Final Exam Statistics 216 Spring 2005

Name _____ Section _____

In each of the following True/False questions (1-10) circle the best answer. Each True/False question is worth 2 points.

1. **TRUE FALSE**
   The α level is the probability of rejecting the null hypothesis if it is false.

2. **TRUE FALSE**
   The population distribution depends on the size of a SRS.

3. **TRUE FALSE**
   It is not possible to make both a Type I and a Type II error on the same test.

4. **TRUE FALSE**
   The P-value is the probability that $\bar{x} = \mu_0$.

5. **TRUE FALSE**
   The mean of a t-distribution with $n - 1$ degrees of freedom is always equal to 0.

6. **TRUE FALSE**
   In a matched pairs study, subjects are matched in pairs and the outcomes are compared within each matched pair.

7. **TRUE FALSE**
   The standard error of the sample mean is a parameter.

8. **TRUE FALSE**
   Formal statistical inference is important because it can help correct for flaws in study design.

9. **TRUE FALSE**
   The total area under a density curve is always equal to 1.

10. **TRUE FALSE**
    A statistical inference procedure is called **robust** if the probability calculations required are insensitive to violations of the assumptions made.

In each of the following multiple choice questions (11-32) choose the **single best** answer. Each multiple choice question is worth 3 points.

11. Which of the following is **not** a parameter in the simple linear regression model?

    A) β_0
    B) β_1
    C) σ
    D) ϵ

12. A test of $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ is to be conducted using a SRS of size n. The population standard deviation is unknown and has to be estimated. Which of the following is the correct pairing of the test statistic and it's sampling distribution under H_0 for this problem?

    A) $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$; $N(0,1)$

    B) $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$; $t(n-1)$

    C) $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$; $N(0,1)$

    D) $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$; $t(n-1)$

13. Other things being equal, which of the following P-values provides the strongest evidence against the null hypothesis?

    A) 0.77

    B) 0.03

    C) 0.45

    D) 0.999

14. A regression analysis has resulted in a least squares regression line of

$$\hat{y} = 10 - 0.75x$$

    with $r^2 = 0.64$. Which of the following is **not** a conclusion we can draw from these results?

    A) There is a strong negative linear relationship between x and y.

    B) The correlation between x and y is -0.8.

    C) The correlation between x and y is -0.75.

    D) 64% of the variation in y is explained by the least squares regression of y on x.

15. For an exam given to a class, the student's scores ranged from a 35 to 98, with a mean of 74. Which of the following is the most realistic value for the standard deviation?

    A) 13

    B) 3

    C) 63

    D) -10

16. Which of the following is **not** an assumption of the simple linear regression model?

    A) The deviations ϵ_i are assumed to be normally distributed with mean 0 and standard deviation σ.

    B) The mean of y varies linearly with x.

    C) The variance of y varies linearly with x.

    D) The deviations ϵ_i are assumed to be independent of one another.

17. A 95% confidence interval for a mean is to be computed using data from a SRS of $n = 11$. The standard deviation is unknown and will be estimated. The correct critical value to use in the confidence interval is

    A) 1.96

    B) 2.201

    C) 2.228

    D) 2.262

18. A call-in poll conducted by *USA Today* concluded that Americans love Donald Trump. *USA Today* later reported that 5640 of the 7800 calls for the polls came from the offices owned by a single fan of Mr. Trump. The results of this poll are probably

    A) reliable because it was conducted by a nationally recognized organization.

    B) biased, but only slightly because the sample size is so large.

    C) biased understating the popularity of Donald Trump.

    D) biased overstating the popularity of Donald Trump.

19. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is

    A) yield of the crop

    B) inches of water

    C) the experimenter

    D) amount of yearly rainfall.

20. The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures motivation, attitude toward school, and study habits of students. Scores range from 0 to 200. The mean score for the population of U.S. students is about 115, and the standard deviation is about 30. A teacher who suspects that older students have better attitudes toward school gives the SSHA to 25 students who are at least 30 years of age. Their mean score is $\bar{x} = 132.2$. What are the correct null and alternative hypotheses?

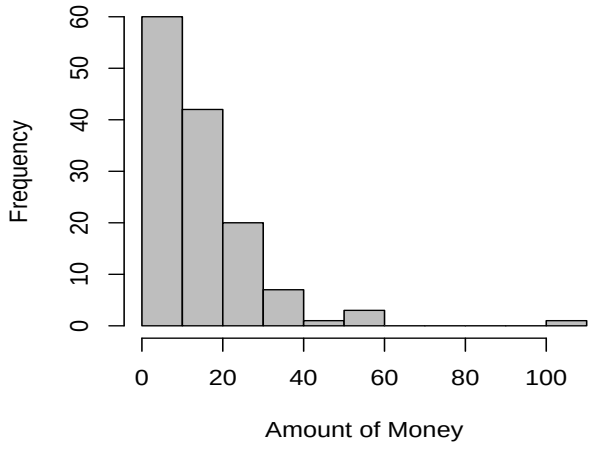
    A) $H_0 : \mu = 115$ vs. $H_a : \mu \neq 115$

    B) $H_0 : \mu \neq 115$ vs. $H_a : \mu > 115$

    C) $H_0 : \mu = 115$ vs. $H_a : \mu = 132.2$

    D) $H_0 : \mu = 115$ vs. $H_a : \mu > 115$

In a large class with 136 students, the professor records how much money each student has in their possession during the first class of the semester. A histogram of the data is shown below. Only the left endpoint is included in each interval. **This graph will be used to answer questions 21 and 22. Note: the vertical axis is frequency, not relative frequency**.



21. The percentage of students with under \$10.00 in their possession is closest to

    A) 35%
    B) 45%
    C) 60%
    D) 70%

22. The above histogram

    A) is skewed right.
    B) has an outlier.
    C) is not symmetric.
    D) all of the above.

23. Which of the following could differ for 2 random samples of the same size from the same population?

    A) The population mean μ.
    B) The population distribution.
    C) The sampling distribution of the sample mean.
    D) The standard error of the sample mean.

24. In 2000, the scores of men on the math part of the SAT approximately followed a normal distribution with a mean of 533 and a standard deviation of 115. A test score is drawn at random from this population and its standardized value is found to be $z = 2$. What is the test score?

A) 763

B) 648

C) 533

D) 115

25. Which of the following is a correct assignment of probabilities to the values of a random variable X that can take on the values 1, 2, 3, 4, or 5?

A)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.2	0.1

B)

Value of X	1	2	3	4	5
Probability	−0.1	0.2	0.3	0.2	0.4

C)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.4	0.5

D)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.2	0.2

26. A random variable X has mean μ_X and standard deviation σ_X. Suppose n independent observations of X are taken and the average of these n observations is computed. We can assert that if n is very large, the sampling distribution of the average is approximately normal. This assertion follows from

A) the law of large numbers

B) the central limit theorem.

C) the definition of sampling distribution.

D) the bell curve.

27. Suppose 50 different survey organizations take simple random samples of 400 persons age 25 and over in the same large city. Each organization computes the mean income of it's sample and a 95% confidence interval for the true unknown population mean income. **The population standard deviation is unknown and must be estimated by each organization**. Which of the following statements is true?

A) The width of the 50 intervals will be the same but they will have different centers.

B) The width and center of the 50 intervals will be the same.

C) The width of the 50 intervals will be the different but they will have the same centers.

D) The width and center of the 50 intervals will all be different.

28. In a statistical test of hypotheses, we say that the data are statistically significant at level α if

    A) $\alpha = 0.05$
    B) α is small
    C) the P-value is less than or equal to α
    D) the P-value is greater than α.

29. When possible, the best way to establish that an observed association is the results of cause-and-effect is by means of

    A) the least-squares regression line.
    B) the correlation r
    C) examining z-scores rather than the original variables.
    D) a well-designed experiment.

30. Four students in a statistics class each took a simple random sample of size 10 from their class and calculated the mean height. They got 64 inches, 67 inches, 60 inches, and 62 inches. This is an example of

    A) sampling variability.
    B) bias.
    C) confounding.
    D) lack of control.

31. A set of 7 systolic blood pressures is

    110, 123, 132, 160, 127, 118, 102

    What is the median systolic blood pressure?

    A) 100
    B) 123
    C) 124.57
    D) 160

32. A researcher has data on 2 populations:

    Population 1 - heights of children in a single kindergarten class
    Population 2 - heights of all children in an elementary school.

    Which of the following is true?

    A) The standard deviation of Population 1 is likely to be greater than the standard deviation of Population 2.
    B) The standard deviation of Population 1 is likely to be less than the standard deviation of Population 2.
    C) The standard deviation of the 2 populations is likely to be about the same.

You must **show all your work** *in any of the following problems requiring computations to receive full or partial credit.*

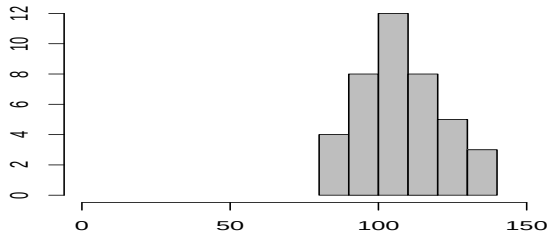
33. A time and motion study measures the time required for an assembly line worker to perform a repetitive task. The data show that the time required to bring a part from a bin to its position on an automobile chassis (X) varies from car to car with mean $\mu_X = 11$ seconds and standard deviation $\sigma_X = 2$ seconds. The time required to attach the part to the chassis (Y) varies with mean $\mu_Y = 20$ seconds and standard deviation $\sigma_Y = 4$ seconds. You can assume that X and Y are independent of one another.

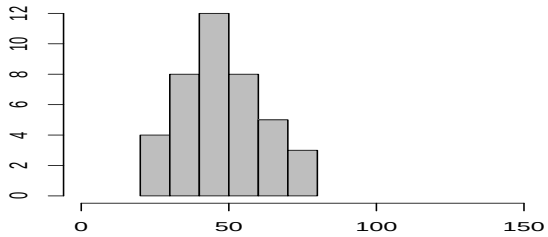
    A) What is mean time required for the entire operation of positioning and attaching the part? (3pts)

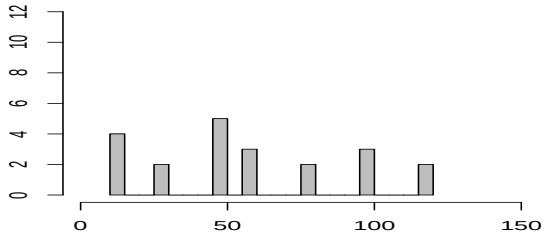
    B) What is the standard deviation of the time required for the entire operation of positioning and attaching the part? (3pts)

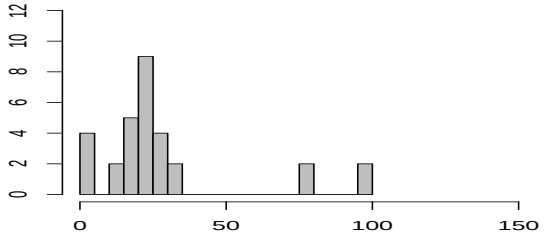
34. In 2000, the scores of men on the math part of the SAT approximately followed a normal distribution with a mean of 533 and a standard deviation of 115. A test score is drawn at random from this population. What is the approximate probability it is less than 418? (3pts)

35. The figure below shows histograms of 4 sampling distributions of statistics intended to estimate the parameter of **50**. Label each distribution relative to the others as high or low bias and high or low variability. (8 pts)









Histogram A _____ bias; _____ variability

Histogram B _____ bias; _____ variability

Histogram C _____ bias; _____ variability

Histogram D _____ bias; _____ variability

36. Read each brief report of statistical research and identify whether it was an observational study or an experiment. (2pts each)

A) Over a 4-month period, among 30 people with bipolar disorder, patients who were given a high dose of omega-3 fats from fish oil improved more than those given a placebo.

B) An examination of the medical records of more than 360,000 Swedish men showed that those who were overweight or who had high blood pressure had a higher risk of kidney cancer.

C) Some people who race greyhounds give the dogs large doses of vitamin C in the belief that the dogs will run faster. Investigators at the University of Florida tried 3 different diets in random order on each of 5 racing greyhounds. They were surprised to find that when the dogs ate high amounts of vitamin C they ran more slowly.

D) The journal *Circulation* reported that among 1900 people who had heart attacks, those who drank an average of 19 cups of tea a week were 44% more likely than non-drinkers to survive at least 3 years after the attack.

37. Product preference depends in part on the age, income, and gender of the consumer. A market researcher selects a large sample of potential car buyers. For each consumer, she records **gender**, **age**, **household income**, and **automobile preference**. Which of these 4 variables are categorical, and which are quantitative? (4pts)

A) Categorical Variables -

B) Quantitative Variables -

38. Data for y =assessed value in dollars and x =size in square feet of 18 houses in Ithaca, NY resulted in the following regression output. You can assume that all assumptions are satisfied.

Predictor	Coef	StDev	T	P
Constant	37108.8	8664	4.283	0.000
Study	11.8987	6.02	?	0.066

A) What is the equation of the least squares line? (2pts)

B) What is the estimated mean increase in assessed value per 1-square foot increase in size.? (2pts)

C) Calculate an approximate 95% confidence interval for β_1, and **interpret the interval in terms of the problem**. (6pts)

D) Using proper statistical notation, write null and alternative hypotheses for assessing the statistical significance of the linear relationship between the response and explanatory variables? (2pts)

E) We omitted from the output the t-statistic for testing the hypotheses in part D. Compute the value of this t-statistic using other information in the output **and** give the degrees of freedom associated with the statistic. (3pts)

F) Using the information in the output carry out a test of the hypotheses you gave in part D. Use $\alpha = 0.01$. Give your decision (reject or fail to reject H_0), justify your decision, and draw a conclusion in terms of the problem. (5pts).

39. The scores on an easy exam have mean = 88, standard deviation =10, minimum = 65, $Q1 = 77$, median=85, $Q3 = 91$, maximum=100. Sketch a box plot, labeling which of these values are used in the plot. (5pts)