

# STAT151A Quiz 5 (Apr 23rd)

Please write your full name and email address:

The OLS estimator is given by  $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ .

**You have 20 minutes for this quiz.**

**There are three parts, (a), (b), and (c), each weighted equally..**

This quiz will use some facts about the standard normal distribution. If  $z \sim \mathcal{N}(0, 1)$ , then:

- $\mathbb{E}[z] = 0$
- $\mathbb{E}[z^2] = 1$
- $\mathbb{E}[z^3] = 0$
- $\mathbb{E}[z^4] = 3$

Recall that the OLS estimator of  $y_n \sim \beta^\top \mathbf{x}_n$  is  $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ .

For this quiz, we will take

$$a_n \sim \mathcal{N}(0, 1) \text{ IID} \quad \text{and} \quad b_n = a_n^2.$$

Note that the pairs  $(a_n, b_n)$  are IID, but  $a_n$  and  $b_n$  are not independent.

Let  $\mathbf{x}_n = (a_n, b_n)^\top$ . Let  $\boldsymbol{\beta} = (\beta_a, \beta_b)^\top$ . Let  $y_n = \boldsymbol{\beta}^\top \mathbf{x}_n + \varepsilon_n$ , where  $\varepsilon_n$  are IID with  $\mathbb{E}[\varepsilon_n] = 0$  and  $\text{Var}(\varepsilon_n) < \infty$ . The residuals  $\varepsilon_n$  and  $a_n$  are all independent of one another.

Note that  $a_n$ ,  $b_n$ ,  $\beta_a$ , and  $\beta_b$  are all scalars, and both  $\mathbf{x}_n$  and  $\boldsymbol{\beta}$  are 2-vectors.

Let  $\hat{\boldsymbol{\beta}}$  denote the OLS estimator of  $y_n \sim \boldsymbol{\beta}^\top \mathbf{x}_n$ , that is, of  $y_n$  regressed on both  $a_n$  and  $b_n$ . **Note that the regression for  $\hat{\boldsymbol{\beta}}$  does not include a constant.**

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  $b_n$ .**

---

**(a)**

Prove that, as  $N \rightarrow \infty$ ,

- $\hat{\alpha} \rightarrow \beta_a$  and
- $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta} = \begin{pmatrix} \beta_a \\ \beta_b \end{pmatrix}$ .

**(b)**

For part (b), let  $\hat{y}_{\text{new}} = \beta_a a_{\text{new}}$ , and assume that  $\beta_b \neq 0$ . **Note that  $\hat{y}_{\text{new}}$  is formed with  $\beta_a$ , not  $\hat{\alpha}$ .** Prove that

$$\mathbb{E}[y_{\text{new}} - \hat{y}_{\text{new}}] \neq 0,$$

where the expectation is taken over  $a_{\text{new}}$ ,  $b_{\text{new}}$ , and  $\varepsilon_{\text{new}}$ . That is, when you exclude  $b_n$  from the regression, the predictions evaluated at the limit  $\beta_a$  are biased.

**(c)**

For part (c), let  $\hat{y}'_{\text{new}} = \beta^\top \mathbf{x}_{\text{new}}$ . **Note that  $\hat{y}'_{\text{new}}$  is formed with  $\beta$ , not  $\hat{\beta}$ .** Prove that

$$\mathbb{E} [y_{\text{new}} - \hat{y}'_{\text{new}}] = 0,$$

where the expectation is taken over  $a_{\text{new}}$ ,  $b_{\text{new}}$ , and  $\varepsilon_{\text{new}}$ . That is, when you include  $b_n$  from the regression, the predictions evaluated at the limit  $\beta$  are unbiased.