Total = 265

STA 6207 – Exam 2 – Fall 2017 **PRINT** Name Answer Key

Conduct all tests at the $\alpha = 0.05$ Significance level.
Be specific on the ranges of the rejection regions.

All Models on test are of the forms:

Scalar: $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon \quad \varepsilon \sim N\left(0, \sigma^2\right)$ independent Matrix: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$

Critical Values for t, χ^2, and F Distributions
F Distributions Indexed by Numerator Degrees of Freedom
CDF - Lower tail probabilities

df	$t_{.95}$	$t_{.975}$	$\chi^2_{.95}$	$F_{.95,1}$	$F_{.95,2}$	$F_{.95,3}$	$F_{.95,4}$	$F_{.95,5}$	$F_{.95,6}$	$F_{.95,7}$	$F_{.95,8}$
1	6.314	12.706	3.841	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883
2	2.920	4.303	5.991	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371
3	2.353	3.182	7.815	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845
4	2.132	2.776	9.488	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041
5	2.015	2.571	11.070	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818
6	1.943	2.447	12.592	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147
7	1.895	2.365	14.067	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726
8	1.860	2.306	15.507	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438
9	1.833	2.262	16.919	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230
10	1.812	2.228	18.307	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072
11	1.796	2.201	19.675	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948
12	1.782	2.179	21.026	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849
13	1.771	2.160	22.362	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767
14	1.761	2.145	23.685	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699
15	1.753	2.131	24.996	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641
16	1.746	2.120	26.296	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591
17	1.740	2.110	27.587	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548
18	1.734	2.101	28.869	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510
19	1.729	2.093	30.144	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477
20	1.725	2.086	31.410	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447
21	1.721	2.080	32.671	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420
22	1.717	2.074	33.924	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397
23	1.714	2.069	35.172	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375
24	1.711	2.064	36.415	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355
25	1.708	2.060	37.652	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337
26	1.706	2.056	38.885	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321
27	1.703	2.052	40.113	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305
28	1.701	2.048	41.337	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291
29	1.699	2.045	42.557	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278
30	1.697	2.042	43.773	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266
40	1.684	2.021	55.758	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180
50	1.676	2.009	67.505	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130
60	1.671	2.000	79.082	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097
70	1.667	1.994	90.531	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074
80	1.664	1.990	101.879	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056
90	1.662	1.987	113.145	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043
100	1.660	1.984	124.342	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032
110	1.659	1.982	135.480	3.927	3.079	2.687	2.454	2.297	2.182	2.094	2.024
120	1.658	1.980	146.567	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016
130	1.657	1.978	157.610	3.914	3.066	2.674	2.441	2.284	2.169	2.081	2.010
140	1.656	1.977	168.613	3.909	3.061	2.669	2.436	2.279	2.164	2.076	2.005
150	1.655	1.976	179.581	3.904	3.056	2.665	2.432	2.274	2.160	2.071	2.001
160	1.654	1.975	190.516	3.900	3.053	2.661	2.428	2.271	2.156	2.067	1.997
170	1.654	1.974	201.423	3.897	3.049	2.658	2.425	2.267	2.152	2.064	1.993
180	1.653	1.973	212.304	3.894	3.046	2.655	2.422	2.264	2.149	2.061	1.990
190	1.653	1.973	223.160	3.891	3.043	2.652	2.419	2.262	2.147	2.058	1.987
200	1.653	1.972	233.994	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985
∞	1.645	1.960	---	3.841	2.995	2.605	2.372	2.214	2.099	2.010	1.938

Q.1. A study measured total phenolic content (X_1) and melanoidin content (X_2) and Oxygen radical absorbence activity (Y) in n = 40 varieties of lager beer. The following linear regression models were fit. The Total (Corrected) SS = 231.31.

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1$ Model 2: $E\{Y\} = \beta_0 + \beta_2 X_2$ Model 3: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Model 4: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

The regression sums of squares for the models are: $SSR_1 = 29.64$, $SSR_2 = 0.98$, $SSR_3 = 46.54$, $SSR_4 = 46.58$.

p.1.a. Compute R^2 for each model.

M1: $\frac{29.64}{231.31}$ $\frac{.98}{231.31}$ $\frac{46.54}{231.31}$ $\frac{46.58}{231.31}$

3 each

Model 1: .1281 Model 2: .0042 Model 3: .2012 Model 4: .2014

p.1.b. Compute the following sums of squares for these models.

4 $R(\beta_1 | \beta_0) =$ 29.64

4 $R(\beta_2 | \beta_0) =$ 0.98

4 $R(\beta_1 | \beta_0, \beta_2) =$ 46.54 − 0.98 = 45.56

4 $R(\beta_2 | \beta_0, \beta_1) =$ 46.54 − 29.64 = 16.90

4 $R(\beta_3 | \beta_0, \beta_1, \beta_2) =$ 46.58 − 46.54 = 0.04

p.1.c. Test $H_0 : \beta_2 = \beta_3 = 0$ versus $H_A : \beta_2$ and/or $\beta_3 \neq 0$

15 $SSR_c = 46.58$ $SSR_R = 29.64$ $SSE_c = 231.31 - 46.58 = 184.73$

$df_c = 40 - 4 = 36$ $df_R = 40 - 2 = 38$

$$F_{obs} = \frac{\frac{\overbrace{46.58 - 29.64}^{4}}{\underbrace{38 - 36}_{2}}}{\frac{184.73}{36} - 4} = \frac{8.47}{5.1314} = 1.6506$$

Test Statistic: $F_{obs} = 1.6506$ Rejection Region: $F_{obs} \geq F_{.05,2,36} \approx 3.274$

3 8

50

Q.2. A linear regression model was fit, relating log of prize winnings per round (Y) for a sample of n = 24 Ladies' PGA golfers to the following predictors: proportion of drives on fairway (X_1), proportion of greens hit in regulation (X_2), and average putts per hole (X_3). The results of the matrix form are given below.

X'X					X'Y
24.00	17.03	15.28	39.96		178.79
17.03	12.16	10.84	28.34		126.99
15.28	10.84	9.76	25.43		114.65
39.96	28.34	25.43	66.55		297.24
INV(X'X)					Beta-hat
152.83	-19.77	-23.42	-74.40		16.07
-19.77	13.75	-1.49	6.59		-2.12
-23.42	-1.49	29.70	3.35		23.51
-74.40	6.59	3.35	40.60		-13.26
Y'IY	Y'J/nY	Y'PY			
1365.91	1331.93	1357.08			

p.2.a. Compute s^2 = MSE

7 $SSE = Y'(I-P)Y = 1365.91 - 1357.08 = 8.83$

$MSE = \frac{8.83}{24-4} = 0.4415$

p.2.b. Obtain a 95% Confidence Interval for β_3.

10 $\hat{se}\{\hat{\beta}_3\} = \sqrt{0.4415(40.60)} = \sqrt{17.9249} = 4.2338$

$t_{.025,20} = 2.086$

95% CI: $-13.26 \pm 2.086(4.2338) = -13.26 \pm 8.83 = (-22.09, -4.43)$

p.2.c. Conduct the test of $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

12 $SSR = Y'(P - \frac{1}{n}J)Y = 1357.08 - 1331.93 = 25.15$ $df_R = 3$

$MSR = \frac{25.15}{3} = 8.3833$ $F = \frac{8.3833}{0.4415} = 18.988$

Test Statistic: $F_{obs} = 18.988$ Rejection Region: $F_{obs} \geq F_{.05,3,20} = 3.098$

p.2.d. As both X_1 and X_2 represent proportions and skills, use the General Linear Test to test $H_0 : \beta_1 = \beta_2$. Show all calculations of the necessary elements of the test.

INV(X'X)					Beta-hat
152.83	-19.77	-23.42	-74.40		16.07
-19.77	13.75	-1.49	6.59		-2.12
-23.42	-1.49	29.70	3.35		23.51
-74.40	6.59	3.35	40.60		-13.26

$$K' = [0 \quad 1 \quad -1 \quad 0] \qquad \underset{\sim}{m} = 0$$

$$K'\hat{\beta} = -2.12 - 23.51 = -25.63$$

$$K'(X'X)^{-1}K = [3.65 \quad 15.29 \quad -31.19 \quad 3.24]\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = 15.29 + 31.19 = 46.48$$

$$[K'(X'X)^{-1}K]^{-1} = \left(\frac{1}{46.48}\right)$$

$$Q = (-25.63)^2\left(\frac{1}{46.48}\right) = 14.1329$$

$$F = \frac{14.1329}{1(.4415)} = 32.011$$

$$F_{.05,1,20} = 4.351$$

Test Statistic: $F_{obs} = 32.011$ Rejection Region: $F_{obs} \geq 4.351$

17 3

20

Q.3. A simple linear regression model is fit, relating berry chewiness (Y) to sugar content (X). There are 6 distinct levels of sugar content, with 15 berries measured at each level.

Let $\bar{Y}_j$ be the mean chewiness at the jth level X_j and $\hat{Y}_j$ be the fitted value from the regression.

p.3.a. Show that $\sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(Y_{ij}-\hat{Y}_j\right)^2=\sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(Y_{ij}-\bar{Y}_j\right)^2+\sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(\bar{Y}_j-\hat{Y}_j\right)^2$ **Show all work**

30

$$Y_{ij}-\hat{Y}_j = (Y_{ij}-\bar{Y}_j)+(\bar{Y}_j-\hat{Y}_j)$$

$$\Rightarrow (Y_{ij}-\hat{Y}_j)^2 = (Y_{ij}-\bar{Y}_j)^2+(\bar{Y}_j-\hat{Y}_j)^2+2(Y_{ij}-\bar{Y}_j)(\bar{Y}_j-\hat{Y}_j)$$

$$\Rightarrow \sum_j\sum_i (Y_{ij}-\hat{Y}_j)^2 = \sum_j\sum_i (Y_{ij}-\bar{Y}_j)^2+\sum_j\sum_i(\bar{Y}_j-\hat{Y}_j)^2$$

$$+2\sum_j(\bar{Y}_j-\hat{Y}_j)\cancel{\sum_i(Y_{ij}-\bar{Y}_j)} \quad 0$$

$$\Rightarrow \sum_j\sum_i (Y_{ij}-\hat{Y}_j)^2 = \sum_j\sum_i (Y_{ij}-\bar{Y}_j)^2+\sum_j\sum_i(\bar{Y}_j-\hat{Y}_j)^2$$

$$= \sum_j\sum_i (Y_{ij}-\bar{Y}_j)^2+\sum_j n_j(\bar{Y}_j-\hat{Y}_j)^2$$

$$= \sum_j (n_j-1)S_j^2+\sum_j n_j(\bar{Y}_j-\hat{Y}_j)^2$$

30

p.3.b. Complete the following table, and test $H_0: \mu_j = \beta_0 + \beta_1 X_j$ versus $H_A: \mu_j \neq \beta_0 + \beta_1 X_j$

$C = 6 \quad n = 6(15) = 90 \quad n - C = 84 \quad C - 2 = 4$

j	X_j	Ybar_j	Yhat_j		ANOVA				Coefficients	andard Err
1	176.5	3.43	3.64			SS		Intercept	7.6629	0.7566
2	192.6	3.41	3.27		Regressio	36.72		sugar	-0.0228	0.0035
3	209.3	2.98	2.89		Residual					
4	225	2.71	2.53		Total	110.85				
5	242.1	2.01	2.14		3					
6	258.5	1.71	~~3~~							

$SSE = 110.85 - 36.72 = 74.13$

$\hat{y}_6 = 7.6629 - .0228(258.5) = 7.6629 - 5.8938 = 1.7691$
(1.77)

j	$\bar{y}_j - \hat{y}_j$	$n_j(\bar{y}_j - \hat{y}_j)^2$
1	-0.21	.6615
2	0.14	.2940
3	0.09	.1215
4	0.18	.4860
5	-0.13	.2535
6	-0.06	.0540

$SSLF = 1.8605$ 12

$SSPE = 74.16 - 1.86$
$= 72.30$

$F_{LOF} = \dfrac{1.8605/\cancel{4}\,3}{72.30/84} = \dfrac{0.465}{0.86} = 0.541$

Test Statistic: $F_{LF} = 0.541$ 4

Rejection Region: $F_{LF} \geq F_{.05,4,84} \doteq 2.479$ 3

39

Q.4. A simple linear regression model is fit, relating NO_X concentration (Y) to Cargo Throughput (X) for a sample of n = 25 ports in People's Republic of China. The largest port in Cargo Throughput is Ningbo-Zhoushan. Below are the model summaries for the models with and without Ningbo-Zhoushan. This port's Cargo Throughput is 8.7347 and it's leverage value is $P_{ii} = v_{ii} = .3505$.

INV(X'X)				
0.1515	-0.0336			
-0.0336	0.0103			
	With Ningbo-Zhoushan		Without Ningbo-Zhoushan	
	Coefficients	*Standard Error*	*Coefficients*	*Standard Error*
Intercept	-0.6831	0.4641	-0.4465	0.5065
CargoTput	0.8505	0.1211	0.7568	0.1462
MSE	1.4220		1.4043	

$S^2_{(i)} = 1.4043$

$S_{(i)} = 1.1850$

p.4.a. What is $V\left\{\hat{Y}_i\right\}$ for Ningbo-Zhoushan?

5

$V\{\hat{Y}_i\} = \sigma^2 P_{ii} = \sigma^2(.3505)$ $\quad \hat{V}\{\hat{Y}_i\} = 1.4043(.3505) = .4922$

p.4.b. Obtain the fitted values for Ningbo-Zhoushan under the two model fits.

4 $\hat{Y}_i = -.6831 + .8505(8.7347) = 6.7458$

4 $\hat{Y}_{i(i)} = -.4465 + .7568(8.7347) = 6.1639$

With : 6.7458 Without: 6.1639

p.4.c. Compute DFFITS for Ningbo-Zhoushan

8

$$DFFITS_i = \frac{6.7458 - 6.1639}{1.1850\sqrt{.3505}} = \frac{.5819}{.7016} = .8294$$

8/8 p.4.d. Compute DFBETAS for Ningbo-Zhoushan

$S_{(i)}\sqrt{.1515} = .4612$ $\quad S_{(i)}\sqrt{.0103} = .1203$

$$\frac{(-.6831) - (-.4465)}{.4612} = -\frac{.2366}{.4612} \qquad \frac{.8505 - .7568}{.1203} = \frac{.0937}{.1203}$$

Intercept -.5130 CargoTput .7791

Q.5. Random samples of 18 men's NBA and 12 women's WNBA games were obtained. Three models were fit, regarding the following variables.

Y = Total Points in Game (100s of points) X_1 = Vegas Over/Under (100s points) X_2 = 1 if men's game, 0 if women's

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

p.5.a. The matrix form was fit, with the following results.

Model 1	All parameters free				
X1'X1					X1'Y
30.00	53.80	18.00	35.56		52.13
53.80	98.13	35.56	70.35		95.27
18.00	35.56	18.00	35.56		34.88
35.56	70.35	35.56	70.35		68.99
INV(X1'X1)					Beta-hat1
36.81	-24.17	-36.81	24.17		-0.16
-24.17	15.90	24.17	-15.90		1.05
-36.81	24.17	77.32	-44.64		0.45
24.17	-15.90	-44.64	26.26		-0.22
Y'IY	Y'J/nY	Y'P1Y	Y'(I-P1)Y		
93.3821	90.5846	92.5234	0.8587		

p.5.a. Give a point prediction for the predicted total score (100s of points) for a women's game with X_1 = 1.50.

5 $\hat{Y} = -0.16 + 1.05(1.50) + 0.45(0) - 0.22(0)$

$= -.16 + 1.575 = 1.415$

p.5.b. Set-up how you would obtain the standard error of the mean of all women's games with X_1 = 1.50 (numbers not symbols)

$X_h' = [1 \quad 1.5 \quad 0 \quad 0]$ $\qquad s^2 = \frac{.8587}{26} = .0330$

$\hat{V}\{X_h'\hat{\beta}\} = s^2 X_h'(X'X)^{-1}X_h = .0330 [1 \quad 1.5 \quad 0 \quad 0](X'X)^{-1}\begin{bmatrix}1\\1.5\\0\\0\end{bmatrix}$

12

$\hat{S}\{X_h'\hat{\beta}\} = \sqrt{\hat{V}\{X_h'\hat{\beta}\}}$

~~= 36.81~~

$.0330[36.81 + 1.5^2(15.90) + 2(1.5)(-24.17)]$

17 $= .04975$

$= .0330(0.075) = .002475$

p.5.c. Set-up how you would construct a 95% Confidence Interval for the ~~mea~~ mean of all women's games with $X_1 = 1.50$ (numbers not symbols) based on your answers to p.5.a. and p.5.b.

7

$t_{.025,26} = 2.056 \qquad S^2 = \frac{.8587}{26} = .0330$

$1.415 \pm 2.056 * \hat{SE}\{X_h'\hat{\beta}\}$ $(.049875)$ $= 1.415 \pm .102$

$= [1.313, 1.517]$

p.5.d. Give the **K'** matrix and **m** vector to test whether the true regression lines are the same for men and women, with intercept of 0, and slope with respect to over/under (X_1) of 1. Compute $\mathbf{K}'\hat{\boldsymbol{\beta}} - \mathbf{m}$

$H_0: E\{Y\} = X_1 \Rightarrow \beta_1 = \beta_2 = \beta_3 = 0, \beta_1 = 1$

8

8

$K' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \underset{\sim}{m} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad K'\hat{\beta} - m = \begin{bmatrix} -.16 \\ 1.05 \\ 0.45 \\ -0.22 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -.16 \\ .05 \\ .45 \\ -.22 \end{bmatrix}$

p.5.e. The error sum of squares for the model that is fit under the restrictions of p.5.d. is SSE = 0.8855.
Test $H_0: \mathbf{K}'\boldsymbol{\beta} - \mathbf{m} = 0$

$SSE_c = .8587 \qquad df_c = 30 - 4 = 26$

$SSE_R = .8855 \qquad df_R = 30 - 0 = 30$

14

$F_{obs} = \frac{\frac{.8855 - .8587}{30 - 26}}{\frac{.8587}{26}} = \frac{\frac{.0268}{4}}{\frac{.8587}{26}} = \frac{.0268(26)}{4(.8587)} = \frac{.6968}{3.4348}$

$= .2029$

$F_{.05,4,26} = 2.743$

Test Statistic: $F_{obs} = .2029$ Rejection Region: $F_{obs} \geq 2.743$

3