STAT151A Final Exam (Spring 2024)

Please write your full name and email address:

- You have 3 hours for this exam.
- You will be allowed one double-sided cheat sheet.
- Please turn in your cheat sheet with the exam.

In the exam, you will find seven questions. From these, please choose **exactly four** to answer. You will be graded only on the four questions you choose to answer.

Please mark the questions you would like graded with an $\times$ in the box provided. For example, to select problems 2, 4, 5, and 7, your exam should look like this:

Question 1 $\bigcirc \leftarrow$ **'X' here to grade this question.**

Question 2 $\otimes \leftarrow$ **'X' here to grade this question.**

Question 3 $\bigcirc \leftarrow$ **'X' here to grade this question.**

Question 4 $\otimes \leftarrow$ **'X' here to grade this question.**

Question 5 $\otimes \leftarrow$ **'X' here to grade this question.**

Question 6 $\bigcirc \leftarrow$ **'X' here to grade this question.**

Question 7 $\otimes \leftarrow$ **'X' here to grade this question.**

If you select more than four questions, we reserve the right to choose which ones we grade.

Note that each of the seven questions has three parts. Read the question completely and carefully before answering. Make sure to answer every part to receive full credit.

In this exam, you may use the following results without proof:

- The continuous mapping theorem.
- The law of large numbers and central limit theorems for independent random variables…
  - For univariate random variables,
  - For multivariate random variables,
  - For identically distributed random variables, and
  - For non-identically distributed random variables.
- The OLS estimator is given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$ when $\boldsymbol{X}$ is full-rank.
- The OLS estimator satisfies $(\boldsymbol{X}^\intercal \boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\intercal \boldsymbol{Y}$.

1 Question 1 ◯ ← ‘X’ here to grade this question.

For this question, we’ll consider the linear model $y_n = \beta_0 + \beta_1 w_n + \beta_2 z_n + \varepsilon_n$.

(1a)

Write the set of equations

$$y_n = \beta_0 + \beta_1 w_n + \beta_2 z_n + \varepsilon_n \quad \text{for } n \in \{1, \ldots, N\}$$

in matrix form. That is, let $\boldsymbol{X}$ denote an $N \times 3$ matrix, $\boldsymbol{Y}$ and $\boldsymbol{\varepsilon}$ length–N column vectors, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\intercal$ a length–3 column vector. Then express the matrices $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{\varepsilon}$ in terms of the scalars y_n, w_n, z_n, and ε_n so that $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is equivalent to the set of regression equations.

(1b)

Define the following quantities:

$$
\begin{aligned}
\overline{z} &:= \frac{1}{N}\sum_{n=1}^{N} z_n & \overline{w} &:= \frac{1}{N}\sum_{n=1}^{N} w_n & & \\
\overline{zz} &:= \frac{1}{N}\sum_{n=1}^{N} z_n^2 & \overline{ww} &:= \frac{1}{N}\sum_{n=1}^{N} w_n^2 & \overline{zw} &:= \frac{1}{N}\sum_{n=1}^{N} z_n w_n \\
\overline{zy} &:= \frac{1}{N}\sum_{n=1}^{N} z_n y_n & \overline{wy} &:= \frac{1}{N}\sum_{n=1}^{N} w_n y_n & \overline{y} &:= \frac{1}{N}\sum_{n=1}^{N} y_n
\end{aligned}
$$

Write an explict expressions for $\frac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{X}$ and $\frac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{Y}$ in terms of these quantities.

(1c)

For this part of the question only, assume that

$$\overline{zw} = \overline{z} = \overline{w} = 0$$

.

In terms of the quantites defined in part (b), write a closed–form expression for $\hat{\boldsymbol{\beta}}$, the OLS estimator of the vector $\boldsymbol{\beta}$.

Hint: The inverse of a 3×3 diagonal matrix is given by

$$\begin{pmatrix} v_1 & 0 & 0 \\ 0 & v_2 & 0 \\ 0 & 0 & v_3 \end{pmatrix}^{-1} = \begin{pmatrix} v_1^{-1} & 0 & 0 \\ 0 & v_2^{-1} & 0 \\ 0 & 0 & v_3^{-1} \end{pmatrix}$$

2 Question 2 ◯ ← 'X' here to grade this question.

For this question, consider the linear models

$$y_n = \boldsymbol{\beta}^\intercal \boldsymbol{x}_n + \varepsilon_n \quad \text{and} \quad y_n = \boldsymbol{\gamma}^\intercal \boldsymbol{z}_n + \eta_n$$

with

$$\boldsymbol{x}_n = \begin{pmatrix} 1 \\ x_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{z}_n = \begin{pmatrix} 1 \\ z_n \end{pmatrix} \text{ where } z_n := 10 x_n$$

Assume that x_n is not a constant (i.e., for at least one pair n and m, $x_n \neq x_m$.).

Let $\boldsymbol{X}$ denote the $N \times 2$ matrix whose n–th row is $\boldsymbol{x}_n^\intercal$, and $\boldsymbol{Z}$ denote the $N \times 2$ matrix whose n–th row is $\boldsymbol{z}_n^\intercal$.

(2a)

Using the definitions above, find a 2×2 matrix $\boldsymbol{A}$ such that $\boldsymbol{z}_n = \boldsymbol{A}\boldsymbol{x}_n$. Then, show that $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A}^\intercal$ for the same matrix $\boldsymbol{A}$.

(2b)

Suppose you know that the OLS estimate of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (4, 30)$. What is the value of $\hat{\boldsymbol{\gamma}}$, the OLS estimate of $\boldsymbol{\gamma}$? Please use the definitions above and your answer for part (a), and justify your answer.

(2c)

Now consider the two models' prediction on a new datapoint with $\boldsymbol{x}_{\text{new}} = (1, 50)^\intercal$ — and so, necessarily, $\boldsymbol{z}_{\text{new}} = (1, 500)^\intercal$ — with respective prediction errors

$$\varepsilon^{\beta}_{\text{new}} := y_{\text{new}} - \hat{\boldsymbol{\beta}}^\intercal \boldsymbol{x}_{\text{new}} \quad \text{and} \quad \varepsilon^{\gamma}_{\text{new}} := y_{\text{new}} - \hat{\boldsymbol{\gamma}}^\intercal \boldsymbol{z}_{\text{new}}.$$

Please select which of (a), (b), (c), or (d) is correct for this particular value of $\boldsymbol{x}_{\text{new}}$ and $\boldsymbol{z}_{\text{new}}$:

a) It is always the case that $\left|\varepsilon^{\beta}_{\text{new}}\right| = |\varepsilon^{\gamma}_{\text{new}}|$

b) It is always the case that $\left|\varepsilon^{\beta}_{\text{new}}\right| < |\varepsilon^{\gamma}_{\text{new}}|$

c) It is always the case that $\left|\varepsilon^{\beta}_{\text{new}}\right| > |\varepsilon^{\gamma}_{\text{new}}|$

d) In general, we cannot determine the relationship between $\left|\varepsilon^{\beta}_{\text{new}}\right|$ and $|\varepsilon^{\gamma}_{\text{new}}|$ using the information provided.

Please justify your answer carefully.

3 Question 3 ◯ $\leftarrow$ 'X' here to grade this question.

For this question, assume that you have access to a function

$$\Phi(z) = \mathbb{P}\left(\tilde{z} \leq z\right),$$

as well as its inverse,

$$\Phi^{-1}(p) = z \text{ such that } \mathbb{P}\left(\tilde{z} \leq z\right) = p, \text{ for } p \in (0, 1),$$

where $\tilde{z} \sim \mathcal{N}\left(0, 1\right)$ denote a scalar-valued standard normal random variable.

(3a)

Suppose that $\tilde{y} \sim \mathcal{N}\left(\mu, \sigma^2\right)$, where μ and σ are known. Using only the functions $\Phi(\cdot)$, $\Phi^{-1}(\cdot)$, and the known quantities μ, and σ, find find a quantity a such that

$$\mathbb{P}\left(\tilde{y} \leq a\right) = 0.90.$$

Note that $\Phi(\cdot)$ is only for a standard normal random variable. Please **do not assume** that you have direct access to the distribution and quantile functions of generic normal random variables.

Please justify your answer carefully.

(3b)

Now, consider the OLS estimator under normal assumptions, so that

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\right).$$

Note that $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ are P–dimensional vectors, σ is a scalar, and $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$ is a $P \times P$ matrix.

Assume that $\boldsymbol{\beta}$, σ, and $\boldsymbol{X}$ are all known. In terms of these quantities, find the distribution of $\hat{\boldsymbol{\beta}}_1$, the first component of $\hat{\boldsymbol{\beta}}$.

(3c)

Combining your answers from parts (a) and (b), find a scalar b such that

$$\mathbb{P}\left(\hat{\boldsymbol{\beta}}_1 \leq b\right) = 0.90.$$

4 Question 4 ◯ ← 'X' here to grade this question.

For this question:

- Let $\boldsymbol{x}_n = (1, z_n)^\intercal$, where $z_n \sim \mathcal{N}(0, 2)$.
- Assume that $y_n = \boldsymbol{\beta}^\intercal \boldsymbol{x}_n + \varepsilon_n$ for each n and some $\boldsymbol{\beta} = (\beta_1, \beta_2)^\intercal$, where $\boldsymbol{\beta}$ is a length 2 column vector.
- Assume the residuals ε_n are IID with $\mathbb{E}[\varepsilon_n] = 0$ and $\mathbb{E}[\varepsilon_n^2] = 1$, but **not necessarily normal.**
- Assume that the residuals ε_n are all independent of all the z_n.

(4a)

Let $\boldsymbol{X}$ denote the $N \times P$ matrix consisting of the observation $\boldsymbol{x}_n^\intercal$ in the n–th row, and let $\boldsymbol{Y}$ denote the N–vector with y_n in the n–th entry.

Write the matrices $\frac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{X}$ and $\frac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{Y}$ in terms of $\boldsymbol{\beta}$, ε_n, z_n, N, and constants. Note that $\boldsymbol{X}^\intercal\boldsymbol{X}$ is a 2×2 matrix and $\boldsymbol{X}^\intercal\boldsymbol{Y}$ is a 2–vector.

(4b)

Evaluate the following limits (which will denote convergence in probability):

$$\frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{X} \to ? \qquad \text{and} \qquad \frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{Y} \to ?$$

Note that your answers may depend on $\boldsymbol{\beta}$, but should not depend on $\boldsymbol{x}_n$ or y_n, since they are limiting quantites that do not depend on the particular dataset.

Justify your conclusion carefully (state which theorems you use).

(4c)

Using your answers from part (a) and (b), find the limit (convergence in probability) of

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\intercal \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y} \rightarrow ?$$

Note that your answer may depend on $\boldsymbol{\beta}$, but should not depend on $\boldsymbol{x}_n$ or y_n, since they are limiting quantites that do not depend on the particular dataset.

Justify your conclusion carefully (state which theorems you use).

5 Question 5 ◯ ← 'X' here to grade this question.

Let $\tilde{\boldsymbol{x}}_n$ denote an IID sequence of random 2–dimensional vectors in $\mathbb{R}^P$ (**not necessarily normal**), with

$$\tilde{\boldsymbol{x}_n} = \begin{pmatrix} \tilde{x}_{n1} \\ \tilde{x}_{n2} \end{pmatrix} \qquad \text{and} \qquad \mathbb{E}\left[\tilde{\boldsymbol{x}}_n\right] = \boldsymbol{0} \qquad \text{and} \qquad \mathrm{Cov}\left(\tilde{\boldsymbol{x}}_n\right) =: \boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}.$$

(5a)

Find the limiting distribution of the vector

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \tilde{\boldsymbol{x}}_n \to ?$$

Justify your conclusion carefully.

(5b)

Using the univariate central limit theorem, find the limiting distributions of the difference between the components of $\tilde{\boldsymbol{x}}_n$:

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tilde{x}_{n1} - \tilde{x}_{n2}) \rightarrow ?$$

Justify your conclusion carefully.

(5c)

Find a vector $\boldsymbol{v}$ such that $\boldsymbol{v}^\intercal \tilde{\boldsymbol{x}}_n = \tilde{\boldsymbol{x}}_{n1} - \tilde{\boldsymbol{x}}_{n2}$. Using this vector, show that the solution to (b) also follows from the solution to (a) and the continuous mapping theorem.

Justify your conclusion carefully.

6 Question 6 $\bigcirc \leftarrow$ **'X' here to grade this question.**

Given a regression on $\boldsymbol{X}$ with P regressors, and the corresponding $\boldsymbol{Y}$, $\hat{\boldsymbol{Y}}$, and $\hat{\varepsilon}$, define the following quantities:

$$
\begin{aligned}
RSS &:= \hat{\boldsymbol{\varepsilon}}^{\intercal}\hat{\boldsymbol{\varepsilon}} && \text{(Residual sum of squares)} \\
TSS &:= \boldsymbol{Y}^{\intercal}\boldsymbol{Y} && \text{(Total sum of squares)} \\
ESS &:= \hat{\boldsymbol{Y}}^{\intercal}\hat{\boldsymbol{Y}} && \text{(Explained sum of squares)} \\
R^2 &:= \frac{ESS}{TSS}.
\end{aligned}
$$

6a

1. Prove that $RSS + ESS = TSS$.
2. Express R^2 in terms of TSS and RSS.

6b

For each of the following questions, **please justify your answer**. An intuitive explanation is enough; a proof or specific example is not necessary.

1. What is R^2 when we include no regressors? ($P = 0$ and $\hat{y}_n = 0$ for all n)
2. What is R^2 when we include N linearly independent regressors? ($P = N$)
3. Can R^2 ever decrease when we add a regressor?
4. Can R^2 ever stay the same when we add a regressor?
5. Can R^2 ever increase when we add a regressor?

6c

For each of the following questions, **please justify your answer**. An intuitive explanation is enough; a proof or specific example is not necessary.

These questions will be about the F-test statistic for the null $H_0 : \boldsymbol{\beta} = \mathbf{0}$,

$$\phi = \frac{\hat{\beta}^\intercal (\boldsymbol{X}^\intercal \boldsymbol{X}) \hat{\beta}}{P \hat{\sigma}^2},$$

where $\hat{\sigma}^2 := \frac{1}{N-P} \sum_{n=1}^{N} \hat{\varepsilon}_n^2$.

1. Write the F-test statistic ϕ in terms of TSS and RSS, and P.
2. Can ϕ ever decrease when we add a regressor?
3. Can ϕ ever stay the same when we add a regressor?
4. Can ϕ ever increase when we add a regressor?

7 Question 7 ◯ ← 'X' here to grade this question.

For this question, we will take

$$a_n \sim \mathcal{N}(0,1) \quad \text{and} \quad b_n = a_n^3.$$

We assume the pairs (a_n, b_n) are IID, but a_n and b_n are not independent. Assume that, for some β_a and β_b,

$$y_n = \beta_a a_n + \beta_b b_n + \varepsilon_n,$$

where ε_n are IID with $\mathbb{E}[\varepsilon_n] = 0$ and $\mathrm{Var}(\varepsilon_n) < \infty$. The residuals ε_n and a_n are all independent of one another. **Note that the residuals are not necessarily normal.**

(7a)

Let $\hat{\alpha}$ denote the OLS estimator of $y_n \sim \alpha a_n$, that is, of y_n regressed on a_n alone. Note that the regression for $\hat{\alpha}$ does not include a constant, and does not include b_n.

Recall that

$$\hat{\alpha} = \frac{\sum_{n=1}^{N} y_n a_n}{\sum_{n=1}^{N} a_n^2},$$

and find the limit

$$\hat{\alpha} \rightarrow ?$$

as $N \rightarrow \infty$.

The answer may depend on the unknown β_a and β_b.

Hint: Standard properties of the normal gives that $\mathbb{E}\left[a_n^3\right] = 0$ and $\mathbb{E}\left[a_n^4\right] = 3$.

(7b)

Letting $\hat{y}_{\mathrm{new}} = \hat{\alpha} a_{\mathrm{new}}$, find

$$\mathbb{E}\left[y_{\mathrm{new}} - \hat{y}_{\mathrm{new}} | a_{\mathrm{new}}, \boldsymbol{Y}, \boldsymbol{A}\right],$$

where $\boldsymbol{A} = (a_1, \ldots, a_N)^\intercal$ is the vector of a_n observations and $\boldsymbol{Y} = (y_1, \ldots, y_N)^\intercal$ is the vector of responses in the training set. Note that the expectation is conditional on the training data and on the new regressor, so the only randomness is in $\varepsilon_{\mathrm{new}}$.

The answer may depend on the unknown β_a and β_b.

(7c)

Assume that $\beta_b \neq 0$, that N is very large.

- What does your result from (a) imply about using $\hat{\alpha}$ for inference on β_a?
- What does your result from (b) imply about using $\hat{\alpha}$ for prediction of y_{new}?