Total points = ~~324~~ 335

STA 6207 – Exam 1 – Fall 2018 PRINT Name ____________________

Conduct all tests at $\alpha = 0.05$ significance level

All Questions are based on the following 2 regression models, where SIMPLE REGRESSION refers to the case where p=1, and X is of full column rank (no linear dependencies among predictor variables).

$$\text{Model 1}: Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad i = 1,\ldots,n \quad \varepsilon_i \sim NID\left(0,\sigma^2\right)$$

$$\text{Model 2}: \mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \quad \mathbf{X} \equiv n \times p' \quad \boldsymbol{\beta} \equiv p' \times 1 \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0},\sigma^2\mathbf{I}\right)$$

Given: $\dfrac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a} \quad \dfrac{d(\mathbf{x}'\mathbf{Ax})}{d\mathbf{x}} = 2\mathbf{Ax}$ (**A** symmetric) $\quad E(\mathbf{Y}'\mathbf{AY}) = tr(\mathbf{AV_Y}) + \boldsymbol{\mu}_\mathbf{Y}'\mathbf{A}\boldsymbol{\mu}_\mathbf{Y}$

Cochran's Theorem: Suppose **Y** is distributed as follows with nonsingular matrix **V**:

$\mathbf{Y} \sim N\left(\boldsymbol{\mu},\sigma^2\mathbf{V}\right) \quad r(\mathbf{V}) = n$ then if **AV** is idempotent:

1. $\mathbf{Y}'\left(\dfrac{1}{\sigma^2}\mathbf{A}\right)\mathbf{Y}$ is distributed non-central χ^2 with: (a) df = $r(\mathbf{A})$ and (b) Noncentrality parameter: $\Omega = \dfrac{1}{2\sigma^2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$

2. $\mathbf{Y'AY}$, $\mathbf{Y'BY}$ independent if $\mathbf{AVB} = \mathbf{0}$ $\quad \mathbf{Y'AY}$, $\mathbf{BY}$ independent if **BVA**=0

Critical Values for t and F-distributions
F-distributions indexed by numerator df across top of table

			1	2	3	4	5	6
df	t(.05)	t(.025)	F(.05,1)	F(.05,2)	F(.05,3)	F(.05,4)	F(.05,5)	F(.05,6)
1	6.314	12.706	161.448	18.513	10.128	7.709	6.608	5.987
2	2.920	4.303	199.500	19.000	9.552	6.944	5.786	5.143
3	2.353	3.182	215.707	19.164	9.277	6.591	5.409	4.757
4	2.132	2.776	224.583	19.247	9.117	6.388	5.192	4.534
5	2.015	2.571	230.162	19.296	9.013	6.256	5.050	4.387
6	1.943	2.447	233.986	19.330	8.941	6.163	4.950	4.284
7	1.895	2.365	236.768	19.353	8.887	6.094	4.876	4.207
8	1.860	2.306	238.883	19.371	8.845	6.041	4.818	4.147
9	1.833	2.262	240.543	19.385	8.812	5.999	4.772	4.099
10	1.812	2.228	241.882	19.396	8.786	5.964	4.735	4.060
11	1.796	2.201	242.983	19.405	8.763	5.936	4.704	4.027
12	1.782	2.179	243.906	19.413	8.745	5.912	4.678	4.000
13	1.771	2.160	244.690	19.419	8.729	5.891	4.655	3.976
14	1.761	2.145	245.364	19.424	8.715	5.873	4.636	3.956
15	1.753	2.131	245.950	19.429	8.703	5.858	4.619	3.938
16	1.746	2.120	246.464	19.433	8.692	5.844	4.604	3.922
17	1.740	2.110	246.918	19.437	8.683	5.832	4.590	3.908
18	1.734	2.101	247.323	19.440	8.675	5.821	4.579	3.896
19	1.729	2.093	247.686	19.443	8.667	5.811	4.568	3.884
20	1.725	2.086	248.013	19.446	8.660	5.803	4.558	3.874
30	1.697	2.042	250.095	19.462	8.617	5.746	4.496	3.808
40	1.684	2.021	251.143	19.471	8.594	5.717	4.464	3.774
50	1.676	2.009	251.774	19.476	8.581	5.699	4.444	3.754
60	1.671	2.000	252.196	19.479	8.572	5.688	4.431	3.740
70	1.667	1.994	252.497	19.481	8.566	5.679	4.422	3.730
80	1.664	1.990	252.724	19.483	8.561	5.673	4.415	3.722
90	1.662	1.987	252.900	19.485	8.557	5.668	4.409	3.716
100	1.660	1.984	253.041	19.486	8.554	5.664	4.405	3.712
110	1.659	1.982	253.157	19.487	8.551	5.661	4.401	3.708
120	1.658	1.980	253.253	19.487	8.549	5.658	4.398	3.705
130	1.657	1.978	253.334	19.488	8.548	5.656	4.396	3.702
140	1.656	1.977	253.404	19.489	8.546	5.654	4.394	3.700
150	1.655	1.976	253.465	19.489	8.545	5.652	4.392	3.698
160	1.654	1.975	253.518	19.489	8.544	5.651	4.390	3.696
170	1.654	1.974	253.565	19.490	8.543	5.649	4.389	3.694
180	1.653	1.973	253.606	19.490	8.542	5.648	4.387	3.693

Q.1. A simple linear regression model was fit, relating the number of airplane tires consumed (Y) to the number of landings (X, in 100s) at an airfield over a series of periods. Based on Model 2, the following results were obtained.

X'X			X'Y		Y'Y
20	95.91		1051		59867
95.91	502.7151		5439.48		
INV(X'X)			Beta-hat		
0.58758	-0.11210		7.7761		
-0.11210	0.02338		9.3366		

3 p.1.a. How many periods were observed by the researchers? $n = 20$

3 p.1.b. How many total aircraft landings were there during the study? $\sum_i X_i = 95.91$ (9591 total landings)

3 p.1.c. How many tires were consumed over the study? $\sum_i Y_i = 1051$

p.1.d. Compute: **Y'IY** 5 **Y'PY** 7 $\mathbf{Y'}\frac{1}{n}\mathbf{JY}$ 7

$\underline{Y}'\underline{I}\underline{Y} = \underline{Y}'\underline{Y} = 59867$

$\underline{Y}'\underline{P}\underline{Y} = \underline{Y}'X(X'X)^{-1}X'\underline{Y} = \hat{\underline{\beta}}'X'\underline{Y} = 7.7761(1051) + 9.3366(5439.48) = 58958.93$

$\underline{Y}'\frac{1}{n}\underline{J}\underline{Y} = \frac{(\sum_i Y_i)^2}{n} = \frac{(1051)^2}{20} = 55230.05$

p.1.e. Compute $SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ 6 and $SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$ 6

$SSE = \underline{Y}'(\underline{I} - P)\underline{Y} = 59867 - 58958.93 = 908.07$

$SSR = Y'(P - \frac{1}{n}J)Y = 58958.93 - 55230.05 = 3728.86$

p.1.f. Compute $\hat{V}\{\hat{\boldsymbol{\beta}}\}$ 6

$S^2 = MSE = \frac{908.07}{20-2} = 50.45$

$$\hat{V}\{\hat{\underline{\beta}}\} = S^2 (X'X)^{-1} = \begin{bmatrix} 29.64 & -5.66 \\ -5.66 & 1.18 \end{bmatrix}$$

46

Q.2. A study in Beijing, China related heights (Y, cm) to mean foot length (X, cm) in a sample of n = 174 male college students. The sample means, standard deviations and correlations are given below. Complete the following table for the simple linear regression relating Height (Y) to mean foot length (X).

$$\overline{X} = 24.93 \quad \overline{Y} = 177.65 \quad s_X = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}} = 0.903 \quad s_Y = 4.810 \quad r_{XY} = 0.483$$

Regression Statistics					
R Square ③	.2333				
Residual Std Error ④	$\sqrt{17.84} = 4.2240$				
Observations ①	20				
ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression	② 1	⑥ 933.80	③ 933.80	③ 52.34	③ ≈ 3.897
Residual	② 18	⑥ 3068.75	② 17.848		
Total	② 19	③ 4002.55			
	Coefficients	*Standard Err*	*t Stat*	*Lower*	*Upper*
Intercept	⑤ 113.51	④ 8.87	② 12.79	② 96.00	③ 131.02
X	⑤ 2.57	④ 0.36	② 7.23	② 1.87	② 3.27

$t_{.025,172} \approx 1.974$

$r^2 = (.483)^2 = .2333$

$SS_{XX} = (174-1)(.903)^2 = 141.0658 \qquad SS_{YY} = 173(4.810)^2 = 4002.5453$

$r = \frac{SS_{XY}}{\sqrt{SS_{XX}SS_{YY}}} \Rightarrow SS_{XY} = .483\sqrt{141.0658(4002.5453)} = 362.9327$

70 $\frac{SSE}{SS_{YY}} = 1 - r^2 = .7667 \Rightarrow SSE = .7667(4002.5423) = 3068.75$

$\Rightarrow SSR = 4002.55 - 3068.75 = 933.80 \qquad s^2 = MSE = 17.8418$

$\hat{\beta}_1 = \frac{362.9327}{141.0658} = 2.5727 \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 177.65 - 2.5727(24.93) = 113.5126$

$\sqrt{17.8418/141.0658} = 0.3556 \quad \sqrt{17.8418\left(\frac{1}{20} + \frac{24.93^2}{141.0658}\right)} = 8.8718$

Q.3. Based on Model 1, defining $Q = \sum_{i=1}^{n} \varepsilon_i^2$: $p = 1$

p.3.a. Derive the normal equations. **Show all work**

20

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \qquad \frac{\partial Q}{\partial \beta_0} = 2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^{2-1}(-1)$$

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^{2-1} (-X_i)$$

setting each to 0

$$-2 \sum_i (Y_i - \beta_0 - \beta_1 X_i) = 0 \Rightarrow \sum_i Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_i X_i \quad (i)$$

$$-2 \sum_i (Y_i - \beta_0 - \beta_1 X_i) X_i = 0 \Rightarrow \sum_i X_i Y_i = \hat{\beta}_0 \sum_i X_i + \hat{\beta}_1 \sum_i X_i^2 \quad (ii)$$

p.3.b. Use the normal equations to show that $\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} \hat{Y}_i e_i = 0$ **Show all work**

20

$$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \Rightarrow \sum_i [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]$$

$$= \sum_i Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_i X_i = 0 \quad \text{from (i)}$$

$$X_i e_i = X_i Y_i - (\hat{\beta}_0 X_i + \hat{\beta}_1 X_i^2) \Rightarrow \sum_i [X_i Y_i - (\hat{\beta}_0 X_i + \hat{\beta}_1 X_i^2)]$$

$$= \sum_i X_i Y_i - \hat{\beta}_0 \sum_i X_i - \hat{\beta}_1 \sum_i X_i^2 = 0 \quad \text{from (ii)}$$

40

$$\hat{Y}_i e_i = (\hat{\beta}_0 + \hat{\beta}_1 X_i) e_i = \hat{\beta}_0 e_i + \hat{\beta}_1 X_i e_i \Rightarrow \sum_i \hat{Y}_i e_i = \hat{\beta}_0 \sum_i e_i + \hat{\beta}_1 \sum_i X_i e_i$$

$$= \hat{\beta}_0(0) + \hat{\beta}_1(0) = 0$$

Q.4. For the simple regression model (scalar form): $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1,\ldots,n \quad \varepsilon_i \sim NID(0,\sigma^2)$

we get: $\hat{\beta}_1 = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{S_{XX}}\right)Y_i, \quad \overline{Y} = \sum_{i=1}^{n}\left(\frac{1}{n}\right)Y_i, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

$$\sum_i (X_i - \overline{X}) = \sum_i X_i - n\overline{X} = \sum_i X_i - n\frac{\sum X_i}{n} = 0$$

$$\Rightarrow \overline{X}\sum_i (X_i - \overline{X}) = 0$$

p.4.a. Derive $E\left(\hat{\beta}_1\right), \quad E\left(\overline{Y}\right), \quad E\left(\hat{\beta}_0\right)$ **Show all work**

12 12 8

$$E\{\hat{\beta}_1\} = \frac{1}{SS_{XX}}\sum_{i=1}^{n}(X_i - \overline{X})E\{Y_i\} = \frac{1}{SS_{XX}}\sum_{i=1}^{n}(X_i - \overline{X})(\beta_0 + \beta_1 X_i)$$

$$= \frac{1}{SS_{XX}}\left[\beta_0 \sum_i (X_i - \overline{X}) + \beta_1 \sum_i (X_i - \overline{X})X_i - \beta_1 \sum_i (X_i - \overline{X})\overline{X}\right]$$

$$= \frac{1}{SS_{XX}}\left[\beta_0 (0) + \beta_1 \sum_i (X_i - \overline{X})(X_i - \overline{X})\right] = \frac{1}{SS_{XX}}\beta_1 SS_{XX} = \beta_1$$

$$E\{\overline{Y}\} = \frac{1}{n}\sum_i E\{Y_i\} = \frac{1}{n}\sum_i(\beta_0 + \beta_1 X_i) = \frac{1}{n}\left[n\beta_0 + \beta_1 \sum_i X_i\right]$$

$$= \beta_0 + \beta_1 \overline{X}$$

$$E\{\hat{\beta}_0\} = E\{\overline{Y} - \hat{\beta}_1\overline{X}\} = E\{\overline{Y}\} - \overline{X}E\{\hat{\beta}_1\}$$

$$= (\beta_0 + \beta_1\overline{X}) - \overline{X}\beta_1 = \beta_0$$

32

p.4.b. Derive $V\left(\hat{\beta}_1\right)$, $V\left(\overline{Y}\right)$, $\mathrm{COV}\left(\hat{\beta}_1, \overline{Y}\right)$, $V\left(\hat{\beta}_0\right)$, $\mathrm{COV}\left(\hat{\beta}_1, \hat{\beta}_0\right)$ **Show all work** $V\{Y_i\} = \sigma^2 \ \forall i$

10 10 10 10 10

$Y_1, \ldots, Y_n$ independent $\Rightarrow$ $V\{\sum_i a_i Y_i\} = \sum_i a_i^2 V\{Y_i\}$

$$\mathrm{Cov}\{\sum_i a_i Y_i, \sum_i b_i Y_i\} = \sum_i a_i b_i V\{Y_i\}$$

$$V\{\hat{\beta}_1\} = V\{\sum_i \frac{(X_i - \bar{X})}{SS_{XX}} Y_i\} = \sum_i \frac{(X_i - \bar{X})^2}{SS_{XX}^2}\sigma^2 = \frac{\sigma^2}{SS_{XX}^2} SS_{XX} = \frac{\sigma^2}{SS_{XX}}$$

$$V\{\bar{Y}\} = V\{\sum_i \frac{1}{n} Y_i) = \sum_i \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n^2} \sum_i 1^2 = \frac{\sigma^2 n}{n^2} = \frac{\sigma^2}{n}$$

$$\mathrm{Cov}\{\hat{\beta}_1, \bar{Y}\} = \sum_i \frac{(X_i - \bar{X})}{SS_{XX}} \left(\frac{1}{n}\right)\sigma^2 = \frac{\sigma^2}{n\, SS_{XX}} \sum_i (X_i - \bar{X}) = 0$$

$$V\{\hat{\beta}_0\} = V\{\bar{Y} - \hat{\beta}_1 \bar{X}\} = V\{\bar{Y}\} + \bar{X}^2 V\{\hat{\beta}_1\} - 2\bar{X}\,\mathrm{Cov}\{\bar{Y}, \hat{\beta}_1\}$$

$$= \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{SS_{XX}} - 2\bar{X}(0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{SS_{XX}}\right]$$

$$\mathrm{Cov}\{\hat{\beta}_1, \hat{\beta}_0\} = \mathrm{Cov}\{\hat{\beta}_1, \bar{Y} - \hat{\beta}_1\bar{X}\} = \mathrm{Cov}\{\hat{\beta}_1, \bar{Y}\} - \mathrm{Cov}\{\hat{\beta}_1, \hat{\beta}_1\bar{X}\}$$

$$= 0 - \bar{X} V\{\hat{\beta}_1\} = -\frac{\bar{X}\sigma^2}{SS_{XX}}$$

50

Q.5. Based on Model 2, with $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$:

$\underset{\sim}{Y} \sim N(X\beta, \sigma^2 I)$

$PP = X(X'X)^{-1}X'X(X'X)^{-1}X' = P$

p.5.a. Derive the sampling distributions of the vectors $\hat{\mathbf{Y}}$, $\mathbf{e}$, $\overline{\mathbf{Y}}$, $\mathbf{Y}-\overline{\mathbf{Y}}$ **Show all work**

$P' = [X(X'X)^{-1}X']' = X(X'X)^{-1}X' = P$

For each Normal (1) Mean (4) Variance (6)

$(I-P)(I-P) = I-P = (I-P)'$

$\frac{1}{n}J\frac{1}{n}J = \frac{1}{n^2}nJ = \frac{1}{n}J$

$E\{\hat{Y}\} = E\{X\hat{\beta}\} = E\{PY\} = PE\{Y\} = PX\beta = X(X'X)^{-1}X'X\beta = X\beta$

$V\{\hat{Y}\} = V\{PY\} = P\sigma^2 IP' = \sigma^2 PP = \sigma^2 P \qquad \hat{Y} \sim N(X\beta, \sigma^2 P)$

$E\{e\} = E\{(I-P)Y\} = (I-P)X\beta = X\beta - PX\beta = X\beta - X(X'X)^{-1}X'X\beta = X\beta - X\beta = 0$

$V\{e\} = (I-P)\sigma^2 I(I-P)' = \sigma^2(I-P) \qquad e\ N(0, \sigma^2(I-P))$

$E\{\frac{1}{n}JY\} = \frac{1}{n}JX\beta \qquad V\{\frac{1}{n}JY\} = \frac{1}{n}J\sigma^2 I(\frac{1}{n}J)' = \sigma^2\frac{1}{n}J$

4 24

$\Rightarrow \frac{1}{n}JY \sim N(\frac{1}{n}JX\beta, \sigma^2\frac{1}{n}J) \qquad Y-\bar{Y} = (I-\frac{1}{n}J)Y \Rightarrow Y-\bar{Y} \sim N(X\beta - \frac{1}{n}JX\beta, \sigma^2(I-\frac{1}{n}J))$

p.5.b. Derive the sampling distributions of $\frac{TSS}{\sigma^2}$ and $\frac{SS\mu}{\sigma^2}$ where $SS\mu = n\overline{Y}^2 = \mathbf{Y'A_\mu Y}$ and show they are independent. **Show all work**

$TSS = Y'(I-\frac{1}{n}J)Y \qquad SS\mu = Y'\frac{1}{n}JY \qquad V\{Y\} = \sigma^2 I$

For each χ^2 (2) df (4) Ω (6)

Independent (5) $TSS: (I-\frac{1}{n}J)I(I-\frac{1}{n}J)I = (I-\frac{1}{n}J)I$

$SS\mu: \frac{1}{n}JI\frac{1}{n}JI = \frac{1}{n}JI$

25

$\Rightarrow \chi^2$ distributions

$\frac{TSS}{\sigma^2}: df_1 = r(I-\frac{1}{n}J) = tr(I-\frac{1}{n}J) = tr(I) - \frac{1}{n}tr(J)$

$= n - \frac{1}{n}(n) = n-1$

$\Omega_1 = \frac{\beta'X'(I-\frac{1}{n}J)X\beta}{2\sigma^2}$

$\frac{SS\mu}{\sigma^2}: df_2 = r(\frac{1}{n}J) = tr(\frac{1}{n}J) = 1$

$\Omega_2 = \frac{\beta'X'\frac{1}{n}JX\beta}{2\sigma^2}$

73

$AVB = (I-\frac{1}{n}J)I(\frac{1}{n}J) = \frac{1}{n}J - \frac{1}{n}J\frac{1}{n}J = \frac{1}{n}J - \frac{1}{n}J = 0$

Q.6. A simple linear regression model is fit, where n=143 cats weighing over 2 kilograms were observed. The response variable was heart weight (Y, in grams) and the ~~response~~ predictor was body weight (X, in kilograms). The estimated regression coefficient vector and its variance covariance matrix are given below.

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.12 \\ 3.85 \end{bmatrix} \qquad \hat{V}\left\{\hat{\boldsymbol{\beta}}\right\} = \begin{bmatrix} 0.4489 & -0.1561 \\ -0.1561 & 0.0576 \end{bmatrix} \qquad \overline{X} = 2.71$$

p.6.a. Obtain the predicted heart weight for a cat weighing 3.0 kilograms.

6

$$\hat{Y}_3 = [1 \quad 3]\begin{bmatrix} 0.12 \\ 3.85 \end{bmatrix} = 0.12 + 11.55 = 11.67$$

p.6.b. Compute a 95% Confidence Interval for the mean heart weight for all cats weighing 3 kilograms.

18

$$\hat{V}\{\hat{Y}_3\} = V\{\hat{\beta}_0 + 3\hat{\beta}_1\} = V\{\hat{\beta}_0\} + 9V\{\hat{\beta}_1\} + 2(3)\,Cov\{\hat{\beta}_0, \hat{\beta}_1\}$$

$$= 0.4489 + 9(0.0576) + 6(-.1561) = .4489 + .5184 - .9366 = .0307$$

$$\hat{SE}\{\hat{Y}_3\} = \sqrt{.0307} = .1752 \qquad t_{.025,141} \approx 1.977$$

95% CI: $11.67 \pm 1.977(.1752) = 11.67 \pm 0.35 = (11.32, 12.02)$

~~p.6.c. Compute a 95% Prediction Interval for the heart weight of a new cat weighing 3 kilograms.~~

24