STAT 343 Final exam—Dec 6 2021

Name: Madhuri Raman

Instructions

1. This is an open-book exam. You may use course materials, your own notes, textbooks, and standard online textbook-type references like Wikipedia. You are not allowed to consult with anyone (aside from the instructor/TAs), or to use online Q&A forums such as Quora, etc. You may need to use a calculator (or you can use R as a calculator).
2. You have 90 minutes to finish the questions. After 90 minutes, please upload your solutions to Gradescope. We understand this may take several minutes; submissions that are over 10 minutes late will be penalized. Please contact the instructor immediately in case of technical difficulties that are preventing you from uploading your solutions on time.
3. For multi-part problems, if you cannot complete an earlier part, please make up numbers for the earlier part so that you can work on the later parts, if needed.
4. For all problems, for full credit you must explain your answer and/or show your work, unless specified otherwise.

Academic integrity

Please complete this section **after you finish your exam**. Initial each statement and then sign.

✓ I have not consulted anyone aside from instructors/TAs while taking this exam, and have not used any resources aside from course materials, my own notes, textbook(s), and a calculator (or R).

✓ I have not discussed this exam, or shared or received any information about this exam, with other students.

✓ After completing this exam, I will not discuss the exam with other students who have not yet taken it.

Signed: ______________

Be sure to include this page when you scan the exam to hand in. If you are not printing the exam, you can complete this section electronically by emailing the instructor — copy-and-paste the text above and sign your name.

1. p large, n, train + test sets

a-1) training error → always will go down (starts high)
↳ by definition

$\boxed{\text{Plot 1}}$

a-2) validation error → will go down then up again (starts high)
↳ can be optimized

$\boxed{\text{Plot 3}}$

b) Transform Y to log(Y)

This will model a "slower" rate of change than the original data so it will help the slope become more constant → linear model

$\boxed{\text{iii}}$

c) Only significant change in Y if A=1, B=1 and C=1

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C$$

all 1: $Y = \beta_0 + (\beta_1 + \beta_2 + \beta_3) \to \neq 0$

any 0: $Y = \beta_0 + \beta_1 + \beta_2 \longrightarrow \beta_1 + \beta_2 = 0$

or $\beta_2 + \beta_3 \longrightarrow \beta_2 + \beta_3 = 0$

or $\beta_1 + \beta_3 \longrightarrow \beta_1 + \beta_3 = 0$

or $\beta_1 \longrightarrow \beta_1 = 0$

or $\beta_2 \longrightarrow \beta_2 = 0$

or $\beta_3 \longrightarrow \beta_3 = 0$

≈ all 0: $Y = \beta_0$ only.

$\boxed{\text{ii}}$ Expected values of $\hat{\beta}_A, \hat{\beta}_B, \hat{\beta}_C$ are nonzero (but may be small) because we know that under $A=1, B=1, C=1$, $\hat{\beta}_A + \hat{\beta}_B + \hat{\beta}_C$ will be nonzero aka it will have a significant change the value of Y, but that is only in 1 of the possible outcomes out of the 8.

2. Y = change_temp

X_1 = out_temp (continuous)

X_2 = insu (Low, High)
$0 \quad 1$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 \underbrace{X_2}_{} + \beta_3 X_1 \cdot X_2 + \epsilon_i$$

0,1   continuous   0,1

a) High insul = 3x Low insul regardless of out temp X_1 "Change in temp inside cooler" = Y_i

low: $Y_i = \beta_0 + \beta_1 X_1$

continuous

high: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 + \beta_3 X_1$

$= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$

$\Rightarrow \beta_0 + \beta_1 X_1 = 3[\beta_0 + \beta_1 X_1 + \beta_2 + \beta_3 X_1]$

$\Rightarrow 2\beta_0 + 2\beta_1 X_1 + 3\beta_2 + 3\beta_3 X_1 = 0$

$\Rightarrow$ $\boxed{H_0: (2\beta_0 + 3\beta_2) + (2\beta_1 + 3\beta_3) X_1 = 0}$

b) "If $X_1 = 60$, $Y_{low} = Y_{high}$:

low: $Y_i = \beta_0 + \beta_1 \cdot 60$

high: $Y_i = \beta_0 + \beta_1 \cdot 60 + \beta_2 + \beta_3 \cdot 60$

$= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \cdot 60$

$\Rightarrow \beta_0 + \beta_1 \cdot 60 = \beta_0 + \beta_2 + (\beta_1 + \beta_3) 60$

$\Rightarrow \cancel{\beta_0} + \cancel{\beta_1 60} = \cancel{\beta_0} + \beta_2 + \cancel{\beta_1 60} + \beta_3 \cdot 60$

$\Rightarrow 0 = \beta_2 + \beta_3 \cdot 60$

$\Rightarrow$ $\boxed{H_0: \beta_2 = -60\beta_3}$

c) Imputation via mean for X_1

But there is actually high correlation btwn X_1 and X_2.

Specifically, $\uparrow X_1 \Leftrightarrow \uparrow$ insul = high ($X_2 = 1$) $\Rightarrow$ imputation by mean will NOT take this relationship btwn X_1 and X_2 into account,

So it is more likely that imputation by mean will <u>underestimate</u> the difference between High and Low since we are imputing the same mean for all values and not respecting the inherent correlation btwn X_1 and X_2.

$\boxed{\text{(ii)}}$

3) Y = response $\quad n=56, \; K=7, \quad n-K=49,$

A = Continuous

L = categorical $\rightarrow$ 7 levels = 6 dummy variables

a) $Y = \beta_0 + \beta_1 A + \beta_2 L_2 \ldots\ldots \beta_7 L_7 + [\text{all}\overset{6}{\ } \text{interactions}]$

WTK F statistic for FULLEST model vs. $Y \sim A$ only (and df of Ftest) WTK

	Df	SumSq	MeanSq	Fvalue	P(>F)
A	1	25.22	25.22		
L	6	13.03	2.172		
A:L	6	7.95	1.325		
Residuals	42	12.52			

$n=56, \; 56-13-1 = 42$

$$F = \frac{(RSS_{partial} - RSS_{full}) / (df_{partial} - df_{full})}{RSS_{full} / df_{full}} = \frac{(13.03 + 7.95)/12}{12.52/42}$$

$$\Rightarrow \boxed{F = 5.865 \quad \text{where } F \underset{\text{under } H_0}{\sim} F_{\underset{df}{12,42}}}$$

b) for each lab $\overset{1\ldots7}{L}$: $\tilde{Y}_i = Y_i - \bar{Y}_{L_i}$ ← sample mean for level of i.

Now:

$$Y_i = \beta_0 + \beta_1 A + \sum_{K=2}^{7} \beta_{LK} \cdot \mathbb{1}\{L_i = K\} + \epsilon_i$$

Let $M_{n\times p} \in \{0,1\}^{n \times L}$ where $M_{i\ell} = 1$ if point i is in level ℓ, $\ell = 1 \ldots L$ ← $L = p-2$

where

$$\tilde{Y}_i = \beta_0 + \beta_1 A + \sum_{K=2}^{7} \beta_{LK} \cdot \mathbb{1}\{L_i = K\} - \bar{Y}_{L_i}$$

$$\Rightarrow \boxed{\tilde{Y}_i = \beta_0 + \beta_1 A + \underset{n\times(p-2)}{M} \cdot \underset{(p-2)\times 1}{\beta_{-0,-1}}}$$

← original β vector excluding 0 and 1, $L = p-2$

c) For this new data set, the moment conditions are
$E[\tilde{\epsilon}_i] = 0$? $\rightarrow$ $E[\tilde{\epsilon}_i] = E[\epsilon_i + \underbrace{(Y_i - \bar{Y}_{l_i})}] = E[\epsilon_i] + E[Y_i - \bar{Y}_{l_i}]$
$Var[\tilde{\epsilon}_i] = \sigma^2$?

$\boxed{\text{No,}}$ they don't, because now our $\tilde{\epsilon}_i$ are not independent of each other since datapoints in the same level will have correlated errors now.

d) Back to original response $\rightarrow Y \sim A + L$
↳ different variance for different levels.
↳ now, different vars for different i's (within each level same though)

The more general model would NOT be equivalent to running 7 separate models of $Y \sim A$ for each lab. This is because the general method, although there are different variances of the noise term for each i, still aims to compare each of the labs to the reference lab and not only accounts for the "within lab level" variance, but we also have "between lab level" variations that do NOT get captured by the model coefficients when we run completely separate $Y \sim A$ regressions for each lab.

e) Yes, we do know that the max/largest pairwise difference gives a p-value of 0.03 which is signif at $\alpha = 0.05$, so we know exactly which pair is signif. different. However we don't know from this if any other pairs are significantly different.

So it is true that the group means are not exactly equal but we don't know exactly which.

4. Lasso Regression, $X_{n \times p}$, $Y_{n \times 1}$, $\hat{\beta}_{p \times 1}$, $\beta_{p \times 1}$

$$\hat{\beta}_\lambda = \underset{\beta}{\operatorname{argmin}} \left\{ \|Y - X\beta\|_2^2 + \lambda \cdot \|\beta\|_1 \right\}$$ ↖ NO INTERCEPT

Special case: columns of X are all orthogonal $\rightarrow X^T X = I$

Linear model moment + normality assumptions satisfied ✓

Show that $\hat{\beta}_1 \ldots \hat{\beta}_p$ are independent.

$$\hat{\beta}_\lambda = \underset{\beta}{\operatorname{argmin}} \left\{ \|Y - X\beta\|_2^2 + \lambda \cdot \|\beta\|_1 \right\}$$

$$= \underset{\beta}{\operatorname{argmin}} \left\{ \|X\hat{\beta}^{OLS} - X\beta\|_2^2 + \lambda \cdot \|\beta\|_1 \right\} \quad \hat{Y} = X\hat{\beta}^{OLS}$$

$$= \underset{\beta}{\operatorname{argmin}} \left\{ \cancel{X^T} \|\hat{\beta}^{OLS} - \beta\|_2^2 \cancel{X} + \lambda \cdot \|\beta\|_1 \right\}$$

since $X^T X = I$

$$= \underset{\beta}{\operatorname{argmin}} \left\{ \|\hat{\beta}^{OLS} - \beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

$\frac{d}{d\beta} \Rightarrow$ so $\hat{\beta}_\lambda = \hat{\beta}^{OLS} - \hat{\beta} + \operatorname{sign}(\hat{\beta})$

$= X^T Y$, $\frac{\lambda}{2}$