Econometrics Homework 1

Tom Augspurger

September 7, 2012

1. Show that $E_X[E[g(Y)|X]] = E[g(Y)]$.

If $g_2(y|x)$ is the conditional expectation of y given x and $f_1(x)$ is the marginal distribution of x, then

$$\begin{aligned}
E_X[E[g(Y)|X]] &= \sum_i E[g(Y)] f_1(x_i) \\
&= \sum_i f_1(x_i) \sum_j g_2(y_j|x_i) g(y_j) \\
&= \sum_i \sum_j g_2(y_j|x_i) g(y_j) f_1(x_i) \\
&= \sum_i \sum_j f(x_i, y_j) g(y_j) = E[g(Y)]
\end{aligned}$$

2. Let X = Male Earnings; Y = Female Earnings. Z = X + Y
We're given that $E(X) = \$40,000$; $E(Y) = \$45,000$ and $\sigma_X = \$12,000$; $\sigma_Y = \$18,000$;

a: The mean of Z is $E(Z) = E(X) + E(Y) = \$85,000$

b: The covariance of male and female earnings is $C(X, Y)$. We know that $\rho(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = 0.8$ so $\sigma_{XY} = 0.8\sigma_X\sigma_Y = \$172,800,000$.

c: The standard deviation of Z is

$$\begin{aligned}
\sigma_Z &= \sigma_{X+Y} \\
&= \sqrt{V(X) + V(Y) + 2C(X, Y)} \\
&= \sqrt{\$345,600,030} \\
&= \$25,314.03
\end{aligned}$$

d: In Euros $E(X) = €68,850$

$C(X, Y) = €139,968,000$

$\sigma_Z = €2,050.44$

3. Y_1 is the value of sales and Y_2 is the value of the costs. Assume that

$$f_1(y_1) = \begin{cases} \frac{1}{6} y_1^3 e^{-y_1} & \text{if } y_1 > 0 \\ 0 & \text{if } y_1 \leq 0 \end{cases}$$

$$f_2(y_2) = \begin{cases} \frac{1}{2} e^{-\frac{y_2}{2}} & \text{if } y_2 > 0 \\ 0 & \text{if } y_2 \leq 0 \end{cases}$$

and let $\Pi = Y_1 - Y_2$.

a: $E[\Pi] = E[Y_1] - E[Y_2] = \frac{1}{6}\int_0^\infty y_1 y_1^3 e^{-y_1}\, dy - \frac{1}{2}\int_0^\infty y_2 e^{-\frac{y_2}{2}}\, dy$
$= 4 - 2 = 2$

b: $V(\Pi) = E(\Pi^2) - E^2(\Pi) = 20 - 8 - 4 = 8$

c: We would expect negative profits to be observed since $E(\Pi) - V(\Pi) < 0$.
First note that since Y_1 and Y_2 are independent $f(y_1, y_2) = f_1(y_1) f_2(y_2) = \frac{1}{12} y_1^3 exp[-y_1 - \frac{y_2}{2}]$. So,

$$\begin{aligned} Pr(Y_1 < Y_2) &= \int_0^\infty \int_0^{y_2} f(y_1, y_2)\, dy_1\, dy_2 \\ &= \int_0^\infty \int_0^{y_2} \frac{1}{12} y_1^3 exp[-y_1 - \frac{y_2}{2}]\, dy_1\, dy_2 \\ &= \frac{16}{81} \approx .197531 \end{aligned}$$

4. Employment Status and College Graduation
a: $E(Y|X = x) = \sum_i y_i Pr(Y = y_i | X = x)$ so

$E(Y|X = 0) = \frac{0.37}{0.659} 0 + \frac{0.622}{0.659} 1 \approx 0.9439$

$E(Y|X = 1) = \frac{0.009}{0.341} 0 + \frac{0.332}{0.341} 1 \approx 0.9736$

b: Find $E^*(Y|X) = \alpha + \beta X$

$$\begin{aligned} \beta &= \frac{C(X, Y)}{V(X)} \\ &= \frac{E(XY) - E(X)E(Y)}{V(X)} \\ &= \frac{\sum_i \sum_j f(x_i y_j) x_i y_j - E(X)E(Y)}{E(X^2) - E^2(X)} \\ &= \frac{0.332 - (0.341)(0.954)}{.341 - .116281} \\ &= \frac{.00669}{.22472} \approx .02977 \end{aligned}$$

so

$$\begin{aligned}\alpha &= E(Y) - \beta E(X) \\ &= .9540 - (.02977)(.341) \approx .94385\end{aligned}$$

Thus, $E^*[Y|X] = 0.94385 + .02977X$

c: $E[Y|X]$ and $E^*[Y|X]$ are the same at the two points they are both defined, $X = 0$ and $X = 1$.

5. I've assumed that the question is asking for a simple regression between average hourly earnings and age, so I haven't controlled for any gender or education in my analysis below. The regression equation is

$$AHE = 2.6235 + .4285 * AGE$$

where AHE is the mean hourly earnings of a worker, in dollars, and AGE is the years a person has lived. On average, mean earnings of a worker increase by about \$0.4285 per year lived. The relationship is statistically significant at the 10%, 5%, and 1% levels with a t-statistic of 16.850 and a p-value of about 0.000. So we conclude that the slope on the AGE variable, (β_1), is significantly different than zero.

A 95% confidence interval around β_1 is (0.379, 0.478) \$/year lived. Note that zero is not contained in this confidence interval.

The averaged-aged worker in the sample is about 29.64 years old. So his predicted earnings is $2.6235 + .4285(29.64)$ or about \$40.20.

Age alone does not seem to account for much of the variance in earnings across individuals with an R^2 of 1.8%.

If the variables gender and education are also used, the regression equation becomes

$$AHE = .0819 + 6.9360 * Education - 2.8045 * Gender + .4530 * Age$$

where Education and Gender are dummy variables (B.A. = 1, Female = 1). All variables are significant at the 1% level, except for the constant (which isn't economically meaningful anyway). A 95% confidence interval for the slope on the Age variable, once education and gender have been controlled for, is (0.407, 0.499) dollars per year.