BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Fall 2003
Problem Set 1
Due at classtime, Tuesday 16 Sep 2003

Problem sets should be your own work. You may work together with classmates, but if you're not figuring this out on your own, you will eventually regret it.

1. Suppose the following equation describes the relationship between the average number of classes missed during a semester (*missed*) and the distance from school (*distance*, measured in miles):

$$missed = 5 + 0.15\ distance$$

(i) Sketch this line, being sure to label the axes. How do you interpret the intercept in this equation?

(ii) What is the average number of classes missed for someone who lives 5 miles away?

(iii) What is the difference in the average number of classes missed for someone who lives 20 miles away?

Answer:

(i)The sketch should have missed classes on the vertical axis and distance to school on the horizontal axis. The intercept is equal to 5 and the slope is equal to .15. The intercept is the number of missed classes for a student who lives on campus.

(ii) $5 + .15(5) = 5.75$ classes

(iii) $15(.15) = 2.25$ classes

2. Suppose the following model describes the relationship between annual salary (*salary*) and the number of previous years of labor market experience (*exper*):

$$\log(salary) = 10.4 + .028\ exper$$

(i) What is the salary when exper = 0? when exper = 5? (Hint:You will need to exponentiate.)

(ii) Use equation (A.28) to approximate the percentage increase in salary when exper increases by five years.

(iii) Use the results of part (i) to compute the exact percentage difference in salary when exper = 5 and exper = 0. Comment on how this compares with the approximation in part (ii).

Answer:

(i) When exper$=0$, $\log(salary) = 10.4$; therefore, $salary = \exp(10.4) \approx \$32,859.63$. When exper$=5$, $salary = \exp[10.4 + .028(5)] \approx \$37,797.57$

(ii) The approximate proportionate increase is $.028(5) = .14$, so the approximate percentage change is 14%

(iii) $100[(37,797.57 - 32,859.63)/32,859.63] \approx 15.03\%$, so the exact percentage increase is about 1.03 percentage points higher.

3. Let *grthemp* denote the proportional growth in employment, at the country level, from 1990 to 1995, and let *salestax* denote the country sales tax rate, stated as a proportion. Interpret the intercept and slope in the equation

$$grthemp = .041 - .79 salestax$$

Answer:

From the given equation, $\triangle grthemp = -.79(\triangle salestax)$. Since both variables are in proportion form, we can multiply the equation through by 100 to turn each variable into percentage form. This leaves the slope as -.79. So, a one percentage point increase in the sales tax rate (say, from 4% to 5%) reduces employment growth by -.79 percentage points.

4. Let X be a random variable distributed as Normal (5,9). Find the probabilities of the following events:

(i) $P(X \leq 6)$

(ii) $P(X > 4)$

(iii) $P(\mid X - 5 \mid > 1)$

Answer:

(i) $P(X \leq 6) = P(\frac{X-5}{3} \leq \frac{6-5}{3}) = P(z \leq .33) \approx .6293$

(ii) $P(X > 4) = 1 - P(X \leq 4) = 1 - P(\frac{X-5}{3} \leq \frac{4-5}{3}) = 1 - P(z \leq -.33) \approx .6293$

(iii) $P(|X-5| > 1) = P(X - 5 > 1) + P(X - 5 < -1) = P(X > 6) + P(X < 4) = P(z > \frac{6-5}{3}) + P(z < \frac{6-5}{3}) =$

$= 2P(z \leq -.33) = 2(1 - P(z \leq .33)) = 2(1 - .6293) = .7414$

5. For a randomly selected county in the United States, let X represent the proportion of adults over age 65 who are employed, or the elderly employment rate. Then, X is restricted to a value between zero and one. Suppose

that the cumulative distribution function for X is given by $F(x) = 3x^2 - 2x^3$ for all $0 \leq x \leq 1$. Find the probability that the elderly employment rate is at least .5 (50%).

Answer:

$P(X > .5) = 1 - P(X \leq .5)$
$P(X \leq .5) = F(.5) = 3(.5)^2 - 2(.5)^3 = .5$
$P(X > .5) = 1 - .5 = .5$

One way to interpret this is that over 50% of all counties have an elderly employment rate of 50% or higher.

6. Suppose that a college student is taking three courses: a two-credit course, a three-credit course, and a four-credit course. The expected grade in the two-credit course is 3.5, while the expected grade in the three- and four-credit courses is 3.3. What is the expected overall grade point average for the semester? (Remember, that each course grade is weighted by its share of the total number of units.)

Answer:

The weights for the two-, three-, and four-credit courses are $2/9, 3/9$, and $4/9$, respectively. Let Y_j be the grade in the j^{th} course, $j = 1, 2$, and 3 and let X be the overall grade point average. Then $X = (2/9)Y_1 + (3/9)Y_2 + (4/9)Y_3$ and the expected value is $E(X) = (2/9)E(Y_1) + (3/9)E(Y_2) + (4/9)E(Y_3) =$

$(2/9)(3.5) + (3/9)(3.3) + (4/9)(3.3) = (7 + 9.9 + 13.2)/9 \approx 3.34$

7. Let Y_1, Y_2, Y_3 and Y_4 be independent, identically distributed random variables from a population with mean μ and variance σ^2. Let $\overline{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ denote the average of these four random variables.

(i) What are the expected value and variance of $\overline{Y}$ in terms of μ and σ^2?

(ii) Now, consider different estimator of μ :

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4$$

This is an example of a *weighted* average of the Y_i . Show that W is also an unbiased estimator of μ. Find the variance of W.

(iii) Based on your answers to parts (i) and (ii), which estimator of μ do you prefer, $\overline{Y}$ or W.

(iv) Now, consider a more general estimation of μ, defined by

$$W_a = a_1Y_1 + a_2Y_2 + a_3Y_3 + a_4Y_4$$

where the a_i are constants. What condition is needed on the a_i for W_a to be an unbiased estimator of μ ?

(v) Compute the variance of the estimator W_a from part (iv)

Answer:

(i) This is just a special case of what we covered in the text, with $n = 4$: $E(\overline{Y}) = \mu$ and $Var(\overline{Y}) = \sigma^2/4$

(ii) $E(W) = E(Y_1)/8 + E(Y_2)/8 + E(Y_3)/4 + E(Y_4)/2 = \mu[(1/8) + (1/8) + (1/4) + (1/2)] = \mu$, which shows that W is unbiased. Because the Y_i are independent,

$Var(W) = Var(Y_1)/64 + Var(Y_2)/64 + Var(Y_3)/16 + Var(Y_4)/4 = \sigma^2[(1/64) + (1/64) + (4/64) + (16/64)] = \sigma^2(22/64) =$

$\sigma^2(11/32)$.

(iii) Because $11/32 > 1/4 = 8/32$, $Var(W) > Var(\overline{Y})$ for any $\sigma^2 > o$, $\overline{Y}$ is preferred to W because each is unbiased.

(iv) The expected value of W_a is $E(W) = a_1E(Y_1) + a_2E(Y_2) + a_3E(Y_3) + a_4E(Y_4) = \mu[a_1 + a_2 + a_3 + a_4]$. For this to equal μ for all μ, we must have $a_1 + a_2 + a_3 + a_4 = 1$

(v) $(Var(W_a) = a_1^2Var(Y_1) + a_2^2Var(Y_2) + a_3^2Var(Y_3) + a_4^2Var(Y_4) = \sigma^2[a_1^2 + a_2^2 + a_3^2 + a_4^2]$

8. Let Y denote the sample average from a random sample with mean μ and variance σ^2. Consider two alternative estimators of μ: $W_1 = [(n-1)/n]\overline{Y}$ and $W_2 = \overline{Y}/3$.

(i) Show that W_1 and W_2 are both biased estimators of μ and find the biases. What happens to the biases as $n \longrightarrow \infty$? Comment on any important differences in bias for the two estimators as the sample size gets large.

(ii) Find the probability limits of W_1 and W_2 {Hint: use properties PLIM.1 and PLIM.2; for W_1 note that plim$[(n-1)/n] = 1$ } Which estimator is consistent?

Answer:

(i) $E(W_1) = [(n-1)/n]E(\overline{Y}) = [(n-1)/n]\mu$, and so $Bias(W_1) = [(n-1)/n]\mu - \mu = -\mu/n$. Similarly, $E(W_2) = E(\overline{Y})/3 = \mu/3$ and so $Bias(W_2) = \mu/3 - \mu = -2\mu/3$. The bias in W_1 tends to zero as $n \to \infty$, while the bias in W_2 is $-2\mu/3$ for all n. This is an important difference.

(ii) plim(W_1) =plim$[(n-1)/n]*$plim$(\overline{Y}) = 1*\mu = \mu, whereas$plim$(W_2)$ =plim$(\overline{Y})/3 = \mu/3$. Because plim$(W_1) = \mu$ and plim$(W_2) = \mu/3$, W_1 is consistent whereas W_2 is inconsistent.

9. Suppose that a military dictator in an unnamed country holds a

plebiscite (a yes/no vote of confidence) and claims that he was supported by 67% of the voters. A human rights group suspects foul play and hires you to test the validity of the dictator's claim. You have a budget that allows you to randomly sample 200 voters from the country.

(i) Let X be the number of yes votes obtained from a random sample of 200 out of the entire voting population. What is the expected value of X if, in fact, 67% of all voters supported the dictator?

(ii) What is the standard deviation of X, again assuming that the true fraction voting yes in the plebiscite is .67?

(iii) Now, you collect your sample of 200, and you find that 115 people actually voted yes. Use the CLT to approximate the probability that you would find 115 or fewer yes votes from a random sample of 200 if, in fact, 67 % of the entire population voted yes.

(iv) How would you explain the relevance of the number in part (iii) to someone who does not have training in statistics?

Answer:

(i) X is distributed as $Binomial(200, .67)$, and so $E(X) = 200(.67) = 134$

(ii) $Var(X) = 200(.67)(1 - .67) = 44.22$, so $sd(x) \approx 6.65$

(iii) $P(X \leq 115) = P[(X - 134)/6.65 \leq (115 - 134)/6.65] \approx P(Z \leq -2.86)$, where Z is a standard normal random variable. From Table G.1, $P(Z \leq -2.86) \approx .0021$

(iv) The evidence is pretty strong against the dictator's claim. If 67% of the voting population actually voted yes in the plebiscite, there is only about a .21% chance of obtaining 115 or fewer voters out of 200 who voted yes.

10. In 1990, Mark Price was the top three–pojnt shooter in the National Basketball Association, making 188 of his 429 attempts. For a given player, the outcome of a shot can be viewed as a Bernoulli random variable. Let θ be the probability of making any particular three–point shot; a natural estimator of θ is the ratio of shots made to shots attempted.

(i) Estimate θ for Mark Price.

(ii) The asymptotic distribution of $(\hat{\theta} - \theta)/se(\hat{\theta})$ is standard Normal, where $se(\hat{\theta}) = \sqrt{\hat{\theta}(1 - \hat{\theta})/n}$ and n is the number of attempts. Use this fact to test $H_0 = \theta = 0.5$ against $H_1 = \theta < 0.5$ for Mark Price, using 5% and 1% significance levels.

Answer:

(i) Let Y_i be the r.v. that Mark's ith shot attempt, $Y_i = 1$ if Mark makes

it and $Y_i = 0$ otherwise. Then $Y_i \sim Bernoulli(\theta)$
Denote $\bar{Y}_i$ the sampling average of Y_i, and $\hat{\theta}$ the unbiased estimation of Y_i. Then $\hat{\theta} = E(\bar{Y}_i) = \frac{188}{429} = 0.438$

(ii) $se(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{0.438(1-0.438)}{429}} \approx 0.024$

$t = \frac{\hat{\theta}-\theta}{se(\hat{\theta})} = \frac{(0.438-0.5)}{0.024} = -2.58$

at 5% significance level, $c = 1.645$

$t < -c$ at 5% significance level, so $\theta = 0.5$ is rejected at this level.

at 1% significance level, $c = 2.326$

$t < -c$ at 1% significance level, so $\theta = 0.5$ is rejected at this level