Econ 300/QAC 201: Quantitative Methods in Economics/Applied Data Analysis

6th Class **6/14/10**

“Statistics are no substitute for judgment.”
Henry Clay

Econ 300: hand out review sheet and answer sheet for test and direct to old test on-line

[show overhead re continuous distributions]

review the discussion of decision making from last time, including risk aversion and repeated game issues

[play a round of "deal or no deal" and discuss relative risk aversion]
[calculate expected value for the later rounds]
sum of 13 on the left side = $2416.01, avg = $185.85
sum of 13 on the right side=$3,416,000, avg = $262,769
sum of all 26 = $3,418,416.01, expected value = avg = $131,477.54
[discuss why there is no set formula for the banker's offers; however pattern is to lowball on the early rounds to keep the game going and in general offer below the expected value]
point out economics paper link on website that uses deal or no deal to study risk aversion

Ch. 5: [important points numbered and kept on board]
--Joint distributions
e.g. X = number of girls and Y = number of runs

<u>Old terminology</u>	<u>New terminology</u>
Pr(G and H) applied to:	Joint [probability] distribution:
Pr(X=1 and Y=2)	p(1,2)
Pr(X=x and Y=y)	p(x,y)

Event G is independent of H if:

$\Pr(G \text{ and } H) = \Pr(G)\Pr(H)$

Variable X is independent of Y if, for all x and y:

$p(x,y) = p(x)p(y)$

--marginal distributions can be calculated from the joint distribution:

1) $p(x) = \sum_{y} p(x,y)$ and $p(y) = \sum_{x} p(x,y)$

these are just the ordinary distributions of X and Y which could be found without reference to the joint distribution (but sometimes one is given the joint distribution to work with) --are placed "in the margins" of the joint distribution table

--can also calculate conditional probabilities (ones inside the table divided by the marginal probabilities) and expected values from the joint distribution

--independence

2) X and Y are [statistically] independent if $p(x,y) = p(x)p(y)$ for all x and y

if given the joint distribution table, have to test this by first calculating the marginal distributions and then multiplying them together to check against each item in the joint distribution table

possible short-cut: if X and Y are independent, rows in the joint distribution will be proportional to each other; so will columns

joint distributions can then be used to calculate the expected value of functions of the two random variables:

3) if $G = g(X,Y)$, then

$$E(G) = \sum g\, p(g), \text{ and equivalently:}$$

$$E[g(X,Y)] = \sum g(x,y)p(x,y)$$

the covariance describes how two variables X and Y vary together, and it is calculated as a function of X and Y:

4) $$\mathrm{Cov}(X,Y) = \sigma_{XY} \equiv E[(X - \mu_X)(Y - \mu_Y)]$$

this can be calculated (often more simply) by using the formula:

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y$$

[show how the definition of covariance reduces to this]

Covariance can be positive, negative, or zero.
If X and Y are independent, then $\sigma_{XY} = 0$ and we say that they are uncorrelated. However, σ_{XY} can be 0 and X and Y may still be related--the positive and negative calculations may just cancel each other out. Then they are uncorrelated, but not necessarily independent.

We usually use a unit-free measure of correlation instead, the correlation coefficient ρ, where ρ is bounded by -1 and 1 inclusive:

5) $$\rho \equiv \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

[show how 1, -1, and 0 cases are calculated]

can also calculate conditional probabilities using the joint distribution and the marginal distributions that are derived from it:

6) $$p(x \mid y) = \frac{p(x,y)}{p(y)} \text{ and } p(y \mid x) = \frac{p(x,y)}{p(x)}$$

While we can use the general method sketched above for calculating the expected value of a function of two random variables, it is easy to calculate the mean and variance of linear combinations of two random variables X and Y:

7) $$E(aX + bY) = aE(X) + bE(Y) ;$$

$Var(aX + bY) = a^2 VarX + b^2 VarY + 2ab\ Cov(X,Y)$

note if X and Y are independent, the last term drops off of the variance expression

example1: relationship between grades in different courses

grade in econometrics E: 3 (B), 4 (A)
grade in statistics S: 3 (B), 4 (A)

	e	
s	3	4
3	.4	.1
4	.2	.3

add marginal distributions, or <u>unconditional</u> distributions:

	e		
s	3	4	p(s)
3	.4	.1	.5
4	.2	.3	.5
p(e)	.6	.4	

then can test for dependence of E and S:
does p(e,s) = p(e)p(s) for all e and s? If one answer is no, then dependent

can use these marginal distributions to calculate [unconditional] expectations:

E(S) = 3.5; E(E) = 3.4

and [unconditional] variances and standard deviations:

$\sigma_S^2 = .25$ and $\sigma_S = .50$; $\sigma_E^2 = .24$ and $\sigma_E = .49$

can calculate table of conditional distribution too for either way:

p(e | s):

s	e = 3	e = 4
3	.8	.2
4	.4	.6

and p(s | e):

s	e = 3	e = 4
3	.67	.25
4	.33	.75

can then use these tables to calculate conditional expectations:

e.g., $E(E \mid S = 3) = \sum_e eP(e \mid 3) = 3*.8 + 4*.2 = 3.2$

and $E(E \mid S = 4) = \sum_e eP(e \mid 4) = 3*.4 + 4*.6 = 3.6$

[note. can also get conditional variances and standard deviations:

$\sigma_{E \mid S=4}{}^2 = .24$ and $\sigma_{E \mid S=4} = .49$]

can also calculate covariance and correlation between econ and stat grades:

need to calculate E(SE) using the joint distribution:

$E(SE) = 9*.4 + 12*.3 + 16*.3 = 12$

then $\sigma_{SE} = E(SE) - E(S)E(E) = 12 - 3.5*3.4 = 12 - 11.9 = .1$

and $\rho = \frac{.1}{.50*.49} = .41$ [note that scaling doesn't always reduce size of correlation below that of the covariance]

can also calculate GPA for the two classes: G = .5S + .5E

since this is a linear function of S and E, can use simple formulas to get the mean and variance of GPA:

$$E(G) = .5E(S) + .5E(E) = .5*3.5 + .5*3.4 = 3.45$$

and $Var(G) = .25Var(S) + .25Var(E) + .5Cov(E,S)$

$$= .25*.25 + .25*.24 + .5*.1 = .1725$$

example2:

Independence: knowing grade in stat contains info on grade in econ, but other course grades may contain no info, e.g. French grade on econ. grade:

	e		
f	3	4	p(f)
3	.3	.2	.5
4	.3	.2	.5
p(e)	.6	.4	

rows and columns are proportional, conditional distribution equals the marginal, p(ef) = p(e)p(f) for all e and f

the covariance and correlation will be 0 between econ and French grades

example3: does knowledge of statistics make you successful?

success scale S: 0 (not successful), 1 (mildly successful), 2 (pretty great)
grade in stats G: 2 (C), 3 (B), 4 (A)

		s	
g	0	1	2
2	.25	.15	0
3	.10	.15	.10
4	0	.05	.20

question: what is $E(S|G)$ for the various levels of G?

first derive the marginal and conditional probabilities from the above table

add column of marginal probability p(g):

		s		
g	0	1	2	p(g)
2	.25	.15	0	.40
3	.10	.15	.10	.35
4	0	.05	.20	.25

table of conditional probabilities:

		s	
g	0	1	2
2	.625	.375	0
3	.2855	.429	.2855
4	0	.20	.80

$E(S \mid G) = \sum_{g} s\, p(s \mid g)$

$E(S \mid G=2) = .375$

$E(S \mid G=3) = 1.00$

$E(S \mid G=4) = 1.8$

so we observe a positive relationship between S and G

question: what is the correlation between S and G?

first find the covariance using E(SG) - E(S)E(G)

find E(SG) using the joint distribution table: E(SG) = 3.15

E(SG) = 2*.15 + 3*.15 + 6*.10 + 4*.05 + 8*.20 = 3.15

using the marginal distribution for G, E(G) = 2.85

E(G) = 2*.40 + 3*.35 + 4*.25 = 2.85

add the row for p(s) to the joint distribution table:

		s		
g	0	1	2	p(g)
2	.25	.15	0	.40
3	.10	.15	.10	.35
4	0	.05	.20	.25
p(s)	.35	.35	.30	

[note it is clear that S and G are not independent--already knew that because $E(S|G)$ was rising with higher values of G--but now can formally show that by multiplying the marginal probabilities together, e.g. $p(g)p(s) \neq p(g,s)$ for p(2,0)]

and calculate E(S) = .95

so COV(SG) = 3.15 - (.95)(2.85) = 0.4425

then scale it using the two s.d.s using the formula for variance $E(X^2) - [E(X)]^2$

calculate $E(G^2)$ and $E(S^2)$ using the marginal distributions:

$E(G^2) = 4*.4 + 9*.35 + 16*.25 = 8.75$ and $E(S^2) = 1.55$

so $\sigma_G^2 = 8.75 - 8.1225 = .6275$ and $\sigma_G = .792$; $\sigma_S^2 = 1.55 - .90 = .65$ and $\sigma_S = .806$

and $\rho = \frac{.4425}{.792*806} = .693$

so does knowledge of stats make you more successful?

Note how statistical calculators work: they store 6 sums as you enter pairs of numbers (X,Y):

$\sum X$
$\sum Y$
$\sum X^2$
$\sum Y^2$
$\sum XY$

and n

Then when you are done entering pairs of numbers, the calculator can give you means, variances, and covariance using these six sums and the following easy formulas:

$$\mu_X = \frac{\sum X}{n}$$

$$\mu_Y = \frac{\sum Y}{n}$$

$$\sigma_X^2 = \frac{n}{n-1} \left[\frac{\sum X^2}{n} - (\mu_X)^2\right]$$ (many calculators let you choose whether to get the mean/expected squared deviation or the variance)

$$\sigma_Y^2 = \frac{n}{n-1} \left[\frac{\sum Y^2}{n} - (\mu_Y)^2\right]$$

$$\sigma_{XY} = \frac{\sum XY}{n} - \mu_X\mu_Y$$

one can then also get standard deviations and the correlation coefficient, often as built-in functions

note in the preceding examples we didn't know n, the actual number of students, just the relative frequencies of grade patterns, so we can only calculate the mean squared deviation, but for n fairly large it doesn't matter; in general since economists deal with large samples we tend to be sloppy about this

for next class read Ch. 6