Дисциплина: «Основы научных исследований»

Лабораторная работа № 2
Проверка статистических гипотез.

УЧЕБНЫЕ ВОПРОСЫ:

1. Статистическая функция распределения случайной величины.
2. Точечные оценки параметров функции распределения случайной величины.

3. Построение доверительного интервала для параметров нормального распределения.
4.Проверка статистических гипотез

Рекомендуемая литература:

Учебное издание

ВЛАСОВ Владимир Михайлович
БОГУМИЛ Вениамин Николаевич
ЕФИМЕНКО Дмитрий Борисович
КОНИН Игорь Валентинович

ОСНОВЫ ОРГАНИЗАЦИИ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ В ОБЛАСТИ ТЕЛЕМАТИКИ НА АВТОМОБИЛЬНОМ ТРАНСПОРТЕ

УЧЕБНОЕ ПОСОБИЕ

Редактор Т.А. Феоктистова

Подписано в печать 02.07.2015 г. Формат 60×84/16.
Усл. печ. л. 7,0. Тираж 300 экз. Заказ . Цена 230 руб.
МАДИ, 125319, Москва, Ленинградский пр-т, 64.

ISBN 978-5-7962-0194-7 © МАДИ, 2015

Учебный вопрос №1 Статистическая функция распределения случайной величины.

Статистической функцией распределения случайной величины X ($F^*(x)$) называется частота события $X < x$ ($P^*(X < x)$) для рассматриваемых выборочных данных, т.е:

$$F^*(x) = P^*(X < x). \quad (3.8)$$

Пример 3.4. Построить статистическую функцию распределения для случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ», рассмотренной в предыдущих примерах. За исходные данные возьмем вариационный ряд, построенный по выборочным значениям случайной величины, показанный в табл. 3.2.

Учебный вопрос №1 Статистическая функция распределения случайной величины.

Решение.

По данным отказов приборов, полученным в результате пробегов автомобилей КАМАЗ, строим график статистической функции распределения случайной величины X.

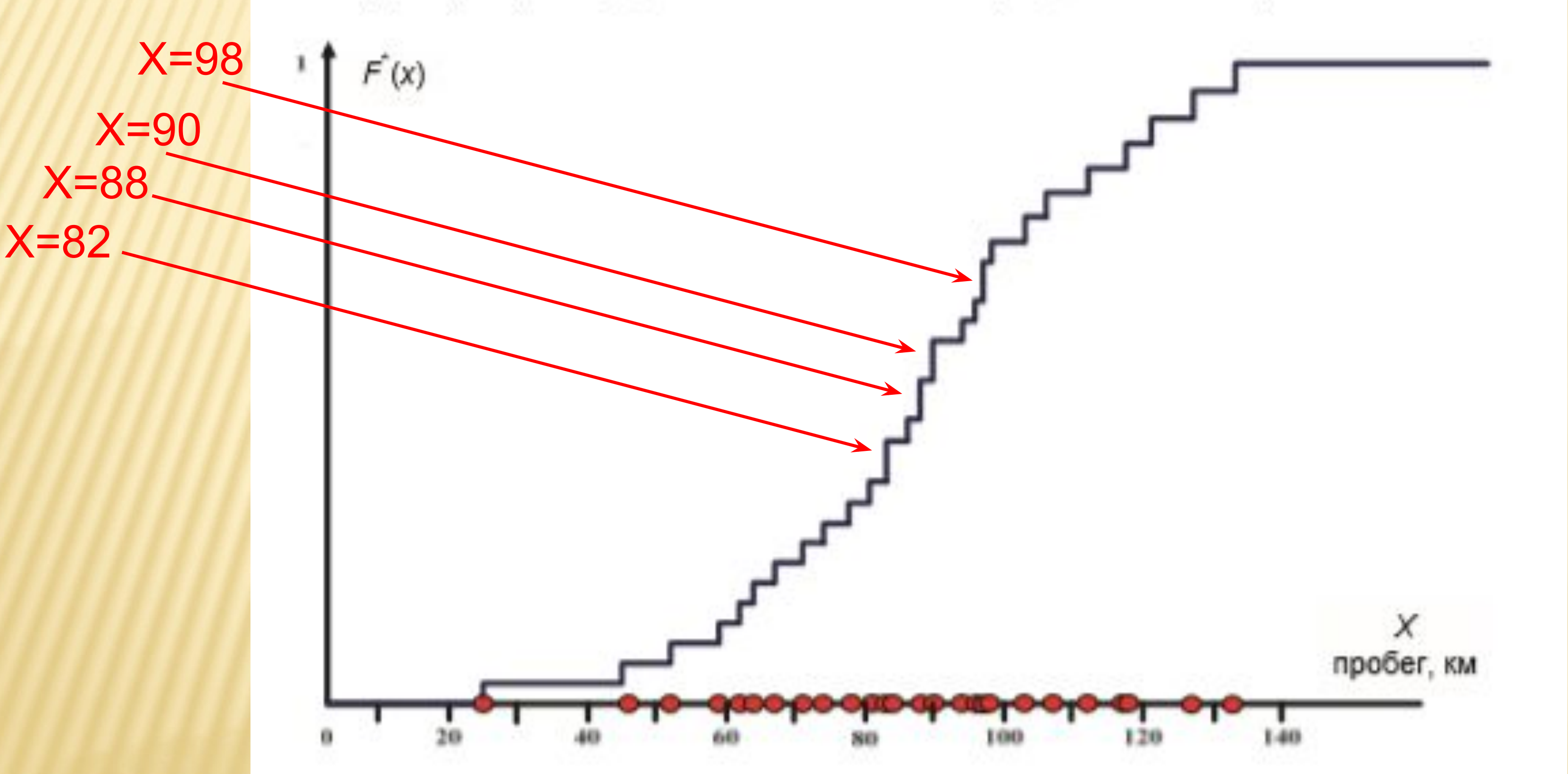

Рис. 3.2. Статистическая функция распределения случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ» (по данным табл. 3.2)

Учебный вопрос №2 Точечные оценки параметров функции распределения случайной величины.

2.1 Основные определения.

Любая функция от выборочных данных, в том числе любые оценки параметров называются *статистикой*, т.е. статистика есть функция от выборочных данных.

Оценкой $\tilde{\theta}_n$ некоторого параметра θ называют всякую функцию (статистику), рассчитываемую по результатам выборочных наблюдений над с.в. X, с помощью которой судят о значении параметра θ. Таким образом:

$$\tilde{\theta}_n = \tilde{\theta}_n(x_1, x_2, \ldots, x_n). \qquad (3.9)$$

Оценка *несмещённая*, если её $M\tilde{\theta}_n = \theta$.

Оценка *асимптотически несмещенная* если $\tilde{\theta}_n \xrightarrow{P} \theta,\ n \to \infty$.

Оценка называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру. Любая состоятельная либо асимптотически несмещённая, либо несмещённая.

Эффективная оценка. Несмещённая оценка $\tilde{\theta}_n$ называется эффективной, если она имеет наименьшую дисперсию среди всех возможных несмещённых оценок параметра θ.

2.2 Нахождение точечной оценки математического ожидания случайной величины по данным выборки.

Пусть имеется выборка объема n из генеральной совокупности: $x_1, x_2, \ldots, x_n$. Тогда *точечной оценкой* математического ожидания случайной величины является арифметическая средняя выборочных значений:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \qquad (3.10)$$

Пример 3.5. Найти точечную оценку математического ожидания случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ», по данным табл. 3.2.

Шаг 1. Найдем сумму выборочных значений:

$$\sum_{i=1}^{30} x_i = 25 + 46 + 52 + 58 + 62 + 63 + 66 + 72 + 74 + 78 + 81 + 82 + 82 + 84 + 88 + 88 + 90 + 90 + 94 + 96 + 98 + 98 + 99 + 104 + 106 + 112 + 116 + 118 + 126 + 134 = 2\,582.$$

Шаг 2. Найдем арифметическое среднее 30 выборочных значений: $\bar{x} = 2\,582/30 = 86{,}07$ тыс. км.

2.3 Нахождение точечной оценки дисперсии, среднего квадратичного отклонения и коэффициента вариации случайной величины по данным выборки

Точечная оценка дисперсии определяется по формуле:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad (3.11)$$

где n – объем выборки, $\bar{x}$ – арифметическая средняя величина выборки, является состоятельной, несмещенной оценкой дисперсии случайной величины.

Оценка среднего квадратичного отклонения s найдется как корень квадратный величины s^2.

Оценка коэффициента вариации v найдется как отношение оценок: среднего квадратичного отклонения к среднему арифметическому.

Пример 3.6. Найти точечную оценку дисперсии, среднего квадратичного отклонения и коэффициента вариации случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ», по данным табл. 3.2.

По данным примера 3.5 средняя арифметическая величина выборки $\bar{x}$ = 86,07 тыс. км.

Шаг 1. Найдем разности значений $(x_i - \bar{x})$, квадраты разностей данных значений $(x_i - \bar{x})^2$. Сведем рассчитанные значения в табл. 3.4.

Таблица 3.4

Расчетные значения $(x_i - \bar{x})$, $(x_i - \bar{x})^2$ (по данным примера 3.5)

№ п/п	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	№ п/п	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	25	(25 – 86,07) = –61,7	3806,89	16	88	(88 – 86,07) = 1,93	3,7249
2	46	(46 – 86,07) = –40,07	1605,6049	17	90	(90 – 86,07) = 3,93	15,4449
3	52	(52 – 86,07) = –34,07	1160,7649	18	90	(90 – 86,07) = 3,93	15,4449
4	58	(58 – 86,07) = –28,07	787,9249	19	94	(94 – 86,07) = 7,93	62,8849
5	62	(62 – 86,07) = –24,07	579,3649	20	96	(96 – 86,07) = 9,93	98,6049
6	63	(63 – 86,07) = –23,07	532,2249	21	98	(98 – 86,07) = 11,93	142,3249
7	66	(66 – 86,07) = –20,07	402,8049	22	98	(98 – 86,07) = 11,93	142,3249
8	72	(72 – 86,07) = –14,07	197,9649	23	99	(99 – 86,07) = 12,93	167,1849
9	74	(74 – 86,07) = –12,07	145,6849	24	104	(104 – 86,07) = 17,93	321,4849
10	78	(78 – 86,07) = –8,07	65,1249	25	106	(106 – 86,07) = 19,93	397,2049
11	81	(81 – 86,07) = –5,07	25,7049	26	112	(112 – 86,07) = 25,93	672,3649
12	82	(82 – 86,07) = –4,07	16,5649	27	116	(116 – 86,07) = 29,93	895,8049
13	82	(82 – 86,07) = –4,07	16,5649	28	118	(118 – 86,07) = 31,93	1019,5249
15	88	(88 – 86,07) = 1,93	3,7249	30	134	(134 – 86,07) = 47,93	2297,2849

Шаг 2. Найдем сумму квадратов разностей $(x_i - \overline{x})^2$:

$$\sum_{i=1}^{30}(x_i - \overline{x})^2 = 3806{,}89 + 1605{,}6049 + 1160{,}7649 + 787{,}9249 + 579{,}3649 +$$
$$+ 532{,}2249 + 402{,}8049 + 197{,}9649 + 145{,}6849 + 65{,}1249 + 25{,}7049 +$$
$$+ 16{,}5649 + 16{,}5649 + 4{,}2849 + 3{,}7249 + 3{,}7249 + 15{,}4449 + 15{,}4449 +$$
$$+ 62{,}8849 + 98{,}6049 + 142{,}3249 + 142{,}3249 + 167{,}1849 + 321{,}4849 +$$
$$+ 397{,}2049 + 672{,}3649 + 895{,}8049 + 1019{,}5249 + 2297{,}2849 =$$
$$= 17197{,}2121.$$

Шаг 3. Найдем величину s^2, поделив сумму квадратов разностей $(x_i - \overline{x})^2$ на величину $(n - 1)$. Получим:

$$s^2 = \frac{17197{,}2121}{(30-1)} =\sim 593{,}00.$$

Шаг 4. Найдем оценку среднего квадратичного отклонения $s = \sqrt{s^2}$:

$$s = \sqrt{593} = 24{,}35 \text{ тыс. км}.$$

Шаг 5. Найдем оценку коэффициента вариации:

$$v = \frac{s}{\overline{\overline{x}}} = \frac{24{,}35}{86{,}07} = 0{,}283.$$

Для нормально распределенной случайной величины значение коэффициента вариации V должно быть меньше 0,3. Исходя из этого, гипотеза о нормальном законе распределения для случайной величины (пробеги автомобилей КАМАЗ до отказа прибора), правдоподобна.

Учебный вопрос №3 Построение доверительного интервала для параметров нормального распределения.

3.1 Основные определения.

Задача интервального оценивания параметра формулируется следующим образом: по данным выборки (x_1, x_2, ..., x_{1n}) построить числовой интервал (θ_1, θ_2), относительно которого с заранее выбранной вероятностью γ можно сказать, что внутри данного интервала находится точное значение оцениваемого параметра.

Интервал (θ_1, θ_2), накрывающий с вероятностью γ истинное значение параметра, называется *доверительным интервалом*, а вероятность γ называется *надежностью оценки* или *доверительной вероятностью*.

Постановка задачи: Пусть случайная величина $X \sim N(a, \sigma)$, a, σ неизвестны. Имеется выборка объема n, по данным которой получена оценка математического ожидания случайной величины a, как среднее арифметическое значений выборки $\overline{X}$. Для заданной доверительной вероятности γ найти такое положительное число ε, чтобы выполнялось соотношение:

$$P\{|\overline{X} - a| < \varepsilon\} = \gamma. \tag{3.12}$$

Введем случайную величину

$$T = \frac{\overline{X} - a}{\dfrac{s}{\sqrt{n}}}, \tag{3.13}$$

где $\overline{X}$ – среднее арифметическое значений выборки.

Состоятельная, несмещенная оценка дисперсии случайной величины имеет специальное обозначение s^2 и определяется из выражения

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{3.14}$$

Выборочная оценка среднего квадратичного отклонения имеет обозначение s и находится как корень квадратный из величины s^2:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{3.15}$$

Доказано, что случайная величина T, определяемая соотношением (3.13), имеет распределение Стьюдента с $n - 1$ степенями свободы. Поделим обе части неравенства в выражении (3.12) на величину $s/\sqrt{n}$. Получим:

$$P\{|\frac{\overline{X} - a}{\frac{s}{\sqrt{n}}}| < \frac{\varepsilon}{\frac{s}{\sqrt{n}}}\} = \gamma. \quad (3.16)$$

Таким образом, исходя из выражения (3.15), можем сказать, что поставленная задача, формально описываемая выражением (3.12), свелась к следующей задаче. Найти величину ε, удовлетворяющую выражению:

$$P\{|T| < \frac{\varepsilon}{\frac{s}{\sqrt{n}}}\} = \gamma \text{ или } P\{|T| < t_\gamma\} = \gamma, \quad (3.17)$$

где $t_\gamma = \frac{\varepsilon\sqrt{n}}{s}$.

Величина t_γ находится из условия:

$$P\{|T| < \gamma\} = \int_{-t_\gamma}^{t_\gamma} f_T(t_{\gamma,n-1})dt = \gamma. \quad (3.18)$$

Поскольку функция симметрична относительно оси OY, равенство (3.18) можно переписать в виде

$$P\{|T| < \gamma\} = 2\int_{0}^{t_\gamma} f_T(t_{\gamma,n-1})dt = \gamma. \quad (3.19)$$

Пользуясь таблицей распределения Стьюдента, находим значение t_γ в зависимости от выбранного значения доверительной вероятности γ и числа степеней свободы, равного $n - 1$. Тогда величина ε из выражения (3.17) определится следующим образом:

$$\varepsilon = \frac{t_\gamma s}{\sqrt{n}}. \qquad (3.20)$$

Доверительный интервал найдется из выражения $(\bar{x} - \varepsilon, \bar{x} + \varepsilon)$.

3.2 Пример построения доверительного интервала.

Пример 3.7. Рассмотрим задачу построения доверительного интервала для неизвестного математического ожидания случайной величины, распределенной нормально, при неизвестной дисперсии, для условий примера 3.6.

В данном примере рассматривалась случайная величина $X \sim N(a, \sigma)$ «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ». При этом параметры a, σ неизвестны. По выборке объема $n = 30$ была получена оценка математического ожидания случайной величины a, как среднее арифметическое значений выборки $\bar{x}$. По данным примера 2.5 средняя арифметическая величина выборки $\bar{x}$ = 86,07 тыс. км. Оценка среднего квадратичного отклонения s = 24,35 тыс. км.

Выберем значение доверительной вероятности γ = 0,9. По таблице распределения Стьюдента при заданной доверительной вероятности найдем значение квантиля t_γ для двусторонней критической области при уровне значимости $\alpha = 1 - 0,9 = 0,1$ и при числе степеней свободы $k = n - 1 = 30 - 1 = 29$. $t_\gamma = 1,7$. По формуле (3.20) получим:

$$\varepsilon = \frac{t_\gamma s}{\sqrt{n}} = \frac{1{,}7 \cdot 24{,}35}{\sqrt{30}} = \frac{41{,}395}{5{,}48} = 7{,}55 \text{ тыс. км.} \qquad (3.21)$$

Доверительный интервал найдется из выражения:

$$(\bar{x} - \varepsilon, \bar{x} + \varepsilon) = (86{,}07 - 7{,}55, 86{,}07 + 7{,}55) = (78, 52, 93, 62). \qquad (3.22)$$

Разность границ доверительного интервала Δ равна:

$$\Delta = 2\varepsilon. \qquad (3.23)$$

Подставляя полученные значения, имеем

$$\Delta = 2 \times 7{,}55 = 15{,}01 \text{ тыс. км.} \qquad (3.24)$$

Относительная погрешность в процентах полученной оценки математического ожидания случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ» при установленной доверительной вероятности, равной 0,9, определиться из соотношения:

$$\delta = \frac{\Delta}{\bar{x}} 100\%. \qquad (3.25)$$

Подставляя найденные значения получим:

$$\delta = \frac{15{,}01}{86{,}07} 100\% = 17{,}4\%. \qquad (3.26)$$

Относительная погрешность полученной оценки достаточно велика.

Значения α-пределов распределения Стьюдента $t_{\alpha,k}$ в зависимости от числа степеней свободы k

k	α				k	α			
	0,1	0,05	0,025	0,01		0,1	0,05	0,025	0,01
1	6,31	12,7	25,5	63,7	14	1,76	2,14	2,51	2,98
2	2,92	4,30	6,21	9,92	16	1,75	2,12	2,47	2,92
3	2,35	3,18	4,18	5,84	18	1,73	2,10	2,44	2,88
4	2,13	2,78	3,50	4,60	20	1,72	2,09	2,42	2,84
5	2,02	2,57	3,16	4,03	22	1,72	2,07	2,40	2,82
6	1,94	2,45	2,97	3,71	24	1,72	2,06	2,39	2,80
7	1,90	2,37	2,84	3,50	26	1,71	2,06	2,38	2,78
8	1,86	2,31	2,75	3,36	28	1,70	2,05	2,37	2,76
9	1,83	2,26	2,68	3,25	30	1,70	2,04	2,36	2,75
10	1,81	2,23	2,63	3,17	∞	1,64	1,96	2,24	2,58
12	1,78	2,18	2,56	3,06					

Учебный вопрос №4 Проверка статистических гипотез

4.1 Основные понятия.

Статистические гипотезы делятся на гипотезы о параметрах распределения известного вида (*параметрические гипотезы*) и о виде неизвестного закона распределения (*непараметрические гипотезы*). Пример непараметрической гипотезы: Случайная величина «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ» имеет нормальное распределение. Параметрическая гипотеза, в которой однозначно фиксируется значение некоторого параметра, называется *простой*. Если в параметрической гипотезе некоторому параметру сопоставляется множество значений, гипотеза называется *сложной*. Пример простой параметрической гипотезы: Случайная величина «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ» имеет математическое ожидание 90,0 тыс. км.

Пример сложной параметрической гипотезы: Случайная величина «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ», имеет математическое ожидание более 90,0 тыс. км.

Выдвинутую гипотезу считают *основной* и обозначают индексированным символом H_0. Формулируют также противоположную гипотезу, которую считают *альтернативной* и обозначают индексированным символом H_1.

Все выше приведенные примеры статистических гипотез можно рассматривать также как примеры основных статистических гипотез. Проверка выдвинутых статистических гипотез может осуществляться только по имеющимся данным выборки ($x_1, x_2, \ldots, x_{1n}$) с помощью специально выбираемой функции от выборочных данных $C_n = C(x_1, x_2, \ldots, x_{1n})$, связанной с проверяемым параметром или законом распределения. Такую функцию называют *статистический критерий*.

Поскольку статистический критерий является функцией выборки, его значения, рассчитанные по выборочным данным, являются случайной величиной. На этом факте основано правило проверки любой статистической гипотезы, которое заключается в следующем. Множество возможных значений критерия разбивается на два не пересекающихся подмножества:

1) критическую область S, – область значений критерия, при которых основная гипотеза H_0 отклоняется, принимается альтернативная гипотеза H_1;

2) область $\overline{S}$ – область значений критерия, при которых основная гипотеза H_0 принимается.

Построение критической области основано на *принципе практической уверенности*, который формулируется следующим образом: если вероятность события A очень мала, то при однократном испытании можно быть уверенным, что событие A не произойдет и в практической деятельности действовать так, как будто событие A невозможно.

Следовательно, в соответствии с принципом практической уверенности, критическая область критерия, – это область маловероятных значений критерия. Таким образом, если в результате опыта рассчитанное значение критерия C_n попало в критическую область, вероятность которого мала, основная гипотеза H_0 будет отклонена.

Поскольку значение критерия случайно, то при его использовании можно получить ошибочное заключение двух родов.

Ошибка первого рода – основная гипотеза H_0 отклоняется, когда на самом деле она верна.

Ошибка второго рода – основная гипотеза H_0 принимается, когда на самом деле верна альтернативная гипотеза H_1.

Ошибки первого и второго рода возникают, когда критические области критериев для основной и альтернативной гипотез пересекаются. Вероятность ошибки первого рода обозначают греческой буквой α. Вероятность ошибки второго рода обозначают греческой буквой β.

Границу критической области обозначают $t_{кр}$. Тогда по расположению значений критерия C_n относительно критической области критерии подразделяются: 1) правосторонние S: $C_n > t_{кр}$; 2) левосторонние S: $C_n < t_{кр}$; 3) двусторонние S: $C_n < t_{кр1}$ и $C_n > t_{кр2}$.

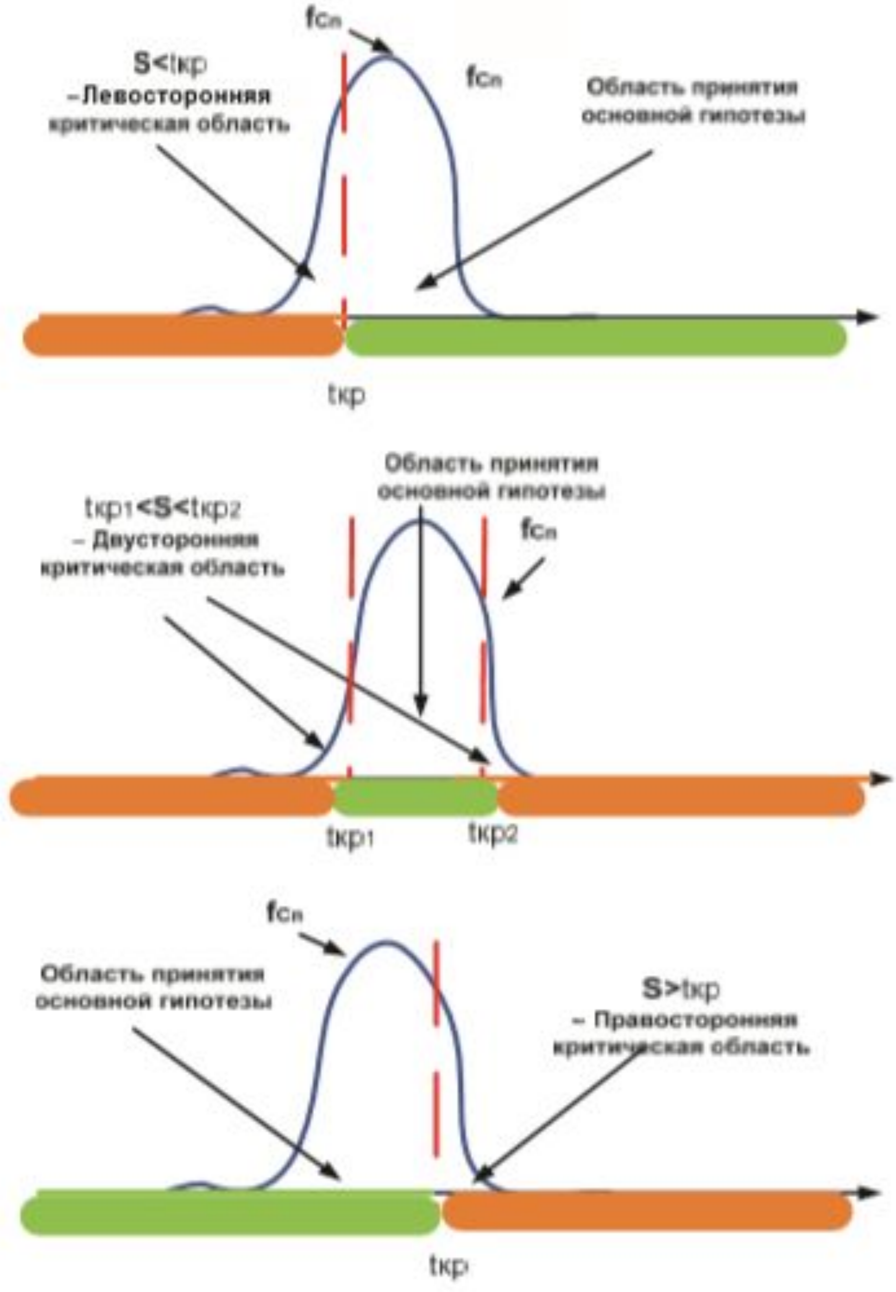

Рис. 3.3. Примеры различных критических областей статистических критериев

4.2 Проверка гипотез о законе распределения с использованием критерия Пирсона χ^2

Как правило, закон распределения случайной величины не известен. Однако можно сделать предположение о виде закона распределения на основании различных признаков, т.е. сформулировать гипотезу H_0 о виде закона распределения:

H_0: $F_x(x) = F_0(x)$.

Альтернативная гипотеза H_1: $F_x(x) \neq F_0(x)$.

Для проверки гипотезы H_0 о распределении случайной величины, разбивают всю область значений случайной величины на m интервалов Δ_1, Δ_2, ..., Δ_m.

Используя предполагаемый закон распределения $F_0(x)$, рассчитывают вероятность p_i попадания случайной величина в интервал Δ_i, $i = 1, 2, \ldots, m$, используя формулу:

$$p_i = P\{\alpha \leq X \leq \beta\} = F_0(\beta) - F_0(\alpha). \quad (3.27)$$

Тогда теоретическое число значений выборки объема n, попавших в интервал Δ_i, определится как $p_i n$.

Получим следующую таблицу теоретического распределения выборочных значений случайной величины по интервалам (табл. 3.5):

Таблица 3.5

Теоретическое распределение выборочных значений случайной величины по интервалам

Δ_1	Δ_2	...	Δ_m
$p_1 n$	$p_2 n$	...	$p_m n$

Британский ученый Карл Пирсон (1857–1936) разработал статистический критерий χ^2 следующего вида:

$$\chi^2 = \sum_{i=1,\dots,m} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1,\dots,m} \frac{n_i^2}{np_i} - n, \qquad (3.28)$$

где n – объем выборки; n_i – количество данных выборки, попавших в интервал Δ_i; p_i – теоретическая вероятность попадания случайной величины в интервал Δ_i; np_i – теоретическое число значений случайной величины из выборки объема n, которые должны попасть в интервал Δ_i.

Пирсоном были разработаны для практического применения таблицы значений квантилей $t_{кр}$ для различных значений k-степеней свободы и различных значений уровня значимости α.

Порядок применения критерия χ^2 Пирсона следующий:

1. Получить выборку объема n.

2. Сформулировать гипотезу H_0 о виде закона распределения, H_0: $Fx(x) = F_0(x)$.

3. Рассчитать точечные оценки параметров распределения по выборочным данным.

4. Разбить область значений случайной величины на m интервалов. При этом в каждый интервал должно попасть не менее пяти выборочных значений. Требования равенства величины интервалов при этом не выставляется.

5. По формуле (3.27) рассчитать теоретические значения вероятностей p_i попадания случайной величины в интервал Δ_i $i = 1, 2, \ldots, m$.

6. Для каждого интервала разбиения рассчитать теоретическое значение количества выборочных данных, попадающих в интервал Δ_i, по формуле $p_i n$.

7. По формуле (3.28) рассчитать значение критерия Пирсона.

8. По таблице квантилей χ^2 распределения Пирсона для заданного значения уровня значимости α и рассчитанного значения числа степеней свободы к найти величину $t_{кр}$.

9. Если рассчитанное по формуле (3.28) значение критерия χ^2 Пирсона меньше табличного значения величины $t_{кр}$, гипотеза H_0 о виде закона распределения принимается. В противном случае отклоняется.

4.3 Пример проверки гипотеза о законе распределения

Пример 3.8. Проверить гипотезу о нормальном законе распределения случайной величины «Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ». Полученные выборочные данные приведены в табл. 3.6.

Рассчитаем среднее арифметическое значений наработки и примем полученное значение, как точечную оценку математического ожидания случайной величины по формуле:

$$\overline{X} = \sum_i \overline{x}_i \frac{n_i}{n}. \qquad (3.29)$$

Таблица 3.6

Пробеги до отказа бортовых навигационно-связных терминалов, установленных на автомобили КамАЗ

x_i	43	62	64	69	72	73	84	87	89
n_i	1	1	2	1	1	2	2	3	3
n_i/n	1/40	1/40	2/40	1/40	1/40	2/40	2/40	3/40	3/40

x_i	92	94	95	96	99	103	107	121	134
n_i	6	3	2	4	2	3	2	1	1
n_i/n	6/40	3/40	2/40	4/40	2/40	3/40	2/40	1/40	1/40

Подставляя табличные значения получим:

$$\bar{\bar{X}} = 43\times1/40 + 62\times1/40 + 64\times2/40 + 69\times1/40 + 72\times1/40 + 73\times2/40 + 84\times2/40 + 87\times3/40 + 89\times3/40 + 92\times6/40 + 94\times3/40 + 95\times2/40 + 96\times4/40 + 99\times2/40 + 103\times3/40 + 107\times2/40 + 120\times1/40 + 134\times1/40 = 90{,}00 \text{ тыс. км.}$$

Рассчитаем оценку дисперсии по формуле:

$$s^2 = \frac{1}{(n-1)}\sum_i (x_i - \bar{X})^2. \qquad (3.30)$$

Подставляя выборочные значения получим:

$$\begin{aligned} s^2 = \frac{1}{(n-1)}\sum_i (x_i - \bar{X})^2 = 1/39[(43-90)^2 + (62-90)^2 + \\ + 2(64-90)^2 + (69-90)^2 + (72-90)^2 + 2(73-90)^2 + \\ + 2(84-90)^2 + 2(87-90)^2 + 3(89-90)^2 + 6(92-90)^2 + \\ + 3(94-90)^2 + 2(95-90)^2 + 4(96-90)^2 + 2(99-90)^2 + \\ + 3(103-90)^2 + 2(107-90)^2 + (121-90)^2 + (134-90)^2 + \\ = 1/39[2209 + 784 + 1352 + 441 + 324 + \\ + 578 + 72 + 18 + 3 + 507 + 578 + 961 + 1936 + 24 + \\ + 48 + 50 + 144 + 162] = 10191/39 = 261{,}3. \end{aligned} \qquad (3.31)$$

Найдем оценку среднего квадратичного отклонения из выражения:

$$s = \sqrt{s^2}. \qquad (3.32)$$

Получим:

$$s = \sqrt{261{,}3} = 16{,}165. \qquad (3.33)$$

Таким образом, если наша гипотеза верна, изучаемая случайная величина распределена нормально: $X \sim N(a, \sigma^2)$ с параметрами $\alpha = 90$; $\sigma^2 = 261,3$.

Распределение количества выборочных данных по равномерным интервалам приведены в табл. 3.7.

Таблица 3.7

Распределение количества выборочных данных по равномерным интервалам

Интервал	Диапазон, тыс. км	Частота, f	n_i/n
2	43–61	1`	1/40
3	61–79	7	7/40
4	79–97	23	23/40
5	97–115	7	7/40
6	115–134	2	2/40

В соответствии с требованиями критерия Пирсона χ^2, перегруппируем данные так, чтобы в каждом интервале было не менее пяти значений.

Получим следующее распределение (табл. 3.8):

Таблица 3.8

Распределение количества выборочных данных по равномерным интервалам

Интервал	Диапазон, тыс. км.	Частота, f	n_i/n
1	43–79	8	0,2
2	79–97	23	0,575
3	97–134	9	0,225

В соответствии с методикой расчета значения критерия χ^2 на следующем шаге необходимо найти вероятности попадания случайной величины в каждый интервал разбиения. Для этого необходимо найти значения функции распределения случайной величины в каждой граничной точке, а именно для $x_1 = 79$, $x_2 = 97$, $x_3 = 134$.

Найдем соответствие между граничными значениями величины X и Y.

Для $x_1 = 79$ $y_1 = (x_1 - \alpha)/\sigma = (79 - 90)/16{,}165 = -0{,}68$.

Для $x_2 = 97$ $y_2 = (x_2 - \alpha)/\sigma = (97 - 90)/16{,}165 = 0{,}433$.

Для $x_3 = 134$ $y_3 = (x_3 - \alpha)/\sigma = (134 - 90)/16{,}165 = 2{,}72$.

Используя табличные значения функции распределения стандартного нормального закона, найдем ее значения в граничных точках и сведем полученные значения в табл. 3.9:

Таблица 3.9

Значения функции распределения в граничных точках

x_i	$x_1 = 79$	$x_2 = 97$	$x_3 = 134$
y_i	$y_1 = -0{,}68$	$y_2 = 0{,}433$	$y_3 = 2{,}72$
$F(y_i) = F(x_i)$	0,2483	0,6664	0,9967

x	$\Phi(x)$	x	$\Phi(x)$
—2,70	0,0035	—1,06	0,1446
—2,60	0,0047	—1,04	0,1492
—2,50	0,0062	—1,02	0,1539
—2,40	0,0082	—1,00	0,1587
—2,30	0,0107	—0,98	0,1635
—2,20	0,0139	—0,96	0,1685
—2,10	0,0179	—0,94	0,1736
—2,00	0,0228	—0,92	0,1788
—1,98	0,0239	—0,90	0,1841
—1,96	0,0250	—0,88	0,1894
—1,94	0,0262	—0,86	0,1949
—1,92	0,0274	—0,84	0,2005
—1,90	0,0288	—0,82	0,2061
—1,88	0,0301	—0,80	0,2119
—1,86	0,0314	—0,78	0,2177
—1,84	0,0329	—0,76	0,2236
—1,82	0,0344	—0,74	0,2297
—1,80	0,0359	—0,72	0,2358
—1,78	0,0375	—0,70	0,2420
—1,76	0,0392	—0,68	0,2483

x	Φ(x)	x	Φ(x)
0	0,5	0,3	0,6179
0,01	0,504	0,31	0,6217
0,02	0,508	0,32	0,6255
0,03	0,512	0,33	0,6293
0,04	0,516	0,34	0,6331
0,05	0,5199	0,35	0,6368
0,06	0,5239	0,36	0,6406
0,07	0,5279	0,37	0,6443
0,08	0,5319	0,38	0,648
0,09	0,5359	0,39	0,6517
0,1	0,5398	0,4	0,6554
0,11	0,5438	0,41	0,6591
0,12	0,5478	0,42	0,6628
0,13	0,5517	0,43	0,6664

Рассчитаем теоретические значения попадания случайной величины в каждый интервал (табл. 3.10):

Таблица 3.10

Теоретические значения попадания случайной величины в каждый интервал

Δ_i	$y_1 - \infty$	$y_2 - y_1$	$y_3 - y_2$
границы	$[-0{,}68, -\infty)$	$[0{,}433, -0{,}68]$	$[2{,}72, -0{,}433]$
$p_i = F(y_{i+1}) - F(y_i)$	0,2483	0,6664 – 0,2483 = = 0,4181	0,9967 – 0,6664 = = 0,3303
np_i	9,932	16,724	13,212

где из условия примера n=40.

Рассчитываем величину критерия χ^2 по формуле (3.28):

$$\chi^2 = \frac{(8-9{,}932)^2}{9{,}932} + \frac{(23-16{,}724)^2}{16{,}724} + \frac{(9-13{,}212)^2}{13{,}212} =$$

$$= 0{,}376 + 2{,}3552 + 1{,}3428 = 4{,}074.$$

Для определения критической области критерия выберем уровень доверия α. Примем $\alpha = 0{,}05$. По таблице критерия χ^2 Пирсона для числа степеней свободы $k = 2$ имеем критическое значение равным 5,99. Поскольку расчетное значение критерия меньше критического значения, гипотеза принимается.

Значения квантилей χ^2_α распределения χ^2

k	Вероятность $P(\chi^2 > \chi^2_\alpha)$									
	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	$3{,}9 \cdot 10^{-5}$	$1{,}6 \cdot 10^{-4}$	$0{,}98 \cdot 10^{-4}$	0,0039	0,016	2,71	3,48	5,02	6,64	7,88
2	0,010	0,020	0,051	0,103	0,211	4,60	5,99	7,38	9,21	10,6
3	0,072	0,115	0,216	0,352	0,584	6,25	7,82	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,8
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,6
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,74	2,09	2,70	3,32	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,86	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,58	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,6	21,0	23,3	26,2	28,3
13	3,56	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,08	4,66	5,63	6,57	7,79	21,1	23,7	26,1	29,1	31,3
16	5,14	5,81	6,91	7,96	9,31	23,5	26,3	28,9	32,0	34,3
18	6,26	7,02	8,23	9,39	10,9	26,0	28,9	31,5	34,8	37,2
20	7,43	8,26	9,59	10,9	12,4	28,4	31,4	34,2	37,6	40,0
22	8,64	9,54	11,0	12,3	14,0	30,8	33,9	36,8	40,3	42,8
24	9,89	10,9	12,4	13,8	15,7	33,2	36,4	39,4	43,0	45,6
26	11,2	12,2	13,8	15,4	17,3	35,6	38,9	41,9	45,6	48,3
28	12,5	13,6	15,3	16,9	18,9	37,9	41,3	44,5	48,3	51,0
30	13,8	15,0	16,8	18,5	20,6	40,3	43,8	47,0	50,9	53,7