STA 35C Statistical Data Science III

Practice midterm exam 2 – Set B

Instructor: Dogyoon Song

Name: _____ **Student ID:** _____

Instructions: This midterm exam is **closed-book**, except for the permitted note sheet described below. You may bring a pen or pencil, one letter-sized sheet of *hand-written* notes (both sides), and a *non-graphing* calculator. No other materials (e.g., textbooks) are allowed. For a timed attempt, allow 50 minutes. Problems 1–4 total **120 points**; Problem 5 offers *up to 10 bonus points*. Once you receive this exam, please confirm that you have all 7 pages.

- Make sure to clearly write your name and student ID above.
- For Problem 1, give only the requested answers; no explanations are required. For Problems 2–5, show the relevant steps for partial credit. Sentence limits apply to explanations, not displayed calculations.
- If necessary, round numerical answers to three decimal places; fractions, logarithms, and exponentials may be left unevaluated.
- Complete Problems 1–4 before attempting the optional bonus questions in Problem 5.

Problem	**Score**
Problem 1	
Problem 2	
Problem 3	
Problem 4	
Problem 5 (bonus)	
Total	

Problem 1 (30 points). Basic concepts

Give only the requested answers. No explanations are required. For each multiple-choice part, **circle all correct options**; more than one may be correct.

(a) (5 points) An OLS model with an intercept already contains X_1. After adding $X_2 = 2X_1$ and refitting on the **same training observations**, can its minimum training RSS decrease? Circle **Yes** or **No**.

(b) (5 points) For model selection and assessment, **circle all correct statements**.

A. Each held-out CV prediction must come from a model fitted without that observation.
B. A validation set remains independent of a model choice made using its responses.
C. After CV tuning, the model may be refitted on all development data before evaluation on an untouched test set.
D. LOOCV always gives a smaller estimated prediction error than five-fold CV.

(c) (5 points) A bootstrap resample draws two observations independently with replacement from the original sample $1, 5$. Let $\overline{X}^*$ be the resample mean. Conditional on this original sample, **compute** $P(\overline{X}^* = 3)$.

(d) (5 points) Adding a predictor to an OLS model leaves its **positive training RSS unchanged** but reduces the residual degrees of freedom by one. On the same observations, does adjusted R^2 **increase, decrease, or stay the same**? Assume both values are defined.

(e) (5 points) Consider ridge and lasso with equally weighted slope penalties and an unpenalized intercept. All predictors have positive sample standard deviations. **Circle all correct statements**.

A. Centering alone removes sensitivity to changing a predictor's measurement unit.
B. Centering and scaling each predictor to unit standard deviation removes this sensitivity.
C. A nonzero lasso coefficient establishes the predictor's statistical significance.
D. The penalty parameter may be chosen using cross-validation error.

(f) (5 points) For multiple testing, define the false-discovery proportion as zero when there are no rejections. **Circle all correct statements**.

A. FDR control at 0.10 guarantees at most 10% false discoveries among the rejections in every dataset.
B. FDR control at 0.10 means each rejected null has probability at most 10% of being true.
C. FDR control at 0.10 bounds the expected false-discovery proportion by 0.10.
D. FWER control at 0.10 is sufficient for FDR control at 0.10.

Problem 2 (30 points). Prediction assessment and uncertainty

Each part concerns a separate analysis.

(a) (10 points) At a fixed input x_0, $f(x_0) = 5$. A new response is $Y_0 = 5 + \varepsilon_0$, where ε_0 has mean 0 and variance 2 and is independent of the training sample. Two methods have:

	Method A	Method B
Mean fitted prediction across training samples	5	4
Variance of the fitted prediction across training samples	4	1

Choose the method with the smaller expected squared prediction error, supporting your choice with both error calculations.

(b) (10 points) A deterministic prediction method is evaluated by leave-one-out CV. Each prediction below is obtained by fitting on the **other three observations**:

Held-out observation	1	2	3	4
Observed response	3	5	4	8
Leave-one-out prediction	2	5	6	7

(i) **Compute the LOOCV MSE**.

(ii) Does the absence of a random split imply that this estimate has no variability across new datasets? Answer in **one sentence**.

(c) (10 points) An i.i.d. sample of delivery times has mean 12 minutes. The bootstrap estimate of the sample mean's standard error is 1.5 minutes. Use the normal-approximation rule

$$\overline{y} \pm 1.96\,\widehat{\mathrm{SE}}_{\mathrm{boot}}.$$

Report the **approximate 95% confidence interval for the population mean delivery time** and a **one-sentence interpretation**.

Problem 3 (30 points). Model selection and shrinkage

Each part concerns a separate analysis.

(a) (10 points) Two polynomial regressions, of degrees 1 and 8, are fitted by ordinary least squares to the **same training observations**, both with an intercept. They are evaluated on the **same independent test sample** from the same population:

Polynomial degree	Training MSE	Test MSE
1	8	10
8	1	14

An analyst claims, "The decrease in training MSE from 8 to 1 shows that squared bias decreased by 7."

Assess this claim and explain how the degree-8 model can have lower training MSE but higher test MSE. Use the bias–variance tradeoff in your explanation, in **at most three sentences**.

(b) (10 points) For a **separate training dataset**, the table gives the training RSS of all $2^4 = 16$ OLS models using subsets of X_1, X_2, X_3, X_4. Every model includes an intercept; $\varnothing$ denotes the intercept-only model.

0–1 predictors		**2 predictors**		**3–4 predictors**	
Predictors	RSS	Predictors	RSS	Predictors	RSS
$\emptyset$	46	X_1, X_2	30	X_1, X_2, X_3	26
X_1	38	X_1, X_3	38	X_1, X_2, X_4	12
X_2	30	X_1, X_4	20	X_1, X_3, X_4	11
X_3	44	X_2, X_3	28	X_2, X_3, X_4	16
X_4	45	X_2, X_4	18	X_1, X_2, X_3, X_4	10
		X_3, X_4	43		

Starting with all four predictors, apply **backward stepwise selection using training RSS** until exactly two predictors remain. **Report the selected subset at each step**, showing the RSS comparisons, and **state whether the resulting pair has the smallest RSS among all two-predictor subsets**, with a one-sentence justification.

(c) (10 points) For a **separate analysis** with four standardized predictors, either ridge or lasso is fitted at three penalty values, with an unpenalized intercept. The table gives full-data slope estimates and CV MSEs computed using the **same folds**; the displayed zeros are exact.

λ	$\widehat{b}_1$	$\widehat{b}_2$	$\widehat{b}_3$	$\widehat{b}_4$	CV MSE (rounded)
6	3.3	0.7	−2.3	0.5	17.0
30	2.5	0	−1.5	0	17.4
120	0	0	0	0	26.8

Identify the method most consistent with the coefficient pattern, and explain when $\lambda = 30$ might be preferred to $\lambda = 6$. Justify both answers in **at most three sentences**, referring to the fitted coefficients and CV errors.

Problem 4 (30 points). Reviewing a customer-renewal analysis

A team predicts whether customers will renew their subscriptions and separately studies associations with renewal.

(a) (10 points) Two classification methods use the **same three validation folds**, which partition 100 customers. Each method is refitted using only the observations outside the current validation fold. The table gives **counts of held-out mistakes**:

Validation fold	1	2	3
Number of customers	20	30	50
Mistakes by Method A	1	3	15
Mistakes by Method B	3	5	8

Choose the method with the smaller CV misclassification rate, supporting your choice with both rates pooled over customers.

(b) (10 points) In a separate study, five prespecified tests have mutually independent, valid p-values. The entire family is listed below in increasing order:

Null hypothesis	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$	$H_{0,4}$	$H_{0,5}$
p-value	0.010	0.045	0.055	0.090	0.300

Apply **Benjamini–Hochberg at target FDR** $q = 0.10$. **Report the rejected null hypotheses**, showing the rank-dependent cutoff comparisons.

(c) (10 points) After training and locking a classifier, the team estimates its error rate using a **separate evaluation sample**. It bootstraps those evaluation rows, keeping the classifier fixed. A colleague claims:

> "Increasing the number of bootstrap replicates from 500 to 2,000, without collecting new observations, halves the sampling uncertainty of our original error-rate estimate."

Assess the claim in at most two sentences, distinguishing the original estimate's sampling uncertainty from Monte Carlo error in the bootstrap approximation.

Problem 5 (10 bonus points). Optional bonus questions

These questions are independent of the required questions and of each other.

(a) (5 points) Two predictors are identical on every training observation: $x_{i1} = x_{i2}$ for all i. With $\lambda > 0$, lasso minimizes

$$\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2})^2 + \lambda(|b_1| + |b_2|).$$

Suppose $(b_0, b_1, b_2) = (5, 2, 0)$ is a minimizer. **Construct a different minimizer with both slopes nonzero**. Justify your answer by comparing its fitted training values and penalty with those of the given minimizer.

(b) (5 points) Let R be the number of rejected null hypotheses and V the number of false rejections. Recall

$$\text{FWER} = P(V \geq 1), \qquad \text{FDR} = \mathbb{E}\left[\frac{V}{\max\{R, 1\}}\right].$$

If **all null hypotheses are true**, what is the relationship between FDR and FWER? **Justify it from these definitions**; no independence assumption is imposed.