STAT151A Quiz 3

Please write your full name and email address here:

Also, please put your intials on each page in case the pages get separated.

You have 30 minutes for this quiz.

There are three questions, each weighted equally..

There are extra pages at the end if you need more space for solutions.

Question 1

Let y_n, x_n, and z_n all denote continuous, scalar–valued variables for $n = 1, \ldots, N$.

Define

$$a_n = \frac{x_n + z_n}{2} \quad \text{and} \quad d_n = \frac{x_n - z_n}{2}.$$

Let

- $\hat{\boldsymbol{Y}}_{1ad}$ denote the fit from the regression $y_n \sim \gamma_0 + \gamma_a a_n + \gamma_b d_n$, and
- $\hat{\boldsymbol{Y}}_{1xz}$ denote the fit from the regression $y_n \sim \beta_0 + \beta_x x_n + \beta_z z_n$.

Please say whether each of the following statements is true, false, or cannot be determined from the information given.

(a) $\left\| \hat{\boldsymbol{Y}}_{1ad} \right\|^2 = \left\| \hat{\boldsymbol{Y}}_{1xz} \right\|^2$.

(b) $\hat{\beta}_0 + \hat{\beta}_x x_n + \hat{\beta}_z z_n = \hat{\gamma}_0 + \hat{\gamma}_a a_n + \hat{\gamma}_b d_n$ for all n.

(c) $\left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}}_{1ad} \right\|^2 < \left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}}_{1xz} \right\|^2$.

(d) All the answers above follow from a single linear algebra fact about the relationship between $(1, x_n, z_n)^\intercal$ and $(1, a_n, d_n)^\intercal$. Briefly state this fact. A single sentence is enough.

Question 2

Recall the definitions:

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

Consider two regressions, $\boldsymbol{Y} \sim \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{Y} \sim \boldsymbol{X}\boldsymbol{\gamma}_z + \boldsymbol{Z}\boldsymbol{\gamma}_z$. That is, the first regression is on $\boldsymbol{X}$ alone and the second regression is on both $\boldsymbol{X}$ and $\boldsymbol{Z}$. You may assume that all the columns of $\boldsymbol{X}$ and $\boldsymbol{Z}$ are mutually linearly independent so that $(\boldsymbol{X}, \boldsymbol{Z})$, the matrix formed by taking the columns of $\boldsymbol{X}$ concatenated horizontally with the columns of $\boldsymbol{Z}$, is full column rank.

Let RSS_X, TSS_X, ESS_X, and R^2_X denote the correponding quantities from $\boldsymbol{Y} \sim \boldsymbol{X}$, and RSS_{XZ}, TSS_{XZ}, ESS_{XZ}, and R^2_{XZ} the corresponding quantities from the regression $\boldsymbol{Y} \sim \boldsymbol{X} + \boldsymbol{Z}$. For example,

$$ESS_X = (\boldsymbol{X}\hat{\boldsymbol{\beta}})^\intercal(\boldsymbol{X}\hat{\boldsymbol{\beta}}) \quad \text{and} \quad ESS_{XZ} = (\boldsymbol{X}\hat{\boldsymbol{\gamma}}_X + \boldsymbol{Z}\hat{\boldsymbol{\gamma}}_Z)^\intercal(\boldsymbol{X}\hat{\boldsymbol{\gamma}}_X + \boldsymbol{Z}\hat{\boldsymbol{\gamma}}_Z),$$

and so on.

Please say whether each of the following statements, true, false, or cannot be determined from the information given.

(a) $RSS_X > RSS_{XZ}$

(b) $ESS_X > ESS_{XZ}$

(c) $R^2_X > R^2_{XZ}$

(d) All the answers above follow from a single linear algebra fact about the column spans of $\boldsymbol{X}$ and $(\boldsymbol{X}, \boldsymbol{Z})$. Briefly state this fact. A single sentence is enough.

Question 3

Suppose you have N observations of the following data about an aircraft:

$$\begin{aligned} y_n &= \text{Drag force (strictly positive)} \\ x_n &= \text{Wind speed (strictly positive)} \\ z_n &= \text{Material type } \in \{0, 1\}. \end{aligned}$$

Here, $z_n = 1$ means that a new experimental material was used, and $z_n = 0$ means the new material was not used. The regressor z_n is either 0 or 1.

You want to test the effect of the experimental material on the drag force taking into account variability in wind speed. Basic physics suggests that the drag force is proportional to x_n^2.

(a)

Suppose you want to estimate the quantity $\boldsymbol{\beta} = (\beta_1, \beta_z, \beta_x)^\intercal$ in the regression

$$y_n = \beta_1 \beta_z^{z_n} x_n^{\beta_x} \exp(\varepsilon_n),$$

where ε_n accounts for the error between y_n and $\beta_1 \beta_z^{z_n} x_n^{\beta_x}$. How can you transform the data in order to write the problem of estimating $\boldsymbol{\beta}$ as a linear regression problem?

(b)

Suppose you define $\ell_n := \log y_n$ run the regression

$$\ell_n \sim \gamma_1 + \gamma_z z_n + \gamma_x \log x_n \quad \text{so that} \quad \hat{\ell}_n = \hat{\gamma}_1 + \hat{\gamma}_z z_n + \hat{\gamma}_x \log x_n.$$

Define $f(x_n, z_n) = \exp(\hat{\ell}_n)$, and write an expression the value of $f(x_n, 1)/f(x_n, 0)$. You may give your answer in terms of the entries of $\boldsymbol{\gamma} = (\gamma_1, \gamma_z, \gamma_x)^\intercal$, x_n, and z_n (though you may not need to use all of them).

(c)

Suppose we assume that

$$f(x_n, z_n) = \tilde{\beta}_1 \tilde{\beta}_z^{z_n} x_n^{\tilde{\beta}_x}$$

holds for all z_n and x_n. Using this, write $\tilde{\beta}_1$, $\tilde{\beta}_x$, and $\tilde{\beta}_z$. in terms of $\hat{\gamma}_1$, $\hat{\gamma}_z$, and $\hat{\gamma}_x$.

(d)

Suppose we estimate that $f(x_n, 1)/f(x_n, 0) < 1$. If we take our estimate to be roughly correct, what does this imply for the effect of our experimental material on the drag?

Extra space for answers (indicate clearly which problem you are working on)

Extra space for answers (indicate clearly which problem you are working on)

Extra space for answers (indicate clearly which problem you are working on)

Extra space for answers (indicate clearly which problem you are working on)