

1 80

Name _____

ID Number _____

**STA 225
Final Examination
April 23, 2003**

Circle the name of your instructor and section:

Hamilton MWF 8:00–9:07 PM

Hamilton MWF 10:40–11:47 AM

Sen TuTh 1:00–2:47 PM

This test contains twenty questions on Part I, fourteen on Part II, and carries a total of 200 points. Please answer ALL questions. Please be sure to write your name and ID number in the spaces above. There are blank pages attached at the end of the examination book which you can use for your work. To get full credit for the questions in Part I, you must show all work. Unsupported answers may receive no credit. Good luck and have a great summer!

Part I ($2 \times 20 = 40$ pts)

For the following **twenty** questions, circle the (only one) letter corresponding to the correct response.

1. Consider the following three variables:

1. race.
2. systolic blood pressure.
3. monthly charge for electricity consumption.

Which of the above are categorical variables?

- (a) only 1. (b) only 2. (c) only 1 and 2. (d) only 1 and 3.
2. The 1958 Detroit Area Study is a major source for the study of the religion and its interaction with other economic and social variables. Questionnaires were filled out by 700 adult residents of the Detroit metropolitan area. One question asked in the Detroit Area Study was "Do you consider yourself a Republican or a Democrat?" Out of the 700 respondents, 27% considered themselves as Republicans. This number 27% is a
(a) correlation. (b) frequency. (c) statistic. (d) parameter.
3. An agronomist wishes to compare the yield of five different varieties of soybean seed under four different fertilizers and a control (no fertilizer). The number of treatments in this experiment is
(a) 2 (b) 5 (c) 20 (d) 25.
4. The histogram for a continuous dataset with mean 65.2 and median 98.6 is most likely
(a) symmetric. (b) skewed to the left. (c) skewed to the right. (d) identical to the normal curve.
5. Upon comparing the course scores among two large sections of STA 225 with equal enrollment, it was found that the upper quartile (Q_3) of Section A was 480 out of a maximum of 600 points and the median of Section B was 480 out of a maximum of 600 points. Which one of the following statements correctly describes the score distributions in these two sections based on this information?
(a) Mean score in Section A is higher than the mean score in Section B.
(b) Standard deviation of scores in Section A is larger than the standard deviation of scores in Section B.
(c) Sections A and B have the same number of students passing the class.
(d) Section A has fewer students scoring 480 or higher than Section B.

6. A ten-question multiple choice exam is given in an STA 225 course at the end of the semester. There is a 25% chance of getting the correct answer to each question by random guessing. If a student in the class attempts to answer each question independently by random guessing, then the probability that she will answer at least one question correctly equals
 (a) $(0.25)^{10}$. (b) $(0.75)^{10}$. (c) $1 - (0.25)^{10}$. (d) $1 - (0.75)^{10}$.
7. A restaurant manager places a card on each table requesting diner's comments on food and service. This is an example of
 (a) randomized block design.
 (b) double blind experiment.
 (c) systematic sampling design.
 (d) voluntary response.
8. In an application form, a question asks for the applicant's "ethnicity". Which of the following is an appropriate graphical display for the frequency distribution of this variable?
 (a) Stem-and-leaf. (b) Histogram. (c) Bar graph. (d) Boxplot.
9. Many board games use a six-faced die to determine the move of each player. Suppose that a die is thrown 5 times and let X equal the number of times a "five" occurs. The mean (μ) and standard deviation (σ) of X are
 (a) $\mu = 1$, $\sigma = 25/36$.
 (b) $\mu = 1/6$, $\sigma = 5/6$.
 (c) $\mu = 5/6$, $\sigma = 25/36$.
 (d) $\mu = 5/6$, $\sigma = 5/6$.
10. A researcher is testing whether the mean age of movie-goers in USA has increased in comparison to that in 1980. She should use
 (a) a two-sided test.
 (b) a one-sided test.
 (c) a confidence interval.
 (d) a sample of size 25.
11. A single dip of a popular brand of ice cream is claimed to weigh 4 ounces by the ice cream store. A regular customer feels that she gets less than 4 ounces of ice cream on the average. Letting μ be the true average amount (ounces) of ice cream in a single scoop, the null and alternative hypothesis for her test should be
 (a) $H_0: \mu \neq 4$, $H_a: \mu = 4$.
 (b) $H_0: \mu \leq 4$, $H_a: \mu > 4$.
 (c) $H_0: \mu \geq 4$, $H_a: \mu < 4$.
 (d) $H_0: \mu = 4$, $H_a: \mu \neq 4$.

12. Which of the following can be a negative number in some instances?

1. P-value.
2. interquartile range of a data set.
3. median of a data set.

(a) Only 1 and 3. (b) Only 2. (c) Only 3. (d) Only 2 and 3.

13. A statistician is interested to know the true proportion (p) of voter support for a candidate for Mayor in his city. From a SRS of 100 voters in his city, he obtains a confidence interval for p as $(0.27, 0.45)$, or equivalently, $0.27 < p < 0.45$. The number of voters in his sample who supported the candidate equals

(a) 36. (b) 27. (c) 45. (d) Can't be determined without further information.

14. A college newspaper wants to determine how the students feel about a proposed new recreation center. In order to conduct such a study, a reporter for the newspaper interviews every 20th student entering the current college gymnasium. This is an example of

- (a) Simple Random Sampling.
- (b) Randomized Block Design.
- (c) Systematic sampling.
- (d) Stratified sampling.

15. If computed from the same data, a 85% confidence interval will be _____ a 95% confidence interval.

- (a) narrower than
- (b) wider than
- (c) of the same width as
- (d) the answer is undeterminable without further information

16. You compute the correlation coefficient between hours of TV watched each week and grade point index for a group of college students, and you obtain $r = -1.25$. This means that

- (a) you made an arithmetic mistake.
- (b) students who watch more TV tend to get lower grades.
- (c) students who watch more TV tend to get higher grades.
- (d) there is little association between TV viewing and grade point index.

17. Let Z be the standard normal random variable. Then $P(Z > 0.59)$ equals

(a) 0.41. (b) 0.23. (c) 0.7224. (d) 0.2776.

18. A sample of production records for an automobile manufacturer shows the following figures for production per shift:

705 700 690 705

The standard deviation for this data is

- (a) 50. (b) 12.25. (c) 7.07. (d) 6.12.
19. For a random experiment with three possible outcomes A_1 , A_2 , and A_3 , which of the following gives a legitimate probability assignment?
- (a) $P(A_1) = 1/2$, $P(A_2) = 1/3$, $P(A_3) = 1/3$.
(b) $P(A_1) = 1/2$, $P(A_2) = 1/6$, $P(A_3) = 1/6$.
(c) $P(A_1) = 1/2$, $P(A_2) = 2/5$, $P(A_3) = 1/10$.
(d) $P(A_1) = 1/2$, $P(A_2) = 3/4$, $P(A_3) = -1/4$.
20. A and B are independent events, such that $P(A) = 0.2$, $P(B) = 0.7$. Which of the following statements is/are correct?

1. $P(A \text{ and } B) = 0.14$.
2. $P(A | B) = 0.2$.
3. $P(A \text{ or } B) = 0.76$.

- (a) only 1. (b) only 1 and 2. (c) only 2 and 3. (d) All of 1, 2 and 3.

Part II

1. ($3 + 3 + 5 + 7 = 18$ pts) The following data represent the lengths of long-distance phone calls (in minutes) made by a business office in one day:

3 7 2 14 4 29 3 9 1 20 10 7 2 41 3 5

- (a) Obtain the mean length of long-distance calls for the above data.
- (b) Obtain the median length of long-distance calls for the above data.
- (c) Obtain the interquartile range for this sample.
- (d) Draw a box plot for the data. Does the data contain any outliers? Indicate outliers (if any) on your boxplot.

2. (6 + 6 = 12 pts) Dr. Paul Oswiecinski randomly selects twenty-five of his 20-29-old-patients and obtains the following data regarding their serum HDL cholesterol:

	71	56	48	48	53	52	66	36	49	35
	28	58	62	45	65	38	73	45	51	56
	46	69	71	33	59					

(a) Construct a stemplot for the data.

(b) Draw a histogram for this data with equal width and class intervals of 20-30, 30-40 etc. Label your histogram properly.

3. (2 + 2 + 3 + 2 = 9 pts) A group of 150 executives is classified according to the status of body weight and incidence of hypertension. The number of executives falling in different categories are presented in the following table.

	Overweight	Normal Weight	Underweight	Total
Hypertensive	15	12	3	30
Not Hypertensive	22	68	30	120
Total	37	80	33	150

A person is selected at random from this group of executives.

- (a) What is the probability that the person selected will have hypertension?
- (b) What is the probability that the chosen person is “Not hypertensive” and “Underweight”?
- (c) What is the probability that selected person is either hypertensive or underweight?
- (d) The selected person is found to be overweight. Given this, what is the conditional probability that the person is also hypertensive?

4. (10 + 2 = 12 pts) For a random sample of seven gasoline fueled subcompact and compact cars, the engine size (x) in cubic inches of cylinder volume, and estimated miles-per-gallon (MPG) ratings (y) according to the 1986 EPA gas mileage guide are obtained with the following summary statistics:

$$r = -0.804, \bar{x} = 153, \bar{y} = 27, s_x = 92.6, s_y = 7.75.$$

- (a) Using the above information, obtain the least squares regression line for predicting the gas mileage of a car from its cylinder volume.

- (b) Based on your regression line, predict the miles-per-gallon rating of a Toyota Corolla whose cylinder volume is 135 cubic inches.

5. ($2 \times 4 = 8$ pts) An engineer in an automobile manufacturing plant wishes to compare the durability of a new front wheel brake pad with that of the standard one. He takes fifteen new cars of identical make and model and installs the new and standard brake pads simultaneously on the two front wheels of each car, one in each, randomizing between the left and right wheels. After a test drive of 1000 miles, he measures the percent corrosion in each brake pad.
- (a) What is the name of this type of comparative experiment?
 - (b) What is the response variable in this experiment?
 - (c) How many treatments are there?
 - (d) What are the experimental units: cars or the wheels?
6. (7 pts) The current version of a certain type of tires is known to give an average of 20000 miles of tread wear with a standard deviation (σ) of 2000 miles. A change in the manufacture of the tire has been proposed which, it is hoped, will increase the average tread wear without changing the standard deviation. What is the minimum number of prototypes that should be manufactured and tested with the new design in order to estimate the true mean tread life to within ± 500 miles with 95% confidence?

7. (8 pts) A new alloy has been devised for use in a space shuttle. Tensile strength measurements are made on a random sample of 64 pieces of the alloy, and the sample mean and the sample standard deviation of these measurements were found to be $\bar{x} = 39.3$ and $s = 2.6$, respectively. Assuming that the tensile strength measurements are normally distributed, find a 90% confidence interval for the true (population) mean strength μ of the alloy.

8. (10 pts) The weight of the eggs produced by a certain breed of hen is normally distributed, with mean $\mu = 65$ grams and standard deviation $\sigma = 5$ grams. Think of carton of such eggs as SRS's of size 12 from the population of all such eggs. What is the probability that the weight of a randomly chosen carton is more than 825 grams?

9. (8 pts) Transmission failures on the XPS automobile are normally distributed with a mean lifetime of $\mu = 70000$ miles and a standard deviation of $\sigma = 10000$ miles. For a randomly chosen automobile of this type, what is the probability that the transmission fails beyond 80000 miles?

10. (10 pts) The time it takes a randomly selected job applicant to perform a certain task is normally distributed with a mean of 120 seconds and a standard deviation of 20 seconds. The fastest 10% are selected for an *Advanced Training* workshop. What task time is the cut-off to qualify for the *Advanced Training* workshop?

11. (12 pts) A recently marketed drug called *Caridez*, is claimed to be more effective than the standard drug in curing a certain kind of gum disease. In a study conducted by the dental researchers, it was found that out of 200 patients with the specific gum disease who were given *Caridez*, 136 were cured. Suppose it is known that the cure rate of the standard drug is 60%. Based on the sample data from the researchers' findings above, you have to find out whether the claim about the new drug is justified. For this purpose, you have to formally perform a statistical test of hypothesis. Answer the following queries which build the steps of the formal test.

1. Write down the appropriate null and alternative hypotheses for this problem.

2. Compute the value of the relevant test statistic.

3. Compute the P-value.

4. Suppose it has been decided to conduct the test at 5% level of significance. Based on your computed P-value, answer the original claim.

12. (7 + 7 = 14 pts) An automobile club wants to estimate the proportions of different types of *Emergency Road Service* (ERS) requests. Upon examining a sample of 2927 ERS calls, it found that 1499 calls related to starting problems, 849 calls involved serious mechanical failures requiring towing, 498 calls involved flat tires or lockouts, and 81 calls were for other reasons.

(a) Calculate a 95% confidence interval for the true proportion of ERS calls that relate to starting problems.

(b) This year, another automobile club wishes to conduct a similar survey. They are most concerned about the accurate estimation of proportion of ERS requests that require towing. After a brainstorming session with their statistician, the management decided that they would like to be 90% confident that the true proportion p of ERS calls requiring towing should not deviate from their estimate by more than .04. If the statistician decides to use the survey result given above in order to calculate her prior guess p^* of p , what is the minimum number of ERS calls she should sample?

13. (12 pts) The Wall Street Journal (Dow Jones Travel Index) reported that the average cost of renting a car in Orlando, Florida is about 38 dollars per day. It is suspected that the actual average is higher than the reported number. To test this suspicion, an independent agency contracted by Florida tourism made a survey of 16 randomly chosen vacationing families from various Disney resorts. They found that for this group of people the car rental fees per day had a sample average of $\bar{x} = 40.40$ dollars with a sample standard deviation of $s = 3.76$ dollars. Assuming that the daily car rental values come from a normally distributed population, test at 5% level of significance whether the suspicion can be substantiated by the data. Show all steps.

Right x Left	26	14	40
Left x Right	12	8	20
Total	38	22	60

The above findings demonstrate an association between the handedness of parents and their biological offspring. In order to answer this question, you are to carry out a chi-square test of independence following the steps below.

Step 1: Specify the null and alternative hypotheses clearly. Verbal description is okay, but you have to specify the hypotheses in the context of the problem.

Step 2: Fill in the following table with the corresponding expected frequencies. Show calculations.

Parents' Handedness	Right	Left	Handedness of Biological Offspring
	Right x Right	Right x Left	
Left	Left x Right	Left x Left	
Total			

14. (20 pts) In a study of possible genetic influence of parental hand preference, a sample of 400 children was classified according to their own handedness and the handedness of their biological parents. The following table documents the findings.

Parents' Handedness (Father × Mother)	Handedness of Biological Offspring		Total
	Right	Left	
Right × Right	302	38	340
Right × Left	26	14	40
Left × Right	12	8	20
Total	340	60	400

Do these findings demonstrate an association between the handedness of parents and their biological offspring? In order to answer this question, you are to carry out a χ^2 test of independence following the steps below.

Step 1: Specify the null and alternative hypotheses clearly. Verbal description is okay, but you have to specify the hypotheses in the context of the problem.

Step 2: Fill in the following table with the corresponding **expected frequencies** (Show calculations).

Parents' Handedness (Father × Mother)	Handedness of Biological Offspring	
	Right	Left
Right × Right		
Right × Left		
Left × Right		

Step 3: Calculate the value of the χ^2 statistic.

Step 4: Carry out the relevant test of hypothesis at 5% level of significance. Clearly state the rejection region and your conclusion.