Final Exam A Statistics 216 Fall 2004

Name _____ Section _____

In each of the following multiple choice questions choose the **single best** answer. Each multiple choice question is worth 4 points.

1. A sample of 9 men between the ages of 30 and 39 years old is asked to do as many situps as they can in one minute. The sample mean number of situps is $\bar{x} = 26.2$ and the sample standard deviation is $s = 6$. A 95% confidence interval is

   A) $26.2 \pm 2.262\frac{6}{\sqrt{9}}$

   B) $26.2 \pm 2.262\frac{6}{\sqrt{8}}$

   C) $26.2 \pm 2.306\frac{6}{\sqrt{9}}$

   D) $26.2 \pm 2.306\frac{6}{\sqrt{8}}$

2. The probability model below describes the probability of the number of customers that a pawn shop may receive during an hour.

No. of Customers	2	3	4	5
Probability	0.1	0.3	0.4	0.2

   The mean number of customers per hour is

   A) 3.7

   B) 3.5

   C) 0.925

   D) 5

3. The probability a randomly chosen male has Type O blood is 0.45. What is the probability a randomly chosen male does not have Type O blood?

   A) 0.45

   B) 0.50

   C) 0.55

   D) Cannot tell because we are not given the probability distribution.

4. A doctor monitoring the level of phosphorus in the blood of a dialysis patient took a SRS of 6 measurements and recorded a sample mean of $\bar{x} = 5.4$ milligrams of phosphorus per deciliter of blood (mg/dl). The sample standard deviation is $s = 0.9$ mg/dl. What is the standard error of the sample mean?

   A) 0.9

   B) 0.402

   C) 0.367

   D) 5.4

5. Leah is flying from Boston to Denver with a connection in Chicago. The probability her first flight leaves on time is 0.15. If the flight is on time, the probability that her luggage will make the connecting flight in Chicago is 0.95, but if the first flight is delayed, the probability that the luggage will make it is only 0.65. Let A be the event her first flight leaves on time and B be the event that her luggage makes the connecting flight. Which of the following statements is true?

   A) The events A and B are independent.

   B) The events A and B are not independent.

   C) The event A is the complement of event B.

   D) The events A and B are disjoint.

6. In 2001 a report in the *Journal of the American Cancer Institute* indicated that women who work nights have a 60% greater risk of developing breast cancer. Researchers based these findings on the work histories of 763 women with breast cancer and 741 women without the disease. This is an example of

   A) an observational study.

   B) a randomized controlled experiment.

   C) a placebo-controlled experiment.

   D) an experiment without double-blinding (the women knew whether or not they had breast cancer).

7. A statistical inference procedure is called **robust** if

   A) the probability calculations required are insensitive to violations of the assumptions made.

   B) the mean of the sampling distribution is equal to the true value of the parameter being estimated.

   C) the underlying population distribution is normal.

   D) the Law of Large Numbers holds.

8. An investigator has a computer file showing family incomes for 1000 subjects in a certain study. These range from \$5,800 a year to \$ 98,600 a year. By accident, the highest income in the file gets changed to \$986,000. Which of the following is a true statement about the effect on the median?

   A) The median increases.

   B) The median stays the same.

   C) The median decreases.

   D) There is not enough information given to answer this question.

9. Which of the following is not correct about the regression model,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i?$$

A) The ϵ_i's are assumed to be independent of one another.

B) The ϵ_i's are assumed to follow a normal distribution.

C) The mean of the responses (the y_i's) is assumed to be 0.

D) The relationship between y and x is assumed to be linear.

10. Suppose that women always married men two years younger than themselves. The correlation between ages of women and their spouses would be

A) between -1 and 0.

B) exactly -1.

C) between 0 and 1.

D) exactly 1.

11. You can graphically examine the relationship between two quantitative variables with a

A) histogram.

B) scatter plot.

C) stem and leaf plot.

D) bar graph.

12. An investigator measuring various characteristics of a large group of athletes found that the correlation between the weight of an athlete and the amount of weight that athlete could lift was 0.60. Which of the following is true?

A) 60% of an athlete's lifting ability can be attributed to weight alone.

B) On average, an athlete can lift 60% of his body weight.

C) If an athlete gains 10 pounds, he can expect to lift an additional 6 pounds.

D) Heavier athletes tend to be able to lift more weight.

13. Each set of five numbers below has mean 6. From inspecting them, which has largest standard deviation? You do not need to compute the standard deviation to do this problem.

A) 6 6 6 6 6

B) 4 4 6 8 8

C) 5 5 6 7 7

D) 2 2 6 10 10

14. A randomly selected US registered voter is asked to respond *"yes"*, *"no"*, or *"maybe"* to the question "Do you intend to vote in the next presidential election?" The sample space is S = {*yes, no, maybe*}. Which of the following represents a legitimate assignment of probabilities for this sample space?

    A) 0.5, 0.5, 0.5
    B) 0.4, 0.4, 0.4
    C) 0.2, 0.3, 0.5
    D) -0.3, 0.3, 1.0

15. A company institutes an exercise break for its workers to see if this will improve job satisfaction, as measured by a questionnaire that assesses worker's satisfaction. Ten randomly selected workers take part in the study and their job satisfaction is assessed prior to and after implementation of the exercise program. The number of degrees of freedom associated with the t test statistic for the difference is

    A) 10
    B) 9
    C) 20
    D) 19

16. A one-sample t-test from a sample of $n = 20$ observations for the two-sided test of $H_0 : \mu = 25$ vs. $H_a : \mu \neq 25$ is $t = 1.65$. Which of the following is a correct statement about the P-value?

    A) $0.05 < P < 0.10$
    B) $0.01 < P < 0.025$
    C) $0.001 < P < 0.0025$
    D) $0.10 < P < 0.20$

17. The significance level α is

    A) the probability of getting a value of the test statistic as extreme or more extreme than the one observed.
    B) the probability of rejecting the null hypothesis if it is true (a Type I error).
    C) the probability of failing to reject the null hypothesis if it is false (a Type II error).
    D) the true probability that the null hypothesis is true.

18. Nitrous oxide emissions for a certain kind of car varies from car to car according to a normal distribution with mean $\mu = 5$ grams/mile and standard deviation $\sigma = 0.5$ grams/mile. A simple random sample of $n = 16$ cars is taken. The sampling distribution of the sample mean is

    A) $N(5, 0.5)$
    B) $N(0, 1)$
    C) $t(5, 0.125)$
    D) $N(5, 0.125)$

19. A 95% confidence interval for a population mean is $(1, 6)$. The margin of error is

    A) 1.96
    B) 2.5
    C) 3.5
    D) 5

20. Other things being equal which of the following P-values provides the weakest evidence against the null hypothesis?

    A) 0.001
    B) 0.03
    C) 0.45
    D) 0.77

In each of the following True/False questions (21-30) circle the best answer. Each True/False question is worth 2 points.

21. **TRUE** **FALSE**
    Hypotheses in significance testing apply only to the samples on which they are based.

22. **TRUE** **FALSE**
    The standard error of the sample mean is a statistic and the standard deviation of the sample mean is a parameter.

23. **TRUE** **FALSE**
    It is not possible to find a statistically significant linear relationship at the $\alpha = 0.05$ level between an explanatory variable x and a response variable y if $r^2 < 0.05$.

24. **TRUE** **FALSE**
    A null hypothesis will be rejected if the P-value is greater than α.

25. **TRUE** **FALSE**
    A P-value is always calculated assuming the null hypothesis is true.

26. **TRUE** **FALSE**
    A P-value in a z-test of a mean is the probability $\bar{x} = \mu_0$.

27. **TRUE** **FALSE**
    The median is a resistant measure of the center of a distribution.

28. **TRUE** **FALSE**
    The t distributions are symmetric, centered at zero, and as n increases, they approach the N(0,1) distribution.

29. **TRUE** **FALSE**
    An alpha level (level of significance) should be specified after a p-value has been found.

30. **TRUE** **FALSE**
    If a distribution is roughly symmetrical with no outliers and the sample is moderate to large then t procedures are robust against non-normality of the population.

Show your work problems and short answer problems.

31. IQ scores from a test are known to be normally distributed with mean 100 and standard deviation of 16.

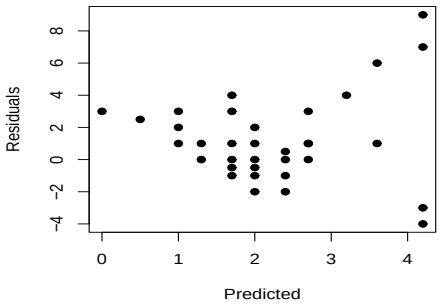
    A) What proportion of the scores is above 116? (3pts)

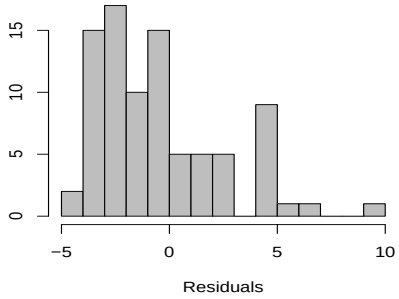
    B) What is the 90th percentile of the population of IQ scores? (3pts)

32. In a clinical trial, data collection usually starts at *baseline* when the subjects are recruited into the trial but before they are assigned to treatment and control groups. Two clinical trials on a new weight loss drug reported baseline data on weight shown below. Investigators in one of these trials were accused of using a biased randomization procedure, i.e. subjects were assigned to treatment and control groups in a biased manner. Which trial and why? (4pts)

Trials	Group	Number of Persons	Average Weight (lbs)	Standard Deviation
1	Treatment	1012	285	25
	Control	997	243	26
2	Treatment	995	266	27
	Control	1017	263	25

33. Below is a residual plot and a histogram of residuals from a regression analysis. What 3 assumptions of the simple linear regression model are violated and why? (6 points)





34. A data set gives y = mean April temperature (degrees Fahrenheit) and x = degrees latitude for 20 U.S. cities. Minitab output for the linear regression is shown below.

Predictor	Coef	SE Coef	T	P
Constant	118.78	4.47	26.57	0.000
latitude	-1.64	0.12	?	0.000

A) Write null and alternative hypotheses for testing whether there is a linear relationship between mean April temperature and degrees latitude. Use proper statistical notation. (2pts)

B) Note that the t-statistic is missing in the above table. Use the information in the table to compute it, and give the degrees of freedom associated with the t-statistic. (3pts)

C) Interpret the value of -1.64 in terms of the problem. (2pts)

D) Write down the equation of the least squares regression line and use it to predict the mean April temperature of a city at 40 degrees latitude. (3pts)

E) Construct a 95% confidence interval for the true slope and interpret it in terms of the problem. (4pts)

35. A plant scientist examines the cellulose content of a variety of alfalfa hay. A previous study claimed that the mean cellulose content was $\mu = 140$ mg/g. The agronomist believes the mean is higher than that and will conduct a test of significance to see if there is evidence for his belief. A SRS of 16 cuttings has mean cellulose content $\bar{x} = 145$ mg/g. The sample standard deviation is $s = 8$. Show work on B and C to receive full credit.

A) State the appropriate null and alternative hypotheses. (2pts)

B) Compute the value of the test statistic. (2pts)

C) Give the narrowest possible bounds on the P-value. (2pts)

D) Assuming an α-level of 0.05 do you reject or fail to reject H_0? You must justify your answer for full credit. (2pts)

E) Draw a conclusion in terms of the problem. (2pts)