Лабораторная работа №5.

Корреляционный анализ

Цели занятия:

- определения законов распределения случайных величин;
- научиться определять условные законы распределения случайных величин;
- приобрести навыки установления функциональной и статистической зависимостей между случайными величинами.

Задание:

- на основе лекционного материала и представленной методики выполнения заданий определить законы распределения случайных величин, ковариацию и коэффициент корреляции в соответствии с заданным вариантом;
- для автоматизации расчётов использовать MS Excel;
- оформить отчёт по лабораторной работе в соответствии с установленными правилами.

Краткие сведения из теории

*Зависимость между двумя случайными величинами называется **вероятностной** (стохастической, статистической), если каждому значению одной из них соответствует определённое (условное) распределение другой.*

***Ковариацией (корреляционным моментом) Kxy** случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий, т.е.* $K_{xy}=M[(X - M(X)(Y - M(Y)]$, *или* $K_{xy}=M[(X - a_x)(Y - a_y)]$.

Из определения: $K_{xy}=K_{yx}$.

Свойства ковариации случайных величин:

1. Ковариация двух независимых случайных величин равна нулю.
2. Ковариация двух случайных величин равна математическому ожиданию их произведения минус произведение их математических ожиданий:
   $K_{xy} = M(XY) - M(X)M(Y)$ **или** $K_{xy} = M(XY) - a_x a_y$.
3. Ковариация двух случайных величин по абсолютной величине не превосходит произведения их средних квадратических отклонений, т.е. $|Kxy| \le \sigma_x\sigma_y$.

***Коэффициентом корреляции** двух случайных величин (**corr(X,Y)**) называется отношение их ковариации к произведению средних квадратических отклонений этих величин:* $\rho_{xy} = \dfrac{k_{xy}}{\sigma_x\sigma_y}$.

Из определения следует $\rho_{xy} = \rho_{yx} = \rho$.

Свойства коэффициента корреляции:

1. $-1 \le \rho \le 1$.
2. Если случайные величины независимы, то $\rho = 0$, так как $K_{xy} = 0$.
3. Если $\rho = |1|$, то между этими случайными величинами существует линейная функциональная зависимость.

Пример 1. Закон распределения дискретной двумерной случайной величины (X,Y) задан в табл. 1.

Таблица 1.

x_i \ y_j	-1	0	1	2
1	0,1	0,25	0,30	0,15
2	0,1	0,05	0	0,05

Найти:

а) закон распределения одномерных случайных величин X и Y;

б) условные законы распределения случайной величины X при условии Y = 2 и случайной величины Y при условии X = 1;

в) вычислить P(Y<X).

г) ковариацию и коэффициент корреляции случайных величин X и Y.

Решение.

а) Случайная величина X может принимать значения:

X=1 с вероятностью p_1 = 0,1+0,25+0,3+0,15 = 0,8;

X=2 с вероятностью p_2 = 0,1+0,05+0+0,05 = 0,2.

Отсюда строится её закон распределения (табл. 2).

Таблица 2

x_i	1	2
p_i	0,8	0,2

Случайная величина Y может принимать значения:

Y=-1 с вероятностью p_1 = 0,1+0,1 = 0,2;

Y=0 с вероятностью p_2 = 0,25+0,05 = 0,3;

Y=1 с вероятностью p_3 = 0,3+0 = 0,3;

Y=2 с вероятностью p_4 = 0,15+0,05 = 0,2.

Отсюда строится её закон распределения (табл. 3).

Таблица 3

y_j	-1	0	1	2
p_j	0,2	0,3	0,3	0,2

б) Условный закон распределения X при условии, что Y = 2 можно получить, если вероятности p_{ij}, стоящие в последнем столбце табл. 1, разделить на их сумму, т.е. на 0,2:

Таблица 4

x_i	1	2
p_i	0,75	0,25

Условный закон распределения Y при условии, что X = 1 можно получить, если вероятности p_{ij}, стоящие в первой строке столбце табл. 1, разделить на их сумму, т.е. на 0,8:

Таблица 5

y_j	-1	0	1	2
p_j	0,125	0,3125	0,375	0,1875

в) Для нахождения вероятности P(Y<X) складываем вероятности событий p_{ij} из табл. 1, для которых $y_j<x_i$. Получим: P(Y<X) = 0,1+0,25+0,1+0,05+0=0,5.

г) $a_x = \sum x_i p_i$ = 1*0,8+2*0,2 = 1,2;

$M(X^2) = \sum x_i^2 p_i = 1^2*0,8+2^2*0,2 = 1,6$;

$D(X) = M(X^2) - a_x^2 = 1,6 - 1,2^2 = 0,16$; σ_x = (D(X))^0,5 = 0,4;

$a_y = \sum y_j p_j$ = (-1)*0,2+0*0,3+1*0,3+2*0,2 = ,5;

$M(Y^2) = \sum y_j^2 p_j = (-1)^2*0,2+0^2*0,3+1^2*0,3+2^2*0,2 = 1,3$;

$D(Y) = M(Y^2) - a_y^2 = 1,3 - 0,5^2 = 1,05$; σ_y = (D(Y))^0,5 = 1,025;

$$M(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j p_{ij} =$$

=1*(-1)*0,1+(-1)*0*0,25+1*1*0,3+1*2*0,15+2*(-1)*0,1+2*0*0,05+2*1*0+2*2*0,05=0,5;

$\boldsymbol{k_{xy}}$ = M(XY) – $\boldsymbol{a_x a_y}$ = 0,5-1,2*0,5 = -0,1.

$\boldsymbol{\rho}$ = $\boldsymbol{k_{xy}/(\sigma_x\sigma_y)}$ = (-0,1)/(0,4*1,025) = -0,244. Следовательно между случайными величинами X и Y существует отрицательная линейная зависимость.

Пример решения этой же задачи в среде MS Excel представлен на рис. 1.

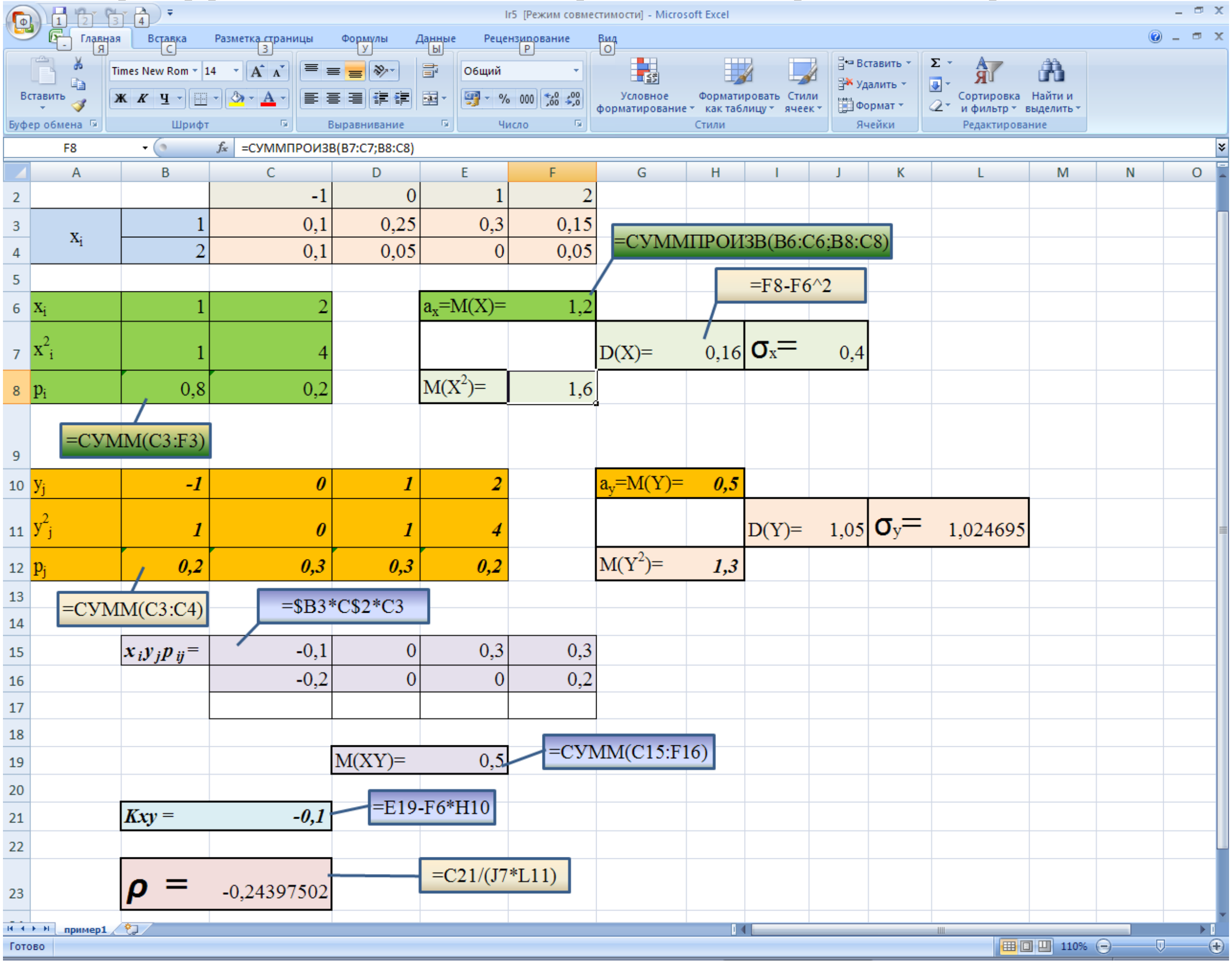

Рис. 1

Варианты заданий

Задание 1

1. Закон распределения дискретной двумерной случайной величины (X,Y) задан в табл. 6.

Таблица 6

Найти: закон распределения одномерных случайных величин X и Y.

x_i \ y_j	0	1	2	3
-1	0,02	0,03	0,09	0,01
0	0,04	0,2	0,16	0,1
1	0,05	0,1	0,15	0,05

2. Закон распределения дискретной двумерной случайной величины (X,Y) задан в табл. 6. Найти условные законы распределения случайной величины X при условии Y = 2 и случайной величины Y при условии X = 1.
3. Закон распределения дискретной двумерной случайной величины (X,Y) задан в табл. 6. Вычислить P(Y>X). (***P(Y>X)=0,81***)

Задание 2

1. Закон распределения дискретной двумерной случайной величины (X,Y) задан в табл. 6. Вычислить ковариацию и коэффициент корреляции случайных величин X и Y. (***Kxy=-0,012; ρ=-0,02006***)
2. Дискретная случайная величина имеет ряд распределения, представленный в табл. 7.

Таблица 7

x_i	-1	0	1	2
p_i	0,2	0,1	0,3	0,4

Найти математическое ожидание и дисперсию случайной величины $Y=2^x$.
(***M(Y)=2,4; D(Y)=1,99***)

3. Имеются две случайные величины X и Y, связанные соотношением ***Y=2-3X***. Известны характеристики величины X: a_x = -1; D(X)=4. Найти математическое ожидание и дисперсию случайной величины Y. (***M(Y)=5; D(Y)=36***)
4. Имеются две случайные величины X и Y, связанные соотношением ***Y=2-3X***. Известны характеристики величины X: a_x = -1; ***D(X)***=4. Найти ковариацию и коэффициент корреляции случайной величины X и Y. (***Kxy=-12; ρ= -1***)
5. Имеется случайная величина X с математическим ожиданием m_x и дисперсией D_x. Найти математическое ожидание и дисперсию следующих случайных величин:
Y = -X; Z = X+2Y-1.
6. Дискретная случайная величина имеет ряд распределения, представленный в табл. 8.

Таблица 8

x_i	-1	0	1	2
p_i	0,2	0,1	0,3	0,4

Найти математическое ожидание и дисперсию случайной величины $Y=e^{-x}$.
(***M(Y)=0,80815; D(Y)= 0,97262***)

7. Имеются две случайные величины X и Y, связанные соотношением ***Y=2X+3***. Известны характеристики величины X: a_x = -1; ***D(X)***=4. Найти ковариацию и коэффициент корреляции случайной величины X и Y. (***Kxy=8; ρ=1***)
8. Дискретная случайная величина имеет ряд распределения, представленный в табл. 9.

Таблица 9

x_i	1	3	5	7
p_i	0,2	0,1	0,3	0,4

Найти математическое ожидание и дисперсию случайной величины $Y=2+x^{1/2}$.
(***M(Y)=4,102326; D(Y)=0,3802254***)