Chapter 1

Two Discrete Random Variables

> If two things are independent, they do not know of each other.
>
> Vladimir Cherniak

Joint Distribution

Joint distribution of two random variables X and Y is probability distribution which shows joint probabilities for each pair of values of X and Y.

GRADES AND HAPPINESS Masha proposes that there is a relation between happiness and academic success. She conducts a study and gathers information on the two variables: G and S. G is a student's grade on Math mock exam, $G = \{1, 2, 3, 4, 5\}$, and S measures the level of student's satisfaction with studies at ICEF from 0 ("totally dissatisfied") to 2 ("happy with the studies"). The table below represents the joint distribution of S and G, based on Masha's observations.

The cells of joint distribution table contain probabilities of *intersections* of corresponding values of S and G.

For example, $P(G = 2 \cap S = 1) = \frac{1}{12}$.

		S		
		0	1	2
G	2	1/16	1/12	5/48
	3	1/16	1/8	1/16
	4	1/16	1/8	1/16
	5	1/16	1/12	5/48

Table 1.1: Joint Distribution of Grade in Math (G) and Satisfaction Level (S)

We can construct probability distribution for each of the variables S and G based on their joint distribution.

For example, we may find $P(G=2)$ by merely summing up the probabilities of all intersections that include $G=2$ (values in the row in front of 2):

$$P(G=2) = \frac{1}{16} + \frac{1}{12} + \frac{5}{48} = \frac{12}{48} = \frac{1}{4}$$

Marginal Distribution

We can calculate probability for each value of S and of G in this way:

		S			
		0	1	2	$P(G)$
G	2	1/16	1/12	5/48	**1/4**
	3	1/16	1/8	1/16	**1/4**
	4	1/16	1/8	1/16	**1/4**
	5	1/16	1/12	5/48	**1/4**
	$P(S)$	**1/4**	**5/12**	**1/3**	**1**

Remember that probability of an event is a sum of probabilities of its elementary outcomes!
You can check that all the probabilities in the table sum up to 1!

As you can see, the right column contains probabilities for each possible value of G, and the lowest row contains probabilities for each value of S. These resulting probability distributions of S and G are called **marginal distributions**. Thus, having joint distribution of two random variables you can always reconstruct their marginal distributions.

For example, below is the marginal distribution of S:

S	0	1	2
$P(S)$	$\frac{1}{4}$	$\frac{5}{12}$	$\frac{1}{3}$

We can calculate expected satisfaction level as $E(S) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{5}{12} + 2 \cdot \frac{1}{3} = \frac{13}{12}$. So, if you randomly choose a student, you would expect that his satisfaction level is slightly above 1.

Conditional Distribution

What if we know that a student has excellent grade in Math? Should we expect him to be more satisfied or less satisfied? To answer this question, you should calculate the conditional expectation, which would take into account information that $G=5$.

Conditional expectation is calculated the same way as usually: it is the sum of possible values multiplied by their probabilities. However, to take into account the known condition ($G=5$), it uses *conditional probabilities* instead of marginal ones in calculation: $E(S|G=5) = \sum_{(i=1)}^{3} S_i P(S_i|G=5)$

Therefore we should first find conditional distribution of S (given $G = 5$):

$$P(S = 0|G = 5) = \frac{P(S = 0 \cup G = 5)}{P(G = 5)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}$$

$$P(S = 1|G = 5) = \frac{P(S = 1 \cup G = 5)}{P(G = 5)} = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{1}{3}$$

$$P(S = 2|G = 5) = \frac{P(S = 2 \cup G = 5)}{P(G = 5)} = \frac{\frac{5}{48}}{\frac{1}{4}} = \frac{5}{12}$$

Thus, we get:

S	0	1	2
$P(S\|G = 5)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{5}{12}$

The new expected satisfaction level is: $E(S|G = 5) = 0 * \frac{1}{4} + 1 * \frac{1}{3} + 2 * \frac{5}{12} = \frac{14}{12} = 1\frac{2}{12}$. As you can see, conditional expectation is not the same as unconditional (initial) one. Since $1\frac{2}{12} > \frac{11}{12}$, Masha might expect that a student with excellent grade has a bit higher study satisfaction than other students, on average. In general, the formula for conditional expectation is:

$$E(X|Y = y_i) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_i)$$

Independent Random Variables

Recall definition of independence for random events. A and B are independent if $P(A|B) = P(A)$. We can view a random variable *taking* a particular value as a random event. Therefore, random variables X and Y are said to be independent, if and only if events$(X = x_i)$ and $(Y = y_j)$ are independent for all i, j:

$$P(X = x_i|Y = y_j) = P(X = x_i)$$

Equivalently:

$$P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

From definition of conditional probability we have: $P(X = x_i \cap Y = y_j) = P(X = x_i|Y = y_j)P(Y = y_j)$. Thus, for independent events: $P(X = x_i \cap Y = y_j) = P(X = x_i)P(Y = y_j)$.

If you know that two random variables X and Y with known probability distributions are *independent*, you can construct their joint distribution:

X	x_1	x_2	...	x_k
P	p_1	p_2	...	p_k

Y	y_1	y_2	...	y_m
P	q_1	q_2	...	q_m

$\Downarrow$ Independence $\Downarrow$

	y_1	y_2	...	y_m
x_1	p_1q_1	p_1q_2	...	p_1q_m
x_2	p_2q_1	p_2q_2	...	p_2q_m
...			...	
x_k	p_kq_1	p_kq_2	...	p_kq_m

Generally (without knowledge of independence), joint distribution cannot be reconstructed based on marginal distributions.

Covariance

In previous chapters we've talked about description and analysis of one isolated variable: its expectation, variance, probability distribution, etc. However, it is also interesting to see how two variables are related.

- Is government spending *related* to inflation rate?
- How is the amount of time a student devotes to studies *connected* to his grades achieved?
- Is there a *relation* between student's grades and happiness level?

To address this kind of questions we need a tool to take into account simultaneous dynamics of two variables. **Covariance** is this kind of tool. Its name speaks for itself: Co-Variance shows "collective variance" of two variables. It reflects whether rise in one variable is accompanied by positive or negative change in the other variable. It is denoted by $Cov(X, Y)$.

To calculate the value of $Cov(X, Y)$ we need a data on *simultaneous dynamics* of X and Y. In other words, we need a set of values $\{x_i, y_i\}$, $i \in [1, N]$ in N periods of time or among N objects in the population.

! By definition covariance of X and Y is the expectation of the product of their deviations:

$$Cov(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

Note that $E(X)$ and $E(Y)$ are just some constants μ_X and μ_Y and equation can be written as:

$$Cov(X, Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)]$$

Covariance may take any value between $-\infty$ and $+\infty$.

How to interpret Cov(X,Y)

First, it is expectation. Expectation is the sum of values of a random variable multiplied by their probabilities. Thus, $Cov(X, Y) = \sum_{i=1}^{N}(x_i - E(X)) \cdot (y_i - E(Y)) \cdot p_i$

Suppose, there is a positive trend between X and Y. That means that when X is above its mean, Y also tends to be above its mean, when X is below its mean, Y also typically falls below $E(Y)$. In both cases the product $(x_i - E(X)) \cdot (y_i - E(Y))$ will be positive. Covariance is the sum of these products. Since there is a positive trend, most of the summands are positive, and thus, covariance will exceed zero. On the contrary, negative dynamics between X and Y should be reflected in negative value of the product $(x_i - E(X)) \cdot (y_i - E(Y))$ for most of i. Thus, the sum of these products will be negative.

Therefore, *the sign of covariance reflects the direction of relations between the two variables.* When covariance equals 0 variables reveal no linear relation.

Two useful formulas with covariance:

1. In many cases it is easier to calculate Covariance using the formula: **!**

$$Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

$$\begin{aligned} Cov(X, Y) &= E[(X - E(X)) \cdot (Y - E(Y))] = \\ &= E[(X \cdot Y - X \cdot E(Y) - E(X) \cdot Y + E(X) \cdot E(Y))] = \\ &= E(X \cdot Y) - E[X \cdot E(Y)] - E[E(X) \cdot Y] + E[E(X) \cdot E(Y))] \end{aligned}$$

   Since E(X) and E(Y) are constants μ_X and μ_Y they can be taken out of expectation operator:
   $E(X \cdot Y) - E(Y) \cdot E(X) - E(X) \cdot E(Y) + E(X) \cdot E(Y) = E(X \cdot Y) - E(X) \cdot E(Y)$
   You know how to calculate $E(X)$ and E(Y), and $E(X \cdot Y)$ requires that you analyse the new variable $X \cdot Y$, draw its probability distribution and calculate its expectation. The example of such calculation is provided on page 9.

2. In Chapter 2 you've already seen the formula for $Var(X + Y)$. Here we provide the proof for it. **!**

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y))$$

$$\begin{aligned} Var(X + Y) &= E(X + Y)^2 - [E(X + Y)]^2 = \\ &= E(X^2 + 2XY + Y^2) - [E(X) + E(Y)]^2 = \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X) \cdot E(Y) - [E(Y)]^2 = \\ &= (E(X^2) - [E(X)]^2) + (E(Y^2) - [E(Y)]^2) + 2(E(XY) - E(X) \cdot E(Y)) = \\ &= Var(X) + Var(Y) + 2Cov(X, Y) \end{aligned}$$

   Analogically, you can get the formula for Var(X-Y)

Properties of Covariance

1. Covariance is symmetric.

$$Cov(X, Y) = Cov(Y, X)$$

   Proof is obvious

2. Covariance of a variable with itself is simply its variance.

$$Cov(X, X) = Var(X)$$

   $Cov(X, X) = E[(X - \mu_X) \cdot (X - \mu_X)] = E[(X - \mu_X)^2] = Var(X)$

3. Covariance of a variable and a constant is zero.

$$Cov(X, a) = 0$$

   $Cov(X, a) = E[((X - E(X))(a - E(a)] = E[(X - E(X))(a - a)] = E(0) = 0$. This property is quite obvious. Might there be a relation between something that changes (X) and something that always takes the same value (a)?

4. Constants can be taken out of the covariance operator.

$$Cov(aX, bY) = abCov(X, Y)$$

   $Cov(aX, bX) = E[(aX - a\mu_X) \cdot (bY - b\mu_Y)] = E[a(X - \mu_X) \cdot b(Y - \mu_Y)] = abE[(X - \mu_X) \cdot (Y - \mu_Y)] = abCov(X, Y)$

5. When calculating covariance of sums work with it as if you are multiplying them sums by each other: $(X + Y)(L - W) = XL - XW + YL - YW$.

$$Cov(X + Y, L - W) = Cov(X, L) - Cov(X, W) + Cov(Y, L) - Cov(Y, W)$$

$$Cov(X+Y, L-W) = E[(X+Y-E(X+Y))((L-W)-E(L-W))] =$$
$$= E[(X-EX)+(Y-EY)][(L-EL)-(W-EW)] =$$
$$= E[[(X-EX)(L-EL)] - [(X-EX)(W-EW)]+$$
$$+ [(Y-EY)(L-EL)] - [(Y-EY)(W-EW)]] =$$
$$= Cov(X,L) - Cov(X,W) + Cov(Y,L) - Cov(Y,W)$$

Correlation

Although covariance is a nice tool that reflects collective dynamics of X and Y, its value is also susceptible to influences of units of measurement, scale and size of deviations of the variables. Due to this scale effect covariance between government spending and GDP is always bigger in a large country than in a small one, although the strength of relations between the variables might be the same.

This is explained by big values of GDP and spending in large countries.

Think of one more example. Let X and Y be the weight and the height of a person. If they are measured in grams and centimeters, correspondingly their deviations $(X-E(X))$ and $(Y-E(Y))$ will give big numbers in absolute value. Conaequently, covariance value will be big. If the same variables are measured in kilograms and meters the resulting covariance will be small. Thus, the value of covariance does *not only* represent the strength of relations between the variables. As a result, when comparing two covariance values you cannot say which pair of variables is stronger related.

Therefore, we need a measure of relation between X and Y that would be pure and free of all influences apart from the relation strength between two variables. Correlation serves this goal.

Look at the formula of covariance: $Cov(X,Y) = E[(X-E(X))\cdot(Y-E(Y))]$. Values $X-E(X)$ and $Y-E(Y)$ depend on the natural deviation in X and Y and also on the size of X and Y. To get rid of this dependence we need to adjust the value of $Cov(X,Y)$ by a number proportional to both the volatility and the size of the variables. Standard deviations σ_X and σ_Y perfectly satisfy these requirements.

Big volatility and small units of measurements are accompanied by proportionally big σ. Thus, whether X and Y are measured so that to produce "big numbers" or not and whether or not they are highly volatile — this would not affect the value $\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}$. Therefore, the value $\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}$ only reflects the collective dynamics of X and Y and nothing else.

Thus, we arrived to the formula of **correlation coefficient**:

!

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

Correlation measures the degree of *linear* relationship between two variables and takes a value *between* -1 *and* 1.

Positive correlation indicates that X and Y tend to move in the same direction. It reflects the degree of positive linear relationship. The closer is the coefficient to 1, the closer is the positive relation of X and Y to linear. When correlation equals 1, there is a strict positive linear relationship between X and Y, so that there are constants a, b, such that: $Y = a + bX$, $b > 0$.

Negative correlation indicates that X and Y tend to change in opposite directions. It reflects the degree of negative linear relationship. When correlation equals -1 all values of X and Y lie on a line with negative slope, so that: $Y = a + bX$, $b < 0$.

Zero correlation suggests that two variables reveal no linear relation. In that case X and Y are called **uncorrelated**. This might happen when X and Y are not related at all *or* when they have *non-linear* relation.

You will find illustrations for different values of correlation in Chapter 12. We will also discuss how to estimate it on sample data.

Correlation does not imply causation!

Recall that:
Independent X and Y require: $P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$, for all i, j,
Uncorrelated X and Y require: $Cov(X, Y) = 0$ or, equivalently[3], $E(XY) = E(X)E(Y)$.

When X and Y are correlated (have non-zero value of ρ_{XY}) this might indicate:

- That X depends on Y
- That Y depends on X
- That both X and Y depend on some other variable W, and this is what makes them related (example below).

SMART PEOPLE WEAR BIGGER SHOES? If you were to analyze the relation between IQ tests of kids and their shoe size you would find incredibly high correlation[2]. Does that mean that guys with big feet are more intelligent? No. Positive correlation is explained by the fact that kids grow up, and in the process their intelligence and size of feet increase simultaneously. The reason for both IQ levels and shoe size getting higher is simply — the age of a pupil! So, be careful when interpreting the value of correlation.

When you find that two variables are correlated all you can say is that they are *not independent*, since they reveal some degree of linear relation. You cannot immediately infer what kind of causation defines the relation you observe (`http://tylervigen.com/spurious-correlations`).

Independent variables are always uncorrelated:

$$\text{Independence} \Rightarrow \text{Uncorrelatedness}$$

[3]It follows from $Cov(X, Y) = E(XY) - E(X) \cdot E(Y)$

[2]The example is taken from a Carnegie Mellon University joky publication, aimed to illustrate the typical misinterpretation of correlation coefficient:`https://www.cmu.edu/CSR/case_studies/shoe_size.html`

Proof

Let's assume for simplicity that both X and Y take only 2 values x_1, x_2 and y_1, y_2, correspondingly (in general case proof is the same). Let's denote $P(X = x_i) = P(x_i)$ and $P(Y = y_i) = P(y_i)$.

$(X \cdot Y)$ takes four possible values:$x_1 \cdot y_1$, $x_1 \cdot y_2$, $x_2 \cdot y_1$, $x_2 \cdot y_2$ with probabilities $P(x_1 \cap y_1)$, $P(x_1 \cap y_2)$, $P(x_2 \cap y_1)$, $P(x_2 \cap y_2)$, correspondingly. Using definition of expectation we get:

$E(X \cdot Y) = (x_1 \cdot y_1) \cdot P(x_1 \cap y_1) + (x_1 \cdot y_2) \cdot P(x_1 \cap y_2) + (x_2 \cdot y_1) \cdot P(x_2 \cap y_1) + (x_2 \cdot y_2) \cdot P(x_2 \cap y_2)$

Since X and Y are independent $P(x_i \cap y_i) = P(x_i) \cdot P(y_i)$ we get:

$E(X \cdot Y) = (x_1 \cdot y_1) \cdot P(x_1) \cdot P(y_1) + (x_1 \cdot y_2) \cdot P(x_1) \cdot P(y_2) + (x_2 \cdot P(x_2) \cdot P(y_1) + (x_2 \cdot y_2) \cdot P(x_2) \cdot P(y_2) = x_1 \cdot P(x_1) \cdot (y_1 P(y_1) + y_2 P(y_2)) + x_2 \cdot P(x_2) \cdot (y_1 P(y_1) + y_2 P(y_2)) = x_1 \cdot P(x_1) \cdot E(Y) + x_2 \cdot P(x_2) E(Y) = E(Y)(x_1 \cdot P(x_1) + x_2 \cdot P(x_2)) = E(Y) \cdot E(X)$.

Since $E(XY) = E(X)E(Y)$, X and Y are uncorrelated.

However, the reverse is not true: correlated variables are not necessarily independent.

$$\text{Uncorrelatedness} \not\Rightarrow \text{Independence}$$

Look at the proof above. Read it on the other way around. You cannot prove that $P(x_i \cap y_j) = P(x_i) \cdot P(y_j)$ for all i, j based on $E(XY) = E(X)E(Y)$, because different combinations of probabilities may ensure the latter equality. Example of uncorrelated but not independent variables is provided on the next page.

The table below represents all possible combinations of correlation and independence statuses:

	Independent	Not independent
Uncorrelated $\rho = 0$	Possible	*Possible!*
Correlated $\rho \neq 0$	Impossible	Possible

What you know for sure? Independent variables are uncorrelated. Correlated variables are not independent. In other cases different combinations of the two qualities are possible.

For uncorrelated variables $Cov(X, Y) = 0$. Since $Cov(X, Y) = E(X \cdot Y) - E(X)E(Y)$ and $Var(X \cdot Y) = V(X) + X(Y) \pm Cov(X, Y)$, you use can use that:

$$E(XY) = E(X) \cdot E(Y)$$

$$Var(X \pm Y) = Var(X) + Var(Y)$$

Since independent variables are uncorrelated, these formulas *are always true for independent X and Y*, but you cannot use them if you don't know that X and Y are independent (or uncorrelated).

Let's go back to the "GRADES AND HAPPINESS" example. Are G and S independent? Are they uncorrelated?

		S			
		0	1	2	$P(G)$
G	2	1/16	1/12	5/48	**1/4**
	3	1/16	1/8	1/16	**1/4**
	4	1/16	1/8	1/16	**1/4**
	5	1/16	1/12	5/48	**1/4**
	$P(S)$	**1/4**	**5/12**	**1/3**	**1**

Let's check independence of G and S:

$$P(G = 3 \cap S = 1) = \frac{1}{8} \neq \frac{5}{12} \cdot \frac{1}{4} = P(G = 3) \cdot P(S = 1)$$

Thus, G and S are not independent.

Let's check correlation of G and S. To find correlation between random variables G and S: $\rho_{GS} = \frac{Cov(G,S)}{\sigma_G \cdot \sigma_S}$ we need first to find the covariance of G and S, $Cov(G, S) = E(GS) - E(G)E(S)$. Let's find $E(G)$, $E(S)$, $E(GS)$. The marginal distribution of G:

G_i	2	3	4	5
p_i	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

So, $E(G) = 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 5 \cdot \frac{1}{4} = \frac{7}{2}$. Analogously, for S:

S_i	0	1	2
p_i	1/4	5/12	1/3

$$E(S) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{5}{12} + 2 \cdot \frac{1}{3} = \frac{13}{12}$$

There are two ways to find $E(GS)$. First, you can simply sum up the products of all possible pairs of G and S values by their joint probabilities: $E(GS) = \sum_{i=1}^{3} \sum_{j}^{4} = 1 s_i \cdot g_j \cdot P(s_i \cap g_j) = 0 \cdot 2 \cdot \frac{1}{16} + 1 \cdot 2 \cdot \frac{1}{12} + \cdots + 1 \cdot 5 \cdot \frac{1}{12} + 2 \cdot 5 \cdot \frac{5}{48} = \frac{91}{24}$.

Alternatively, to find the $E(GS)$ we need to find the distribution of GS: all possible values of the product GS and the corresponding probabilities. The table contains the values of G multiplied by S for different values of S and G.

		S		
		0	1	2
G	2	0	2	4
	3	0	3	6
	4	0	4	8
	5	0	5	10

By summing up probabilities from joint distribution of G and S corresponding to each particular value of GS we find the probability distribution of GS. For example, $GS = 4$ in two cases. Thus, $P(GS = 4) = P(G = 4 \cap S = 1) + P(G = 2 \cap S = 2) = \frac{1}{8} + \frac{5}{48} = \frac{11}{48}$.

GS	0	2	3	4	5	6	8	10
$P(GS)$	1/4	1/12	1/8	11/48	1/12	1/16	1/16	5/48

Note that the sum of all probabilities equals 1. Always check that after calculations to make sure you didn't conduct a mistake.

$E(GS) = 0 + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{11}{48} + 5 \cdot \frac{1}{12} + 6 \cdot \frac{1}{16} + 8 \cdot \frac{1}{16} + 10 \cdot \frac{5}{48} = \frac{91}{24}$

Now, we can calculate covariance: $Cov(G, S) = \frac{91}{24} - \frac{13}{12} \cdot \frac{7}{2} = 0$.

$\rho_{GS} = \frac{Cov(G,S)}{\sigma_G \cdot \sigma_S} = 0$. This means that G and S are uncorrelated. This is an example of two variables being uncorrelated but not independent.

You MUST BE ABLE TO REPRODUCE even being half-awake:

- X and Y are independent if and only if: $P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$ for all i and j
- $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$
- $Cov(X, X) = Var(X)$
- $Cov(aX, bY) = abCov(X, Y)$
- $Cov(X + Y, L + W) = Cov(X, L) - Cov(X, W) + Cov(Y, L) - Cov(Y, W)$
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
- $Corr(X, Y) = \rho_{XY} = Cov(X, Y)/(\sigma_X \cdot \sigma_Y), -1 \leqslant \rho_{XY} \leqslant 1$
- Independence $\Rightarrow$ Uncorrelatedness
- Uncorrelatedness $\nRightarrow$ Independence

Calculator BOX

To find $E(XY)$:

1. Represent your joint distribution table as a list of intersections of values of X and Y and their corresponding probabilities. Put values of X into `List 1`, values of Y into `List 2` and the probabilities of corresponding intersections into `List 3`. For example for the "GRADES AND HAPPINESS" example (G and S) the data would look like:

List1	List2	List3
2	0	1/16
2	1	1/12
2	2	5/48
3	0	1/16
3	1	1/8
3	2	1/16
...	...	...

2. CALC $\to$ SET

2 Var	XList	List1
2 Var	YList	List2
2 Var	Freq	List3

3. $\to$ EXIT, now you have results:

Distribution	What you want	Where it is
X	$E(X)$ σ	$\bar{x} = \sum x$ σ_x
Y	(the same as for X)	
Joint	$E(XY)$	$\sum xy$

Sample AP problems with solutions

Problem 1. AP 2015 №3 *You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.*

A shopping mall has three automated teller machines (ATMs). Because the machines receive heavy use, they sometimes stop working and need to be repaired. Let the random variable X represent the number of ATMs that are working when the mall opens on a randomly selected day. The table shows the probability distribution of X.

Number of ATMs working when the mall opens	0	1	2	3
Probability	0.15	0.21	0.4	0.24

(a) What is the probability that at least one ATM is working when the mall opens?

(b) What is the expected value of the number of ATMs that are working when the mall opens?

(c) What is the probability that all three ATMs are working when the mall opens, given that at least one ATM is working?

(d) Given that at least one ATM is working when the mall opens, would the expected value of the number of ATMs that are working be less than, equal to, or greater than the expected value from part (b)? Explain.

Solution.

(a) Let X be the number of ATM machines working when the mall opens.
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.15 = \mathbf{0.85}$

(b) $E(X) = \sum_{i=1}^{4} = 0 \cdot 0.15 + 1 \cdot 0.21 + 2 \cdot 0.4 + 3 \cdot 0.24 = \mathbf{1.73}$

(c) $P(X = 3|X \geq 1) = \frac{P(X=3 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X=3)}{P(X \geq 1)} = \frac{0.24}{0.85} \approx \mathbf{0.282}$

(d) $E(X|X \geq 1) = \sum_(i = 1)^4 x_i \cdot P(X = x_i|X \geq 1) = \sum_(i = 1)^4 x_i \cdot \frac{P(X=x_i)}{P(X \geq 1)} = \frac{1}{P(X \geq 1)} \cdot \sum_(i = 1)^4 x_i \cdot P(X = x_i) = \frac{1}{P(X \geq 1)} \cdot E(X)$
Since $\frac{1}{P(X \geq 1)} > 1$, $E(X|X \geq 1) > E(X)$. So, conditional expectation will be **greater** than unconditional.

Problem 2. AP 2011 №2

You'll have about 9 minutes to solve this problem. It will bring you 10% of score for Free Response section.

The table below shows the political party registration by gender of all 500 registered voters in Franklin Township.

PARTY REGISTRATION — FRANKLIN TOWNSHIP

	Party W	Party X	Party Y	Total
Female	60	120	120	300
Male	28	124	48	200
Total	88	244	168	500

(a) Given that a randomly selected registered voter is a male, what is the probability that he is registered for Party Y?

(b) Among the registered voters of Franklin Township, are the events "is a male" and "is registered for Party Y" independent? Justify your answer based on probabilities calculated from the table above.

(c) See Chapter 6.

Solution

Let W, X, Y denote that a randomly selected registered voter is registered for party W, X and Y, correspondingly.

Let M be "a randomly selected registered voter is a male"

(a) Of the 200 male registered voters in Franklin Township, 48 are registered for Party Y. Therefore the conditional probability that a randomly selected voter is registered for Party Y, given that the voter is a male, is $\frac{48}{200} = 0.24$.
Alternatively[d], $P(Y|M) = \frac{P(Y \cap M)}{P(M)} = \frac{\frac{48}{500}}{\frac{200}{500}} = \mathbf{0.24}$

(b) Since $P(Y \cap M) = \frac{48}{500} = 0.096$ and $P(Y) \cdot P(M) = \frac{168}{500} \cdot \frac{200}{500} = 0.1344$.
Since $P(Y \cap M) \neq P(Y) \cdot P(M)$ M and Y are **not independent** events.

[d]Note that here we calculate probabilities as relative frequencies. For example, P(M) is found as a fraction of males among all the voters, $\frac{200}{500}$.

Problem 3. AP 2003 Form B №2 *You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.*

A simple random sample of adults living in a suburb of a large city was selected. The age and annual income of each adult in the sample were recorded. The resulting data are summarized in the table below.

	Annual Income			
Age Category	\$25,000-\$35,000	\$35,000-\$50,000	Over \$50,000	Total
21-30	8	15	27	50
31-45	22	32	35	89
46-60	12	14	27	53
Over 60	5	3	7	15
Total	47	64	96	207

(a) What is the probability that a person chosen at random from those in this sample will be in the 31-45 age category?

(b) What is the probability that a person chosen at random from those in this sample, whose incomes are over $ 50,000 will be in the 31-45 age category? Show your work.

(c) Based on your answers in parts (a) and (b), is annual income independent of age category for those in this sample? Explain.

Solution

(a) Let A denote the age of an adult. Let I denote the income.
$P(A = \text{"}31-45\text{"}) = \frac{89}{207} \approx \mathbf{0.430}$

(b) $P(A = \text{"}31-45\text{"} | I = \text{"over \$50,000"}) = \frac{P(A=\text{"31-45"} \cap \text{I="over \$50,000"})}{P(I=\text{"over \$50,000"})} = \frac{\frac{35}{207}}{\frac{96}{207}} \approx 0.365$
Since unconditional probability of age category "31-45" is not equal to its probability conditional on income category "over $ 50,000", A and I are **not independent**.

Practice AP problems

Problem 1. AP 2014 №1 5min 5%

An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise.

A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

- Are you an on campus student or an off campus student?
- In how many extracurricular activities do you participate?

The responses of the 100 students are summarized in the frequency table shown.

Level of Participation in Extracurricular Activities	Residential Status		Total
	On campus	Off campus	
No activities	9	30	39
One activity	17	25	42
Two or more activities	7	12	19
Total	33	67	100

(a) Calculate the proportion of on campus students in the sample who participate in at least one extracurricular activity and the proportion of off campus students in the sample who participate in at least one extracurricular activity.

(b) See Chapter 6.

(c) See Chapter 12.

Problem 2. AP 2010 Form B №5 10min 11%

An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

	HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT			
Primary Source for News	Not High School Graduate	High School Graduate But Not College	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1437	693	2500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the Internet?

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

(d) See Chapter 12.

Answers to the practice problems:

Problem 1. a) on campus: 0.727; off campus: 0.552

Problem 2. a) 0.454 b) 0.354 c) not independent