

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего профессионального образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

Методические указания
по выполнению практических работ
по дисциплине
**Методы планирования и анализа эксперимента с применением
цифровых технологий**

Направление подготовки 13.03.02 Электроэнергетика и
электротехника Направленность (профиль) «Электроэнергетические
системы и сети»
Квалификация выпускника бакалавр
Изучается в 3 семестре

Ставрополь
2026

Практическое занятие №1.

Определение основных числовых характеристик совокупности случайных величин

ОСНОВНЫЕ СВЕДЕНИЯ

При измерении свойств продуктов легкой промышленности и разработке методов оценки технологических параметров, как правило, изучается совокупность случайных величин, которая может быть определена основными числовыми характеристиками: **средним, дисперсией, коэффициентом вариации, квадратической неровнотой и т.д.** Известно, что числовые характеристики меняются от выборки к выборке и являются также случайными величинами, которые варьируют с заданной доверительной вероятностью в определенном интервале. Чем больше ошибка числовой характеристики, тем шире интервал. Точность каждой числовой характеристики определяется ее **ошибкой**, а надежность – **доверительной вероятностью**. Задаваясь точностью и надежностью при известной дисперсии случайной величины, можно определить **доверительный объем** испытаний для оценки числовой характеристики.

1.1 Получение совокупности случайных величин

Для ознакомления с методикой определения основных числовых характеристик совокупности случайных величин необходимо получить данную совокупность. Она может быть получена на разрывной машине (прочность, удлинение), весах (масса отрезков пряжи, полосок ткани или трикотажа), круткомере (крутка пряжи) и других приборах. Воспользуемся совокупностями, приведенными в приложении А.

1.2 Расчет оценок математического ожидания, дисперсии и среднего квадратического отклонения

Математическое ожидание $\bar{Y}$ (среднее значение) определяет центр распределения случайных величин, около которого группируется большая их часть. Абсолютными характеристиками рассеяния случайной величины Y около центра распределения $\bar{Y}$ является дисперсия $S^2\{Y\}$ и среднее квадратическое отклонение $S\{Y\}$.

Расчет оценок математического ожидания, дисперсии и среднего квадратического отклонения для анализируемой выборки осуществляется по следующим формулам:

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i; \quad (1.1)$$

$$S^2\{Y\} = \frac{1}{m-1} \cdot \sum_{i=1}^m (Y_i - \bar{Y})^2; \quad (1.2)$$

$$S\{Y\} = \sqrt{S^2\{Y\}}. \quad (1.3)$$

1.3 Исключение резко выделяющихся экспериментальных данных

Совокупность полученных экспериментальных данных часто имеет значения, резко выделяющиеся относительно других, что приводит к постановке вопроса об их исключении из дальнейшей обработки. Причиной появления таких данных может быть изменение условий проведения опыта в момент наблюдения, ошибочная регистрация параметра (по вине оператора) и т.п. Независимо от причин получения резко выделяющихся данных они могут существенно исказить числовые характеристики. С другой стороны, при необоснованном исключении таких данных числовые характеристики также будут искажены.

Самый надежный метод определения возможности исключения резко выделяющихся данных – это анализ условий, при которых они были получены. Если условия существенно отличаются от стандартных (или установленных по плану эксперимента), то данные необходимо исключить из дальнейшей обработки независимо от их величины. Если определение существенности изменения условий эксперимента невозможно или представляет большие трудности, то используют статистический метод исключения данных, сущность которого заключается в следующем:

- находят в совокупности максимальную и минимальную величины и определяют расчетные значения критерия Смирнова-Грabbса:

$$V_{R \max} = \frac{Y_{i \max} - \bar{Y}}{S\{Y\}} \cdot \sqrt{\frac{m}{m-1}}; \quad (1.4)$$

$$V_{R \min} = \frac{\bar{Y} - Y_{i \min}}{S\{Y\}} \cdot \sqrt{\frac{m}{m-1}}. \quad (1.5)$$

- сравнивают полученные значения с табличным V_T (приложение Б), если

$V_{R \max}$ или $V_{R \min}$ больше V_T , то соответствующее значение Y_i необходимо исключить из совокупности, а затем повторить расчет оценок $\bar{Y}$, $S^2\{Y\}$ и $S\{Y\}$;

- процедуру повторяют до полного исключения резко выделяющихся значений из совокупности.

1.4 Расчет относительных характеристик рассеяния случайной величины

Относительной характеристикой рассеяния случайной величины является коэффициент вариации $CV\{Y\}$:

$$CV\{Y\} = V\{Y\} = \frac{S\{Y\}}{\bar{Y}}. \quad (1.6)$$

Если данная величина выражается в процентах, то она называется квадратической неровнотой $C\{Y\}$:

$$C\{Y\} = \frac{S\{Y\}}{\bar{Y}} \cdot 100(\%). \quad (1.7)$$

1.5 Определение ошибки среднего и границ доверительного интервала

В результате измерений исследуемого параметра возникают ошибки (погрешности измерения), для описания которых введены оценки абсолютной ε_i и относительной δ_i погрешности. Абсолютная и относительная доверительные ошибки, допущенные при оценке математического ожидания, определяются по формулам:

$$\varepsilon\{\bar{Y}\} = \frac{2 \cdot S\{Y\}}{\sqrt{m}}; \quad (1.8)$$

$$\delta\{\bar{Y}\} = \frac{2 \cdot \tilde{N}\{Y\}}{\sqrt{m}}. \quad (1.9)$$

Двусторонним доверительным интервалом называется интервал, который покрывает неизвестный параметр распределения с заданной доверительной вероятностью P_D :

$$\bar{Y} - \varepsilon\{Y\} \leq \bar{Y} \leq \bar{Y} + \varepsilon\{Y\}. \quad (1.10)$$

В практике текстильных исследований при статистической обработке обычно принимают $P_D = 0,95$. Величина, равная $\alpha = 1 - P_D$, называется уровнем значимости и иногда выражается в процентах.

1.6 Доверительный объем испытаний

Анализируя точность оценки среднего значения, можно решить, является ли она достаточной или требуется увеличение объема измерений. Задаваясь требуемой величиной относительной ошибки (например, $\delta = 3\%$) и приняв квадратическую неровноту по данным предыдущих опытов или другой априорной информации, можно рассчитать доверительный объем выборки:

$$m\{\bar{Y}\} \geq \left(\frac{u\{P_D\} \cdot C\{Y\}}{\delta\{\bar{Y}\}} \right)^2, \quad (1.11)$$

где $u\{P_D\}$ – квантиль нормального распределения случайной величины (при $P_D = 0,954$ и $u\{P_D\} = 2$).

Практическое занятие №2.

Упрощенный способ вычисления статистических характеристик вариационных рядов

При больших значениях вариантов и соответствующих им частот вычисление выборочного среднего, дисперсии и выборочных моментов по приведенным ниже формулам приводит к громоздким вычислениям.

В этом случае используют условные варианты u_i , определяемые по формулам: $u_i = \frac{x_i - c}{h}$, где числа c и h выбираются произвольно.

Чтобы упростить вычисления в качестве c выбирают вариант, который имеет наибольшую частоту или находится в середине ряда. Число c называется «ложным нулем». В качестве h выбирают число равное длине интервала (в случае интервального ряда) или наибольший общий делитель разностей $(x_i - c)$.

Для вычисления числовых характеристик выборки составляем табл. 1.

Таблица 1.

u_i	n_i	$u_i n_i$	$u_i^2 n_i$	$u_i^3 n_i$	$u_i^4 n_i$	$(u_i + 1)^4 n_i$
u_1	n_1					
$\vdots$	$\vdots$					
u_m	n_m					
	$\sum_{i=1}^m n_i = n$	$\sum_{i=1}^m u_i n_i$	$\sum_{i=1}^m u_i^2 n_i$	$\sum_{i=1}^m u_i^3 n_i$	$\sum_{i=1}^m u_i^4 n_i$	$\sum_{i=1}^m (u_i + 1)^4 n_i$

Контроль:

$$\sum_{i=1}^m (u_i + 1)^4 n_i = \sum_{i=1}^m u_i^4 n_i + 4 \sum_{i=1}^m u_i^3 n_i + 6 \sum_{i=1}^m u_i^2 n_i + 4 \sum_{i=1}^m u_i n_i + n$$

С помощью сумм, полученных в нижней строке таблицы, находим условные моменты:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m u_i n_i, \quad M_2^* = \frac{1}{n} \sum_{i=1}^m u_i^2 n_i,$$

$$M_3^* = \frac{1}{n} \sum_{i=1}^m u_i^3 n_i, \quad M_4^* = \frac{1}{n} \sum_{i=1}^m u_i^4 n_i.$$

Числовые характеристики выборки вычисляем по формулам:

$$\overline{x}_e = M_1^* h + c; \quad D_e = \left(M_2^* - (M_1^*)^2 \right) \cdot h^2; \quad \sigma_e = \sqrt{D_e};$$

$$A_s^* = \frac{\mu_3^*}{\sigma_e^3}; \quad E_k^* = \frac{\mu_4^*}{\sigma_e^4} - 3,$$

где μ_3^* и μ_4^* находим по формулам:

$$\mu_3^* = \left(M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3 \right) \cdot h^3,$$

$$\mu_4^* = \left(M_4^* - 4M_1^* M_3^* + 6(M_1^*)^2 \cdot M_2^* - 3(M_1^*)^4 \right) \cdot h^4.$$

Пример. Вычислить числовые характеристики выборки, для которой построен интервальный ряд.

○ В качестве вариантов x_i возьмем середины интервалов. Перейдем к условным вариантам.

Вариант, значение которого 0,04, имеет наибольшую частоту и находится в середине ряда. Примем его за «ложный ноль» (начало отсчета).

Условные варианты найдем по формуле:

$$u_i = \frac{x_i - c}{h},$$

где $c = 0,04$, $h = 0,6$.

Составим расчетную табл. 2 по форме табл. 1

Таблица 2

x_i	n_i	u_i	$u_i \cdot n_i$	$u_i^2 \cdot n_i$	$u_i^3 \cdot n_i$	$u_i^4 \cdot n_i$	$(u_i + 1)^4 \cdot n_i$
-1,76	2	-3	-6	18	-54	162	32
-1,16	6	-2	-12	24	-48	96	6
-0,56	11	-1	-11	11	-11	11	0
0,04	15	0	0	0	0	0	15
0,64	11	1	11	11	11	11	176
1,24	3	2	6	12	24	48	243
1,84	2	3	6	18	54	162	512
Σ	50		-6	94	-24	490	984

Контроль: $\sum_{i=1}^m (u_i + 1)^4 n_i = \sum_{i=1}^m u_i^4 n_i + 4 \sum_{i=1}^m u_i^3 n_i + 6 \sum_{i=1}^m u_i^2 n_i + 4 \sum_{i=1}^m u_i n_i + n =$
 $= 490 + 4 \cdot (-24) + 6 \cdot 94 + 4 \cdot (-6) + 50 = 984$. Расчеты проведены верно.

По данным табл. 8 находим условные моменты:

$$M_1^* = -\frac{6}{50} = -0,12, \quad M_2^* = \frac{94}{50} = 1,88,$$

$$M_3^* = -\frac{24}{50} = -0,48, \quad M_4^* = \frac{490}{50} = 9,8.$$

Находим числовые характеристики выборки:

$$\overline{x}_g = \mathbf{M}_1^* \mathbf{h} + \mathbf{c} = -0,12 \cdot 0,6 + 0,04 = -0,032$$

$$\mathbf{D}_g = \left(\mathbf{M}_2^* - (\mathbf{M}_1^*)^2 \right) \cdot \mathbf{h}^2 = (1,88 - (-0,12)^2) \cdot 0,6^2 = 0,6716$$

$$\sigma_g = \sqrt{\mathbf{D}_g} = \sqrt{0,672} = 0,8195$$

Вычислим центральные моменты третьего и четвертого порядка:

$$\mu_3^* = \left(\mathbf{M}_3^* - 3\mathbf{M}_1^* \mathbf{M}_2^* + 2(\mathbf{M}_1^*)^3 \right) \cdot \mathbf{h}^3 =$$

$$= (-0,48 - 3 \cdot (-0,12) \cdot 1,88 + 2(-0,12)^3) \cdot 0,6^3 = 0,0418$$

$$\mu_4^* = \left(\mathbf{M}_4^* - 4\mathbf{M}_1^* \mathbf{M}_3^* + 6(\mathbf{M}_1^*)^2 \cdot \mathbf{M}_2^* - 3(\mathbf{M}_1^*)^4 \right) \cdot \mathbf{h}^4 =$$

$$= (9,8 - 4 \cdot (-0,12) \cdot (-0,48) + 6(-0,12)^2 \cdot 1,88 - 3(-0,12)^4) \cdot 0,6^4 = 1,2127$$

Вычислим выборочные коэффициенты асимметрии и эксцесса:

$$\mathbf{A}_s^* = \frac{\mu_3^*}{\sigma_g^3} = \frac{0,0418}{0,8195^3} = 0,0759$$

$$\mathbf{E}_k^* = \frac{\mu_4^*}{\sigma_g^4} - 3 = \frac{1,2127}{0,8195^4} - 3 = -0,3112. \bullet$$

Практическое занятие №3.
Графическое изображение статистических данных

Эмпирическая функция распределения

Пусть получено статистическое распределение выборки и каждому варианту из этой выборки поставлена в соответствие его частость.

Определение. Эмпирической функцией (функцией распределения выборки) называется функция $F^*(x)$, определяющая для каждого значения x частость события $\{X < x\}$,

$$F^*(x) = \frac{n_x}{n},$$

- где n - объем выборки, n_x - число наблюдений, меньших x ($x \in R$).

При увеличении объема выборки частость события $\{X < x\}$ приближается к вероятности этого события. Эмпирическая функция $F^*(x)$ является оценкой интегральной функции $F(x)$ в теории вероятностей.

Функция $F^*(x)$ обладает теми же свойствами, что и функция $F(x)$:

1. $0 \leq F^*(x) \leq 1$;
2. $F^*(x)$ - неубывающая функция;
3. $F^*(-\infty) = 0$, $F^*(+\infty) = 1$.

Пример.

Построить эмпирическую функцию и ее график по данным табл.1

○

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ 0,2 & \text{при } 0 < x \leq 2; \\ 0,4 & \text{при } 2 < x \leq 3; \\ 0,7 & \text{при } 3 < x \leq 4; \\ 0,8 & \text{при } 4 < x \leq 10; \\ 1 & \text{при } x > 10; \end{cases}$$

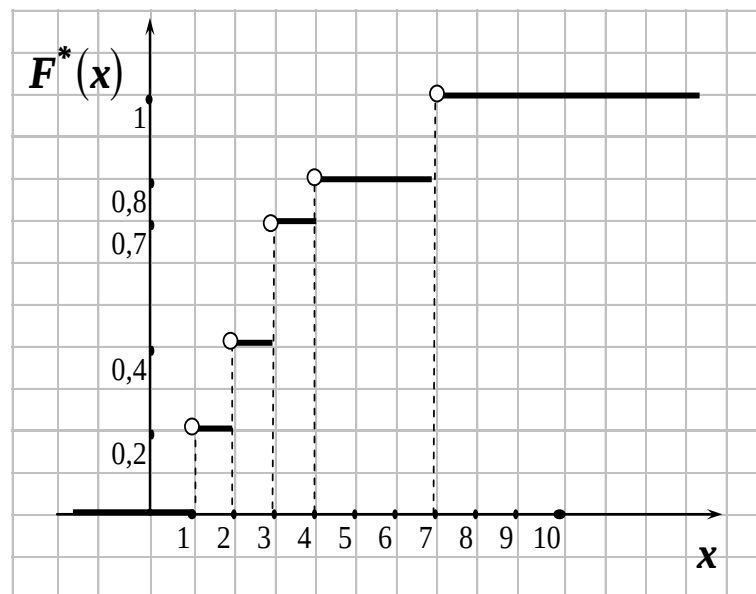


Рис. 1

Эмпирическая плотность распределения

Для интегральной функции распределения $F(x)$ справедливо приближенное равенство: $F(x + \Delta) - F(x) \approx f(x) \cdot \Delta x$,

где $f(x)$ - дифференциальная функция распределения (функция плотности вероятности).

Потому естественно выборочным аналогом функции $f(x)$ считать функцию:

$$f^*(x) = \frac{F^*(x + \Delta x) - F^*(x)}{\Delta x},$$

где $F^*(x + \Delta x) - F^*(x)$ - частота попадания наблюдаемых значений случайной величины X в интервал $[x; x + \Delta x)$. Таким образом, значение $f^*(x)$ характеризует плотность частоты на этом интервале.

Пусть наблюдаемые значения непрерывной случайной величины представлены в виде интервального вариационного ряда.

Полагая, что p_i^* - частота попадания наблюдаемых значений в интервал $[a_i; a_i + h)$, где h - длина частичного интервала, выборочную функцию плотности $f^*(x)$ можно задать соотношением

$$f^*(x) = \begin{cases} 0, & \text{при } x < a_1, \\ \frac{p_i^*}{h}, & \text{при } a_i \leq x < a_{i+1}, \quad i = 1, 2, \dots, m, \\ 0, & \text{при } x \geq a_{m+1}, \end{cases}$$

где a_{m+1} - конец последнего m -го интервала.

Так как функция $f^*(x)$ является аналогом распределения плотности случайной величины, площадь области под графиком этой функции равна 1.

Графическое изображение статистических данных

Статистическое распределение изображается графически с помощью полигона и гистограммы.

Определение. Полигоном частот называют ломаную, отрезки которой соединяют точки с координатами (x_i, n_i) ; полигоном частостей – с координатами (x_i, p_i^*) , где $p_i^* = \frac{n_i}{n}$, $i = \overline{1, m}$.

Полигон служит для изображения дискретного статистического ряда.

Полигон частот является аналогом многоугольника распределения дискретной случайной величины в теории вероятностей.

Определение. Гистограммой частот (частостей) называют ступенчатую фигуру, состоящую из прямоугольников, основания которых расположены на оси **Ox** и длины их равны длинам частичных интервалов (**h**), а высоты равны отношению:

$$\frac{n_i}{h} - \text{для гистограммы частот}; \quad \frac{n_i}{n \cdot h} - \text{для гистограммы частостей}.$$

Гистограмма является графическим изображением интервального ряда.

Площадь гистограммы частот равна **n**, а гистограммы частостей равна 1.

Можно построить полигон для интервального ряда, если его преобразовать в дискретный ряд. В этом случае интервалы заменяют их серединными значениями и ставят в соответствие интервальные частоты (частости). Полигон получим, соединив отрезками середины верхних оснований прямоугольников гистограммы.

Пример. 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

Требуется:

- построить дискретный вариационный ряд;
- найти размах варьирования **R**, моду **M₀**, медиану **M_e**;
- построить полигон частостей.

○ 1) Ранжируем выборку : 12, 12, 13, 13, 13, 14, 14, 14, 14, 14,
15, 15, 16, 16, 17, 17, 17, 18, 18, 19.

2) Находим частоты вариантов и строим дискретный вариационный ряд (табл.3)

Таблица 3.

Значения вариантов x_i	12	13	14	15	16	17	18	19	$\sum_{i=1}^8 n_i = 20$
Частоты n_i	2	3	5	2	2	3	2	1	
Частоты $p_i^* = \frac{n_i}{n}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{5}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\sum_{i=1}^8 p_i = 1$

3) По результатам таблицы 3 находим:

$$R = 19 - 12 = 7, \quad M_0 = 14, \quad M_e = \frac{x_{10} + x_{11}}{2} = \frac{14 + 15}{2} = 14,5$$

4) Строим полигон частотей.

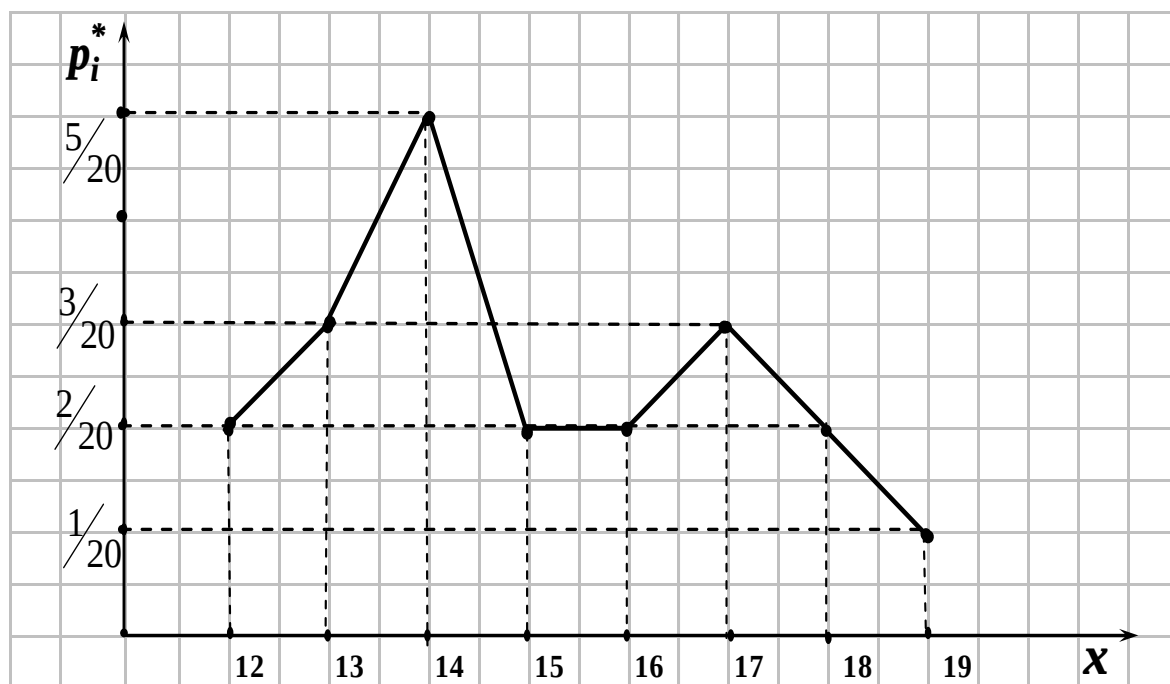


Рис. 2

Пример. Результаты измерений отклонений от нормы диаметров

50 подшипников дали численные значения (в мкм), приведенные в табл. 4.

Таблица 4.

-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 интервальный вариационный ряд;
 - построить гистограмму и полигон частостей.

○ 1. Строим интервальный ряд.

По данным таблицы 4 определяем: $x_{\min} = -1,76$; $x_{\max} = 1,74$

Для определения длины интервала h используем формулу Стерджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg 50}.$$

Число интервалов $m \approx 1 + 3,322 \lg 50$.

$$h = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg 50} = \frac{1,74 - (-1,76)}{1 + 3,322 \lg 50} \approx \frac{3,5}{1 + 3,322 \lg 50} \approx \frac{3,5}{6,644} \approx 0,526$$

Примем $h = 0,6$, $m = 7$.

За начало первого интервала примем величину

$$x_{\text{нач}} = x_{\min} - \frac{h}{2} = -1,76 - 0,3 = -2,06.$$

Конец последнего интервала должен удовлетворять условию:

$$x_{\text{кон}} - h \leq x_{\max} < x_{\text{кон}}.$$

Действительно, $2,14 - 0,6 \leq 1,74 < 2,14$; $1,54 \leq 1,74 < 2,14$.

Строим интервальный ряд (табл. 5).

Таблица 5.

Интервалы	$[-2,06; -1,46)$	$[-1,46; -0,86)$	$[-0,86; -0,26)$	$[-0,26; 0,34)$
Подсчет частот	• •	┌──┐ • ┘	⊠ •	⊠ ┌──┐ • ┘
Частоты n_i	2	6	11	15
Частости p_i	$\frac{2}{50}$	$\frac{6}{50}$	$\frac{11}{50}$	$\frac{15}{50}$

Интервалы	$[0,34; 0,94)$	$[0,94; 1,54)$	$[1,54; 2,14)$
Подсчет частот	⊠ •	• • •	• •
Частоты n_i	11	3	2
Частости p_i	$\frac{11}{50}$	$\frac{3}{50}$	$\frac{2}{50}$

$$\sum_{i=1}^7 n_i = 50;$$

$$\sum_{i=1}^7 p_i = 1.$$

Строим гистограмму частостей.

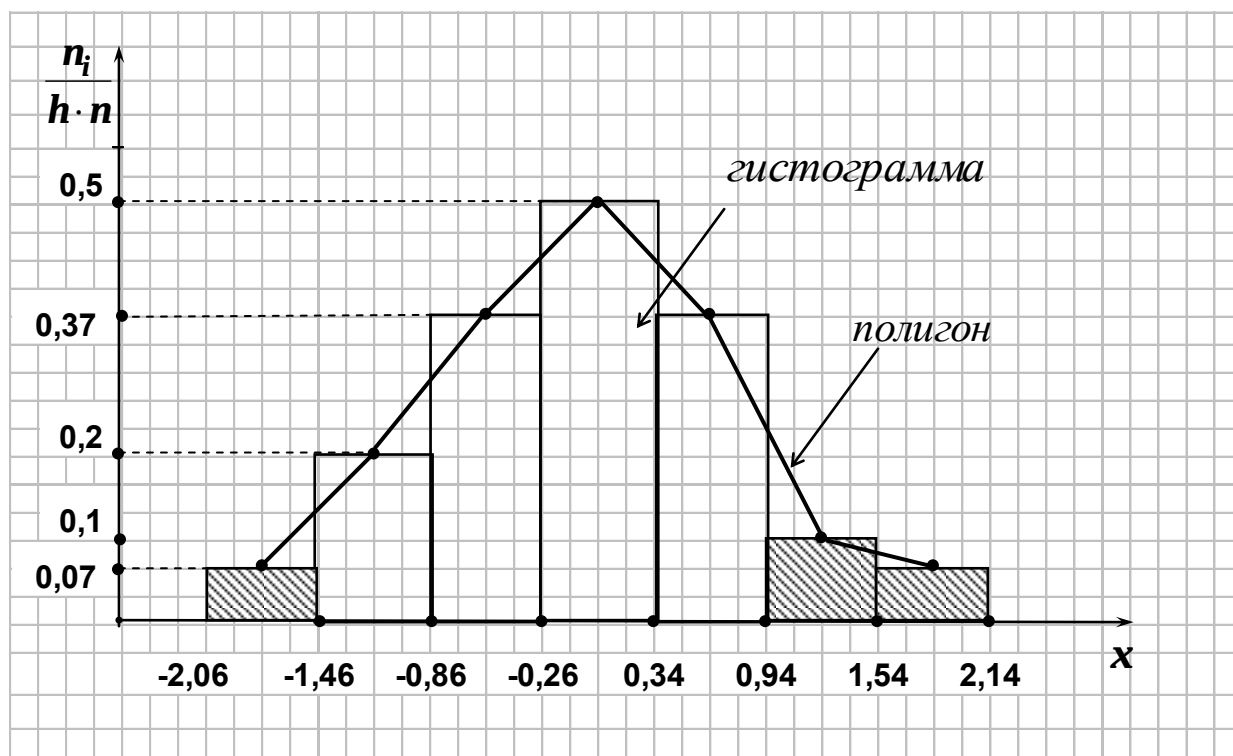


Рис.3

Вершинами полигона являются середины верхних оснований прямоугольников гистограммы.

Убедимся, что площадь гистограммы равна 1.

$$S = h \cdot \left(\frac{n_1 + n_2 + \dots + n_m}{n \cdot h} \right)$$

$$S = 0,6(0,07 + 0,2 + 0,37 + 0,5 + 0,37 + 0,1 + 0,07) = 0,6 \cdot 1,68 = 1,008 \approx 1 \bullet$$

Практическое занятие №4. ГРУБЫЕ ПОГРЕШНОСТИ И МЕТОДЫ ИХ ИСКЛЮЧЕНИЯ

Методические указания

Грубой погрешностью (промахом) называется погрешность, существенно превышающая значение ожидаемой погрешности при данных условиях проведения измерительного эксперимента. Обычно грубая погрешность является следствием значительного внезапного изменения условий эксперимента: броска тока источника электропитания; не учтенное экспериментатором изменение температуры окружающей среды (при длительном эксперименте); неправильный отсчет показаний из-за отвлечения внимания экспериментатора и др.

При однократных измерениях обнаружить промах не представляется возможным. Для уменьшения вероятности появления промахов измерения проводят два-три раза и за результат принимают среднее арифметическое полученных отсчетов. При многократных измерениях для обнаружения промахов используют статистические критерии, такие как критерий Романовского, критерий Шарлье, критерий Диксона.

Для выявления грубых погрешностей задаются уровнем значимости q (вероятностью P) того, что сомнительный результат действительно мог иметь место в данной совокупности результатов измерений.

Критерий Романовского

Критерий Романовского применяется, если число измерений $n < 20$. При этом вычисляется отношение

$$\beta = \frac{|x_i - \bar{x}|}{S_x}, \quad (1)$$

x_i – проверяемое значение (наименьший/наибольший результат измерения); $\bar{x}$ – среднее арифметическое значение измеряемой величины; S_x – среднее квадратическое отклонение (СКО).

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad (2)$$

n – количество измерений.

В зависимости от выбранного уровня значимости, т. е. от желания экспериментатора получить уверенный результат проверки гипотезы, и числа измерений n из табл. 1 находят теоретический критерий Романовского β_T , и сравнивают с ним расчетное значение β . Если $\beta \geq \beta_T$, то результат x_i считается промахом и отбрасывается.

Таблица 1

Значения критерия Романовского $\beta_T = f(n)$

q	$n = 4$	$n = 6$	$n = 8$	$n = 10$	$n = 12$	$n = 15$	$n = 20$
0,01	1,73	2,16	2,43	2,62	2,75	2,90	3,08
0,02	1,72	2,13	2,37	2,54	2,66	2,80	2,96
0,05	1,71	2,10	2,27	2,41	2,52	2,64	2,78
0,10	1,69	2,00	2,17	2,29	2,39	2,49	2,62

Пример решения

При шестикратном измерении расстояний между ориентирами осей зданий получены следующие результаты:

i	1	2	3	4	5	6
x_i , м	25,155	25,150	25,165	25,165	25,160	25,180

Последний результат вызывает сомнения. Произведем проверку по критерию Романовского, не является ли он промахом?

Находим среднее арифметическое значение:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{150,957}{6} = 25,163 \text{ м.}$$

По формуле (2) определяем среднее квадратическое отклонение. Для удобства вычислений составим табл. 2.

Оценка СКО:

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 0,0107 \text{ м.}$$

Обработка результатов измерений

№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	25,155	-0,008	0,000064
2	25,15	-0,013	0,000169
3	25,165	0,002	0,000004
4	25,165	0,002	0,000004
5	25,16	-0,003	0,000009
6	25,18	0,017	0,000289
	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{150,957}{6} = 25,163$ и		$\sum_{i=1}^n (x_i - \bar{x})^2 = 0,000539$

Вычисляем β для сомнительного результата измерения (при $n = 6$)

$$\beta = \frac{|x_i - \bar{x}|}{S_x} = \frac{|25,18 - 25,163|}{0,0107} = 1,58.$$

Выводы: критическое значение β при уровне значимости $q = 0,05$ (доверительная вероятность $P = 0,95$) для количества измерений $n = 6$ составляет 2,1. Поскольку $1,58 < 2,1$ ($\beta < \beta_T$), результат не является промахом и не исключается из результатов измерений.

Критерий Шарлье

Критерий Шарлье используется, если число измерений велико ($n > 20$). Тогда по теореме Бернулли число результатов, превышающих по абсолютного значению среднее арифметическое значение на величину $\hat{E}_0 \cdot S_x$, будет $n[1 - \Phi(\hat{E}_0)]$, где $\Phi(\hat{E}_0)$ – значение нормированной функции Лапласа для $X = K_{ш}$.

Если сомнительным в ряду результатов наблюдений является один результат, то

$$n[1 - \Phi(\hat{E}_0)] = 1.$$

Отсюда

$$\Phi(K_{ш}) = \frac{n-1}{n}.$$

Значения критерия Шарлье приведены в таблице 3.

Таблица 3

Значения критерия Шарлье

n	5	10	20	30	40	50	100
$K_{ш}$	1,3	1,65	1,96	2,13	2,24	2,32	2,58

Пользуясь данным критерием, отбрасывается результат, для значения которого выполняется неравенство [1]

$$|x_i - \bar{x}| > \hat{E}_0 \cdot S_x.$$

Пример решения

При измерении расстояний между колоннами были получены следующие результаты (тал. 4, значения x_i).

Таблица 4

Обработка исходных данных

№ п/п	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	Проверка по критерию Шарлье
1	23,67	0,00	0,0000	не является промахом
2	23,68	0,01	0,0001	не является промахом
3	23,66	-0,01	0,0001	не является промахом
4	23,67	0,00	0,0000	не является промахом
5	23,67	0,00	0,0000	не является промахом
6	23,68	0,01	0,0001	не является промахом
7	23,67	0,00	0,0000	не является промахом
8	23,68	0,01	0,0001	не является промахом
9	23,67	0,00	0,0000	не является промахом
10	23,68	0,01	0,0001	не является промахом
11	23,66	-0,01	0,0001	не является промахом
12	23,67	0,00	0,0000	не является промахом
13	23,67	0,00	0,0000	не является промахом
14	23,68	0,01	0,0001	не является промахом
15	23,68	0,01	0,0001	не является промахом
16	23,68	0,01	0,0001	не является промахом
17	23,67	0,00	0,0000	не является промахом
18	23,68	0,01	0,0001	не является промахом
19	23,68	0,01	0,0001	не является промахом
20	23,67	0,00	0,0000	не является промахом
21	23,68	0,01	0,0001	не является промахом
22	23,67	0,00	0,0000	не является промахом
23	23,67	0,00	0,0000	не является промахом
24	23,67	0,00	0,0000	не является промахом
25	23,68	0,01	0,0001	не является промахом
26	23,66	-0,01	0,0001	не является промахом
27	23,68	0,01	0,0001	не является промахом

28	23,67	0,00	0,0000	не является промахом
29	23,67	0,00	0,0000	не является промахом
30	23,68	0,01	0,0001	не является промахом
	$\bar{x} = 23,6733$		$\sum_{i=1}^{30} (x_i - \bar{x})^2 = 0,0016$	$S_x = 0,0074$ $\hat{E}_0 \cdot S_x = 0,0158$

Обработка результатов измерений

Находим СКО:

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{0,0016}{29}} = 0,0074.$$

Проверяем ряд измерений на наличие промаха. Если условие $|x_i - \bar{x}| > \hat{E}_0 \cdot S_x$ выполняется, то результат измерения x_i отбрасывается.

Критерий Шарлье для числа измерений $n = 30$

$$K_{ш} = 2,13.$$

Таким образом, проверяемые значения не являются промахом (см. табл. 4) и не отбрасываются из ряда измерений.

Критерий Диксона

Вариационный критерий Диксона удобный и достаточно мощный (с малыми вероятностями ошибок). При его применении полученные результаты наблюдений записывают в вариационный возрастающий ряд $x_1, x_2, \dots, x_n (x_1 < x_2 < \dots < x_n)$. Критерий Диксона определяется как

$$\hat{E}_A = \frac{x_n - x_{n-1}}{x_n - x_1}.$$

Критическая область для этого критерия $P(K_A > Z_q) = q$. Значения Z_q приведены в табл. 5.

Таблица 5

Значения критерия Диксона

n	Z_q при q , равном			
	0,10	0,05	0,02	0,01
4	0,68	0,76	0,85	0,89
5	0,56	0,64	0,78	0,82
6	0,48	0,56	0,64	0,70
8	0,40	0,47	0,54	0,59
10	0,35	0,41	0,48	0,53
14	0,29	0,35	0,41	0,45
16	0,28	0,33	0,39	0,43
18	0,26	0,31	0,37	0,41
20	0,26	0,30	0,36	0,39

30	0,22	0,26	0,31	0,34
----	------	------	------	------

Пример решения

Было проведено пять измерений напряжения в электросети. Получены следующие данные:

<i>i</i>	1	2	3	4	5
<i>x_i</i>, В	127,1	127,2	126,9	127,6	127,2

Результат 127,6 В существенно (на первый взгляд) отличается от остальных. Необходимо проверить, не является ли он промахом.

Составим вариационный ряд из результатов измерений напряжения в электросети:

<i>i</i>	1	2	3	4	5
<i>x_i</i>, В	126,9	127,1	127,2	127,2	127,6

Для крайнего члена этого ряда (127,6) критерий Диксона

$$\hat{E}_A = \frac{127,6 - 127,2}{127,6 - 126,9} = \frac{0,4}{0,7} = 0,57.$$

Как следует из табл. 5, по этому критерию результат 127,6 В может быть отброшен как промах лишь на уровне значимости ***q*** = 0,10.

Таблица 6

Исходные данные

Вариант	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆
1	484	485	484	485	483	492	485	484	485	482	481	484	494	485	484	483
2	15,1	15,2	15,5	15,4	15,5	15,6	15,3	15,4	15,4	15,5	15,3	15,5	15,4	15,6	16,2	15,4
3	5,8	6,1	5,7	5,6	5,4	5,6	5,5	5,4	5,6	5,5	5,3	5,1	5,6	5,4	5,5	5,4
4	1,6	1,5	1,7	1,5	1,4	1,6	1,5	1,8	2,2	1,5	1,6	1,7	1,4	1,5	1,4	1,5
5	6,6	6,5	6,8	6,9	6,4	6,5	6,6	6,5	6,7	6,5	7,3	6,4	6,5	6,5	6,5	6,6
6	10,3	10,1	10,2	10,1	10,3	10,2	10,9	11,2	10,4	10,3	10,4	10,3	10,2	10,1	10,3	10,2
7	15,5	15,3	15,3	15,4	15,3	15,2	15,6	15,4	15,3	15,2	15,8	15,4	16,2	15,5	15,3	15,4
8	11,8	11,7	11,8	11,9	11,6	11,5	11,6	11,8	11,7	11,8	11,6	11,9	11,7	10,6	11,6	11,9
9	5,6	5,5	5,8	5,3	5,5	5,6	5,4	5,9	5,5	5,6	5,7	5,4	5,5	5,7	6,3	5,4
10	4,8	4,6	4,7	4,8	4,6	4,8	4,9	4,6	4,8	4,7	4,8	4,6	4,8	3,9	4,7	4,5
11	2,5	2,7	2,8	2,5	2,3	2,2	2,5	2,3	2,4	2,5	2,6	2,9	3,2	2,6	2,1	2,5
12	4,5	4,3	4,1	4,8	4,6	4,8	4,9	4,6	4,8	4,7	4,8	4,7	4,8	3,9	4,5	4,6
13	12,6	12,8	12,4	12,5	12,5	12,2	12,4	12,6	12,2	12,4	11,5	12,3	12,5	12,7	12,4	12,3
14	9,3	9,4	9,1	9,2	9,5	9,2	9,4	9,3	9,4	9,5	10,6	9,4	9,2	9,5	9,3	9,2
15	5,8	5,9	6,2	5,8	5,6	5,7	6,1	5,9	5,8	6,9	5,8	5,7	5,8	5,7	5,9	5,8
16	4,3	4,4	4,6	4,2	4,3	4,6	4,5	4,3	4,6	4,9	4,3	4,6	4,3	3,8	4,7	4,9
17	3,1	3,4	3,2	3,5	3,1	3,6	3,2	3,3	3,4	3,3	3,4	3,5	3,3	3,2	3,5	3,3
18	10,6	10,2	10,5	10,3	10,4	10,3	10,5	10,3	10,6	10,1	10,5	10,4	11,4	10,5	10,3	10,4
19	54,8	54,6	54,7	54,6	54,9	54,7	54,7	54,8	53,9	54,6	54,7	54,8	54,7	54,5	54,6	54,6
20	25,6	25,5	25,3	25,5	25,6	25,9	25,5	25,7	25,4	25,7	25,4	25,5	25,7	26,3	25,4	25,6

Практическое занятие №5.

Точечные оценки

Статистической оценкой Θ^* неизвестного параметра Θ теоретического распределения называют функцию

$f(X_1, X_2, \dots, X_n)$ от наблюдаемых случайных величин $X_1, X_2, \dots, X_n$.

Точечной называют статистическую оценку, которая определяется одним числом $\Theta^* = f(x_1, x_2, \dots, x_n)$, где $x_1, x_2, \dots, x_n$ - результаты n наблюдений над количественным признаком X (выборка).

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя

$$\bar{x}_B = (\sum_{i=1}^k n_i x_i) / n,$$

где x_i - варианты выборки, n_i - частота варианты x_i , $n = \sum_{i=1}^k n_i$ - объем выборки.

Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

$$D_B = (\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2) / n;$$

эта оценка является смещенной, так как

$$M[D_B] = \frac{n-1}{n} D_{\Gamma}.$$

Более удобна формула

$$D_B = \bar{x}^2 - [\bar{x}]^2 = \frac{\sum n_i x_i^2}{n} - \left[\frac{\sum n_i x_i}{n} \right]^2.$$

Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum n_i (x_j - \bar{x}_B)^2}{n-1},$$

1. Из генеральной совокупности извлечена выборка объема $n=50$:

варианта x_i	2	5	7	10
частота n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

2. Задано распределение первоначальных вариантов выборки объема n :

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Доказать, что

$$\bar{x}_B = C + (\sum n_i u_i) / n,$$

где условные варианты $u_i = x_i - C$.

3. Найти выборочную среднюю по данному распределению выборки объема $n=10$:

x_i	1250	1270	1280
n_i	2	5	3

4. Найти выборочную среднюю по данному распределению выборки объема $n=20$:

x_i	25601	2600	2620	2650	2700
n_i	2	3	10	4	1

Указание. Перейти к условным вариантам $u_i = x_i - 2620$.

5. По выборке объема $n=41$ найдена смещенная оценка $DB=3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

6. Ниже приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Рост	154-158	158-162	162-166	166-170	170-174	174-178	178-182
Число студентов	10	14	26	28	12	8	2

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

Указание. Найти середины интервала и принять их в качестве вариант.

Метод моментов

Метод моментов точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов соответствующим эмпирическим моментам того же порядка.

Если распределение определяется одним параметром, то для его отыскания приравнивают один теоретический момент одному эмпирическому моменту того же порядка. Например, можно приравнять начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка: $\nu_1 = M_1$. Учитывая, что $\nu_1 = M(X)$ и $M_1 = \bar{x}_B$, получим

$$M(X) = \bar{x}_B. \quad (*)$$

Математическое ожидание является функцией от неизвестного параметра заданного распределения, поэтому, решив уравнение (*) относительно неизвестного параметра, тем самым получим его точечную оценку.

1. Случайная величина X распределена по закону Пуассона

$$P_m(x_i) = \lambda^{x_i} e^{-\lambda} / x_i!,$$

где m - число испытаний, произведенных в одном опыте; x_i - число появлений события в i -м опыте.

Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ , определяющего распределение Пуассона.

2. Случайная величина X (число семян сорняков в пробе зерна) распределена по закону Пуассона. Ниже приведено распределение семян сорняков в $n=1000$ пробах зерна (в первой строке указано количество x_i сорняков в одной пробе; во второй строке указана частота n_i - число проб, содержащих x_i семян сорняков):

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

3. Случайная величина X (число нестандартных изделий в партии изделий) распределена по закону Пуассона. Ниже приведено распределение нестандартных изделий в $n=200$ партиях (в первой строке указано количество x_i нестандартных изделий в одной партии; во второй строке указана частота n_i - число партий, содержащих x_i нестандартных изделий):

x_i	0	1	2	3	4
n_i	132	43	20	3	2

Найти методом моментов точечную оценку неизвестного параметра λ распределения Пуассона.

4. Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку параметра p биномиального распределения

$$P_m(x_i) = C_m^{x_i} p^{x_i} (1-p)^{m-x_i},$$

где x_i - число появлений события в i -м опыте ($i = 1, 2, \dots, n$), m - количество испытаний в одном опыте.

Указание. Приравнять начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка.

5. Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x}$ ($x \geq 0$).

Метод наибольшего правдоподобия

Метод наибольшего правдоподобия точечной оценки неизвестных параметров заданного распределения сводится к отысканию максимума функции одного или нескольких оцениваемых параметров.

А. Дискретные случайные величины. Пусть X - дискретная случайная величина, которая в результате n опытов приняла возможные значения $x_1, x_2, \dots, x_n$. Допустим, что вид закона распределения величины X задан, но неизвестен параметр Θ , которым определяется этот закон: требуется найти его точечную оценку $\Theta^* = \Theta^*(x_1, x_2, \dots, x_n)$.

Обозначим вероятность того, что в результате испытания величина X примет значение x_i через $p(x_i; \Theta)$.

Функцией правдоподобия дискретной случайной величины X называют функцию аргумента Θ :

$$L(x_1, x_2, \dots, x_n; \Theta) = p(x_1; \Theta)p(x_2; \Theta) \dots p(x_n; \Theta).$$

Оценкой наибольшего правдоподобия параметра Θ называют такое его значение Θ^* , при котором функция правдоподобия достигает максимума.

Функции L и $\ln L$ достигают максимума при одном и том же значении Θ , поэтому вместо отыскания максимума функции L ищут, что удобнее, максимум функции $\ln L$.

Б. Непрерывные случайные величины. Пусть X - непрерывная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2, \dots, x_n$. Допустим, что вид плотности распределения - функции $f(x)$ - задан, но неизвестен параметр Θ , которым определяется эта функция.

Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента Θ :

$$L(x_1, x_2, \dots, x_n; \Theta) = f(x_1; \Theta)f(x_2; \Theta) \dots f(x_n; \Theta).$$

1. Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра p (вероятность появления события в одном испытании) биномиального распределения:

$$P_m(x_i) = C_m^{x_i} p^{x_i} (1-p)^{m-x_i},$$

где x_i - число появлений события в i -м опыте, m - количество испытаний в одном опыте, n - число опытов.

2. Случайная величина X (число появлений события A в m независимых испытаниях) подчинена закону распределения Пуассона с неизвестным параметром λ :

$$P(X = x_i) = \lambda^{x_i} e^{-\lambda} / x_i!,$$

где m - число испытаний в одном опыте, x_i - число появлений события в i -м опыте ($i = 1, 2, \dots, n$).

Найти методом наибольшего правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ распределения Пуассона.

3. Найти методом наибольшего правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x} (x \geq 0)$.

4. Найти методом наибольшего правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечную оценку параметра β гамма-распределения (параметр α известен), плотность которого

$$f(x) = \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} x^{\alpha} e^{-x/\beta} \quad (\alpha > -1, \beta > 0, x \geq 0).$$

5. Найти методом наибольшего правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечную оценку параметра p геометрического распределения:

$$P(X = x_i) = (1-p)^{x_i-1} p,$$

где x_i - число испытаний, произведенных до появления события; p - вероятность появления события в одном испытании.

Практическое занятие №6.

Интервальные оценки. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии

При выборке малого объема точечная оценка может существенно отличаться от оцениваемого параметра. В этом случае целесообразно использовать интервальные оценки.

Определение. Интервальной называют оценку, которая определяется двумя числами – концами интервала.

Пусть найденная по данным выборки величина $\bar{\theta}$ служит оценкой неизвестного параметра θ . Оценка $\bar{\theta}$ определяет θ тем точнее, чем меньше $|\theta - \bar{\theta}|$, то есть чем меньше δ в неравенстве $|\theta - \bar{\theta}| < \delta$ ($\delta > 0$).

Поскольку $\bar{\theta}$ - случайная величина, то и разность $|\theta - \bar{\theta}|$ - случайная величина. Поэтому неравенство $|\theta - \bar{\theta}| < \delta$, при заданном δ может выполняться только с некоторой вероятностью.

Определение. Доверительной вероятностью (надежностью) оценки $\bar{\theta}$ параметра θ называется вероятность γ , с которой выполняется неравенство $|\theta - \bar{\theta}| < \delta$.

Обычно задается надежность γ и определяется δ . Чаще всего надежность задается значениями от 0,95 и выше, в зависимости от конкретно решаемой задачи.

Неравенство $|\theta - \bar{\theta}| < \delta$ можно записать $\bar{\theta} - \delta < \theta < \bar{\theta} + \delta$.

Определение. Доверительным интервалом называется интервал $(\bar{\theta} - \delta; \bar{\theta} + \delta)$, который покрывает неизвестный параметр с заданной надежностью γ .

Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии

Пусть случайная величина $\mathbf{X}$ имеет нормальное распределение: $N(\mathbf{a}; \sigma)$.

Известно значение σ и задана доверительная вероятность (надежность) γ . Требуется построить доверительный интервал для параметра $\mathbf{a}$ по выборочному среднему $\overline{\mathbf{X}}_g$.

Чтобы подчеркнуть случайный характер $\overline{\mathbf{X}}_g$ обозначим его $\overline{\mathbf{X}}_g$.

Примем без доказательства, что если случайная величина $\mathbf{X}$ распределена нормально, то и выборочное среднее $\overline{\mathbf{X}}_g$, найденное по независимым наблюдениям, также распределено нормально.

Параметры распределения $\overline{\mathbf{X}}_g$ таковы: $M(\overline{\mathbf{X}}_g) = \mathbf{a}$; $\sigma(\overline{\mathbf{X}}_g) = \frac{\sigma}{\sqrt{n}}$.

Из теории вероятности известна формула для нормально распределенной случайной величины $\mathbf{X}$:

$$P(|\mathbf{X} - \mathbf{a}| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right),$$

где $\Phi(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \int_0^{\mathbf{x}} e^{-\frac{t^2}{2}} dt$ - функция Лапласа, значение которой в точке

$$\frac{\delta}{\sigma}$$

находим по таблице (Приложение 2).

Учитывая, что $\overline{\mathbf{X}}_g$ имеет нормальное распределение можно записать

$$P(|\overline{\mathbf{X}}_g - \mathbf{a}| < \delta) = 2\Phi\left(\frac{\delta}{\sigma(\overline{\mathbf{X}}_g)}\right) = \gamma \quad \text{или} \quad \gamma = 2\Phi\left(\frac{\delta \cdot \sqrt{n}}{\sigma}\right) = 2\Phi(\mathbf{t}),$$

где $\frac{\delta \cdot \sqrt{n}}{\sigma} = t$

Из последнего равенства по таблице Лапласа находим t (Приложение 2).

Тогда $\delta = t \cdot \frac{\sigma}{\sqrt{n}}$ и доверительный интервал

$$\left(\overline{X}_e - t \cdot \frac{\sigma}{\sqrt{n}}; \overline{X}_e + t \cdot \frac{\sigma}{\sqrt{n}} \right)$$

покрывает с надежностью γ математическое ожидание a .

Пример. Случайная величина имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$. Найти доверительный интервал оценки неизвестного математического ожидания по выборочной средней $\overline{X}_e$, если объем выборки $n = 36$, а надежность оценки $\gamma = 0,95$.

○ 1. Находим t : $2\Phi(t) = 0,95$ $\Phi(t) = 0,475$

По таблице значений функции Лапласа $t = 1,96$.

2. Определяем $\delta = t \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{3}{\sqrt{36}} = 0,98$.

Доверительный интервал запишется в виде: $(\overline{x}_e - 0,98; \overline{x}_e + 0,98)$. ●

Практическое занятие №7.

Интервальные оценки. Доверительный интервал для оценки математического ожидания при неизвестной дисперсии

Пусть случайная величина X имеет нормальное распределение: $N(a; \sigma)$, причем σ - неизвестно, γ - задана.

Если $D(X)$ неизвестна, то пользуются оценкой S^2 .

Введем случайную величину
$$T = \frac{\overline{X}_e - a}{\frac{S}{\sqrt{n}}},$$

где S - исправленное среднее квадратическое отклонение случайной величины X , вычисленное по выборке:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_e)^2};$$

Случайная величина T имеет распределение Стьюдента с $(n-1)$ степенью свободы.

Тогда доверительный интервал для оценки $a = M(X)$ имеет вид:

$$\left(\overline{X}_e - t_j \cdot \frac{S}{\sqrt{n}}; \overline{X}_e + t_j \cdot \frac{S}{\sqrt{n}} \right),$$

где $\overline{X}_e$ - выборочное среднее;

S - исправленное среднее квадратическое отклонение;

t_j - находим по таблице квантилей распределения Стьюдента (Приложение 4) в зависимости от числа степеней свободы и доверительной вероятности γ .

Пример. Произведено пять независимых наблюдений над случайной величиной $X \sim N(a; \sigma)$. Результаты наблюдений таковы:

$$\mathbf{x}_1 = 35, \mathbf{x}_2 = 20, \mathbf{x}_3 = 15, \mathbf{x}_4 = -12, \mathbf{x}_5 = 42.$$

Построить для неизвестного $\mathbf{M}(\mathbf{x}) = \mathbf{a}$ доверительный интервал, если $\gamma = 0,95$.

$$\bigcirc 1. \text{ Находим } \overline{\mathbf{x}_g}: \quad \overline{\mathbf{x}_g} = \frac{1}{5}(-35 + 20 + 15 - 12 + 42) = \frac{1}{5} \cdot 30 = 6$$

$$\underline{\overline{\mathbf{x}_g} = 6}$$

2. Находим $\mathbf{S}^2$:

$$\begin{aligned} \mathbf{S}^2 &= \frac{1}{4}((-35-6)^2 + (20-6)^2 + (15-6)^2 + (-12-6)^2 + (42-6)^2) = \\ &= \frac{1}{4}((-41)^2 + 16^2 + 9^2 + (-18)^2 + 36^2) = \frac{1}{4}(1681 + 256 + 81 + 324 + 1296) = \\ &= \frac{1}{4}3638 = 909,5 \end{aligned}$$

$$\underline{\mathbf{S} = \sqrt{909,5} \approx 30,2}$$

3. По таблице квантилей распределения Стьюдента (Приложение 4) для $\gamma = 0,95$ и $\mathbf{n}-1 = 4$ находим $\mathbf{t}_j$:

$$\underline{\mathbf{t}_j = 2,78}$$

Доверительный интервал:

$$\left(6 - 2,78 \frac{30,2}{2,24}; 6 + 2,78 \frac{30,2}{2,24} \right) \quad \text{или} \quad (31,5; 43,5). \bullet$$

Практическое занятие №8.

Интервальные оценки. Доверительный интервал для оценки среднего квадратического отклонения нормального распределения

1. Если $M(\mathbf{X}) = \mathbf{a}$ неизвестно, то доверительный интервал для оценки $\sigma(\mathbf{X})$ имеет вид:

$$\left(\frac{\sqrt{n-1} \cdot \mathbf{S}}{\chi_2}; \frac{\sqrt{n-1} \cdot \mathbf{S}}{\chi_1} \right)$$

где n - объем выборки; $\mathbf{S}$ - исправленное среднее квадратическое отклонение:

$$\mathbf{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}}_e)^2,$$

$$\chi_1^2 = \chi_{\frac{1+\gamma}{2}; n-1}^2, \quad \chi_2^2 = \chi_{\frac{1-\gamma}{2}; n-1}^2 - \text{квантили } \chi^2 - \text{распределения,}$$

определяемые по таблице $\chi_{\alpha, k}^2$ (Приложение 3)

$$\text{при } k = n-1 \text{ и } \alpha = \frac{1+\gamma}{2}, \quad \alpha = \frac{1-\gamma}{2}.$$

Пример. Для оценки параметра $\sigma(\mathbf{X})$ нормально распределенной случайной величины была сделана выборка объема в 25 единиц и вычислено $\mathbf{S} = 0,8$.

Найти доверительный интервал, покрывающий σ с вероятностью $\gamma = 0,95$.

○ Имеем $n = 25$, $\gamma = 0,95$.

$$\chi_1^2 = \chi_{\frac{1+0,95}{2}; 25-1}^2 = \chi^2(0,975; 24) = 12,4$$

$$\chi_2^2 = \chi_{\frac{1-0,95}{2}; -1}^2 = \chi^2(0,025; 24) = 39,4$$

Доверительный интервал имеет вид:

$$\left(\frac{\sqrt{24} \cdot 0,8}{\sqrt{39,4}}; \frac{\sqrt{24} \cdot 0,8}{\sqrt{12,4}} \right) \quad \text{или} \quad (0,79; 1,4) . \bullet$$

2. Другой вид доверительного интервала для оценки $\sigma(\mathbf{X})$ нормального распределения имеет вид:

$$S(1-q) < \sigma < S(1+q) \quad \text{при} \quad q < 1;$$

$$0 < \sigma < S(1+q) \quad \text{при} \quad q > 1;$$

где S - исправленное среднее квадратическое отклонение;

$q = q(\gamma; n)$ находим по таблице значений (Приложение 5).

Пример. Для оценки параметра нормально распределенной случайной величины была сделана выборка объема в 25 единиц и вычислено $S = 0,8$.

Найти доверительный интервал, покрывающий σ с вероятностью $\gamma = 0,95$.

○ Имеем $n = 25$, $\gamma = 0,95$, $S = 0,8$

По таблице значений $q = q(\gamma; n)$ находим $q = 0,32$.

Доверительный интервал имеет вид:

$$(0,8(1-0,32); 0,8(1+0,32)) \quad \text{или} \quad (0,544; 1,056) . \bullet$$

Замечание. Доверительные интервалы в примерах 8 и 9 получили разные при одинаковых данных, но они с вероятностью $\gamma = 0,95$ покрывают среднее квадратическое отклонение $\sigma(\mathbf{X})$.

Практическое занятие №9.

Сравнение двух дисперсий нормальных генеральных совокупностей

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Нулевой (основной) называют выдвинутую гипотезу H_0 . *Конкурирующей* (альтернативной) называют гипотезу H_0 , которая противоречит нулевой.

Простой называют гипотезу, содержащую только одно предположение. *Сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

В итоге проверки гипотезы могут быть допущены ошибки двух родов. Ошибка *первого рода* состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность ошибки первого рода называют уровнем значимости и обозначают через α . Ошибка *второго рода* состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность ошибки второго рода обозначают через β .

Синтаксическим критерием (или просто критерием) называют случайную величину K , которая служит для проверки гипотезы.

Наблюдаемым (эмпирическим) значением $K_{набл.}$ называют то значение критерия, которое вычислено по выборкам.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия принадлежит критической области, то нулевую гипотезу отвергают; если наблюдаемое значение критерия принадлежит области принятия гипотезы, то гипотезу принимают.

Критическими точками (границами) $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ - положительное число.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр}$ - отрицательное число.

Двусторонней называют критическую область, определяемую неравенством $K < k_1, K > k_2$, где $k_2 > k_1$.

Для отыскания критической области задаются уровнем значимости α и ищут критические точки, исходя из следующих соотношений:

а) для правосторонней критической области

$$P(K > k_{кр.}) = \alpha, (k_{кр.} > 0);$$

б) для левосторонней критической области

$$P(K < k_{кр.}) = \alpha, (k_{кр.} < 0);$$

в) для двусторонней симметричной области

$$P(K > k_{кр.}) = (\alpha/2) (k_{кр.} > 0), P(K < k_{кр.}) = \alpha/2.$$

Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Другими словами, мощность критерия есть вероятность того, что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.

Сравнение двух дисперсий нормальных генеральных совокупностей

По независимым выборкам, объемы которых n_1, n_2 , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии S^2_X и S^2_Y . Требуется сравнить эти дисперсии.

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: D(X)=D(Y)$ о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе $H_1: D(X)>D(Y)$, надо вычислить наблюдаемое значение критерия (отношение большей исправленной дисперсии к меньшей) $F_{набл} = s_B^2/s_M^2$ и по таблице критических точек распределения Фишера-Снедекора, по заданному уровню значимости α и числам степеней свободы $k_1=n_1-1, k_2=n_2-1$ (k_1 - число степеней свободы большей исправленной дисперсии) найти критическую точку $F_{кр}(\alpha; k_1; k_2)$. Если $F_{набл} < F_{кр}$ - нет оснований отвергнуть нулевую гипотезу. Если $F_{набл} > F_{кр}$ - нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1: D(X) \neq D(Y)$ критическую точку $F_{кр}(\alpha/2; k_1; k_2)$ ищут по уровню значимости $\alpha/2$ (вдвое меньшему заданного) и числам степеней свободы k_1 и k_2 (k_1 – число степеней свободы большей дисперсии). Если $F_{набл} < F_{кр}$ – нет оснований отвергнуть нулевую гипотезу. Если $F_{набл} > F_{кр}$ – нулевую гипотезу отвергают.

Пример. По двум независимым выборкам, объемы которых $n_1=11$ и $n_2=14$, извлеченным из нормальных генеральных совокупностей X и Y , найдены исправленные выборочные дисперсии $s_X^2 = 0,76$ и $s_Y^2 = 0,38$. При уровне значимости $\alpha=0,05$, проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о равенстве генеральных дисперсий, при конкурирующей гипотезе $H_1: D(X) > D(Y)$.

Решение. Найдем отношение большей исправленной дисперсии к меньшей: $F_{набл} = 0,76/0,38 = 2$. По условию конкурирующая гипотеза имеет вид $D(X) > D(Y)$, поэтому критическая область – правосторонняя.

По таблице приложения В (Критические точки распределения F Фишера – Снедекора), по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n_1 - 1 = 11 - 1 = 10$ и $k_2 = n_2 - 1 = 14 - 1 = 13$ находим критическую точку $F_{кр}(0,05; 10; 13) = 2,67$.

Так как $F_{набл} < F_{кр}$ – нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначимо.

1. По двум независимым выборкам, объемы которых $n_1=9$ и $n_2=16$, извлеченным из нормальных генеральных совокупностей X и Y , найдены исправленные выборочные дисперсии $S_x^2=34,02$ и $S_y^2=12,15$. при уровне значимости $0,01$, проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X) > D(Y)$.

Пример. По двум независимым выборкам, объемы которых $n_1=14$ и $n_2=10$, извлеченным из нормальных генеральных совокупностей X и Y , найдены исправленные выборочные дисперсии $s_X^2 = 0,84$ и $s_Y^2 = 2,52$. При уровне значимости $\alpha=0,1$, проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о

равенстве генеральных дисперсий, при конкурирующей гипотезе $H_1: D(X) \neq D(Y)$.

Решение. Найдем отношение большей исправленной дисперсии к меньшей: $F_{набл} = 2,52/0,84 = 3$. По условию конкурирующая гипотеза имеет вид $D(X) \neq D(Y)$, поэтому критическая область—двусторонняя. В соответствии с правилом 2 при отыскании критической точки следует брать уровень значимости, вдвое меньший заданного.

По таблице приложения В, по уровню значимости $\alpha/2 = 0,1/2 = 0,05$ и числам степеней свободы $k_1 = 10 - 1 = 9$ и $k_2 = 14 - 1 = 13$, находим критическую точку $F_{кр}(0,05; 9; 13) = 2,72$.

Так как $F_{набл.} > F_{кр}$ — нулевую гипотезу о равенстве генеральных дисперсий отвергаем.

2. Двумя методами проведены измерения одной и той же физической величины. Получены следующие результаты:

а) в первом случае $x_1=9,6$; $x_2=10,0$; $x_3=9,8$; $x_4=10,2$; $x_5=10,6$;

б) во втором случае $y_1=10,4$; $y_2=9,7$; $y_3=10,0$; $y_4=10,3$.

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости $\alpha = 0,1$? Предполагается, что результаты измерений распределены нормально и выборки независимы.

Решение. Будем судить о точности методов по величинам дисперсий. Таким образом, нулевая гипотеза имеет вид $H_0: D(X) = D(Y)$. В качестве конкурирующей примем гипотезу $H_1: D(X) \neq D(Y)$. Найдем выборочные дисперсии. Для упрощения вычислений перейдем к условным вариантам:

$u_i = 10x_i - 100$, $v_i = 10y_i - 100$. В итоге получим условные варианты

u_i	-4	0	-2	2	6
v_i	4	-3	0	3	

Найдем исправленные выборочные дисперсии:

$$s_u^2 = \frac{\sum u_i^2 - [\sum u_i]^2 / n_1}{n_1 - 1} = \frac{(16 + 4 + 4 + 36) - 2^2 / 5}{5 - 1} = 14,8;$$

$$s_v^2 = \frac{\sum v_i^2 - [\sum v_i]^2 / n_2}{n_2 - 1} = \frac{(16 + 9 + 9) - 4^2 / 4}{4 - 1} = 10.$$

Сравним дисперсии. Найдем отношение большей исправленной дисперсии к меньшей (каждая из дисперсий увеличилась в 10^2 раз, но их отношение не изменилось):

$$F_{набл} = S_X^2 / S_Y^2 = S_u^2 / S_v^2 = 14,8 / 10 = 1,48.$$

По условию конкурирующая гипотеза имеет вид $D(X) \neq D(Y)$, поэтому критическая область двусторонняя и в соответствии с правилом 2 при отыскании критической точки следует брать уровень значимости, вдвое меньший заданного.

По таблице приложения В, по уровню значимости $\alpha/2 = 0,1/2 = 0,05$ и числам степеней свободы $k_1 = n_1 - 1 = 5 - 1 = 4$ и $k_2 = n_2 - 1 = 4 - 1 = 3$ находим критическую точку $F_{кр}(0,05; 4; 3) = 9,12$.

Так как $F_{набл} < F_{кр}$ — нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий. Другими словами, исправленные дисперсии различаются незначимо и, следовательно, оба метода обеспечивают одинаковую точность измерений.

3. Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объемы которых $n_1 = 10$ и $n_2 = 8$. В результате измерения контролируемого размера отобранных изделий получены следующие результаты:

x_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
y_i	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,28		

Можно ли считать, что станки обладают одинаковой точностью $H_0: D(X) = D(Y)$, если принять уровень значимости $\alpha = 0,1$ и в качестве конкурирующей гипотезы $H_1: D(X) \neq D(Y)$?

Указание. Для упрощения вычислений перейти к условным вариантам $u_i = 100x_i - 124$; $v_i = 100y_i - 126$.

Практическое занятие №10.

Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона

Пусть выборка из генеральной совокупности X задана в виде статистического интервального ряда:

$[x_1, x_2)$	$[x_2, x_3)$	$\dots$	$[x_m, x_{m+1})$
n_1	n_2	$\dots$	n_m

где n_i - интервальные частоты, $\sum_{i=1}^m n_i = n$ - объем выборки,

m - число интервалов, h - длина интервала, x_i - середина интервала.

Требуется проверить гипотезу H_0 о том, что генеральная совокупность X распределена по нормальному закону, применяя критерий Пирсона. (К.Пирсон, 1857-1936 г; английский математик, биолог, философ).

Правило проверки

1. Вычисляем $\overline{x_g}$ и σ_g (см. Пример 5).
2. Находим теоретические частоты n_i' .

Их можно вычислить двумя способами.

Первый способ

$$n_i' = \frac{n \cdot h}{\sigma_g} \cdot \varphi(t_i),$$

где n - объем выборки, h - шаг, $t_i = \frac{x_i - \overline{x_g}}{\sigma_g}$;

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} - \text{функция Гаусса, значение которой в точке } t_i$$

находим по таблице (Приложение 1).

$$P_i = \frac{\varphi(t_i)}{\sigma_{\varepsilon}} \cdot h - \text{вероятность попадания значений случайной}$$

величины X в i - й интервал.

Для вычисления n_i' составляем табл. 1.

Таблица 1

i	x_i	n_i	$x_i - \overline{x_{\varepsilon}}$	t_i	$\varphi(t_i)$	P_i	$n_i' = P_i \cdot n$
1	x_1	n_1	$x_1 - \overline{x_{\varepsilon}}$	t_1	$\varphi(t_1)$	P_1	$n_1' = P_1 \cdot n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	x_m	n_m	$x_m - \overline{x_{\varepsilon}}$	t_m	$\varphi(t_m)$	P_m	$n_m' = P_m \cdot n$
Σ		n				1	n

Второй способ.

$$n_i' = P_i \cdot n$$

где n - объем выборки, $z_i = \frac{x_i - \overline{x_{\varepsilon}}}{\sigma_{\varepsilon}}$,

$$P_i = \Phi(z_{i+1}) - \Phi(z_i) - \text{вероятность попадания } X \text{ в } i - \text{й интервал,}$$

$\Phi(z)$ - значение функции Лапласа (Приложение 2).

Полагают $z_1 = (-\infty)$, $z_{m+1} = +\infty$.

Для вычисления n_i' составляем табл. 10.

Таблица 2

i	Границы интервала		n_i	Границы интервала		$\Phi(z_i)$	$\Phi(z_{i+1})$	P_i	n_i'
	x_i	x_{i+1}		z_i	z_{i+1}				
1	x_1	x_2	n_1	$-\infty$	z_2	-0,5	$\Phi(z_2)$	P_1	n_1'
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	x_m	x_{m+1}	n_m	z_m	$+\infty$	$\Phi(z_m)$	0,5	P_m	n_m'
Σ			n					1	n

3. Сравниваем эмпирические (n_i) и теоретические (n_i') частоты с помощью критерия Пирсона.

Для этого:

1) составляем расчетную табл.3, по которой находим

$$\chi^2_{набл} - \text{наблюдаемое значение критерия} \quad \chi^2_{набл} = \sum_{i=1}^m \frac{(n_i - n_i')^2}{n_i'}$$

Таблица 3.

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	n_1	n_1'	$n_1 - n_1'$	$(n_1 - n_1')^2$	$\frac{(n_1 - n_1')^2}{n_1'}$	n_1^2	$\frac{n_1^2}{n_1'}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	n_m	n_m'	$n_m - n_m'$	$(n_m - n_m')^2$	$\frac{(n_m - n_m')^2}{n_m'}$	n_m^2	$\frac{n_m^2}{n_m'}$
Σ	n				$\chi^2_{набл}$		

$$\text{Контроль: } \chi^2_{набл} = \sum_{i=1}^m \frac{n_i^2}{n_i'} - n.$$

2) Находим число степеней свободы k : $k = m - r - 1$

где ***m***- число интервалов; ***r*** - число параметров предполагаемого распределения,

Для нормального распределения **$k = m - 3$** , так как **$r = 2$** (нормальный закон распределения характеризуется двумя параметрами ***a*** и ***σ***).

4. В таблице критических точек (квантилей) распределения χ^2

(Приложение 3) по заданному уровню значимости **α** и числу степеней свободы

находим $\chi_{кр}^2(\alpha; k)$ правосторонней критической области.

Если $\chi_{набл}^2 < \chi_{кр}^2$ - нет оснований отвергнуть гипотезу **H_0**

о нормальном распределении генеральной совокупности.

Если $\chi_{набл}^2 > \chi_{кр}^2$ - гипотезу отвергаем.

Замечание.

1) Объем выборки должен быть достаточно велик (**$n \geq 50$**).

2) Малочисленные частоты (**$n_i < 5$**) следует объединить. В этом случае и соответствующие им теоретические частоты также надо сложить.

Если производилось объединение частот, то при определении числа степеней свободы по формуле **$k = m - 3$** следует в качестве ***m*** принять число интервалов, оставшихся после объединения частот.

Пример. Пусть из генеральной совокупности ***X*** задана выборка объемом 50. Требуется проверить гипотезу **H_0** о нормальном распределении генеральной совокупности по данной выборке.

○ 1. Из рассмотренных выше примеров известно:

- интервальный ряд табл. 4

Таблица 4

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

$$\sum_{i=1}^7 n_i = 50.$$

- числовые характеристики выборки $\bar{x}_e = -0,032$, $\sigma_e = 0,8195$,

$A_S^* = 0,0759$, $E_k^* = 0,3112$ (см. Пример 5).

2. Проверим гипотезу H_0 с помощью средних квадратических отклонений коэффициентов A_S^* и E_k^* .

Критерием распределения выборки по нормальному закону является равенство нулю коэффициентов A_S^* и E_k^* .

Если они отличны от нуля, то для предварительного выбора закона распределения вычисляют средние квадратические отклонения для A_S^* и E_k^* :

$$\sigma_{A_S^*} = \sqrt{\frac{6(n-1)}{(n+1) \cdot (n+3)}} \quad \sigma_{E_k^*} = \sqrt{\frac{24n \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}}$$

Если A_S^* и E_k^* отличаются по модулю от нуля не более чем на удвоенные средние квадратические отклонения, то есть $|A_S^*| \leq 2\sigma_{A_S^*}$ и $|E_k^*| \leq 2\sigma_{E_k^*}$, то можно предположить, что данная выборка распределена по нормальному закону.

3) Сравним эмпирические (n_i) и теоретические (n_i') частоты. Для этого составляем расчетную табл.6 по форме табл.5

Таблица 6

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	8	7,810	0,190	0,0361	0,0046	64	8,1946
2	11	11,675	-0,675	0,4556	0,0390	121	10,3640
3	15	14,195	0,805	0,6480	0,0457	225	15,8507
4	11	10,470	0,530	0,2809	0,0268	121	11,5568
5	5	5,850	-0,850	0,7225	0,1235	25	4,2735
Σ					0,2396		50,2396

Контроль: $\sum_{i=1}^5 \frac{n_i^2}{n_i'} - n = 50,2396 - 50 = 0,2396$

$\chi_{набл}^2 = 0,2396$. Расчеты проведены верно.

4) Зададим $\alpha = 0,05$.

Вычислим число степеней свободы $k = 5 - 3 = 2$ и найдем $\chi_{кр}^2(0,05; 2) = 6,0$ (Приложение 3). Получим $\chi_{набл}^2 < \chi_{кр}^2$.

Следовательно, нет оснований отвергать гипотезу H_0 о нормальном распределении генеральной совокупности X .

Другими словами различие между эмпирическими (n_i) и теоретическими (n_i') частотами незначительное (случайное), которое можно объяснить малым объемом выборки.

Построим нормальную кривую. Для этого составим табл.7.

Таблица 7

Середины интервалов	-1,76	-1,16	-0,56	0,04	0,64	1,24	1,84
$\frac{p_i}{h}$	0,05	0,19	0,39	0,52	0,34	0,14	0,03

Практическое занятие №11.

Определение вида дифференциального закона распределения совокупности случайных величин

Цель работы: анализ закона распределения генеральной совокупности случайных величин.

ОСНОВНЫЕ СВЕДЕНИЯ

Наиболее полной характеристикой совокупности случайных величин являются **дифференциальная или интегральная функции распределения**, устанавливающие зависимость между значением (или интервалом значений) случайной величины и вероятностью появления данного значения в заданном интервале. Для определения вида распределения в исследуемой совокупности используются критерии согласия Пирсона, Колмагорова, Смирнова и др. Ограничимся применением **критерия Пирсона** для проверки гипотезы о нормальном распределении генеральной совокупности.

Совокупность случайных величин может быть получена на разрывной машине (прочность, удлинение), весах (масса образцов), твердомере и других приборах. Воспользуемся совокупностями, приведенными в приложении В.

Формирование частотной таблицы

В тех случаях, когда выборка имеет большой объем, т.е. число значений более 30, для упрощения расчетов применяют «способ отсчета от условного нуля». Полученный ряд экспериментальных значений делят на классы (интервалы). Исходя из количества элементов совокупности ***m***, число классов ***k*** определяют по формуле (с округлением до целого):

$$k = 3,332 \cdot \lg (m + 1) \quad \text{при } 50 < m < 200;$$

$$k = 4 \cdot \sqrt[5]{0,75 \cdot (m - 1)^2} \quad \text{при } m > 200.$$

Например, для ***m* = 50** принимаем ***k* = 7**.

Находим в анализируемой выборке максимальное Y_{max} и минимальное Y_{min} значения и определяем величину интервала:

$$\Delta Y = \frac{Y_{max} - Y_{min}}{k}. \quad (1)$$

Составляем таблицу и разносим все значения анализируемой совокупности по соответствующим классам.

№ класса	1	2	3	K	k
Границы класса	$Y_{min} - (Y_{min} + \Delta Y)$	$(Y_{min} + \Delta Y) - (Y_{min} + 2 \cdot \Delta Y)$	$(Y_{min} + 2 \cdot \Delta Y) - (Y_{min} + 3 \cdot \Delta Y)$	K	$(Y_{min} + (k + 1) \cdot \Delta Y) - \Delta Y$
Значения Y_i					
Частота m_i					
Среднее Y_i^*					

Количество случайных величин в каждом классе m_i называется частотой. После сортировки значений определяем частоту m_i и математическое ожидание (среднее) Y_i^* в каждом классе.

Дальнейшие расчеты сводим в таблицу.

	Границы классов	m_i	Y_i^*	y_i	$m_i \cdot y_i$	y_i^2	$m_i \cdot y_i^2$	m_i^T	$\frac{(m_i - m_i^T)^2}{m_i}$
1				-3					
2				-2					
3				-1					
4				0					
5				1					
6				2					
7				3					
Σ	—		—	—		—		—	

Значение Y_i^* в том классе, где m_i принимает максимальное значение, называется условным нулем выборки Y_0^* .

Значения y_i находятся по формуле (и округляются до ближайшего целого):

$$y_i = \frac{Y_i^* - Y_0^*}{\Delta Y}. \quad (2)$$

Определение оценок математического ожидания, среднего квадратического отклонения и квадратической неровноты

По способу отсчета от условного нуля находим среднее значение выборки:

$$\bar{Y} = Y_0^* + \frac{\Delta Y}{m} \sum_{i=1}^k m_i \cdot y_i. \quad (3)$$

Находим среднее квадратическое отклонение и квадратическую неровноту:

$$S\{Y\} = \frac{\Delta Y}{\sqrt{m}} \sqrt{\sum_{i=1}^k m_i \cdot y_i^2 - \frac{1}{m} \left(\sum_{i=1}^k m_i \cdot y_i \right)^2}; \quad (4)$$

$$C\{Y\} = \frac{S\{Y\}}{\bar{Y}} \cdot 100\%. \quad (5)$$

Определение закона распределения исследуемой величины

Задаемся видом предполагаемой дифференциальной или интегральной функции распределения. Как правило, случайные величины, являющихся предметом анализа при исследовании технологических процессов легкой промышленности, отвечают нормальному закону распределения:

$$\varphi\{Y\} = \frac{1}{\sqrt{2\pi}} \cdot \exp \left[-\frac{(Y_i^* - \bar{Y})^2}{2 \cdot S^2\{Y\}} \right]. \quad (6)$$

Вычисляем теоретические частоты m_i^T в каждом классе:

$$m_i^T = \frac{m \cdot \Delta Y}{S\{Y\}} \cdot \phi\{Y\}. \quad (7)$$

Полученные значения заносим в таблицу и определяем наблюдаемое значение критерия Пирсона:

$$\chi_{i\hat{a}\hat{a}\hat{e}}^2 = \sum_{i=1}^k \frac{(m_i - m_i^T)^2}{m_i}. \quad (8)$$

По таблице приложения Г определяем критическое значение критерия Пирсона $\chi_{\hat{e}\hat{d}\hat{e}\hat{d}}^2$ при условии, что доверительная вероятность $P_D = 0,95$ и число степеней свободы $f = k - 2$.

Если $\chi_{i\hat{a}\hat{a}\hat{e}}^2 \leq \chi_{\hat{e}\hat{d}\hat{e}\hat{d}}^2$, то анализируемую величину можно считать распределенной по нормальному закону. Если $\chi_{i\hat{a}\hat{a}\hat{e}}^2 \geq \chi_{\hat{e}\hat{d}\hat{e}\hat{d}}^2$, необходимо использовать другие функции (логнормальную, экспоненциальную, показательно-степенную и др.) до нахождения распределения, адекватного исследуемой случайной величине.

Построение графика функции распределения

Наглядное представление о различиях между экспериментальными значениями и теоретической функцией распределения можно получить путем построения частотного полигона (рисунок 1).

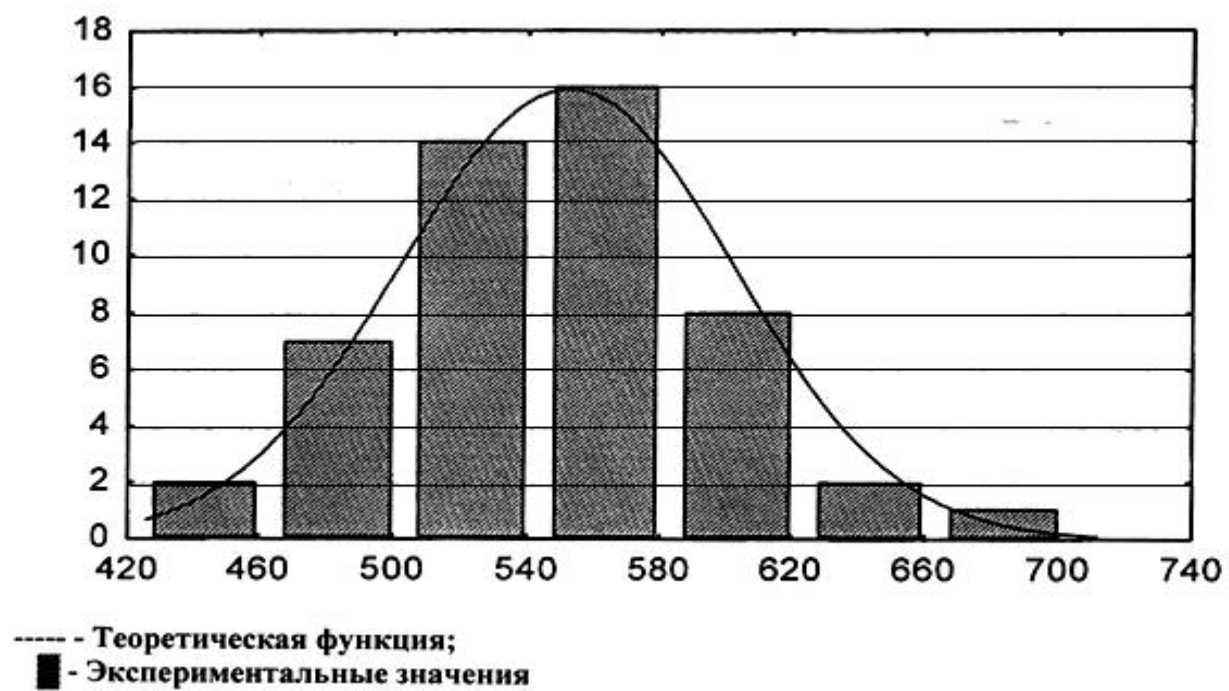


Рисунок 1 – Функции распределения (частотный полигон)

Практическое занятие №12.

Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности

Обозначим через объем выборки n , по которой найдена исправленная дисперсия S^2 .

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2$ о равенстве неизвестной генеральной дисперсии σ^2 гипотетическому (предполагаемому) значению σ_0^2 при конкурирующей гипотезе $H_0: \sigma^2 > \sigma_0^2$, надо вычислить наблюдаемое значение критерия

$$\chi_{\text{набл}}^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

и по таблице критических точек распределения χ^2 , по заданному уровню значимости α и числу степеней свободы $k = n-1$ найти критическую точку $\chi_{кр}^2(\alpha; k)$. Если $\chi_{\text{набл}}^2 < \chi_{кр}^2$ - нет оснований отвергнуть нулевую гипотезу. Если $\chi_{\text{набл}}^2 > \chi_{кр}^2$ - нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1: \sigma^2 \neq \sigma_0^2$ находят левую $\chi_{\text{лев.кр}}^2(1-\alpha/2; k)$ и правую $\chi_{\text{прав.кр}}^2(\alpha/2; k)$ критические точки. Если $\chi_{\text{лев.кр}}^2 < \chi_{\text{набл}}^2 < \chi_{\text{прав.кр}}^2$ - нет оснований отвергнуть нулевую гипотезу. Если $\chi_{\text{набл}}^2 < \chi_{\text{лев.кр}}^2$ или $\chi_{\text{набл}}^2 > \chi_{\text{прав.кр}}^2$ - нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: \sigma^2 < \sigma_0^2$ находят критическую точку $\chi_{кр}^2(1-\alpha; k)$. Если $\chi_{\text{набл}}^2 > \chi_{кр}^2(1-\alpha; k)$ - нет оснований отвергнуть нулевую гипотезу. Если $\chi_{\text{набл}}^2 < \chi_{кр}^2(1-\alpha; k)$ - нулевую гипотезу отвергают.

Замечание. Если число степеней свободы $k > 30$, то критическую точку $\chi_{кр}^2(\alpha; k)$ можно найти из равенства Уилсона-Гильферти:

$$\chi_{кр}^2(\alpha; k) = k \left[1 - 2/9k + z_\alpha \sqrt{2/(9k)} \right]^3$$

где Z_α находят, используя функцию Лапласа (см. приложение Б), из равенства $\Phi(Z_\alpha) = (1 - 2\alpha)/2$.

Пример. Из нормальной генеральной совокупности извлечена выборка объема $n = 21$ и по ней найдена исправленная выборочная дисперсия

$S^2 = 16,2$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2 = 15$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 15$.

Решение. Найдем наблюдаемое значение критерия:

$$\chi_{набл}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(21-1) \cdot 16,2}{15} = 21,6.$$

По условию, конкурирующая гипотеза имеет вид $\sigma^2 > 15$, поэтому критическая область – правосторонняя (правило 1). По таблице приложения Г, по уровню значимости 0,01 и числу степеней свободы $k = n - 1 = 21 - 1 = 20$ находим критическую точку $\chi_{кр}^2(0,01; 20) = 37,6$.

Так как $\chi_{набл}^2 < \chi_{кр}^2$ – нет оснований отвергнуть нулевую гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 = 15$. Другими словами, различие между исправленной дисперсией (16,2) и гипотетической генеральной дисперсией (15) незначимо.

2. Из нормальной генеральной совокупности извлечена выборка объема $n = 17$ и по ней найдена исправленная выборочная дисперсия $S_2 = 0,24$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2 = 0,18$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 0,18$.

3. Из нормальной генеральной совокупности извлечена выборка объема $n = 31$:

варианты x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
частоты n_i	1	3	7	10	6	3	1

Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2 = 0,18$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 0,18$.

Указание. Принять условные варианты $u_i = 10x_i - 110$; вычислить сначала $s_u^2 = \frac{\sum n_i u_i^2 - [\sum n_i u_i]^2}{n-1} / 2$, а затем $s_x^2 = \frac{s_u^2}{10^2}$.

4. Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать $\sigma_0^2 > 0,1$. Взята проба из 25 случайных отобранных изделий, причем получены следующие результаты измерений:

контролируемый размер изделий пробы, x_i	3,0	3,5	3,8	4,4	4,5
частота, n_i	2	6	9	7	1

Требуется при уровне значимости 0,05 проверить, обеспечивает ли станок требуемую точность.

5. Партия изделий принимается, если дисперсия контролируемого размера значимо не превышает 0,2. Исправленная выборочная дисперсия, найденная по выборке объема $n = 121$, оказалась равной $s_x^2 = 0,3$. Можно ли принять партию при уровне значимости 0,01?

Сравнение двух средних генеральных совокупностей, дисперсии которых известны (большие независимые выборки)

Обозначим через n и m объемы больших ($n > 30$, $m > 30$) независимых выборок, по которым найдены соответствующие выборочные средние $\bar{x}$ и $\bar{y}$. Генеральные дисперсии $D(X)$ и $D(Y)$ известны.

Правило 1. Для того чтобы при заданном уровне значимости проверить нулевую гипотезу $H_0: M(X) = M(Y)$ о равенстве математических ожиданий (генеральных средних) двух нормальных генеральных совокупностей с известными дисперсиями (в случае больших выборок) при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$, надо вычислить наблюдаемое значение критерия

$$z_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{D(X)/n + D(Y)/m}}$$

и по таблице функции Лапласа найти критическую точку $Z_{кр}$ из равенства

$$\Phi(Z_{кр}) = (1 - \alpha) / 2.$$

Если $|Z_{набл}| < Z_{кр}$ - нет оснований отвергнуть нулевую гипотезу. Если $|Z_{набл}| > Z_{кр}$ - нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1: M(X) > M(Y)$ находят критическую точку $Z_{кр}$ по таблице функции Лапласа из равенства

$$\Phi(Z_{кр}) = (1 - 2\alpha) / 2.$$

Если $Z_{набл} < Z_{кр}$ - нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} > Z_{кр}$ - нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: M(X) < M(Y)$ находят «вспомогательную точку» $Z_{кр}$ по правилу 2. Если $Z_{набл} > -Z_{кр}$ - нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} < -Z_{кр}$ - нулевую гипотезу отвергают.

Пример. По двум независимым выборкам, объемы которых $n=40$ и $m=50$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\bar{x}=130$, и $\bar{y}=140$. Генеральные дисперсии известны: $D(X)=80$, $D(Y)=100$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: M(X)=M(Y)$ при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Решение. Найдем наблюдаемое значение критерия:

$$Z_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{D(X)/n + D(Y)/m}} = \frac{130 - 140}{\sqrt{80/40 + 100/50}} = -5.$$

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область – двусторонняя.

Найдем правую критическую точку из равенства

$$\Phi(z_{кр}) = (1 - \alpha) / 2 = (1 - 0,01) / 2 = 0,495.$$

По таблице функции Лапласа находим $z_{кр} = 2,58$. Так как $|Z_{набл}| > z_{кр}$, то в соответствии с правилом 1 нулевую гипотезу отвергаем. То есть выборочные средние различаются значимо.

1. По выборке объема $n = 30$ найден средний вес $\bar{x}=130$ г. изделий, изготовленных на первом станке; по выборке объема $m = 40$ найден средний вес $\bar{y}=125$ г. изделий, изготовленных на втором станке. Генеральные дисперсии известны: $D(X)=60$ г², $D(Y)=80$ г². Требуется, при уровне значимости 0,05, проверить нулевую гипотезу $H_0: M(X)=M(Y)$ при конкурирующей гипотезе $M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

2. По выборке объема $n=50$ найден средний размер $\bar{x}=20,1$ мм диаметра валиков, изготовленных автоматом №1; по выборке объема $m=50$ найден средний размер $\bar{y}=19,8$ мм диаметра маликов, изготовленных автоматом №2. Генеральные дисперсии известны: $D(x)=1,750$ мм², $D(Y)=1,375$ мм². Требуется, при уровне значимости 0.05, проверить нулевую гипотезу $H_0: M(X)= M(Y)$ при

конкурирующей гипотезе $M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

Практическое занятие №13.

Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны и одинаковы (малые независимые выборки)

Обозначим через n и m объемы малых независимых выборок ($n < 30$, $m < 30$), по которым найдены соответствующие выборочные средние $\bar{x}$ и $\bar{y}$ и исправленные выборочные дисперсии s_x^2 и s_y^2 . Генеральные дисперсии хотя и неизвестны, но предполагаются одинаковыми.

Правило 1. Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0: M(X)=M(Y)$ о равенстве математических ожиданий (генеральных средних) двух нормальных совокупностей с неизвестными, но одинаковыми дисперсиями (в случае малых независимых выборок) при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$, надо вычислить наблюдаемое значение критерия

$$T_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} \sqrt{\frac{nm(n+m-1)}{n+m}}$$

и по таблице критических точек распределения Стьюдента, по заданному уровню значимости α , помещенному в верхней строке таблицы приложения 6, и числу степеней свободы $k=n+m-2$ найти критическую точку $t_{\text{двуст.кр.}}(\alpha; k)$. Если $|T_{\text{набл}}| < t_{\text{двуст.кр.}}(\alpha; k)$ - нет оснований отвергнуть нулевую гипотезу. Если $|T_{\text{набл}}| > t_{\text{двуст.кр.}}(\alpha; k)$ - нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $M(X) > M(Y)$ находят критическую точку $t_{\text{правост.кр.}}(\alpha; k)$ по таблице приложения Д (распределение Стьюдента) по уровню значимости α , помещенному нижней строке таблицы, и числу степеней свободы $k=n+m-1$. Если $T_{\text{набл.}} < t_{\text{правост.кр.}}$ - нет оснований отвергнуть нулевую гипотезу. Если $T_{\text{набл.}} > t_{\text{правост.кр.}}$ - нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $M(X) < M(Y)$ находят сначала критическую точку $t_{\text{правост.кр.}}$ по правилу 2 и полагают $t_{\text{левост.кр.}} = -t_{\text{правост.кр.}}$. Если $T_{\text{набл.}} > -t_{\text{правост.кр.}}$ - нет оснований отвергнуть нулевую гипотезу. Если $T_{\text{набл.}} < -t_{\text{правост.кр.}}$ - нулевую гипотезу отвергают.

Пример. По двум независимым малым выборкам, объемы которых $n=12$ и $m=18$, извлеченным из нормальных генеральных совокупностей X и Y , найдены выборочные средние: $\bar{x} = 31,2$, $\bar{y} = 29,2$ и исправленные дисперсии: $S^2_X = 0,84$ и $S^2_Y = 0,4$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: M(X) = M(Y)$ при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Решение. Исправленные дисперсии различны, поэтому проверим предварительно гипотезу о равенстве генеральных дисперсий, используя критерий Фишера-Снедекора.

Найдем отношение большей дисперсии к меньшей: $F_{набл.} = 0,84/0,40 = 2,1$. Дисперсия S^2_X значительно больше дисперсии S^2_Y , поэтому в качестве конкурирующей примем гипотезу $H_1: D(X) > D(Y)$. В этом случае критическая область – правосторонняя. По таблице приложения В (Критические точки распределения F Фишера – Снедекора), по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n-1 = 12-1 = 11$ и $k_2 = m-1 = 18-1 = 17$ находим критическую точку $F_{кр}(0,05; 11; 17) = 2,41$.

Так как $F_{набл.} < F_{кр}$ – нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий. Предположение о равенстве генеральных дисперсий выполняется, поэтому сравним средние.

Вычислим наблюдаемое значение критерия Стьюдента:

$$T_{набл.} = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}} \cdot \sqrt{\frac{n \cdot m(n+m-1)}{n+m}}.$$

Подставив числовые значения в эту формулу получим $T_{набл.} = 7,1$.

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область – двусторонняя. По уровню значимости 0,05 и числу степеней свободы $k = n+m-2 = 12+18-2 = 28$ находим по таблице приложения Д (распределение Стьюдента) критическую точку $t_{двуст.кр.}(0,5; 28) = 2,05$. Так как $T_{набл.} > t_{двуст.кр.}$ – нулевую гипотезу о равенстве средних отвергаем. Другими словами, выборочные средние различаются значимо.

1. По двум независимым малым выборкам, объемы которых $n = 10$ и $m = 8$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\bar{x} = 142,3$, $\bar{y} = 145,3$ и исправленные дисперсии: $S^2_X = 2,7$ и $S^2_Y = 3,2$. При уровне значимости 0,01 проверить нулевую гипотезу $H_0: M(X) = M(Y)$ при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Указание. Предварительно проверить равенство дисперсий, используя критерий Фишера-Снедекора.

2. Из двух партий изделий, изготовленных на двух одинаково настроенных станках, извлечены малые выборки, объемы которых $n=10$ и $m=12$. Получены следующие результаты:

контролируемый размер изделий первого станка	x_i	3,4	3,5	3,7	3,9
частота (число изделий)	n_i	2	3	4	1
контролируемый размер изделий второго станка	y_i	3,2	3,4	3,6	
частота	m_i	2	2	8	

Требуется при уровне значимости 0,02 проверить гипотезу $H_0: M(X)=M(Y)$ о равенстве средних размеров изделий при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально.

Решение. По формулам $\bar{x} = (\sum n_i x_i) / n$ и $\bar{y} = (\sum m_i y_i) / m$ найдем выборочные средние: $\bar{x}=3,6$, $\bar{y}=3,5$.

Для упрощения вычислений исправленных дисперсий перейдем к условным вариантам $u_i=10x_i-36$, $v_i=10y_i-35$.

По формулам

$$S_u^2 = \frac{\sum n_i u_i^2 - [\sum n_i u_i]^2 / n}{n-1} \quad \text{и} \quad S_v^2 = \frac{\sum m_i v_i^2 - [\sum m_i v_i]^2 / m}{m-1}$$

найдем $S_u^2=2,67$ и $S_v^2=2,55$. Следовательно,

$$S^2_X = S_u^2 / 10^2 = 2,67 / 100 = 0,0267; \quad S^2_Y = S_v^2 / 10^2 = 2,55 / 100 = 0,0255.$$

Таким образом, исправленные дисперсии различны; рассмотренный выше критерий предполагает, что генеральные дисперсии одинаковы, поэтому надо сравнивать дисперсии, используя критерий Фишера-Снедекора. Сделаем это, приняв в качестве конкурирующей гипотезы $H_1: D(X) \neq D(Y)$.

Найдем наблюдаемое значение критерия: $F_{набл} = 0,0267 / 0,0255 = 1,05$. По таблице приложения В находим $F_{кр}(0,01; 9; 11) = 4,63$. Так как $F_{набл} < F_{кр}$ –

дисперсии различаются незначимо и, следовательно, можно считать, что допущение о равенстве генеральных дисперсий выполняется.

Сравним средние, для чего вычислим наблюдаемое значение критерия Стьюдента:

$$T_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \cdot \sqrt{\frac{n \cdot m(n+m-2)}{n+m}}.$$

Подставив в эту формулу числовые значения входящих в нее величин, получим $T_{\text{набл.}}=1,45$.

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область – двусторонняя. по уровню значимости 0,02 и числу степеней свободы $k=n+m-2=10+12-2=20$ находим по таблице приложения Д критическую точку $t_{\text{двуст.кр.}}(0,02; 20)=2,53$.

Так как $T_{\text{набл.}} < t_{\text{двуст.кр.}}$ – нет оснований отвергнуть гипотезу о равенстве средних. Таким образом, средние размеры изделий существенно не различаются.

3. На уровне значимости 0,05 требуется проверить нулевую гипотезу $H_0: M(X)=M(Y)$ о равенстве средних нормальных совокупностей X и Y при конкурирующей гипотезе $H_1: M(X) > M(Y)$ по малым независимым выборкам, объемы которых $n=10$ и $m=16$. Получены следующие результаты:

x_i	12,3	12,5	12,8	13,0	13,5	y_i	12,2	12,3	13,0
n_i	1	2	4	5	1	m_i	6	8	2

Указание. Предварительно проверить нулевую гипотезу $H_0: D(X)=D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X) > D(Y)$.

Практическое занятие №14.

Определение корреляционных однофакторных моделей по данным пассивного эксперимента

ОСНОВНЫЕ СВЕДЕНИЯ

При исследовании технологических процессов и объектов часто оказывается, что выходной параметр и фактор (входной параметр) оказываются случайными величинами. В результате дискретных измерений фактора X (например, массы 500-миллиметрового отрезка пряжи) и выходного параметра Y (например, разрывной нагрузки вышеупомянутого отрезка) получают две последовательности сопряженных случайных чисел:

$$X_1, X_2, \dots, X_m;$$

$$Y_1, Y_2, \dots, Y_m.$$

Каждой паре полученных значений соответствует определенная точка в корреляционном поле точек. Для оценки степени взаимосвязи двух случайных величин X и Y рассчитывают числовую характеристику r_{YX} , называемую **коэффициентом парной корреляции**.

Для корреляционной взаимосвязи двух случайных величин характерно наличие двух зависимостей $\bar{Y}(X)$ и $\bar{X}(Y)$, которые в корреляционном поле точек изображаются в виде сопряженных прямых. Причем, чем меньше разброс точек в корреляционном поле, тем сильнее теснота связи между случайными величинами и тем меньше угол φ (рисунок 9.1) между сопряженными прямыми.

В практике исследований процессов легкой промышленности корреляционная связь между случайными величинами считается:

- | | | |
|-----------------|-----|------------------------|
| • слабой | при | $0,3 < r_{YX} < 0,4$ |
| • средней | при | $0,4 < r_{YX} < 0,7$ |
| • сильной | при | $0,7 < r_{YX} < 0,9$ |
| • очень сильной | при | $0,9 < r_{YX} $ |

Для определения коэффициентов парной корреляции и построения однофакторной корреляционной модели необходимо получить две совокупности сопряженных случайных величин (т.е. совокупность пар случайных значений). Воспользуемся совокупностями случайных величин, приведенными в приложении А.

Расчет основных статистических характеристик

Определяем средние значения ($\bar{X}$ и $\bar{Y}$) и дисперсии ($S^2\{X\}$ и $S^2\{Y\}$) для совокупностей:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i; \quad (1)$$

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i; \quad (2)$$

$$S^2\{X\} = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2; \quad (3)$$

$$S^2\{Y\} = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2. \quad (4)$$

Расчет коэффициентов парной корреляции и определение их значимости

Значение коэффициентов парной корреляции рассчитываем по формуле

$$r_{XY} = \frac{\sum_{i=1}^m (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{(m-1) \cdot S\{X\} \cdot S\{Y\}}. \quad (5)$$

По значению коэффициентов парной корреляции можно сделать определенные выводы о тесноте корреляционной взаимосвязи между случайными величинами X и Y .

Для определения значимости коэффициента корреляции определяем расчётное значение **критерия Стьюдента**:

$$t_R(r_{YX}) = \frac{r_{YX} \sqrt{m-1}}{\sqrt{1-r_{YX}^2}}. \quad (6)$$

Теоретическое значение критерия t_T определяем по таблице приложения Д при условии, что $P_D = 0,95$ и $f = m - 2$. Если $t_R\{r_{YX}\} > t_T$, то гипотеза о наличии корреляционной взаимосвязи между X и Y не отвергается.

Определение линейной модели корреляционной взаимосвязи

Рассчитываем значения коэффициентов линейных уравнений сопряженных прямых:

$$d_{0X} = \bar{Y}; \quad d_{1X} = \frac{r_{YX} \times S\{Y\}}{S\{X\}}; \quad (7)$$

$$d_{0Y} = \bar{X}; \quad d_{1Y} = \frac{r_{YX} \times S\{X\}}{S\{Y\}}. \quad (8)$$

Подставляем полученные значения в соответствующие уравнения:

$$Y = d_{0X} + d_{1X}(X - \bar{X}); \quad (9.9)$$

$$X = d_{0Y} + d_{1Y}(Y - \bar{Y}).$$

Раскрываем скобки и получаем уравнения прямых. Строим оси координат, наносим корреляционное поле точек, а затем строим сопряженные прямые с углом φ между ними.

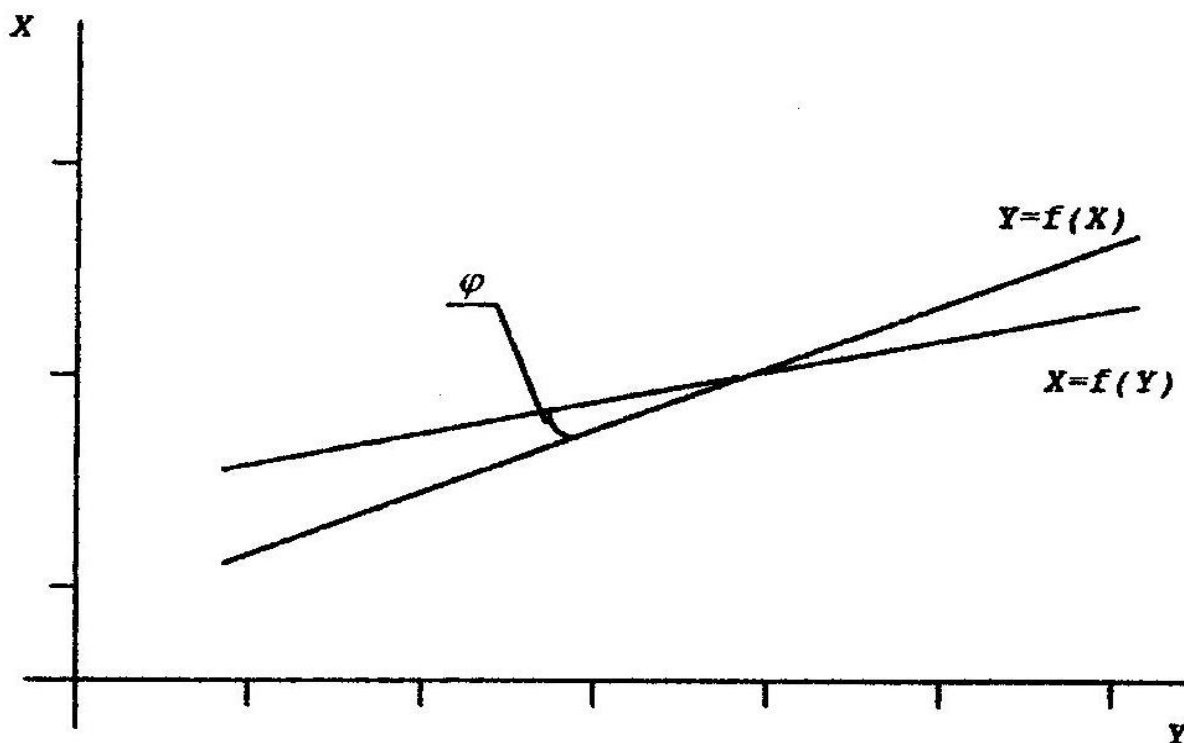


Рисунок 1 – Линейные модели корреляционной взаимосвязи

Практическое занятие №15.

Определение статических корреляционных многофакторных моделей по данным пассивного эксперимента

Цель работы: расчет парных коэффициентов корреляции, множественного коэффициента корреляции, определение его значимости и линейной модели корреляционной взаимосвязи.

ОСНОВНЫЕ СВЕДЕНИЯ

В том случае, если требуется проанализировать зависимость одной случайной величины (Y) от нескольких случайных величин $X_1, X_2, \dots, X_M$, необходимо определить **корреляционную многофакторную модель**:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_MX_M.$$

Методику рассмотрим на примере разработки двухфакторной корреляционной модели. В результате дискретных измерений факторов X_1, X_2 и выходного параметра Y получают совокупность сопряженных случайных чисел (можно воспользоваться совокупностями, приведенными в приложении А).

10.1 Расчет основных статистических характеристик

Определяем средние значения и дисперсии:

$$\bar{X}_1 = \frac{1}{m} \sum_{i=1}^m X_{1i}; \quad \bar{X}_2 = \frac{1}{m} \sum_{i=1}^m X_{2i}; \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i; \quad (1)$$

$$S^2\{X_1\} = \frac{1}{m-1} \sum_{i=1}^m (X_{1i} - \bar{X}_1)^2; \quad S^2\{X_2\} = \frac{1}{m-1} \sum_{i=1}^m (X_{2i} - \bar{X}_2)^2; \quad S^2\{Y\} = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2. \quad (2)$$

10.2 Расчет парных коэффициентов корреляции

Значения **парных коэффициентов корреляции** отражают тесноту взаимосвязи двух параметров и определяются для каждой двух переменных:

$$r_{YX_1} = \frac{\sum_{i=1}^m (X_{1i} - \bar{X}_1) \cdot (Y_i - \bar{Y})}{(m-1) \cdot S\{X_1\} \cdot S\{Y\}}; \quad (3)$$

$$r_{YX_2} = \frac{\sum_{i=1}^m (X_{2i} - \bar{X}_2) \cdot (Y_i - \bar{Y})}{(m-1) \cdot S\{X_2\} \cdot S\{Y\}}; \quad (4)$$

$$r_{X_1X_2} = \frac{\sum_{i=1}^m (X_{1i} - \bar{X}_1) \cdot (X_{2i} - \bar{X}_2)}{(m-1) \cdot S\{X_1\} \cdot S\{X_2\}}. \quad (5)$$

Расчет множественного коэффициента корреляции и определение его значимости

Теснота линейной связи между случайными величинами X_1 , X_2 и Y определяется **множественным коэффициентом корреляции**. Этот коэффициент определяет силу совместного влияния всех факторов на выходной параметр и для двухфакторной модели имеет вид:

$$R_{YX_1X_2} = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1} \cdot r_{YX_2} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2}}. \quad (6)$$

Используя критерий Стьюдента, определяем значимость найденного коэффициента

$$t_R\{R_{YX_1X_2}\} = \frac{R_{YX_1X_2}}{S\{R_{YX_1X_2}\}}. \quad (7)$$

Среднее квадратическое отклонение определяется по формуле

$$S\{R_{YX_1X_2}\} = \frac{(1 - R_{YX_1X_2}^2)}{\sqrt{m - M - 1}}, \quad (8)$$

где $M = 2$ (число факторов) и $m = 10$ (количество испытаний).

Теоретическое значение критерия Стьюдента t_T определяется по таблице приложения Д при условии, что $P_D = 0,95$ и $f = m - M - 2$. Если $t_R(R_{YX_1X_2}) > t_T$, то множественный коэффициент корреляции значим.

Определение линейной модели корреляционной взаимосвязи

Искомая модель имеет вид:

$$Y = a_0 + a_1X_1 + a_2X_2,$$

где a_0, a_1, a_2 – коэффициенты с натуральными значениями факторов, которые определяются по следующим формулам

$$a_1 = \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2} \cdot \frac{S\{Y\}}{S\{X_1\}}; \quad (9)$$

$$a_2 = \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2} \cdot \frac{S\{Y\}}{S\{X_2\}}; \quad (10)$$

$$a_0 = \bar{Y} - \sum_{i=1}^M a_i \cdot \bar{X}_i = \bar{Y} - a_1\bar{X}_1 - a_2\bar{X}_2. \quad (11)$$

Подставляем значения полученных коэффициентов в уравнение и получаем корреляционную модель в натуральных значениях.

Практическое занятие №16.

Разработка регрессионной однофакторной модели по данным активного эксперимента

Цель работы: построение однофакторной регрессионной модели методом наименьших квадратов и определение ее адекватности.

ОСНОВНЫЕ СВЕДЕНИЯ

Одной из задач обработки экспериментальных данных является определение количественной зависимости показателей качества объекта исследований от значений входных факторов и характеристик внешней среды. Другими словами, необходимо найти вид и параметры зависимости выходного параметра от значений входных факторов. В случае если Y является случайной величиной, а $X_1, X_2, \dots, X_k$ – величины неслучайные, для разработки искомой математической модели вида $Y = f(X_1, X_2, \dots, X_k)$ применяется **регрессионный анализ**.

Применение регрессионного анализа правомерно при выполнении следующих условий:

1. Значения выходного параметра Y в каждом опыте матрицы планирования эксперимента представляют собой независимые, нормально распределенные случайные величины.
2. Дисперсии выходного параметра в различных опытах матрицы однородны.
3. Значения уровней факторов не являются линейной комбинацией от уровней остальных факторов.
4. Точность определения значений выходного параметра значительно ниже точности определения величины уровня фактора.

Если одно из приведенных выше условий не будет выполняться, эффективность анализа значительно снижается и по найденной модели могут быть получены неверные технологические выводы.

В данной работе разработку линейной однофакторной модели (модели первого порядка) проведем методом наименьших квадратов.

Условия проведения активного эксперимента

Цель данной работы – получить модель вида

$$Y_R = a_0 + a_1 X. \quad (1)$$

Для получения данной модели проводят активный эксперимент в широком диапазоне изменения фактора X . Обычно применяют число уровней фактора, т.е. число опытов в матрице планирования $N \geq 5$. Для повышения точности определения выходного параметра Y каждый опыт матрицы повторяется несколько раз ($m \geq 2$).

По результатам эксперимента заполняем таблицу, используя значения, приведенные в приложении Е.

X_u	u	Y_{ui}					$\sum_{i=1}^m Y_{ui}$	$\bar{Y}_i$	$S_u^2\{Y\}$
		$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$			

Итого:

Нахождение статистических характеристик

Находят средние значения функции отклика по строкам $\bar{Y}_i$ и построчные дисперсии $S_u^2\{Y\}$ по формулам

$$\bar{Y}_i = \frac{1}{m} \sum_{i=1}^m Y_{ui} ; \quad (2)$$

$$S_u^2\{Y\} = \frac{1}{m-1} \sum_{i=1}^m (Y_{ui} - \bar{Y}_u)^2 . \quad (3)$$

где m – число повторений опытов

Проверка гипотезы об однородности дисперсий

Если число повторных опытов m одинаково для всех опытов матрицы, то для проверки однородности дисперсий применяется **критерий Кочрена**, расчетное значение которого определяется по формуле:

$$G_R = \frac{S_u^2 \max\{Y\}}{\sum_{u=1}^N S_u^2\{Y\}}. \quad (4)$$

Расчетное значение G_R сравнивают с табличным значением G_T , которое определяют по таблице приложения Ж, в зависимости от числа опытов в матрице N и числа степеней свободы дисперсии $f\{S_u^2\} = m - 1$ для заданной доверительной вероятности.

Если $G_R < G_T$, то гипотеза об однородности дисперсий принимается, если нет – следует применить методику исключения резко выделяющихся величин или найти причину возникновения большой дисперсии в u -м опыте, а затем повторить (полностью или частично) экспериментальную часть работы.

Если число повторных опытов m различно для разных опытов матрицы, то для проверки гипотезы об однородности дисперсий в опытах матрицы применяется критерий Бартлета.

Вычисление дисперсии воспроизводимости выходного параметра в опытах матрицы

Если в опытах матрицы дисперсии однородны и число повторных опытов одинаково, то средняя дисперсия определяется по формуле:

$$S_{\bar{a}\bar{n}}^2\{Y\} = \frac{1}{N} \sum_{i=1}^m S_u^2\{Y\}. \quad (5)$$

Вычисление коэффициентов искомого уравнения (модели) и их дисперсий

Для получения искомого уравнения предварительно находят коэффициенты уравнения:

$$Y_R = d_0 + d_1(X - \bar{X}), \quad (6)$$

где

$$\bar{X} = \frac{1}{N} \sum_{u=1}^N X_u; \quad (7)$$

$$d_0 = \frac{1}{N} \sum_{u=1}^N \bar{Y}_u = \bar{Y}; \quad (8)$$

$$d_1 = \frac{\sum_{u=1}^N (X_u - \bar{X}) \cdot \bar{Y}_u}{\sum_{u=1}^N (X_u - \bar{X})^2}. \quad (9)$$

Расчеты необходимых сумм сводим в таблицу

u	X_u	$X_u - \bar{X}$	$(X_u - \bar{X})^2$	$\bar{Y}_u$	$(X_u - \bar{X}) \cdot \bar{Y}_u$
1					
2					
3					
4					
5					
$\sum_{i=1}^N$					

Преобразуем уравнение (11.6) в уравнение (11.1).

Проверка адекватности полученной модели

Вначале определяется дисперсия неадекватности:

$$S_{\text{над}}^2 = \frac{m \sum_{u=1}^N (\bar{Y}_u - Y_{Ru})^2}{N - 2}, \quad (10)$$

где Y_{Ru} – возвращаемые моделью расчетные значения выходного параметра, которые определяют для каждого опыта путем подстановки в полученное уравнение соответствующих значений входных параметров.

Расчеты необходимых сумм сводим в таблицу.

u	X_u	$d_1 X_u$	Y_{Ru}	$\bar{Y}_u$	$\bar{Y}_u - Y_{Ru}$	$(\bar{Y}_u - Y_{Ru})^2$
1						
2						
3						
4						
5						
$\sum_{i=1}^N$						

Определяем расчетное значение критерия Фишера:

$$F_R = \frac{S_{i\hat{a}\hat{a}}^2\{Y\}}{S_{\hat{a}\hat{n}}^2\{\bar{Y}\}}, \text{ если } S_{i\hat{a}\hat{a}}^2\{Y\} > S_{\hat{a}\hat{n}}^2\{\bar{Y}\}; \quad (11)$$

$$F_R = \frac{S_{\hat{a}\hat{n}}^2\{\bar{Y}\}}{S_{i\hat{a}\hat{a}}^2\{Y\}}, \text{ если } S_{\hat{a}\hat{n}}^2\{\bar{Y}\} > S_{i\hat{a}\hat{a}}^2\{Y\}. \quad (12)$$

Расчетное F_R значение критерия сравнивают с табличным F_T , которое определяют по таблице приложения 3 при условии, что $P_D = 0,95$, $f\{S^2_{восп}\} = N(m - 1)$, $f\{S^2_{над}\} = N - 2$. Если $F_R < F_T$, то с вероятностью P_D гипотеза об адекватности полученной модели принимается. Если гипотеза об адекватности отвергается, необходимо переходить к описанию процесса моделью более высокого порядка на базе другого вида эксперимента.

Оценка значимости полученных коэффициентов регрессии

Значимость полученных коэффициентов оценивается с помощью критерия Стьюдента, расчетное значение которого (для каждого коэффициента) определяется по формуле:

$$T_R\{d_i\} = \frac{|d_i|}{\sqrt{S^2\{d_i\}}}, \quad (13)$$

где

$$S^2\{d_0\} = \frac{S^2\{Y\}}{m \cdot N}; \quad (14)$$

$$S^2\{d_1\} = \frac{S^2\{Y\}}{m \cdot \sum_{u=1}^N (X_u - \bar{X})^2}; \quad (15)$$

$$S^2\{Y\} = \frac{(m-1) \cdot N \cdot S_{\bar{a}\bar{a}}^2\{Y\} + (N-2) \cdot S_{i\bar{a}\bar{a}}^2\{Y\}}{m \cdot N - 2}. \quad (16)$$

Полученное расчетное значение t_R сравнивается с табличным t_T которое определяют по таблице приложения Д при условии, что $P_D = 0,95$ и число степеней свободы $f\{S^2\} = N \cdot m - 2$. Если $t_R\{d_i\} > t_T$ для обоих коэффициентов значимы, то линейная связь между X и Y значима.

Практическое занятие №17.

Полный факторный эксперимент

Полным факторным экспериментом (ПФЭ) называется эксперимент, реализующий все возможные неповторяющиеся комбинации опытов n независимых управляемых факторов, каждый из которых варьируется на k уровнях. Необходимое число опытов N при ПФЭ равно $N = k^n$. Если число уровней составляет 2, то $N = 2^n$.

Рассмотрим случай двухфакторного эксперимента, для которого уравнение регрессии неполной квадратичной модели имеет вид:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \quad (1)$$

где x_1 и x_2 - значения факторов;

b_0 - свободный член, равный отклику при $x_1 = x_2 = 0$;

b_1, b_2 - коэффициенты регрессии, показывающие степень влияния соответствующих факторов на параметр оптимизации;

b_{12} - коэффициент, указывающий на наличие эффекта взаимодействия двух факторов (парного взаимодействия).

Если в качестве базовой модели выбран полином первой степени, то каждый фактор варьируется на двух уровнях, например, $x_1 = x_1^{\min}$ и $x_1 = x_1^{\max}$, а интервал варьирования равен:

$$\Delta x_1 = (x_1^{\max} - x_1^{\min}) / 2.$$

Если выбран полином второй степени, то каждый фактор необходимо варьировать уже на трех уровнях:

$$x_1 = x_1^{\max}; \quad x_1 = x_1^0; \quad x_1 = x_1^{\min}.$$

При планировании ПФЭ преобразуют размерные (натуральные) значения факторов x_i в безразмерные (кодовые) X_i по следующей зависимости:

$$X_i = (x_i - x_i^0) / \Delta x_i, \quad i = 1, 2, 3, \dots, n. \quad (2)$$

где x_i^0 - базовые значения i -го фактора (значение нулевого или основного уровня).

Такое нормирование дает возможность легко построить ортогональную матрицу планирования и значительно облегчить дальнейшие расчеты.

В этом случае верхние и нижние уровни варьирования X_i^s и X_i^h в относительных единицах равны соответственно $+1$ и -1 независимо от

физической природы факторов, значений базовых уровней и интервалов варьирования Δx_i .

При увеличении числа факторов $n > 2$ проводят последовательное достраивание матриц ПФЭ (табл. 1).

Таблица 1

Планы двухфакторных экспериментов

План			№ опыта	X ₁	X ₂	X ₃	X ₄
2 ⁴	2 ²		1	+	+	+	+
			2	-	+	+	+
			3	+	-	+	+
			4	-	-	+	+
			5	+	+	-	+
			6	-	+	-	+
			7	+	-	-	+
			8	-	-	-	+
			9	+	+	+	-
			10	-	+	+	-
			11	+	-	+	-
			12	-	-	+	-
			13	+	+	-	-
			14	-	+	-	-
			15	+	-	-	-
			16	-	-	-	-

Отметим следующие два важных свойства рассмотренных планов:

1) Оси симметричны относительно центра эксперимента

$$\sum_{j=1}^N x_{ij} = 0; \quad i=1, 2, 3, \dots, n; \quad j=1, 2, 3, \dots, N.$$

Сумма значений любого столбца равна нулю.

2) Эти планы нормированы

$$\sum_{j=1}^N x_{ij}^2 = N; \quad i = 0, 1, 2, 3, \dots, n; \quad j = 0, 1, 2, 3, \dots, N.$$

Сумма квадратов элементов любого столбца равна числу опытов.

Основным преимуществом планов, обладающих указанным свойством, является возможность проведения родственной (независимой) оценки коэффициентов регрессии.

После составления плана эксперимента приступают к его реализации. На исследуемый процесс влияют не только выбранные факторы x_i , но и еще ряд факторов, которые могут быть вообще неизвестны исследователю. Для того, чтобы внести элемент случайности влияния этих факторов на результат эксперимента (а это необходимо для обоснованного использования аппарата математической статистики), приходится проводить m параллельных опытов и устанавливать случайный порядок проведения опытов во времени. Эта процедура называется рандомизацией (перемешиванием) и выпроступают к его реализации. На исследуемый процесс влияют не только выбранные факторы x_i , но и еще ряд факторов, которые могут быть вообще неизвестны исследователю. Для того, чтобы внести элемент случайности влияния этих факторов на результат экспериментахрена:

$$G_0 = \frac{\sigma_{j_{\max}}^2}{\sum_{j=1}^N \sigma_j^2} \leq G_{(0,05;\phi_1;\phi_2)} \quad (3)$$

где σ_j^2 - выборочная дисперсия выходной величины y по j строке матрицы планирования, полученная из m параллельных опытов:

$$\sigma_j^2 = \frac{1}{m-1} \sum_{i=1}^m (y_{ij} - \bar{y}_j)^2 \quad (4)$$

где y_{ij} - значение выходной величины по j строке матрицы планирования (j изменяется от 1 до N) из i -го параллельного опыта (i изменяется от 1 до m);

$\bar{y}_j$ - среднее значение выходной переменной, полученное из параллельных опытов по j строке матрицы планирования;

$\sigma_{j_{\max}}^2$ - наибольшая из дисперсий в строках плана;

$G_{(0,05;\phi_1;\phi_2)}$ - табличное значение критерия Кохрена при 5% уровне значимости;

$$\Phi_1 = m - 1; \Phi_2 = N.$$

Если вычисленное значение G_0 окажется меньше табличного, то гипотеза об однородности дисперсий принимается. Делается вывод о воспроизводимости экспериментов. В противном случае необходимо принять меры по повышению воспроизводимости опытов.

В случае воспроизводимости процесса рассчитывают коэффициенты уравнения регрессии. Независимые оценки b_0, b_i, b_{ik} соответствующих коэффициентов $\beta_0, \beta_i, \beta_{ik}$, ($b_0 \rightarrow \beta_0, b_i \rightarrow \beta_i, b_{ik} \rightarrow \beta_{ik}$) находят по следующим формулам:

$$b_0 = \frac{1}{N} \sum_{j=1}^N \bar{y}_j \quad \text{или} \quad b_0 = \frac{\bar{y}_1 + \bar{y}_2 + \dots + \bar{y}_N}{N} \quad (5)$$

$$b_i = \frac{1}{N} \sum_{j=1}^N x_{ij} \bar{y}_j \quad \text{или} \quad b_i = \frac{x_{i1}\bar{y}_1 + x_{i2}\bar{y}_2 + \dots + x_{iN}\bar{y}_N}{N} \quad (6)$$

$$b_{ik} = \frac{1}{N} \sum_{j=1}^N x_{ij} x_{kj} \bar{y}_j \quad \text{или} \quad b_{ik} = \frac{x_{i1}x_{k1}\bar{y}_1 + x_{i2}x_{k2}\bar{y}_2 + \dots + x_{iN}x_{kN}\bar{y}_N}{N} \quad (7)$$

После определения оценок коэффициентов регрессии b_{ik} необходимо проверить гипотезу об их значимости. Проверяют гипотезу с помощью критерия Стьюдента. Если выполняется условие:

$$|b_i| \geq t_{\Phi; \alpha} \sqrt{\sigma_{\epsilon}^2}, \quad \Phi = N(m - 1) \quad (8)$$

$$\text{где } \sigma_{\epsilon}^2 = \frac{\sigma_y^2}{N \cdot m} \quad - \quad \text{дисперсия коэффициентов,} \quad (9)$$

$$\text{а } \sigma_y^2 = \frac{\sum \sigma_j^2}{N} \quad - \quad \text{дисперсия воспроизводимости,} \quad (10)$$

то данный коэффициент является статистически значимым. Все коэффициенты, для которых это условие не выполняется, являются незначимыми и они из уравнения регрессии исключаются вместе с соответствующим фактором, при этом оставшиеся коэффициенты не пересчитываются.

Проверяют гипотезу об адекватности полученной математической модели результатом эксперимента с использованием критерия Фишера. Модель адекватна, если выполняется неравенство:

$$F_0 = \frac{\sigma_y^2}{\sigma_{ad}^2} \geq F_{(\Phi_1, \Phi_2, \alpha)}, \quad \text{если} \quad \sigma_{ad}^2 < \sigma_y^2 \quad (11)$$

где σ_{ad}^2 - дисперсия адекватности (остаточная дисперсия)

$$\sigma_{ad}^2 = \frac{1}{N-d} \sum_{j=1}^N (\bar{y}_j - \hat{y}_j)^2 \quad (12)$$

где ***d*** – число значимых коэффициентов уравнения регрессии, включая и ***b*₀**.

$\hat{y}_j$ - рассчитанное по модели значение отклика (параметра оптимизации) в ***j*** опыте.

$$\Phi_1 = N(m-1); \Phi_2 = N - d.$$

Если же $\sigma_{ad}^2 > \sigma_y^2$, то рассчитывают критерий Фишера, исходя из отношения:

$$F_0 = \frac{\sigma_{ad}^2}{\sigma_y^2} \quad (13)$$

Гипотеза об адекватности может быть принята, если

$$F_0 \leq F_{(\Phi_1; \Phi_2; \alpha)}, \quad \text{при этом} \quad \Phi_1 = N - d, \Phi_2 = N(m-1).$$

В том случае, когда гипотеза об адекватности модели отвергается, необходимо переходить к более сложной форме математического

описания, либо, если это возможно, провести эксперимент с меньшим интервалом варьирования.

После проверки на адекватность, выполняют переход от кодированных значений факторов к натуральным по уже известной зависимости

$$x_i = \frac{x_i - x_i^0}{\Delta x_i}, \quad i = 1, 2, \dots, n. \quad (14)$$

Для двухфакторного эксперимента:

$$y = b_0 + b_1 \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) + b_2 \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) + b_{12} \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) \quad (15)$$

Пример планирования полного факторного эксперимента

Сопротивление деформации σ_s алюминиевого сплава 1915 в наибольшей степени зависит от температуры θ и скорости деформации ξ . Необходимо получить математическую модель вида $\sigma_s = \sigma_s(\theta, \xi)$ для последующей оптимизации параметров процесса пластической обработки.

Экспериментальное исследование условий горячего прессования алюминиевого сплава 1915 позволило установить технологически разумные пределы, в которых могут изменяться факторы: температура от 370⁰С до 430⁰С; скорость деформации от 8 до 12 с⁻¹. Для решения задачи моделирования принято решение провести ПФЭ 2².

Опыты проводятся путем растяжения образцов на пластометре. Условия эксперимента приведены в табл. 4.2, а матрица плана и результаты экспериментов в табл. 4.3. Проводилось по три параллельных опыта ($m=3$) с рандомизацией.

Таблица 2

Условия эксперимента

Уровень фактора	$\theta, ^\circ\text{C}$	ξ, c^{-1}
Основной $x_i = 0$	400	10
Интервал варьирования Δx_i	30	2
Нижний $x_i = -1$	370	8
Верхний $x_i = +1$	430	12
Кодовые обозначения	X_1	X_2

Таблица 3

План эксперимента

№ опыта j	X_1	X_2	X_1 X_2	Параллельные опыты $\sigma_s, \text{Мпа}$			$\bar{y}_j$	σ_j^2	$\hat{y}_j$	$(\bar{y}_j - \hat{y}_j)^2$
				Y_1	Y_2	Y_3				
1	+	+	+	106	108	105	106.3	2.33	106.3	0
2	-	+	-	156	154	158	156	4.01	156.0	0
3	+	-	-	99	100	96	98.3	4.34	98.0	0.09
4	-	-	+	139	141	141	140.3	1.34	140.3	0
Σ								12.02		0.09
5	0	0	0	122.5	124.5	123	123.3	1.085	125.22	3.69
Σ								13.10		3.78

1. Определяем $\bar{y}_j$ и выборочные дисперсии σ_j^2 для каждого опыта по формуле (4.7). Результаты расчета сводим в таблицу.
2. Проводим проверку воспроизводимости опытов по критерию Кохрена (3) $\sigma_{j\max}^2 = 4,34$.

$$G_0 = 4,34/12,02 = 0,36.$$

Табличное значение критерия Кохрена для уровня значимости

$\alpha = 0,05$ и степеней свободы $\Phi_1 = m - 1 = 3 - 1 = 2$, $\Phi_2 = N = 4$

$G_{(0,05, 2, 4)} = 0,7679$ (табл.5 Приложения).

Сравниваем G_0 и $G_{(\alpha, \Phi_1, \Phi_2)}$.

$G_0 < G_{(0,05,2,4)}$, следовательно, дисперсии однородны, опыты воспроизводимы.

3. По формуле (4.13) находим дисперсию воспроизводимости:

$$\sigma_y^2 = (2,33 + 4,01 + 4,34 + 1,34) / 4 = 3,01$$

Степень свободы дисперсии воспроизводимости равна

$$\Phi = N(m - 1) = 4(3 - 1) = 8.$$

4. Определяем коэффициенты уравнения регрессии, которое в общем случае имеет вид:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2$$

Для нахождения коэффициентов b_0 , b_1 , b_2 и b_{12} используем соответственно зависимости (4.8), (4.9) и (4.10):

$$b_0 = (106,3 + 156 + 98,3 + 140,3) / 4 = 125,22$$

$$b_1 = (106,3 - 156 + 98,3 - 140,3) / 4 = -22,92$$

$$b_2 = (106,3 + 156 - 98,3 - 140,3) / 4 = 5,92$$

$$b_{12} = (106,3 - 156 - 98,3 + 140,3) / 4 = -1,92.$$

Уравнение регрессии примет вид:

$$\hat{y} = 125,22 - 22,92 X_1 + 5,92 X_2 - 1,92 X_1 X_2$$

5. По формуле (4.12) находим дисперсию коэффициентов

$$\sigma_b^2 = \frac{3,01}{4 \cdot 3} = 0,25 \quad \text{и, исходя из зависимости (2) оцениваем значимость}$$

коэффициентов уравнения регрессии. Табличное значение критерия Стьюдента для уровня значимости $\alpha = 0,05$ и степени свободы

$$\Phi = N(m - 1) = 4(3 - 1) = 8 \text{ равно } t_{0,05, 8} = 2,31 \text{ (табл.2 Приложения).}$$

$$\text{Произведение } t_{\alpha, \Phi} \cdot \sqrt{\sigma_b^2} = 2,31 \cdot \sqrt{0,25} = 1,15$$

Все коэффициенты по абсолютной величине превышают это значение. Следовательно мы должны признать их значимыми.

6. Проверяем адекватность полученного уравнения экспериментальным результатам. В нашем случае число значимых коэффициентов уравнения регрессии равно числу опытов, т.е. степень свободы дисперсии адекватности (6) равна нулю. Поэтому мы вынуждены поставить дополнительный опыт на нулевом уровне. Результаты опыта заносим в план эксперимента. При этом число опытов N становится равным пяти, а дисперсия воспроизводимости (4)

$$\sigma_y^2 = (2,33+4,01+4,34+1,34+1,085)/5 = 2,6.$$

По уравнению регрессии рассчитываем значения $\hat{y}_j$ и определяем сумму квадратов отклонений $\sum_{j=1}^N (\hat{y}_j - \bar{y}_j)^2$. Результаты расчета заносим в таблицу плана эксперимента. Определяем дисперсию адекватности (6) для $N=5$ и $d=4$.

$$\sigma_{ad}^2 = 3,78/(5-4) = 3,78.$$

Тогда **F**-отношение (расчетное значение критерия Фишера) (7):

$$F_0 = 3,78/2,60 = 1,45.$$

Табличное значение критерия Фишера для $\alpha=0,05$,

$$\Phi_1=N-d=1, \Phi_2 = N(m-1)=10, F_{(0,05;1,10)} = 5,0 \text{ (табл.3 Приложения).}$$

Получаем $F_0 < F_{(0,05;1,10)}$ и, следовательно, уравнение регрессии адекватно экспериментальным результатам.

7. Выполняем переход от кодированных значений факторов к натуральным по уравнениям (8) и (9)

$$\begin{aligned} \sigma_s &= 125,22 - 22,92 \left(\frac{\theta - 400}{30} \right) + 5,92 \left(\frac{\xi - 10}{2} \right) - 1,92 \left(\frac{\theta - 400}{30} \right) \left(\frac{\xi - 10}{2} \right) = \\ &= 273,2 - 0,444\theta + 15,78\xi - 0,032\theta \cdot \xi \end{aligned}$$

Задания

Построить математическую модель по матрице планирования эксперимента.

№ опыта	X ₁	X ₁	X ₁ X ₁	Параллельные опыты λ_p			$\bar{y}_j$	σ_j^2	$(\hat{y}_j - \bar{y}_j)^2$
				y ₁	y ₂	y ₃			
1	+	+		4,59	4,48	4,72			
2	-	+		3,54	3,50	3,67			
3	+	-		5,10	5,03	5,20			
4	-	-		3,90	4,00	3,97			

Практическое занятие №18.

Дробный факторный эксперимент

ПФЭ требует большого числа опытов, причем часть из них несет мало информации. Дробный факторный эксперимент (ДФЭ) позволяет сократить число опытов и в то же время получить основной объем необходимой информации.

Эксперимент, составляющий по объему только часть ПФЭ, называется дробным факторным экспериментом или дробной репликой. Существует $\frac{1}{2}$ реплики, $\frac{1}{4}$ реплики, $\frac{1}{8}$ реплики и т.д. Условные обозначения дробных реплик и количество опытов даны в табл.1.

Таблица 1

Дробный факторный эксперимент

Кол-во факторов	Дробная реплика	Условное обозначение ДФЭ	Кол-во опытов	
			Для дробной реплики	Для ПФЭ
3	$\frac{1}{2}$ реплики от 2^3	2^{3-1}	4	8
4	$\frac{1}{2}$ реплики от 2^4	2^{4-1}	8	16
5	$\frac{1}{2}$ реплики от 2^5	2^{5-1}	16	32
5	$\frac{1}{4}$ реплики от 2^5	2^{5-2}	8	32
6	$\frac{1}{2}$ реплики от 2^6	2^{6-1}	32	64
6	$\frac{1}{4}$ реплики от 2^6	2^{6-2}	16	64
6	$\frac{1}{8}$ реплики от 2^6	2^{6-3}	8	64
7	$\frac{1}{2}$ реплики от 2^7	2^{7-1}	64	128
7	$\frac{1}{4}$ реплики от 2^7	2^{7-2}	32	128
7	$\frac{1}{8}$ реплики от 2^7	2^{7-3}	16	128
7	$\frac{1}{16}$ реплики от 2^7	2^{7-4}	8	128

При образовании реплик необходимо помнить, что количество опытов

должно быть хотя бы на единицу больше, чем количество факторов в ДФЭ.

Эффективность применения дробных реплик зависит от удачного выбора системы смешивания линейных эффектов с эффектами взаимодействий. Реплики, у которых линейные эффекты смешаны с взаимодействиями наивысшего порядка, являются наиболее эффективными, они обладают наибольшей разрешающей способностью и называются главными.

В реальных условиях разработчик может не иметь твердой уверенности в отсутствии того или иного взаимодействия факторов. В этом случае надо знать, когда и какие эффекты определяются совместно, определить разрешающую способность дробных реплик. Для этого пользуются понятиями «определяющие контрасты» и «генерирующее соотношение».

Рассмотрим эти понятия на примере полуреплики 2^{3-1} . Соотношение показывающее, с каким из эффектов смешан данный эффект, или какое взаимодействие факторов заменено данным фактором, называется генерирующим соотношением:

$$X_3 = X_1X_2 \quad (1)$$

Покажем эту полуреплику в качестве таблицы:

Таблица 2

План ДФЭ 2^{3-1}

№ опы та	$X_3=X_1X_2$			
	X_1	X_2	X_3	$X_1X_2X_3$
1	+	+	+	+
2	-	+	-	+
3	+	-	-	+
4	-	-	+	+

Для произведения трех столбцов матрицы в каждом опыте имеем:

$$+1 = X_1 X_2 X_3 \quad (2)$$

эти же уравнения можно получить из генерирующего соотношения, умножением левой и правой его части на X_3 .

Произведение столбцов матрицы равное +1 называется определяющим контрастом. Контраст помогает найти смешанные эффекты. Для того, чтобы определить какой эффект смешан с данным, нужно умножить обе части определяющего контраста на фактор, соответствующий эффекту. Так для полуреплики 2^{3-1} с определяющим контрастом

$1 = X_1 X_2 X_3$ имеем:

$$X_1 = X_1^2 X_2 X_3 = X_2 X_3 \quad (3)$$

$$X_2 = X_1 X_2^2 X_3 = X_1 X_3 \quad (4)$$

$$X_3 = X_1 X_2 X_3^2 = X_1 X_2 \quad (5)$$

Это значит, что коэффициенты линейного уравнения будут оценками линейных эффектов:

$$b_1 = \beta_1 + \beta_{23}; \quad b_2 = \beta_2 + \beta_{13}; \quad b_3 = \beta_3 + \beta_{12} \quad (6)$$

Эффект смешивания в принципе снижает точность оценок. Однако, поскольку мы считаем модель линейной и взаимодействия пренебрежимо малыми, то точность оценок будет достаточной.

Пример планирования дробного факторного эксперимента

Необходимо спланировать эксперимент с целью выбора оптимальных параметров устройства для получения максимального значения выходной характеристики y . Исходные значения факторов и интервалы варьирования заданы в таблице.

Условия эксперимента

Фактор	Уровни факторов			Интервал варьирования
	-1	0	+1	
X_1	200	220	240	20
X_2	3	6	9	3
X_3	40	100	160	60
X_4	1	2	3	1

Для матрицы планирования выбираем полуреплику от ПФЭ 2^4 , заданную генерирующим соотношением $X_4 = X_1X_2X_3$. Определяющим контрастом является $I = X_1X_2X_3X_4$. Умножая определяющий контраст последовательно на X_1 , X_2 , X_3 и X_4 , найдем совместно оценки линейных эффектов и взаимодействий:

$$b_1 = \beta_1 + \beta_{234} \qquad b_{12} = \beta_{12} + \beta_{34}$$

$$b_2 = \beta_2 + \beta_{134} \qquad b_{23} = \beta_{23} + \beta_{14}$$

$$b_3 = \beta_3 + \beta_{124} \qquad b_{13} = \beta_{13} + \beta_{24}$$

$$b_4 = \beta_4 + \beta_{123}$$

Составляем таблицу матрицы планирования в которую вносим результаты эксперимента.

План эксперимента

№ опыта	X ₀	X ₁	X ₂	X ₃	X ₄ = X ₁ X ₂ X ₃	X ₁ X ₂ =X ₃ X ₄	X ₁ X ₃ =X ₂ X ₄	X ₂ X ₃ =X ₁ X ₄	$\bar{y}_j$
1	+	+	+	-	-	+	-	-	10
2	+	-	-	-	-	+	+	+	9
3	+	+	-	-	+	-	-	+	15
4	+	-	+	-	+	-	+	-	25
5	+	+	+	+	+	+	+	+	26
6	+	-	-	+	+	+	-	-	14
7	+	+	-	+	-	-	+	-	5
8	+	-	+	+	-	-	-	+	20
b _i	15	-1,5	4,75	0,75	4,5	-0,75	0,75	2,0	

Коэффициенты **b_i** рассчитываются по формулам приведенным в ПЗ №17.

$$b_0 = \frac{\sum \bar{y}_j}{N} = (10 + 9 + 15 + 25 + 26 + 14 + 5 + 20) / 8 = 15$$

$$b_1 = \frac{\sum x_{ij} \bar{y}_j}{N} = (10 - 9 + 15 - 25 + 26 - 14 + 5 - 20) / 8 = -1,5$$

и т.д.

Таким образом, аналитическое выражение для y принимает вид:

$$\hat{y} = 15 - 1,5X_1 + 4,75X_2 + 0,75X_3 + 4,5X_4 - 0,75X_1X_2 + 0,75X_1X_3 + 2X_2X_3$$

Дальнейший анализ подобных выражений, их изменение (переход от кодированных факторов к натуральным), проведение новых экспериментов ведется в соответствии с порядком, изложенным для ПФЭ.

Задания

Построить математическую модель по матрице планирования эксперимента.

Практическое занятие № 19

Введение в вероятностные методы анализа данных. Числовые характеристики выборки

Цель работы

Целью занятия является получение студентом основных сведений о типах данных и числовых характеристиках выборок.

Краткие теоретические сведения

Анализ данных (data analysis) – это процесс исследования, очистки, преобразования и моделирования данных с целью обнаружения нужной информации, получения информированных выводов и поддержки принятия решений.

Исходные данные, анализ которых проводится, представляются в виде наборов $x_1, x_2, \dots, x_n$; $y_1, y_2, \dots, y_m$, $z_1, z_2, \dots, z_l$ и т.д., которые отражают значения переменных или признаков x, y, z объектов изучения. Например, если объектом изучения являются пациенты больницы, в качестве x и y могут выступать рост и вес пациента. Тогда каждый пациент i характеризуется парой значений (x_i, y_i) .

Пусть рассматривается один признак x , характеризующий объект исследования. При вероятностном анализе данных целью исследователя является установление свойств генеральной совокупности объектов путём анализа ограниченной выборки измерений. Генеральной совокупностью называется множество (конечное или бесконечное) всех сходных объектов, представляющих интерес для исследователя. Выборкой называется подмножество генеральной совокупности $x_1, x_2, \dots, x_n$, выбранное для представления генеральной совокупности в статистическом анализе.

Переменная x может иметь один из следующих распространённых типов.

Категориальная переменная содержит данные с ограниченным числом уникальных значений или категорий, в текстовом или числовом формате. Такие данные также называют качественными.

- *Номинальный тип* категориальной переменной представляет собой категории без естественного упорядочения, например, название компании сотрудника, регион, почтовый индекс.
- *Порядковый тип* категориальной переменной представляет категории с естественным упорядочением, например, состояние пациента (от крайне плохого до отличного).

Непрерывная (интервальная) переменная – это переменная, принимающая числовые значения на некотором промежутке. Для таких переменных есть и порядок значений. Примерами интервальных переменных являются рост и вес, температура воздуха.

Для выборок с непрерывными значениями возможно посчитать числовые характеристики. Самыми популярными в практических приложениях являются следующие характеристики:

- Выборочное среднее $\bar{x}$
- Выборочная дисперсия s_x^2
- Выборочное стандартное отклонение s_x
- Медиана h_x

Команды на языке Python для вычисления соответствующих характеристик выборки с использованием библиотеки numpy:

- `numpy.mean` / `numpy.nanmean`
- `numpy.var` / `numpy.nanvar`
- `numpy.std` / `numpy.nanstd`
- `numpy.median` / `numpy.nanmedian`

Варианты команд с префиксом `nan` игнорируют значения NaN (пропуски в данных).

Пример листинга программы с вычислением и выводом выборочного среднего – см. листинг 1.

```
import numpy as np
a = np.array([1,2,3,4,5])
print(np.mean(a))
```

Листинг 1. Вычисление и вывод выборочного среднего.

Ход работы

1. Найти в открытых источниках или придумать выборку непрерывных величин
2. Реализовать скрипт в Python для загрузки данных (соответствующие команды найти самостоятельно), посчитать и вывести основные числовые характеристики выборки (среднее, дисперсию, стандартное отклонение, медиану).

Вопросы:

1. Задачи анализа данных. Генеральная совокупность и выборка.
2. Типы переменных.
3. Основные числовые характеристики выборки.

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
2. Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2009.
3. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Т. 1: Основные понятия теории вероятностей и математической статистики. М.: МЦНМО, 2007

Практическое занятие № 20

Первичный и разведочный анализ данных

Цель работы

Целью занятия является получение студентом навыков первичного и разведочного анализа данных, включая обнаружение выбросов, графический анализ, выдвижение гипотез о наличии связи между переменными.

Краткие теоретические сведения

Первичный анализ данных может включать в себя следующие этапы:

- Просмотр таблицы данных – в случае, если объём данных невелик.
- Графический анализ данных – построение линейчатых графиков, гистограмм, диаграмм рассеяния и диаграмм размаха.

Первичный анализ даёт возможность оценить характер данных и зафиксировать наиболее очевидные ошибки в данных, вызванные неправильным вводом или импортом данных (например, текстовые значения вместо численных), а также пропуски в данных.

Графический анализ позволяет оценить вид распределений величин в выборках (гистограммы), диапазон значений переменных (диаграммы размаха), обнаружить выбросы в данных (диаграммы размаха, линейные графики), оценить наличие и характер связи между парами переменных (диаграммы рассеяния).

Выбросом (грубой ошибкой, англ. outlier) называется результат измерения, выделяющийся из общей выборки. Наличие выбросов в данных может привести к некорректным результатам их статистического анализа, поэтому заблаговременное нахождение и удаление выбросов является важной задачей.

На листинге 2 приведён код для построения линейного графика. Результат выполнения кода приведён на рисунке 1.

```
import matplotlib.pyplot as plt

x = [1,2,3,4,5,6,7,8,9,10]
y = [10,20,34,38,358,59,71,82,89,101]

plt.plot(range(len(y)),y, '- ')
plt.xlabel('x')
plt.ylabel('y')
```

```
plt.show()
plt.close()
```

Листинг 2. Код для построения линейного графика

На рисунке видно наличие выброса – элемента выборки у номер 5. Требуется проверить данный элемент на корректность значения (при наличии информации об исходных данных и методе их сбора) и, возможно, удалить его из выборки.

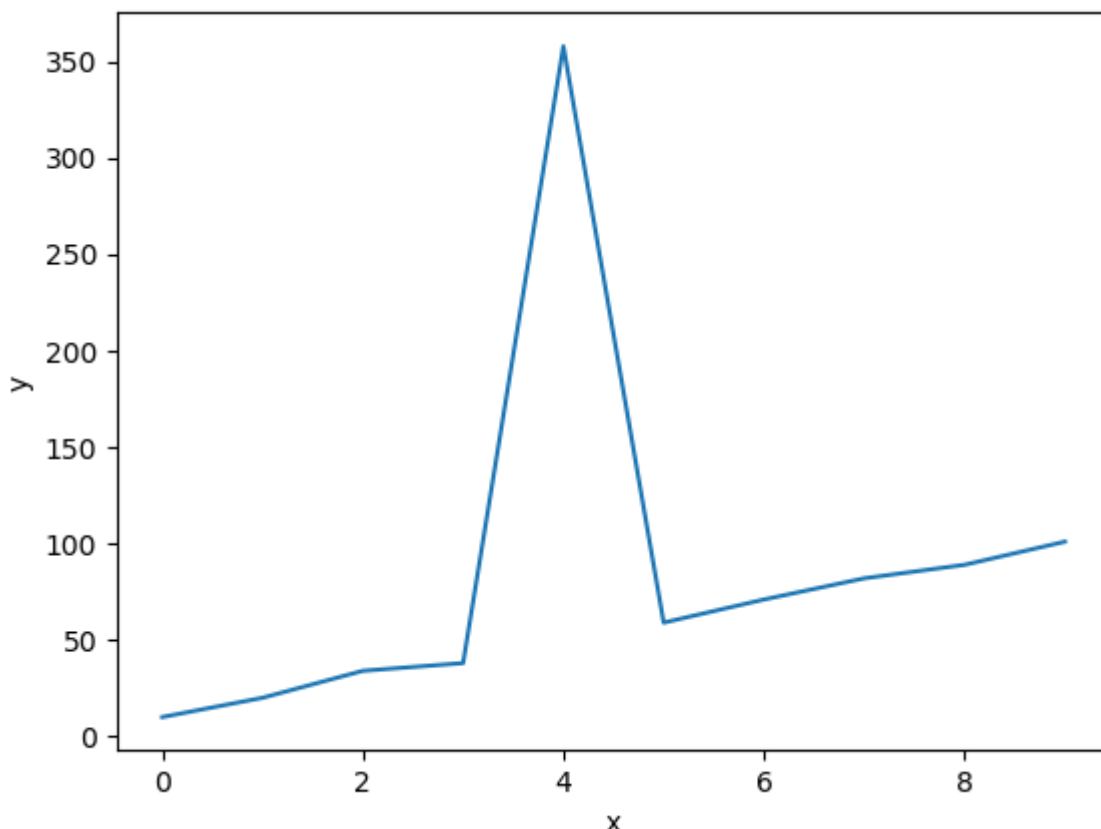


Рисунок 1. Выявление выброса с помощью графического анализа

Простейшим способом обнаружения выбросов в данных является использование межквартильного расстояния (interquartile range, IQR). Для произвольной случайной величины оценка интерквартильное расстояние рассчитывается по формуле $IQR = Q3 - Q1$, где $Q1$ и $Q3$ – первый и третий quartили соответственно. Ошибками считаются точки, находящиеся вне диапазона $[Q1 - k \cdot IQR, Q3 + k \cdot IQR]$, k выбирается равным фиксированному значению – наиболее часто используются значения 1.5 и 3.

Для нахождения IQR на Python используются команды поиска процентилей. Например, для нахождения 75-процентиля (0.75-квантиля) применяется команда `numpy.percentile(data, 75)` библиотеки `numpy`. Аналогично находится 25-процентиль. Пример кода для поиска интерквартильного

расстояния и границ отрезка, внутри которого должны лежать точки, не являющиеся ошибками, приведён на листинге 3.

```
import numpy

data = [10,20,34,38,357,59,71,82,89,101]
IQR = numpy.percentile(data, 0.75) - numpy.percentile(data, 0.25)
upper_limit = numpy.percentile(data, 0.75) + (IQR * 1.5)
upper_limit_extreme = numpy.percentile(data, 0.75) + (IQR * 3)
print(upper_limit, upper_limit_extreme)

lower_limit = numpy.percentile(data, 0.25) - (IQR * 1.5)
lower_limit_extreme = numpy.percentile(data, 0.25) - (IQR * 3)
print(lower_limit, lower_limit_extreme)
```

Листинг 3. Пример кода для поиска границ отрезка, внутри которого должны лежать корректные данные

Для попарного графического анализа двух переменных и поиска потенциальной связи между ними удобно использовать диаграмму рассеяния (облако точек). По форме полученного облака можно делать предварительные выводы о наличии или отсутствии связи, а также о её характере. На листинге 4 приведён код на Python для построения облака точек двух переменных. Следует обратить внимание на то, что размеры выборок данных для обеих переменных должны совпадать. На рисунках 2-4 приведены примеры графиков для различных вариантов данных. Судя по виду графика, на рисунке 2 – демонстрируется наличие явной зависимости ординат точек от их абсцисс – можно предположить наличие линейной связи между отображаемыми переменными а и b. На рисунке 3 также прослеживается связь, но она не является линейной. На рисунке 4 наличие связи между переменными маловероятно, поскольку точки с разными ординатами (значение элемента выборки y) могут иметь одинаковые абсциссы (значение элемента выборки x).

```
import matplotlib.pyplot as plt

a = [1,2,3,4,5,6,7,8,9,10]
b = [10,20,34,38,58,59,71,82,89,101]

plt.plot(a,b,'o')
plt.xlabel('a')
plt.ylabel('b')
plt.show()
plt.close()
```

Листинг 4. Код для построения облака точек двух переменных

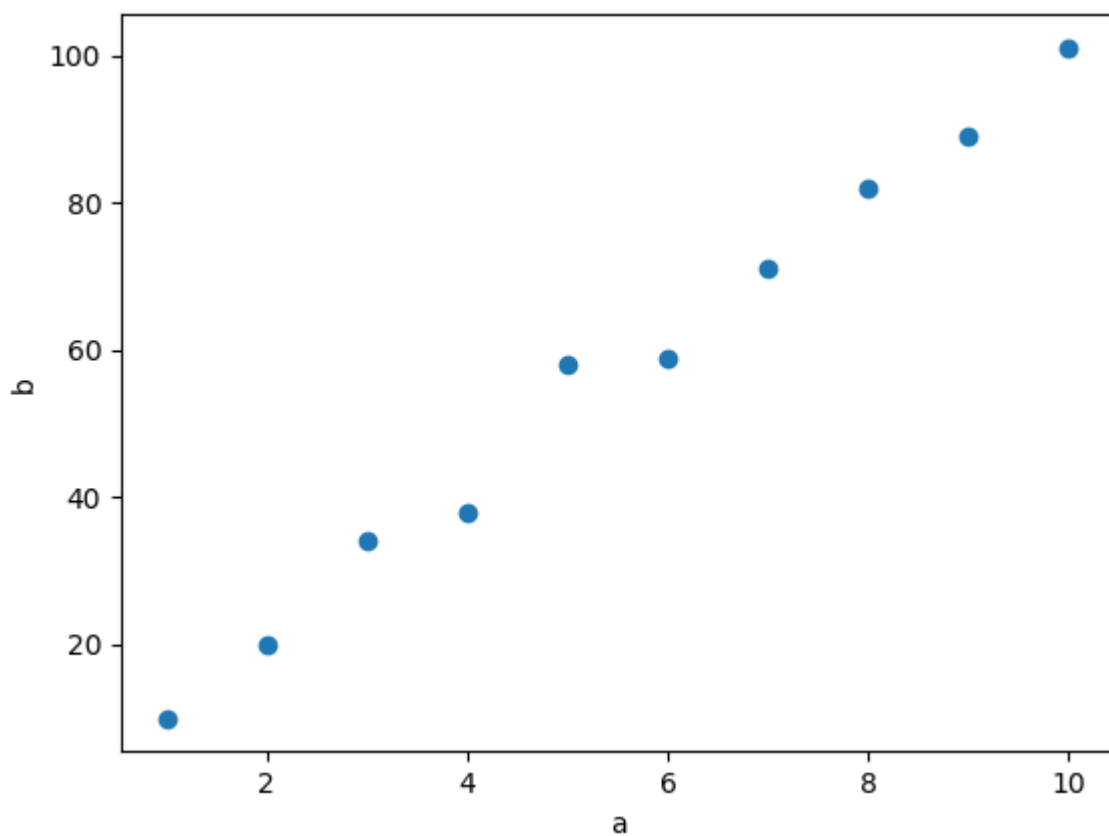


Рисунок 2. Облако точек с предполагаемой линейной связью между переменными a и b .

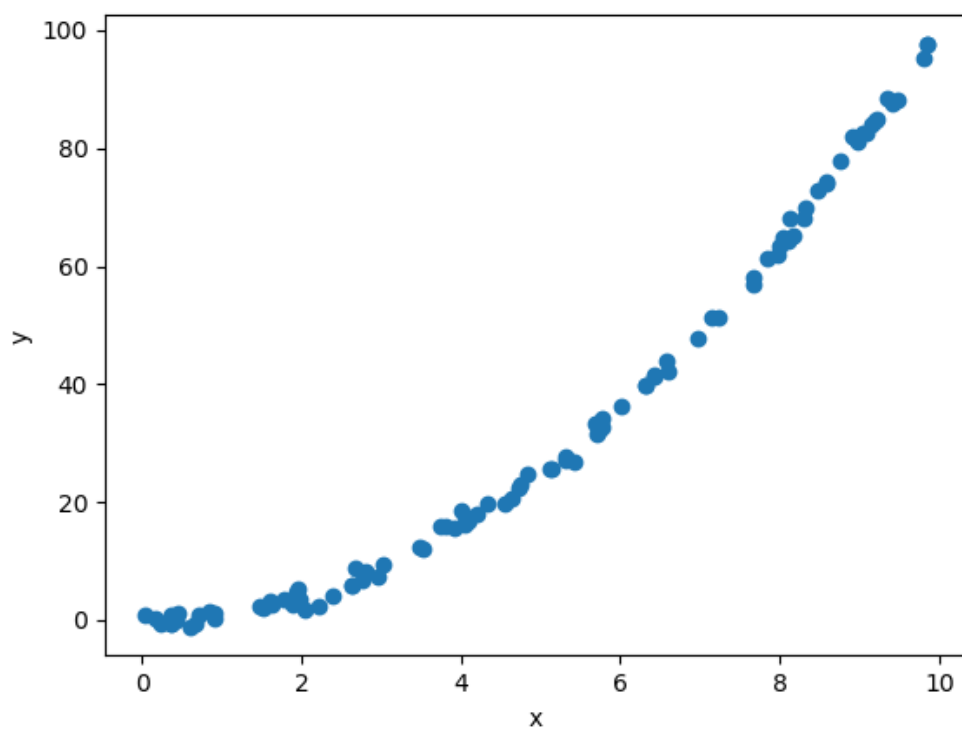


Рисунок 3. Облако точек с предполагаемой нелинейной связью между переменными x и y .

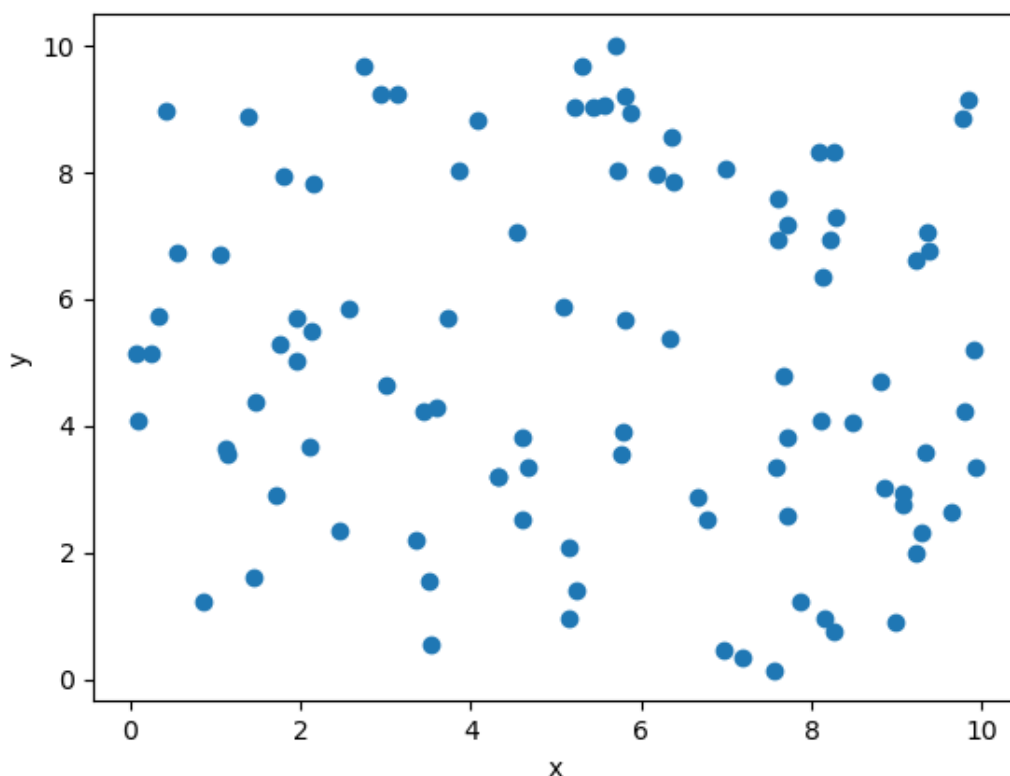


Рисунок 4. Облако точек с предполагаемым отсутствием связи между переменными x и y .

Поиск связи между переменными графическим анализом с помощью построения облаков точек может также производиться в исключительных случаях для трёх переменных, при этом облако точек становится трёхмерным. Однако визуальный анализ таких облаков обычно затруднён.

При необходимости поиска связи между множеством пар переменных в большом массиве данных анализ облаков точек нецелесообразен из-за большой трудоёмкости этой операции. В таких случаях следует опираться на другие методы, в частности, на корреляционный анализ.

Ход работы

1. Найти в открытых источниках или придумать две выборки непрерывных измерений x , y (размер выборки $n > 20$).
2. Отобразить значения графически на одномерном интервале и проверить выборки на наличие грубых ошибок.
3. Построить облако пар точек (x, y) и выдвинуть гипотезу о наличии или отсутствии связи между переменными.

Вопросы

1. Основные способы обнаружения выбросов.
2. Способы графического анализа.

3. Выдвижение гипотез о наличии связи между переменными.

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
2. Крамер Г. Математические методы статистики. М.: Мир, 1975. - 648 с
3. Ллойд Э., Ледерман У. (Ред.). Справочник по прикладной статистике. В 2-х тт. Пер. с англ. Т.1-2. 1990.

Практическое занятие № 21

Проверка статистических гипотез

Цель работы

Целью занятия является получение студентом навыков выдвижения и проверки статистических гипотез на примере применения критерия хи-квадрат для проверки гипотезы о законе распределения случайной величины.

Краткие теоретические сведения

При проведении научного исследования исследователь, как правило, имеет некоторые предварительные представления о данных перед началом их сбора и обработки, которые могут быть сформулированы в виде определённой гипотезы об изучаемом объекте. Эта гипотеза может возникать как результат теоретических исследований или являться следствием предшествующих этапов работы по изучению объекта. Пусть выдвигаемая гипотеза обозначается H_0 .

Для принятия или отклонения гипотезы H_0 на основе имеющихся данных, представленных в виде выборки, требуется ответить на вопрос, согласуется ли гипотеза H_0 с рассматриваемой выборкой. Если ответ на данный вопрос положительный, то гипотеза принимается. В противном случае можно принять некую альтернативную гипотезу H_1 . Статистические методы, которые позволяют ответить на указанный вопрос, называются методами проверки гипотез.

При проверке гипотезы с помощью статистических методов на основании имеющихся данных последние позволяют получить так называемое p -значение. Данное значение отражает вероятность того, что имеющуюся выборку данных возможно наблюдать при условии, что гипотеза H_0 верна, т.е. вероятность согласования наблюдений с первоначальным представлением об объекте. Предполагается, что все события, вероятность наступления которых очень мала, в практических приложениях можно считать невозможными. В таком случае малое значение p означает необходимость отвергнуть гипотезу H_0 . Пороговое значение α , отделяющее возможные события от невозможных, называется уровнем значимости критерия, по которому проверяется гипотеза H_0 . На практике обычно берут $\alpha \cdot 100\% = 5\%, 1\%, 0,1\%$, что соответствует

значениям $\alpha = 0,05; 0,01; 0,001$. Если окажется, что $p \geq \alpha$, то гипотеза H_0 принимается. Если окажется, что $p < \alpha$, то гипотеза H_0 отклоняется в пользу некоторой альтернативной гипотезы H_1 .

В процессе проверки гипотез с помощью специальных статистических методов возможны ошибки, обусловленные конечностью выборочных данных.

Если верная гипотеза H_0 была отклонена статистическим методом, соответствующая ошибка называется ошибкой I рода. Если же неверная гипотеза H_0 была принята, возникает ошибка II рода. Вероятность ошибки I рода совпадает с уровнем значимости критерия α . Число $\alpha \cdot 100\%$ задает риск исследователя отклонить верную гипотезу H_0 на основании обработки исходных данных. Отклонение гипотезы H_0 на уровне значимости $\alpha \cdot 100\% = 5\%$ называют значимым, на уровне $\alpha \cdot 100\% = 1\%$ – статистически значимым, на уровне $\alpha \cdot 100\% = 0,1\%$ – высоко статистически значимым.

Примером статистического критерия является критерий хи-квадрат проверки гипотезы о том, что рассматриваемая выборка извлечена из генеральной совокупности с заданным законом распределения. Общая идея критерия хи-квадрат состоит в следующем. Отрезок, содержащий выборочные значения, разбивается на непересекающиеся интервалы (для непрерывных случайных величин) либо на отдельные категории (для дискретных случайных величин). При ручном выборе разбиения важно обеспечить число элементов выборки в каждом интервале, не меньшее пяти. Изучается соответствие наблюдаемого числа попаданий ожидаемому согласно выдвинутой гипотезе о законе распределения. Если вероятность того, что наблюдаемое распределение числа попаданий по интервалам возможно при предполагавшемся законе распределения, превышает значение уровня значимости α , то гипотеза H_0 принимается. В противном случае гипотеза отвергается и предполагается, что выборка извлечена из генеральной совокупности с каким-то другим распределением.

В данной лабораторной работе предлагается в качестве примера выдвижения и проверки статистической гипотезы применить критерий хи-квадрат проверки, извлечена ли выборка из генеральной совокупности с заданным распределением. Для этого требуется совершить следующие шаги:

- Найти произвольную выборку данных.
- Выдвинуть гипотезу H_0 относительно закона распределения, например, «выборка извлечена из генеральной совокупности со стандартным нормальным распределением».
- Построить эмпирическую и теоретическую оценки плотности распределения. На листинге 5 и рисунке 5 приведён пример для нормального и асимметричного нормального распределения.
- Зафиксировать уровень значимости критерия α .
- Применить критерий хи-квадрат (можно использовать встроенные функции в Python или Excel) и получить p-value.
- Сравнить p-value с выбранным уровнем значимости α .

- Отклонить или принять гипотезу H_0 в зависимости от знака неравенства $p \geq \alpha$ ($p < \alpha$).

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

m =
[328,454,312,625,609,546,502,736,485,766,429,313,328,344,360,453,563,343
,375,28,312,361,297,437,328,328,328,297,359,328,361,703,500,344,329,312,
328,547,314,328,439,359,126,408,360,346,328,392,453,359]

samples=np.array(m)
mean=np.mean(samples)
var=np.var(samples)
std=np.sqrt(var)

x=np.linspace(min(samples), max(samples),12)
y_pdf=stats.norm.pdf(x,mean,std)
y_skew_pdf=stats.skewnorm.pdf(x,*stats.skewnorm.fit(samples))
l1=plt.plot(x,y_pdf, label='Нормальное распределение')
l2=plt.plot(x,y_skew_pdf, label='Асимметричное нормальное
распределение')
n,
bins,patches=plt.hist(samples,12,density=True,facecolor='g',edgecolor='b
lack', alpha=0.75)

plt.xlabel('Значение величины')
plt.ylabel('Частота')

plt.legend((l1,l2),(l1.get_label(), l2.get_label()), loc='upper right')

plt.axis([126, 766, 0, 0.01])
plt.show()
```

Листинг 5. Код для отрисовки теоретических оценок плотностей распределений и эмпирической функции плотности

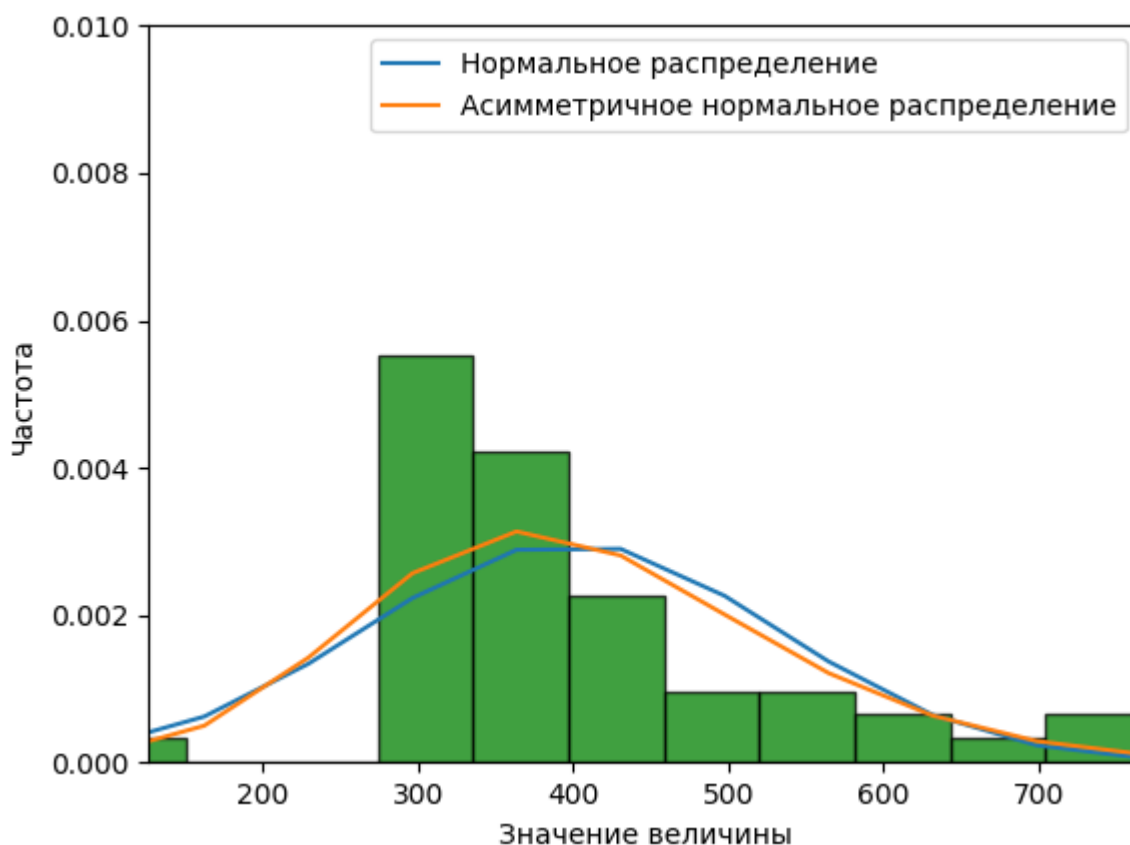


Рисунок 5. График теоретических оценок плотностей распределений и эмпирической функции плотности

Ход работы

1. Построить гистограмму плотности случайной величины и выдвинуть три разные гипотезы о законе её распределения.
2. Проверить выдвинутые гипотезы с помощью критерия хи-квадрат.
3. Самостоятельно найти функцию для проверки гипотезы о законе распределения случайной величины с помощью критерия Колмогорова-Смирнова в библиотеке `scipy.stats`. Применить её для рассматриваемых данных и сравнить результаты.

Вопросы

1. Проверка статистических гипотез.
2. Применение критерия хи-квадрат для проверки гипотезы о законе распределения случайной величины.

Литература

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика, Основы моделирования и первичная обработка данных. М.: 1983.

2. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
3. Корольюк В.С., Портенко Н.И., Скороход А.В. Турбин А.Ф. Справочник по теории вероятностей и математической статистике. М.: 1985.

Практическое занятие № 22

Корреляционные зависимости

Цель работы

Целью занятия является получение студентом навыков поиска связи между выборками и оценки силы найденной связи с помощью коэффициентов корреляции.

Краткие теоретические сведения

Рассмотрим переменные x и y , отражающие свойства некоторого объекта исследования. Будем говорить, что переменные x и y являются зависимыми, если допустимые значения пары $(x; y)$ согласованы друг с другом по некоторому закону. Зависимости могут иметь скрытый характер и сложный вид, в силу этого их приходится искать в приближённой форме.

Для оценки наличия и силы зависимости между переменными используются выборочные коэффициенты корреляции. Коэффициенты корреляции могут принимать значения от -1 до 1, знак коэффициента указывает на направление связи (прямая или обратная). В случае если коэффициент корреляции близок к нулю, связь между переменными отсутствует. Значимость коэффициентов корреляции (значимое отличие их от нуля) может быть проверено с помощью специальных статистических критериев. В прикладных задачах обычно выбирается некоторое пороговое значение коэффициента корреляции (например, 0,6), которое позволяет говорить о наличии значимой связи между переменными. Выбор порогового значения зависит от области исследования и интерпретации рассматриваемого параметра.

Существует несколько разных коэффициентов корреляции, выбор между которыми осуществляется исходя из задачи исследователя. Если между некоторыми переменными x и y интервального или порядкового типа предполагается связь в монотонной форме, то существует некоторая монотонная функция $y \approx g(x)$. Для установления или отрицания факта такой связи используется понятие ранговой корреляции по Спирмену. Формула для вычисления выборочного коэффициента ранговой корреляции по Спирмену имеет следующий вид:

$$\hat{r}_s = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (\hat{x}_i - \hat{y}_i)^2,$$

где n – объём выборок (должен совпадать для x и y), $\hat{x}_i$ и $\hat{y}_i$ – ранги i -х элементов выборок x и y .

Для установления или отрицания факта связи между x и y интервального типа в линейной форме, т.е. в виде $y \approx a + bx$, где a , b - некоторые константы, используется понятие корреляции по Пирсону. Значение выборочного коэффициента корреляции по Пирсону находится с помощью следующей формулы:

$$\bar{r}_{xy} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{s_x s_y},$$

где n – объём выборок, $\bar{x}$, $\bar{y}$ – выборочные средние, s_x , s_y – среднеквадратичные отклонения для выборок x и y соответственно.

Поскольку коэффициент корреляции по Пирсону менее устойчив к выбросам в данных, чем коэффициент корреляции по Спирмену, и может быть рассчитан только для переменных интервального типа, в практических задачах обычно имеет смысл отдавать предпочтение коэффициенту корреляции по Спирмену.

Команды на языке Python для вычисления коэффициентов корреляции по Пирсону и Спирмену с использованием библиотеки `scipy.stats`:

- `scipy.stats.pearsonr`
- `scipy.stats.spearmanr`

Ход работы

1. Рассмотреть следующие две выборки:
 $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 $Y = \{10, 34, 46, 55, 61, 76, 84, 99, 103, 110\}$
2. Построить облако точек для переменных X и Y .
3. Найти значения коэффициентов корреляции по Пирсону и по Спирмену между указанными переменными.

Вопросы

1. Зависимость между переменными. Формулы зависимости.
2. Ранговый коэффициент корреляции по Спирмену.
3. Коэффициент корреляции по Пирсону.

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
2. Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2009.

Практическое занятие № 23

Линейная регрессия

Цель работы

Целью занятия является получение студентом навыков поиска формулы линейной связи между выборками x и y на основе формулы парной регрессии.

Краткие теоретические сведения

Рассмотрим переменные x и y , отражающие свойства некоторого объекта исследования (например, рост и вес пациентов). В случае, если переменные x и y , являются зависимыми и связь между ними имеет линейный вид, возможен поиск конкретной формулы этой связи на основе линейной регрессии. Форма связи между зависимой переменной y и объясняющей переменной x в виде парной регрессии имеет следующий вид:

$$y = a + bx + u,$$

где a , b – параметры регрессии, u – скрытая переменная, отражающая влияние неизвестных факторов на y . Предполагается, что математическое ожидание переменной u равно нулю, а дисперсия $D(u)$ является постоянной величиной.

Значения параметров регрессии a и b могут быть найдены на основе значений элементов выборок x и y с помощью метода наименьших квадратов. После нахождения оценочных значений параметров a^* и b^* формируется выборка остатков на основе следующей формулы:

$$e_{ii} = y_{ii} - (a^* + b^* x_{ii}), i = 1, 2, \dots, n.$$

Выборка остатков отражает значения скрытой переменной u и требуется для проверки вышеуказанных априорных предположений относительно этой переменной.

Код с примером расчёта параметров парной линейной регрессии и построения прогноза с её помощью приведён на листинге 6.

```
import numpy as np
from sklearn.linear_model import LinearRegression

X = np.array([[1, 1], [1, 2], [2, 2], [2, 3]])
y = np.dot(X, np.array([1, 2])) + 3
reg = LinearRegression().fit(X, y)
print(reg.score(X, y))
print(reg.coef_)
print(reg.intercept_)
print(reg.predict(np.array([[3, 5]])))
```

Листинг 6. Код для расчёта параметров парной регрессии и построения прогноза

Для оценки качества описания связи между выборками с помощью найденной регрессии можно использовать значение коэффициента детерминации R^2 , отражающего, насколько лучше регрессия с объясняющей переменной x описывает значения элементов выборки y по сравнению с выборочной оценкой математического ожидания μ_y . В языке программирования Python значение коэффициента детерминации может быть рассчитано с помощью команды `r2_score` библиотеки `sklearn.metrics`.

Ход работы

Даны две выборки:

$X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$Y = \{10, 34, 46, 55, 61, 76, 84, 99, 103, 110\}$

Построить линейную регрессию $Y=f(X)$ для приведенных выборок, нарисовать её график и привести формулу. Посчитать коэффициент детерминации R^2 .

Вопросы

1. Линейная регрессия. Формула парной регрессии.
2. Выборка остатков. Коэффициент детерминации.

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
2. Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2009.
3. Крамер Г. Математические методы статистики. М.: Мир, 1975. - 648 с

Практическое занятие № 24

Анализ влияния факторов

Цель работы

Целью лабораторной работы является получение студентом навыков оценки влияния фактора, задаваемого категориальной переменной, на значения интервальной переменной на основе использования критериев Манна-Уитни и Краскела-Уоллиса.

Краткие теоретические сведения

Пусть изучается влияние фактора A на значения некоторой переменной S (например, влияние диеты на вес пациентов, влияние терапии на артериальное давление, влияние широты города на дневную температуру). Пусть влияние фактора отражается некоторым конечным числом уровней, тогда A представляется категориальной переменной. Задача оценки влияния фактора может быть решена с помощью подтверждения или опровержения гипотезы о неизменности генерального среднего $M(S)$ при изменении значения A . В ситуации, когда о законе распределении выборочных данных ничего не известно заранее, для решения указанной задачи могут применяться непараметрические методы.

Случай 1. Фактор A имеет два уровня, A_1 и A_2 . Предположим, что значения переменной S при уровнях фактора A_1 и A_2 задаются выборками x и y соответственно. Тогда анализ влияния фактора A на переменную S сводится к проверке гипотезы H_0 , заключающейся в том, что $M(x)=M(y)$, т.е. математическое ожидание переменной S не зависит от уровня фактора. Данная гипотеза может быть проверена с помощью критерия Манна-Уитни. Процедура применения критерия совпадает с описанной в лабораторной работе 2 и заключается в следующем:

- Фиксируется уровень значимости критерия α .
- В результате применения критерия Манна-Уитни к выборкам x и y (в языке программирования Python может быть применена команда `mannwhitneyu` библиотеки `scipy.stats`) получается p -значение критерия.
- По результатам сравнения p -значения с выбранным уровнем значимости α делается вывод о принятии или отклонении гипотезы H_0 .

Случай 2. Фактор A имеет несколько уровней, $A_1, A_2, \dots, A_k$. Предположим, что значения переменной S при уровнях фактора A_i задаются выборками $x_i, i=1,2,\dots,k$. Анализ влияния фактора A на переменную S сводится к проверке гипотезы H_0 , заключающейся в равенстве математических ожиданий $M(x_1), M(x_2), \dots, M(x_k)$. Для проверки данной гипотезы может быть использован

критерий Краскела-Уоллиса. Последовательность шагов для принятия или отклонения гипотезы аналогична случаю двух уровней факторов. Применение критерия Краскела-Уоллиса к проверяемым выборкам x_i на языке программирования Python осуществляется командой `kruskal` библиотеки `scipy.stats`.

Несложно заметить, что критерий Краскела-Уоллиса может быть применен для двух уровней факторов, тем самым дополняя критерий Манна-Уитни.

Ход работы

Заданы выборки X, Y, Z цен на определенное лекарство в аптеках г. Москвы, Санкт-Петербурга и Омска:

$X = \{100, 129, 205, 134, 0, 130, 156, 130, 141\}$,

$Y = \{98, 110, 102, 96, 97, 93, 101, 90, -110, 91, 94, 105, 90\}$,

$Z = \{56, 78, 96, 76, 69, 89, 61, 63, 60, 71, 68\}$.

1. Проанализировать данные на наличие грубых ошибок, выдвинуть и обосновать предположения относительно элементов выборок, которые следует исправить или удалить.
2. С помощью применения статистических критериев ответить на вопрос: «В каком из трёх городов наиболее выгодно покупать данное лекарство?» Обосновать выбор использованных критериев.

Вопросы

1. Оценка влияния факторов. Непараметрические методы.
2. Метод Манна-Уитни для двух уровней факторов. Метод Краскела-Уоллиса для нескольких уровней факторов.

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика — 12-е изд. — Москва : Издательство Юрайт, 2014. — 479 с.
2. Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2009.
3. Кобзарь А. И. Прикладная математическая статистика. — М.: Физматлит, 2006. — 466—468 с.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего профессионального образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

Демин М.С.

**ПЛАНИРОВАНИЕ И ОБРАБОТКА ЭКСПЕРИМЕНТОВ С ПРИМЕНЕНИЕМ
ЦИФРОВЫХ ТЕХНОЛОГИЙ**

Методические указания по организации и проведению самостоятельной работы

Направление подготовки 13.03.02 Электроэнергетика и электротехника
Направленность (профиль) «Электроэнергетические системы и сети»

Квалификация выпускника бакалавр
Изучается в 3 семестре

Ставрополь 2025

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего профессионального образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Демин М.С.

ПЛАНИРОВАНИЕ И ОБРАБОТКА ЭКСПЕРИМЕНТОВ С ПРИМЕНЕНИЕМ
ЦИФРОВЫХ ТЕХНОЛОГИЙ

Методические указания по организации и проведению самостоятельной работы

Направление подготовки 13.03.02 Электроэнергетика и электротехника
Направленность (профиль) «Электроэнергетические системы и сети»

Квалификация выпускника бакалавр
Изучается в 3 семестре

Ставрополь 2025

Введение

Самостоятельная работа является внеаудиторной и предназначена для самостоятельного изучения определенных тем дисциплины по материалам, рекомендованным преподавателем, а также в порядке подготовки к практическим занятиям.

Цель дисциплины «Планирование и обработка экспериментов с применением цифровых технологий» – формирование согласно ОП ВО набора общепрофессиональных и профессиональных компетенций ПК-1, 4 будущего бакалавра по направлению подготовки 13.03.02 Электроэнергетика и электротехника.

Для достижения этой цели в процессе изучения дисциплины решаются следующие задачи: ознакомление с основными понятиями и терминами изучаемой дисциплины, которыми будущий специалист будет оперировать в своей практической деятельности; освоение базовых принципов работы с современными системами автоматизированного проектирования.

Компетенции обучающегося

№ п/п	Содержание компетенции	Шифр
	<u>Профессиональные компетенции</u>	<u>ПК-(№)</u>
1	Способен участвовать в научно-исследовательской деятельности в области электроэнергетики	1
2	Способен применять интеллектуальные системы управления в электроэнергетике	4

В результате освоения дисциплины обучающийся должен:

ЗНАТЬ	– основы теории научного эксперимента; – способы представления результатов эксперимента; – методики формирования математических моделей;
УМЕТЬ	–проводить экспериментальные исследования по заданной методике; – планировать экспериментальные исследования; – обрабатывать результаты измерения;
ВЛАДЕТЬ	– навыками использования программного обеспечения для обработки результатов экспериментов; – разработки планов, программ и методики проведения испытаний.

1 Общая характеристика самостоятельной работы студента при изучении дисциплины

Основная задача высшего образования заключается в формировании творческой личности специалиста, способного к саморазвитию, самообразованию, инновационной деятельности. Решение этой задачи вряд ли возможно только путем передачи знаний в готовом виде от преподавателя к студенту. Для решения этой задачи в учебные планы всех направлений подготовки (специальностей) включена самостоятельная работа, составляющая около 50% учебного времени студента.

В широком смысле под самостоятельной работой студентов следует понимать совокупность всей самостоятельной деятельности студентов, как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие.

В учебном процессе выделяют два вида самостоятельной работы:

- аудиторная;
- внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа выполняется студентом по заданию преподавателя, но без его непосредственного участия.

Содержание внеаудиторной самостоятельной работы определяется в соответствии с рекомендуемыми видами заданий согласно примерной и рабочей программ учебной дисциплины. Видами заданий для внеаудиторной самостоятельной работы по дисциплине «Планирование и обработка экспериментов с применением цифровых технологий» являются:

- для овладения знаниями: чтение текста (учебника, первоисточника, дополнительной литературы), составление плана текста, графическое изображение структуры текста, конспектирование текста, выписки из текста, работа со словарями и справочниками, ознакомление с нормативными документами, учебно-исследовательская работа, использование аудио- и видеозаписей, компьютерной техники и Интернета и др.;

- для закрепления и систематизации знаний: работа с конспектом лекции, обработка текста, повторная работа над учебным материалом (учебника, первоисточника, дополнительной литературы, аудио и видеозаписей, составление плана, составление таблиц для систематизации учебного материала, ответ на контрольные вопросы, заполнение рабочей тетради, аналитическая обработка текста (аннотирование, рецензирование, реферирование, конспект-анализ и др.);

– для закрепления навыков – подготовка к практическим занятиям, оформление отчетов по практическим занятиям, выполнение дополнительных заданий, выданных преподавателем.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов. Контроль результатов внеаудиторной самостоятельной работы студентов может осуществляться в пределах времени, отведенного на обязательные учебные занятия по дисциплине и внеаудиторную самостоятельную работу студентов по дисциплине, может проходить в письменной, устной или смешанной форме.

Следует с первых дней обучения в вузе приучать себя к целесообразному распределению занятий по месяцам и неделям семестра. При этом очень важна равномерная работа как основное условие успешных занятий в вузе.

Высшей ступенью обучения является приобретение навыков самостоятельной творческой работы, а творческому применению знаний нужно учиться параллельно с их приобретением.

Задачами самостоятельной работы студентов по дисциплине «Планирование и обработка экспериментов с применением цифровых технологий» являются:

- формирования готовности студентов к поиску, обработке и применению информации для решения профессиональных задач;
- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- развитие познавательных способностей и активности студентов, творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- выработка навыков эффективной самостоятельной профессиональной деятельности.

2 План-график выполнения самостоятельной работы

Общий объём самостоятельной работы в рамках изучения дисциплины «Планирование и обработка экспериментов с применением цифровых технологий» составляет 54 часов.

В структуре самостоятельной работы предусмотрено самостоятельное изучение теоретического материала, подготовка к практическим занятиям.

В рамках самостоятельной работы предусмотрена подготовка студента к практическим занятиям.

В таблице 2.1 приведен план график выполнения самостоятельной работы.

Таблица 2.1 – План график выполнения самостоятельной работы

Шифр компетенции	Вид деятельности студентов	Результат самостоятельной работы	Форма контроля	Объем часов
ПК-1, 4	Изучение тем разделов теоретического курса	Конспект	Собеседование, тестирование	18
ПК-1, 4	Подготовка к практическим занятиям	Отчеты по практическому занятию	Собеседование, тестирование	18
ПК-1, 4	Подготовка к контрольной работе	Контрольная работа	Собеседование, тестирование	18
Итого за 4 семестр				54

3 Контрольные точки и виды отчетности по ним

Успеваемость студентов по дисциплине оценивается в ходе текущего и промежуточного контроля.

Текущий контроль осуществляется в течение семестра на практических занятиях.

В начале семестра студенческая группа получает график контрольных мероприятий с указанием форм и сроков их проведения, с указанием баллов за каждое контрольное мероприятие.

Проверка подготовки к практическим занятиям проводится в форме защиты отчетов по данным виду занятия.

Допуск к защите практических занятий происходит при наличии у студентов печатного варианта отчета. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя.

Максимальное количество баллов студент получает, если оформление отчета соответствует установленным требованиям, студент может правильно ответить на контрольные вопросы, а отчет полностью раскрывает суть работы. Основанием для снижения оценки являются:

- неполное выполнение задания;
- некорректность результатов и выводов;
- затруднение в ответах при защите;

Отчет может быть отправлен на доработку в следующих случаях:

- оформление не соответствует действующим требованиям;
- нет необходимых расчетных данных.

Контроль самостоятельной работы студентов

Контроль самостоятельной работы проводится преподавателем.

Предусмотрены следующие виды контроля: собеседование, тестирование.

Подробные критерии оценивания компетенций приведены в Фонде оценочных средств для проведения текущей и промежуточной аттестации.

Вопросы для собеседования

I. Основы эксперимента

1. Что такое эксперимент, и какую роль он играет в научных исследованиях?
2. В чем отличие пассивного эксперимента от активного?
3. Какие основные этапы включает в себя проведение эксперимента?

4. Какие бывают ошибки в эксперименте и как их минимизировать?
5. Что такое воспроизводимость эксперимента, и почему она важна?
6. Как классифицируются эксперименты в зависимости от условий их проведения?
7. Какие факторы могут влиять на достоверность экспериментальных данных?
8. Какие требования предъявляются к выбору оборудования для эксперимента?
9. В чем разница между лабораторным и полевым экспериментом?
10. Как определить, является ли эксперимент корректным?

II. Предварительная обработка экспериментальных данных

11. Какие этапы включает в себя предварительная обработка экспериментальных данных?
12. Как выявлять и устранять выбросы в данных?
13. Какие методы используются для заполнения пропущенных значений?
14. В чем разница между нормализацией и стандартизацией данных?
15. Какие графические методы используются для визуализации экспериментальных данных?
16. Как вычисляются основные статистические показатели (среднее, дисперсия, стандартное отклонение)?
17. Как понять, что данные требуют преобразования перед анализом?
18. Какие виды ошибок измерений бывают?
19. Как проверить данные на наличие систематических ошибок?
20. В каких случаях может потребоваться фильтрация данных?

III. Анализ результатов пассивного эксперимента. Эмпирические зависимости

21. В чем отличие пассивного эксперимента от активного?
22. Какие методы применяются для анализа данных пассивного эксперимента?
23. Как строить эмпирические зависимости на основе экспериментальных данных?
24. Какие методы используются для аппроксимации экспериментальных данных?
25. Как оценить качество полученной модели?
26. Что такое коэффициент детерминации, и как он интерпретируется?
27. В чем разница между интерполяцией и экстраполяцией?
28. Какие математические методы помогают выявлять закономерности в экспериментальных данных?
29. Как проверить адекватность эмпирической модели?
30. Как можно интерпретировать полученные закономерности в данных?

IV. Методы планирования экспериментов

31. Что такое планирование эксперимента, и зачем оно нужно?
32. Какие основные принципы планирования экспериментов?

33. Что представляет собой полный факторный эксперимент?
34. В чем разница между полным и дробным факторным экспериментом?
35. Как работает метод центрального композиционного планирования?
36. Какие преимущества дают ортогональные планы экспериментов?
37. Какие факторы важно учитывать при планировании эксперимента?
38. Как оценить точность и эффективность выбранного плана эксперимента?
39. В каких случаях используется случайный план эксперимента?
40. Как влияет число факторов на сложность планирования эксперимента?

V. Оптимизационные эксперименты

41. Что такое оптимизационный эксперимент?
42. Какие основные этапы включает в себя оптимизационное исследование?
43. Какие методы поиска оптимальных условий используются в экспериментах?
44. Как работает метод градиентного спуска в оптимизации?
45. В чем суть метода отклика поверхности (Response Surface Methodology, RSM)?
46. Какие методы применяются для многокритериальной оптимизации?
47. Как можно использовать математические модели для оптимизации?
48. Как оценить достоверность результатов оптимизационного эксперимента?
49. Как балансировать точность и затраты при проведении оптимизационного эксперимента?
50. Какие примеры оптимизационных исследований можно привести из реальной практики?

VI. Применение цифровых технологий для обработки и анализа данных

51. Какие цифровые технологии используются для обработки экспериментальных данных?
52. Какие преимущества дает автоматизация обработки данных?
53. Какие программы чаще всего применяются для статистического анализа данных?
54. Как используются методы машинного обучения в анализе экспериментальных данных?
55. Что такое облачные технологии, и как они помогают в анализе данных?
56. Какие существуют базы данных для хранения результатов экспериментов?
57. В чем преимущества Python для обработки экспериментальных данных?
58. Какие существуют методы интеллектуального анализа данных?
59. Как цифровые технологии помогают в визуализации результатов эксперимента?
60. Как использование цифровых технологий влияет на точность и достоверность эксперимента?

2. Описание шкалы оценивания

Текущее контрольное мероприятие считается сданным, если студент получил за него не менее 60% от установленного для этого контроля максимального балла. Рейтинговый балл, выставляемый студенту за текущее контрольное мероприятие, сданное студентом в установленные графиком контрольных мероприятий сроки, определяется следующим образом:

Уровень выполнения контрольного задания	Рейтинговый балл (в % от максимального балла за контрольное задание)
Отличный	100
Хороший	80
Удовлетворительный	60
Неудовлетворительный	0

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедура проведения данного оценочного мероприятия включает в себя устный ответ на практическом занятии по предложенным вопросам.

Для подготовки к данному оценочному мероприятию необходимо изучить рекомендованную основную и дополнительную литературу, материалы интернет-ресурсов, ознакомиться с медиа-текстами.

При проверке задания, оцениваются степень владения теоретическим материалом, умение применять теоретические знания на практике. Умение грамотно, аргументировано и непротиворечиво излагать теоретический материал и собственные суждения, оценки, примеры.

Оценочный лист

Тема занятия:								
№	ФИО студента	Базовый уровень				Повышенный уровень	Сумма баллов	
		Степень владения теоретическим материалом (0-5 баллов)	Примеры (0-5 баллов)	Уровень аргументации (0-5 баллов)	Собственная точка зрения (0-5 баллов)	Апелляция к самостоятельно найденным научным источникам (0-5 баллов)		

Тесты

Номер задания	Содержание вопроса
1.	Каким методом находятся коэффициенты регрессионной модели при многофакторном эксперименте? а) ковариационным анализом б) дисперсионным анализом в) методом корреляционного анализа г) наименьших квадратов
2.	Для проверки согласованности мнений специалистов вычисляют: а) коэффициент конкордации б) критерий Стьюдента в) коэффициент Фишера г) коэффициент детерминации
3.	Какой доверительный интервал следует выбрать для обеспечения нахождения в нем случайных величин с вероятностью не выше 0,683? а) $\pm 1\sigma$ б) $\pm 2\sigma$ в) $\pm 3\sigma$ г) $\pm 5\sigma$
4.	<i>Число опытов в плане ДФЭ 2^{6-2} меньше, чем в плане ПФЭ 2^6:</i> а) в два раза б) в четыре раза в) на восемь опытов г) на шестнадцать опытов
5.	Основопологающей идеей метода ДФЭ является: а) формальное приравнивание суммы нескольких факторов фактору, не входящему в эту сумму б) формальное приравнивание произведения нескольких факторов одному из факторов, входящему в это произведение в) формальное приравнивание произведения нескольких факторов фактору, не входящему в это произведение г) формальное приравнивание произведения всех факторов фактору, входящему в это произведение.
6.	Что такое сверхнасыщенные экспериментальные планы: а) планы, в которых число опытов равно числу факторов б) планы, в которых число опытов меньше числа факторов в) планы, в которых число опытов больше числа факторов г) планы, в которых число степеней свободы положительно.
7.	Вычисленные моменты распределения являются а) точечными оценками выборочных величин б) распределительными оценками вычисляемых величин в) квадратичным отклонением при вычислении точечных оценок г) дисперсией
8.	Какие библиотеки Python можно использовать для работы с вероятностными методами анализа данных?
9.	Как рассчитать среднее значение выборки в Python?
10.	Каковы основные шаги первичного вероятностного анализа данных в контексте электроэнергетики?
11.	Приведите пример, как числовая характеристика выборки может быть использована для прогнозирования нагрузки на энергосистему.

Номер задания	Содержание вопроса
12.	Какие методы Python используются для визуализации распределения данных?
13.	Как провести первичный анализ данных с использованием Python для датасета с данными по потреблению электроэнергии?
14.	Какие визуализационные библиотеки Python наиболее подходят для разведочного анализа данных в области электроэнергетики?
15.	Какие статистические метрики наиболее важны при первичном анализе данных в области электроэнергетики?
16.	Приведите пример использования гистограммы для анализа распределения частоты сбоев в сети.
17.	Как в Python проверить гипотезу о равенстве средних двух выборок данных по потреблению электроэнергии?
18.	Какой тест использовать для проверки нормальности распределения данных в Python?
19.	Приведите пример статистической гипотезы, которую можно проверить на данных электроэнергетики.
20.	Как в Python рассчитать коэффициент корреляции Пирсона?
21.	Приведите пример, как корреляционный анализ может помочь в оптимизации распределения электроэнергии.
22.	Какие выводы можно сделать, если корреляция между двумя переменными близка к 1?
23.	Какие особенности корреляционного анализа важно учитывать при его применении в электроэнергетике?
24.	Какие еще методы кроме Пирсона можно использовать для определения корреляционных зависимостей?
25.	Как построить модель линейной регрессии в Python для прогнозирования потребления электроэнергии?
26.	Какие метрики качества модели линейной регрессии наиболее важны в контексте анализа данных в электроэнергетике?
27.	Приведите пример использования линейной регрессии для определения влияния температуры на потребление электроэнергии.
28.	Как проверить адекватность модели линейной регрессии для данных о потреблении электроэнергии?
29.	Как использовать анализ влияния факторов в Python для оценки влияния различных условий на потребление электроэнергии?
30.	Приведите пример, как методы анализа влияния факторов могут быть использованы для улучшения эффективности энергопотребления.
31.	Какие статистические методы в Python позволяют оценить взаимодействие между несколькими факторами в исследованиях в области электроэнергетики?
32.	В чем заключается метод анализа дисперсии (ANOVA) и как он применяется для анализа влияния факторов?
33.	Какие выводы можно сделать на основе анализа влияния факторов для планирования нагрузки на электросеть?
34.	Какая библиотека в Python наиболее часто используется для статистического анализа данных? a) TensorFlow b) PyTorch c) Pandas d) Matplotlib

Номер задания	Содержание вопроса
35.	Какой метод используется для вычисления корреляции Пирсона между двумя переменными в Pandas? a) <code>.corr(method='pearson')</code> b) <code>.correlation()</code> c) <code>.pearson()</code> d) <code>.calculate_correlation()</code>
36.	Какую функцию из <code>scipy.stats</code> используют для проверки нормальности распределения выборки? a) <code>normtest()</code> b) <code>kurtosistest()</code> c) <code>shapiro()</code> d) <code>pearsonr()</code>
37.	Для визуализации распределения частоты использования электроэнергии подходит: a) Гистограмма b) Линейный график c) Круговая диаграмма d) Столбчатая диаграмма
38.	Какой метод в Python позволяет легко удалить строки с пропущенными значениями из DataFrame? a) <code>.dropna()</code> b) <code>.fillna()</code> c) <code>.remove_missing()</code> d) <code>.delete_nulls()</code>
39.	Какая функция из библиотеки <code>matplotlib</code> используется для создания гистограмм? a) <code>plot()</code> b) <code>scatter()</code> c) <code>hist()</code> d) <code>bar()</code>
40.	Какую библиотеку в Python обычно используют для выполнения линейной регрессии? a) Numpy b) Scipy c) Matplotlib d) Scikit-learn
41.	Какой параметр в функции <code>train_test_split</code> из <code>sklearn.model_selection</code> задает размер тестовой выборки? a) <code>test_size</code> b) <code>train_size</code> c) <code>size</code> d) <code>split_size</code>
42.	Какой тест используется для определения статистической значимости различий между средними двух выборок? a) ANOVA b) Тест Шапиро-Уилка c) t-тест d) Хи-квадрат тест

Критерии оценивания компетенций

Оценка «отлично» выставляется студенту, если он правильно выполнил 88-100 % тестовых заданий

Оценка «хорошо» выставляется студенту, если он правильно выполнил 72-87 % тестовых заданий

Оценка «удовлетворительно» выставляется студенту, если он правильно выполнил 53-71 % тестовых заданий

Оценка «неудовлетворительно» выставляется студенту, если он правильно выполнил менее 53 % тестовых заданий

4 Методические рекомендации по изучению теоретического материала

Большая часть самостоятельной работы студента состоит в изучении литературы. Одна из задач студента – научиться самостоятельно работать с книгой, а это требует определенных затрат энергии и времени. Самостоятельная работа с книгой может быть успешной, если текст не только прочитан, но и законспектирован. Существует несколько форм записей, но любая форма записи не даст нужного результата, если не будет пробуждать мысли того, кто ее ведет, если отсутствует активная работа ума и формирование своих выводов из прочитанного.

Выбор формы записи зависит от индивидуальных особенностей человека, его образованности и опыта. При этом не меньшую роль играет назначение записей, то есть то, какие задачи ставит перед собой человек (для самообразования, для выступления на семинаре, для использования в будущем).

Введение записей мобилизует наряду со зрительной памятью, также и моторную память. Кроме того, у человека, систематически ведущего записи изучаемой литературы, создается свой фонд материалов для быстрого повторения и мобилизации накопленных знаний.

Все записи должны быть убористыми и компактными. Интервалы между строками должны быть достаточными, чтобы вписывать дополнения. Рекомендуется вести записи ручкой, а карандашом или ручкой другого цвета пользоваться для отметок и выделений при последующей работе. Полезно также датировать записи.

Записи могут носить различный характер: план, выписки, тезисы, аннотирование, конспектирование.

1. План - наиболее краткая формой записи. Это перечень вопросов, рассматриваемых в книге или статье. План обычно раскрывает структуру произведения, логику автора, способствует лучшей ориентации в содержании.

Так, составленным планом можно воспользоваться, чтобы вспомнить прочитанное или быстро отыскать в книге нужное место. Представление об основных пунктах плана дает оглавление книги, поэтому во многих случаях наименования глав и разделов можно использовать в качестве пунктов. Составление плана приучает логически мыслить, вырабатывать умение сжато и последовательно излагать суть вопроса в письменной и устной форме.

Существует два способа составления плана: работа над ним по ходу чтения и составление плана после ознакомления с произведением. При этом план получается более последовательным и стройным.

Трудность составления плана состоит в том, что надо выяснить для себя, прежде всего, построение изучаемого текста, ход мыслей автора и лишь затем изложить содержание работы кратко и ясно. При всей своей краткости план дает представление о содержании прочитанного. Следует учесть, что форма плана не исключает цитирования отдельных мест и обобщений. Различают простой и развернутый план. В отличие от простого плана развернутый план не только содержит перечисление вопросов, но и раскрывает основные идеи произведения, может включать выдержки из него, схемы, таблицы. Планом, особенно развернутым, необходимо пользоваться при написании выступления или статьи.

В целом развернутый план дает гораздо большее представление о произведении, его основных идеях, задачах, которые в нем решаются. Он может включать положения, замечания, собственные мысли студента.

Важно знать, что составление планов помогает вырабатывать способность к отвлеченному, абстрактному мышлению, но наибольшую пользу составление плана даст подготовленным лицам, которые бывают достаточно лишь взглянуть на перечень основных вопросов, чтобы воспроизвести содержание прочитанного.

2. Тезисы – более сложная и совершенная форма записи, чем составление плана.

Это сжатое изложение основных мыслей прочитанного произведения или подготовляемого выступления. Особенностью тезисов является их утвердительный характер.

В них сосредотачивается самое главное, только выводы и обобщения, в них меньше доказательств, иллюстрации и пояснений. Тезисы не должны повторять дословно текст, но в ряде мест могут быть близки к нему, воспроизводя некоторые характерные выражения автора, важные для понимания хода его мыслей. Составление тезисов помогает глубже понять основные идеи произведения, выделить главное в нем; приучают сжато, точно и четко сформулировать свои мысли, повышает культуру речи и письма. При составлении тезисов учитывают следующее. Прежде всего, если произведение небольшое, необходимо внимательно изучать его в целом, если большое – изучать по главам и разделам. Затем, когда будут ясны основные идеи, кратко и последовательно излагать их в виде пунктов.

Различают простые и сложные, развернутые тезисы. Если записывают только утверждение чего – либо, такой тезис называют простым, а сложным тезисом будет выражение главной мысли, содержащее, кроме утверждения, еще и краткое ее доказательство.

Часто тезисы формулируются самим автором как выводы и обобщения в заключении книги или разделах книги. Нередко тезисы выделяются в тексте другим шрифтом.

Рекомендуется делать тезисные записи своими словами, причем можно записывать один абзац за другим, учитывая смысловую связь между ними. Но в большинстве случаев следует составлять сводный тезис, сложный по форме. При этом объединяется несколько утверждений, тесно связанных между собой.

Тезисы по содержанию очень близки к **конспекту**, но конспект носит более описательный характер, и его положения не столь категоричны, как в тезисах. Кроме того, конспект представляет собой более полную форму записи.

Следует отметить, что различие между формами записей условно, но в любой форме запись – важнейшая часть самостоятельной работы с книгой.

3. Выписки. Это записи текста из книги: теоретических положений, статистических данных, имеющих по мнению читателя важное значение.

Достоинство выписок состоит в точности воспроизведения текста книги, удобстве пользования записями при последующей работе, в накоплении обобщений и фактического материала. Выписки полезны для повторения, освежения в памяти прочитанного, для быстрой мобилизации своих знаний, когда необходимо в короткий срок вспомнить материал. Выписки выделяют из текста самое главное и тем самым помогают глубже понять его. Без них трудно обойтись при подготовке доклада, реферата, выступления. Выписки следует рассматривать как составную часть тезисов и конспектов.

Выписывать текст можно и по ходу чтения и после его завершения. В последнем случае надо замечать места, которые потом будут выписаны. Необходимо каждую выписку снабжать ссылкой на источник с указанием соответствующей страницы. Это нужно, чтобы в последствии можно было быстро найти в книге соответствующее место. Целесообразно выписывать из текста только такие места, в которых содержится самое главное, суть вопроса. Выписки должны быть ориентированы на изучение произведения в целом, а не отдельных мест, поскольку положения, вырванные из общего контакта, понимаются нередко совсем не так, как этого хотел автор. Иначе говоря, отдельно взятые, лишенные пояснений выдержки могут быть не поняты или поняты неправильно.

Выписки бывают дословные (цитаты) и «свободные», когда мысли автора излагаются своими словами. Следует учесть, что большие отрывки, которые трудно цитировать, целесообразнее в краткой форме переложить своими словами, но «яркие» и важные места лучше выписывать дословно. Каждую цитату следует заключать в кавычки. Если ее берут из середины предложения, то после вводных кавычек ставят три точки. Ставят их и в конце цитаты, если из предложения опущены последние слова.

Следует знать, что какого-либо единого метода выписок, годного для всех случаев, не существует, поскольку у каждого человека свои особенности мышления и восприятия, свой подход к теме. Все это влияет на содержание и характер выписок.

Выписки рекомендуется хранить в картотеке, конвертах или папках, на которых следует обозначить общую тему.

4. Аннотация – еще одна форма записи, являющаяся кратким обобщением содержания книги. Ею удобно пользоваться, если имеется намерение вернуться к изучаемому произведению. Аннотация может быть необходима и для того, чтобы не забыть о нем.

Для составления аннотации надо сначала полностью прочитать и глубоко продумать произведение. При всей своей краткости аннотация может содержать отдельные фрагменты авторского текста, а не только оценку книги или статьи.

5. Резюме очень близко к аннотации. Это запись, являющаяся краткой оценкой прочитанного материала. Различие между ними состоит в том, что аннотация сжато характеризует произведение в целом, а резюме концентрирует внимание на его выводах, главных итогах.

6. Конспект – наиболее совершенная и наиболее сложная форма записи. Слово «конспект» происходит от латинского «conspectus», что означает «обзор, изложение». В правильно составленном конспекте обычно выделено самое основное в изучаемом тексте, сосредоточено внимание на наиболее существенном, в кратких и четких формулировках обобщены важные теоретические положения.

Конспект представляет собой относительно подробное, последовательное изложение содержания прочитанного. На первых порах целесообразно в записях ближе держаться тексту, прибегая зачастую к прямому цитированию автора. В дальнейшем, по мере выработки навыков конспектирования, записи будут носить более свободный и сжатый характер.

Конспект книги обычно ведется в тетради. В самом начале конспекта указывается фамилия автора, полное название произведения, издательство, год и место издания. При цитировании обязательная ссылка на страницу книги. Если цитата взята из собрания сочинений, то необходимо указать соответствующий том. Следует помнить, что четкая ссылка на источник – неперемutable правило конспектирования. Если конспектируется статья, то указывается, где и когда она была напечатана.

Конспект подразделяется на части в соответствии с заранее продуманным планом. Пункты плана записываются в тексте или на полях конспекта. Писать его рекомендуется четко и разборчиво, так как небрежная запись с течением времени становится

малопонятной для ее автора. Существует правило: конспект, составленный для себя, должен быть по возможности написан так, чтобы его легко прочитал, и кто-либо другой.

Формы конспекта могут быть разными и зависят от его целевого назначения (изучение материала в целом или под определенным углом зрения, подготовка к докладу, выступлению на занятии и т.д.), а также от характера произведения (монография, статья, документ и т.п.). Если речь идет просто об изложении содержания работы, текст конспекта может быть сплошным, с выделением особо важных положений подчеркиванием или различными значками.

В случае, когда не ограничиваются переложением содержания, а фиксируют в конспекте и свои собственные суждения по данному вопросу или дополняют конспект соответствующими материалами их других источников, следует отводить место для такого рода записей. Рекомендуется разделить страницы тетради пополам по вертикали и в левой части вести конспект произведения, а в правой свои дополнительные записи, совмещая их по содержанию.

Конспектирование в большей мере, чем другие виды записей, помогает вырабатывать навыки правильного изложения в письменной форме важные теоретических и практических вопросов, умение четко их формулировать и ясно излагать своими словами.

Таким образом, составление конспекта требует вдумчивой работы, затраты времени и труда. Зато во время конспектирования приобретаются знания, создается фонд записей.

Конспект может быть текстуальным или тематическим. В текстуальном конспекте сохраняется логика и структура изучаемого произведения, а запись ведется в соответствии с расположением материала в книге. За основу тематического конспекта берется не план произведения, а содержание какой-либо темы или проблемы.

Текстуальный конспект желательно начинать после того, как вся книга прочитана и продумана, но это, к сожалению, не всегда возможно. В первую очередь необходимо составить план произведения письменно или мысленно, поскольку в соответствии с этим планом строится дальнейшая работа. Конспект включает в себя тезисы, которые составляют его основу. Но, в отличие от тезисов, конспект содержит краткую запись не только выводов, но и доказательств, вплоть до фактического материала. Иначе говоря, конспект – это расширенные тезисы, дополненные рассуждениями и доказательствами, мыслями и соображениями составителя записи.

Как правило, конспект включает в себя и выписки, но в него могут войти отдельные места, цитируемые дословно, а также факты, примеры, цифры, таблицы и схемы, взятые из книги. Следует помнить, что работа над конспектом только тогда будет творческой, когда

она не ограничена текстом изучаемого произведения. Нужно дополнять конспект данными из другими источников.

В конспекте необходимо выделять отдельные места текста в зависимости от их значимости. Можно пользоваться различными способами: подчеркиваниями, вопросительными и восклицательными знаками, репликами, краткими оценками, писать на полях своих конспектов слова: «важно», «очень важно», «верно», «характерно».

В конспект могут помещаться диаграммы, схемы, таблицы, которые придадут ему наглядность.

Составлению тематического конспекта предшествует тщательное изучение всей литературы, подобранной для раскрытия данной темы. Бывает, что какая-либо тема рассматривается в нескольких главах или в разных местах книги. А в конспекте весь материал, относящийся к теме, будет сосредоточен в одном месте. В плане конспекта рекомендуется делать пометки, к каким источникам (вплоть до страницы) придется обратиться для раскрытия вопросов. Тематический конспект составляется обычно для того, чтобы глубже изучить определенный вопрос, подготовиться к докладу, лекции или выступлению на семинарском занятии. Такой конспект по содержанию приближается к реферату, докладу по избранной теме, особенно если включает и собственный вклад в изучение проблемы.

5 Методические указания по подготовке к практическим занятиям

В рамках самостоятельной работы предусмотрена подготовка студента к выполнению практических занятий. Практические занятия позволяют интегрировать теоретические знания и формировать практические умения и навыки студентов в процессе учебной деятельности.

Цели практических занятий по дисциплине «Планирование и обработка экспериментов с применением цифровых технологий»:

- 1) закрепление теоретического материала путем систематического контроля за самостоятельной работой студентов;
- 2) формирование умений использования теоретических знаний в процессе выполнения практических занятий;
- 3) развитие аналитического мышления путем обобщения результатов практических занятий;
- 4) формирование навыков оформления результатов практических работ в виде таблиц, графиков, выводов.

На практических занятиях осуществляются следующие формы работ со студентами: индивидуальная (оценка знаний, выполненных тестовых заданий, проверка рабочих тетрадей).

6 Методические указания по подготовке к экзамену

Готовиться к экзамену необходимо в течение всего учебного времени, т.е. с первого дня очередного семестра: вся работа студента на лекциях, семинарских занятиях, консультациях, а также написание рефератов и выполнение курсовых работ и т.п. - это и есть этапы подготовки студента к зачетам и экзаменам. Подготовка к сессии должна быть нацелена не столько на приобретение новых знаний, сколько на закрепление ранее изученного материала и повторение его. Сумму полученных знаний студенту перед сессией надо разумно обобщить, привести в систему, закрепить в памяти, для чего ему надо использовать учебники, лекции, консультации, курсовые работы, рефераты и т.п., а также методические пособия и различного рода руководства. Повторение необходимо производить по разделам, темам.

Методика подготовки к сдаче зачета и экзамена одинакова и состоит из двух взаимосвязанных этапов.

1. Регулярное посещение всех учебных занятий в течение всего семестра: лекций, семинарских занятий, консультаций и т.п., а также активное изучение рекомендованной литературы, отработка (в случае необходимости) в установленные сроки всех пропущенных учебных занятий.

2. Непосредственная подготовка к зачету и экзамену, когда студенту нужно в короткий срок (2-4 дня) охватить весь изученный материал по предмету и успешно сдать экзамен. А для того, чтобы это успешно сделать, студент, в первую очередь, должен мысленно в спокойной обстановке вспомнить весь материал, изученный за семестр с тем, чтобы выявить разделы курса (предмета), слабо изученные или плохо понятые при первоначальном изучении с целью устранения пробелов в своих знаниях. Для успешной подготовки к экзаменам студенту необходимо составить себе своеобразный рабочий график, в котором отразился бы последовательный переход от темы к теме, от раздела к разделу. Во время подготовки к экзаменам могут появиться вопросы. Их нужно записать и получить ответ на предэкзаменационных консультациях. На консультациях преподаватель интересуется, как студентами изучены темы, пройденные в учебном процессе, при необходимости он разъясняет отдельные вопросы этих тем. Иногда преподаватель делает сжатый обзор важнейших тем курса, отмечает те вопросы, на которые студенты, ранее сдававшие экзамен, отвечали плохо. Во время подготовки к экзамену студенту необходимо просмотреть и собственные конспекты прослушанных лекций и самостоятельно проработанных тем семинарских занятий. Это позволит ему восстановить в памяти ранее изученные положения, выявить пробелы в своих знаниях и восполнить их из других

источников. Значение предэкзаменационных консультаций очень велико. Они призваны: 1) помочь устранить пробелы в знаниях; 2) помочь систематизировать весь ранее изученный материал; 3) информировать студентов о новейших сведениях по тому предмету, который изучен последним.

Экзамен проводится в форме устного опроса.

Вопросы к экзамену:

1. Общие свойства объектов эксперимента, подходы к их изучению, модели, основные этапы проведения эксперимента.
2. Общие сведения теории статистических выводов: основные разделы, свойства, методы.
3. Методы многомерного анализа: виды, характеристики, свойства.
4. Типы эксперимента: определения, свойства, методология.
5. Основы теории подобия: основные определения, критерии, моделирование.
6. Точность и достоверность результатов эксперимента: требования и методы достижения.
7. Ошибки и гипотезы: определения и характеристики, критические и допустимые области.
8. Проверка статистических гипотез и критериев: принципы, решаемые задачи, основной алгоритм.
9. Критерий Стьюдента: задачи, свойства, методология.
10. Критерий Фишера, критерий Кохрена, Критерий Пирсона: задачи, свойства, методология.
11. Характеристика объекта исследования: свойства, характеристики. Выбор выходных параметров (оптимизации): принципы, классификация параметров, требования к ним, методы сокращения их количества.
12. Исследуемые факторы: определение, требования к ним, классификация.
13. Отбор факторов методом экспертных оценок: определение, методология.
14. Статистический анализ экспериментальных данных: определение, основные методы, методология, достоинства и недостатки.
15. Дисперсионный анализ экспериментальных данных: определение, назначение, модель эксперимента, сущность метода, основное уравнение дисперсионного анализа.
16. Дисперсионный анализ экспериментальных данных: основное уравнения дисперсионного анализа, его составляющие и их вычисление, применение критерия Фишера.

17. Корреляционный анализ экспериментальных данных: свойства, определения, формы представления корреляционной связи, виды корреляции, корреляционное отношение, его смысл.
18. Корреляционный анализ экспериментальных данных: определение коэффициента корреляции и корреляционного отношения, оценка их достоверности и доверительные интервалы.
19. Регрессионный анализ экспериментальных данных: назначение, уравнения регрессии, теоретические и выборочные коэффициенты регрессии, определение степени уравнения и коэффициентов регрессии.
20. Регрессионный анализ при одном факторе: уравнение регрессии, вывод уравнений для определения коэффициентов регрессии.
21. Установление связи остаточной дисперсии и коэффициента корреляции.
22. Проверка значимости и адекватности модели, значимости коэффициентов регрессии.
23. Основы планирования эксперимента: активный эксперимент, методы, модель объекта, ее параметры, функция отклика, виды планов эксперимента, концепции планирования эксперимента.
24. Построение плана активного эксперимента: матрица планирования, построение плана эксперимента (основные шаги).
25. Полный факторный эксперимент: определение, уровни факторов, типы ПФЭ, матрица планирования эксперимента, геометрическая интерпретация плана эксперимента, матричная форма представления плана эксперимента.
26. Полный факторный эксперимент: построение линейной модели, эффекты фактора и взаимодействия, МПЭ.
27. Свойства матрицы ПФЭ типа 2^k .
28. Дробный факторный эксперимент: отличия от ПФЭ, построение плана, реплики, генерирующие соотношения, смешанные и несмешанные коэффициенты регрессии, разрешающая способность дробной реплики.
29. Определение коэффициентов регрессии (смешанных и несмешанных) линейной многофакторной модели методами ПФЭ и ДФЭ.
30. Проверка результатов факторного эксперимента: воспроизводимости опытов; значимости коэффициентов регрессии (полинома); адекватности модели.
31. Планы второго порядка: основные подходы, уравнение регрессии, «звездные точки», ядро планирования, виды планов, их свойства и отличия.
32. Ортогональные центральные композиционные планы: определение, свойства, построения.

33. Ротатабельные центральные композиционные планы: определение, свойства, построения.
34. Оптимальные планы эксперимента: виды, их характеристики и назначения.
35. Метод крутого восхождения (Бокса-Уилсона): основные понятия, методология.
36. Симплексный метод планирования: основные понятия, методология.
37. Цифровые технологии в экспериментальных исследованиях: классификация, области применения, особенности интеграции.
38. Автоматизация обработки экспериментальных данных: цели, методы, преимущества и риски.
39. Программное обеспечение для статистического анализа: обзор современных инструментов, их функциональные возможности и сравнительная характеристика.
40. Методы машинного обучения в анализе экспериментальных данных: принципы, типы задач, примеры применения, алгоритмы.
41. Облачные технологии в экспериментальной аналитике: архитектура, принципы работы, преимущества для обработки и хранения данных.
42. Базы данных в экспериментальных исследованиях: типы, структура, методы организации хранения и доступа к данным.
43. Python в статистическом анализе и обработке экспериментальных данных: библиотеки, методы, преимущества и ограничения.
44. Интеллектуальный анализ данных (Data Mining): основные методы, этапы анализа, примеры применения в рамках экспериментальных задач.
45. Визуализация экспериментальных данных: методы, инструменты, принципы построения графиков, визуальных моделей и панелей мониторинга.
46. Влияние цифровых технологий на точность, достоверность и воспроизводимость результатов эксперимента: критерии оценки, примеры, направления повышения эффективности.

7. Методические рекомендации по выполнению контрольной работы

Контрольная работа является важным элементом самостоятельной работы студентов. Основной целью работы является получение будущим специалистом практических навыков по решению задач, возникающих в практической деятельности, а также умения работать с литературой, делать на основе ее изучения выводы и обобщения.

Контрольная работа включает себя практические вопросы по разделу: «Методы планирования экспериментов».

Контрольная работа с учетом ее содержания оценивается по пятибалльной системе. Контрольная работа должна быть написана в сроки, устанавливаемые кафедрой. Работу, которую преподаватель признал неудовлетворительной, возвращается для переработки с учетом высказанных в отзыве замечаний.

Список рекомендуемой литературы

Основная литература

1. Химченко, А. В. Планирование эксперимента : учебное пособие / А. В. Химченко, Н. И. Мищенко, В. В. Быков. — Саратов : Вузовское образование, 2021. — 127 с. — ISBN 978-5-4487-0793-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/110117.html>. — Режим доступа: для авторизир. пользователей
2. Абрамова, И. В. Теория планирования эксперимента : учебное пособие / И. В. Абрамова, З. В. Шилова. — Соликамск : Соликамский государственный педагогический институт (филиал) ФГБОУ ВО «Пермский государственный национальный исследовательский университет», 2020. — 157 с. — ISBN 978-5-91252-120-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/104339.html>. — Режим доступа: для авторизир. Пользователей
3. Протодьяконов А.В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / Протодьяконов А.В., Пылов П.А., Садовников В.Е.. – Москва, Вологда : Инфра-Инженерия, 2022. – 392 с. – ISBN 978-5-9729-1006-9. – Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. – URL: <https://www.iprbookshop.ru/124000.html>. – Режим доступа: для авторизир. пользователей

Дополнительная литература

1. Казаков, В. Г. Планирование экспериментальных исследований и статистическая обработка данных. Основы научных исследований в промышленной теплоэнергетике : учебное пособие / В. Г. Казаков, Е. Н. Громова. — Санкт-Петербург : Санкт-Петербургский государственный университет промышленных технологий и дизайна, 2020. — 85 с. — ISBN 978-5-91646-221-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/118407.html>. — Режим доступа: для авторизир. пользователей. - DOI: <https://doi.org/10.23682/118407>
2. Мусина, О. Н. Планирование и постановка научного эксперимента : учебно-методическое пособие / О. Н. Мусина. — Москва ; Берлин : Директ-Медиа, 2015. — 89 с. : ил. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=274057>. — Библиогр. в кн. — ISBN 978-5-4475-2569-9. — DOI 10.23681/274057. — Текст : электронный.
3. Медведев, П. В. Математическое планирование эксперимента : учебное пособие / П. В. Медведев, В. А. Федотов ; Оренбургский государственный университет. — Оренбург : Оренбургский государственный университет, 2017. — 98 с. : табл., граф., схем., ил. — Режим

доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=481785>. – Библиогр.: с. 72-74. – ISBN 978-5-7410-1759-3. – Текст : электронный.

4. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. – 2-е изд. – Саратов : Профобразование, 2019. – 482 с. – ISBN 978-5-4488-0046-7. – Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. – URL: <https://www.iprbookshop.ru/88752.html>. – Режим доступа: для авторизир. Пользователей.