**Midterm Practice Problems**
Economics 140A
August 23, 2017

1. 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 $\mu$.

(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 $\mu$? Show all work.

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

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?

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

2. Suppose someone suggests to you an alternative linear estimator given by:

$$
\tilde{\beta} = \frac{\sum x_i^2 y_i}{\sum x_i^3}
$$

Assume:

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

(a) Find $\mathbb{E}[\tilde{\beta}]$. Is $\tilde{\beta}$ an unbiased estimator?

$$
\begin{aligned}
\mathbb{E}[\tilde{\beta}] &= \mathbb{E}\left[\frac{\sum x_i^2 y_i}{\sum x_i^3}\right] \\
&= \mathbb{E}\left[\frac{\sum x_i^2(\beta x_i + \varepsilon_i)}{\sum x_i^3}\right] && \text{(A1)} \\
&= \mathbb{E}\left[\frac{\sum(\beta x_i^3 + \varepsilon_i x_i^2)}{\sum x_i^3}\right] \\
&= \mathbb{E}\left[\beta\frac{\sum x_i^3}{\sum x_i^3} + \frac{\sum \varepsilon_i x_i^2}{\sum x_i^3}\right] \\
&= \mathbb{E}\left[\beta\frac{\cancel{\sum x_i^3}}{\cancel{\sum x_i^3}} + \frac{\sum \varepsilon_i x_i^2}{\sum x_i^3}\right] \\
&= \mathbb{E}\left[\beta\right] + \mathbb{E}\left[\frac{\sum \varepsilon_i x_i^2}{\sum x_i^3}\right] \\
&= \beta + \frac{1}{\sum x_i^3}\mathbb{E}\left[\sum \varepsilon_i x_i^2\right] && \text{(A2)} \\
&= \beta + \frac{1}{\sum x_i^3}\sum x_i^2 \mathbb{E}[\varepsilon_i] && \text{(A2)} \\
&= \beta + \frac{1}{\sum x_i^3}\sum x_i^2 \cancelto{0}{\mathbb{E}[\varepsilon_i]} && \text{(A3)} \\
&= \beta
\end{aligned}
$$

The estimator is unbiased.

(b) You now run the regression,

$$wage_i = \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 female_i + u_i,$$

where $wage_i$ is the hourly wage of an individual, $age_i$ is an individual's age in years, $educ_i$ is an individual's education in years, and $female_i$ is equal to 1 if the individual is female and 0 if the individual is male. Interpret the coefficients $\beta_1$ and $\beta_3$.

An additional year of age increases your hourly wage by $\hat{\beta}_1$ on average holding all other covariates constant. On average, the wage difference between males and females with the same age and years of education equals $\hat{\beta}_3$.

3. You wish to examine the relationship between the number of planets a spaceship can visit in one year and the number of crew members on the spaceship. You set up the following model where $p_i$ is the number of planets a spaceship visits in a year and $x_i$ is the number of crew members on the ship.

$$p_i = \beta x_i + u_i$$

Assume:

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

(a) Give the OLS estimator for $\beta$. Show that the OLS estimator of $\beta$ is unbiased. Make sure to be clear about when you use each assumption and show all work.

$$\begin{aligned}
\hat{\beta} &= \frac{\sum x_i p_i}{\sum x_i^2} \\
\mathbb{E}[\hat{\beta}] &= \mathbb{E}\left[\frac{\sum x_i p_i}{\sum x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\sum x_i(\beta x_i + u_i)}{\sum x_i^2}\right] && \text{(A1, 2pts)} \\
&= \mathbb{E}\left[\frac{\sum(\beta x_i^2 + u_i x_i)}{\sum x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\sum \beta x_i^2}{\sum x_i^2} + \frac{\sum u_i x_i}{\sum x_i^2}\right] \\
&= \mathbb{E}\left[\beta \frac{\cancel{\sum x_i^2}}{\cancel{\sum x_i^2}}\right] + \mathbb{E}\left[\frac{\sum u_i x_i}{\sum x_i^2}\right] \\
&= \mathbb{E}[\beta] + \frac{\sum \mathbb{E}[u_i] x_i}{\sum x_i^2} && \text{(A2, 2pts)} \\
&= \beta + \frac{\sum \cancelto{0}{\mathbb{E}[u_i]} x_i}{\sum x_i^2} && \text{(A3, 2pts)} \\
&= \beta
\end{aligned}$$

$\beta$ is unbiased.(4pts)

(b) Suppose you now observe $w_i$, the weight of the ship, a characteristic you believe affects $p_i$. You now use the following model.

$$p_i = \beta_1 x_i + \beta_2 w_i + \varepsilon_i$$

Interpret the coefficient $\beta_1$, explain what the coefficient means in terms a non-statistician would understand.

An additional crew member increases the number of planets visited in a year by $\hat{\beta}_1$ on average holding all other covariates constant.(10 pts)

(c) You believe $x_i$ and $w_i$ have a positive covariance, explain what this means in terms a non-statistician would understand.

A positive covariance tells us that when one variable is above the mean the other variable is also likely to be above the mean.(10 pts)

4. A spaceship is stranded in outerspace, but there is hope for a rescue as it sends back encrypted signals. In order to decrypt the signal, you need to figure out the distribution of the signal. After some analysis, the possibilities are narrowed down to two distributions: a uniform distribution in $(0, 1)$, i.e. the signal takes any value in (0,1) with equal probability; and a normal $\mathbb{N}(0.7, 0.3)$ distribution, i.e. the signal is distributed normal with mean 0.7 and variance 0.3. Your job is to determine the correct distribution.

(a) One way of determining the correct distribution is by looking at the moments. The first moment is the mean. What is the mean of each distribution?

The mean of the uniform distribution is 0.5 (5 pts) The mean of the normal distribution is 0.7 (5pts)

(b) You receive 5 signals: 0.1, 0.3, 0.6, 1.1, 0.4. What is your estimate for the mean equal to? Based on the information, which do you think is the correct distribution?

The correct distribution is normal distribution since uniform (0,1) cannot take value 1.1 (10pts)

(c) It turns out that the signals are received with some noise: $X_i = X_i^* + u_i$ where $X_i^*$ is the true signal with one of the two distributions above and the noise $u_i$ are independent and identically distributed with normal $\mathbb{N}(0, 0.05)$ (a normal distribution with mean 0 and variance 0.05). $X_i^*$ and $u$ are also independent. Given this, is your mean estimator unbiased? Does this change your answer to part (b)? Show necessary details of your argument.

Our mean estimator is $\frac{1}{n}\sum_{i=1}^{n} X_i$ (2pts). It is unbiased:

$$\begin{aligned}
\mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) &= \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n}(X_i^* + u_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^*) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(u_i) \\
&= \frac{1}{n}n\mathbb{E}(X_i^*) + 0 \\
&= \mathbb{E}(X^*)(4\text{pts})
\end{aligned}$$

This changes the answer to part (b) because now $X_i$ can take values in any real number no matter what the distribution of $X_i^*$ is. Since the estimate is equal to 0.5, we would guess that the true signal has mean 0.5, hence a uniform (0,1) distribution. (4pts)