on an airplane that can seat only 350. (Past studies have revealed that only 85% of the booked passengers actually arrive for the flight.) Estimate the probability that, if Air America books 400 people, not enough seats will be available.

6. Workers at a business are classified according to religion and gender. The results are shown in the table below.

	Male	Female
Protestant	30	20
Catholic	45	30
Jewish	45	30
Other	7	8

A worker is randomly selected from this business. What is the probability the worker is

- a) Catholic
- b) Catholic, given the worker is a female
- c) Are the events Catholic and Female independent? Why or why not?
- d) Jewish and Male
- e) Jewish or Male
- f) Are the events Jewish and Male mutually exclusive? Why or why not?
- g) Male, given the worker is Jewish

7.

a) Calculate the mean, median and standard deviation. OK to use these functions on your calculator.

1, 3, 4, 7, 10

b) Indicate which histogram (A or B) displays a data set with larger standard deviation.



- c) Would the Empirical Rule apply to the data represented in Histogram A? Why or why not?
- d) Give an example of a data set for which the mean is larger than the median.
- e) Suppose a computer file contains a data set of the family annual incomes for 1000 subjects. The incomes range from \$2000 to \$78,000. By accident, the highest income gets changes to \$786,000. Which of the following summary numbers are affected by the mistake: mean median standard deviation Q_1
- f) Score on an exam have a mean of 75 and a standard deviation of 5. What does Chebychev's Rule tell us about the percent of data that will fall between 55 and 95 points?

- 8. When a coin is tossed 200 times, 120 heads are obtained.
 - a) Obtain a 92% confidence interval for the true proportion of heads.
 - b) Based on your results in part (a), do you think this coin is fair?
- 9. You are about to skydive and mentally perform the following hypothesis test:

H₀: The parachute will not open H_a: The parachute will open

Explain what Type I and Type II errors would be in the context of this problem (no generic definitions, refer to this specific context). What would be possible consequences of each?

Would you prefer $\alpha = 0.01$ and $\beta = 0.05$ or $\alpha = 0.05$ and $\beta = 0.01$? Justify your choice.

- 10. A study is conducted to compare the effectiveness of two vaccines, Vaccine A and Vaccine B. Of the 200 people given Vaccine A, 80 were infected. On the other hand, with Vaccine B, 105 out of 300 were infected.
 - a) Construct a 99% confidence interval for p_A-p_B. Does the interval indicate there is a significant difference in the vaccines?
 - b) Are we guaranteed that p_A - p_B fall in the confidence interval? Explain.
- **11.** A hat contains four slips of paper numbered 1,2,3 and 4. Suppose two slips of paper are selected *with replacement* from the hat.
 - a) List the sample space, S, for this experiment. (Hint: A tree diagram might help.)
 - b) Let the random variable \overline{x} be the mean of the values on the two slips of paper. Give the probability distribution of \overline{x} .
 - c) Find the mean of the distribution in part (b) using the general definition $\mu = \sum_{x} x \cdot p(x)$. Is there an easier way to find this mean?
- **12.** In order to estimate the true mean yield of apple trees in a huge orchard, 40 trees are randomly selected. The mean yield for the sampled trees is 200 pounds with standard deviation 19 pounds.
 - a) Construct a 95% confidence interval estimate of the true mean yield of an apple tree in this orchard.
 - b) Interpret the interval in part (a).
 - c) Do we need to assume apple tree yields are normally distributed to do part (a)? Why or why not?
 - d) True or False. 95% of all apple trees will have a yield that falls in the interval in part (a). Explain.
- **13.** At Santa's Workshop, Inc., the true mean amount of time for an elf to assemble a "Dancing Donkey on a Trampoline" toy is 12.3 minutes. An efficiency expert suggests a new method of assembling the dancing donkey, which he claims will reduce the true mean assembly time. The elves assemble 50 dancing donkey toys using the new method. The average assembly time under the new method is 12.0 minutes with a standard deviation of 1.1 minute. Test the hypothesis that the new method reduces the true mean assembly time at the 0.01 level of significance.

14. A study to estimate the mean body temperature of human adults is conducted. Body temperatures from a random sample of adults are taken and entered into Webstat. The Webstat output is shown below.

μ - mean of variable								
Variable	n	Sample Mean	Std. Err.	L. Limit	U. Limit			
temperature	62	98.30161	0.08241759	98.140076	98.46315			

95% Confidence interval results:

- a) Recall that the Std. Err. = standard error = s/sqrt(n). Find s.
- b) It is commonly believed that, on average, healthy adults have a body temperature of 98.6 degrees Fahrenheit. Conduct a hypothesis test at the 0.10 level of significance to determine if the true mean body temperature differs from 98.6 degrees Fahrenheit.
- **15.** The scatterplot below shows X = the number of people per TV, Y = the average life expectancy for 22 countries.
 - c) Which of the following is the most likely correlation coefficient for the plot? 1.00, 0.80, 0.02, -0.80, -1.00
 - d) From the scatterplot, it appears that countries with fewer people per TV (in other words -- more TVs per capita) generally have greater life expectancy. Can we conclude from this data that access to TV causes people to live longer? Why or why not?
 - e) Suppose the data point labeled Yemen were removed from the plot. The correlation coefficient of the remaining points would most likely be: -1.00, -0.86, -0.60, 0.05, 0.75

