Daniel Cullen
Econ 140A
dancullen.me/teaching/140A

Final Review Spring 2019

1. Suppose that you wish to estimate a regression in deviations-from-means form. The deviations-from-means model is given by $y_i = \beta x_i + \varepsilon_i$ and all standard assumptions hold (i.e, x_i is nonrandom, $\mathbb{E}(\varepsilon_i) = 0$, $var(\varepsilon_i) = \sigma^2$ for all i, and $cov(\varepsilon_i, \varepsilon_j) = 0$ for all observations).

   (a) Derive the Ordinary Least Squares (OLS) estimator $\hat{\beta}^{OLS}$. Show all work and simplify completely to receive full credit.

$$\widehat{\beta}^{OLS} \equiv \underset{\widehat{\beta}}{\text{argmin}} \sum_{i=1}^{n} \left(y_i - \widehat{\beta} x_i\right)^2 \qquad \text{(recalling that } e_i = y_i - \widehat{\beta} x_i\text{)}$$

Differentiate with respect to $\widehat{\beta}$, set it equal to zero, and solve for $\widehat{\beta}$.

$$\text{F.O.C.:} \qquad \sum_{i=1}^{n} -2(y_i - \widehat{\beta} x_i) x_i = 0$$

$$2\sum_{i=1}^{n} y_i x_i - 2\widehat{\beta} \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i = \widehat{\beta} \sum_{i=1}^{n} x_i^2$$

$$\boxed{\widehat{\beta}^{OLS} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}}$$

(b) Determine if $\hat{\beta}^{OLS}$ is an unbiased estimator of β. Make sure to be clear about when you use an assumption, show all work, and simplify completely to receive full credit.

$$\begin{aligned}
&\underline{A1}: \quad && y_i = \beta x_i + \varepsilon_i \text{ is the true DGP} \\
&\underline{A2}: \quad && x_i \text{ is nonrandom} \\
&\underline{A3}: \quad && \mathbb{E}[\varepsilon_i] = 0 \ \ \forall i \\
&\underline{A4}: \quad && Var(\varepsilon_i) = \sigma^2 \ \ i = 1, \dots, n \\
& && Cov(\varepsilon_i, \varepsilon_j) = 0 \ \ \forall i \neq j
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}^{OLS}\right] &= \mathbb{E}\left[\frac{\sum_i y_i x_i}{\sum_i x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\sum_i (\beta x_i + \varepsilon_i) x_i}{\sum_i x_i^2}\right] && \text{(A1)} \\
&= \mathbb{E}\left[\frac{\sum_i (\beta x_i^2 + \varepsilon_i x_i)}{\sum_i x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\sum_i \beta x_i^2}{\sum_i x_i^2} + \frac{\sum_i \varepsilon_i x_i}{\sum_i x_i^2}\right] \\
&= \mathbb{E}\left[\beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}\right] \\
&= \beta + \mathbb{E}\left[\frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}\right] \\
&= \beta + \frac{\sum_i x_i \mathbb{E}[\varepsilon_i]}{\sum_i x_i^2} && \text{(A2)} \\
&\boxed{\mathbb{E}\left[\widehat{\beta}^{OLS}\right] = \beta} && \text{(A3)}
\end{aligned}$$

$$\implies \widehat{\beta}^{OLS} \text{ is an unbiased estimator of } \beta$$

(c) Find the variance of $\hat{\beta}^{OLS}$. Make sure to be clear about when you use an assumption, show all work and simplify completely to receive full credit.

$$
\begin{aligned}
Var\left(\widehat{\beta}^{OLS}\right) &= Var\left(\frac{\sum\limits_i x_i y_i}{\sum\limits_i x_i^2}\right) \\
&= Var\left(\frac{\sum\limits_i x_i(\beta x_i + \varepsilon_i)}{\sum\limits_i x_i^2}\right) && \text{(A1)} \\
&= Var\left(\beta + \frac{\sum\limits_i x_i \varepsilon_i}{\sum\limits_i x_i^2}\right) \\
&= Var\left(\frac{\sum\limits_i x_i \varepsilon_i}{\sum\limits_i x_i^2}\right) && \text{(the variance of a constant is zero)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} Var\left(\sum_i x_i \varepsilon_i\right) && \text{(pull constant out and square)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} \left[\sum_i Var(x_i \varepsilon_i) + \sum_{i \neq j} Cov(x_i \varepsilon_i, x_j \varepsilon_j)\right] && \text{(property of variance)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} \left[\sum_i x_i^2 Var(\varepsilon_i) + \sum_{i \neq j} x_i x_j Cov(\varepsilon_i, \varepsilon_j)\right] && \text{(A2)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} \left[\sigma^2 \sum_i x_i^2\right] && \text{(A4)}
\end{aligned}
$$

$$
\boxed{Var\left(\widehat{\beta}^{OLS}\right) = \frac{\sigma^2}{\sum\limits_i x_i^2}}
$$

2. Consider the following assumptions:

$$
\begin{aligned}
A1 :& y_i = 5(\mu + \varepsilon_i) \\
A2 :& \mathbb{E}[\varepsilon_i] = 0 \text{ for all } i \\
A3 :& Var(\varepsilon_i) = \sigma^2 \text{ for all } i \\
A4 :& Cov(\varepsilon_i, \varepsilon_j) = 0 \text{ for } i \neq j \\
A5 :& \varepsilon_i \sim \text{ Normal}
\end{aligned}
$$

Suppose you are interested in generating an estimate for μ.

(a) What is the expected value of the sample mean estimator, $\hat{\mu} = \frac{1}{n}\sum y_i$, under these assumptions? Is $\hat{\mu}$ an unbiased estimator for μ? Show all work.

$$
\begin{aligned}
\mathbb{E}[\hat{\mu}] &= \mathbb{E}\left[\frac{1}{n}\sum y_i\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum 5(\mu + \varepsilon_i)\right] && \text{(A1)} \\
&= \mathbb{E}\left[\frac{1}{n}\sum 5\mu + \frac{1}{n}\sum 5\varepsilon_i\right] \\
&= \mathbb{E}\left[\frac{5}{n}\sum \mu\right] + \mathbb{E}\left[\frac{5}{n}\sum \varepsilon_i\right] \\
&= \frac{5}{n}\sum \mathbb{E}[\mu] + \frac{5}{n}\sum \cancelto{0}{\mathbb{E}[\varepsilon_i]} && \text{(A2)} \\
&= \frac{5}{n}\sum \mu \\
&= \frac{5}{n}n\mu \\
&= 5\mu
\end{aligned}
$$

The estimator is biased.

(b) Derive the variance for the sample mean under these assumptions. Show all work.

$$
\begin{aligned}
Var(\hat{\mu}) &= Var\left(\frac{1}{n}\sum y_i\right) \\
&= Var\left(\frac{1}{n}\sum 5(\mu + \varepsilon_i)\right) && \text{(A1)} \\
&= Var\left(\frac{1}{n}\sum 5\mu + \frac{1}{n}\sum 5\varepsilon_i\right) \\
&= Var\left(\frac{5}{n}\sum \varepsilon_i\right) \\
&= \frac{25}{n^2}\sum Var\left(\varepsilon_i\right) && \text{(A4)} \\
&= \frac{25}{n^2}\sum \sigma^2 && \text{(A3)} \\
&= \frac{25}{n^2} n\sigma^2 \\
&= \frac{25\sigma^2}{n}
\end{aligned}
$$

(c) Given your derivations thus far, and the assumptions listed above, what is the distribution of the sample mean?

$$
\hat{\mu} \sim \mathcal{N}\left(5\mu, \frac{25\sigma^2}{n}\right)
$$

3. You are asked to determine if the majority of the people of Westeros support Daenerys to sit on the Iron Throne. We know there are 25 million people in Westeros, let y_i equal 1 if a person supports Daenerys and 0 otherwise. Define $d = \sum_{i=1}^{25,000,000} y_i$ to be the number of people in Westeros who support Daenerys and $p = \frac{d}{25,000,000}$ to be the proportion of people who support Daenerys. p is the variable we want to learn.

We randomly choose with replacement 2 people and ask them if they support Daenerys. Let Y_1 and Y_2 denote these answers. We know that $\mathbb{E}(Y_1) = p$ and $V(Y_1) = p(1-p)$ and $\mathbb{E}(Y_2) = p$ and $V(Y_2) = p(1-p)$. We also know that Y_1 and Y_2 are independent.

(a) Is $10Y_1 - 9Y_2$ a biased or unbiased estimator of p? Justify each step in your derivation.

Answer:

$$\begin{aligned}
\mathbb{E}(10Y_1 - 9Y_2) &= \mathbb{E}(10Y_1) - \mathbb{E}(9Y_2) && \text{(P1}\mathbb{E}\text{ and P2}\mathbb{E}\text{)} \\
&= 10\mathbb{E}(Y_1) - 9\mathbb{E}(Y_2) && \text{(P2}\mathbb{E}\text{)} \\
&= 10p - 9p && (\mathbb{E}(Y_i) = p) \\
&= p && \text{(unbiased)}
\end{aligned}$$

(b) Compute $V(10Y_1 - 9Y_2)$. Justify all steps.

Answer:

$$\begin{aligned}
V(10Y_1 - 9Y_2) &= V(10Y_1) + V(-9Y_2) && \text{(P2Var \& independence)} \\
&= 100\mathbb{E}(Y_1) + 81\mathbb{E}(Y_2) && \text{(P2Var)} \\
&= 100p(1-p) + 81p(1-p) && (Var(Y_i) = p(1-p)) \\
&= 181p(1-p)
\end{aligned}$$

(c) Suppose we randomly choose 3600 people and asked them if they support Daenerys. Let $Y_1, Y_2, \ldots, Y_{3600}$ denote the answers of these 3600 people. Assume that you find that $\frac{1}{3600}\sum_{i=1}^{3600} Y_i = 0.48$, meaning 48% of those polled support Daenerys. Compute the 95% confidence interval of p.

Answer: The 95% confidence interval of p is $\left[\bar{Y} - 1.96\sqrt{\frac{\bar{Y}(1-\bar{Y}}{n}}, \bar{Y} + 1.96\sqrt{\frac{\bar{Y}(1-\bar{Y}}{n}}\right]$. So the confidence interval is $\left[0.49 - 1.96\sqrt{\frac{0.49(1-0.49)}{3600}}, 0.49 + 1.96\sqrt{\frac{0.49(1-0.49)}{3600}}\right] = [0.47367, 0.50633]$.

4. You want to know if tutors help students pass a class. To answer that question, you conduct a randomized controlled trial. You run an OLS regression of Y_i on a constant and D_i. Let $\hat{\beta}_1$ denote the coefficient of D_i in that regression, this coefficient is the Average Treatment Effect (ATE). The ATE can be written as $\frac{1}{n}\sum_{i=1}^{n} y_i(1) - \frac{1}{n}\sum_{i=1}^{n} y_i(0)$

   (a) Prove that $\hat{\beta}_1$ is an unbiased estimator of the ATE.

   Answer:

$$
\begin{aligned}
E(\widehat{\beta}_1) &= E\left(\frac{1}{n_1}\sum_{i=1}^{n} y_i(1)D_i - \frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)(1-D_i)\right) \\
&= E\left(\frac{1}{n_1}\sum_{i=1}^{n} y_i(1)D_i\right) - E\left(\frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)(1-D_i)\right) \\
&= \frac{1}{n_1}E\left(\sum_{i=1}^{n} y_i(1)D_i\right) - \frac{1}{n-n_1}E\left(\sum_{i=1}^{n} y_i(0)(1-D_i)\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n} E\left(y_i(1)D_i\right) - \frac{1}{n-n_1}\sum_{i=1}^{n} E\left(y_i(0)(1-D_i)\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n} y_i(1)E\left(D_i\right) - \frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)E\left(1-D_i\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n} y_i(1)E\left(D_i\right) - \frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)\left(1-E(D_i)\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n} y_i(1)\frac{n_1}{n} - \frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)\left(1-\frac{n_1}{n}\right) \\
&= \frac{1}{n_1}\sum_{i=1}^{n} y_i(1)\frac{n_1}{n} - \frac{1}{n-n_1}\sum_{i=1}^{n} y_i(0)\frac{n-n_1}{n} \\
&= \frac{1}{n_1}\frac{n_1}{n}\sum_{i=1}^{n} y_i(1) - \frac{1}{n-n_1}\frac{n-n_1}{n}\sum_{i=1}^{n} y_i(0) \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i(1) - \frac{1}{n}\sum_{i=1}^{n} y_i(0) \\
&= ATE.
\end{aligned}
$$

1st equality: result from question 2. 2nd equality: P1Exp and P2Exp. 3rd equality: P2Exp. 4th equality: P3Exp. 5th equality: P2Exp. 6th equality: P1Exp and P2Exp. 7th equality: $E(D_i) = \frac{n_1}{n}$. 9th equality: P2Sum.