Математическая статистика

Математическая статистика — наука о математических методах систематизации и использования статистических данных для решения научных и практических задач.

Математическая статистика тесно примыкает к теории вероятностей и базируется на ее понятиях.

Однако главным в математической статистике является не распределение случайных величин, а анализ статистических данных и выяснение, какому распределению они соответствуют.

Предположим, что необходимо изучить множество объектов по какому-либо признаку.

Большая совокупность объектов для исследования, называется ***генеральной совокупностью.***

Для генеральной совокупности можно определить ***генеральную среднюю*** — среднее арифметическое значение всех величин, составляющих эту совокупность.

Учитывая большой объем этой совокупности, можно полагать, что генеральная средняя равна математическому ожиданию.

$$\overline{x}_{\Gamma} = \mathrm{M(X)}$$

Рассеяние значений изучаемого признака генеральной совокупности от их генеральной средней оценивают ***генеральной дисперсией***, или ***генеральным средним квадратическим отклонением***.

Рассеяние значений изучаемого признака генеральной совокупности от их генеральной средней оценивают ***генеральной дисперсией***, или ***генеральным средним квадратическим отклонением***.

$$\bar{x}_{\Gamma} = \mathrm{M(X)} = x_1 \frac{n_1}{N} + x_2 \frac{n_2}{N} + \cdots + x_k \frac{n_k}{N},$$

$$D_{\Gamma}(\mathrm{X}) = {x_1}^2 \frac{n_1}{N} + {x_2}^2 \frac{n_2}{N} + \cdots + {x_k}^2 \frac{n_k}{N} - [\bar{x}_{\Gamma}]^2,$$

$$\sigma_{\Gamma}(x) = \sqrt{D_{\Gamma}(x)}$$

где *N* — объем генеральной совокупности.

Часто возникает ситуация, при которой изучить всю генеральную совокупность практически невозможно.

Тогда изучают не всю генеральную совокупность, а только ее часть и по полученным результатам делают вывод о всей генеральной совокупности и ее числовых характеристиках.

Множество объектов, отобранные из ***генеральной совокупности***, называются ***выборкой,*** или ***выборочной совокупностью***.

Свойство объектов выборки должно соответствовать свойству объектов генеральной совокупности, или, как принято говорить, выборка должна быть представительной *(**репрезентативной**).*

Для дальнейшего изучения значений случайной величины служат числовые характеристики выборки.

Эти характеристики вычисляются по статистическим данным, т.е. данным, полученным в результате наблюдений, поэтому их называют статистическими.

Основные характеристики, которые находятся в теории вероятности с помощью вероятности ***p***, в статистике находятся с помощью ***относительной частоты*** $w = \frac{n_i}{n}$ по тем же формулам.

Нам в лабораторной работе понадобятся:

математическое ожидание $M(x)$,

дисперсия $D(x)$ и

среднее квадратичное отклонение $\sigma(x)$.

Для дискретной случайной величины X, принимающей конечное число значений, законом распределения в теории вероятностей считается таблица.

Необходимые нам характеристики находятся по следующим формулам

$$M(x) = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k,$$

$$D(x) = M(x^2) - [M(x)]^2 = {x_1}^2 p_1 + {x_2}^2 p_2 + \cdots + {x_k}^2 p_k - [M(x)]^2,$$

$$\sigma(x) = \sqrt{D(x)}.$$

В статистике, представленные данные измерений составляют похожую таблицу

(статистический закон распределения).

Здесь n_i показывает сколько раз встретилось в выборке значение x_i,

а N дает общее количество данных – объем выборки.

Для выборки в качестве основных характеристик выступают:

выборочное среднее $\overline{x}_{\text{в}}$ (аналог математического ожидания), выборочная дисперсия $D_{\text{в}}(x)$ (аналог дисперсии) и выборочное среднее квадратичное отклонение

$$\sigma_{\text{в}}(x) = \sqrt{D_{\text{в}}(x)}.$$

Находятся они по тем же формулам, что и их аналоги, но, как мы уже упоминали, вероятность меняется на ее приближенное значение – относительную частоту.

$$\bar{x}_{\text{в}} = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \cdots + x_k \frac{n_k}{n},$$

$$D_{\text{в}}(x) = {x_1}^2 \frac{n_1}{n} + {x_2}^2 \frac{n_2}{n} + \cdots + {x_k}^2 \frac{n_k}{n} - [\bar{x}_{\text{в}}]^2,$$

$$\sigma_{\text{в}}(x) = \sqrt{D_{\text{в}}(x)}.$$

Указанные характеристики дают хорошее приближение для *математического ожидания* и *дисперсии* изучаемой случайной величины.

Однако ***дисперсия*** имеет небольшую, но постоянную ошибку (говорят является смещенной оценкой).

Чтобы этого избежать, используют следующие исправленные оценки

$$D_x = \frac{n}{n-1} D_{\text{в}};$$

$$s_x = \sqrt{D_{\text{и}}}.$$

Эти формулы называют точечными оценками числовых характеристик.

Задания к лабораторной работе

В лесу проведена рубка деревьев на пробном участке и проведены измерения диаметра при основании (X см) и длины этих деревьев (Y м).

Приводятся результаты 100 измерений двумерной случайной величины ($X;Y$).

I) Для каждой из двух случайных величин X и Y провести статистический анализ:

Построить интервальные и вариационные ряды;

Построить гистограмму и полигон относительных частот;

Найти числовые характеристики выборки;

Проверить с помощью критерия Пирсона гипотезу о нормальном распределении генеральной совокупности при $\alpha = 0{,}05$;

Построить гистограмму и выровненную нормальную кривую;

Найти интервальные оценки нормального распределения.

II) Для двумерной случайной величины (*X*;*Y*) провести корреляционный анализ:

Составить интервальную и вариационную корреляционные таблицы;

Найти коэффициент корреляции r_B;

Найти функции регрессии *X* на *Y* и *Y* на *X;*

Построить корреляционное поле и графики функций регрессии.

Пусть мы имеем следующий набор данных.

Отметим, что

величина ***x*** представляет измерения диаметра древесного ствола у случайно выбранных деревьев, а

величина ***y*** измерения высоты того же дерева в метрах.

x	y	x	y	x	y	x	y	x	y
31,4	7,64	37,6	8,42	37,7	8,84	38	8,94	32,6	6,83
18,3	3,68	29,8	7,20	21,9	5,37	29,1	6,16	25,5	5,75
33,7	7,90	32,7	7,23	34	8,76	31,5	6,56	35,8	9,61
28,9	7,46	27,5	5,46	29,9	5,71	22,6	6,36	27,7	6,04
36,9	8,78	35,9	10,32	33	9,31	36,1	9,74	30,3	6,36
29,1	5,90	18,6	4,22	18,7	4,79	29,9	7,69	16,9	2,19
30	7,40	44	10,00	38,7	10,97	33,7	8,12	37,7	10,18
26,2	7,06	21,2	5,59	20,4	3,62	26,2	5,12	28	7,87
35,1	7,76	33,2	8,41	34,2	9,18	34,2	9,91	30,4	5,85
21,3	6,04	20,8	4,38	29,9	8,58	26,3	6,82	27,5	6,93
30,2	7,31	43,8	11,40	31,4	6,89	36,1	8,27	31	8,33
26,2	5,55	17	4,25	18,2	3,01	27,9	6,79	19,2	3,43
30	7,55	34,7	9,77	32,5	7,70	30	7,59	31,1	7,00
28,7	5,28	21,8	4,22	27,9	5,88	23,7	4,91	20,9	3,09
34,3	7,08	33,6	8,21	30,7	8,31	39,2	10,87	33,4	8,34
29	7,45	21,3	6,27	26,7	5,46	21,3	5,89	27,3	6,72
40,3	11,0	33,6	8,21	31,9	7,84	33,3	7,65	30,8	6,15
14,2	3,08	13	1,32	24,5	4,06	27,2	4,90	28,8	6,04
40,6	10,7	42,7	10,78	31,8	7,54	33,2	7,78	35	8,52
24,6	6,73	21	4,26	23,3	4,48	28,7	8,10	26,3	6,43

Данные измерений очевидно содержат ошибки.

Сюда входят как ошибки измеряющих, так и вытекающие из ограниченной точности приборов, используемых для измерений.

Мы разбиваем имеющиеся данные на несколько интервалов и на каждом интервале усредняем имеющиеся значения.

Это позволяет частично сгладить возможные ошибки. Процесс усреднения заключается в том, что мы рассматриваем сколько данных попало на интервал, но конкретные данные заменяем на их среднее.

Для определения оптимальной длины интервала используем формулу Стерджесса

$$h = \frac{x_{max} - x_{min}}{1 + 3.32 Lg(n)}.$$

где x_{max} , x_{min} – соответственно максимальное и минимальное значения выборки (представленных данных), а n объем выборки (общее количество данных).

В нашем случае $n = 100$ и $h \cong 4{,}1$.

Минимальное значение x равно 13, прибавляем $h \cong 4{,}1$, получаем первый интервал $(13; 17{,}1)$.

В нашей выборке четыре данных попадают в этот интервал. Продолжая процесс, получаем следующую таблицу (***интервальная таблица распределения частот***) .

	4	10	10	22	27	17	7	3

Завершая усреднение, выберем представителем каждого интервала его середину

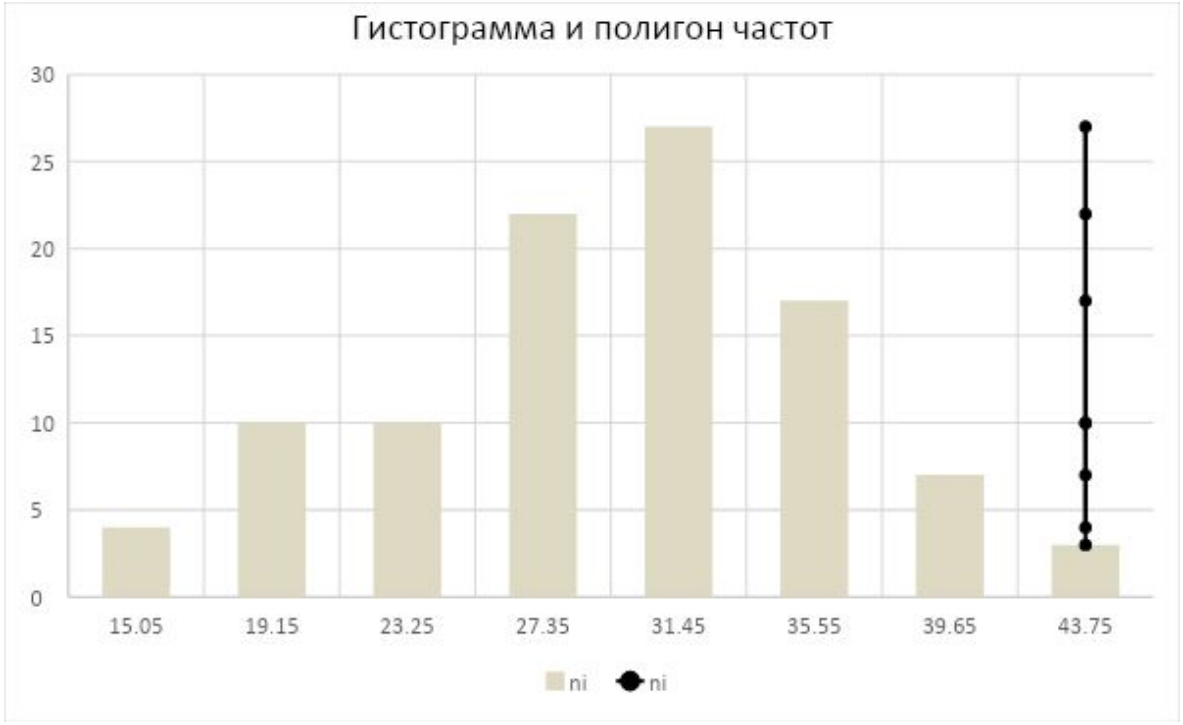
$$x_i' = \frac{x_i + x_{i+1}}{2}$$

и будем считать, что она встретилась столько раз, сколько данных попало на данный интервал.

В результате получим следующий статистический закон распределения (***дискретная таблица распределения частот***)

	15,05	19,15	23,25	27,35	31,45	35,55	39,65	43,75
	4	10	10	22	27	17	7	3

По интервальной таблице распределения частот построим гистограмму и полигон частот.



Вычисление основных характеристик выборки.

Чтобы завершить предварительную обработку данных, по формулам, указанным выше, найдем точечные оценки числовых характеристик

$$\bar{x}_{\text{в}} = x_1\frac{n_1}{n} + x_2\frac{n_2}{n} + \cdots + x_k\frac{n_k}{n} = 15{,}05\frac{4}{100} + 19{,}15\frac{10}{100} + \cdots \cong 29{,}48,$$

$$D_{\text{в}} = {x_1}^2\frac{n_1}{n} + {x_2}^2\frac{n_2}{n} + \cdots + {x_k}^2\frac{n_k}{n} - [\bar{x}_{\text{в}}]^2 \cong 44{,}54,$$

$$\sigma_{\text{в}}(x) = \sqrt{D_{\text{в}}(x)} \cong 6{,}67,$$

$$D_x = \frac{n}{n-1}D_{\text{в}} = \frac{100}{99}\cdot 44{,}54 = 44{,}98;$$

$$s_x = \sqrt{D_{\text{в}}} = \sqrt{44{,}98} = 6{,}71$$

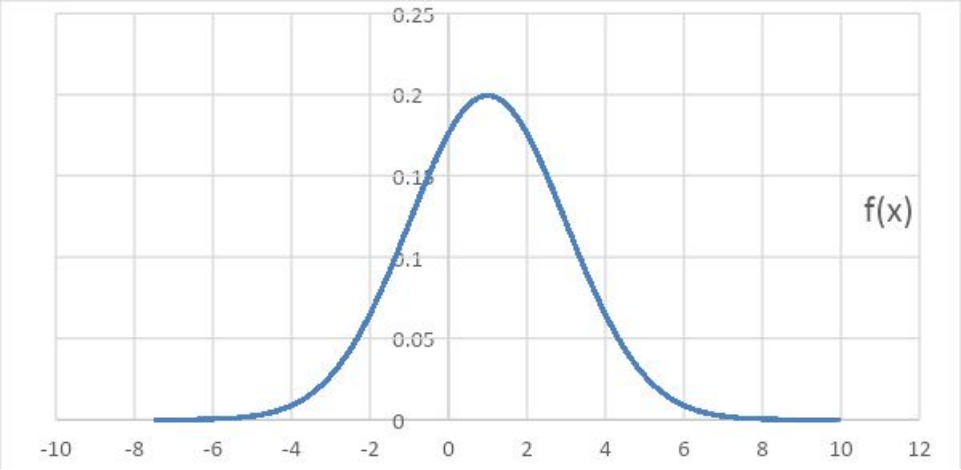
Проверка гипотезы о нормальном распределении.

Закон распределения непрерывной случайной величины называется нормальным, если функция плотности задается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Причем параметр $\boldsymbol{a}$ является математическим ожиданием случайной величины, а параметр $\boldsymbol{\sigma}$ средним квадратичным отклонением. График функции плотности имеет следующий вид.

Здесь $\boldsymbol{a = 1}$ и $\boldsymbol{\sigma = 2}$

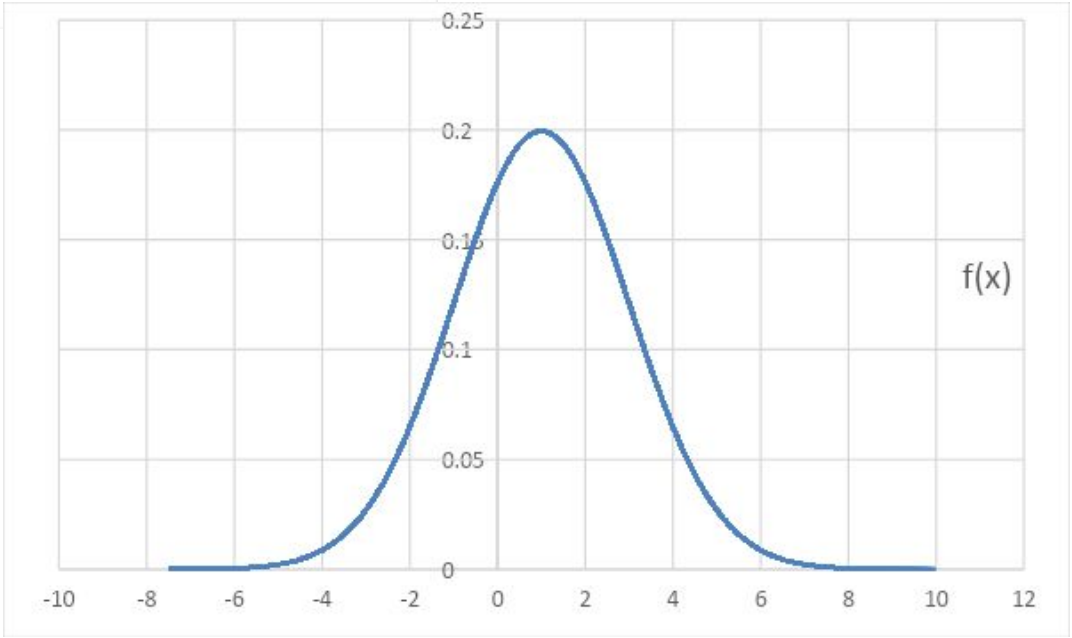


По интервальной таблице распределения частот мы построили гистограмму и полигон частот.

Если получившаяся конструкция похожа на «шапочку» как у графика плотности нормального закона мы выдвинем гипотезу о нормальном законе изучаемой случайной величины и, в дальнейшем, проверим это.

Если не похожа, то мы, используя рассмотренный далее критерий, убедимся в необоснованности предположения о нормальном распределении.


Гистограмма и полигон частот
30
25
20
15
10
5
0
15.05
19.15
23.25
27.35
31.45
35.55
39.65
43.75
ni
ni


0.25
0.2
0.15
0.1
0.05
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
f(x)

Выдвигаем гипотезу о нормальном распределении изучаемой случайной величины. Причем параметры a и σ считаем равными

$$a = \bar{x}_{\text{в}}, \qquad \sigma = s_x.$$

Функция плотности нормального распределения, соответственно, задается следующей формулой

$$f(x) = \frac{1}{s_x\sqrt{2\pi}} e^{-\frac{(x-\bar{x}_{\text{в}})^2}{2{s_x}^2}} =$$

$$= \frac{1}{6{,}7 \cdot \sqrt{2\pi}} e^{-\frac{(x-29{,}48)^2}{89{,}98}}.$$

Используя функцию плотности, задающую предполагаемый закон распределения, мы можем найти теоретические частоты, то есть сколько должно появляться данных на каждом интервале, если закон распределения задается принятой нами формулой.

Теоретические частоты n_i' находятся для каждого интервала по следующей формуле.

$$n_i' = \frac{nh}{s_x} \cdot \varphi\left(\frac{x_i - \bar{x}_в}{s_x}\right),$$

где

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

После подсчетов получаем таблицу.

	15,05	19,15	23,25	27,35	31,45	35,55	39,65	43,75
	4	10	10	22	27	17	7	3
	2,41	7,44	15,84	23,18	23,36	16,2	7,73	2,54

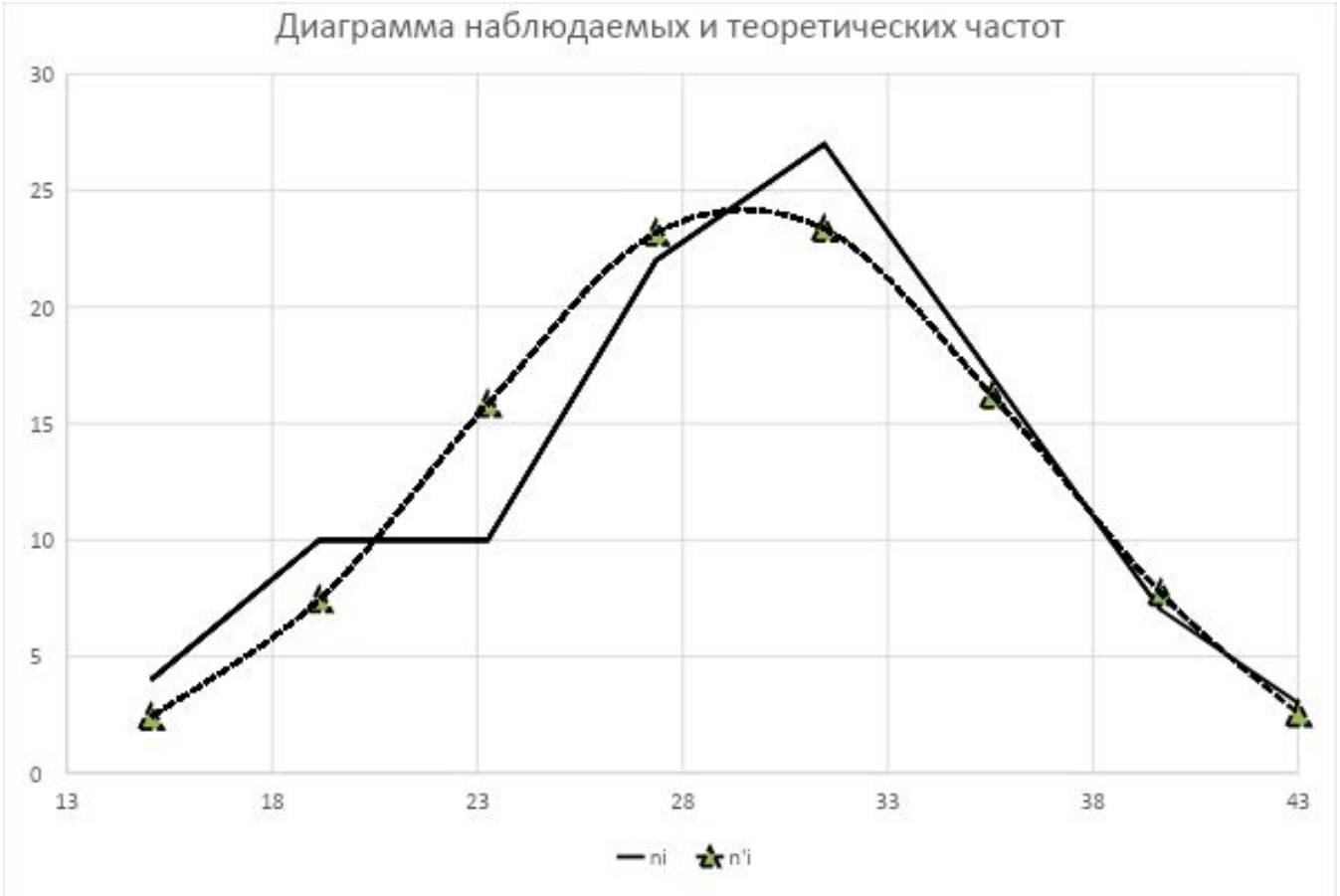


Диаграмма наблюдаемых и теоретических частот
30
25
20
15
10
5
0
13
18
23
28
33
38
43
ni
n'i

Проверка гипотезы о нормальном распределении будет проводится с помощью критерия Пирсона.

Этот метод оценивает суммарную погрешность между наблюдаемыми частотами n_i и теоретическими частотами n_i'.

Если суммарная погрешность больше критического значения, то гипотеза отвергается. В ином случае, гипотеза принимается.

Точная формула по которой находится значение критерия Пирсона.

$$X^2_{\text{набл.}} = \sum_{i=1}^{i=8} \frac{\left(n_i - n_i'\right)^2}{n_i'}.$$

Результаты вычислений удобно представлять в виде таблицы.

	15,05	19,15	23,25	27,35	31,45	35,55	39,65	43,75
	4	10	10	22	27	17	7	3
	2,41	7,44	15,84	23,18	23,36	16,2	7,73	2,54

i					
1	4	2,409019	1,590981	2,531221	1,050727
2	10	7,445548	2,554452	6,525225	0,876393
3	10	15,83736	-5,83736	34,07477	2,151544
4	22	23,18454	-1,18454	1,403144	0,060521
5	27	23,35845	3,641553	13,26091	0,567714
6	17	16,19642	0,803581	0,645742	0,039869
7	7	7,729007	-0,72901	0,531451	0,068761
8	3	2,538388	0,461612	0,213085	0,083945
		98,69873			

Критическое значение критерия $X^2_{\text{кр.}}$ находится из таблицы с таким же названием и зависит от двух параметров.

Первый параметр s называется «число степеней свободы» и находится из равенства $s = k - 3$, где k это количество рассматриваемых интервалов. В нашем случае $k = 8$ и $s = 5$.

Второй параметр $\propto$ называется уровень значимости (или критерий значимости). Этот параметр показывает вероятность отвергнуть правильную гипотезу.

По условию задачи: Проверить с помощью критерия Пирсона гипотезу о нормальном распределении генеральной совокупности при $\alpha = 0{,}05$.

Для заданных параметров

$$\boldsymbol{X^2_{\text{кр.}} = 11,1}.$$

Так как $$X^2_{\text{набл.}} = 4{,}899 < X^2_{\text{кр.}} = 11.1,$$

то у нас нет оснований отвергнуть гипотезу.

Гипотеза принимается.

Отыскание интервальных оценок параметров нормального распределения.

Найдём интервальные оценки математического ожидания и среднего квадратического отклонения генеральной совокупности X.

Для математического ожидания

$$\bar{x}_в - t_\gamma \cdot \frac{s_x}{\sqrt{n}} < a < \bar{x}_в + t_\gamma \cdot \frac{s_x}{\sqrt{n}}$$

где n—объём выборки,

$\bar{x}_в$ —выборочное среднее,

s_x —исправленное среднее квадратическое отклонение.

t_γ— находится по заданной надёжности α (доверительной вероятности) и объёму выборки n по приложению 3 (см. Гмурман В.Е.). Приняв за надёжность 0,95 (в соответствии с заданием $\gamma = 1 - \alpha$), n = 100, получаем: 1,984.

Тогда $28.15 < a < 30.81$

Сделаем окончательный вывод.

Проведенные исследования показали, что генеральная совокупность случайной величины *x*, выражающей диаметры ствола деревьев изучаемого лесного массива из которой взята выборка, распределена по нормальному закону, плотность вероятности которого

$$f(x) = \frac{1}{6{,}7 \cdot \sqrt{2\pi}} e^{-\frac{(x-29{,}48)^2}{89{,}98}}.$$

Среднее значение диаметра ствола деревьев у основания составляет 29,48 сантиметра, причем с вероятностью 0,95 оно лежит в интервале (28,15 ; 30,81),

Среднее отклонение $s_x = 6{,}7$ сантиметра.

В основном (68%) значения диаметра ствола деревьев лежат в интервале (22,78 ; 36,18) сантиметра.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Вернемся к исходным данным - двумерной случайной величине (*X*;*Y*).

x	y	x	y	x	y	x	y
31,4	7,64	37,6	8,42	37,7	8,84	38	8,94
18,3	3,68	29,8	7,20	21,9	5,37	29,1	6,16
33,7	7,90	32,7	7,23	34	8,76	31,5	6,56
28,9	7,46	27,5	5,46	29,9	5,71	22,6	6,36
36,9	8,78	35,9	10,32	33	9,31	36,1	9,74
29,1	5,90	18,6	4,22	18,7	4,79	29,9	7,69
30	7,40	44	10,00	38,7	10,97	33,7	8,12
26,2	7,06	21,2	5,59	20,4	3,62	26,2	5,12
35,1	7,76	33,2	8,41	34,2	9,18	34,2	9,91
21,3	6,04	20,8	4,38	29,9	8,58	26,3	6,82
30,2	7,31	43,8	11,40	31,4	6,89	36,1	8,27
26,2	5,55	17	4,25	18,2	3,01	27,9	6,79
30	7,55	34,7	9,77	32,5	7,70	30	7,59
28,7	5,28	21,8	4,22	27,9	5,88	23,7	4,91
34,3	7,08	33,6	8,21	30,7	8,31	39,2	10,87
29	7,45	21,3	6,27	26,7	5,46	21,3	5,89
40,3	11,09	33,6	8,21	31,9	7,84	33,3	7,65
14,2	3,08	13	1,32	24,5	4,06	27,2	4,90
40,6	10,76	42,7	10,78	31,8	7,54	33,2	7,78
24,6	6,73	21	4,26	23,3	4,48	28,7	8,10
27,3	6,72	31	8,33	16,9	2,19	32,6	6,83
30,8	6,15	19,2	3,43	37,7	10,18	25,5	5,75
28,8	6,04	31,1	7,00	28	7,87	35,8	9,61
35	8,52	20,9	3,09	30,4	5,85	27,7	6,04
26,3	6,43	33,4	8,34	27,5	6,93	30,3	6,36

Для каждой из двух случайных величин ***X*** и ***Y*** отдельно мы построили интервальные и вариационные ряды, нашли числовые характеристики.

X. инт.	13-17	17-21	21-25	25-29	29-33	33-37	37-41	41-45
Ni	3	9	12	19	25	21	8	3

X ср	15	19	23	27	31	35	39	43
Ni	3	9	12	19	25	21	8	3

$$\bar{x}_{\text{в}} \cong 29{,}56, \quad D_x = 42{,}51; \quad s_x = 6{,}52$$

Y инт.	1,32-2,62	2,62-3,92	3,92-5,22	5,22-6,52	6,52-7,82	7,82-9,12	9,12-10,42	10,42-11,72
Ni	2	6	11	21	26	19	9	6

Y ср	1,97	3,27	4,57	5,87	7,17	8,47	9,77	11,07
Ni	2	6	11	21	26	19	9	6

$$\bar{y}_{\text{в}} = \cong 6{,}99, \qquad D_x = \frac{n}{n-1} D_{\text{в}} = 4{,}40, \qquad s_y = \sqrt{D_{\text{в}}} = 2{,}1$$

Построим ***интервальную корреляционную таблицу***.

По горизонтали отложим интервалы *X,* а по вертикали по *Y.* В ячейки таблицы запишем количество пар (*X*;*Y*) , для которых *X* попадает в интервал (X_i;X_{i+1}), а *Y* в интервал (Y_i;Y_{i+1}).

Например, первая пара (31,4; 7,64) по *X* попадает в интервал (29; 33), а по *Y* в интервал (6,52; 7,82).

Для контроля просуммируем строки и столбцы. Должны получиться интервальные ряды для каждой из двух случайных величин *X* и *Y* отдельно.

Получаем интервальную корреляционную таблицу.

Y\X	13-17	17-21	21-25	25-29	29-33	33-37	37-41	41-45	
1,32-2,62	2								2
2,62-3,92	1	5							6
3,92-5,22		4	5	2					11
5,22-6,52			6	9	6				21
6,52-7,82			1	6	16	3			26
7,82-9,12				2	3	12	2		19
9,12-10,42						6	2	1	9
10,42-11,72							4	2	6
	3	9	12	19	25	21	8	3	100

Заменив интервалы на их середины, получим вариационную корреляционную таблицу.

Y\X	15	19	23	27	31	35	39	43	
1,97	2								2
3,27	1	5							6
4,57		4	5	2					11
5,87			6	9	6				21
7,17			1	6	16	3			26
8,47				2	3	12	2		19
9,77						6	2	1	9
11,07							4	2	6
	3	9	12	19	25	21	8	3	100

Рассчитываем показатель тесноты связи между случайными величинами *X* и *Y*. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

$$r_B = \frac{\sum n_{ij} \cdot x_i \cdot y_j - n \cdot \bar{x}_{\text{в}} \cdot \bar{y}_{\text{в}}}{n \cdot s_x \cdot s_y}$$

где n_{ij} – количество пар, попавших, в ячейку с координатами $\left(x_i; y_j\right)$ в вариационной корреляционной таблице.

В нашем примере

$$\sum n_{ij} \cdot x_i \cdot y_j = 2 \cdot 15 \cdot 1{,}97 + 1 \cdot 15 \cdot 3{,}27 + \cdots + 2 \cdot 43 \cdot 11{,}07 = 21878{,}32$$

$$r_B = \frac{21878{,}32 - 100 \cdot 29{,}56 \cdot 6{,}99}{100 \cdot 6{,}52 \cdot 2{,}1} = 0{,}89$$

Линейный коэффициент корреляции принимает значения от –1 до +1.

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются по шкале Чеддока:

0.1< r_B< 0.3: слабая;
0.3 < r_B < 0.5: умеренная;
0.5 < r_B < 0.7: заметная;
0.7 < r_B < 0.9: высокая;
0.9 < r_B < 1: весьма высокая.

Если $|r_B|$ близок к единице, то зависимость мало отличается от линейной. В этом случае связь между величинами достаточно точно описывается линейной функцией.

В нашем примере $r_B = 0{,}89$, поэтому связь между признаком *Y* и фактором *X* высокая и в качестве уравнения регрессии (зависимости), мы можем рассматривать линейную функцию.

Уравнение линейной регрессии *Y* на *X* имеет вид

$$y = \bar{y}_B + r_B \frac{s_y}{s_x}(x - \bar{x}_в)$$

$$y = 6{,}99 + 0{,}89\frac{2{,}1}{6{,}52}(x - 29{,}56) = 0{,}59x - 1{,}48$$

$$y = 0{,}59x - 1{,}48$$

Уравнение линейной регрессии *X* на *Y* имеет вид

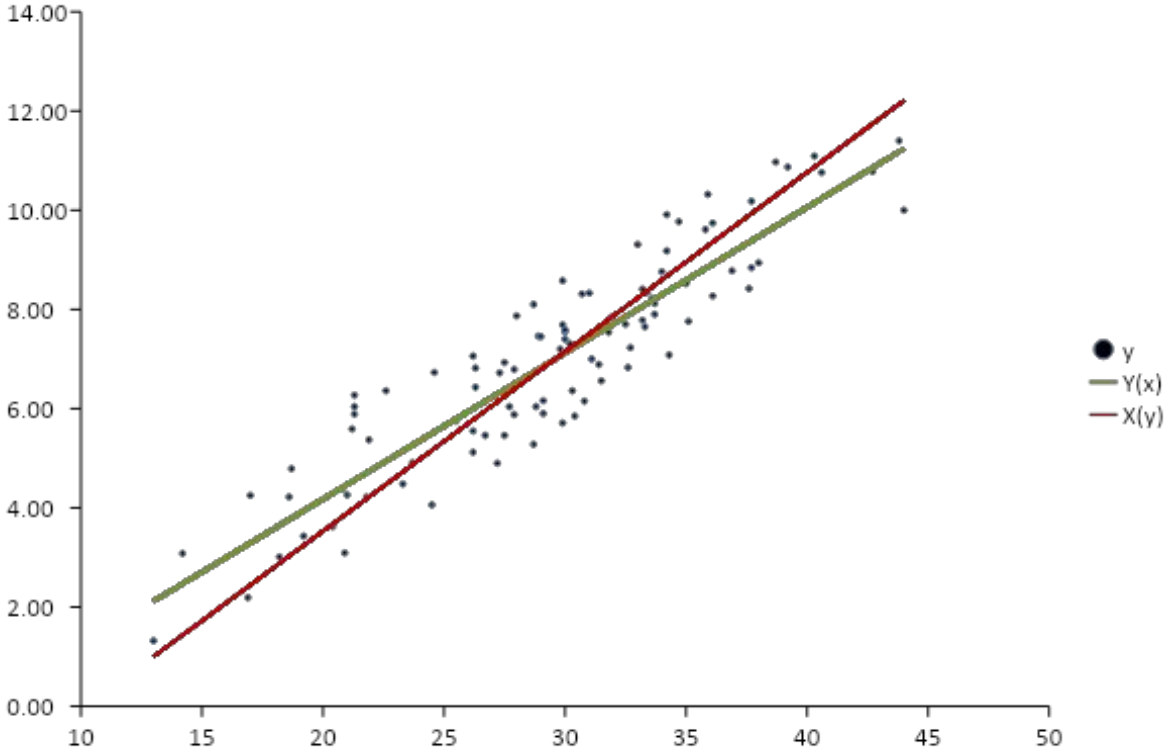
$$x = \bar{x}_B + r_B \frac{s_x}{s_y}\left(y - \bar{y}_{\text{в}}\right)$$

$$y = 29{,}56 + 0{,}89\frac{6{,}52}{2{,}1}(x - 6{,}99) = 2{,}76x + 10{,}24$$

$$y = 2{,}76x + 10{,}24$$

На одних осях координат построим три графика:

1) корреляционное поле (пары (*X*;*Y*) указанные в задании) – на графике изображаются точками,
2) график линейной регрессии *Y* на *X* – прямая Y(x),
3) график линейной регрессии *X* на *Y*– прямая X(y).



Вывод.

Длина ствола дерева Y и диаметр основания X зависят между собой линейно ($y = ax + b$) и связь между величинами *X* и *Y* высокая (результаты измерений лежат в узкой полосе вокруг прямой).

В среднем, длина ствола дерева Y(м) вычисляется через диаметр основания X(см) по формуле

$$y = 0{,}59x - 1{,}48,$$

а диаметр основания X(см) вычисляется через длину ствола дерева Y(м) по формуле

$$y = 2{,}76x + 10{,}24.$$