Hypotheses on correlation. χ^2-tests

Statistics

Anton Afanasev

Higher School of Economics

DSBA 211
February 11, 2023

Seminar Overview

1. Quiz
2. Hypotheses on correlation
3. Spearman's rank correlation
4. χ^2-tests
   - Association tests
   - Goodness-of-fit tests

Quiz

Let X is the number of calls to reception per 10 minutes. Random variable X follows Poisson distribution with mean λ.
To test null hypothesis $H_0 : \lambda = \lambda_0$ against alternative $H_1 : \lambda > \lambda_0$, the following decision rule is used: null hypothesis is rejected if $X_1 + X_2 \geq c$, where $\{X_1, X_2\}$ is a random sample of size $n = 2$.
Let $\lambda_0 = 0.2$ and $c = 2$.

1. Find the significance level of the test.
2. Find the probability of type II error if the alternative is $\lambda = 1.5$.

Problem 1

The accompanying table shows percentage change (x_i) in the *Dow-Jones* index over the first five trading days of each of thirteen years, and also the corresponding percentage change (y_i) in the index over the whole year.

x_i	y_i	x_i	y_i	x_i	y_i
1.5	14.9	−1.6	27.7	1.4	27.0
0.2	−9.2	−1.3	22.6	1.5	−4.3
−0.1	19.6	5.6	2.3	−4.7	20.3
2.8	20.3	−1.4	11.9	1.1	4.2
2.2	−3.7				

1. Calculate the sample correlation.
2. Test at the 10% significance level, against a two-sided alternative, the null hypothesis that the population correlation is 0.

Problem 2

The table presents scores of the first and the second game bowled by the same person on six consecutive Tuesday evenings. Assume a bivariate normal distribution

Game 1	170	190	200	183	187	178
Game 2	197	178	150	176	205	153

1. Test $H_0 : \rho = 0$ against alternative $H_1 : \rho \neq 0$ at 10% significance level.
2. Test $H_0 : \sigma_1^2 = \sigma_2^2$ against alternative $H_1 : \sigma_1^2 \neq \sigma_2^2$ at 10% significance level.

Problem 3

An aptitude test has been designed to examine a prospective salesman's ability to sell. Ten current staff sit the test. Instead of putting achieved scores in the computer, a research assistant ranks the individuals in ascending order in terms of the test as well as productivity. The data are:

Staff member	A	B	C	D	E	F	G	H	I	J
Rank order in test	2	3	5	1	4	9	10	6	7	8
Rank order in productivity	1	2	3	4	5	6	7	8	9	10

Does it look as if the test is a good predictor of sales ability?

Problem 4

Consider observations in the table below:

x	0	2	6	-3	4	1	-2	5	-1
y	8	2	0	6	1	5	7	3	4

1. Find Spearman's rank correlation coefficient r_s.
2. Find sample correlation coefficient r_{xy} and compare it with r_s.

Problem 4

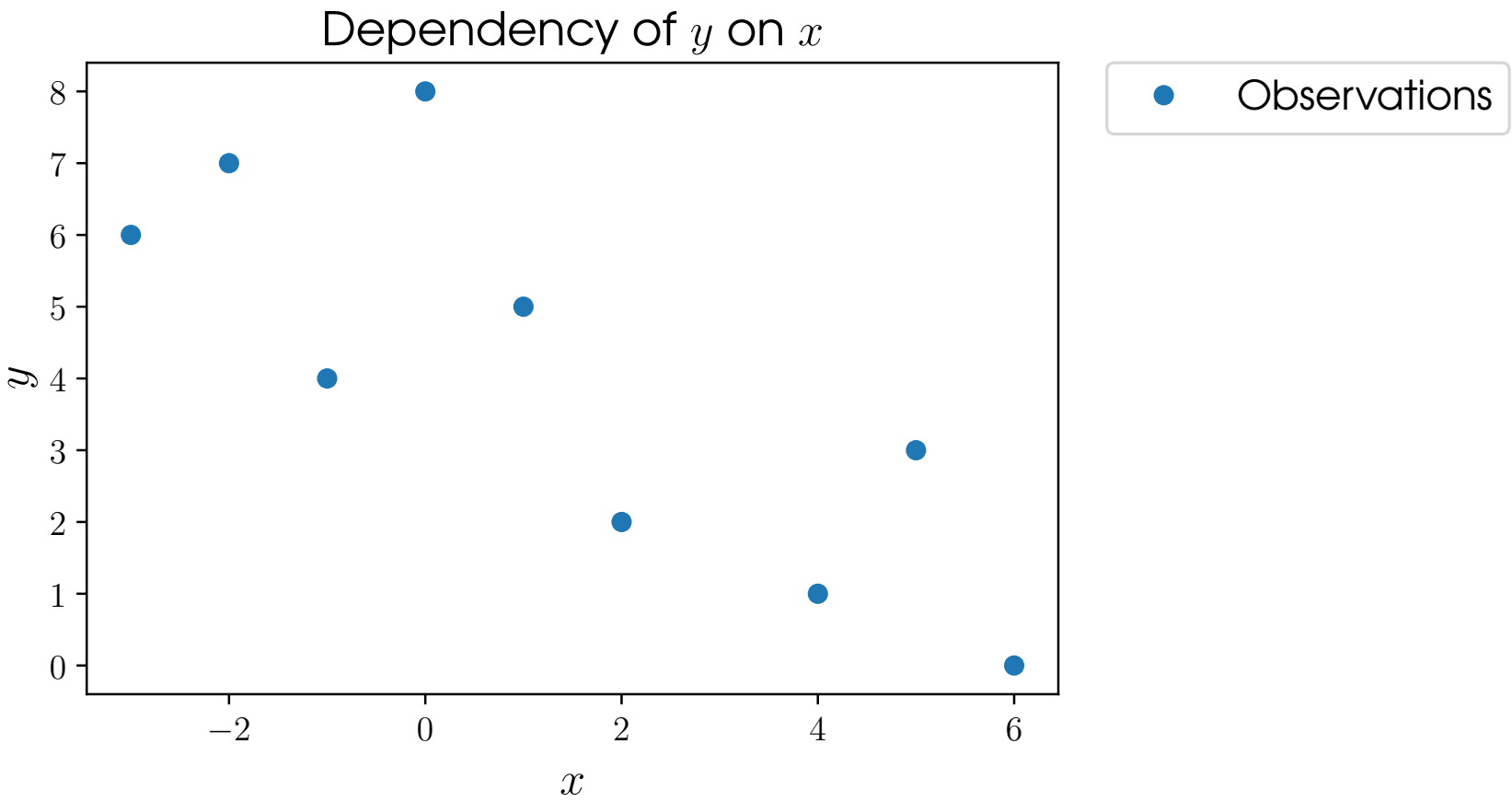

Figure: Observations scatter plot

Problem 5

A random survey of 100 students asked each student to select the most preferred form of recreational activity from 5 choices. Test whether the choice is independent of the gender of the respondent. Approximate the p-value of the test. Would we reject the hypothesis at $\alpha = 0.05$?

Gender	Recreational Choice					Total
	Basketball	Softball	Swimming	Running	Tennis	
Male	21	5	9	12	13	60
Female	9	3	1	15	12	40
Total	30	8	10	27	25	100

Problem 6

A die was tossed 60 times, results are in the table

Face	1	2	3	4	5	6	Total
Number of results	0	10	10	10	15	15	60

Test if the die is a fair die. Use 5% significance level.

Problem 7

Parents of blood type AB will produce children of three different types: AA, AB, and BB. If the Mendelian inheritance hypothesis is true, these three types will be born 25%, 50%, and 25% of the time in the long run. The following data gives the blood types of the 284 children born of 100 AB couples. What p-value does it yield for the Mendelian hypothesis?

Blood Type	Number of children
AA	65
AB	152
BB	67

Problem 8

In the Michigan daily lottery one three-digit integer is generated 6 days a week. Twenty weeks of lottery numbers have been classified by the day of the week and the magnitude of the number. At the 10% significance level, test whether these attributes of classification are independent.

Days	Magnitude of Number		Total
	$000-499$	$500-999$	
Monday and Tuesday	22	18	40
Wednesday and Thursday	19	21	40
Friday and Saturday	13	27	40
Total	54	66	120

Problem 9

A company produces candies of red, yellow and green color. An official of a company claims that the proportion of red and yellow candies is 2 : 3. In a random sample of 500 candies there are 120 red candies, 210 yellow candies and 170 green candies. Test at 5% significance level the official's claim.

Problem 10

A survey is carried out to determine whether the proportion of employees in favor of working from home on two days per week is the same for all age groups. The raw results of the survey are as follows:

Age group	$18-25$	$26-35$	$36-50$	Over 50
In favor	20	62	40	40
Surveyed	52	112	86	120

1. Test the claim that there is no difference in the proportions in favor by age groups at the 5% significance level.
2. What proportions are likely to be different?

Look at the time!