22S:008 Final Exam, May 12, 2000 Name:_____

Form A ID: _____

Please circle one. Discussion Section: 9:30, 10:30, 11:30, or 12:30

Please enter all of your answers on these exam pages. There are 50 questions.
The Defective Question Report, Formula sheet, and Tables will be handed out separtely.

1. Consider a *t*-distribution with 15 degrees of freedom. Which of the following is the value we look up in the *t*-distribution table to construct a 99% confidence interval for μ?

   A) 2.602
   B) 2.624
   C) 2.947
   D) 2.977
   E) 3.000

2. Consider testing the null hypothesis H_0: $\mu = 50$. True or false? The test statistic $\sqrt{n}((\bar{y} - 50)/\sigma)$ is called the ***t-statistic***.

   A) True B) False

3. In the beer-tasting example, *Consumer Reports* said "Although the 11 correct choices were an improvement over guesswork, we don't consider that to be statistically significant evidence that a beer drinker can tell domestic *Lowenbrau* from *Miller High Life*." True or false: This means ***they rejected the null hypothesis***.

   A) True B) False

4. If a process is stable (that is, in control), the process will produce goods or services that meet the customers' specifications.

   A) True B) False

5. Failing to reject a true null hypothesis is a correct decision.

   A) True B) False

6. A normal distribution has a mean of 100 and a standard deviation of 16. What percentage of the data values lie above the value 108?

   A) 31%
   B) 69%
   C) 95%
   D) 99%
   E) 99.7%

7. The probability of a Type II error (computed under the assumption that the null hypotesis is true) is called the ***significance level*** of the statistical test.

   A) True B) False

8. ***Sampling Error*** is the difference between the value of a sample estimate and the corresponding value in the population that is due to misleading and ambiguous questions.

   A) True B) False

9. The distribution of a sample statistic over all possible samples is called a ***sampling distribution***.

   A) True B) False

10. The set of all measurements on a variable in a universe is called a ***population***.

    A) True B) False

11. Tampering with a process that is in statistical control will usually increase the variability in the process.

    A) True B) False

12. Consider 4 Bernoulli trials with success probability $\pi = 0.2$. Which of the following gives the probability of observing ***at least three successes***?

    A) $(0.2)^4$

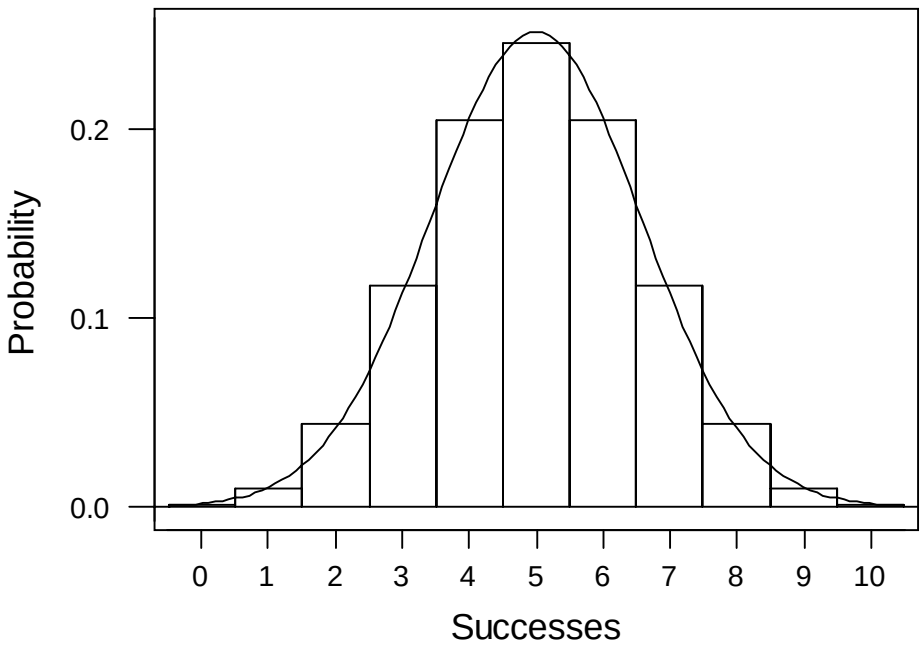
    B) $4(0.2)^3(0.8)$

    C) $4(0.2)^3(0.8) + (0.2)^4$

    D) $(0.8)^4$

    E) $1 - (0.8)^4$

13. The graph at the below displays the Binomial distribution with 10 trials and success probability .5. A normal curve with the same mean and same standard deviation is also shown. The shaded area gives the normal approximation for the chance of



    A) exactly 7 successes using the continuity correction.

    B) 7 or more successes using the continuity correction.

    C) exactly 7 successes without using the continuity correction.

    D) 7 or more successes without using the continuity correction.

    E) exactly 7 successes with no approximation.

14. Control charts reflect both the cross-sectional and longitudinal aspects of a process.

    A) True B) False

15. The standard deviation of the distribution of the number of successes in *n* trials of a Bernoulli process is largest when $\pi = 0.5$.

    A) True B) False

16. In a simple random sample, increasing the sample size will increase the margin of sampling error (other things being equal).

    A) True B) False

17. A freight train pulls 25 cars. Historically, the distribution of weights for individual cars has a mean of 50 tons and a standard deviation of 10 tons. If the total weight of the cars exceeds 1500 tons, the train cannot travel safely through the mountains. What is the chance that the train will not be able to travel safely through the mountains?

    A) 1%

    B) 5%

    C) 16%

    D) 325

    E) The chance is effectively zero.

18. A politician is interested in her constituents views on Proposition X. In particular, she wants to know if a majority of her constituents favor Proposition X. Let π denote the proportion of her constituents who favor Proposition X. Which of the following give appropriate hypotheses for her to test?

    A) $H_0: \pi = 0.5$ versus $H_1: \pi < 0.5$

    B) $H_0: \pi = 0.5$ versus $H_1: \pi > 0.5$

    C) $H_0: \pi < 0.5$ versus $H_1: \pi = 0.5$

    D) $H_0: \pi > 0.5$ versus $H_1: \pi = 0.5$

    E) $H_0: \pi \geq 0.5$ versus $H_1: \pi \leq 0.5$

19. In a taste test subjects were given identical looking glasses of regular *Pepsi* and the new low calorie *Pepsi One*. They were asked to identify the *Pepsi One*. Suppose we decide to reject the null hypothesis of random guessing ($\pi = 0.5$) only if we get 9 or 10 correct responses. Which of the following expressions gives the significance level of the test?

    A) $(0.5)^{10}$

    B) $9(0.5)^{10}$

    C) $10(0.5)^{10}$

    D) $11(0.5)^{10}$

    E) None of the above.

20. In multiple regression modeling, the most basic residual plot is the plot of residuals versus the corresponding fitted values.

    A) True B) False

21. A large organization surveyed their employees with a random sample of size 400. Only 20% of the sample said they were in favor of changing to a new pension plan. What is the upper endpoint of a 95% confidence interval for the proportion of all employees who favor the new pension plan?

    A) 0.20

    B) 0.24

    C) 0.25

    D) 0.95

    E) 0.975

22. The standard deviation of a statistic is also called the ***standard error*** of the statistic.

    A) True B) False

23. A company selected a random sample of 100 accounts receivable. For these accounts they calculated a mean of $340 and a standard deviation of $70. What is the lower endpoint of a 95% confidence interval for the mean of the population of all accounts receivable? (Assume a large population.)

A) $200
B) $326
C) $340
D) $354
E) $480

24. A survey organization would like to estimate a population percentage in a large population to an accuracy of 5 percentage points with 95% confidence. They have no current knowledge of the size of the percentage. How large should a random sample be to acheive their goal for the margin of error?

A) 100
B) 200
C) 400
D) 1100
E) None of the above.

25. Five percent of the products produced by a certain stable process are defective. If a random sample of 4 items is taken, what is the probability that the sample contains ***at most one*** defective? (to the nearest hundredth)

A) 0.17
B) 0.19
C) 0.81
D) 0.83
E) 0.99

26. Suppose that 30 subgroups (samples) each of size 50 are selected from a Bernoulli process. If there are 150 defectives overall, what is the value of the upper control limit (UCL) for the *p*-chart for fraction defective? (Rounded to the nearest hundredth.)

A) 0.10
B) 0.13
C) 0.16
D) 0.23
E) 0.26

27. The regression model $\hat{y} = b_0 + b_1 x + b_2 z$ where x and is a continuous variable, and z is a binary indicator variable is best described as

A) one straight line
B) two parallel lines
C) two non-parallel lines
D) a quadratic curve
E) a regression plane

28. A ***statistic*** is a numerical characteristic of a sample.

A) True B) False

29. The Central Limit Effect applied to a sample proportion p from n trials of a Bernoulli process says that for large n the distribution of p is approximately normal with mean π and standard deviation $\sqrt{(\pi(1-\pi))/n}$.

A) True B) False

30. If data have a very strong upward but curved relationship, it is possible to get a correlation coefficient that is larger than +1.

A) True B) False

31. In general, the larger the residual standard deviation, the better the regression model.

A) True B) False

32. Each year the General Social Survey, GSS, selects 1500 households throughout the U.S. using complicated randomization methods. People in the selected households are interviewed extensively to gather social science data. Such a study is best classified as

A) a designed experiment with blinding of the subjects
B) an observational study
C) a designed study
D) a double blind, designed experiment
E) a designed experiment with neither blinding nor double blinding

33. In the Physicians' Health Study 22,000 male doctors were used as subjects to measure the effect of a simple drug in the prevention of heart attacks. In this study aspirin was used as a **placebo**.

A) True B) False

34. The Physicians' Health Study used **double blinding**.

A) True B) False

35. The Physicians' Health Study failed to use **randomization** to decide which subjects received the real drug and which received the placebo.

A) True B) False

36. A prediction based on a regression line is likely to be very precise when

I. the residual sum of squares is small
II. the residual standard deviation is small
III. the value of R^2 is close to 100%

A) I only
B) II only
C) III only
D) I, II, and III
E) None of the above

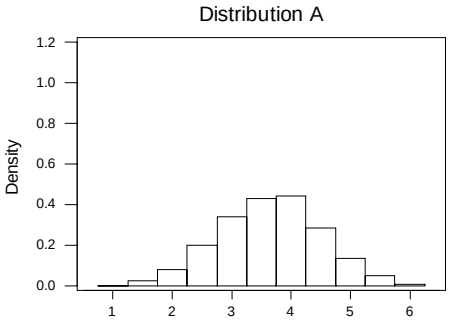
37. In a perfectly symmetric distribution the distance from the first quartile to the median is the same as the distance from the median to the third quartile.

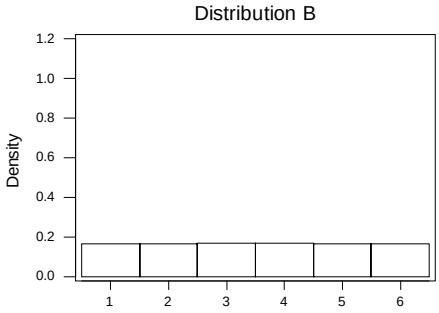
A) True B) False

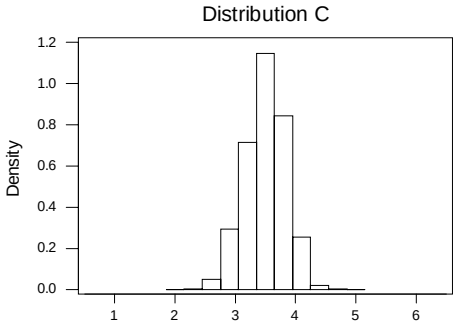
38. What is the 90th percentile of a normal distribution with mean 75 and standard deviation 8.6?

A) 1.28
B) 10
C) 86
D) 90
E) 92

39. Three distributions are shown at the right. One represents the distribution of individual values, one the distribution of means from random samples of size 4, and one the distribution of means from random samples of size 25. Which is the **correct order** for: individuals, mean of 4, and mean of 25?

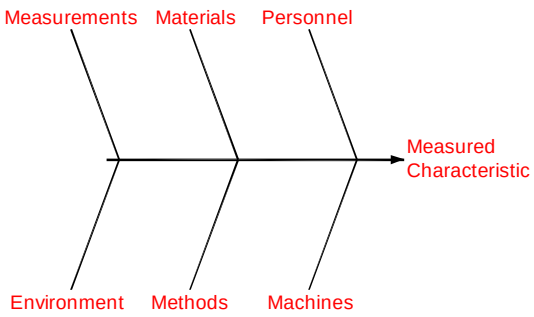






A) A, B, C
B) C, B, A
C) C, A, B
D) B, C, A
E) B, A, C

40. The diagram at the right is called:



A) A fishbone diagram
B) An Ishikawa diagram
C) A Cause-and-Effect diagram
D) All of the above
E) None of the above

41. A longitudinal study studies a process variable over time.

A) True B) False

42. Seasonality in a sequence means that observations that are close in time are also close in value. Observations that are far apart in time may be very different in value.

A) True B) False

43. Special causes are considered to be due to chance and to remain in the system unless the process is itself altered.

A) True B) False

44. Which of the following statistics are resistant to outliers?

I. The median
II. The interquartile range
III. The standard deviation

A) I and II only
B) I and III only
C) II and III only
D) I, II, and III
E) None of the above.

45. The table below gives the breakdown of shoe sale transactions over a long period for a large shoe chain. An auditor selects one transaction at random from all of the transactions.

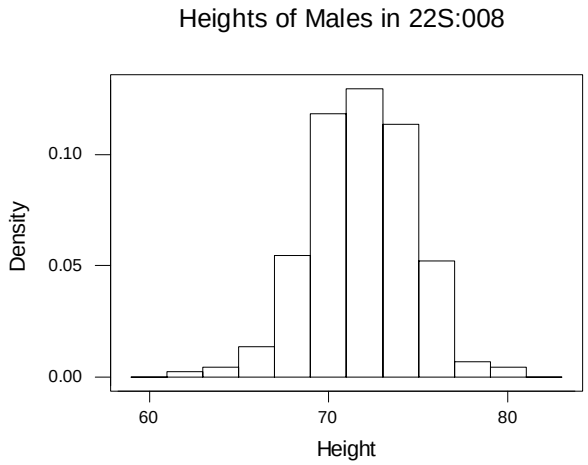
		Brand Purchased		
		Reebok	**Nike**	**Other**
Type of Purchase	**Cash**	5%	10%	5%
	Credit Card	40%	15%	25%

What is the chance that a Credit Card transaction is selected?

A) 20%
B) 35%
C) 55%
D) 80%
E) None of the above.

46. The density histogram at the right displays the distribution of heights of the 220 males in a class. The mean is 71.3 inches. Which of the following values could be the standard deviation of the distribution? (Notice that the distribution is quite mound-shaped.)

A) 1.0
B) 2.7
C) 7.9
D) 11.3
E) 15.2



47. If you are interested in predicting future behavior of a ***meandering process***, it would be more helpful to look at a histogram of the data rather than a sequence plot.

A) True B) False

48. A regression model was used to explain young girls' Heights (in inches) from their Age (in years). The Minitab results are shown below.

```
The regression equation is
Height = 27.6 + 2.58 Age

Predictor        Coef       StDev          T        P
Constant      27.6242      0.6716      41.13    0.000
Age           2.58424     0.08163      31.66    0.000

S = 1.349       R-Sq = 98.7%     R-Sq(adj) = 98.6%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      1824.7      1824.7   1002.20    0.000
Residual Error    13        23.7         1.8
Total             14      1848.4
```

From this regression output we see that the **residual standard deviation** is

A) 1.349

B) 1.8

C) 23.7

D) 1848.4

E) None of the above.

49. Refer to question 48. From this regression output we see that the least-squares **slope** coefficient is

A) 1.349

B) 2.58

C) 23.7

D) 27.6

E) 98.7%

50. Refer to question 48. True or false: From this regression output we see that the model is slightly better if we **do not adjust** the R^2 value.

A) True B) False

Defective Question Report

Name:_____
Section:_____

If you believe that a test question is defective in some way, please list your complaint here. All complaints will be considered in our interpretation of the test results.

Please tell us your test Form: A, B, C, D

Question number: _____ Your answer: _____
Your complaint:

Question number: _____ Your answer: _____
Your complaint:

Question number: _____ Your answer: _____
Your complaint: