STA 6207 – Exam 1 – Fall 2013

Name

Conduct all tests at the \( \alpha = 0.05 \) Significance level.

All Models on test are of the forms:

Scalar: \( Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon \)  \( \varepsilon \sim N(0, \sigma^2) \) independent  \n\nMatrix: \( Y = X\beta + \varepsilon \)  \( \varepsilon \sim N(0, \sigma^2I) \)

Critical Values for \( t \), \( \chi^2 \), and F Distributions

<table>
<thead>
<tr>
<th>( \chi^2 )</th>
<th>( F_{0.01} )</th>
<th>( F_{0.01} )</th>
<th>( F_{0.025} )</th>
<th>( F_{0.05} )</th>
<th>( F_{0.01} )</th>
<th>( F_{0.01} )</th>
<th>( F_{0.025} )</th>
<th>( F_{0.05} )</th>
<th>( F_{0.01} )</th>
<th>( F_{0.01} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.015</td>
<td>2.571</td>
<td>11.070</td>
<td>6.608</td>
<td>5.786</td>
<td>5.409</td>
<td>5.192</td>
<td>5.050</td>
<td>4.950</td>
<td>4.876</td>
</tr>
<tr>
<td>13</td>
<td>1.689</td>
<td>2.160</td>
<td>22.435</td>
<td>4.055</td>
<td>3.467</td>
<td>3.134</td>
<td>2.910</td>
<td>2.769</td>
<td>2.687</td>
<td>2.605</td>
</tr>
<tr>
<td>16</td>
<td>1.621</td>
<td>2.120</td>
<td>26.769</td>
<td>3.681</td>
<td>3.276</td>
<td>2.971</td>
<td>2.763</td>
<td>2.632</td>
<td>2.557</td>
<td>2.478</td>
</tr>
<tr>
<td>17</td>
<td>1.602</td>
<td>2.110</td>
<td>28.274</td>
<td>3.582</td>
<td>3.201</td>
<td>2.905</td>
<td>2.699</td>
<td>2.571</td>
<td>2.497</td>
<td>2.419</td>
</tr>
<tr>
<td>18</td>
<td>1.586</td>
<td>2.101</td>
<td>29.814</td>
<td>3.495</td>
<td>3.078</td>
<td>2.788</td>
<td>2.585</td>
<td>2.460</td>
<td>2.388</td>
<td>2.312</td>
</tr>
<tr>
<td>19</td>
<td>1.572</td>
<td>2.093</td>
<td>31.384</td>
<td>3.418</td>
<td>2.982</td>
<td>2.694</td>
<td>2.494</td>
<td>2.372</td>
<td>2.302</td>
<td>2.229</td>
</tr>
<tr>
<td>20</td>
<td>1.560</td>
<td>2.086</td>
<td>32.991</td>
<td>3.347</td>
<td>2.693</td>
<td>2.407</td>
<td>2.209</td>
<td>2.089</td>
<td>2.022</td>
<td>1.953</td>
</tr>
<tr>
<td>21</td>
<td>1.549</td>
<td>2.080</td>
<td>34.630</td>
<td>3.281</td>
<td>2.624</td>
<td>2.341</td>
<td>2.145</td>
<td>2.028</td>
<td>2.059</td>
<td>1.993</td>
</tr>
<tr>
<td>22</td>
<td>1.539</td>
<td>2.074</td>
<td>36.300</td>
<td>3.223</td>
<td>2.566</td>
<td>2.287</td>
<td>2.091</td>
<td>1.977</td>
<td>2.005</td>
<td>1.939</td>
</tr>
<tr>
<td>23</td>
<td>1.530</td>
<td>2.069</td>
<td>38.007</td>
<td>3.165</td>
<td>2.515</td>
<td>2.241</td>
<td>2.046</td>
<td>1.934</td>
<td>1.964</td>
<td>1.899</td>
</tr>
</tbody>
</table>

F Distributions Indexed by Numerator Degrees of Freedom

CDF - Lower tail probabilities

Total = 29 + 23 + 37 + 36 + 8 + 36 + 27 = 229 229
Q.1. Regression models were fit, relating height \(Y\) (in mm) to hand length \(X_1\) (in mm), foot length \(X_2\) (in mm) and gender \((X_3=1 \text{ if male}, 0 \text{ if female})\) based on a sample of 80 males and 75 females. Consider these 4 models:

Model 1: \(E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{23} X_2 X_3\)

Model 2: \(E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3\)

Model 3 (Males Only): \(E\{Y\} = \delta_0 + \delta_1 X_1 + \delta_2 X_2\)

Model 4 (Females Only): \(E\{Y\} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2\)

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>Model1</th>
<th></th>
<th>ANOVA</th>
<th>Model2</th>
<th></th>
<th>ANOVA</th>
<th>Model3</th>
<th></th>
<th>ANOVA</th>
<th>Model4</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>SS</td>
<td></td>
<td>df</td>
<td>SS</td>
<td></td>
<td>df</td>
<td>SS</td>
<td></td>
<td>df</td>
<td>SS</td>
</tr>
<tr>
<td>Regression</td>
<td>5</td>
<td>1201091</td>
<td></td>
<td>Regression</td>
<td>3</td>
<td>1193101</td>
<td></td>
<td>Regression</td>
<td>7</td>
<td>208298</td>
</tr>
<tr>
<td>Residual</td>
<td>149</td>
<td>157138</td>
<td></td>
<td>Residual</td>
<td>151</td>
<td>165128</td>
<td></td>
<td>Residual</td>
<td>77</td>
<td>88305</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>1358229</td>
<td></td>
<td>Total</td>
<td>154</td>
<td>1358229</td>
<td></td>
<td>Total</td>
<td>79</td>
<td>296603</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>Coefficients</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>744.14 63.68</td>
<td>Intercept</td>
<td>582.16 60.55</td>
<td>Intercept</td>
<td>439.42 97.49</td>
<td>Intercept</td>
<td>744.14 75.57</td>
</tr>
<tr>
<td>Hand</td>
<td>2.38 0.51</td>
<td>Hand</td>
<td>2.81 0.34</td>
<td>Hand</td>
<td>3.29 0.47</td>
<td>Hand</td>
<td>2.38 0.49</td>
</tr>
<tr>
<td>Foot</td>
<td>1.73 0.39</td>
<td>Foot</td>
<td>2.06 0.26</td>
<td>Foot</td>
<td>2.38 0.35</td>
<td>Foot</td>
<td>1.73 0.37</td>
</tr>
<tr>
<td>Male</td>
<td>-304.72 125.47</td>
<td>Male</td>
<td>39.61 8.90</td>
<td>Male</td>
<td></td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>MaleHand</td>
<td>0.91 0.68</td>
<td>MaleHand</td>
<td></td>
<td></td>
<td></td>
<td>MaleHand</td>
<td></td>
</tr>
<tr>
<td>MaleFoot</td>
<td>0.65 0.52</td>
<td>MaleFoot</td>
<td></td>
<td></td>
<td></td>
<td>MaleFoot</td>
<td></td>
</tr>
</tbody>
</table>

p.1.a. Confirm the equivalence of the regression coefficients (but not standard errors) based on the appropriate models (Hint: set up the fitted equations based on the two models):

Females:

\[
\hat{Y}_{\text{Model 1}} = 744.14 + 2.38 H + 1.73 F + 0
\]

\[
\hat{Y}_{\text{Model 4}} = 744.14 + 2.38 H + 1.73 F
\]

Males:

\[
\hat{Y}_{\text{Model 1}} = (744.14 - 304.72) + (2.38 + 0.91) H + (1.73 + 0.65) F
\]

\[
= 439.42 + 3.29 H + 2.38 F = \hat{Y}_{\text{Model 3}}
\]

p.1.b. Test \(H_0: \beta_{13} = \beta_{23} = 0\) (No interactions between Hand and Gender or Foot and Gender).

\[
F_{\text{obs}} = \frac{165148 - 157138}{151 - 149} = \frac{157138}{149} = \frac{3995}{1054.62} = 3.788
\]

\[
F(95; 2, 149) = 3.056
\]

Statistic: \(F_{\text{obs}} = 3.788\)

Rejection Region: \(F_{\text{obs}} \geq 3.056\)

p-value > 0.05
p.1.c. Use Hartley's Test to test whether the error variances among the individual regressions are equal:

\[ B = \frac{1}{C} \left[ \nu \ln(MSE) - \sum_{i=1}^{t} \nu_i \ln(s_i^2) \right] \]
\[ C = 1 + \frac{1}{3(t-1)} \left[ \sum_{i=1}^{t} \nu_i^{-1} - \nu^{-1} \right] \]

\[ N = 149 \quad MSE = 1054.62 \]
\[ N_1 = 77 \quad s_1^2 = \frac{883.05}{77} = 11.6.82 \]
\[ N_2 = 72 \quad s_2^2 = \frac{6883.33}{72} = 95.6.01 \]

\[ C = 1 + \frac{1}{3(2-1)} \left[ \left( \frac{1}{77} + \frac{1}{72} \right) - \frac{1}{149} \right] = 1 + 0.0067 = 1.0067 \]

\[ B = \frac{1}{1.0067} \left[ 149 \ln(1054.62) - \left( 77 \ln(11.682) + 72 \ln(95.601) \right) \right] \]

\[ = \frac{1}{1.0067} \left[ 1037.18 - (542.45 + 597.43) \right] = 0.606 \]

Test Statistic \( B = 0.601 \)

Rejection Region: \( B \geq \chi^2(0.05, 1) = 3.841 \)

p.1.d. What fraction of the total variation in height is explained by the set of predictors: hand length, foot length, and gender (but no interactions)?

\[ R^2 = \frac{1193101}{1358229} = 0.8784 \]

p.1.e. Compute the standard deviations among the 80 Male heights and among the 75 Female heights (ignoring hand and foot length).

\[ s_m = \sqrt{\frac{246603}{79}} = 61.27 \]
\[ s_f = \sqrt{\frac{179385}{74}} = 49.27 \]

Males: SD = 61.27 \quad Females: SD = 49.27
Q.2. A study related Personal Best Shot Put distance (Y, in meters) to best preseason power clean lift (X, in kilograms). The following models were fit, based on a sample of n = 24 male collegiate shot putters:

Model 1:  \( E\{Y\} = \beta_0 + \beta_1 X \)  \( \text{SSE}_1 = 43.41 \)  \( R^2 = .686 \)  \( \hat{Y}(X) = 4.4353 + 0.0898X \)

Model 2:  \( E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X^2 \)  \( \text{SSE}_2 = 37.41 \)  \( R^2 = .729 \)  \( \hat{Y}(X, X^2) = 12.08 + 0.3285X - 0.00084X^2 \)

p.2.a. Use Model 2 to test \( H_0: \beta_1 = \beta_2 = 0 \) (Y is not related to X)

\[ 12 \]

\[ .729 = \theta - \frac{37.41}{SSR} = \Rightarrow SSR = \frac{37.41}{.271} = 138.04 \]

\[ = \Rightarrow SS_{R^2} = 138.04 - 37.41 = 100.63 \]

\[ F_{obs} = \frac{100.63/2}{37.41/21} = \frac{50.315}{1.781} = 28.27 \]

Test Statistic: \( F_{obs} = 28.27 \)  Rejection Region: \( F_{obs} > F(.05, 2, 21) = 3.467 \)  Reject \( H_0 \)?  Yes or No

p.2.b. Use Models 1 and 2 to test \( H_0: \beta_2 = 0 \) (Y is linearly related to X)

\[ 12 \]

\[ F_{obs} = \frac{(43.41-37.41)/1}{1.781} = \frac{6.00}{1.781} = 3.369 \]

\( F(.05, 1, 21) = 4.325 \)

Test Statistic: \( F_{obs} = 3.369 \)  Rejection Region: \( F_{obs} > 4.325 \)  Reject \( H_0 \)?  Yes or No

p.2.c. Give an estimate of the level of X is that maximizes \( E\{Y\} \).

\[ 12 \]

\[ \frac{d\hat{Y}}{dx} = 0.3285 - .00168x = 0 \]

\[ \Rightarrow x^* = \frac{-3285}{.00168} = 195.54 \]

\[ x^* = 195.54 \]

\[ 37 \]
Q.3. A study was conducted to determine whether having been exposed to an advertisement claiming a natural ingredient is contained in a perfume had an effect on subjects’ rating of the perfume’s scent. There were 112 subjects of which, 56 were exposed to the ad, and 56 were not. We fit the following regression model:

\[ Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, \ldots, 112 \]

\[ X_i = \begin{cases} 
1 & \text{if Subject } i \text{ was exposed to the ad} \\
0 & \text{if Subject } i \text{ was not exposed to the ad} 
\end{cases} \]

<table>
<thead>
<tr>
<th>X'X</th>
<th>X'Y</th>
<th>Y'Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>56</td>
<td>3683.05</td>
</tr>
</tbody>
</table>

p.3.a. First, we fit a model with only an intercept term, what will \( P_0 = X_0 (X_0' X_0)^{-1} X_0' \) be (symbolically, do not write out a 112x112 matrix!)? Compute \( R(\beta_0) \).

\[ X_0 = \begin{bmatrix} 1 \\
\vdots \\
1 \\end{bmatrix} \quad X_0' X_0 = 112 \quad X_0 (X_0' X_0)^{-1} X_0' = \frac{1}{112} J_{112} \]

\[ Y'P_0 Y = \frac{1}{112} (2 \cdot 36) = \frac{(587)^2}{112} = 3076.51 \]

\[ P_0 = \frac{1}{112} J_{112} \quad R(\beta_0) = 3076.51 \]

p.3.b. Compute \((X'X)^{-1}\) and \(\hat{\beta} \) NOTE: Write \((X'X)^{-1}\) as \( \frac{1}{|X'X|} A \) for the appropriate \( A \)

\[ |X'X| = 112(56) - 56^2 = 3136 \quad (X'X)^{-1} = \frac{1}{3136} \begin{bmatrix} 56 & -56 \\ -56 & 112 \end{bmatrix} \]

\[ \hat{\beta} = (X'X)^{-1} X' Y = \frac{1}{3136} \begin{bmatrix} 56(587) - 56(337) \\ -56(587) + 112(337) \end{bmatrix} = \frac{1}{3136} \begin{bmatrix} 17000 \\ 4872 \end{bmatrix} = \begin{bmatrix} 4.464 \\ 1.554 \end{bmatrix} \]

\[ 36 \]
p.3.c. Compute $R(\beta_0, \beta_1)$, $R(\beta_1 | \beta_0)$, and MSResidual

$$R(\beta_0, \beta_1) = Y'X(X'X)^{-1}X'Y = \hat{\beta}'X'Y$$

$$= \frac{1}{3136} \begin{bmatrix} 14000 & 4872 \end{bmatrix} \begin{bmatrix} 587 \\ 337 \end{bmatrix} = \frac{985964}{3136} = 3144.09$$

$$R(\beta_1 | \beta_0) = 3144.09 - 3076.51 = 67.58$$

$$SSRes = Y'Y - \hat{\beta}'X'Y = Y'(I - \hat{\beta})Y = 3683.05 - 3144.09 = 538.96$$

$$MSRes = \frac{538.96}{112-2} = 4.90$$

$$R(\beta_0, \beta_1) = 3144.09 \quad R(\beta_1 | \beta_0) = 67.58 \quad MSResidual = 4.90$$

p.3.d. Use the t-test and the F-test to test $H_0: \beta_1 = 0 \quad vs \quad H_A: \beta_1 \neq 0$

$$t-test: \quad t_{0.05} = \frac{1.554}{\sqrt{4.90 (12/3136)}} = \frac{1.554}{0.418} = 3.715$$

$$F-test: \quad F_{0.05} = \frac{67.58/1}{4.90} = 13.992$$

$$t\text{-Statistic: } 3.715 \quad \text{Rejection Region: } |t_{0.05}| > t(.05,110) = 1.982$$

$$F\text{-Statistic: } 13.992 \quad \text{Rejection Region: } F_{0.05} > F(.05,1,110) = 3.927$$
Q.4. An experiment was conducted to measure the subsoil pressure of a steel ground roller. **There were 3 replicates at each of 4 depths (X=5, 10, 15, 20 cm)**. The response was measured force (100s of Newtons).

The fitted regression equation is \( \hat{Y} = 49.371 - 2.036X \)

We want to test \( H_0: E(Y_j) = \beta_0 + \beta_1X_j \) \( H_A: E(Y_j) = \mu \neq \beta_0 + \beta_1X_j \)

<table>
<thead>
<tr>
<th>j</th>
<th>X_j</th>
<th>Ybar_j</th>
<th>SD_j</th>
<th>Yhat_j</th>
<th>Pure Error</th>
<th>Lack of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>40.38</td>
<td>4.32</td>
<td>39.191</td>
<td>37.325</td>
<td>4.241</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>28.87</td>
<td>6.83</td>
<td>29.011</td>
<td>93.296</td>
<td>0.060</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>16.23</td>
<td>3.76</td>
<td>18.831</td>
<td>28.275</td>
<td>20.294</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>11.15</td>
<td>4.48</td>
<td>8.651</td>
<td>40.141</td>
<td>18.735</td>
</tr>
</tbody>
</table>

p.4.a. Compute the Pure Error Sum of Squares and Degrees of Freedom. Hint: What is \( SD_j \) equal to?

\[
SSPE = 199.039 \\
df_{PE} = n - c = 12 - 4 = 8
\]

\[
SSPE = \frac{199.039}{10} \\
df_{PE} = 6
\]

p.4.b. Compute the Lack-of-Fit Sum of Squares and Degrees of Freedom.

\[
SSLF = 43.332 \\
df_{LF} = c - 2 = 4 - 2 = 2
\]

\[
SSLF = \frac{43.332}{10} \\
df_{LF} = 2
\]

p.4.c. Conduct the F-test for Lack-of-Fit

\[
F_{LS} = \frac{43.332/2}{199.039/8} = \frac{21.666}{43.880} = 0.871
\]

\[
F(0.05; 2, 8) = 4.459
\]

**Test Statistic:** \( F_{LF} = 0.871 \)  
**Rejection Region:** \( \geq 4.459 \)

\( \text{Reject } H_0? \) Yes \( \bigcirc \) No \( \bigcirc \)
Q.5. A series of models were fit, relating Average January High Temperature ($Y$, in degrees F) to Elevation ($X_1$, in 100s ft above sea level), and Latitude ($X_2$ degrees North Lat) for $n = 369$ weather stations in Texas. Latitude and Elevation were centered in the regression models.

$$C_p = \frac{SS(Res)_{Model}}{MS(Res)_{Complete}} + 2p' - n$$

$$AIC = n \ln(SS(Res)_{Model}) + 2p' - n \ln(n)$$

$$SBC = BIC = n \ln(SS(Res)_{Model}) + \ln(n) p' - n \ln(n)$$

<table>
<thead>
<tr>
<th>Variables in Model</th>
<th>SS(RES)</th>
<th>$C_p$</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELEV (E)</td>
<td>7986.3</td>
<td>4764.2</td>
<td>1138.6</td>
<td>1146.4</td>
</tr>
<tr>
<td>LAT (L)</td>
<td>1168.0</td>
<td>385.1</td>
<td>429.2</td>
<td>437.0</td>
</tr>
<tr>
<td>E,L</td>
<td>616.2</td>
<td>32.8</td>
<td>195.2</td>
<td>207.0</td>
</tr>
<tr>
<td>E,L,E*L</td>
<td>603.9</td>
<td>26.9</td>
<td>189.8</td>
<td>205.4</td>
</tr>
<tr>
<td>E,L,E^2,L^2</td>
<td>575.0</td>
<td>10.3</td>
<td>173.7</td>
<td>193.2</td>
</tr>
<tr>
<td>E,L,E*L,E^2,L^2</td>
<td>565.2</td>
<td>6.0</td>
<td>169.3</td>
<td>192.8</td>
</tr>
</tbody>
</table>

p.5.a. Complete the table.

\[ C_p (LAT) = \frac{1168}{565.2} + 2(2) - 369 = 750.15 + 4 - 369 = 385.1 \]

\[ AIC (E,L,E*L) = 369 \ln(603.9) + 2(4) - 369 \ln(369) = 189.8 \]

\[ SBC (E,L,E^2,L^2) = 369 \ln(575.0) + \ln(369)(5) - 369 \ln(369) = 193.2 \]

p.5.b. Based on each criteria, which model do you choose?

$p.c.$: $\overline{E,L,E^2,L^2}$

AIC: Same

SBC: Same