

Stat 324: Lecture 12

Correlation

Moo K. Chung
mchung@stat.wisc.edu

March 8, 2005

Concepts

1. The *conditional probability density* of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}.$$

The *conditional expectation* of Y given $X = x$ is

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy.$$

2. The covariance between X and Y is

$$\mathbf{Cov}(X, Y) = \mathbb{E}(X - \mu_x)(Y - \mu_y),$$

where $\mu_x = \mathbb{E}X, \mu_y = \mathbb{E}Y$.

Example. Mathematical achievement test scores (x_i) and calculus grades (y_i) for 10 college freshmen?

$x \parallel 39, 43, 21, 64, 57, 47, 28, 75, 34, 52$

$y \parallel 65, 78, 52, 82, 92, 89, 73, 98, 56, 75$

We will assume $p(x_i, y_i) = 0.1$ and $p(x_i, x_j) = 0$ for $i \neq j$.

```
> x<-c(39,43,21,64,57,47,28,75,34,52)
> y<-c(65,78,52,82,92,89,73,98,56,75)
> sum(x*y)/10-mean(x)*mean(y)
[1] 189.4
```

3. *Correlation coefficient* is given by

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}(X)\mathbf{Var}(Y)}}$$

```
> mean(x^2)-mean(x)^2
[1] 247.4
> mean(y^2)-mean(y)^2
[1] 205.6
> 189.4/(sqrt(247.4)*sqrt(205.6))
[1] 0.8397859
> cor(x,y)
[1] 0.8397859
```

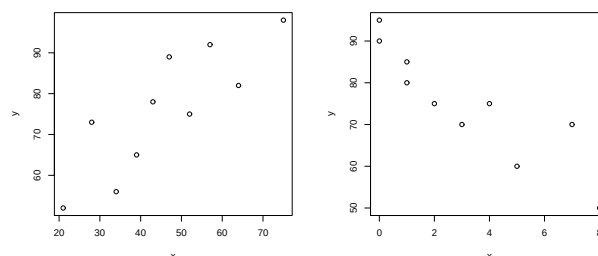


Figure 1: Left: calculus grade example showing positive correlation. Right: missing homework example showing negative correlation.

Example. Negative correlation. The number of times you miss submitting homeworks (x_i) and the final grade (y_i)?

$x \parallel 03, 04, 01, 02, 01, 00, 00, 07, 08, 05$

$y \parallel 70, 75, 80, 75, 85, 95, 90, 70, 50, 60$

```
> x<-c(03,04, 01,02,01,00,00,07,08,05)
> y<-c(70,75,80,75,85,95,90,70,50,60)
> cor(x,y)
[1] -0.8938321
```

4.

$$\mathbf{Cov}(aX + b, cY + d) = ac\mathbf{Cov}(X, Y).$$

$$\mathbf{Cov}(X, X) = \mathbf{Var}X.$$

$$\rho(aX + b, cY + d) = \rho(X, Y).$$

Problems

1. Suppose that the joint distribution of X and Y is a uniform distribution over $x^2 + y^2 \leq 1$. Determine the correlation coefficient of X and Y .

Solution. The joint density is $f(x, y) = \frac{1}{\pi}$, for $(x, y) \in \{(x, y) | x^2 + y^2 \leq 1\}$ and 0 otherwise.

The marginal density is

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi}, -1 \leq x \leq 1.$$

Since $f_X(x)$ is an even function, $xf_X(x)$ is odd so from symmetry, $\mathbb{E}X = 0$. Similarly, $\mathbb{E}Y = 0$. Also you should be able to show $\mathbb{E}(XY) = 0$. Therefore, $\rho = 0$.

2. Suppose that X and Y are identically distributed Bernoulli random variables with parameter $0 < p < 1$ such that $\text{Cov}(X, Y) = 0$. Show that X and Y are independent.

solution. The condition $\text{Cov}(X, Y) = 0$ is equivalent to $P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = p^2$. Then it follows $P(X = 1, Y = 0) = P(X = 1) - P(X = 1, Y = 1) = p(1 - p) = P(X = 1)P(Y = 0)$ and similarly for other cases. So we have $P(X = i, Y = j) = P(X = i)P(Y = j)$ for all $i, j = 0, 1$.

3. For random variables X and Y , prove or disprove that X and $Y - \mathbb{E}(Y|X)$ are correlated.

Solution.

$$\begin{aligned} \text{cov}(X, Y - \mathbb{E}(Y|X)) &= \mathbb{E}\{X(Y - \mathbb{E}(Y|X))\} - \mathbb{E}(X)\mathbb{E}\{Y - \mathbb{E}(Y|X)\} \\ &= \mathbb{E}(XY) - \mathbb{E}\{X\mathbb{E}(Y|X)\} - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}\{\mathbb{E}(Y|X)\} \\ &= \mathbb{E}(XY) - \mathbb{E}\{\mathbb{E}(XY|X)\} - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = 0 \end{aligned}$$

Assigned Problems: Please read Chapter 5. Exercise 5.32., 5.36. Second midterm is tentatively scheduled to be on April 5th. It will cover from Lecture 1.