Review Time

1) A dummy variable can be assigned up to three values.
   a) True
   b) False

2) Transformations may be used when nonlinear relationships exist between the response and explanatory variable when performing regression.
   a) True
   b) False

3) The value of the coefficient of determination can never decrease when more variables are added to the model.
   a) True
   b) False

Review Time

4) For statistical tests of significance about the regression coefficients, the null hypothesis is that the slope is 1.
   a. True
   b. False

5) If the assumptions of regression have been met, residuals plotted against the independent variable(s) will typically show patterns.
   a) True
   b) False

6) The noise in a regression model is assumed to have zero variance.
   a. True
   b. False

Review Time

11) If the equation of the least squares regression line was computed to be y=45.7+3.1x, then the correlation cannot be less than 0.

a. True
b. False

12) If the equation of the regression line that relates percent blood alcohol (x) to reaction time in milliseconds (y) is y=36 - 1.3x, then the slope tells us that for every percent increase in blood alcohol, we can expect reaction time to go down by 1.3 milliseconds

a. True
b. False

13) A researcher found the correlation between age of death and number of cigarettes smoked per day to be -0.95. Based just on this information, the researcher can justly conclude that smoking causes early death.

a. True
b. False

Review Time

18) A least-squares regression line is not just any line drawn through the points of a scatterplot. What is special about a least-squares regression line?

a) It passes through all the points.
b) It minimizes the squared values of the data.
c) It has slope equal to the correlation between the two variables.
d) It minimizes the sum of the squared vertical distances of the data points from the line.

Review Time

20) Suppose that the least-squares regression line for predicting y from x is $y = 100 + 1.3x$. Which of the following is a possible value for the correlation between x and y?

a) 1.3
b) –1.3
c) 0
d) –0.5
e) 0.5

Review Time

25) Which of the following is NOT an assumption of the Binomial distribution?

a) All trials must be identical.
b) All trials must be independent.
c) Each trial must be classified as a success or a failure
d) The number of successes in the trials is counted.
e) The probability of success is equal to .5 in all trials.

Review Time

34) The weight of a gum drop (piece of candy) in ounces is normally distributed with mean 2 and standard deviation 0.25. A bag contains 10 independent gum drops. The probability that the total weight of the gum drops in the bag exceeds 20 ounces is

a) 0.25
b) 0.5
c) 0.33
d) 0.75
e) 0.35

```
1-pnorm(2,2,.25/sqrt(10))
[1] 0.5
```

Review Time

36) The purpose of hypothesis testing is to help the researcher reach a conclusion about _____ by examining the data contained in _____.

a) a population, a sample
b) an experiment, a computer printout
c) a population, an event
d) a sample, a population

Review Time

37) If the coefficient of determination (R^2) is 0.80, then which of the following is true regarding the slope of the regression line?

a) All we can tell is that it must be positive.
b) It must be 0.80
c) It must be 0.89.
d) Cannot tell the sign or the value.
e) The slope must be significant.

Review Time

39) A multiple regression model with two independent variables exhibits a highly significant F-ratio, but each variable's individual t-statistic is insignificant. The most likely cause of such a situation is

a) Heteroskedasticity
b) Homoskedasticity
c) Multicollinearity
d) Non-normality of residuals

Review Time

41) What is the meaning of the term "heteroscedasticity"?

a) The variance of the errors is not constant
b) The variance of the dependent variable is not constant
c) The errors are not linearly independent of one another
d) The errors have non-zero mean

Review Time

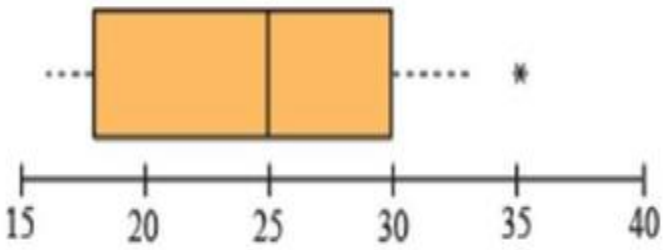
61) Suppose we obtain the following regression model for baseball bat sales (Y) when regressed against seasonal indicator variables; $\hat{y} = 100 - 40Spring + 20Wtr - 15Fall$. If we decide to make the baseline season Fall, what would then be the resulting coefficient for Winter (Wtr)?

a) 25
b) -40
c) 30
d) 15
e) None of the above

85+? = 120 so ? = 35

Review Time

65) Season's Pizza delivers food items to homes in their local area. The following box-and-whisker plot describes the distribution for delivery times in minutes.



Based on this plot, which one of the following statements is correct?

A) The average delivery time is 25 minutes.
B) There are no outliers in this data set.
C) The 75th percentile in this data set is 30 minutes.
D) The second quartile is approximately 18 minutes.
E) None of the above

Review Time

43) Which of the following can NOT be answered from a regression equation?

a) Predict the value of y at a particular value of x.
b) Estimate the slope between y and x.
c) Estimate whether the linear association is positive or negative.
d) Estimate whether the association is linear or non-linear

Review Time

42) Suppose you have estimated wage = 5 + 3education + 2gender – edu*gender, where gender is one for male and zero for female. Suppose instead that gender had been one for female and zero for male. Under this coding what would be the the sum of the coefficients for the gender and interaction variables? (that is we want $b_{gender} + b_{edu*gender}$)

a) -3
b) -1
c) 0
d) 1
e) 2

Original model
Females: w=5+3edu
Men: w=7+2edu

New model
W = 7+2edu-2gender+edu*gender