

Please note: This sample Final Exam is from Dr. Scott Stevens. It has 55 questions. Each will be 0.182 points for a total 10 points.

COB 191
Statistical Methods
Final Exam

Name _____
Dr. Scott Stevens

You can see the answers to these questions by putting your cursor over the yellow boxes. Try it with this box. (Note: With Office XP, the comments will appear on the side of the page.)

DO NOT TURN TO THE NEXT PAGE UNTIL YOU ARE INSTRUCTED TO DO SO!

The following exam consists of 55 questions, each worth 1.8 points. (This totals to 99 points; you get an additional one point for stacking your scantron in the same orientation as the others in the pile.) You will have 120 minutes complete the test. This means that you have, on average, a bit over 2 minutes per question.

- Record your answer to each question on the scantron sheet provided. You are welcome to write on this exam, but your scantron will record your graded answer.
- Read carefully, and check your answers. Don't let yourself write nonsense.
- Keep your eyes on your own paper. If you believe that someone sitting near you is cheating, raise your hand and quietly inform me of this. I'll keep an eye peeled, and your anonymity will be respected.
- If any question seems unclear or ambiguous to you, raise your hand, and I will attempt to clarify it.
- Be sure you correctly record your student number on your scantron, and blacken in the corresponding digits. **Failure to do so will cost you 10 points on this exam!**

Pledge: On my honor as a JMU student, I pledge that I have neither given nor received unauthorized assistance on this examination.

Signature _____

1. Using information obtained from a sample to estimate a characteristic of a population is called
 - a. descriptive statistics
 - b. inferential statistics
 - c. popular statistics
 - d. analytical statistics
 - e. enumerative statistics

2. Which of the following can be reduced by proper interviewer training?
 - a. Sampling error
 - b. Measurement error
 - c. Population standard deviation
 - d. Confidence level
 - e. All of the above

Refer to the following for questions 3 and 4.

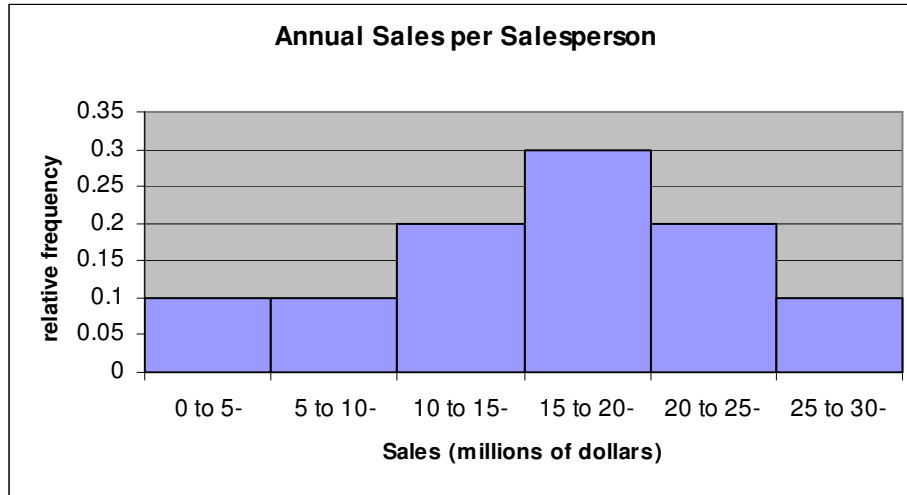
The percentage distribution below represents results from surveying 200 people with a second job and determining how much yearly income this second job generated.	
<u>yearly income</u>	<u>percentage distribution</u>
\$0 but less than \$3000	40%
\$3000 but less than \$6000	30%
\$6000 but less than \$9000	10%
\$9000 but less than \$12000	20%

3. Of the 200 people surveyed, how many earned at least \$6000 in their second job?
 - a. 20 people
 - b. 30 people
 - c. 40 people
 - d. 60 people
 - e. 140 people

4. What percent of the people earned at least \$3000 but less than \$9000 in their second job?
 - a. 20%
 - b. 40%
 - c. 50%
 - d. 60%
 - e. 80%

Refer to this chart for questions 5 - 7.

The histogram below represents the annual sales of a sample of 200 sales employees for a large firm.



5. What percent of employees sold less than \$20 million?
- a. 20%
 - b. 30%
 - c. 60%
 - d. 70%
 - e. 80%
6. How many employees had at least \$10 million in sales?
- a. 40 employees
 - b. 60 employees
 - c. 120 employees
 - d. 160 employees
 - e. 180 employees
7. 90% of the salespeople had at least _____ in annual sales.
- a. \$5 million
 - b. \$10 million
 - c. \$15 million
 - d. \$20 million
 - e. \$25 million

Refer to the following data for questions 8, 9, and 10.

The 11 numbers below represent the number of grams of carbohydrates
in one serving of 11 different breakfast cereals.

13 15 24 30 21 16 19 24 16 24 18

The mean of these values is 20.

8. The median number of grams of carbohydrate is:
- a. 16
 - b. 17.5
 - c. 18.5
 - d. 19
 - e. 24
9. The range of the sample is
- a. 5
 - b. 10
 - c. 11
 - d. 17
 - e. 30
10. The standard deviation of this sample, to three decimal places, is
- a. 4.86
 - b. 5.10
 - c. 26.0
 - d. 260
 - e. 265
11. Suppose A and B are independent events where $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \text{ and } B) = 0.2$ then $P(A | B) =$
- a. 0.02
 - b. 0.10
 - c. 0.20
 - d. 0.40
 - e. 0.80

Questions 12 - 14 refer to the information below.

A survey is taken among customers of a popular nightclub to determine patrons' age and marital status. Of the 140 respondents selected, 91 were under 30 and 105 were single. 21 of the married customers were 30 or older.

12. What is the probability that a randomly selected individual is married?
- a. $35/140$
 - b. $21/140$
 - c. $91/140$
 - d. $105/140$
 - e. $77/140$
13. What is the probability that a randomly selected individual is single and under 30?
- a. $14/140$
 - b. $21/140$
 - c. $28/140$
 - d. $77/140$
 - e. $105/140$
14. If a randomly selected individual is married, what is the probability that this individual is 30 or older?
- a. $14/35$
 - b. $77/105$
 - c. $28/105$
 - d. $21/35$
 - e. $21/49$
15. If two events are mutually exclusive, the probability of their intersection, that is, $P(A \text{ and } B)$, is equal to
- a. 0
 - b. 0.50
 - c. $P(A) \times P(B)$
 - d. $P(A) + P(B)$
 - e. Cannot be determined from the information given.

16. The marketing manager of a toy firm is planning to introduce a new toy. In the past, 40% of the toys introduced by the company have been successful. Before the toy is actually marketed, market research is conducted and a report, either favorable or unfavorable, is compiled. It has been shown that 80% of the successful toys received a favorable market research report. It has also been shown that 40% of the unsuccessful toys also received favorable reports. Suppose that market research gives a favorable report for this latest toy. What is the probability that the toy will be successful? (Apply Bayes' Theorem.)
- 0.43
 - 0.57
 - 0.64
 - 0.75
 - 0.80
17. In a binomial distribution
- the random variable X is continuous.
 - the probability of success p is stable from trial to trial.
 - the number of trials n must be at least 30.
 - the results of one trial are dependent on the results of the other trials.
 - All of these statements (a-d) are true.
18. On the average, 1.8 customers per minute arrive at any one of the checkout counters of a grocery store. What type of probability distribution can be used to find out the probability that there will be no customer arriving at a checkout counter during a given minute of time?
- binomial distribution
 - Poisson distribution
 - normal distribution
 - F distribution
 - chi-squared distribution
19. The local police department must write on average 5 tickets a day to keep department revenues at budgeted levels. Suppose the number of tickets written per day follows a Poisson distribution with a mean of 6 tickets per day. Interpret the value of the mean.
- The number of tickets that is written most often is 6 tickets per day.
 - About half of the days have less than 6 tickets written and about half of the days have more than 6 tickets written.
 - If we recorded the number of tickets written each day for a long period of time, the arithmetic average of these values would be very close to 6.
 - 6 tickets are written every day.
 - The number of tickets written on any day is within 6 tickets of the number of tickets written on any other day.

20. The standard error of the mean

- a. is never larger than the standard deviation of the population.
- b. decreases as the sample size increases.
- c. measures the variability of the mean from sample to sample.
- d. is used to construct the confidence interval for μ .
- e. All of the above.

21. The Central Limit Theorem is important in statistics because it says that

- a. for a large n , the population is always approximately normal.
- b. for any population, the sampling distribution of the mean is approximately normal.
- c. for a large n , the sampling distribution of the mean is approximately normal.
- d. for any binomial population, the sampling distribution of the mean is approximately normal.
- e. if the sampling distribution of the mean is normal, then so is the population.

22. Which of the following four statements about the sampling distribution of the sample mean is **incorrect**? Assume all samples discussed have the same size, n . If none of the four statement is incorrect, choose answer e.

- a. The sampling distribution of the mean is approximately normal whenever the sample size is sufficiently large ($n \geq 30$).
- b. The sampling distribution of the mean is generated by repeatedly taking samples of size n and computing the sample means.
- c. The mean of the sampling distribution of the mean is equal to μ .
- d. The standard deviation of the sampling distribution of the mean is equal to σ .
- e. None of the four statements above is incorrect; they all are true.

23. You and a friend will play a game of “matching coins”. You both flip coins of equal value, and if the two coins are both heads or both tails, your friend wins. If one coin shows heads and the other tails, you win. The winner gets to keep both coins. You are going to play 10 times. Let W be the total number of cents you win in the 10 rounds. Note that if you lose money overall, W will be negative. Which of the following statements are then necessarily true?

- I. The mean of W is 0.
- II. W is a random variable.
- III. The size of σ_W does not depend on whether you play for pennies or nickels.

- a. I only b. II only c. III only d. I and II only e. I, II and III

24. Why is the Central Limit Theorem so important to the study of sampling distributions of the mean and the proportion?
- It allows us to disregard the size of the sample selected when the population is not normal.
 - It allows us to disregard the shape of the sampling distribution when the size of the population is large.
 - It allows us to disregard the size of the population we are sampling from when it is large.
 - It allows us to disregard the shape of the population when n is large.
 - It allows us to compute the area under any section of the standard normal curve.
25. At a computer manufacturing company, the actual size of computer chips is normally distributed with a mean of 1 centimeter and a standard deviation of 0.1 centimeter. A random sample of 12 computer chips is taken. What is the standard error for the sample mean?
- 0.029
 - 0.050
 - 0.091
 - 0.100
 - 0.120
26. In a certain problem, the standard error of the mean for a sample of 100 is 30. In order to cut the standard error of the mean to 15, we would
- increase the sample size to 115.
 - increase the sample size to 200.
 - increase the sample size to 400.
 - decrease the sample size to 50.
 - decrease the sample to 25.
27. The number of violent crimes committed in a day possesses a distribution with a population mean of 1.5 crimes per day and a population standard deviation of 4 crimes per day. Suppose a large number of random samples of 100 days each were drawn from this population and for each sample the mean number of crimes per day was calculated. If a histogram of these sample means were created we would expect it to be:
- approximately normal with mean = 1.5 standard deviation = 4
 - shape unknown with mean = 1.5 and standard deviation = 4
 - approximately normal with mean = 1.5 and standard deviation = 0.4
 - shape unknown with mean = 1.5 and standard deviation = 0.4
 - approximately normal, with a mean of 0.15 and a standard deviation of 4.

28. If we repeatedly draw (from a population) samples of size n and calculate the median for each sample, then the accumulation of these sample medians would result in _____.
- a confidence interval for the population variation
 - the sampling distribution for the population mean
 - the nonrejection region for the median
 - an estimate of the sample median
 - the sampling distribution of the median
29. X is normally distributed with a mean of 32 and a standard deviation of 8. What Excel calculation would give the probability that X is greater than 43?
- =NORMSINV(0.25)
 - =NORMSDIST(1.375)
 - =NORMSINV(5.375)
 - =1 - NORMSDIST(1.375)
 - =1 - NORMSDIST(5.375)
30. Consider this question: “What interval (centered on 100) will contain 80% of all observations in a normally distributed population with a mean of 100 and a standard deviation of 10?” Which of the following calculations would provide the **upper cutoff** for this range?
- = NORMSINV (0.8) * 10
 - = NORMSINV(0.9) * SQRT(10)
 - = TINV(0.8, 99) * 100
 - = 100 + NORMSINV(0.8) * SQRT(10)
 - = 100 + NORMSINV(0.9) * 10
31. Consider this question: “According to government data, 25% of employed women have never been married. If 5 employed women are selected at random, what is the probability that exactly 2 have never been married?” Which of the following calculations would provide the answer to this question? (You may assume that the population of employed women is essentially infinite.)
- = NORMSINV(0.25) * SQRT((0.4*0.6)/5)
 - = NORMSINV(0.4) * SQRT((0.25*0.75)/5)
 - = BINOMDIST(2, 5, 0.25, TRUE)
 - = BINOMDIST(2, 5, 0.25, FALSE)
 - = BINOMDIST(0.4, 5, 0.75, TRUE)

32. Consider this question: “*Suppose that, on average, three customers arrive per minute at a bank during the noon hour. What is the probability that two customers will arrive in a given minute?*” An Excel computation giving the answer is

- a. = BINOMDIST(2, 3, 2, FALSE)
- b. = BINOMDIST(2, 3, 2, TRUE)
- c. = POISSON(2, 3, FALSE)
- d. = NORMSDIST(2/3)
- e. = TDIST(2, 3, 2)

33. A _____ is the corresponding statistic that is used to estimate the parameter of interest.

- a. point estimate
- b. sampling fraction
- c. sample coefficient
- d. population proxy
- e. null parameter

Use this information for problems 34 and 35.

The quality control manager at a light bulb factory needs to estimate the average life of a shipment of light bulbs. The shipment standard deviation for the life of these light bulbs is known to be 100 hours. A random sample of 50 light bulbs taken from the shipment yields a sample average life of 350 hours. The manager wishes to set up an 85% confidence interval estimate of the true average life of light bulbs in this shipment.

34. What is the parameter of interest for this problem?

- a. population percentage
- b. sample standard deviation
- c. population mean
- d. sample mean
- e. population standard deviation

35. If the manager constructs a confidence interval for this problem, what would be the purpose of doing so?

- a. to estimate the parameter of interest
- b. to estimate the sample proportion
- c. to estimate the sample mean
- d. both a and b
- e. both b and c

36. When reporting the results of a confidence interval estimation, what three pieces of information would it be most important to include?
- The parameter being estimated, the confidence level and the limits of the interval.
 - The true value of the parameter being estimated, the value of the statistic that estimates that parameter, and the level of confidence.
 - The sample size, the sample mean, and the point estimate.
 - The standard error, the point estimate, and the margin of error.
 - The limits of the confidence interval, the size of the sample, and the size of the population.
37. The owner of a small deli wants to estimate the population mean expenditure for customers at the deli. The population standard deviation is \$2.25. The expenditure per customer is normally distributed. A sample of size 16 is selected and yields a mean expenditure per customer of \$7.20. The 95% confidence interval for the population mean is
- \$3.51 to \$10.89
 - \$4.95 to \$9.45
 - \$6.10 to \$8.30
 - \$6.27 to \$8.13
 - \$6.48 to \$7.92
38. A researcher will do an interval estimate of the mean income of employees. If a maximum error in estimation of $\pm \$7$ is desired with a confidence level of 90%, what size sample would be used? Assume that the population standard deviation is approximately \$30.
- about 30
 - about 40
 - about 50
 - about 60
 - about 70
39. Which of the following is most closely synonymous with the term “standard error”?
- standard deviation of the population
 - standard deviation of the sample
 - standard deviation of the sampling distribution
 - sampling error
 - estimator bias

40. Which of the following statements could correctly be used as the null hypothesis for an hypothesis test?

- I. The mean of the population is equal to 55.
- II. The mean of the sample is equal to 55.
- III. The mean of the population is greater than 55.

a. I only b. II only c. III only d. I and III only e. II and III only

41. Which of these tools is the most useful when studying the relationship between responses to two **categorical** variables?

- a. paired differences
- b. scatter plot
- c. crosstab table
- d. Fisher's F statistic
- e. hypothesis test for difference of two means

42. If an economist wishes to determine if there is evidence that average family income in a community exceeds \$25,000

- a. either a one-tailed or two-tailed test could be used.
- b. a one-tailed test should be used, with rejection occurring in the upper tail.
- c. a one-tailed test should be used, with rejection occurring in the lower tail.
- d. a two-tailed test should be used.
- e. the number of tails used depends on whether variances can be assumed equal.

Use this information for questions 43, 44, and 45:

A chemical process is set to produce an average of 8.81 tons per hour with a population standard deviation of 0.20 tons. The population is normally distributed. The most recent sample of $n = 24$ show an average production of 8.89 tons. Hypothesis testing at the 0.1 level of significance will be performed by the manager to see if the process average has changed.

43. What is the null and hypothesis?

- a. $H_0: \mu = 8.81$
- b. $H_0: \mu = 8.89$
- c. $H_0: \mu \neq 8.81$
- d. $H_0: \mu \neq 8.89$
- e. $H_0: \mu \geq 8.89$

44. Which calculation would give the (upper) **critical value** for the test statistic?

- a. = NORMSINV(0.11)
- b. = NORMSINV(0.89)
- c. = NORMSINV(0.945)
- d. = TINV(0.11, 23)
- e. = TINV(0.89, 23)

45. Which calculation would give the test statistic value (t or z) for this sample?

- a. = $8.89 - 8.81 / (24 * 0.2)$
- b. = $(8.89 - 8.81) / \text{SQRT}(0.2)$
- c. = $(8.89 - 8.81) / \text{SQRT}(24)$
- d. = $(8.89 - 8.81) / (0.2 / \text{SQRT}(24))$
- e. = $(8.89 - 8.81) / 0.2$

Use this information for questions 46 and 47:

A manufacturer of television picture tubes has a production line that is designed to produce an average of 100 tubes per day. A new safety device has been installed, and the manufacturer believes that it may have reduced the average daily output. A random sample of 30 days output after the installation was taken. The manager conducts the appropriate (lower tailed) hypothesis test at the 0.10 level of significance, based on this sample.

46. The manager conducts this hypothesis test using a piece of software which reports both the one-tailed and two-tailed p values for the sample. Under which of the following circumstances should he conclude that the average daily output has dropped below 100 tubes per day? [You may assume that the average output of the sample is, indeed, less than 100.] Be careful!

- I. The one tailed (lower tail) p value of the sample is 0.003
- II. The one tailed (upper tail) p value of the sample is 0.933
- III. The two tailed p value of the sample is 0.121

- a. I only b. II only c. III only d. I and III e. I, II, and III

47. The appropriate hypothesis test for this problem is a(n)

- a. Fisher F test
- b. t test
- c. z test
- d. χ^2 test
- e. binomial test

48. If p is very small for a two-tailed test, we should

- a. reject the null hypothesis.
- b. reject the alternate hypothesis.
- c. fail to reject the null hypothesis
- d. fail to reject the alternate hypothesis.
- e. take another sample.

Use this information for questions 49 through 52:

The marketing branch of the Mexican Tourist Bureau would like to increase the percentage of tourists who purchase silver jewelry while in Mexico from its present estimated population percentage of 40%. Toward this end, promotional literature describing the beauty of the jewelry is distributed to all passengers on airplanes arriving at a seaside resort during a one-week period. A sample of 500 passengers returning at the end of the one-week period is randomly selected, and 220 of these passengers indicate that they have purchased jewelry. At the .05 level of significance, the Tourist Bureau wishes to conduct an hypothesis to see if there is evidence that the proportion of tourists buying silver jewelry has increased above the previous value of 40%.

49. What are the null and alternate hypothesis?

- a. $H_0: p \leq 0.44$ versus $H_1: p > 0.44$
- b. $H_0: p \leq 0.40$ versus $H_1: p > 0.40$
- c. $H_0: p < 0.40$ versus $H_1: p \geq 0.40$
- d. $H_0: p < 0.44$ versus $H_1: p \geq 0.44$
- e. $H_0: p = 0.40$ versus $H_1: p \neq 0.40$

50. If you were to do this hypothesis test by using the p -value approach, the first step toward computing that p would be to compute the standard error of the proportion. What calculation would provide the value of the standard error of the proportion?

- a. $= \text{SQRT}(0.4 \cdot 0.6 / 500)$
- b. $= \text{SQRT}(0.44 \cdot 0.56 / 500)$
- c. $= (0.44 - 0.40) / \text{SQRT}(500)$
- d. $= (220 - 200) / \text{SQRT}(500)$
- e. $= \text{SQRT}(200 / 220)$

51. If you were to do this hypothesis test by using the p-value approach, your second step would be to compute the z-score of your test statistic. You would compute the z rather than the t because
- the population standard deviation is known
 - the population standard deviation is unknown
 - the problem involves a single population
 - the problem involves two populations
 - the problem involves proportions
52. If you were to do this hypothesis test by using the p-value approach, your third step would be to compute the p statistic from your sample based on the z score that you found in problem 51. Pretend that this z-statistic turned out to be 3.5. Then correct p score for this sample in this upper tailed test is given by the Excel calculation
- = NORMSDIST(3.5)
 - = 1 - NORMSDIST(3.5)
 - = 2 * NORMSDIST(3.5)
 - = 0.5 * NORMSDIST(3.5)
 - = 0.4 * NORMSDIST(3.5)
53. The graph of X,Y pairs represented by dots is called
- a scatter diagram or scatter plot
 - a frequency histogram
 - a coefficient plot
 - a binomial probability diagram or plot
 - a pareto diagram
54. The _____ measures the strength of the linear relationship between two numerical variables.
- coefficient of variation
 - coefficient of correlation
 - level of significance
 - mode
 - power test
55. Which of the following quantities does not make sense for ordinal data?
- mean
 - median
 - mode
 - all of these quantities make sense for ordinal data
 - none of these quantities make sense for ordinal data