STA 3024 **Practice Problems Exam 2**

NOTE: These are just Practice Problems. This is NOT meant to look just like the test, and it is NOT the only thing that you should study. Make sure you know all the material from the notes, quizzes, suggested homework and the corresponding chapters in the book.

1. The parameters to be estimated in the simple linear regression model $Y=\alpha+\beta x+\varepsilon$ $\varepsilon\sim N(0,\sigma)$ are:
a) α, β, σ b) $\alpha, \beta, \varepsilon$ c) a, b, s d) ε, 0, σ

2. We can measure the proportion of the variation explained by the regression model by:
a) r b) R^2 c) σ^2 d) F

3. The MSE is an estimator of:
a) ε b) 0 c) σ^2 d) Y

4. In multiple regression with p predictor variables, when constructing a confidence interval for any β_i, the degrees of freedom for the tabulated value of t should be:
a) n-1 b) n-2 c) n- p-1 d) p-1

5. In a regression study, a 95% confidence interval for β_1 was given as: (-5.65, 2.61). What would a test for H_0: $\beta_1=0$ vs H_a: $\beta_1 \neq 0$ conclude?
a) reject the null hypothesis at $\alpha=0.05$ and all smaller α
b) fail to reject the null hypothesis at $\alpha=0.05$ and all smaller α
c) reject the null hypothesis at $\alpha=0.05$ and all larger α
d) fail to reject the null hypothesis at $\alpha=0.05$ and all larger α

6. In simple linear regression, when β is **not** significantly different from zero we conclude that:
a) X is a good predictor of Y b) there is no linear relationship between X and Y
c) the relationship between X and Y is quadratic d) there is no relationship between X and Y

7. In a study of the relationship between X=mean daily temperature for the month and Y=monthly charges on electrical bill, the following data was gathered:

X	20	30	50	60	80	90
Y	125	110	95	90	110	130

Which of the following seems the most likely model?
a) $Y=\alpha+\beta x+\varepsilon$ $\beta<0$
b) $Y=\alpha+\beta x+\varepsilon$ $\beta>0$
c) $Y=\alpha+\beta_1 x+\beta_2 x^2+\varepsilon$ $\beta_2<0$
d) $Y=\alpha+\beta_1 x+\beta_2 x^2+\varepsilon$ $\beta_2>0$

8. If a predictor variable x is found to be highly significant we would conclude that:
a) a change in y causes a change in x b) a change in x causes a change in y
c) changes in x are not related to changes in y d) changes in x are associated to changes in y

9. At the same confidence level, a prediction interval for a new response is always;
a) somewhat larger than the corresponding confidence interval for the mean response
b) somewhat smaller than the corresponding confidence interval for the mean response
c) one unit larger than the corresponding confidence interval for the mean response
d) one unit smaller than the corresponding confidence interval for the mean response

10. Both the prediction interval for a new response and the confidence interval for the mean response are narrower when made for values of x that are:
a) closer to the mean of the x's b) further from the mean of the x's
c) closer to the mean of the y's d) further from the mean of the y's

11. In the regression model $Y = \alpha + \beta x + \varepsilon$ the change in Y for a one unit increase in x:
a) will always be the same amount, α b) will always be the same amount, β
c) will depend on the error term d) will depend on the level of x

12. In a regression model with a dummy variable **without** interaction there can be:

a) more than one slope and more than one intercept
b) more than one slope, but only one intercept
c) only one slope, but more than one intercept
d) only one slope and one intercept

13. In a multiple regression model, where the x's are predictors and y is the response, multicollinearity occurs when:

a) the x's provide redundant information about y
b) the x's provide complementary information about y
c) the x's are used to construct multiple lines, all of which are good predictors of y
d) the x's are used to construct multiple lines, all of which are bad predictors of y

14. Compute the simple linear regression equation if:

	mean	stdev	correlation
x	163.5	16.2	-0.774
y	874.1	54.2	

15. Match the statements below with the corresponding terms from the list.

a) multicollinearity
b) extrapolation
c) R^2 adjusted
d) quadratic regression
e) interaction
f) residual plots
g) fitted equation
h) dummy variables
i) cause and effect
j) multiple regression model
k) R^2
l) residual
m) influential points
n) outliers

____ Used when a numerical predictor has a curvilinear relationship with the response.

____ Worst kind of outlier, can totally reverse the direction of association between x and y.

____ Used to check the assumptions of the regression model.

____ Used when trying to decide between two models with different numbers of predictors.

____ Used when the effect of a predictor on the response depends on other predictors.

____ Proportion of the variability in y explained by the regression model.

____ Is the observed value of y minus the predicted value of y for the observed x..

____ A point that lies far away from the rest.

____ Can give bad predictions if the conditions do not hold outside the observed range of x's.

____ Can be erroneously assumed in an observational study.

____ $y= \alpha +\beta_1x_1+\beta_2x_2+...+\beta_px_p+\varepsilon \quad \varepsilon \sim N(0,\sigma^2)$

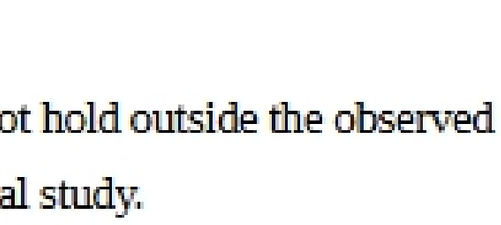
____ $\hat{y} =a+b_1x_1+b_2x_2+...+b_px_p$

____ Problem that can occur when the information provided by several predictors overlaps.

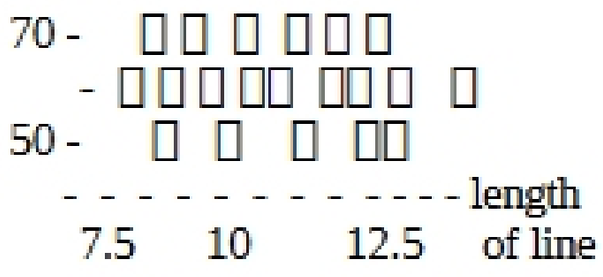
____ Used in a regression model to represent categorical variables.

Questions 16 - 19 Palm readers claim to be able to tell how long your life will be by looking at a specific line on your hand. The following is a plot of age of person at death (in years) vs length of life line on the right hand (in cm) for a sample of 28 (dead) people.

16. If we fit a simple linear regression model to these data, what would the value of r be?



a) close to -1
b) close to 0
c) close to 1
d) it's impossible to tell



17. Would you say:
a) length of life line is a very good predictor of age of person at death
b) length of life line is a poor predictor of age of person at death
c) length of life line is a reasonably good predictor of age of person at death
d) cannot determine how good a predictor length of life line is of age of person at death

18. The ANOVA p-value will be around
a) 1.00 b) 0.000 c) 0.05 d) 0.01

19. A better way of modeling age of person at death using this data set would be to use:
a) a nonparametric procedure
b) the average age at death
c) a contingency table
d) quadratic regression

20. According to the null hypothesis of the ANOVA F test, which predictor variables are providing significant information about the response?
a) most of them b) none of them c) all of them d) some of them

21. According to the alternative hypothesis of the ANOVA F test, which predictor variables are providing significant information about the response?
a) most of them b) none of them c) all of them d) some of them

22. In general, the Least Squares Regression approach finds the equation:
a) that includes the best set of predictor variables
b) of the best fitting straight line through a set of points
c) with the highest R^2, after comparing all possible models
d) that has the smallest sum of squared errors

23. Studies have shown a high positive correlation between the number of firefighters dispatched to combat a fire and the financial damages resulting from it. A politician commented that the fire chief should stop sending so many firefighters since they are clearly destroying the place. This is an example of:
a) extrapolation
b) dummy variables
c) misuse of causality
d) multicollinearity

24. The following appeared in the magazine *Financial Times*, March 23, 1995: "When Elvis Presley died in 1977, there were 48 professional Elvis impersonators. Today there are an estimated 7328. If that growth is projected, by the year 2012 one person in four on the face of the globe will be an Elvis impersonator." This is an example of:
a) extrapolation
b) dummy variables
c) misuse of causality
d) multicollinearity

Questions 25 – 43 Most supermarkets use scanners at the checkout counters. The data collected this way can be used to evaluate the effect of price and store's promotional activities on the sales of any product. The promotions at a store change weekly, and are mainly of two types: flyers distributed outside the store and through newspapers (which may or may not include that particular product), and in-store displays at the end of an aisle that call the customers' attention to the product. Weekly data was collected on a particular beverage brand, including sales (in number of units), price (in dollars), flyer (1 if product appeared that week, 0 if it didn't) and display (1 if a special display of the product was used that week, 0 if it wasn't).

As a preliminary analysis, a simple linear regression model was done.