
When exams are fast approaching but you have no idea what's going on in class because you spent the entire semester looking at memes
@BetaSalmon
[SCREAMING INTERNALLY]
ifunny.co

Today's Agenda

Last time:

- (Joint) expectations

Today (Lec 32, 07/20):

- Comments on the correlation coefficient
- Linear combinations

If $g : \mathbb{R}^n \to \mathbb{R}$, and $X_1, ..., X_n$ are discrete random variables with joint probability function $f(x_1, ..., x_n)$, then

$$\mathrm{E}\left[g(X_1, ..., X_n)\right] = \sum_{(x_1,...,x_n)} g(x_1, ..., x_n) f(x_1, ..., x_n).$$

Definition

If X and Y are jointly distributed, then $Cov(X, Y)$ denotes the **covariance** between X and Y. It is defined by

$$Cov(X, Y) = \mathrm{E}\left[(X - E(X))(Y - E(Y))\right].$$

Shortcut formula:

$$Cov(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

Definition

The **correlation** of X and Y is denoted $corr(X, Y)$, and is defined by

$$corr(X, Y) = \rho = \frac{Cov(X, Y)}{SD(X)SD(Y)}.$$

It follows from the Cauchy-Schwarz inequality that

$-1 \leq corr(X, Y) \leq 1$, and if $|corr(X, Y)| = 1$, $X = aY + b$.

Definition

We say that X and Y are uncorrelated if $Cov(X, Y) = 0$ (or $corr(X, Y) = 0$).

Remark:

- If X and Y are independent, then X and Y are uncorrelated.
- $Cov(X, X) = Var(X)$
- The correlation is unit-free.

Properties of correlation:

a) $\rho = corr(X, Y)$ has the same sign as $Cov(X, Y)$

b) $-1 \leq \rho \leq 1$

c) $|\rho| = 1 \implies X = aY + b$

d) X, Y independent $\implies corr(X, Y) = 0$

e) $corr(X, X) = cov(X, X)/SD(X)^2 = Var(X)/Var(X) = 1$

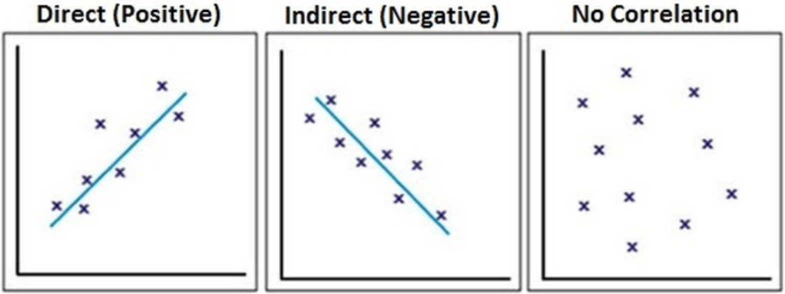
Properties 2 and 3 can be proved using Cauchy-Schwarz inequality.

ASIDE: While correlation is useful, it can be misleading as well.

This is more of a statistics issue than probability issue, but visualise the data first. Do not trust correlation blindly.

Take a look at the following link:
`https://www.autodeskresearch.com/publications/samestats`

Basic examples of correlation



The previous three plots are the “ideal” examples for correlation, because its value describes the pattern pretty accurately.

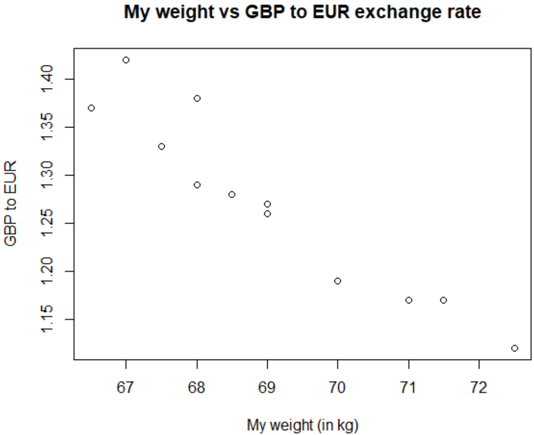
However, there are lots of cases where correlation itself can be very misleading.

This is more of a statistics issue than probability issue, but visualise the data first. Do not trust correlation blindly.

Take a look at the following link:
`https://www.autodeskresearch.com/publications/samestats`

“Correlation doesn’t imply causation”

Two variables being correlated does not always imply that one variable causes another to behave in certain ways.



Question

Suppose X, Y, and Z are jointly distributed random variables such that $corr(X, Y) = 1$, and $corr(Y, Z) \neq 0$. Which of the following is not necessarily true:

A X and Z are dependent

B $corr(X, Z) \neq 0$

C Y and Z are dependent

D $|corr(X, Z)| \neq |corr(Y, Z)|$

E $X = Y$

Definition

Suppose that $X_1, ..., X_n$ are jointly distributed RVs with joint probability function $f(x_1, ..., x_n)$. A **linear combination** of the RVs $X_1, ..., X_n$ is any random variable of the form

$$\sum_{i=1}^{n} a_i X_i$$

where $a_1, ..., a_n \in \mathbb{R}$. If $\mathbf{X} = (X_1, ..., X_n)^\top$, $\mathbf{a} = (a_1, ..., a_n)^\top$, then a linear combination is

$$\mathbf{X}^\top \mathbf{a}.$$

Example

Some "famous" linear combinations (ones you will see in STAT 231/241) are

a) The total

$$T = \sum_{i=1}^{n} X_i \quad a_i = 1, \;\; 1 \leq i \leq n$$

b) The sample mean

$$\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i \quad a_i = \frac{1}{n}, \;\; 1 \leq i \leq n$$

Expected Value of a Linear Combination

Theorem

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

This follows directly from the linearity of expected value.

Example

Let $P_1, P_2, \ldots, P_7$ represent the number of cans of pop that Harold drinks each day (from day 1 to day 7). If each random variable P_i has mean $\mu = 6$, what is the expected number of cans consumed per day during those 7 days?

Question

Suppose $X \sim N(1,1)$ and $Y \sim U(0,1)$. Compute $E(2X - 4Y)$.

A $E(2X - 4Y) = 2$

B $E(2X - 4Y) = -4$

C $E(2X - 4Y) = 3$

D $E(2X - 4Y) = -1$

E $E(2X - 4Y) = 0$

Bilinearity of Cov

Theorem
Let X, Y, U, V be random variables, and $a, b, c, d \in \mathbb{R}$. Then,

$$\begin{aligned} &Cov(aX + bY, cU + dV) \\ &= acCov(X, U) + adCov(X, V) + bcCov(Y, U) + bdCov(Y, V) \end{aligned}$$

Proof: Exercise.

Remark: You can generalise this to a linear combination of arbitrary length, but the formula becomes messy.

Variance of a linear combination

The following result shows how the variance of a linear combination is "broken down" into pieces.

Theorem

Let X, Y be random variables, and $a, b \in \mathbb{R}$. Then,

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

In general,

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j Cov(X_i, X_j)$$

Example

Let X, Y be independent random variables, and $a, b \in \mathbb{R}$. What is $Var(aX + bY)$?

Example

Suppose $X \sim N(3, 4)$, and $Y \sim U(0, 1)$, and $Cov(X, Y) = -0.1$, compute $Var(2X - Y)$.

An immediate, yet very useful, application of these results comes from a linear combination of normally-distributed random variables.

The following results about normal random variables are not only useful in STAT 230, but they will appear throughout your statistics education, and they are used in cutting-edge research very often.

Theorem ("A linear function of normal is normal")

Let $X \sim N(\mu, \sigma^2)$ *and* $Y = aX + b$*, where* $a, b \in \mathbb{R}$*. Then,*

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Theorem ("A linear combination of normal is normal")

Let $X_i \sim N(\mu_i, \sigma_i^2),\ \ i = 1, 2, \ldots, n$ **independently**, *and* $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{R}$. *Then,*

$$\sum_{i=1}^{n} a_i X_i + b_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i + b_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Theorem ("Sample mean of normal is normal")

Let $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ **independently**. *Then,*

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2),$$

and

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

It may not be obvious why we have $n\sigma^2$ and σ^2/n in the third theorem when $Var(aX) = a^2\sigma^2$. Let's see why.

Example

Suppose that $X_i \sim N(\mu, \sigma^2)$, and that $X_1, ..., X_n$ are independent. Compute $Var\left(\sum_{i=1}^{n} X_i\right)$ and $Var(\bar{X})$.

Example

Three cylindrical parts are joined end to end to make up a shaft in a machine: 2 type-A parts and 1 type-B part. The lengths of the parts vary a little, and have the following distributions:

$$A \sim N(6, 0.4), \quad B \sim N(35.2, 0.6).$$

The overall length of the assembled shaft must lie between 46.8 and 47.5 or else the shaft has to be scrapped. Assume the lengths of different parts are independent.

a) What percent of assembled shafts have to be scrapped?

b) What is the scrapping percentage if we reduce the variance of A and B by 50% each?

This kind of problem is an example of “process improvement” or “industrial engineering”. The purpose of this field is to optimise the current process while minimising the cost of improvement.

Here’s an example where process improvement techniques can be useful: `http://fortune.com/2018/07/16/recall-fda-valsartan-blood-pressure-medication/`

The statistics department offers a course on it: STAT 435 - Statistical Methods for Process Improvements.

Example

Suppose that the height of adult males in Canada is normally distributed with a mean of 70 inches and variance of 4^2 inches, and let $X_1, ..., X_{10}$ denote the heights of a random sample of adult males. Suppose $\bar{X}_{10}$ denotes the sample mean of these heights.

a) Compute the probability that X_3 exceeds 75.
b) Compute the probability that $\bar{X}_{10}$ exceeds 75