

Stat 311-Lecture 14 Correlation

Nov 22, 2005

1. The covariance between X and Y is

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

where $\mu_x = \mathbb{E}X, \mu_y = \mathbb{E}Y$. It measures the relationship between two random variables.

2. Correlation coefficient is given by

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

. It is the normalized version of the covariance.

- 3.

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

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

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

4. 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, $\rho = 0$.

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

Solution. The condition $\mathbf{Cov}(X, Y) = 0$ 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 $i, j = 0, 1$.

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

Solution.

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

7. Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then are $\bar{X}$ and $X_j - \bar{X}$ independent?

solution. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Since $X_j - \bar{X}$ is a linear combination of X_i 's, it is also distributed as normal. Any uncorrelated normal distributions are independent. So we only need to check if $\mathbf{Cov}(\bar{X}, X_j - \bar{X}) = 0$.

$$\begin{aligned} & \mathbf{Cov}(\bar{X}, X_j - \bar{X}) \\ &= \mathbf{Cov}\left(\frac{1}{n} \sum_{i=1}^n X_i, X_j - \frac{1}{n} \sum_{k=1}^n X_k\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Cov}(X_i, X_j) - \frac{1}{n^2} \sum_{i,k=1}^n \mathbf{Cov}(X_i, X_k) \\ &= \frac{1}{n} \mathbb{V}(X_j) - \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) \\ &= \frac{1}{n} \mathbb{V}(X_j) - \frac{1}{n} \mathbb{V}(X_j) = 0 \end{aligned}$$

Note. 1. The second midterm exam result: mean 25.2 ± 8.6 . max 40, min 5. 12 students above 30. 18 students between 20 and 30. 11 students below 20.

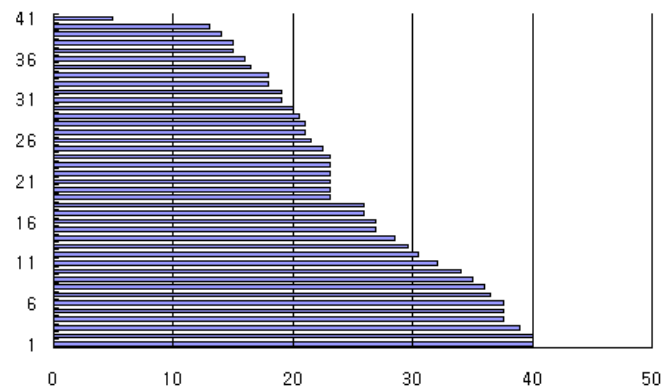


Figure 1: Number of students (vertical) below the given exam score (horizontal).

Note. 2. The final homework problem will be posted in the TA's class website by this Thursday. Solve 9 problems in Chapter 6 and 7. Due date: December 15 (Thursday).