STAT151A Homework 2: Due February 9th

Your name here

1 Transformation of variables

Consider two different regressions, $\boldsymbol{Y} \sim \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{Y} \sim \boldsymbol{Z}\boldsymbol{\alpha}$, with the same $\boldsymbol{Y}$, where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are both $N \times P$ and are both full-rank. Let the n–th row of $\boldsymbol{X}$ be written $\boldsymbol{x}_n^\intercal$, and the n–th row of $\boldsymbol{Z}$ be $\boldsymbol{z}_n^\intercal$.

(a)

Suppose $\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{z}_n$ for an invertible $\boldsymbol{A}$ and for all $n = 1, \ldots, N$. Find an expression for $\hat{\alpha}$ in terms of $\hat{\beta}$ that does not explicitly use $\boldsymbol{Y}$, $\boldsymbol{X}$, or $\boldsymbol{Z}$.

Solution:

This means $\boldsymbol{x}_n^\intercal = \boldsymbol{z}_n^\intercal \boldsymbol{A}^\intercal$ so $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{A}^\intercal$. Therefore the fits are the same when $\boldsymbol{X}\beta = \boldsymbol{Z}\alpha = \boldsymbol{X}\boldsymbol{A}^\intercal\alpha$, which are the same when $\boldsymbol{A}^\intercal\alpha = \beta$. Therefore $\hat{\alpha} = (\boldsymbol{A}^\intercal)^{-1}\hat{\beta}$.

(b)

Suppose that, for all $n = 1, \ldots, N$, $\boldsymbol{x}_n = f(\boldsymbol{z}_n)$ for some invertible but non-linear function $f(\cdot)$. In general, can you find an expression for $\hat{\alpha}$ in terms of $\hat{\beta}$ that does not explicitly use $\boldsymbol{Y}$, $\boldsymbol{X}$, or $\boldsymbol{Z}$? Prove why or why not. (To prove that you cannot, finding a single counterexample is enough.)

Solution:

You cannot. It's easiest to show this with a counterexample, of which there are many.

(c)

Now consider only the regression $\boldsymbol{Y} \sim \boldsymbol{X}\boldsymbol{\beta}$, but suppose we are not interested in $\boldsymbol{\beta}$, but rather some other $\boldsymbol{\gamma} = \phi(\boldsymbol{\beta})$, where ϕ is an invertible function. Prove that the least squares estimator of $\boldsymbol{\gamma}$ is given by $\hat{\boldsymbol{\gamma}} = \phi(\hat{\boldsymbol{\beta}})$.

Solution:

The least squares estimate of $\boldsymbol{\gamma}$ is $\underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \frac{1}{N}\sum_{n=1}^{N}(y_n - \boldsymbol{x}_n^\intercal \phi^{-1}(\boldsymbol{\gamma}))$. The minimizer cannot be other than $\phi(\hat{\beta})$, for if it were, then taking $\beta = \phi^{-1}(\hat{\gamma})$ would provide a better fit than $\hat{\beta}$, which is a contradiction.

(d)

Prove that result (a) is special case of the result (c). (Hint: find the corresponding ϕ.)

Solution:

Take $\phi(\beta) = (A^\intercal)^{-1}\alpha$. The point is that when the data transform is linear, then it is equivalent to a parameter transpose by the associative property of matrix multiplication.

2 Spaces of possible estimators.

Consider the simple linear model $y_n = \beta_0 + \beta_1 z_n + \varepsilon_n$. Assume that $\frac{1}{N}\sum_{n=1}^{N} z_n \neq 0$.

(a)

Fix $\beta_0 = \frac{1}{N}\sum_{n=1}^{N} y_n$ and find a value of β_1 such that $\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n = 0$. How does your answer depend on whether or not $\frac{1}{N}\sum_{n=1}^{N} z_n = 0$?

Solution:

$\beta_1 = 0$ is always a solution. If $\frac{1}{N}\sum_{n=1}^{N} z_n = 0$ then any β_1 is a solution.

(b)

Fix $\beta_1 = 10,000,000$ and find a value of β_0 such that $\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n = 0$.

Solution:

$$\frac{1}{N}\sum_{n=1}^{N} y_n = \beta_0 + \beta_1 \frac{1}{N}\sum_{n=1}^{N} z_n + \frac{1}{N}\sum_{n=1}^{N}\varepsilon_n$$

so for any β_1, including the one given, we can take

$$\beta_0 = \frac{1}{N}\sum_{n=1}^{N} y_n - \beta_1 \frac{1}{N}\sum_{n=1}^{N} z_n.$$

(c)

In general, how many different choices of β_0 and β_1 can you find that satisfy $\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n = 0$? Are all of them reasonable? Are any of them reasonable?

Solution:

As seen above, setting $\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n = 0$ provides a single equation to identify two unknowns, and so there are an infinite number of solutions. The family is one-dimensional unless $\frac{1}{N}\sum_{n=1}^{N} z_n = 0$, in which case it is two-dimensional. Many are unreasonable, some are reasonable, since the OLS solution is included.

(d)

Find an N–dimensional vector $\boldsymbol{v}$ such that

$$\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n = 0 \quad \Leftrightarrow \quad \boldsymbol{v}^\intercal \boldsymbol{\varepsilon} = \mathbf{0}.$$

Solution:

$\boldsymbol{v} = \mathbf{1}$, or any proportional vector. (These are vectors; bold isn't rendering in Quarto latex for me right now)

(e)

Suppose I give you a general N–dimensional vector $\boldsymbol{v}$ and a scalar a. How many different choices of β_0 and β_1 can you find such that $\boldsymbol{v}^\intercal \boldsymbol{\varepsilon} = a$?

Solution:

Again, this gives one equation to identify two unknowns, so there are an infinite number.

(f) (Optional — this will not be graded)

Suppose I give you two different vectors, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Under what circumstances can you find β_0 and β_1 such that

$$\boldsymbol{v}_1^\intercal \boldsymbol{\varepsilon} = \mathbf{0} \quad \text{and} \quad \boldsymbol{v}_2^\intercal \boldsymbol{\varepsilon} = \mathbf{0}?$$

When are there infinitely many solutions? When is there only one solution? (Hint: what if $\boldsymbol{v}_1^\intercal \mathbf{1} = \boldsymbol{v}_2^\intercal \mathbf{1} = 0$?)

(g)

Now, consider the general linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$. Prove that there always exists $\boldsymbol{\beta}$ and ε so that the $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$.

Solution:

Take $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$ for any $\boldsymbol{\beta}$.

(h) (Optional — this will not be graded)

Suppose, for the general linear model, that the matrix $\boldsymbol{X}$ is full-rank (that is, of rank P, where P is the number of columns of $\boldsymbol{X}$). Suppose I give you a $N \times D$ matrix $\boldsymbol{V}$, and ask you to find $\boldsymbol{\beta}$ such that $\boldsymbol{V}^\intercal \boldsymbol{\varepsilon} = \mathbf{0}$. Under what circumstances are there no solutions? A single solution? An infinite set of solutions? (Hint: you already answered this question for $P = 2$, now you just need to state the result in matrix form.)

3 Collinear regressors

Suppose that $\boldsymbol{X}$ does not have full column rank — that is, $\boldsymbol{X}$ is $N \times P$ but has column rank $Q < P$.

(a)

How many solutions $\hat{\beta}$ are there to the least-squares problem

$$\hat{\beta} := \underset{\beta}{\operatorname{argmin}} \, \|\boldsymbol{Y} - \boldsymbol{X}\beta\|_2^2?$$

Solution:

Our first-order condition gives than the sum of squares is minimized when $\boldsymbol{X}^\intercal \boldsymbol{X}\beta = \boldsymbol{X}^\intercal \boldsymbol{Y}$. Any β in the nullspace of $\boldsymbol{X}$ leaves this identifying condition invariant. So there is an $P - Q$ dimensional family of solutions.

(b)

Relate the solutions $\hat{\beta}$ from part (a) to spaces spanned by eigenvectors of $\boldsymbol{X}^\intercal \boldsymbol{X}$. Among the solutions, identify the one with the smallest norm, $\left\|\hat{\beta}\right\|_2^2$.

Solution:

The component of $\hat{\beta}$ is determined in the span of eigenvectors with nonzero eigenvalues, and free to vary in the span of zero eigenvectors (the nullspace). The minimum norm solution sets the component in the nullspace to zero by the triangle inequality.

(c)

Suppose that $\boldsymbol{X}'$ is a full column-rank $N \times Q$ matrix with the same column span as $\boldsymbol{X}$, and let $\hat{\gamma}$ be the OLS estimator for the regression $\boldsymbol{Y} \sim \boldsymbol{X}'\gamma$. Compare the fits $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta}$ and $\hat{\boldsymbol{Y}}' = \boldsymbol{X}'\hat{\gamma}$, and compare the sum of squared residuals for the two regressions.

Solution:

The fits and sum of squared residuals are the same.