K. Two Random Variables.

1. Regression (Summary).

2. Covariance (σ_{xy} and s_{xy})

a. Population Covariance

The population covariance is defined, using probability, as
$Cov(x,y) = \sigma_{xy} = E[(x-\mu_x)(y-\mu_y)] = E(xy) - \mu_x\mu_y$. This can be used to describe the relationship between x and y. If the covariance is positive we can say that x and y tend to move together, while if it is negative we can say that they tend to move in opposite directions. In order to use this formula we must realize that $E(xy) = \Sigma\Sigma(xyP(x,y))$. This means that we must add together the product of x and y, together with their joint probability, for each possible pair of values of x and y. For example, assume that x and y are related by the following joint probability table:

		x: 400	600	800
y	400	.12	.15	.18
	600	.10	.05	.08
	800	.16	.07	.09

We begin by taking the upper left hand probability, .12, which is the probability that both x and y are 400, and multiplying it by 400 twice. Then we take the next probability in the same row, .15, which is the probability that x is 600 and y is 400, and multiply it by both 600 and 400. If we continue in this way we get

$$E(xy) = \sum\sum xyP(xy) = \begin{bmatrix} .12(400)(400) & +.15(600)(400) & +.18(800)(400) \\ +.10(400)(600) & +.05(600)(600) & +.08(800)(600) \\ +.16(400)(800) & +.07(600)(800) & +.09(800)(800) \end{bmatrix}$$

$$= \begin{bmatrix} 19200 & +36000 & +57600 \\ +24000 & +18000 & +38400 \\ +51200 & +33600 & +57600 \end{bmatrix} = 335600.$$

We can now use the following tableau to compute the means and variances of x and y.

		x					
		400	600	800	$P(y)$	$yP(y)$	$y^2P(y)$
	400	.12	.15	.18	.45	180	72000
y	600	.10	.05	.08	.23	138	82800
	800	.16	.07	.09	.32	256	204800
	$P(x)$	.38	.27	.35	1.00	574	359600
	$xP(x)$	152 +	162 +	280 =	594		
	$x^2P(x)$	60800 +	97200 +	224000 =	382000		

To summarize $\sum P(x) = 1$ (a check), $\mu_x = E(x) = \sum xP(x) = 594$, $E(x^2) = \sum x^2 P(x) = 382000$, $\sum P(y) = 1$, $\mu_y = E(y) = \sum yP(y) = 574$ and $E(y^2) = \sum y^2 P(y) = 359600$

We will need the variances below. To complete what we have done, write

$$\sigma_{xy} = Cov(xy) = E(xy) - \mu_x \mu_y = 335600 - (594)(574) = -5356$$

b. The Sample Covariance

The sample covariance is much easier to compute, the formula being

$$s_{xy} = \frac{\sum[(x - \bar{x})(y - \bar{y})]}{n-1} = \frac{\sum xy - n\bar{x}\bar{y}}{n-1}.$$

For example, assume that we have data on income (X) and savings (Y)(in thousands) for 5 families.

Family	x	y	x^2	y^2	xy
1	1.9	0.0	3.61	0.00	0.00
2	12.4	0.9	153.76	0.81	11.16
3	6.4	0.4	40.96	0.16	2.56
4	7.0	1.2	49.00	1.44	8.40
5	7.0	0.3	49.00	0.09	2.10
Sum	34.7	2.8	296.33	2.50	24.22

Then $\bar{x} = \frac{34.7}{5} = 6.94$ and $\bar{y} = \frac{2.8}{5} = 0.56$.

$$s_x^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{296.33 - 5(6.94)^2}{4} = 13.878,$$

$$s_y^2 = \frac{\sum y^2 - n\bar{y}^2}{n-1} = \frac{2.50 - 5(0.56)^2}{4} = 0.2330$$ and since

$\sum xy = 24.22$, $s_{xy} = \frac{24.22 - 5(6.94)(0.56)}{5-1} = 1.197$.

The positive sign of s_{xy}, the sample covariance, indicates that X and Y tend to move together.