Statistics 206 (Solution)

Fall 2022

Midterm: Oct. 19, 10:00am - 11:50am, TLC 3214

Print name:

Print ID (all digits):

Sign name:

Instructions: This is an open notes exam. No mobile device of any kind is allowed. A handheld calculator is allowed. The duration of the exam is 110 minutes which include time for distributing and collecting the exam.

The total score is **100**. You must show your work for full credit. Partial credit can only be given if your thoughts can be followed. Make sure your name is written on the first page and all the additional pages attached by yourself (if any). Tables that may be useful are attached at the end of the exam sheet.

You must not show this exam to anyone outside of this class or post it anywhere.

Score:
1:

2:

3:

4:

Total:

1. (25 points) Tell true or false of the following statements with regard to the simple regression model and briefly explain your answer.

   (a) True. $\hat{\beta}_1 = r_{XY}\frac{s_Y}{s_X}$. So, if $s_Y^2 = s_X^2$, then the slope $\hat{\beta}_1 = r_{XY} \leq 1$.

   (b) True. $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} = 0$ if $\bar{X} = 0$.

   (c) False. The interval $[-1, 1]$ is a particular realization of the confidence interval, it either contains β_1 or not.

   (d) True. By $H^2 = H$ and $H' = H$, so $\sum_{j=1}^n h_{ij}^2 = h_{ii}$. We then have $h_{ii}(1 - h_{ii}) = \sum_{j \neq i} h_{ij}^2 \geq 0$. So, $0 \leq h_{ii} \leq 1$.

   Alternatively, $Var(\hat{Y}_i) = h_{ii}\sigma^2 \geq 0$ and $Var(e_i) = (1 - h_{ii})\sigma^2 \geq 0$. Thus, $0 \leq h_{ii} \leq 1$.

   (e) False. If this is true, then $Var(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H}) = c\mathbf{I}$. Because $\mathbf{I} - \mathbf{H}$ is a projection matrix, so $c = \sigma^2$. This leads to $\mathbf{H} = \mathbf{0}_{n\times n}$. But, this leads to a contradiction as the rank of $\mathbf{H}$ is at least 1 (due to the column $\mathbf{1}_n$).

2. (a) $\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}\{\bar{Y} + \hat{\beta}_1(X_i - \bar{X})\} = \bar{Y} + \hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) = \bar{Y}$.

   Alternatively, $\bar{Y} - \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \frac{1}{n}\sum_{i=1}^{n}e_i = 0$.

   (b) From lecture/homework, we know that $Cov(e_i, \hat{Y}_j) = 0$ for any $1 \le i, j \le n$. Since e_is and $\hat{Y}_j$s are jointly normally distributed, the set of e_is and the set of $\hat{Y}_j$s are independent. Therefore, their functional forms, $SSE = \sum_{i=1}^{n}e_i^2$ and $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n}\hat{Y}_j$ (from a.), are independent as well. Thus, $Cov(SSE, \bar{Y}) = 0$.

   Alternatively, from homework, we know that under the normal error model, $Cov(SSE, \hat{\beta}_0) = 0$ and $Cov(SSE, \hat{\beta}_1) = 0$. So, $Cov(SSE, \bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}Cov(SSE, \hat{Y}_i) = \frac{1}{n}\sum_{i=1}^{n}Cov(SSE, \hat{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n}X_iCov(SSE, \hat{\beta}_1) = 0$.

3. (a) Verify that $(\frac{1}{n}\mathbf{J}_n)' = \frac{1}{n}\mathbf{J}_n$ and $(\frac{1}{n}\mathbf{J}_n)^2 = \frac{1}{n}\mathbf{J}_n$ directly.

(b) Because all columns of $\frac{1}{n}\mathbf{J}_n$ equal $\frac{1}{n}\mathbf{1}_n$, thus its rank is one.

Alternatively, as a projection matrix $rank(\frac{1}{n}\mathbf{J}_n) = trace(\frac{1}{n}\mathbf{J}_n) = 1$.

4. (a) From the R output,
$y = -0.1284 + 1.7574x$

(b) First, from the R output, $\sqrt{MSE} = 1.004$, so $MSE = 1.004^2 \approx 1.008$ with degrees of freedom $n - 2 = 48$. Thus $SSE = 48 \times MSE \approx 48.384$. Because $R^2 = 1 - \frac{SSE}{SSTO}$ and from R output $R^2 = 0.7473$, we have $SSTO = \frac{SSE}{1-R^2} \approx 191.471$. Last, $SSR = SSTO - SSE \approx 143.087$ with the degrees of freedom 1, and thus $MSR = SSR$. Now, put everything into an ANOVA table:

Source of Variation	Sum of Squares	degrees of freedom	Mean Squares
Regression	SSR = 143.087	1	MSR= 143.087
Error	SSE = 48.384	48	MSE = 1.008
Total	SSTO = 191.471	49	

(c) By $Cov(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}$, $Var(\hat{Y}_1) = h_{11}\sigma^2 = 0.03 \times \sigma^2$, $Cov(\hat{Y}_1, \hat{Y}_2) = h_{12}\sigma^2 = 0.01 \times \sigma^2$. Because the fitted value and the residual are uncorrelated, $Cov(\hat{Y}_1, e_1) = 0$. Thus, $Var(\hat{Y}_1)$ is the larges among the three, and $Cov(\hat{Y}_1, e_1)$ is the smallest among the three.

(d) By $Var(\mathbf{e}) = \sigma^2(\mathbf{I}_n - \mathbf{H})$, $Var(e_3) = (1 - h_{33})\sigma^2$, and thus, the standard error is $s.e.(e_3) = \sqrt{(1 - 0.021)MSE} \approx 0.9934$.

(e) $\sum_{i=1}^{50}\sum_{j=1}^{50} h_{ij} = \mathbf{1}_n'\mathbf{H}\mathbf{1}_n = \mathbf{1}_n'\mathbf{1}_n = n = 50$.

(f) The test statistic is $T^* = \frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = 11.915$ (from the R output). At level 0.01, we should reject the null hypothesis $H_0 : \beta_1 = 0$ if $T^* > t(0.99, 48) = 2.407$. So here, at the significance level 0.01, we reject $H_0 : \beta_1 = 0$ and conclude that $\beta_1 > 0$.

(g) $\hat{Y}_h \pm t(0.95, 48)\sqrt{MSE\left(1 + \frac{1}{n} + \frac{(0-\bar{X})^2}{(n-1)s_x^2}\right)} = -0.1284 \pm 1.677\sqrt{1.008\left(1 + \frac{1}{50} + \frac{0.13^2}{49 \times 0.946}\right)} \approx [-1.8291, 1.5723]$.

(h) We need the Normal error assumption.

END OF EXAM.

Table 1: t table: $t(A;df)$

$A \backslash df$	43	44	45	46	47	48	49	50
0.90	1.302	1.301	1.301	1.300	1.300	1.299	1.299	1.299
0.95	1.681	1.680	1.679	1.679	1.678	1.677	1.677	1.676
0.975	2.017	2.015	2.014	2.013	2.012	2.011	2.010	2.009
0.99	2.416	2.414	2.412	2.410	2.408	2.407	2.405	2.403
0.995	2.695	2.692	2.690	2.687	2.685	2.682	2.680	2.678