Econometrics – Problem Set 1

Junwoo Yang

March 15, 2021

1 Yes, Economics Is a Science

It is often seen that there are disagreements among economists. Skeptics say this discrepancy cannot be a useful basis for economics to make policy decisions. Identifying the cause or determinant of a Macro question is challenging. But medicine is not much different. Just by looking at how diet and lifestyle affect health and aging for more than a century. To overcome the cost of several experiments, economists began to develop tools that were close to scientific experiments. In the past, theorists such as Krugman and Yellen were famous, but today prominent economists are empiricists such as Card and Duflo. Data-based studies have also shown that extending unemployment benefits does not significantly increase unemployment rate while providing assistance to unemployed people, and have also identified causal effects by separating certain factors from others. Thus, economics also falls into the category of science and will become a more empirical and scientific field as the availability of data increases.

2 Exercises

#2.6

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

	Unemployed ($Y = 0$)	Employed ($Y = 1$)	Total
Non-college grads ($X = 0$)	0.026	0.576	0.602
College grads ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

a. Compute $E(Y)$.

Proof. $E(Y) = \sum_{y=0}^{1} yP(Y = y) = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = 0 \times 0.035 + 1 \times 0.965 = 0.965.$ □

b. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.

Proof. Unemployment Rate $= \dfrac{\#\text{ unemployed}}{\#\text{ labor force}} = P(Y = 0) = 0.035 = 1 - 0.965 = 1 - E(Y).$ □

c. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.

Proof. First, let calculate the conditional probability:

$$
\begin{aligned}
P(Y=0|X=0) &= \frac{P(X=0,Y=0)}{P(X=0)} = \frac{0.026}{0.602} = 0.0432, \\
P(Y=1|X=0) &= \frac{P(X=0,Y=1)}{P(X=0)} = \frac{0.576}{0.602} = 0.9568, \\
P(Y=0|X=1) &= \frac{P(X=1,Y=0)}{P(X=1)} = \frac{0.009}{0.398} = 0.0226, \\
P(Y=1|X=1) &= \frac{P(X=1,Y=0)}{P(X=1)} = \frac{0.389}{0.398} = 0.9774.
\end{aligned}
$$

The conditional expectations are

$$
\begin{aligned}
E(Y|X=1) &= 0 \cdot P(Y=0|X=1) + 1 \cdot P(Y=1|X=1) = 0 \times 0.0226 + 1 \times 0.9774 = 0.9774, \\
E(Y|X=0) &= 0 \cdot P(Y=0|X=0) + 1 \cdot P(Y=1|X=0) = 0 \times 0.0432 + 1 \times 0.9568 = 0.9568.
\end{aligned}
$$

□

d. Calculate the unemployment rate for (1) college graduates and (2) non-college graduates.

Proof. By using result of **b.**, unemployment rate for college graduates is

$$1 - E(Y|X=1) = 1 - 0.9774 = 0.0226,$$

and one for non-college graduates is

$$1 - E(Y|X=1) = 1 - 0.9568 = 0.0432.$$

□

e. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non-college graduate?

Proof. The probability that a randomly selected worker who is reported being unemployed is a college graduate is

$$P(X=1|Y=0) = \frac{P(X=1,Y=0)}{P(Y=0)} = \frac{0.009}{0.035} = 0.2571.$$

The probability that this worker is a non-college graduate is

$$P(X=0|Y=0) = \frac{P(X=0,Y=0)}{P(Y=0)} = \frac{0.026}{0.035} = 0.7429.$$

□

f. Are educational achievement and employment status independent? Explain.

Proof. To be independent between random variables X and Y, it should hold

$$P(A \subset X, B \subset Y) = P(A \subset X)P(B \subset Y)$$

where A, B are subsets of $\{0,1\}$. It is trivial for $\emptyset$ and $\{0,1\}$. Thus, we only need to check for point subsets. However,

$$0.026 = P(X=0,Y=0) \neq P(X=0)P(Y=0) = 0.602 \times 0.035 = 0.021.$$

Thus, X and Y are not independent. $\square$

#2.7

In a given population of two-earner male-female couples, male earnings have a mean of \$40,000 per year and a standard deviation of \$12,000. Female earnings have a mean of \$45,000 per year and a standard deviation of \$18,000. The correlation between male and female earnings for a couple is 0.80. Let C denote the combined earnings for a randomly selected couple.

a. What is the mean of C?

Proof. Let M, F be random variable with $\mu_M = \$40,000$, $\sigma_M = \$12,000$, $\mu_F = \$45,000$, and $\sigma_F = \$18,000$. Let C be summation of M and F, that is, $C = M + F$. Thus, mean of C, which is expectation of C, is

$$\mu_C = E(C) = E(M + F) = E(M) + E(F) = \mu_M + \mu_F = \$40,000 + \$45,000 = \$85,000. \quad \square$$

b. What is the covariance between male and female earnings?

Proof. By using the definition of correlation, $\rho_{M,F} = \frac{\operatorname{Cov}(M,F)}{\sigma_M \sigma_F}$, we get

$$\operatorname{Cov}(M, F) = \rho_{M,F}\sigma_M\sigma_F = 0.8 \times \$12,000 \times \$18,000 = \$172,800,000. \quad \square$$

c. What is the standard deviation of C?

Proof. By the fact that $\operatorname{Var}(C) = \operatorname{Var}(M + F) = \operatorname{Var}(M) + \operatorname{Var}(F) + 2\operatorname{Cov}(M, F)$, we get

$$\sigma_C = \sqrt{\sigma_M^2 + \sigma_F^2 + 2\operatorname{Cov}(M, F)} = \sqrt{\$12,000^2 + \$18,000^2 + 2 \times \$172,800,000} \approx \$28,524. \quad \square$$

d. Convert the answers to (a) through (c) from U.S. dollars (\$) to euros (€).

Proof. Let the exchange rate of US dollar to euro be 0.84. Then, all the answers are just a result multiplied by 0.84.

$$\mu_C = €71,400, \quad \operatorname{Cov}(M, F) = €145,152,000, \quad \sigma_C \approx €23,960. \quad \square$$

#2.10

Compute the following probability:

a. If Y is distributed $N(1, 4)$, find $P(Y \leq 3)$.

Proof. These probabilities can be obtained directly through program tools such as Excel, Python, and R. However, it can also be obtained through standardization and a standard normal table (Z table). By standardization, $\frac{Y-\mu_Y}{\sigma_Y}$, and Z table, we get

$$P(Y \leq 3) = P\left(\frac{Y-1}{2} \leq \frac{3-1}{2}\right) = P(Z \leq 1) = \Phi(1) = 0.84134$$

where Z is standard normal random variable, Φ is cumulative distribution function of Z. $\square$

b. If Y is distributed $N(3, 9)$, find $P(Y > 0)$.

Proof.
$$\begin{aligned} P(Y > 0) &= P\left(\frac{Y-3}{3} > \frac{0-3}{3}\right) = P(Z > -1) \\ &= 1 - \Phi(-1) = 1 - 0.15866 = 0.84134. \end{aligned}$$
□

c. If Y is distributed $N(50, 25)$, find $P(40 \le Y \le 52)$.

Proof.
$$\begin{aligned} P(40 \le Y \le 52) &= P\left(\frac{40-50}{5} \le \frac{Y-50}{5} \le \frac{52-50}{5}\right) = P(-2 \le Z \le 0.4) \\ &= \Phi(0.4) - \Phi(-2) = 0.65542 - 0.02275 = 0.63267. \end{aligned}$$
□

d. If Y is distributed $N(5, 2)$, find $P(6 \le Y \le 8)$.

Proof.
$$\begin{aligned} P(6 \le Y \le 8) &= P\left(\frac{6-5}{\sqrt{2}} \le \frac{Y-5}{\sqrt{2}} \le \frac{8-5}{\sqrt{2}}\right) \approx P(0.71 \le Z \le 2.12) \\ &= \Phi(2.12) - \Phi(0.71) = 0.98300 - 0.76115 = 0.22185. \end{aligned}$$
□

#2.13

X is a Bernoulli random variable with $P(X = 1) = 0.99$; Y is distributed $N(0, 1)$; W is distributed $N(0, 100)$; and X, Y, and W are independent. Let $S = XY + (1 - X)W$. (That is, $S = Y$ when $X = 1$, and $S = W$ when $X = 0$.)

a. Show that $E(Y^2) = 1$ and $E(W^2) = 100$.

Proof.
$$\begin{aligned} E(Y^2) &= \mathrm{Var}(Y) + E(Y)^2 = 1 + 0 = 1, \\ E(W^2) &= \mathrm{Var}(W) + E(W)^2 = 100 + 0 = 100. \end{aligned}$$
□

b. Show that $E(Y^3) = 0$ and $E(W^3) = 0$.

Proof. To prove this, we need the moment generating function (mgf) of normal random variable, and its derivatives. Consider random variable Z distributed $N(\mu, \sigma^2)$. The probability density function (pdf) of Z is

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < z < \infty.$$

The mgf of Z is defined as follows:

$$\begin{aligned} M_Z(t) = E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tz} f(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} \frac{e^{tz}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} \mathrm{d}z \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{z^2 - 2(\mu + \sigma^2 t)z + \mu^2}{2\sigma^2}\right\} \mathrm{d}z \\ &= \exp\left(\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(z - (\mu + \sigma^2 t))^2}{2\sigma^2}\right\}}_{\text{pdf of } N(\mu+\sigma^2 t, \sigma^2)} \mathrm{d}z \\ &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right), \quad -\infty < t < \infty. \end{aligned}$$

The first derivative, $M_Z^{(1)}(t)$

$$M_Z^{(1)}(t) = \frac{d}{dt}M_Z(t) = (\mu + \sigma^2 t)\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) = (\mu + \sigma^2 t)M_Z(t).$$

The second derivative, $M_Z^{(2)}(t)$

$$M_Z^{(2)}(t) = \frac{d^2}{dt^2}M_Z(t) = \{\sigma^2 + (\mu + \sigma^2 t)^2\}\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) = \{\sigma^2 + (\mu + \sigma^2 t)^2\}M_Z(t).$$

The third derivative, $M_Z^{(3)}(t)$

$$\begin{aligned} M_Z^{(3)}(t) = \frac{d^3}{dt^3}M_Z(t) &= 2\sigma^2(\mu + \sigma^2 t)M_Z(t) + \{\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3\}M_Z(t) \\ &= \{3\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3\}M_Z(t). \end{aligned}$$

Note that

$$M_Z^{(n)}(0) = E\left[\frac{d^n}{dt^n}e^{tZ}\bigg|_{t=0}\right] = E\left[Z^n e^{tZ}\big|_{t=0}\right] = E[Z^n]. \tag{1}$$

Because Y and W are also normal random variables, mgf are shown as above. Thus,

$$\begin{aligned} E(Y^2) &= M_Y^{(2)}(0) = \{\sigma_Y^2 + (\mu_Y + \sigma_Y^2 t)^2\}M_Y(t)\Big|_{t=0} = \sigma_Y^2 = 1, \\ E(W^2) &= M_W^{(2)}(0) = \{\sigma_W^2 + (\mu_W + \sigma_W^2 t)^2\}M_W(t)\Big|_{t=0} = \sigma_W^2 = 100. \end{aligned}$$

This is alternative proof of **a.**. Now, for third moment,

$$\begin{aligned} E(Y^3) = M_Y^{(3)}(0) &= \{3\sigma_Y^2(\mu_Y + \sigma_Y^2 t) + (\mu_Y + \sigma_Y^2 t)^3\}M_Y(t)\big|_{t=0} = 0, \\ E(W^3) = M_W^{(3)}(0) &= \{3\sigma_W^2(\mu_W + \sigma_W^2 t) + (\mu_W + \sigma_W^2 t)^3\}M_W(t)\big|_{t=0} = 0. \end{aligned}$$ □

c. Show that $E(Y^4) = 3$ and $E(W^4) = 3 \times 100^2$.

Proof. (1) says that $E(Z^4) = M_Z^{(4)}(0)$. The fourth derivative, $M_Z^{(4)}$

$$\begin{aligned} M_Z^{(4)} &= \{3\sigma^4 + 3\sigma^2(\mu + \sigma^2 t)^2\}M_Z(t) + \{3\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4\}M_Z(t) \\ &= \{3\sigma^4 + 6\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4\}M_Z(t). \end{aligned}$$

If we apply this to Y and W,

$$\begin{aligned} E(Y^4) = M_Y^{(4)}(0) &= \{3\sigma_Y^4 + 6\sigma_Y^2(\mu_Y + \sigma_Y^2 t)^2 + (\mu_Y + \sigma_Y^2 t)^4\}M_Y(t)\big|_{t=0} = 3\sigma_Y^4 = 3, \\ E(W^4) = M_W^{(4)}(0) &= \{3\sigma_W^4 + 6\sigma_W^2(\mu_W + \sigma_W^2 t)^2 + (\mu_W + \sigma_W^2 t)^4\}M_W(t)\big|_{t=0} = 3\sigma_W^4 = 3 \times 100^2. \end{aligned}$$ □

d. Derive $E(S)$, $E(S^2)$, $E(S^3)$, and $E(S^4)$.

Proof.

$$\begin{aligned} E(S) &= E(S|X = 0)P(X = 0) + E(S|X = 1)P(X = 1) \\ &= E(W)P(X = 0) + E(Y)P(X = 1) = 0, \end{aligned}$$

$$
\begin{aligned}
E(S^2) &= E(S^2|X=0)P(X=0) + E(S^2|X=1)P(X=1) \\
&= E(W^2)P(X=0) + E(Y^2)P(X=1) = 100 \times 0.01 + 1 \times 0.99 = 1.99, \\
E(S^3) &= E(S^3|X=0)P(X=0) + E(S^3|X=1)P(X=1) \\
&= E(W^3)P(X=0) + E(Y^3)P(X=1) = 0, \\
E(S^4) &= E(S^4|X=0)P(X=0) + E(S^4|X=1)P(X=1) \\
&= E(W^4)P(X=0) + E(Y^4)P(X=1) = 3 \times 100^2 \times 0.01 + 3 \times 0.99 = 302.97. \qquad \square
\end{aligned}
$$

e. Derive the skewness and kurtosis for S.

Proof. Since $\mu_S = E(S) = 0$, $\sigma_S^2 = E(S^2) - E(S)^2 = 1.99$,

$$
\begin{aligned}
\text{skewness} &= E\left[\left(\frac{S-\mu_S}{\sigma_S}\right)^3\right] = \frac{E(S^3)}{\sigma_S^3} = 0, \\
\text{kurtosis} &= E\left[\left(\frac{S-\mu_S}{\sigma_S}\right)^4\right] = \frac{E(S^4)}{\sigma_S^4} = \frac{302.97}{1.99^2} = 76.5. \qquad \square
\end{aligned}
$$

#**2.14**

In a population, $\mu_Y = 100$ and $\sigma_Y^2 = 43$. Use the central limit theorem to answer the following questions:

a. In a random sample of size $n = 100$, find $P(\overline{Y} \leq 101)$.

Proof. We know $\overline{Y}$ distributed $(\mu_Y, \frac{\sigma_Y^2}{n})$. The actual distribution of $\overline{Y}$ is unknown, but if the sample is large, we can approximate it with a normal distribution by CLT. Under CLT, we get

$$
P(\overline{Y} \leq 101) = P\left(\frac{\overline{Y}-100}{\sqrt{43/100}} \leq \frac{101-100}{\sqrt{43/100}}\right) \approx P(Z \leq 1.52) = \Phi(1.52) = 0.93574. \qquad \square
$$

b. In a random sample of size $n = 165$, find $P(\overline{Y} > 98)$.

Proof.
$$
\begin{aligned}
P(\overline{Y} > 98) &= P\left(\frac{\overline{Y}-100}{\sqrt{43/165}} > \frac{98-100}{\sqrt{43/165}}\right) \\
&\approx P(Z > -3.92) = 1 - \Phi(-3.92) = 1 - 0.00004 = 0.99996. \qquad \square
\end{aligned}
$$

c. In a random sample of size $n = 64$, find $P(101 \leq \overline{Y} \leq 103)$.

Proof.
$$
\begin{aligned}
P(101 \leq \overline{Y} \leq 103) &= P\left(\frac{101-100}{\sqrt{43/64}} \leq \frac{\overline{Y}-100}{\sqrt{43/64}} \leq \frac{103-100}{\sqrt{43/64}}\right) \\
&\approx P(1.22 \leq Z \leq 3.66) \\
&= \Phi(3.66) - \Phi(1.22) = 0.99987 - 0.88877 = 0.1111. \qquad \square
\end{aligned}
$$