

Math 1180:Lab7

Kyle Gaffney

February 19th, 2014

1 Covariance

Covariance is a measure comparing two random variables. A positive covariance means that both random variables tend to be large or small in value at the same time, a negative covariance on the other hand indicates that one variable tends to be large while the other is small.

$$Cov(X, Y) = \sum_{j=1}^m \sum_{i=1}^n (x_i - \bar{X})(y_j - \bar{Y})p_{ij} = \sum_{j=1}^m \sum_{i=1}^n x_i y_j p_{ij} - \bar{X}\bar{Y}$$

In order to calculate the covariance we are going to need a joint distribution, the marginal distribution and the expected values. Consider the example from last lab of the dart throwers:

For example let us consider the following situation for dart players which perform differently under pressure. Each player independent of the other will hit the board (score=1) with probability 0.8. However player 1 gets nervous under pressure and this causes him to only hit the board 40 percent of the time if his opponent has already hit the board. Player 2 on the other hand excels under pressure and increases his hit percentage to 90 percent when his opponent hits the board first. Let us construct data.frames describing the joint probabilities.

Case 1: Player 1 goes first.

$$Pr(S_1 = 1 \text{ and } S_2 = 1) = Pr(S_1 = 1) * Pr(S_2 = 1|S_1 = 1) = (0.8) * (0.9) = 0.72$$

$$Pr(S_1 = 0 \text{ and } S_2 = 1) = Pr(S_1 = 0) * Pr(S_2 = 1|S_1 = 0) = (0.2) * (0.8) = 0.16$$

$$Pr(S_1 = 1 \text{ and } S_2 = 0) = Pr(S_1 = 1) * Pr(S_2 = 0|S_1 = 1) = (0.8) * (0.1) = 0.08$$

$$Pr(S_1 = 0 \text{ and } S_2 = 0) = Pr(S_1 = 0) * Pr(S_2 = 0|S_1 = 0) = (0.2) * (0.2) = 0.04$$

Which we can construct in R as a data.frame with the following commands:

```
P1S1=c(0.72,0.08)
P1S0=c(0.16,0.04)
P1goesfirst=data.frame(P1S1,P1S0,row.names=c('P2S1','P2S0'))
```

Now like last time let us generate the Marginal distributions and add them to the data.frame:

```

P1S1=c(P1S1,sum(P1S1))
P1S0=c(P1S0,sum(P1S0))
Marginal=c(sum(P1goesfirst[1,]),sum(P1goesfirst[2,]),1)
P1gf=data.frame(P1S1,P1S0,Marginal,row.names=c('P2S1','P2S0','Marginal'))

```

We have now found the Joint Distributions and the Marginal Distributions for the dart problem, all that we have left is to calculate the expected values for the scores of Players 1 and 2.

```

ESP1=1*P1gf[3,1]+0*P1gf[3,2]
ESP2=1*P1gf[1,3]+0*P1gf[2,3]

```

Now let us use our expected values and our data frame of our joint and marginal distributions to calculate the Covariance.

```

COV=0
for (i in 1:2){
for (j in 1:2){COV=COV+P1gf[i,j]*(2-i)*(2-j)}}
COV=COV-ESP1*ESP2

```

Is this a value that you would expect? What does this tell you about the 2 players?

Another way to view this is like the tables in the book. Let us attempt to recreate one of these for our example. Our columns will be S1, S2, Product, and Joint Probability.

```

S1=0
S2=0
Product=0
Joint=0
n=2
for (i in 1:n){
for (j in 1:2){
S1[(i-1)*n+j]=2-j
S2[(i-1)*n+j]=2-i
Product[(i-1)*2+j]=S1[(i-1)*2+j]*S2[(i-1)*2+j]
Joint[(i-1)*2+j]=P1gf[i,j]}}}

```

Then to calculate the Covariance we can use the following command

```
COV2=sum(Product*Joint)-ESP1*ESP2
```

How does this calculation compare?

2 Correlation

Correlation is a scaled Covariance.

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Where σ_X, σ_Y are the standard deviations of the random variables X, Y

How to find the Correlation of 2 random variables:

1. Find the Marginal Distributions of each variable
2. Compute the expectation of each from the Marginal Distribution
3. Compute the variance of each from the Marginal Distribution, and take the square root of the variance to get the standard deviation.
4. Compute the covariance using a formula like the one above
5. Divide the covariance by the product of the standard deviations to find the correlation.

Recall theorem 6.4: $\sigma_X^2 = E(X^2) - E(X)^2 = E(X^2) - \bar{X}^2 = \sum_{i=1}^n x_i^2 p_i - \bar{X}^2$

For our example problem we already calculated the covariance, so to get the correlation all we need to do is calculate the standard deviations of S_1 and S_2

$$\sigma_{S_1}^2 = \sum_{i=1}^2 x_i^2 p_i - \bar{X}_1^2 = 1^2 * .8 + 0^2 * 0.2 - 0.8^2 = .8 + 0 - .64 = .16$$

$$\sigma_{S_2}^2 = \sum_{i=1}^2 x_i^2 p_i - \bar{X}_2^2 = 1^2 * .88 + 0^2 * 0.12 - 0.88^2 = .88 + 0 - .7744 = .1056$$

So $\sigma_{S_1} = \sqrt{.16} = .4$ and $\sigma_{S_2} = \sqrt{.1056} = .325$

So our Correlation between the scores of Player 1 and Player 2 is:

$$\frac{.016}{.4 * .325} = 0.123$$

3 Assignment for the Week

Suppose you are given the following data.frame from last week describing data collected on a third bird, Bird C.

```
BirdCLO=c(.1, .2, .1)
```

```
BirdCL1=c(.1,.05,.1)
BirdCL2=c(.15,.1,.1)
C=data.frame(BirdCL0,BirdCL1,BirdCL2,row.names=c('Mites0','Mites1','Mites2'))
```

1. Using your marginal probabilities from before calculate the Covariance by hand. Then calculate it in R using either the table method or the for loop code both of which you will need to alter to work for this data.frame.
2. Calculate the Variances and Standard deviations for lice and mites.
3. Finally using your solutions to 1,2 calculate the correlation between lice and mites. What does that tell you about the relationship between lice and mites for this species of bird?