Instructor: Fa-Hsiang Chang
Terms: Fall, 2013

ECO 480A Econometrics

Homework 2 Solution

Due Date: 10/18/2013

Chap 5

5-1

a. The probabilities are just the relative frequencies, obtained by dividing by N = 100,000,000, that is, just inserting an initial decimal point:

	y			b.	d.	
x	30	45	70	p(x)	xp(x)	$(x-\mu)^2 p(x)$
0	.01	.02	.05	.08	0	.271
1	.03	.06	.10	.19	.19	
2	.18	.21	.15	.54	1.08	or, by
3	.07	.08	.04	.19	.57	calculator
b. p(y)	.29	.37	.34	1.00	1.84	$\sigma_x^2 = .674$
					$\mu_x = 1.84$	$\sigma_x = .82$

c. In real life, two variables practically never are perfectly independent, because independence requires it to be exactly true that

$$p(x,y) = p(x)p(y) \text{ for all } x \text{ and } y \qquad (5\text{-}6)$$

Let us try the first cell:

L.S. = .01
R.S. = (.08)(.29) = .023 ≠ L.S.

This one exception to (5-6) rules out independence. We need go no further. X and Y are dependent.

5-4

(a)

X \ Y	1	2	P(X)
1	0.1	0.2	0.3
2	0.3	0.4	0.7
P(Y)	0.4	0.6	1

$$P(X,Y) \neq P(X)P(Y)$$

Not independent

(b)

X \ Y	0	1	2	P(X)
0	0.1	0.2	0.1	0.4
1	0.15	0.2	0.25	0.6
P(Y)	0.25	0.4	0.35	1

$$P(X,Y) \neq P(X)P(Y)$$

Not independent

(c)

X \ Y	0	1	2	P(X)
1	0	0.1	0.1	0.2
2	0.1	0.4	0.1	0.6
3	0.1	0.1	0	0.2
P(Y)	0.2	0.6	0.2	1

$$P(X,Y) \neq P(X)P(Y)$$

Not independent

(d)

Y \ X	1	2	3	4	P(X)
1	0.12	0.03	0.06	0.09	0.3
2	0.20	0.05	0.10	0.15	0.5
3	0.08	0.02	0.04	0.06	0.2
P(Y)	0.4	0.1	0.2	0.3	1

$$P(X,Y) = P(X)P(Y)$$

Independent

5-9 For part (b) we will need the distribution of V anyhow (to calculate its standard deviation), so we might as well calculate it from the beginning. For part (d), we also derive the marginal distribution of D and of H:

d	h = 20	25	p(d)
1.00	.16	.09	.25
1.25	.15	.30	.45
1.50	.03	.17	.20
1.75	.00	.10	.10
p(h)	.34	.66	1.00

a.

v	p(v)	vp(v)
8.00	.16	1.28
10.00	.09	.90
12.50	.15	1.88
15.62	.30	4.69
18.00	.032	.54
22.50	.17	3.82
30.63	.10	3.06
		μ = 16.17

a.b. Using the distribution of V above, our calculator gives $\mu = 16.17$, $\sigma = 6.63$.

c. Since $\mu = \dfrac{\text{population total}}{\text{population size}}$ like (2-4)

Thus population total = μ (population size)

= 16.17 (380) = 6145 cubic feet

d. From the marginal distribution in the very first table above, our calculator gives E(D) = 1.288 and E(H) = 23.3. Or, here is the detail by hand: We start by copying down the marginals in the table above:

d	p(d)	dp(d)
1.00	.25	.25
1.25	.45	.563
1.50	.20	.30
1.75	.10	.175
		E(D)=1.288

h	p(h)	hp(h)
20	.34	6.80
25	.66	16.50
		E(H) = 23.30

Will $E(V) = .4[E(D)]^2E(H)$? Since volume is a nonlinear function, we wouldn't expect it to be true. Calculation bears this out:

From part (a), E(V) = 16.17
From part (d), $.4\,[E(D)]^2 = .4\,(1.288)^2(23.30) = 15.46$

5-10

The joint distribution

X \ Y	1	2	3	P(X)
1	1/9	1/9	1/9	1/3
2	1/9	1/9	1/9	1/3
3	1/9	1/9	1/9	1/3
P(Y)	1/3	1/3	1/3	1

(a)

$$E(X)=\frac{1}{3}(1+2+3)=2$$

$$\text{var}(X)=(1-2)^2*1/3+(2-2)^2*1/3+(3-2)^2*1/3=2/3$$

$$\sigma_x=\sqrt{2/3}=0.82$$

(b)

$$E(Y)=\frac{1}{3}(1+2+3)=2$$

$$\text{var}(Y)=(1-2)^2*1/3+(2-2)^2*1/3+(3-2)^2*1/3=2/3$$

$$\sigma_y=\sqrt{2/3}=0.82$$

(c)

$$E(X+Y)=\frac{1}{9}*2+\frac{2}{9}*3+\frac{3}{9}*4+\frac{2}{9}*5+\frac{1}{9}*6=4$$

$$\text{var}(X+Y)=(2-4)^2*1/9+(3-4)^2*2/9+(4-4)^2*3/9+(5-4)^2*2/9+(6-4)^2*1/9=4/3$$
$$\sigma_{x+y}=\sqrt{4/3}=1.155$$

(d)

(i) Yes, E(S)=E(X)+E(Y)=4

(ii) Yes, Var(S)=Var(X)+Var(Y)=4/3

(iii) No, $\sigma_{x+y}\neq\sigma_x+\sigma_y$

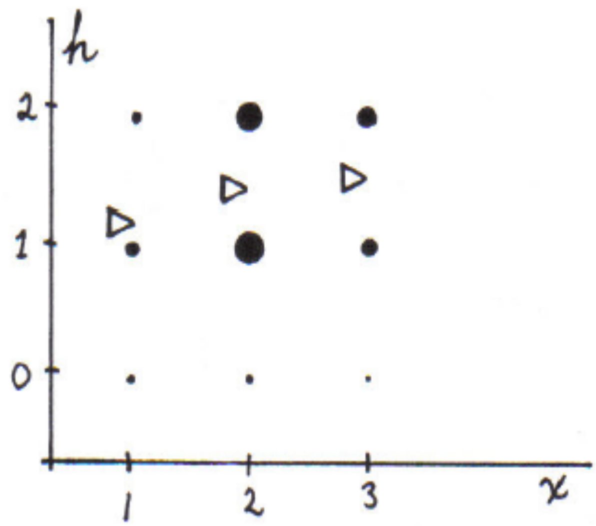
5-13

a. In part (c) of Problem 5-12, $\rho = 0$, and X and Y are independent. In part (d), $\rho = 0$, yet X and Y are not independent.

b. Statement (1) is true, in fact is just the statement (5-12). It is illustrated by part (c) of Problem 5-12, where X and Y were independent. Then $\rho = 0$, as promised.
Statement (2) is false, as part (d) of Problem 5-12 shows: $\rho = 0$, yet it did not follow that X and Y were independent.

$E(H|X=2) = 1.42$
$E(H|X=3) = 1.50$

As X increases, note that H slightly increases on average, as the graph will illustrate:



b.

x	h: 0	1	2	p(x)
1	.02	.08	.05	.15
2	.02	.28	.25	.55
3	.01	.13	.16	.30
p(h)	.05	.49	.46	1.00 √

From the marginal distribution our calculator gives:

$\mu_X = 2.15, \quad \sigma_X = .654, \quad \mu_H = 1.41, \quad \sigma_H = .585$

For σ_{XH}, we likewise calculate E(XH) from the table of xhp(x,h):

x	h: 0	1	2
1	0	.08	.10
2	0	.56	1.00
3	0	.39	.96

E(XY) = 3.09

Thus $\sigma_{XH} = E(XH) - \mu_X \mu_H$ like (5-11)

$= 3.09 - (2.15)(1.41) = .0585$

Finally $\rho = \frac{\sigma_{XH}}{\sigma_X \sigma_H} = \frac{.0585}{(.654)(.585)} = .153$

c. (i) True

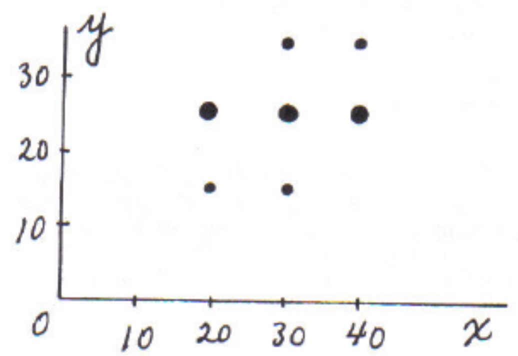
5-17

a. Tally of the frequency:

x	y: 15	25	35
20	/	//	
30	/	//	/
40		//	/

This yields the following relative frequency (probability) distribution, and marginal distributions for part b:

x	y: 15	25	35	p(x)
20	.1	.2	0	.3
30	.1	.2	.1	.4
40	0	.2	.1	.3
p(y)	.2	.6	.2	



b. From the margins of the bivariate distribution in part (a), our calculator gives:

$\mu_X = 30.0, \quad \sigma_X = 7.75, \quad$ hence $\sigma_X^2 = 60$

$\mu_Y = 25.0, \quad \sigma_Y = 6.32, \quad$ hence $\sigma_Y^2 = 40$

c. Since the means are such easy numbers, we will use the definition $\sigma_{XY} = E(X-\mu_X)(Y-\mu_Y)$:

Table of $(X-\mu_X)(Y-\mu_Y)p(x,y)$

x \ y	15	25	35
20	10	0	0
30	0	0	0
40	0	0	10

σ_{XY} = sum = 20

d. First we shall do it the hard way (so we can appreciate the easy way later). We begin by deriving the distribution of S from the bivariate distribution:

x \ y	15	25	35
20	.1	.2	
30	.1	.2	.1
40		.2	.1

s	p(s)
35	.1
45	.3
55	.2
65	.3
75	.1

hence $\mu_S = 55$, $\sigma_S = 11.8$

By contrast, the easy way is to use the individual means and variances:

$$E(X + Y) = E(X) + E(Y) = 30 + 25 = 55 \qquad (5\text{-}16)$$
$$\text{var}(X + Y) = \text{var } X + \text{var } Y + 2 \text{ cov}(X, Y) \qquad (5\text{-}18)$$
$$= 60 + 40 + 2(20) = 140$$

e.
$$E(.6X + .8Y) = .6\,E(X) + .8\,E(Y) \qquad \text{like } (5\text{-}17)$$
$$= .6(30) + .8(25) = 38$$
$$\text{var}(.6X + .8Y) = .6^2 \text{ var } X + .8^2 \text{ var } Y + 2(.6)(.8) \text{ cov}(X,Y) \qquad (5\text{-}19)$$
$$= .36(60) + .64(40) + .96(20) = 66.4$$

f.
$$E(X-Y) = E(X) - E(Y) = 30 - 25 = 5.0 \qquad \text{like } (5\text{-}17)$$
$$\text{var}(X-Y) = 1^2 \text{ var } X + (-1)^2 \text{ var } Y + 2(1)(-1) \text{ cov}(X,Y) \qquad (5\text{-}19)$$
$$= 60 + 40 - 2(20) = 60$$

g. Although it is incontestable that wives earn less (by 5 thousand dollars) than husbands on the average, there is no evidence given that this is <u>because</u> of sex discrimination. There are many other possible explanations for wives having a lower average income in an observational study like this. For example, wives may be younger on average (and people tend to earn more as they gain experience).

5-18

By 5-17, S=X+Y E(S)=55 Var(S)=140

(a)

T=0.2S

E(T)=0.2E(S) =(1/5)*55=11

$\text{Var(T)}=(0.2)^2\text{Var(S)}=0.04*140=5.6$

$\sigma_T = \sqrt{5.6} = 2.37$

(b)

T=0.5S-7.5

E(T)=0.5E(S)-7.5=(1/2)*55-7.5=20

$\text{Var(T)}=(0.5)^2\text{Var(S)}=0.25*140=35$

$\sigma_T = \sqrt{35} = 5.92$

(c)

$$E(T) = \frac{1}{10}*5 + \frac{3}{10}*7 + \frac{2}{10}*11 + \frac{3}{10}*16 + \frac{1}{10}*22 = 11.8$$

$$\text{var}(T) = (5^2*0.1 + 7^2*0.3 + 11^2*0.2 + 16^2*0.3 + 22^2*0.1) - 11.8^2 = 27.36$$

$\sigma_y = \sqrt{27.36} = 5.23$

Chap 6

6-6

(a)

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0.8$$

(b)

The same as part (a) since SE only depends on sample size, not the population size.

6-7 **a.** $SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{780,000}} = .009$ thousand = \$9

b. $SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{78,000}} = .029$ thousand = \$29

Note that it is n that makes the difference, not the sampling fraction (1% in both cases.)

6-9 **a.** ... an expected value of μ = \$30,000 with a standard error of $\sigma/\sqrt{n}$ = \$9000/$\sqrt{25}$ = \$1800, and with a distribution shape that is normal.

b. "10% high" is an income of: 30,000 + 30,000 x .10 = \$33,000

To calculate the chance of exceeding this, we standardize it:

$$Z = \frac{\bar{X} - \mu}{SE} = \frac{33,000 - 30,000}{9,000/\sqrt{25}} = 1.67 \qquad (6\text{-}11)$$

$$\Pr(\bar{X} > 33,000) = \Pr(Z > 1.67) = .047$$

6-10

$\mu = 470 \quad \sigma = 120$

(a)

$$\Pr(X > 500) = \Pr(Z > \frac{500-470}{120}) = \Pr(Z > 0.25) = 0.401$$

(b)

$$n = 250 \quad SE = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{250}} = 7.6$$

$$\Pr(460 < \overline{X} < 480) = \Pr(\frac{460-470}{7.6} < Z < \frac{480-470}{7.6}) = \Pr(-1.32 < Z < 1.32) = 0.814$$

6-11 $\bar{X}$ has an approximately normal distribution, with mean μ = 16.20 and standard error = $\sigma/\sqrt{n}$ = $.12/\sqrt{n}$. We use these to standardize the critical value $\bar{X}$ = 16:

$$Z = \frac{\bar{X} - \mu}{SE} \qquad (6\text{-}11)$$

$$= \frac{16 - 16.20}{.12/\sqrt{n}} = -1.67\sqrt{n}$$

a. n = 1, and so Z = -1.67. Thus

$$Pr(\bar{X} < 16) = Pr(Z < -1.67) = .047$$

b. n = 4, $Pr(Z < -3.33) < .00048$

c. n = 16, $Pr(Z < -6.67) \simeq 0$

Notice that increasing the sample size decreases the probability of drawing a discrepant sample with $\bar{X}$ < 16.0.

6-13

a. Our calculator gives μ = 34.0, σ = 8.0

b. From 9 to 5 pm is 8 hours. The number of minutes available for 4 men is therefore:

8 x 60 x 4 = 1920 minutes.

For 50 jobs, this is an average of:

$$\bar{X} = \frac{1920}{50} = 38.4 \text{ minutes per job}$$

To find the chance of exceeding this critical average, we first standardize:

$$Z = \frac{\bar{X} - \mu}{SE} = \frac{38.4 - 34.0}{8.00/\sqrt{50}} = 3.9$$

Thus $Pr(\text{Total} > 1920) = Pr(\bar{X} > 38.4)$

$$= Pr(Z > 3.9)$$

$$= .000048$$

c. We assumed the 50 mufflers were a <u>random</u> sample from the year's given record. But some cold and wet winter days, for example, may be very time consuming days for mufflers; then the chance of not finishing on time would be much higher.

Also, some workers tend to be slower. On the days when slow workers are scheduled, watch out again.

Of course, we assumed myriad other things, such as no time off for coffee or lunch, etc.

6-16

$\pi = 0.02 \quad n = 1000$

$SE = \sqrt{\pi(1-\pi)/n} = 0.0044$

$$\Pr(P > 40/1000) = \Pr(P > 0.04) = \Pr(Z > \frac{0.04 - 0.02}{0.0044}) = \Pr(Z > 4.5) = 0.0000034$$

6-17 We express the total as a proportion: P = 5/50 = .10. To standardize, we use π = .20:

$$Z = \frac{P - \pi}{SE} = \frac{.10 - .20}{\sqrt{\frac{.20(.80)}{50}}} = -1.77$$

Pr(P < .10) = Pr(Z < -1.77) = .038

For the continuity correction, we can re-express "no more than 5" as "fewer than 6". Taking 5 1/2 as our compromise, we finally get:

Pr(P < 5.5/50) = Pr(Z < -1.59) = .056

6-19 An erroneous prediction occurs if the proportion of Republicans is less than a majority: P < .50. To find the chance of this, we standardize using π = .539:

$$Z = \frac{P - \pi}{SE} = \frac{.50 - .539}{\sqrt{\frac{.539(.461)}{1000}}} = -2.47$$

Pr(P < .50) = Pr(Z < -2.47) = .007

For such a large n, the continuity correction is hardly worthwhile.

6-20

(a) n=10 P=1/2

$$\Pr(X \geq 7) = C_7^{10}(\frac{1}{2})^7(\frac{1}{2})^3 + C_8^{10}(\frac{1}{2})^8(\frac{1}{2})^2 + C_9^{10}(\frac{1}{2})^9(\frac{1}{2})^1 + C_{10}^{10}(\frac{1}{2})^{10}(\frac{1}{2})^0 = 0.172$$

(b) $\pi = 0.5 \quad n = 10$

$SE = \sqrt{\pi(1-\pi)/n} = 0.158$

$$\Pr(P \geq 7/10) = \Pr(P \geq 0.7) = \Pr(Z \geq \frac{P-\pi}{SE}) = \Pr(Z \geq 1.26) = 0.104$$

(c) $\pi = 0.5 \quad n = 10$

$$SE = \sqrt{\pi(1-\pi)/n} = 0.158$$

$$\Pr(P \geq 6.5/10) = \Pr(P \geq 0.65) = \Pr(Z \geq \frac{P-\pi}{SE}) = \Pr(Z \geq 0.95) = 0.171$$

Chap 7

7-1 a. True

b. Interchange μ and $\bar{X}$

c. If we quadruple the sample size ...

d. True

7-6

(a)Yes.

$E(X) = 8*0.25 + 10*0.25 + 11*0.5 = 10$

(b)Yes.

$E(Y) = 4*0.5 + 6*0.5 = 5$

(c)Yes.

If X, Y are independent, then we have

E[XY]=E[X]E[Y]

E[A]=10*5=E[X]E[Y]

7-9 MSE = variance of estimator + bias^2

$$= \frac{\sigma^2}{n} + \text{bias}^2$$

Presumably only one measurement is taken, so n=1 and the MSE reduces to:

$$\text{MSE} = \sigma^2 + \text{bias}^2$$

for A, $= 10^2 + 0 = 100$
for B, $= 0 + (-10)^2 = 100$
for C, $= 5^2 + 5^2 = 50$
for D, $= 8^2 + 2^2 = 68$

So gauge C has greatest accuracy.

Remarks Note that if there are 10 units of error to be split between bias and standard deviation, the greatest accuracy comes from the compromise 5/5 split.

7-10

(a)

$\mu = (0*0.04)+(1*0.24)+(2*0.2)+(3*0.12)+(4*0.04)=1.16$

(b)

Case (i)

$\overline{X} = (0*0.62)+(1*0.21)+(2*0.12)+(3*0.04)+(4*0.01)=0.61$

Bias=0.61-1.16=-0.55

Case (ii)

$\overline{X} = (0*0.04)+(1*0.24)+(2*0.2)+(3*0.12)+(4*0.04)=1.16$

Bias=1.16-1.16=0

(c)

Case (i)

$Var(\overline{X}) = (1*0.21)+(4*0.12)+(9*0.04)+(16*0.01)-0.61^2 = 1.15$

MSE=1.15/200+$(-0.55)^2$=0.30825

Case (ii)

$Var(\overline{X}) = (1*0.24)+(4*0.2)+(9*0.12)+(16*0.04)-1.16^2 = 1.41$

MSE=1.41/25+0=0.0564

Case (ii) is more accuracy.