Linear Regression **Name**______________________________________

Chapter 3 Test

Use a pencil. Show your work. Be neat.

1. The simple-linear, normal-error model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ for $i = 1, ..., n$ assumes

   (i) that the ϵ_i are independent and normally distributed with mean 0 and constant variance σ^2;
   True False

   (ii) that the Y_i are independent normal random variable with mean 0 and constant variance σ^2;
   True False

   (iii) that $E(Y_i)$ is linearly related to X_i;
   True False

   (iv) the sum of the error terms ϵ_i equals 0;
   True False

   (v) the errors terms ϵ_i are positively correlated with the X_i;
   True False

2. The modified Levene test compares the mean absolute deviation of residuals for low X values, $\bar{d}_1$, to the mean absolute deviation of residuals for high X values, $\bar{d}_2$. The test statistic has an approximate t distribution when H_o is true.

   (i) State the null and alternative hypotheses for the modified Levene test.

   H_o:

   H_a:

   (ii) If the statistical distance between $\bar{d}_1$ and $\bar{d}_2$ is relatively small, or if the p-value is greater than α, we conclude that

3. The Breusch-Pagan (or Cook-Weisberg) test for constancy of error variance involves regressing the squared residuals e_i^2 on the predictor variable values X_i. If resulting regression sum of squares, SSR^*, is relatively large, we conclude that

4. With regards to the simple-linear, normal-error regression model ...

(i) the formal F test for lack of fit has hypotheses: (complete...)

H_o: $E(Y)$ $\beta_0 + \beta_1 X$

H_a: $E(Y)$ $\beta_0 + \beta_1 X$

(ii) The SSE is partitioned into SSPE and SSLF. If all X levels have only single observations, then SSLF is equal to

(a) 1 (b) 0 (c) SSE (d) SSR/SSTO (e) SSPE

(iii) Suppose there are c levels for X. The F test ratio for a lack of fit test is given by

$$F^* = \frac{\text{MSLF}}{\text{MSPE}} =$$

(a) $\frac{\sum\sum(\bar{Y}_j - \hat{Y}_{ij})/(c-2)}{\sum\sum(Y_{ij} - Y_j)/(n-c)}$ (b) $\frac{\sum\sum(Y_{ij} - \hat{Y}_{ij})^2/(c-2)}{\sum\sum(Y_{ij} - Y_j)^2/(n-c)}$ (c) $\frac{\sum\sum(\bar{Y}_j - \hat{Y}_{ij})^2/(c-2)}{\sum\sum(Y_{ij} - Y_j)^2/(n-c)}$

(iv) If for every level j of X the deviations $|\bar{Y}_j - \hat{Y}_{ij}|$ are relatively small, then we have evidence pointing in the direction of

(a) lack of fit (b) no significant lack of fit (c) a quadratic linear relationship

(v) If the SSPE makes up the bulk of the SSE, then we have evidence pointing in the direction of

(a) lack of fit (b) no significant lack of fit (c) a quadratic linear relationship

(vi) The general linear test formulation provides the test statistic $F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{SSE(F)/df_F}$.

For the lack of fit test, SSE(R) $=$

(a) SSR (b) SSTO (c) SSE (d) SSPE (e) SSLF

5. Let $e_{(k)}$ represent the kth smallest residual. Assuming normality or error terms, an estimate of the expected value of $e_{(k)}$ is $\sqrt{\text{MSE}} \cdot z_{\text{c}}$ where z_{c} is a quantile from the standard normal distribution that depends on k and the sample size n.

(i) If $n = 20$ and $k = 5$, use a standard normal table to find z_{c}. $z_{\text{c}} =$

(ii) Write MINITAB command that produces the z_{c} (the normal scores) for a column of data. Suppose the residual data is in C5 and you want to store the normal scores in c15.

MTB>

(iii) A normal probability plot of residuals is scatter plot of

the__________________________ versus _____________________________________

(iv) Lack of linearity in the residual normal probability plot is indicative of

____

.

(v) Suppose the correlation coefficient, r, of the ordered residuals and their estimated expected values under H_o turns out to be $r = .9317$. If the sample size is 20, choose the best conclusion (use $\alpha = .05$):

(a) Since r $\leq$ the critical value from Table B6, we accept the null hypothesis of normal errors.
(b)Since r $\leq$ the critical value from Table B6, we reject the null hypothesis of normal errors.
(c)Since r $>$ the critical value from Table B6, we accept the null hypothesis of normal errors.
(d) Since r $>$ the critical value from Table B6, we reject the null hypothesis of normal errors.

(vi) A random sample from a normal distribution will always have a symmetric looking boxplot and a bell-shaped histogram. True False

(vii) Since $\sum e_i = 0$, the residuals can never be completely independent. True False

(viii) A boxplot of the residuals provides more information about possible normality of error terms than the correlation test for normality. True False

6. For the following Y vs. X scatterplot patterns, provide a simple transformation of X that may linearize the regression relationship.

(i)

```
    20.0+                                                     *  *
        -                                                *  *
Y       -                                             *
        -                                       *  *
        -                                 *  *
    15.0+                              *
        -                         *  *
        -                      *
        -                   *
        -                *
    10.0+             *
        -          *
        -        *
        -     *
        -
     5.0+  *
        -
          --------+---------+---------+---------+---------+--------X

                 35        70       105       140       175
```

7. The Box-Cox procedure for determining a transformation of Y that may help linearize a curvilinear relationship and may remedy unequal error variances and nonnormality of the error terms is based on finding a transformation of the form $\mathrm{Y}' = \mathrm{Y}^{\lambda}$ that minimizes SSE.

(i) What does a minimum SSE suggest?

(ii) Below is the output from the Minitab macro boxcox.MTB. What transformation is suggested?

ROW	LAMBDAS	SSES
1	-3.0	59533950976
2	-2.0	909786496
3	-1.0	19201312
4	-0.5	3076121
5	0.5	3432
6	1.0	439162
7	2.0	8343774
8	3.0	96799376
9	0.0	374251

8. Below is a plot of residuals versus an omitted variable.

```
     300+                                          *
        - *
residual-
        -
        - *                                   *
     150+
        -  *                              *
        -
        -
        -   *                          *
       0+                           *
        -    *
        -     *                  *
        -
        -      **             *
    -150+         *   * *  *
        -           *
          +---------+---------+---------+---------+----X2

          0       100       200       300       400
```

The plot suggests that the variable X_2

(a) should be included in the model since there is a distinct pattern in the plot.

(b) should not be included in the model since the residual plot does not indicate random scattering in a more-or-less horizontal band about 0.

9. Below is Minitab output from the macro levene.MTB that performs a modified Levene test.

Below are absolute value of Levene test statistic and p-value.

MTB > print k93 k97

|levt*| 0.0453620

LTpvalue 0.964316

If the p-value above is less than .05, we may conclude that the error variances are nonconstant.

Write your conclusion in plain English.