STAT151A Homework 1: Due Jan 26th

Your name here

1 Simple regression in matrix form

Consider the simple linear model $y_n = \beta_0 + \beta_1 z_n + \varepsilon_n$.

Let $\bar{y} := \frac{1}{N} \sum_{n=1}^{N} y_n$ and $\bar{z} := \frac{1}{N} \sum_{n=1}^{N} z_n$. Recall that the ordinary least squares estimates are given by

$$\hat{\beta}_1 = \frac{\frac{1}{N} \sum_{n=1}^{N} (z_n - \bar{z})(y - \bar{y})}{\frac{1}{N} \sum_{n=1}^{N} (z_n - \bar{z})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{z}.$$

(a)

Write the set of equations

$$y_n = \beta_0 + \beta_1 z_n + \varepsilon_n$$

for $n \in \{1, \ldots, N\}$ in matrix form. That is, let $\boldsymbol{X}$ denote an $N \times 2$ matrix, $\boldsymbol{Y}$ and $\boldsymbol{\varepsilon}$ denote $N \times 1$ matrices, $\_ = (\beta_0, \beta_1)^{\mathsf{T}}$, and express the matrices $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{\varepsilon}$ in terms of the scalars y_n, z_m, and ε_n so that $\boldsymbol{Y} = \boldsymbol{X}\_ + \boldsymbol{\varepsilon}$ is equivalent to the set of regression equations.

Solution:

$$\boldsymbol{X} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_N \end{pmatrix} \qquad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

(b)

Define

$$\overline{zz} := \frac{1}{N} \sum_{n=1}^{N} z_n^2 \quad \text{and} \quad \overline{zy} := \frac{1}{N} \sum_{n=1}^{N} z_n y_n$$

Write an explict expressions for $\boldsymbol{X}^\intercal \boldsymbol{X}$, $\boldsymbol{X}^\intercal \boldsymbol{Y}$, and $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$, all in terms of $\bar{y}$, $\bar{z}$, $\overline{zz}$, $\overline{zy}$, and N. Verify that the inverse is correct by direct multiplication.

Solution:

$$\boldsymbol{X}^\intercal \boldsymbol{X} = N \begin{pmatrix} 1 & \bar{z} \\ \bar{z} & \overline{zz} \end{pmatrix} \qquad \boldsymbol{X}^\intercal \boldsymbol{Y} = N \begin{pmatrix} \bar{y} \\ \overline{zy} \end{pmatrix} \qquad (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} = \frac{1}{N} \frac{1}{\overline{zz} - \bar{z}^2} \begin{pmatrix} \overline{zz} & -\bar{z} \\ -\bar{z} & 1 \end{pmatrix}$$

(c)

Compute $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$. Show that the first row is equal to $\hat{\beta}_0$ and the second row is equal to $\hat{\beta}_1$ as given by the ordinary least squares formula in the problem statement above.

Solution: Just do matrix multiplication correctly. The calculation is also done in Lecture 3.

2 Mean zero residuals.

Consider the model $y_n = \beta z_n + \varepsilon_n$. Let $\hat{\beta}$ denote the least squares estimator and $\hat{\varepsilon}_n = y_n - \hat{\beta} z_n$.

(a)

Suppose z_n is not a constant. Is it necessarily the case that $\frac{1}{N} \sum_{n=1}^{N} \hat{\varepsilon}_n = 0$? Prove your answer.

Solution: It is not. A counterexample is enough. For example, take $y_n = 1$, N even and $z_n = (-1)^n$. Then $\hat{\beta} = \sum_{n=1}^{N} y_n z_n / \sum_{n=1}^{N} z_n^2 = 0$, so $\hat{\varepsilon}_n = 1$.

(b)

Suppose z_n is a constant, but $z_n \equiv 5$ for every $n \in \{1, \ldots, N\}$. Is it necessarily the case that $\frac{1}{N} \sum_{n=1}^{N} \hat{\varepsilon}_n = 0$? Prove your answer.

Solution: Yes. You can either do the calculation directly, or you can use the fact that $\boldsymbol{\varepsilon} \perp \boldsymbol{z}$.

2

(c)

Now the model $y_n = \beta_1 z_{n1} + \beta_2 z_{n2} + \varepsilon_n$. Suppose that $z_{n1} = 1$ is n is even, and is 0 otherwise. Similarly, suppose that $z_{n2} = 1$ is n is odd, and is 0 otherwise. Let N be even. Is it necessarily the case that $\frac{1}{N} \sum_{n=1}^{N} \hat{\varepsilon}_n = 0$? Prove your answer.

Solution: It is the case. The ones vector $\mathbf{1}$ is in the linear span of $\boldsymbol{z}$ since $z_{n1} + z_{n2} = 1$, so $\varepsilon \perp \mathbf{1}$.

3 Inner products and covariances

Let $\boldsymbol{z} = (z_1, \ldots, z_N)$ and $\boldsymbol{y} = (y_1, \ldots, y_N)$. Let $\boldsymbol{X}$ denote an $N \times P$ matrix whose n–th row is the transpose of the P-vector $\boldsymbol{x}_n^\mathsf{T}$.

(Note: this question involves limits of random variables, and there are many distinct ways that random variables can converge to limits. If you're familiar with these different modes of probabilisitic convergence, feel free to state what mode of convergence applies, but if you are not, don't worry — modes of convergence will not matter much for this class, and you can state your result heuristically.)

For a set of quantities (numbers, vectors, pairs of vectors, etc), the "empirical distribution" over that set refers to drawing an element with replacement from the set with equal probability given to each entry. For example, if $\mathcal{Z}'$ is a drawn from the empirical distribution over the set $\{z_1, \ldots, z_N\}$, then $\mathbb{P}(\mathcal{Z}' = z_n) = 1/N$ for each n. Similarly, if $(\mathcal{Z}', \mathcal{Y}')$ is drawn from the empirical distribution over the pairs $\{(z_1, y_1), \ldots, (z_N, y_N)\}$, then $\mathbb{P}((\mathcal{Z}', \mathcal{Y}') = (z_n, y_n)) = 1/N$ for all n.

(Hint: it may help to recall that the bootstrap uses draws from the empirical distribution, and that, in the empirical distribution, the elements of the set are fixed and not random.)

(a)

Let $(\mathcal{Z}', \mathcal{Y}')$ denote a draw from the empirical distribution over the set $\{(y_1, z_1), \ldots, (y_N, z_N)\}$.

Prove that $\frac{1}{N} \boldsymbol{z}^\mathsf{T} \boldsymbol{y} = \mathbb{E}[\mathcal{Z}' \mathcal{Y}']$. Then prove that $\frac{1}{N} \mathbf{1}^\mathsf{T} \boldsymbol{z} = \mathbb{E}[\mathcal{Z}']$ as a special case.

Solution: Recognize the inner product as a sum. Then both follow from the definition of the expectation with respect to the empirical distribution.

(b)

Now suppose that the entries of $\boldsymbol{z}$ are independent and identically distributed (IID) realizations of the random variable $\mathcal{Z}$, and that the entries of $\boldsymbol{y}$ are similarly IID realizations of a random variable $\mathcal{Y}$. Assuming that $\mathbb{E}\left[|\mathcal{Z}|\right] < \infty$ and $\mathbb{E}\left[|\mathcal{Y}|\right] < \infty$, prove that

$$\frac{1}{N}\boldsymbol{z}^\mathsf{T}\boldsymbol{y} \to \mathbb{E}\left[\mathcal{Z}\mathcal{Y}\right] \text{ as } N \to \infty$$

(Hint: don't prove this from scratch, appeal to a probability theorem.)

Solution: Recognize the inner product as a sum, then appeal to the law of large numbers (weak or strong is okay).

(c)

Using only inner products involving $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{1}$, write an expression for $\mathrm{Cov}\left(\mathcal{Y}', \mathcal{Z}'\right)$. Prove that the expression converges with probability one to $\mathrm{Cov}\left(\mathcal{Y}, \mathcal{Z}\right)$. (Hint: again, use your previous results and a theorem from probability.)

Solution: Using our previous results,

$$
\begin{aligned}
\mathrm{Cov}\left(\mathcal{Y}', \mathcal{Z}'\right) &= \mathbb{E}\left[(\mathcal{Y}' - \mathbb{E}\left[\mathcal{Y}'\right])(\mathcal{Z}' - \mathbb{E}\left[\mathcal{Z}'\right])\right] \\
&= \mathbb{E}\left[(\mathcal{Y}' - N^{-1}\boldsymbol{1}^\mathsf{T}\boldsymbol{y})(\mathcal{Z}' - N^{-1}\boldsymbol{1}^\mathsf{T}\boldsymbol{z})\right] \\
&= N^{-1}(\boldsymbol{y} - N^{-1}\boldsymbol{1}^\mathsf{T}\boldsymbol{y})^\mathsf{T}(\boldsymbol{z} - N^{-1}\boldsymbol{1}^\mathsf{T}\boldsymbol{z}).
\end{aligned}
$$

(d)

Now, let $(\mathcal{X}', \mathcal{Y}')$ denote a draw from the empirical distribution over $\{(x_1, y_1), \ldots, (x_N, y_N)\}$. (Recall that the vector x_n is a length–P column vector, and x_n^T is the n–th row of the matrix $\boldsymbol{X}$.)

$$\frac{1}{N}\boldsymbol{X}^\mathsf{T}\boldsymbol{X} = \mathbb{E}\left[\mathcal{X}'\mathcal{X}'^\mathsf{T}\right] \quad \text{and} \quad \frac{1}{N}\boldsymbol{X}^\mathsf{T}y = \mathbb{E}\left[\mathcal{X}'\mathcal{Y}'\right].$$

Solution: Apply the previous results componentwise.

(e)

Now, suppose that rows of $\boldsymbol{X}$ are IID realizations of the random P–vector $\mathcal{X}$, and that $\mathbb{E}\left[\|\mathcal{X}_p\|\right] < \infty$ for each $p \in \{1, \ldots, P\}$. Assume, as above, that $\mathbb{E}\left[\|\mathcal{Y}\|\right] < \infty$.

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

$$\frac{1}{N}\boldsymbol{X}^\mathsf{T}\boldsymbol{X} \to \mathbb{E}\left[\mathcal{X}\mathcal{X}^\mathsf{T}\right] \quad \text{and} \quad \frac{1}{N}\boldsymbol{X}^\mathsf{T}\boldsymbol{Y} \to \mathbb{E}\left[\mathcal{X}\mathcal{Y}\right],$$

where both limits are with probability one.

Solution: Apply the previous results componentwise.

(f)

Now assume that, for each $p \in \{1, \ldots, P\}$ and $q \in \{1, \ldots, P\}$, $\mathbb{E}\left[\left|\mathcal{X}'_p\right|\left|\mathcal{X}'_q\right|\mathcal{Y}^2\right] < \infty$. Prove that, as $N \to \infty$,

$$\frac{1}{\sqrt{N}}\left(\boldsymbol{X}^\mathsf{T}\boldsymbol{Y} - \mathbb{E}\left[\boldsymbol{X}^\mathsf{T}\boldsymbol{Y}\right]\right) \to \mathcal{Z},$$

where $\mathcal{Z}$ is a multivariate normal random variable. What is the covariance of $\mathcal{Z}$? (Hint: again, appeal to a probability theorem.)

Solution: Appeal to the multivariate central limit theorem.