STA 3024 **Practice Problems for Exam 2** **Fall 2007**

Please Note: Answering all of these questions is not necessarily sufficient for a decent grade. You must first study, using your book as well as your notes. These questions are not intended to cover everything, but aim to give you an idea about the questions you may have in the test. You are responsible from everything in the text as well as the notes. Solve as many problems as you can from your text AFTER CAREFULLY READING IT.

Read questions very carefully. For each question, choose the best answer then explain (to yourself) in a sentence why you chose that as the correct answer and why the other options are wrong. Solutions will be posted soon.

1. The parameters to be estimated in the simple linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ with the assumption $\varepsilon \sim N(0, \sigma)$ are:
   a) $\beta_0, \beta_1, \sigma^2$ b) $\beta_0, \beta_1, \varepsilon$ c) b_0, b_1, s^2 d) $\varepsilon, 0, \sigma^2$

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 is the decision for testing H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$?
   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 we decide that β_1 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
   e) Y is a good predictor of X

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 were 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) $Y = \beta_0 + \beta_1 x + \varepsilon$ with $\beta_1 < 0$
b) $Y = \beta_0 + \beta_1 x + \varepsilon$ with $\beta_1 > 0$
c) $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ with $\beta_2 < 0$
d) $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ with $\beta_2 < 0$

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 $Y = \beta_0 + \beta_1 x + \varepsilon$ the change in Y for a one unit increase in x:
   a) Will always be the same amount, β_0 b) Will always be the same amount, β_1
   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. Write the prediction equation given the following results from a sample of 100 (NO FORMULAS GIVEN)

	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

i. _____ Used when a numerical predictor has a curvilinear relationship with the response
ii. _____ Worst kind of outlier, can totally reverse the direction of association between x and y.
iii. _____ Used to check the assumptions of the regression model.
iv. _____ Used when trying to decide between two models with different numbers of predictors.
v. _____ Used when the effect of a predictor on the response depends on other predictors.
vi. _____ Proportion of the variability in y explained by the regression model.
vii. _____ Is the observed value of y minus the predicted value of y for the observed x..
viii. _____ a point that lies far away from the rest.
ix. _____ can give bad predictions if the conditions do not hold outside the observed range of x's.
x. _____ can be erroneously assumed in an observational study.
xi. _____ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$
xii. _____ $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$
xiii. _____ Problem that can occur when the information provided by several predictors overlaps.
xiv. _____ Used in a regression model to represent categorical variables.