Кузбасский гуманитарно-педагогический институт
Кемеровский государственный университет
1939

Кузбасский гуманитарно-педагогический институт
федерального государственного бюджетного
образовательного учреждения
высшего образования
«Кемеровский государственный университет»
654027, г. Новокузнецк, пр-т Пионерский, 13

Факультет психологии и педагогики

Кафедра психологии и общей педагогики

Вячкин Евгений Сергеевич

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания
к семинарским занятиям
для обучающихся по специальности
37.05.02 Психология служебной деятельности

Новокузнецк
2024

Настоящие методические указания являются составной частью методического обеспечения учебной дисциплины «Математическая статистика» и содержат рекомендации обучающимся к семинарским занятиям.

Методические указания предназначены для студентов, обучающихся по специальности 37.05.02 Психология служебной деятельности

Текст представлен в авторской редакции

Вячкин Е.С.

Математическая статистика: метод.указания по работе на семинар. занятиях для студентов ф-та психологии и педагогики, обучающихся по специальности 37.05.02 Психология служебной деятельности / Е.С. Вячкин. – перераб. и доп. - Кузбасский гум-пед. инст. Кемеров. гос. ун-та.– Новокузнецк: КГПИ КемГУ, 2023. – 28с. – Текст : непосредственный.

Рекомендовано
на заседании кафедры
психологии и общей педагогики
(протокол № 7 от 14.03.2024 г.)
Заведующая кафедрой
А.И. Алонцева

Утверждено
методической комиссией
факультета психологии и педагогики
(протокол № 5 от 20.03.2024 г.)
Председатель метод.комисии
Е.В. Дворцова

Оглавление

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Настоящие методические указания по учебной дисциплине «Математическая статистика» являются составной частью нормативно-методического обеспечения требований, предусмотренных ФГОС ВО по специальности подготовки 37.05.02 Психология служебной деятельности специализации подготовки «Морально-психологическое обеспечение служебной деятельности».

Методические указания направлены на освоение обучающимися общекультурных, общепрофессиональных, профессиональных, профессионально-специализированных компетенций, предусмотренных основной профессиональной образовательной программой высшего образования, в том числе учебным планом и рабочей программой дисциплины.

Методические указания к практическим занятиям предназначены для студентов очной формы обучения направлений 37.05.02 Психология служебной деятельности, профиль «Морально-психологическое обеспечение служебной деятельности». Дисциплина «Математическая статистика» включена в образовательную программу 37.05.02 Психология служебной деятельности, профиль «Морально-психологическое обеспечение служебной деятельности» и входит в базовую часть дисциплин.

Методические указания содержат сведения о пяти темах курса «Математическая статистика», каждая из которых включает вопросы для самопроверки, подробно описанное решение демонстрационного примера и задания для решения различного уровня сложности. В конце методических указаний приведен список рекомендуемой к изучению литературы.

Выполнение практических заданий начинается с ответов на вопросы для самопроверки, затем необходимо подробно изучить предложенный демонстрационный пример, после чего можно приступать к решению предложенных задач.

Практическая работа 1. Основы математической статистики

Цель: научиться собирать статистические данные для оценки числовых характеристик случайной величины.

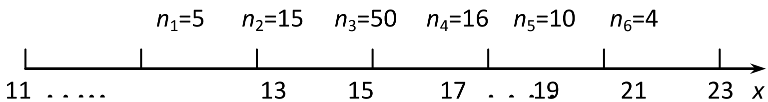
Вопросы для теоретической подготовки

1. Что называется простой выборкой из генеральной совокупности? Постановка задачи статистического оценивания. Какая оценка называется несмещенной, эффективной, состоятельной?
2. Как проводится группировка значений и строится статистическое распределение выборки, полигон и гистограмма частот?

Демонстрационные примеры

<u>Пример 1.</u> Сгруппировать выборку объема $n = 100$ значений случайной величины X, изображенных на рисунке 3 точками числовой оси Ox.

Рисунок 3 – Группировка выборочных данных



<u>Решение.</u> Для разбивки выборки на k частей найдем число k из неравенств $ln(n) < k < \sqrt{n}$. В нашем случае $5 \le k \le 10$. Пусть, например, $X_{\max} = 22{,}8$ и $X_{\min} = = 11{,}4$. Выбираем длину каждого интервала целым числом, например, $h = 2$; тогда $k = 6$. Количество точек в i-м интервале обозначим n_i. Если точка попала на границу между интервалами, то ее относим к интервалу с большим количеством точек. Частоту n_i присвоим точке, расположенной в середине i-го интервала. Координата этой точки называется вариантой x_i . Полученные данные

заносим в таблицу 1, которая называется статистическим распределением выборки.

Таблица 1.Статистическое распределение выборки

Номер i	1	2	3	4	5	6
Варианта x_i	12	14	16	18	20	22
Частота n_i	5	15	50	16	10	4

Пример 2. Построить гистограмму и полигон частот согласно статистическому распределению выборки из примера 1.

Решение. Построим гистограмму частот с помощью рисунка 3. На горизонтальной оси откладываются точки разбиения значений выборки, а на вертикальной оси – частоты соответствующих вариант, деленные на шаг разбиения h. Сумма площадей всех прямоугольников на гистограмме равна объему выборки. Аналогично рисунку 4 строится гистограмма относительных частот. В этой гистограмме на вертикальной оси откладываются частоты соответствующих вариант, деленные на шаг разбиения и на объем выборки. Поэтому сумма площадей всех прямоугольников на гистограмме относительных частот равна единице. Формы этих гистограмм одинаковы.

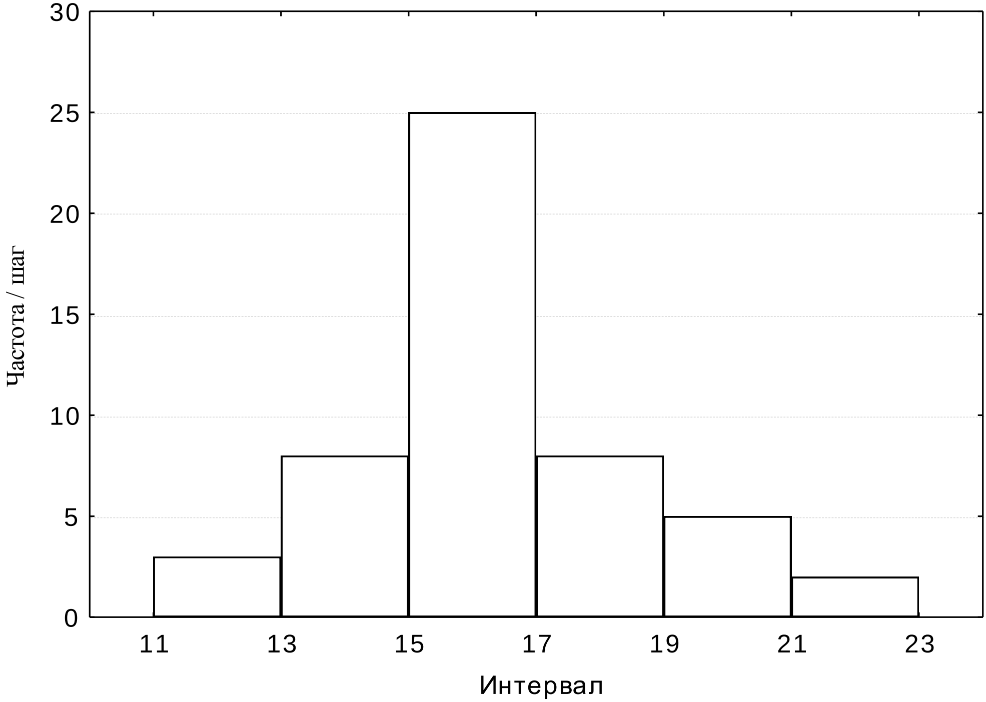


Рисунок 4 – Гистограмма частот

Построим полигон частот (рисунок 5) с помощью таблицы 1.

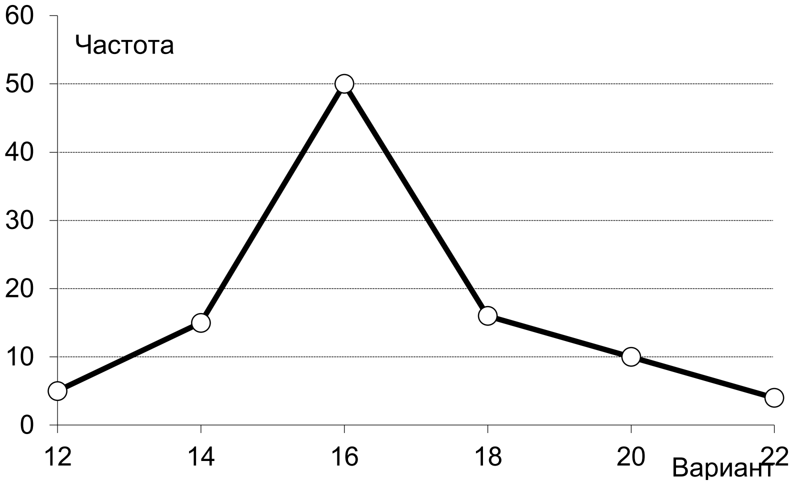


Рисунок 5 – Полигон частот

Задания для самостоятельного решения

1. Дана выборка: 2,4,7,3,1,1,3,2,7,3. Построить статистическое распределение выборки, полигон и гистограмму частот.

2. Дана выборка значений случайной величины X объема 20:

12, 14, 19, 15, 14, 18, 13, 16, 17, 12
18, 17, 15, 13, 17, 14, 14, 13, 14, 16

Построить статистическое распределение выборки, полигон и гистограмму частот.

3. Даны результаты измерений отклонений от нормы диаметров 50 подшипников:

-1,760	-0,291	-0,110	-0,450	0,512
-0,158	1,701	0,634	0,720	0,490
1,531	-0,433	1,409	1,740	-0,266
-0,058	0,248	-0,095	-1,488	-0,361

0,415	-1,382	0,129	-0,361	-0,087
-0,329	0,086	0,130	-0,244	-0,882
0,318	-1,087	0,899	1,028	-1,304
0,349	-0,293	0,105	-0,056	0,757
-0,059	-0,539	-0,078	0,229	0,194
0,123	0,318	0,367	-0,992	0,529

Для данной выборки построить статистическое распределение выборки, полигон и гистограмму частот.

Практическая работа 2. Числовые характеристики выборки

Цель: научиться использовать статистические данные для оценки числовых характеристик случайной величины.

Вопросы для теоретической подготовки

1. Метод произведений расчета начальных и центральных моментов ***k***-го порядка.
2. Статистические оценки: выборочная средняя, выборочная исправленная дисперсия, выборочная асимметрия и эксцесс, их свойства и отыскание с помощью ложного нуля.
3. Интервальная оценка для математического ожидания нормальной случайной величины. Точность и достоверность оценки в случае известной и неизвестной дисперсии.

Демонстрационные примеры

Пример 1. Методом произведений найти выборочную среднюю, выборочную дисперсию, асимметрию и эксцесс по данным выборки из таблицы 1.

Решение. Составим расчетную таблицу 2. Для этого:

1) дополним таблицу 1 еще пятью строчками;

2) в качестве ложного нуля C выберем варианту (16), которая имеет наибольшую частоту (в качестве ложного нуля C можно взять любую варианту, расположенную примерно в середине строки); в клетке четвертой строки, которая принадлежит столбцу, содержащему ложный нуль, пишем 0; слева от нуля последовательно записываем –1, –2, а справа от нуля 1, 2, 3;

3) произведения частот n_i на условные варианты u_i запишем в пятой строке; отдельно находим сумму (–25) отрицательных чисел и отдельно сумму (48) положительных чисел; сложив эти числа, их сумму (23) помещаем в последнюю клетку строки;

4) произведения частот на квадраты условных вариант, то есть $n_iu_i^2$ запишем в шестой строке (удобнее перемножить числа каждого столбца четвертой и пятой строк: $u_i \cdot n_iu_i = n_iu_i^2$); сумму чисел строки (127) помещаем в последнюю клетку строки и так далее.

В итоге получим расчетную таблицу 2.

Таблица 2.Расчетная таблица

Номер i	1	2	3	4	5	6	
Варианта x_i	12	14	16	18	20	22	
Частота n_i	5	15	50	16	10	4	$n = 100$
Условная u_i	–2	–1	0	1	2	3	
n_iu_i	–10	–15	0	16	20	12	$\Sigma n_iu_i = 23$
$n_iu_i^2$	20	15	0	16	40	36	$n_iu_i^2 = 127$
$n_iu_i^3$	–40	–15	0	16	80	108	$n_iu_i^3 = 149$
$n_iu_i^4$	80	15	0	16	160	324	$n_iu_i^4 = 595$

Вычислим условные начальные и центральные моменты первого, второго, третьего и четвертого порядков:

$$\overline{u} = (\Sigma n_i u_i)/n = 23/100 = 0{,}23; \quad \overline{u^2} = (\Sigma n_i u_i^2)/n = 127/100 = 1{,}27;$$

$$\overline{u^3} = (\Sigma n_i u_i^3)/n = 149/100 = 1{,}49; \quad \overline{u^4} = (\Sigma n_i u_i^4)/n = 595/100 = 5{,}95;$$

$$\mu_2 = \overline{u^2} - \left(\overline{u}\right)^2 = 1{,}217; \quad \mu_3 = \overline{u^3} - 3\overline{u^2}\,\overline{u} + 2\left(\overline{u}\right)^3 = 0{,}638;$$

$$\mu_4 = \overline{u^4} - 4\overline{u^3}\,\overline{u} + 6\overline{u^2}\left(\overline{u}\right)^2 - 3\left(\overline{u}\right)^4 = 4{,}974.$$

Вычислим искомые выборочные величины, учитывая, что h = 2, а ложный нуль (варианта, которая имеет наибольшую частоту) C = 16

$$\overline{x} = \overline{u} \cdot h + C = 0{,}23 \cdot 2 + 16 = 16{,}46; \quad D_x = \mu_2 h^2 = 4{,}87;$$

$$A_s = \frac{\mu_3}{\mu_2 \sqrt{\mu_2}} = 0{,}475; \quad E_k = \frac{\mu_4}{\left(\mu_2\right)^2} - 3 = 0{,}358.$$

Пример 4. По данным выборки из таблицы 1 найти доверительный интервал с надежностью 0,95 для математического ожидания m нормально распределенной случайной величины X.

Решение. Выборочные среднюю и дисперсию используем из примера 2. Исправленное среднее квадратическое отклонение найдем по формуле

$$s = \sqrt{\frac{n}{n-1} D_x} = \sqrt{\frac{100}{99} 4{,}87} = 2{,}218.$$

Значение t_γ найдем с помощью таблицы приложения 4 и значений γ = 0,95 и n = 100: t_γ = 1,984. Найдем доверительный интервал

$$\bar{x} - \frac{t_\gamma s}{\sqrt{n}} < m < \bar{x} + \frac{t_\gamma s}{\sqrt{n}}.$$

Подставляя $\bar{x} = 16{,}46;\ t_\gamma = 1{,}984;\ s = 2{,}218;\ n = 100$, получим искомый доверительный интервал $16{,}02 < m < 16{,}90$.

Задания для самостоятельного решения

1. Вычислить числовые характеристики для статистического ряда:

x_i	1	2	3	4	7
n_i	2	2	3	1	2

2. Вычислить числовые характеристики для статистического ряда:

x_i	12	13	14	15	16	17	18	19
n_i	2	3	5	2	2	3	2	1

3. Вычислить числовые характеристики для статистического ряда:

Интервалы	$[-2{,}06; -1{,}46)$	$[-1{,}46; -0{,}86)$	$[-0{,}86; -0{,}26)$	$[-0{,}26; 0{,}34)$
Частоты n_i	2	6	11	15
Интервалы	$[0{,}34; 0{,}94)$	$[0{,}94; 1{,}54)$	$[1{,}54; 2{,}14)$	
Частоты n_i	11	3	2	

Практическая работа 3. Проверка статистических гипотез

Цель: познакомиться с типами статистических гипотез и критериями их проверки.

Вопросы для теоретической подготовки

1. Основная и альтернативная гипотеза. Статистический критерий и его критическая область. Виды критических областей.
2. Ошибки 1-го и 2-го рода. Уровень значимости и мощность критерия.
3. Критерий согласия Пирсона проверки гипотезы о нормальном законе распределения генеральной совокупности.
4. Критерий Фишера-Снедекора проверки гипотезы о совпадении двух дисперсий нормальных генеральных совокупностей. Условия применения критерия.
5. Критерий Стьюдента проверки гипотезы о совпадении двух математических ожиданий нормальных генеральных совокупностей в случае известных и неизвестных, но равных дисперсий.
6. Критерий согласия Пирсона проверки гипотезы о независимости двух признаков с помощью таблицы сопряженности признаков.
7. Критерий Стьюдента проверки гипотезы об отсутствии корреляционной зависимости между двумя нормальными генеральными совокупностями.

Демонстрационные примеры

Пример 1. Используя критерий согласия Пирсона χ^2 при уровне значимости 0,05, проверить, согласуется ли гипотеза о нормальном распределении случайной величины с заданным распределением выборки

Номер i	1	2	3	4	5	6
Варианта x_i	12	14	16	18	20	22
Частота n_i	5	15	50	16	10	4

<u>Решение.</u> Используя метод произведений, найдем выборочные среднюю $\bar{x} = 16{,}46$ (см. пример 1, практическая работа 7) и исправленное среднее квадратическое отклонение $s = 2{,}218$. Вычислим теоретические частоты, учитывая, что n=100, h=2, s=2,218, по формуле

$$n_i' = \frac{nh}{s}\varphi(u_i) = \frac{100 \cdot 2}{2{,}218}\varphi(u_i) = 90{,}17 \cdot \varphi(u_i), \quad u_i = \frac{x_i - \bar{x}}{s}.$$

Составим расчетную таблицу 3 значений эмпирических n_i и теоретических n_i' частот.

Таблица 3.Расчетная таблица

Номер i	1	2	3	4	5	6	Σ
Варианта x_i	12	14	16	18	20	22	
Частота n_i	5	15	50	16	10	4	100
u_i	–2,01	–1,11	–0,21	0,694	1,596	2,498	
$\varphi(u_i)$	0,053	0,215	0,39	0,313	0,098	0,018	
n_i'	4,78	19,39	35,17	28,22	8,84	1,62	
$n_i - n_i'$	0,22	–4,39	14,83	–12,2	1,16	2,38	
$(n_i - n_i')^2$	0,048	19,27	220	149	1,35	5,66	
$(n_i - n_i')^2/n_i'$	0,01	0,99	6,26	5,28	0,15	3,49	16,18

Из таблицы 3 находим наблюдаемое значение критерия согласия Пирсона $\chi^2_{набл} = \Sigma(n_i - n_i^{'})^2 / n_i^{'} = 16{,}18$, которое характеризует степень отличия теоретических и эмпирических частот. По таблице критических точек распределения χ^2, согласно уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = 6 - 3 = 3$ (6 – максимальный номер i) находим критическую точку $\chi^2_{кр}(0{,}05;3) = 7{,}8$. Так как $\chi^2_{набл} > \chi^2_{\text{кр}}$, то теоретические и эмпирические частоты отличаются значимо на уровне α , а гипотеза о нормальном распределении случайной величины не согласуется с заданным распределением выборки.

<u>Пример 2.</u> В результате обработки двух независимых выборок, объемы которых соответственно равны $n_x = 5$ и $n_y = 6$, извлеченных из нормальных генеральных совокупностей X и $У$, найдены выборочные исправленные дисперсии $s_x^2 = 0{,}25; \quad s_y^2 = 0{,}108$. Используя критерий Фишера-Снедекора проверить гипотезу о равенстве дисперсий двух нормальных совокупностей $\boldsymbol{H_0 : D(X) = D(У)}$ при конкурирующей гипотезе $\boldsymbol{H_1 : D(X) > D(У)}$ и уровне значимости $\alpha = 0{,}05$.

Следует отметить, что в качестве совокупности ***X*** выбирается та, для которой исправленная дисперсия больше.

<u>Решение.</u> Используя критерий Фишера-Снедекора, вычислим наблюдаемое значение критерия по формуле $F_{набл} = s_x^2 / s_y^2 = 2{,}31$. По таблице критических точек распределения Фишера–Снедекора, согласно уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = 5 - 1 = 4$, $k_2 = 6 - - 1 = 5$ находим критическую точку распределения $F_{кр} = 5{,}19$. Так как $F_{набл} < F_{\text{кр}}$, то выборочные исправленные дисперсии отличаются незначимо на уровне значимости α и нет оснований отвергнуть гипотезу о равенстве дисперсий двух нормальных генеральных совокупностей.

Пример 3. Используя критерий Стьюдента проверить гипотезу о равенстве математических ожиданий двух нормальных совокупностей с неизвестными, но равными дисперсиями $\boldsymbol{H_0 : M(X) = M(У)}$ при конкурирующей гипотезе $\boldsymbol{H_1 : M(X) \neq M(У)}$ и уровне значимости $\alpha = 0{,}05$.

В результате обработки двух независимых выборок, объемы которых соответственно равны $n_x = 5$ и $n_y = 6$, извлеченных из нормальных генеральных совокупностей X и $У$, найдены выборочные средние и исправленные дисперсии $\overline{x} = 3{,}3;\quad \overline{y} = 2{,}48;\quad s_x^2 = 0{,}25;\quad s_y^2 = 0{,}108.$

Решение. Используя критерий Стьюдента, вычислим наблюдаемое значение критерия по формуле

$$t_{набл} = \frac{\overline{x} - \overline{y}}{\sqrt{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}.$$

Подставляя заданные значения в эту формулу получим $t_{набл} = 3{,}27$. Критическое значение определяется в случае двусторонней критической области, а число степеней свободы $k = n_x + n_y - 2$. По таблице критических точек распределения Стьюдента, согласно уровню значимости $\alpha = = 0{,}05$ и числу степеней свободы $k = 5 + 6 - 2 = 9$ находим критическую точку $t_{кр} = 2{,}26$. Так как $t_{набл} > t_{кр}$, то выборочные средние отличаются значимо на уровне α , а гипотеза о равенстве математических ожиданий двух нормальных совокупностей не согласуется с заданными распределениями выборок.

Пример 4. Используя критерий согласия Пирсона χ^2 при уровне значимости 0,05 проверить, согласуется ли гипотеза о независимости уровня успеваемости от фамилии студента с результатами успеваемости двух студентов разных курсов.

Таблица 4.Эмпирические частоты оценок

Фамилия	Оценки студентов				Всего
	5	4	3	2	
Иванов	30	10	10	0	50
Петров	24	32	3	1	60
Итого	54	42	13	1	110

Результаты успеваемости студентов приведены в таблице 4 сопряженности рассматриваемых признаков, где указаны частоты полученных оценок за период обучения.

Решение. Введем следующие обозначения: i – номер строки таблицы, который равен 1, 2; j – номер столбца, который равен 1, 2, 3, 4; N_{ij} – эмпирическая частота, расположенная на пересечении i -й строки и j -го столбца. Для оценивания доли количества оценок каждого студента используются отношения $N_{i\cdot}/N$, где $N_{i\cdot} = \sum_j N_{ij}$, $N = \sum_{ij} N_{ij}$ =110. Аналогично доли всех “пятерок”, “четверок” и так далее оцениваются отношениями $N_{\cdot j}$ $/N$, где $N_{\cdot j} = \sum_i N_{ij}$. Если уровни успеваемости студентов отличаются незначимо, то приведенные в таблице 4 частоты приближаются теоретическими частотами $N'_{ij} = \dfrac{N_{i\cdot} N_{\cdot j}}{N}$. Составим таблицу 5 теоретических частот.

Таблица 5.Теоретические частоты оценок

Фамилия	Оценки студентов				Всего
	5	4	3	2	
Иванов	24,55	19,09	5,91	0,45	50
Петров	29,45	22,91	7,09	0,55	60
Итого	54	42	13	1	110

Наблюдаемое значение критерия с учетом теоретических частот $\chi^2_{набл} = \sum_i \sum_j \left(N_{ij} - N'_{ij}\right)^2 \left(N'_{ij}\right)^{-1} =$ 16,18, которое характеризует степень отличия теоретических и эмпирических частот. По таблице критических точек распределения χ^2, согласно уровню значимости α = 0,05 и числу степеней свободы k = (4 – 1)(2 – 1) = 3 (4 – максимальный номер j, 2 – максимальный номер i) находим критическую точку $\chi^2_{кр}(0{,}05;3) = 7{,}8$. Так как $\chi^2_{набл} > \chi^2_{\text{кр}}$, то теоретические и эмпирические частоты отличаются значимо на уровне α , а гипотеза о независимости уровней успеваемости студентов не согласуется с данными успеваемости.

Пример 5. В результате обработки двумерной выборки объема n = 100, извлеченной из нормальной генеральной совокупности ($X, У$), методом произведений найден выборочный коэффициент корреляции $r_в$ = 0,76.

Используя критерий Стьюдента, проверить гипотезу об отсутствии корреляционной зависимости между двумя нормальными генеральными совокупностями $\boldsymbol{H_0 : r_{xy} = 0}$ при конкурирующей гипотезе

$\boldsymbol{H_1 : r_{xy} \neq 0}$ и уровне значимости α = 0,05 .

Решение. Используя критерий Стьюдента, вычислим наблюдаемое значение критерия по формуле $t_{набл} = r_в\sqrt{n-2} / \sqrt{1-(r_в)^2}$ = 11,58. Критическое значение определяется в случае двусторонней критической области, а число степеней свободы k = $n-2$. По таблице критических точек распределения Стьюдента , согласно уровню значимости α = 0,05 и числу степеней свободы k = 100 – 2 = 98 находим критическую точку $t_{кр}$ = 1,985. Так как $t_{набл} > t_{\text{кр}}$, то выборочный коэффициент корреляции отличаются от нуля значимо на уровне α, а гипотеза о об отсутствии корреляционной зависимости между двумя переменными $X, У$ не согласуется с данными выборки.

Задания для самостоятельного решения

1. В результате обработки выборки из генеральной совокупности получены следующие частоты вариант (таблица 6).

Таблица 6.Эмпирические и теоретические частоты

Эмпирические	6	13	38	74	106	85	30	14
Теоретические	3	14	42	82	99	76	37	13

При уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении случайной величины с заданным распределением выборки ($\chi^2_{набл}$ = 7,19 < $\chi^2_{кр}$(0,05; 5) = 11,1).

2. В результате обработки двух независимых выборок, объемы которых n_x = n_y = 10, извлеченных из нормальных генеральных совокупностей X и $У$, найдены выборочные средние и исправленные дисперсии $\bar{x} = 17{,}4;\quad \bar{y} = 15{,}1;\quad s_x^2 = 7{,}25;\quad s_y^2 = 6{,}45$. Используя критерий Стьюдента, проверить гипотезу о равенстве математических ожиданий двух нормальных совокупностей с неизвестными, но равными дисперсиями $\boldsymbol{H_0}$: $\boldsymbol{M(X) = M(У)}$ при конкурирующей гипотезе $\boldsymbol{H_1}$: $\boldsymbol{M(X) \neq M(У)}$ и уровне значимости 0,05 ($t_{набл} = 1{,}97 < t_{\text{кр}}(0{,}05;18) = 2{,}1$).

3. Анализ ранних произведений Эрнеста Хемингуэя показал, что предложения, включающие разное число слов, распределены в них следующим образом (таблица 7)

Таблица 7
Относительные частоты предложений

Число слов	2–3	4–5	6–8	9–12	13–16	17–20	21–24	25–28	> 28
Частота	0,01	0,034	0,067	0,091	0,21	0,174	0,181	0,143	0,09

Наследники автора заявили о находке рукописи ранее неопубликованного произведения своего предка. В выборке, состоящей из 2000 предложений, было обнаружено следующее распределение фраз по длине (таблица 8).

Таблица 8.Частоты предложений

Число слов	2–3	4–5	6–8	9–12	13–16	17–20	21–24	25–28	> 28
Частота	15	51	118	227	476	401	352	239	121

Могут ли эти данные подтвердить заявление наследников? ($\chi^2_{набл}$ = 61,4, что больше критического значения при любом α).

4. Используя критерий согласия Пирсона χ^2 при уровне значимости 0,05, проверить, согласуется ли гипотеза о независимости уровня успеваемости от фамилии студента с результатами успеваемости трех студентов разных курсов. Результаты успеваемости студентов приведены в таблице сопряженности рассматриваемых признаков, где указаны частоты полученных оценок за период обучения ($\chi^2_{набл}$ = 8,92 < $\chi^2_{кр}(0{,}05;\ 6)$ = 12,53).

5.

Таблица 9.Эмпирические частоты оценок

Фамилия	Оценки студентов				Всего
	5	4	3	2	
Иванов	30	10	10	0	50
Петров	30	20	10	0	60
Сидоров	18	17	4	1	40
Итого	78	47	24	1	150

6. Используя критерий Стьюдента для проверки гипо-

тезы об отсутствии корреляционной зависимости между двумя нормальными генеральными совокупностями $\boldsymbol{H_0}$: $\boldsymbol{r_{xy}} = \boldsymbol{0}$ при конкурирующей гипотезе $\boldsymbol{H_1}$: $\boldsymbol{r_{xy} \neq 0}$ и уровне значимости α = 0,05, найти критическое значение выборочного коэффициента корреляции для выборки объема $\boldsymbol{n}$ = 100 (0,196).

Практическая работа 4. Линейные статистические модели

Цель: научить устанавливать корреляционную зависимость между двумя нормальными случайными величинами.

Вопросы для теоретической подготовки

1. Какая зависимость называется корреляционной и чем она отличается от функциональной?
2. Компоненты двумерной случайной величины и их числовые характеристики: условные плотность и математическое ожидание.
3. Корреляционный момент и коэффициент корреляции, их вероятностный смысл и свойства.
4. Выборочное уравнение прямой линии регрессии и его числовые характеристики.

Демонстрационный пример

X	*У*	*X*	*У*	*X*	*У*	*X*	*У*	*X*	*У*
21,47	20	32	32	40	36	25	23,93	34	48,01
42	45,48	24,14	13	28,64	46,61	34,03	36,02	30	26,91
24,23	18	19,85	18,02	26	23,01	23,74	17,85	29,24	43,37
41	53,39	18,53	19,46	30,07	43,17	28,5	36,29	36,81	45,36
30	46,96	35	41,89	32,4	37,49	31,17	38	35,28	57,47
24	27,11	29,95	30	40	55,72	29,95	38,19	28	40
24,13	12	28,45	35	35,18	40,14	26	28,62	23,83	15,5
30	31,25	29	32,38	40	49,12	30,66	29	26	24,51

31	35,26	28	31,64	28	36,74	29	35	38	40
25	25,3	28	39	28,5	32,09	30,25	29	32	35,33
40	34,94	39	57,79	29,8	33,71	32	33	35,47	46
29,5	31,62	30,97	30	39	38	19,52	18,32	41	35
36,09	46	41	53	29	38	26,49	23,73	38,5	42,65
31,13	27,67	31	38,74	28,84	29	39,5	45	35,89	45
32,06	22,49	34,9	44,15	41	45,67	31	32	33,71	45,63
39	32,91	30	36	27,5	39,26	30,33	38,59	31	40
31,84	39	32,19	39,54	40,36	33,36	30,43	28,75	40,5	40,32
30	39,56	34,19	35,6	38	36,47	40	45	33,18	43,88
30,5	38,23	36,38	44,08	30,5	39,66	22,73	20,05	29	40
25	29,35	28,58	30,05	35,59	47,63	36,21	47,95	42	56

Найти выборочное уравнение прямой линии регрессии *Y* на *X* согласно протоколу наблюдений двумерной выборки, приведенному в таблице.

Решение. Наблюдения двумерной выборки в виде 100 точек с координатами изображаются на координатной плоскости и заключаются в прямоугольник, который разбивается на прямоугольные клетки со сторонами h_x и h_y. На рисунке 6 выбрано k = 5 , а шаги $h_x = 5$, $h_y = 10$.

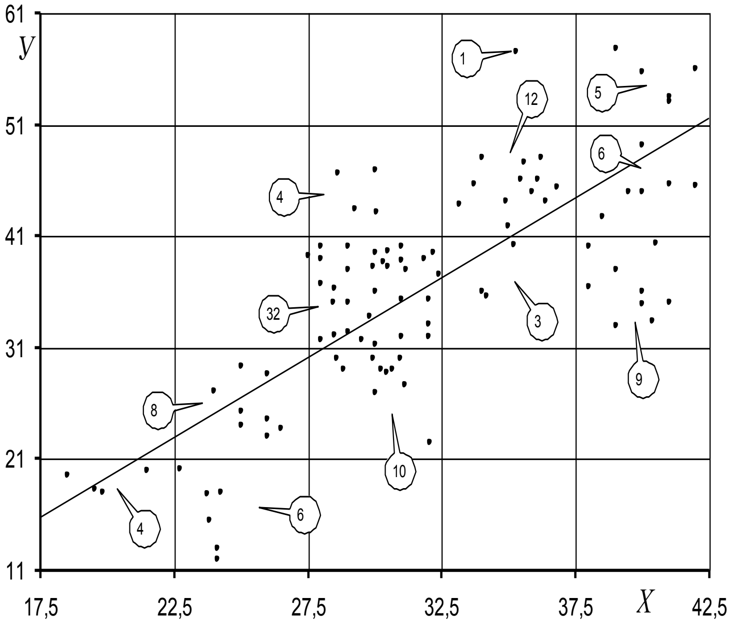


Рисунок 6 – Группировка наблюдений двумерной выборки

Количество точек, попавших в клетку, отмечено в выносках и рассматривается как частота nxy центра соответствующей клетки. Если точка расположена на границе, разделяющей две клетки, то она относится к той клетке, в которой находится больше точек. Найденные таким образом значения частот приведены в клетках корреляционной таблицы 4, обведенных двойной линией. Сверху и слева указаны соответствующие координаты центра клетки (варианты X, У), а снизу и справа – суммарные частоты вариант.

Таблица 10. Корреляционная таблица

У	*X*	n_y

	20	25	30	35	40	
16	4	6	-	-	-	10
26	-	8	10	-	-	18
36	-	-	32	3	9	44
46	-	-	4	12	6	22
56	-	-	-	1	5	6
n_x	4	14	46	16	20	n = 100

Для каждой случайной величины Х и У вводятся условные варианты *u* и *v*. Составим корреляционную таблицу 11 в условных вариантах

Таблица 11.Корреляционная таблица в условных вариантах

v	u					n_v	$U=\Sigma n_{uv}u$	vU
	–2	–1	0	1	2			
–2	–8 4	–6 6				10	–14	28
–1		–8 8	0 10			18	–8	8
0			0 32	3 3	18 9	44	21	0
1			0 4	12 12	12 6	22	24	24
2				1 1	10 5	6	11	22
n_u	4	14	46	16	20	100	$\Sigma vU = 82$	

Методом произведений найдем выборочные средние: сначала $\bar{u} = 0{,}34$ и $\overline{u^2} = 1{,}26$; затем $\bar{v} = -0{,}04$ и $\overline{v^2} =$

Выборочные средние квадратические отклонения условных вариант *u* и *v* найдем по формулам:

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = 1{,}07; \quad \sigma_v = \sqrt{\overline{v^2} - (\bar{v})^2} = 1{,}02.$$

Затем проверяется гипотеза о нормальном распределении генеральных совокупностей X и У. Только после этого отыскивается выборочный коэффициент корреляции.

Выборочный коэффициент корреляции найдем по формуле

$$r_{\text{в}} = \frac{\Sigma n_{uv}uv - n\overline{u}\overline{v}}{n\sigma_u\sigma_v} = \frac{82 - 100 \cdot 0{,}34 \cdot (-0{,}04)}{100 \cdot 1{,}07 \cdot 1{,}02} = 0{,}76.$$

Поскольку значение $r_{\text{в}} > 0{,}196$, то с уровнем значимости 0,05 можно утверждать о зависимости У от X. Если это условие не выполняется, то можно принять гипотезу об отсутствии корреляционной зависимости с отмеченным уровнем значимости.

Найдем выборочные средние $\overline{x}$ и $\overline{y}$, учитывая, что шаги $h_x = 5$, $h_y = 10$, а ложные нули $C_x = 30$; $C_y = 36$:

$$\overline{x} = \overline{u} \cdot h_x + C_x = 0{,}34 \cdot 5 + 30 = 31{,}7; \quad \overline{y} = \overline{v} \cdot h_y + C_y = -0{,}04 \cdot 10 + 36 = 3$$

Затем найдем выборочные средние квадратические отклонения:

$\sigma_x = h_x\ \sigma_u = 5 \cdot 1{,}07 = 5{,}35$; $\sigma_y = h_y\ \sigma_v = 10 \cdot 1{,}02 = 10{,}2$.

Подставим найденные величины в уравнение регрессии Y на X

$$\overline{y}_x - \overline{y} = r_{\text{в}} \frac{\sigma_y}{\sigma_x}(x - \overline{x}). \quad \text{Получим} \quad \overline{y}_x - 35{,}6 = 0{,}76\frac{10{,}2}{5{,}35}(x - 31{,}$$

или окончательно $\overline{y}_x = 1{,}45x - 10{,}36.$ График полученной корреляционной зависимости приведен на рисунке 6. Этот график разбивает облако данных на две симметричные части и наилучшим образом приближает эти данные.

Задания для самостоятельного решения

1. В результате независимых испытаний получены пары значений случайных величин X и Y :

x_i	10	20	25	28	30
y_i	4	8	7	12	14

В таблице значения X расставлены в возрастающем порядке.

Найти выборочное уравнение линейной регрессии и выборочный коэффициент корреляции. Построить прямые регрессии.

2. В результате независимых испытаний получены пары значений случайных величин X и Y :

x_i	12	28	22	27	40
y_i	5	9	12	19	29

В таблице значения X расставлены в возрастающем порядке.

Найти выборочное уравнение линейной регрессии и выборочный коэффициент корреляции. Построить прямые регрессии

3. В результате независимых испытаний получены пары значений случайных величин X и Y :

x_i	10	20	30	40	50
y_i	2	8	12	15	17

В таблице значения X расставлены в возрастающем порядке.

Найти выборочное уравнение линейной регрессии и выборочный коэффициент корреляции. Построить прямые регрессии

СПИСОК УЧЕБНОЙ ЛИТЕРАТУРА ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Основная литература

1. Горлач, Б.А. Теория вероятностей и математическая статистика : Учебник / Б.А. Горлач – Москва: Лань, 2013. – URL: http://e.lanbook.com/reader/book/4864/ – Текст : электронный.

Дополнительная литература

1. Емельянов, Г.В. Задачник по теории вероятностей и математической статистике : Учебник / Г.В. Емельянов, В.П. Скитович – Москва: Лань, 2007. – URL: http://e.lanbook.com/reader/book/141/ – Текст : электронный.
2. Кибзун, А.И. Теория вероятностей и математическая статистика . Базовый курс с примерами и задачами / А.И. Кибзун, Е.Р. Горяинова, А.В. Наумов – Москва: Лань, 2005. – URL: http://e.lanbook.com/reader/book/2198/ – Текст : электронный.
3. Хрущева, И.В. Теория вероятностей : Учебник / И.В. Хрущева – – Москва: Лань, 2009. – URL: http://e.lanbook.com/reader/book/425/ – Текст : электронный.
4. Туганбаев, А.А. Теория вероятностей и математическая статистика : Учебник / А.А. Туганбаев, В.Г. Крупин – Москва: Лань, 2011. – URL: http://e.lanbook.com/reader/book/652/ – Текст : электронный.
5. Бородин, А.Н. Элементарный курс теории вероятностей и математической статистики : Учебник / А.Н. Бородин – – Москва: Лань, 2011. – URL: http://e.lanbook.com/reader/book/2026/ – Текст : электронный.
6. Палий, И. А. Теория вероятностей : Учебное пособие / И.А. Палий. – Москва: ИНФРА-М, 2012. –

URL:http://znanium.com/bookread2.php?book=225156 – Текст : электронный.

7. Хуснутдинов, Р. Ш. Теория вероятностей : Учебник / Р.Ш. Хуснутдинов. – М.: НИЦ ИНФРА-М, 2013. – URL: http://znanium.com/bookread2.php?book=363773 – Текст : электронный.

8. Ермаков, В. И. Теория вероятностей и математическая статистика : Учеб. пособие / Под ред. В.И. Ермакова.– Москва: ИНФРА-М, 2004. – URL:http://znanium.com/bookread2.php?book=76845 – Текст : электронный.

9. Павлов, С. В. Теория вероятностей и математическая статистика : Учебное пособие / С.В. Павлов. – Москва: ИЦ РИОР: ИНФРА-М, 2010. – URL:http://znanium.com/bookread2.php?book=217167 – Текст : электронный.