

Санкт-Петербургский политехнический университет  
Петра Великого Физико-механический институт Высшая  
школа прикладной математики и вычислительной  
физики

Отчёт  
по лабораторным работам №1-8  
по дисциплине  
«Математическая статистика»

Выполнил студент:  
**Басалаев Даниил**  
группа:  
5030102/10201  
Проверил:  
доцент Баженов Александр Николаевич

Санкт-Петербург

# 1 Лабораторные работы №1-2

## 1.1 Постановка задачи

### 1.1.1 Описательная статистика

Для 5 распределений:

- Нормальное распределение  $N(x, 0, 1)$
- распределение Коши  $C(x, 0, 1)$
- Распределение Стьюдента  $t(x, 0, 3)$  с тремя степенями свободы
- Распределение Пуассона  $P(k, 10)$
- Равномерное распределение  $U(x, -\sqrt{3}, \sqrt{3})$

Сгенерировать выборки размером 10, 50, 1000 элементов.

Построить на одном рисунке гистограмму и график плотности распределения.

### 1.1.2 Точечное оценивание характеристик положения и рассеяния

Сгенерировать выборки размером 10, 50, 1000 элементов.

Для каждой выборки вычислить следующие статистические характеристики положения данных:  $\bar{x}$ ,  $med\ x$ ,  $z_Q$ ,  $z_R$ ,  $z_{tr}$ . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:  $E(z) = \bar{z}$ . Вычислить оценку дисперсии по формуле  $D(z) = \bar{z}^2 - \bar{z}^2$ .

## 1.2 Теоретическое обоснование

### 1.3 Функции распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (1)$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad (2)$$

- Распределение Стьюдента  $t(x, 0, 3)$  с тремя степенями свободы

$$t(x, 0, 3) = \frac{6\sqrt{3}}{\pi(3 + t^2)^2} \quad (3)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (4)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при } |x| \leq \sqrt{3} \\ 0 & \text{при } |x| > \sqrt{3} \end{cases} \quad (5)$$

### 1.3.1 Характеристики положения и рассеяния

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

- Выборочная медиана

$$\text{med } x = \begin{cases} x_{(l+1)} & \text{при } n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2} & \text{при } n = 2l \end{cases} \quad (7)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (8)$$

- Полусумма квартилей

Выборочная квартиль  $z_p$  порядка  $p$  определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } np \text{ дробном} \\ x_{(np)} & \text{при } np \text{ целом} \end{cases} \quad (9)$$

Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (10)$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \quad (11)$$

- Среднее характеристики

$$E(z) = \bar{z} \quad (12)$$

- Оценка дисперсии

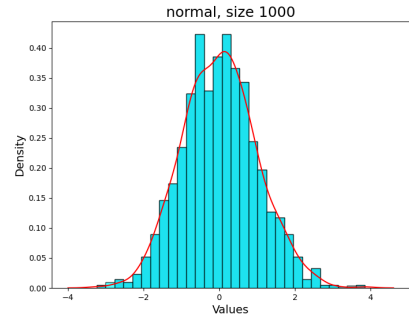
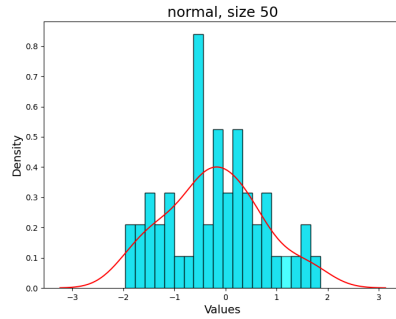
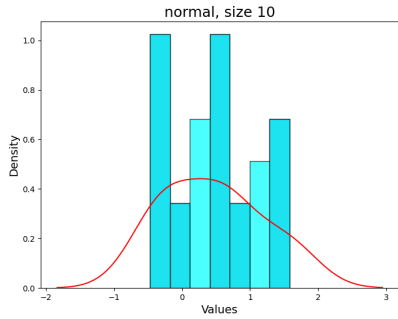
$$D(z) = \overline{z^2} - \bar{z}^2 \quad (13)$$

## 1.4 Ход работы

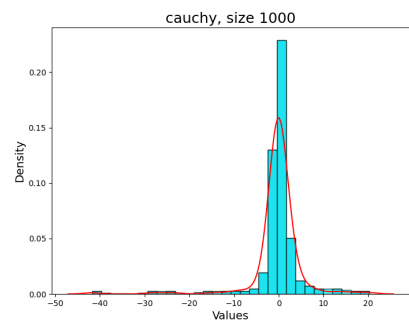
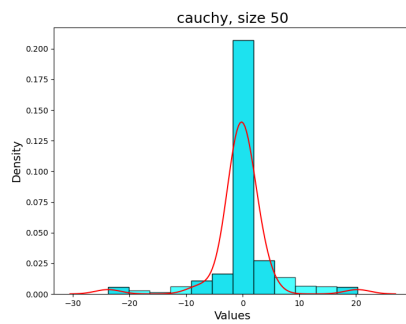
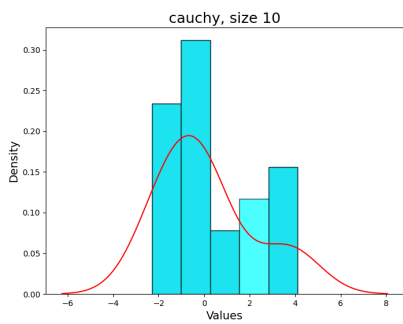
Во время выполнения лабораторной работы все расчеты значений и построения графиков были выполнены на языке программирования Python с использованием таких библиотек, как matplotlib, statistics, pandas, numpy, math.

## 1.5 Графики

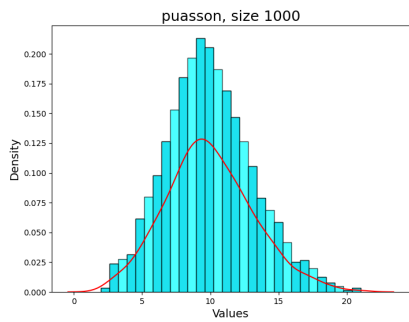
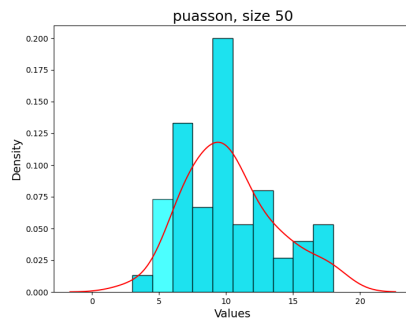
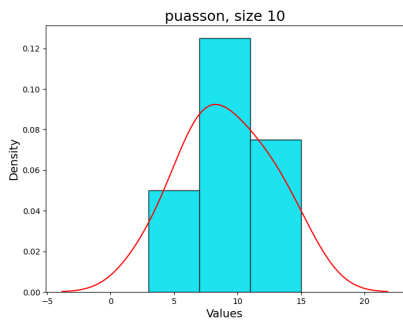
### 1.5.1 Нормальное распределение



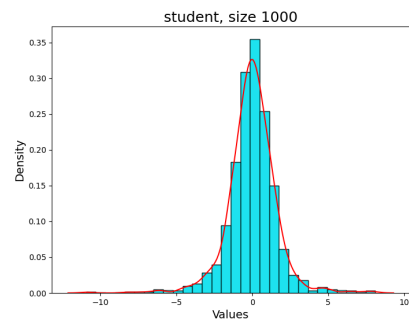
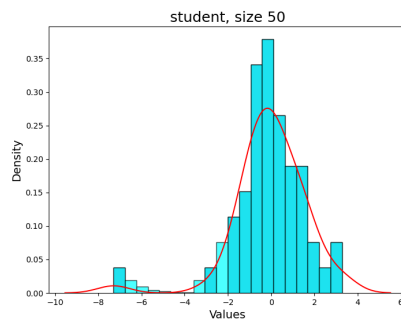
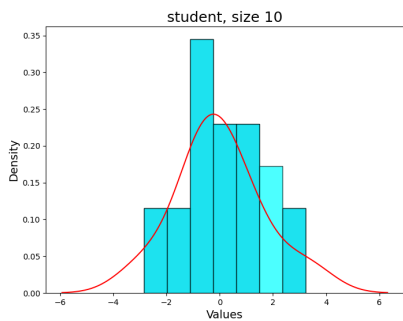
### 1.5.2 Коши



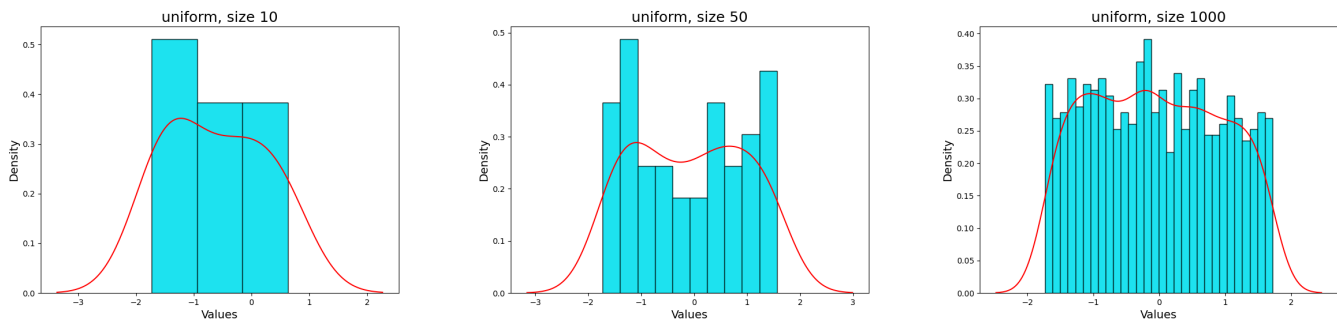
### 1.5.3 Пуассон



### 1.5.4 Стьюдент



## 1.5.5 Равномерное распределение



## 1.6 Таблицы

### 1.6.1 Нормальное распределение

normal n = 10	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0029	-0.0049	0.0071	0.0066	0.0015
$D(z)$	0.101	0.1403	0.3761	0.2337	0.1132
normal n = 50	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.0034	-0.004	-0.0075	-0.0027	-0.005
$D(z)$	0.0199	0.0277	0.2403	0.0491	0.0219
normal n = 1000	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0005	-0.0003	0.0018	-0.0017	-0.0001
$D(z)$	0.001	0.0016	0.125	0.0025	0.0012

### 1.6.2 Коши

cauchy n = 10	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.8588	0.0018	-4.0287	-0.0241	-0.0063
$D(z)$	584.999	0.3453	27944.2716	2.7613	0.5354
cauchy n = 50	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-1.1602	0.0034	-27.9837	-0.0054	-0.0015
$D(z)$	1476.3684	0.0488	1825915.1848	0.2	0.053
cauchy n = 1000	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.2727	-0.0014	161.0178	-0.0064	-0.0023
$D(z)$	1258.889	0.0024	624762602.4426	0.0095	0.0024

### 1.6.3 Пуассон

poisson n = 10	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	9.9945	9.8645	10.279	9.937	9.9002
$D(z)$	100.9016	98.7267	215.26	199.953	99.1721
poisson n = 50	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	9.9754	9.8225	10.74	9.88	9.8427
$D(z)$	99.7091	96.8537	232.936	195.799	97.1256
poisson n = 1000	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	9.9981	9.996	11.6705	9.9955	9.8567
$D(z)$	99.9713	99.924	273.7785	199.8245	97.1652

### 1.6.4 Стьюдент

student n = 10	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0071	-0.0032	0.0172	0.0008	0.0018
$D(z)$	0.2774	0.1671	4.1802	0.3923	0.1526
student n = 50	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0011	0.0043	-0.028	0.0039	0.0025
$D(z)$	0.0583	0.0345	9.5648	0.0759	0.0293
student n = 1000	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0006	-0.0001	-0.0726	-0.0015	0.0006
$D(z)$	0.0031	0.0018	101.8467	0.0037	0.0016

### 1.6.5 Равномерное распределение

uniform n = 10	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.0085	-0.0142	-0.0102	-0.0016	-0.0073
$D(z)$	0.1058	0.2442	0.0979	0.2827	0.1714
uniform n = 50	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0024	0.0017	-0.0015	0.0038	0.0043
$D(z)$	0.0209	0.0587	0.0045	0.0619	0.0399
uniform n = 1000	$MX$	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.0003	0.0002	-0.0001	-0.001	0.0005
$D(z)$	0.001	0.0028	0.0	0.003	0.002

## 1.7 Вывод

В ходе выполнения лабораторной работы были построены гистограммы и графики плотности на выборках разной мощности (10, 50, 1000) для 5 распределений: Коши, нормального, Пуассона, Стьюдента и равномерного.

В ходе проведения лабораторной работы эти распределения также были исследованы. Суть исследования заключалась в нахождении мат. ожидания и дисперсии 5 величин: выборочного среднего  $M[X]$ , выборочной медианы  $med x$ , полусуммы выборочных элементов  $z_R$ , полусуммы квартилей  $z_p$  и усеченного среднего  $z_{tr}$ .

По результатам проведенных исследований можно сделать следующие выводы:

- По полученным графикам легко видеть, что с увеличением мощности выборки характер распределения (форма гистограммы и графика распределения) приближается к теоретическому.
- Для распределения Коши полученные значения исследуемых величин сильно отличаются от теоретических.
- Распределение Стьюдента при малых мощностях выборки дает отличные от теоретических значения, но с ростом мощности они становятся точнее.
- Для нормального, равномерного распределений и распределения Пуассона полученные значения характеристик оказались близки к теоретическим.
- Переходя к оценке поведения исследуемых характеристик, можно сказать, что выборочное среднее является наиболее чувствительным к экстремальным значениям по сравнению с медианой, особенно в выборках малых мощностей. Однако с увеличением мощности выборки влияние этих экстремальных значений на среднее значение уменьшается. В то же время, медиана обычно более устойчива к выбросам и мало варьирует с изменением размера выборки.
- Медиана же является чувствительной к виду распределения: в распределении Коши она дает надежные, устойчивые к выбросам оценки, в нормальном и распределении Стьюдента медиана равна среднему, в Пуассоновском приближается к среднему, и в равномерном равна половине суммы минимального и максимального значений.

## 2 Лабораторные работы №3-4

### 2.1 Постановка задачи

#### 2.1.1 Боксплот Тьюки

Сгенерировать выборки размером 20 и 100 элементов. Построить для них боксплот Тьюки.

#### 2.1.2 Доверительные интервалы для параметров нормального распределения

Сгенерировать выборки размером 20 и 100 элементов. Вычислить параметры положения и рассеяния:

- для нормального распределения,
- для произвольного распределения.

### 2.2 Теоретическое обоснование

#### 2.2.1 Функции распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (14)$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad (15)$$

- Распределение Стюдента  $t(x, 0, 3)$  с тремя степенями свободы

$$t(x, 0, 3) = \frac{6\sqrt{3}}{\pi(3 + t^2)^2} \quad (16)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (17)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} \quad (18)$$

#### 2.2.2 Боксплот Тьюки

Боксплот (англ. box plot) — график, использующихся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выброса). Длину «усов» определяют разность первого квартиля и полутора межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \quad (19)$$

где  $X_1$  — нижняя граница уса,  $X_2$  — верхняя граница уса,  $Q_1$  — первый квартиль,  $Q_3$  — третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков. Выбросами считаются величины, такие что:

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (20)$$



### 2.2.3 Доверительные интервалы для параметров нормального распределения

Пусть  $st_{1-\alpha/2}(n-1)$  — квантиль распределения Стьюдента с  $n-1$  степенями свободы и порядка  $1-\alpha/2$ , где  $\alpha$  — выбранный уровень значимости:  $2F_T(x) - 1 = 1 - \alpha$ . Тогда получаем

$$P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha, \quad (21)$$

что и даст доверительный интервал для  $m$  с доверительной вероятностью  $\gamma = 1 - \alpha$  для нормального распределения. Случайная величина  $n\frac{s^2}{\sigma^2}$  распределена по закону  $\chi^2$  с  $n-1$  степенями свободы. Тогда

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi_{1-\alpha/2}^2(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi_{\alpha/2}^2(n-1)}}\right) = 1 - \alpha, \quad (22)$$

что и даст доверительный интервал для  $\sigma$  с доверительной вероятностью  $\gamma = 1 - \alpha$  для нормального распределения.

### 2.2.4 Доверительные интервалы для параметров произвольного распределения

Пусть  $u_{1-\alpha/2}$  — квантиль нормального распределения порядка  $1-\alpha/2$ , где  $\alpha$  — выбранный уровень значимости:  $2\Phi(x) - 1 = 1 - \alpha$ . Тогда получаем

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} + \frac{su_{1-\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha, \quad (23)$$

что и даст доверительный интервал для  $m$  с доверительной вероятностью  $\gamma = 1 - \alpha$  для произвольного распределения. Тогда для оценки  $\sigma$  справедлива следующая формула

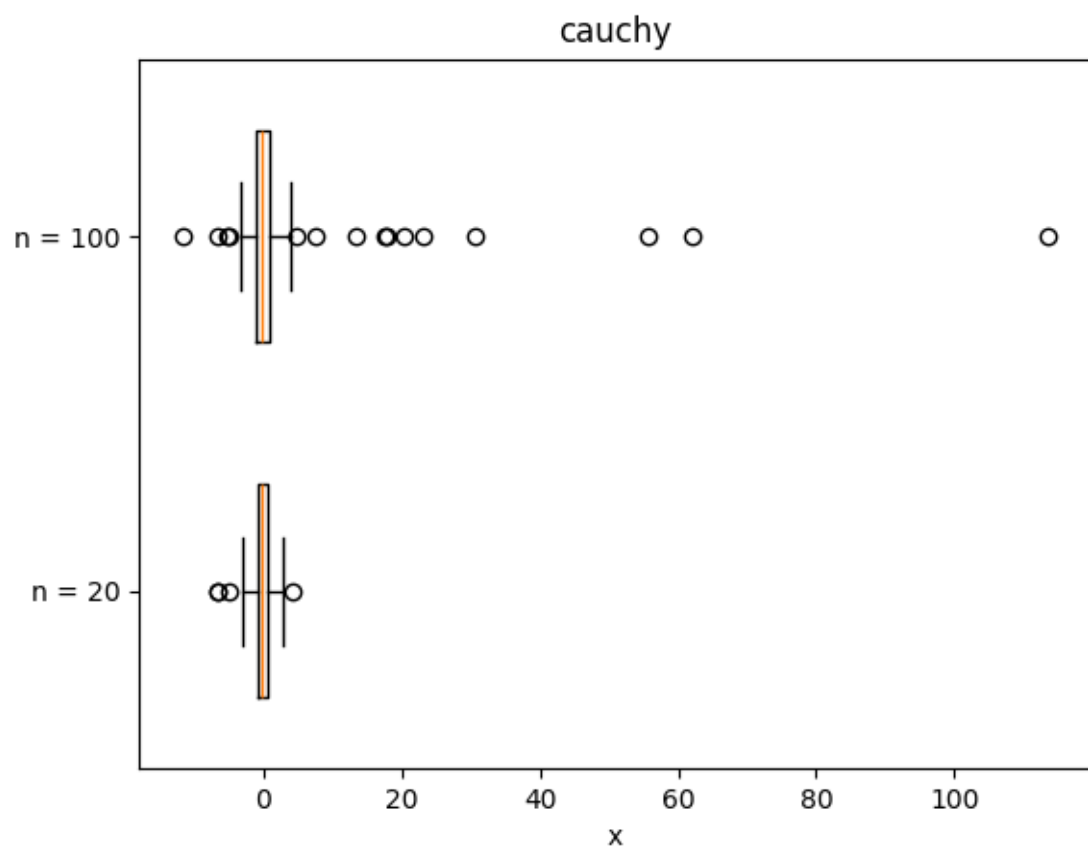
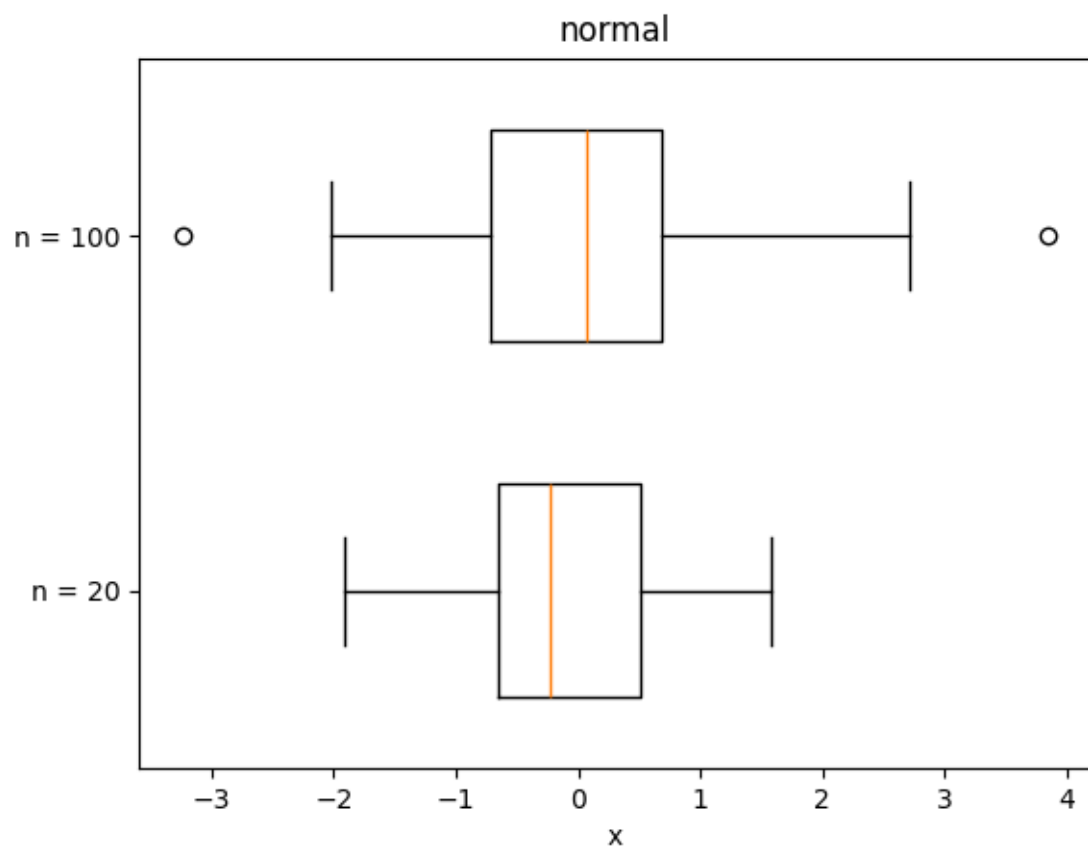
$$P\left(1 - \frac{0.5u_{1-\alpha/2}\sqrt{e+2}}{\sqrt{n}} < \sigma < 1 + \frac{0.5u_{1-\alpha/2}\sqrt{e+2}}{\sqrt{n}}\right) = 1 - \alpha, \quad (24)$$

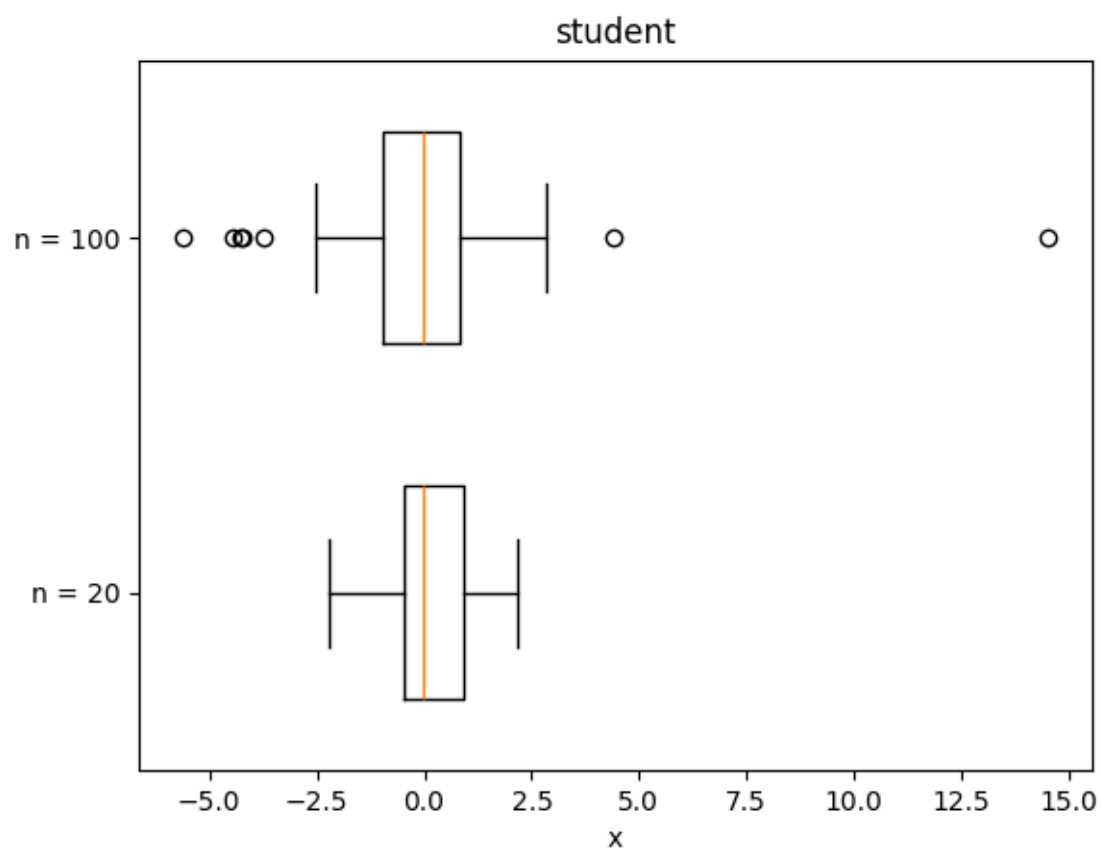
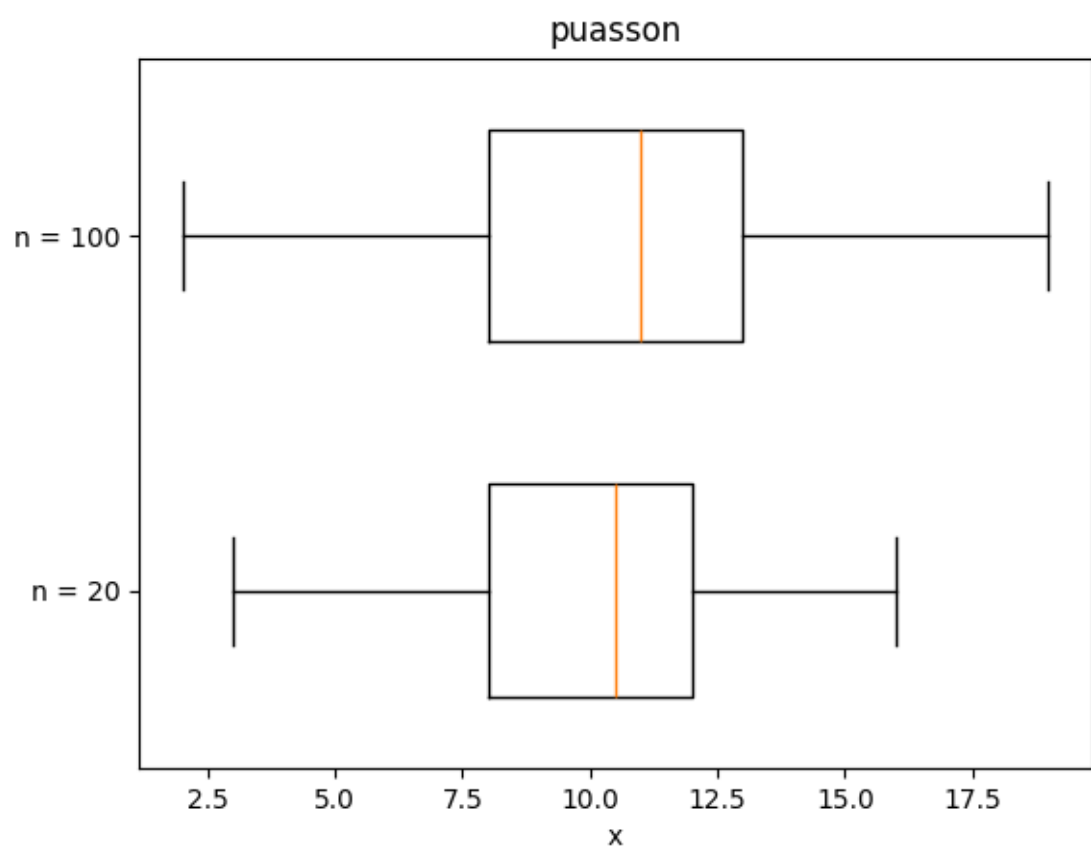
что и даст доверительный интервал для  $\sigma$  с доверительной вероятностью  $\gamma = 1 - \alpha$  для произвольного распределения

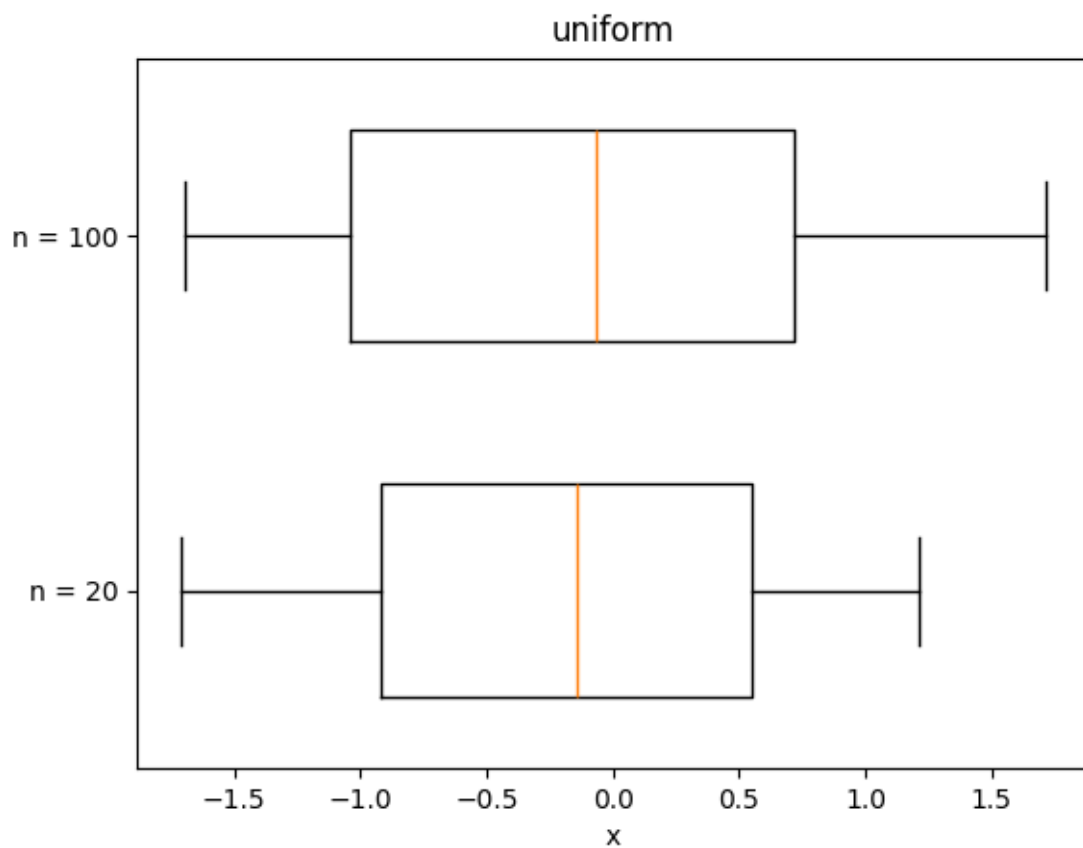
## 2.3 Ход работы

Во время выполнения лабораторной работы все расчеты значений и построения графиков были выполнены на языке программирования Python с использованием таких библиотек, как matplotlib, statistics, pandas, numpy, math.

## 2.4 Бокс-плот Тьюки







## 2.5 Доверительные интервалы для параметров распределений

n = 20	$m$	$\sigma$
	$-0.113 < m < 0.422$	$0.836 < \sigma < 1.237$
n = 100	$m$	$\sigma$
	$-0.157 < m < 0.258$	$0.919 < \sigma < 1.116$

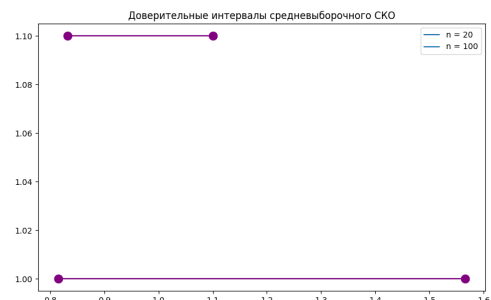
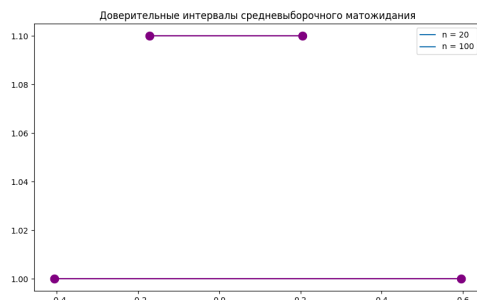
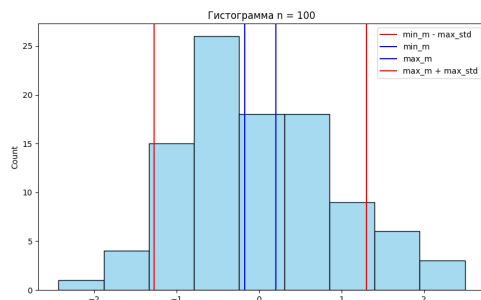
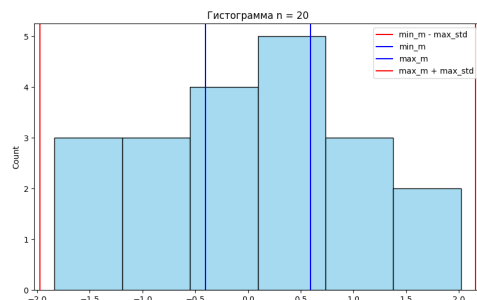
Таблица 1: Доверительные интервалы для параметров нормального распределения (14)

n = 20	$m$	$\sigma$
	$7.576 < m < 10.624$	$2.212 < \sigma < 4.568$
n = 100	$m$	$\sigma$
	$9.587 < m < 10.813$	$2.663 < \sigma < 3.559$

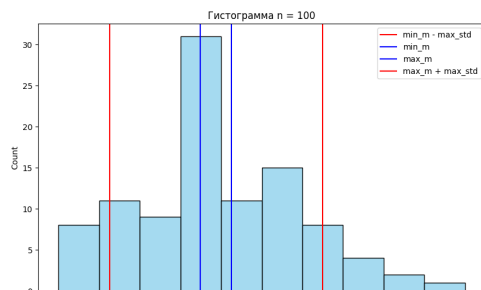
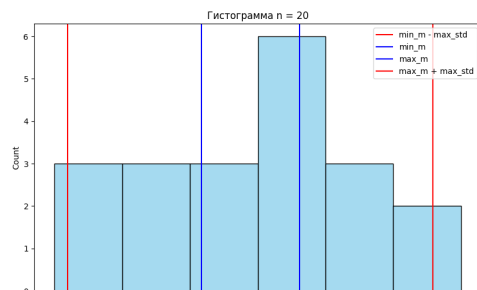
Таблица 2: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

## 2.6 Гистограммы

### 2.6.1 Гистограммы и оценки для параметров нормального распределения



### 2.6.2 Гистограммы и оценки для параметров произвольного распределения



## 2.7 Выводы

По результатам выполнения лабораторной работы были сгенерированы выборки размером 20 и 100 элементов и построены для них боксплоты Тьюки.

Боксплот позволяет наглядно представить основные характеристики выборки - медиану, квартили, межквартильный размах и выбросы. На основе построенных графиков можно увидеть разницу в распределении данных для двух выборок.

Для выборки размером в 100 элементов представленные метрики имеют более проработанный вид, ведь с увеличением размера выборки улучшается точность оценок параметров распределения.

Также в ходе выполнения лабораторной работы были сгенерированы две выборки размерами 20 и 100 элементов для нормального и произвольного распределения. Затем для каждой из них были вычислены параметры распределения: среднее значение и дисперсия.

Результаты, представленные графически, демонстрируют, что количество элементов в выборке влияет на точность оценок параметров. Более большое количество наблюдений (т.е. 100 элементов) приводит к более точным и стабильным оценкам среднего и дисперсии, как для нормального, так и для произвольного распределения. Для выборки с меньшим количеством элементов (20 элементов) оценки могут сильно варьироваться в зависимости от конкретной выборки, что также наглядно отображено на графиках.

Лабораторная работа иллюстрирует важнейший статистический принцип: точность статистической оценки увеличивается с ростом объема выборки. Результаты этого исследования подчеркивают значимость использования достаточно больших выборок для надежного анализа данных.

## 3 Лабораторные работы №5-6

### 3.1 Постановка задачи

#### 3.1.1 Коэффициент корреляции

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения  $N(x, y, 0, 0, 1, 1, \rho)$ . Коэффициент корреляции  $\rho$  взять равным 0, 0.5, 0.9. Каждая выборка генерируется 1000 раз и для неё вычисляются: среднее значение, среднее значение квадрата и дисперсия коэффициентов корреляции Пирсона, Спирмена и квадрантного коэффициента корреляции. Повторить все вычисления для смеси нормальных распределений:

$$f(x, y) = 0.9 * N(x, y, 0, 0, 1, 1, 0.9) + 0.1 * N(x, y, 0, 0, 10, 10, -0.9).$$

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

#### 3.1.2 Простая линейная регрессия

Найти оценки коэффициентов линейной регрессии  $y_i = a + bx_i + e_i$ , используя 20 точек на отрезке  $[-1.8; 2]$  с равномерным шагом равным 0.2. Ошибку  $e_i$  считать нормально распределённой с параметрами  $(0, 1)$ . В качестве эталонной зависимости взять  $y_i = 2 + 2x_i + e_i$ . При построении оценок коэффициентов использовать два критерия: критерий наименьших квадратов и критерий наименьших модулей. Прodelать то же самое для выборки, у которой в значения  $y_1$  и  $y_{20}$  вносятся возмущения 10 и -10.

### 3.2 Теоретическое обоснование

#### 3.2.1 Двумерное нормальное распределение

Двумерная случайная величина  $(X, Y)$  называется распределённой нормально (или просто нормальной), если её плотность вероятности определена формулой

$$N(x, y, \bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} * \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho \frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} \right] \right\} \quad (25)$$

Компоненты  $X, Y$  двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями  $\bar{x}, \bar{y}$  и средними квадратическими отклонениями  $\sigma_x, \sigma_y$  соответственно.

Параметр  $\rho$  называется коэффициентом корреляции.

#### 3.2.2 Корреляционный момент (ковариация) и коэффициент корреляции

Корреляционный момент, иначе ковариация, двух случайных величин  $X$  и  $Y$ :

$$K = \text{cov}(X, Y) = M[(X - \bar{x})(Y - \bar{y})] \quad (26)$$

Коэффициент корреляции  $\rho$  двух случайных величин  $X$  и  $Y$ :

$$\rho = \frac{K}{\sigma_x\sigma_y} \quad (27)$$

#### 3.2.3 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{K}{s_X s_Y}, \quad (28)$$

где  $K, s_X^2, s_Y^2$  — выборочные ковариации и дисперсии случайных величин  $X$  и  $Y$ .

### 3.2.4 Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n}, \quad (29)$$

где  $n_1, n_2, n_3, n_4$  — количество точек с координатами  $(x_i, y_i)$ , попавшими, соответственно, в I, II, III, IV квадранты декартовой системы с осями  $x' = x - \text{med}x$ ,  $y' = y - \text{med}y$ .

### 3.2.5 Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной  $X$ , через  $u$ , а ранги, соответствующие значениям переменной  $Y$ , — через  $v$ .

Выборочный коэффициент ранговой корреляции Спирмена:

$$r_S = \frac{\frac{1}{n} \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum (u_i - \bar{u})^2 \frac{1}{n} \sum (v_i - \bar{v})^2}}, \quad (30)$$

где  $\bar{u} = \bar{v} = \frac{1+2+\dots+n}{n} = \frac{n+1}{2}$  — среднее значение рангов.

### 3.2.6 Эллипсы рассеивания

Рассмотрим поверхность распределения, изображающую функцию (1). Она имеет вид холма, вершина которого находится над точкой  $(\bar{x}, \bar{y})$ .

В сечении поверхности распределения плоскостями, параллельными оси  $N(x, y, \bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho)$ , получаются кривые, подобные нормальным кривым распределения. В сечении поверхности распределения плоскостями, параллельными плоскости  $xOy$ , получаются эллипсы. Напишем уравнение проекции такого эллипса на плоскость  $xOy$ :

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2\rho \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = \text{const} \quad (31)$$

Уравнение эллипса 31 можно проанализировать обычными методами аналитической геометрии. Применяя их, убеждаемся, что центр эллипса 31 находится в точке с координатами  $(\bar{x}, \bar{y})$ ; что касается направления осей симметрии эллипса, то они составляют с осью  $Ox$  углы, определяемые уравнением

$$\text{tg}(2\alpha) = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \quad (32)$$

Это уравнение дает два значения углов:  $\alpha$  и  $\alpha_1$ , различающиеся на  $\frac{\pi}{2}$ .

Таким образом, ориентация эллипса 31 относительно координатных осей находится в прямой зависимости от коэффициента корреляции  $\rho$  системы  $(X, Y)$ ; если величины не коррелированы (т.е. в данном случае и независимы), то оси симметрии эллипса параллельны координатным осям; в противном случае они составляют с координатными осями некоторый угол.

Пересекая поверхность распределения плоскостями, параллельными плоскости  $xOy$ , и проектируя сечения на плоскость  $xOy$  мы получим целое семейство подобных и одинаково расположенных эллипсов с общим центром  $(\bar{x}, \bar{y})$ . Во всех точках каждого из таких эллипсов плотность распределения  $N(x, y, \bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho)$  постоянна. Поэтому такие эллипсы называются эллипсами равной плотности или, короче эллипсами рассеивания. Общие оси всех эллипсов рассеивания называются главными осями рассеивания.

### 3.2.7 Метод наименьших квадратов

При оценивании параметров регрессионной модели используют различные методы. Один из наиболее распространённых подходов заключается в следующем: вводится мера (критерий) рассогласования отклика и регрессионной функции, и оценки параметров регрессии определяются так, чтобы сделать это рассогласование наименьшим. Достаточно простые расчётные формулы для оценок получают при выборе критерия в виде суммы квадратов отклонений значений отклика от значений регрессионной функции (сумма квадратов остатков):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \rightarrow \min_{\beta_0, \beta_1}. \quad (33)$$

Задача минимизации квадратичного критерия (33) носит название задачи метода наименьших квадратов (МНК), а оценки  $\beta_0, \beta_1$  параметров  $\beta_0, \beta_1$ , реализующие минимум критерия (33), называют МНК-оценками.



### 3.2.8 Метод наименьших модулей

Робастность оценок коэффициентов линейной регрессии (т.е. их устойчивость по отношению к наличию в данных редких, но больших по величине выбросов) может быть обеспечена различными способами. Одним из них является использование метода наименьших модулей вместо метода наименьших квадратов:

$$\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i| \rightarrow \min_{\beta_0, \beta_1} . \quad (34)$$

## 3.3 Результаты

### 3.3.1 Коэффициент корреляции

$n = 20, \rho = 0$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$6.161 * 10^{-3}$	$4.842 * 10^{-3}$	$5.663 * 10^{-4}$
Среднее квадратов	$4.982 * 10^{-2}$	$4.971 * 10^{-2}$	$1.014 * 10^{-1}$
Дисперсия	$4.982 * 10^{-2}$	$4.971 * 10^{-2}$	$1.011 * 10^{-1}$
$n = 20, \rho = 0.5$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$4.919 * 10^{-1}$	$4.642 * 10^{-1}$	$4.664 * 10^{-1}$
Среднее квадратов	$2.717 * 10^{-1}$	$2.488 * 10^{-1}$	$3.148 * 10^{-1}$
Дисперсия	$2.994 * 10^{-2}$	$3.327 * 10^{-2}$	$9.723 * 10^{-2}$
$n = 20, \rho = 0.9$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$8.953 * 10^{-1}$	$8.659 * 10^{-1}$	$9.774 * 10^{-1}$
Среднее квадратов	$8.040 * 10^{-1}$	$7.543 * 10^{-1}$	$1.026 * 10^0$
Дисперсия	$2.454 * 10^{-3}$	$5.210 * 10^{-3}$	$5.804 * 10^{-2}$
$n = 60, \rho = 0$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$6.997 * 10^{-4}$	$1.996 * 10^{-4}$	$2.828 * 10^{-4}$
Среднее квадратов	$1.794 * 10^{-2}$	$1.734 * 10^{-2}$	$3.373 * 10^{-2}$
Дисперсия	$1.794 * 10^{-2}$	$1.734 * 10^{-2}$	$3.373 * 10^{-2}$
$n = 60, \rho = 0.5$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$4.910 * 10^{-1}$	$4.694 * 10^{-1}$	$4.618 * 10^{-1}$
Среднее квадратов	$2.586 * 10^{-1}$	$2.373 * 10^{-1}$	$2.505 * 10^{-1}$
Дисперсия	$1.000 * 10^{-2}$	$1.094 * 10^{-2}$	$3.256 * 10^{-2}$
$n = 60, \rho = 0.9$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$8.979 * 10^{-1}$	$8.810 * 10^{-1}$	$9.937 * 10^{-1}$
Среднее квадратов	$8.069 * 10^{-1}$	$7.774 * 10^{-1}$	$1.004 * 10^0$
Дисперсия	$7.297 * 10^{-4}$	$1.202 * 10^{-3}$	$1.700 * 10^{-2}$
$n = 100, \rho = 0$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$5.121 * 10^{-3}$	$3.510 * 10^{-3}$	$3.677 * 10^{-3}$
Среднее квадратов	$1.043 * 10^{-2}$	$1.036 * 10^{-2}$	$2.078 * 10^{-2}$
Дисперсия	$1.041 * 10^{-2}$	$1.034 * 10^{-2}$	$2.077 * 10^{-2}$
$n = 100, \rho = 0.5$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$5.013 * 10^{-1}$	$4.812 * 10^{-1}$	$4.723 * 10^{-1}$
Среднее квадратов	$2.568 * 10^{-1}$	$2.375 * 10^{-1}$	$2.407 * 10^{-1}$
Дисперсия	$5.481 * 10^{-3}$	$6.013 * 10^{-3}$	$1.762 * 10^{-2}$
$n = 100, \rho = 0.9$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$8.999 * 10^{-1}$	$8.866 * 10^{-1}$	$1.003 * 10^0$
Среднее квадратов	$8.103 * 10^{-1}$	$7.868 * 10^{-1}$	$1.017 * 10^0$
Дисперсия	$4.017 * 10^{-4}$	$6.665 * 10^{-4}$	$1.049 * 10^{-2}$

Таблица 3: Характеристики нормального двумерного распределения

$n = 20$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$-9.160 * 10^{-2}$	$-8.827 * 10^{-2}$	$-9.419 * 10^{-2}$
Среднее квадратов	$6.124 * 10^{-2}$	$6.149 * 10^{-2}$	$1.217 * 10^{-1}$
Дисперсия	$5.285 * 10^{-2}$	$5.370 * 10^{-2}$	$1.128 * 10^{-1}$
$n = 60$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$-8.399 * 10^{-2}$	$-7.915 * 10^{-2}$	$-7.806 * 10^{-2}$
Среднее квадратов	$2.347 * 10^{-2}$	$2.271 * 10^{-2}$	$3.973 * 10^{-2}$
Дисперсия	$1.641 * 10^{-2}$	$1.645 * 10^{-2}$	$3.364 * 10^{-2}$
$n = 100$	$r$ (28)	$r_S$ (30)	$r_Q$ (29)
Среднее	$-1.010 * 10^{-1}$	$-9.533 * 10^{-2}$	$-8.666 * 10^{-2}$
Среднее квадратов	$2.098 * 10^{-2}$	$1.983 * 10^{-2}$	$2.849 * 10^{-2}$
Дисперсия	$1.078 * 10^{-2}$	$1.075 * 10^{-2}$	$2.098 * 10^{-2}$

Таблица 4: Характеристики смеси нормальных распределений

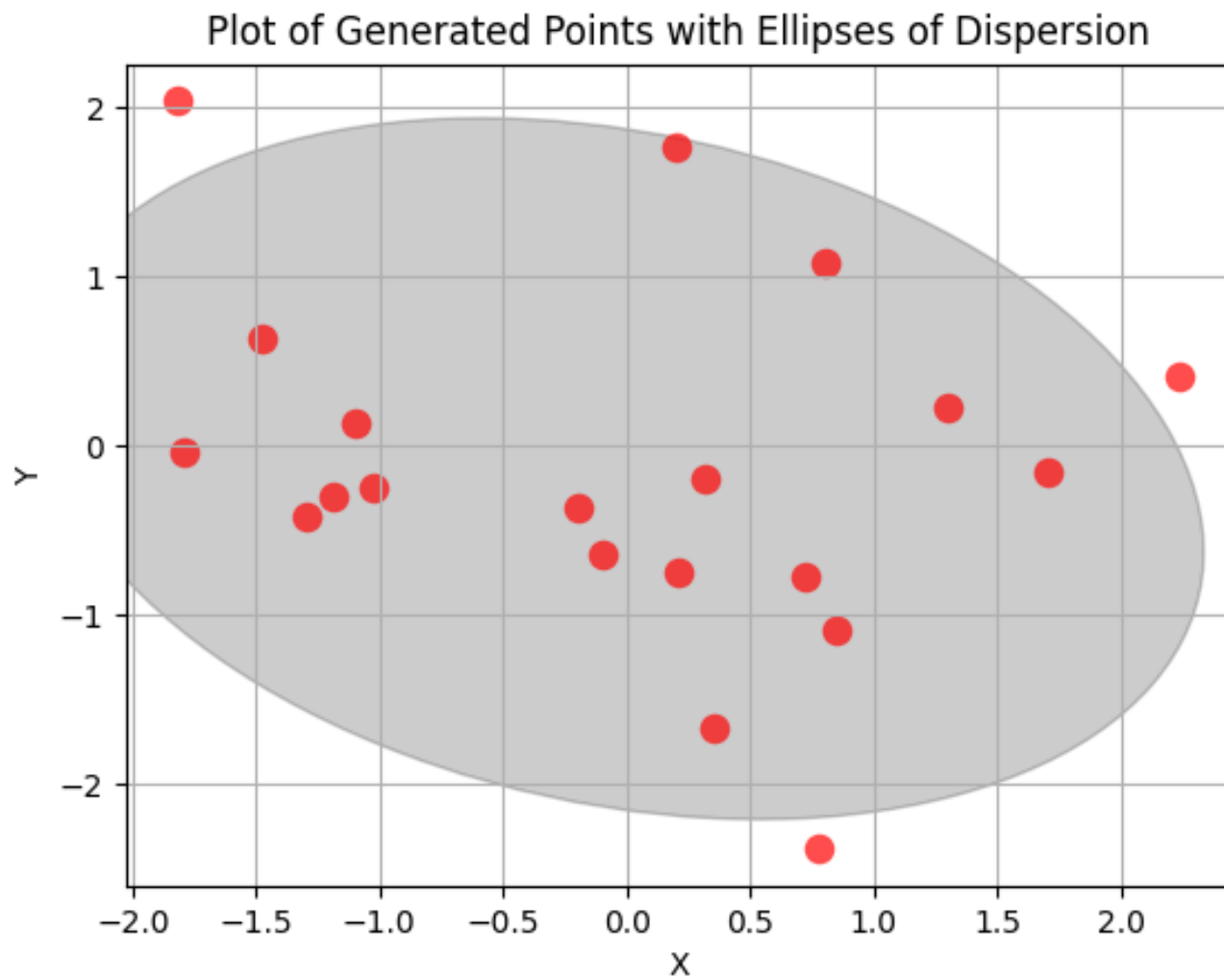


Рис. 1: Эллипсы равновероятности для выборки размером 20,  $\rho = 0$

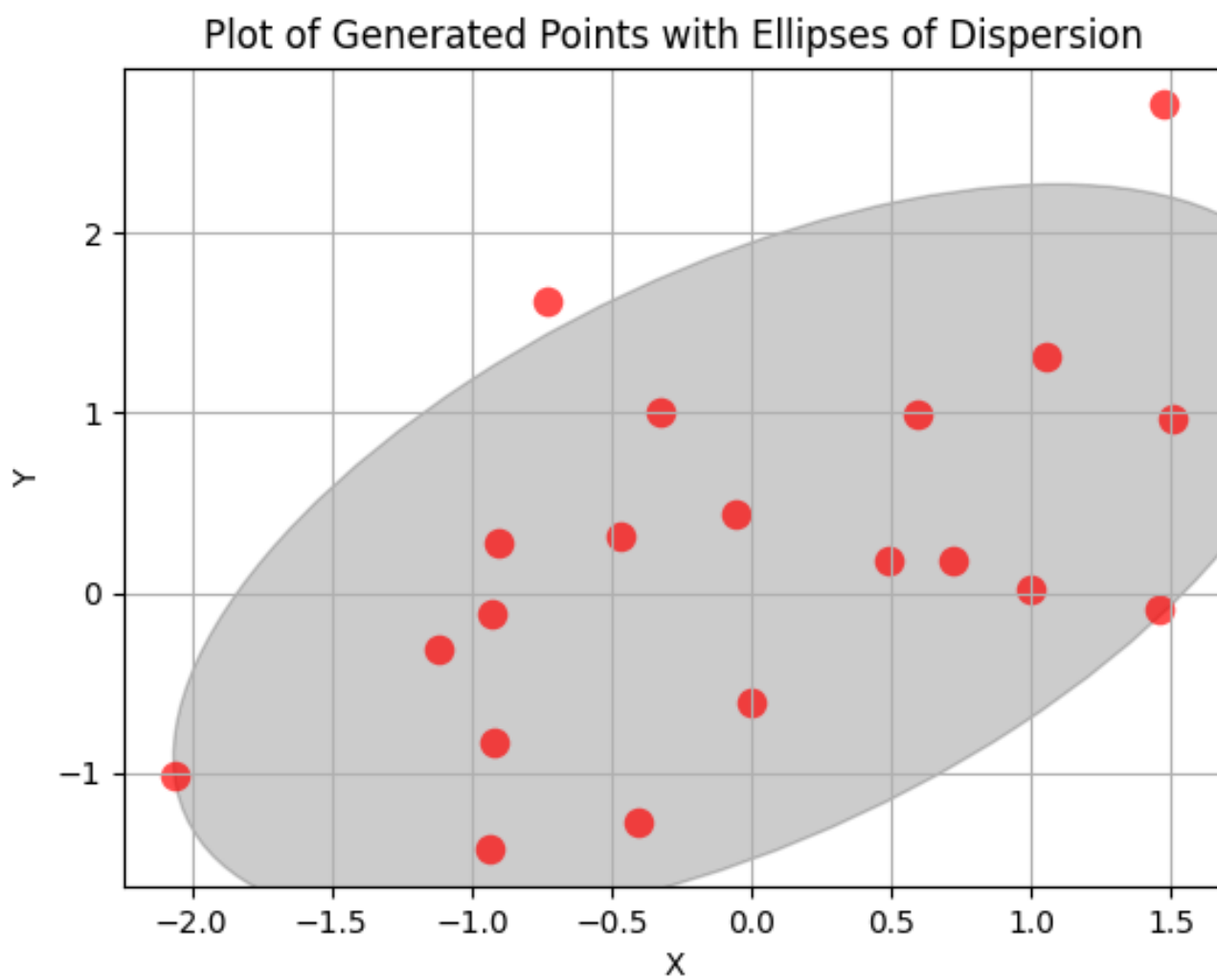


Рис. 2: Эллипсы равновероятности для выборки размером 20,  $\rho = 0,5$

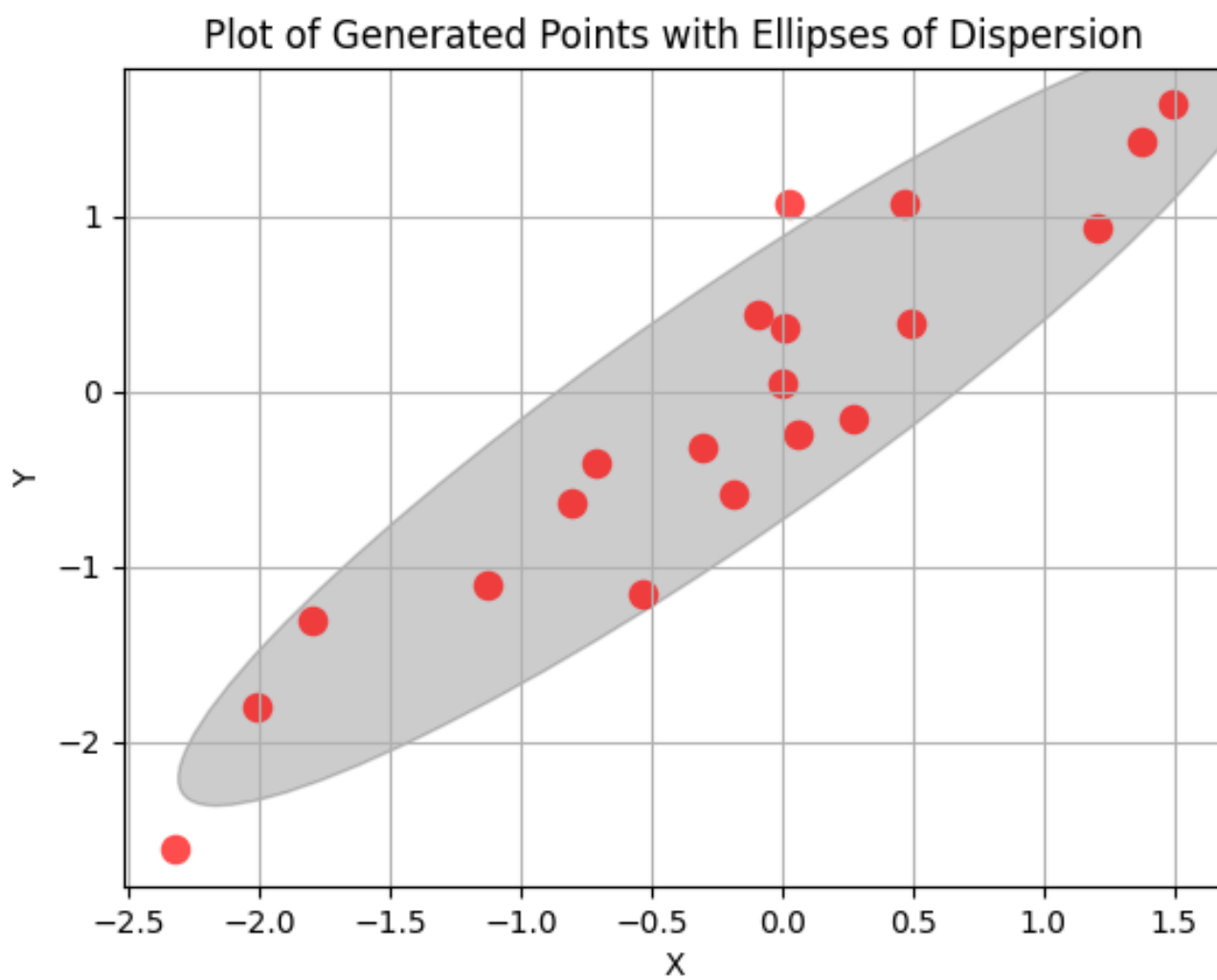


Рис. 3: Эллипсы равновероятности для выборки размером 20,  $\rho = 0,9$

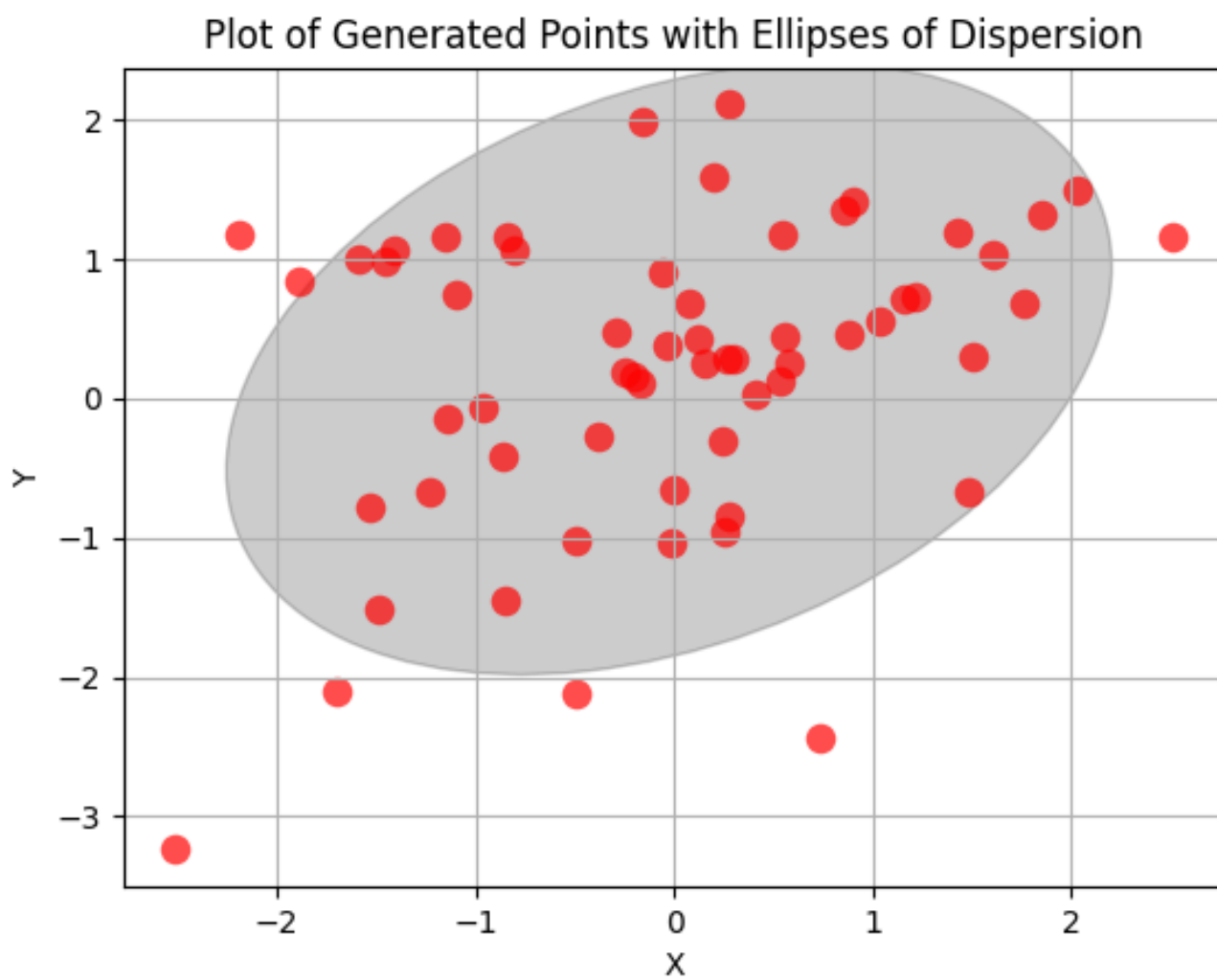


Рис. 4: Эллипсы равновероятности для выборки размером 60,  $\rho = 0$

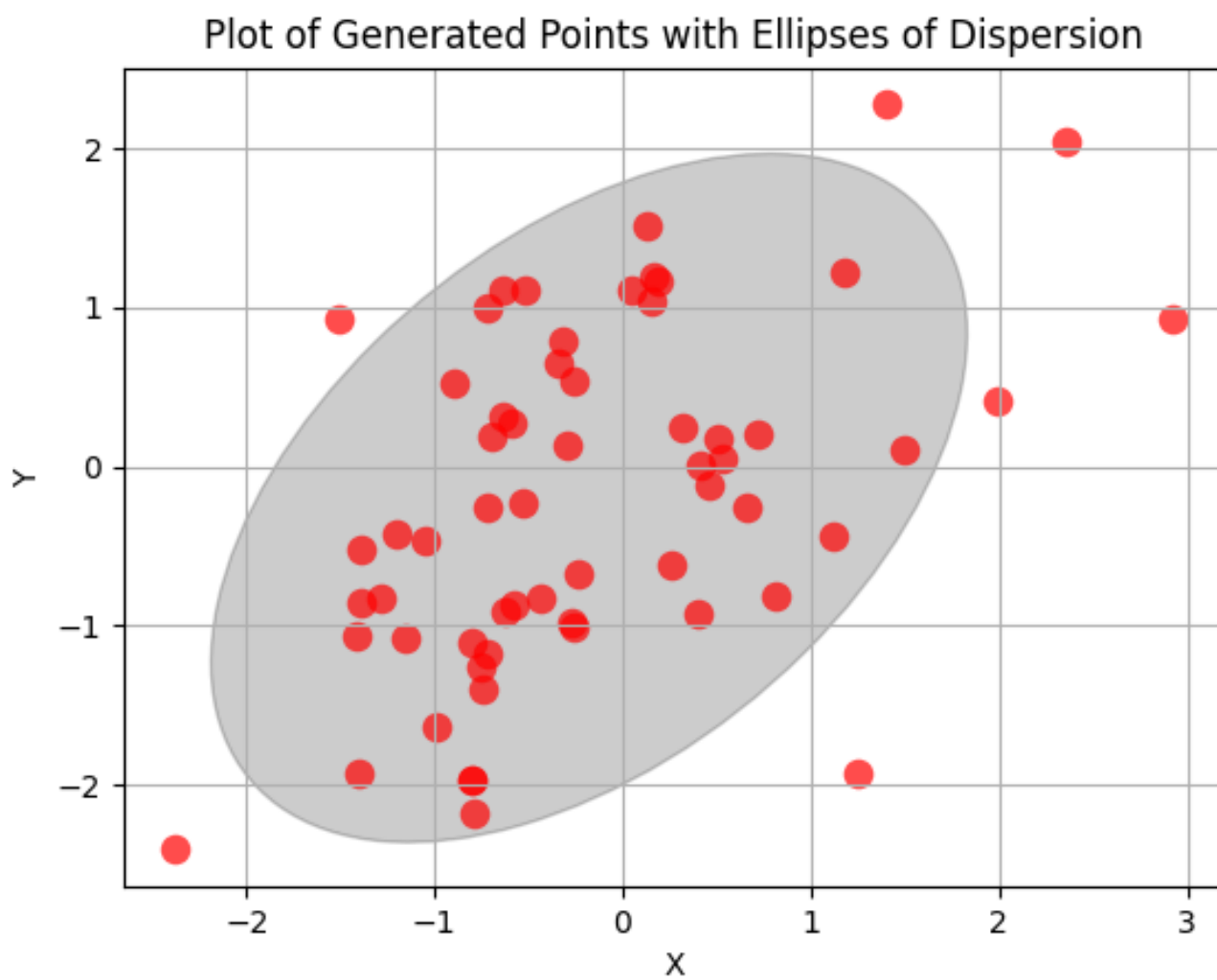


Рис. 5: Эллипсы равновероятности для выборки размером 60,  $\rho = 0,5$

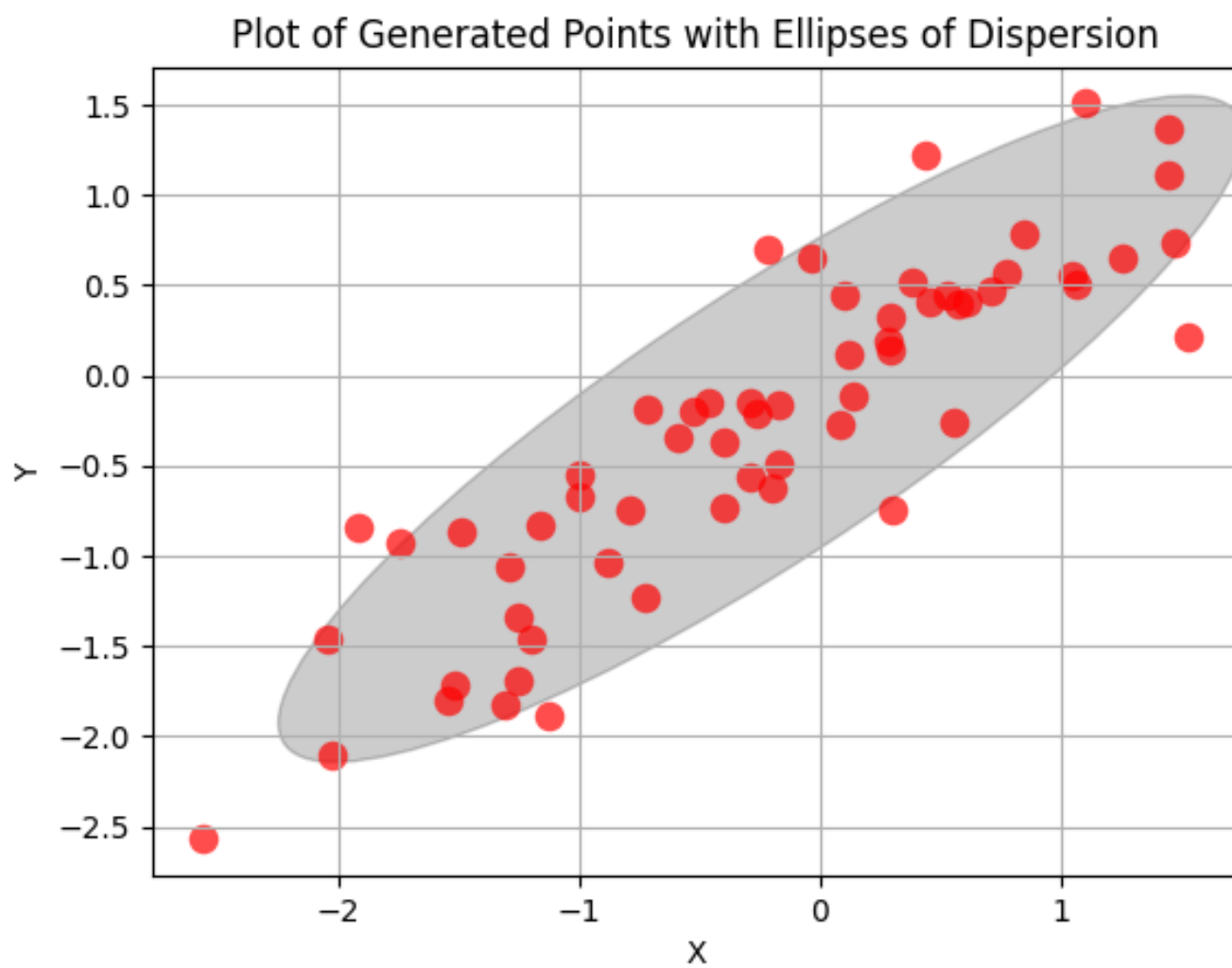


Рис. 6: Эллипсы равновероятности для выборки размером 60,  $\rho = 0,9$



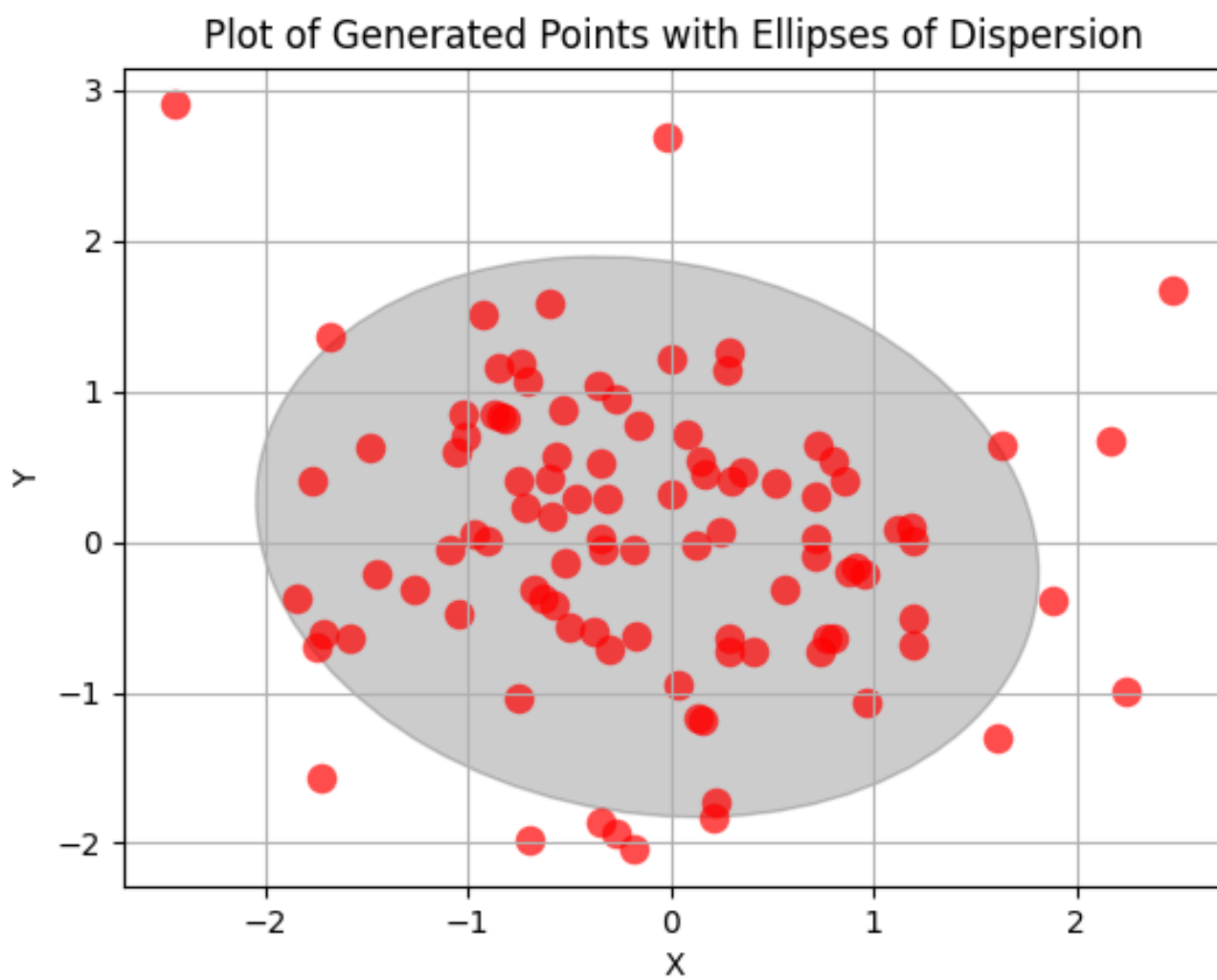


Рис. 7: Эллипсы равновероятности для выборки размером 100,  $\rho = 0$

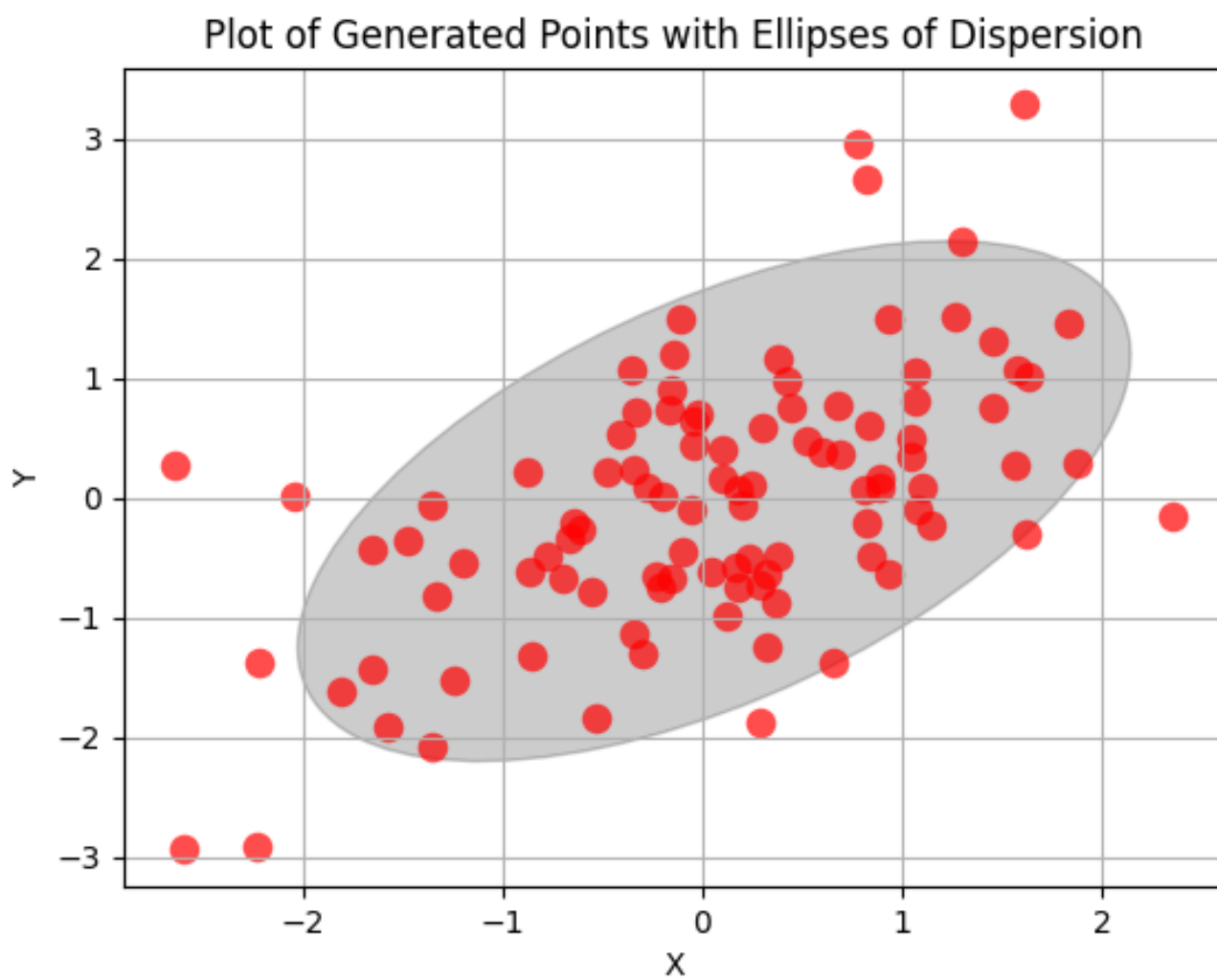


Рис. 8: Эллипсы равновероятности для выборки размером 100,  $\rho = 0,5$

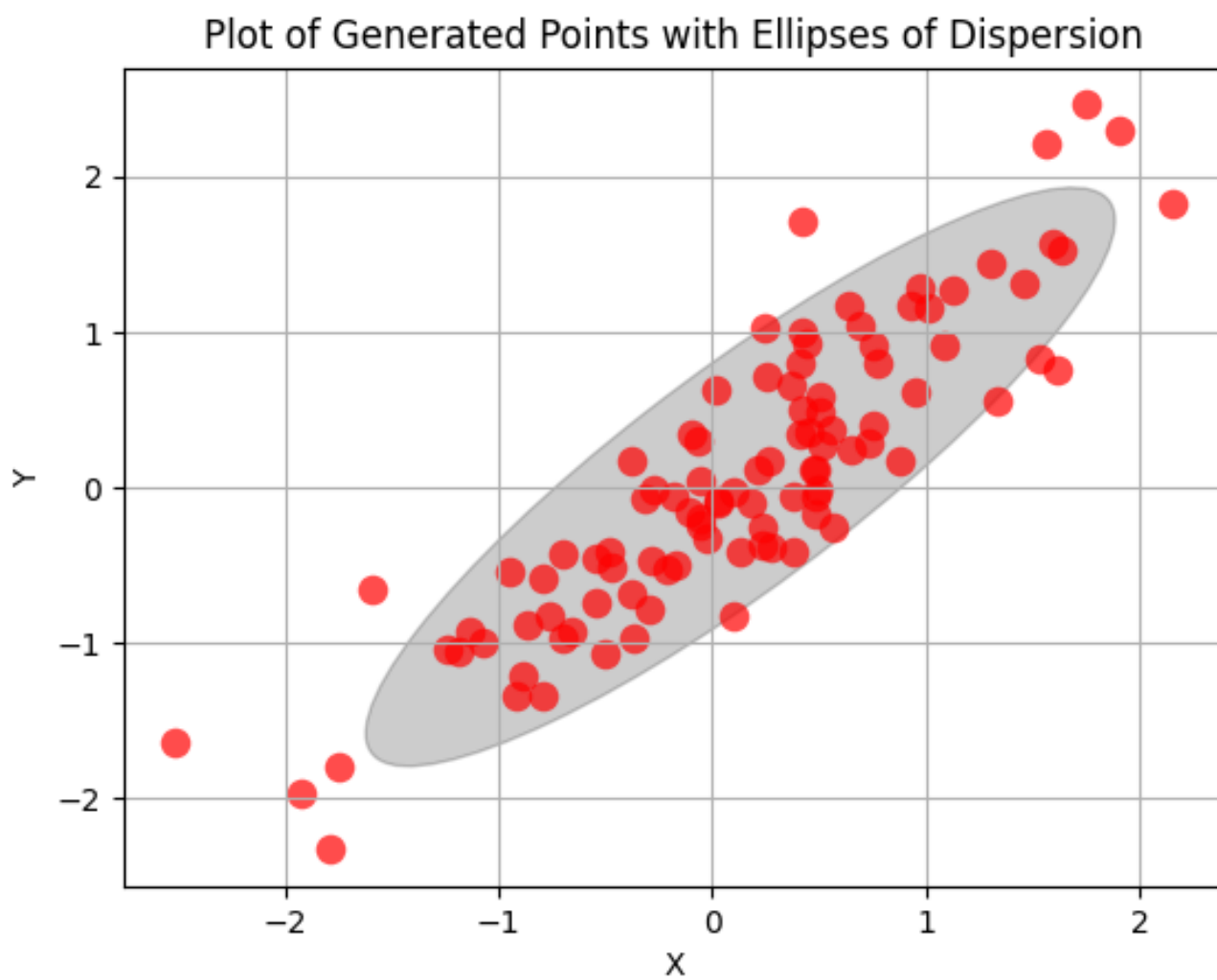


Рис. 9: Эллипсы равновероятности для выборки размером 100,  $\rho = 0,9$

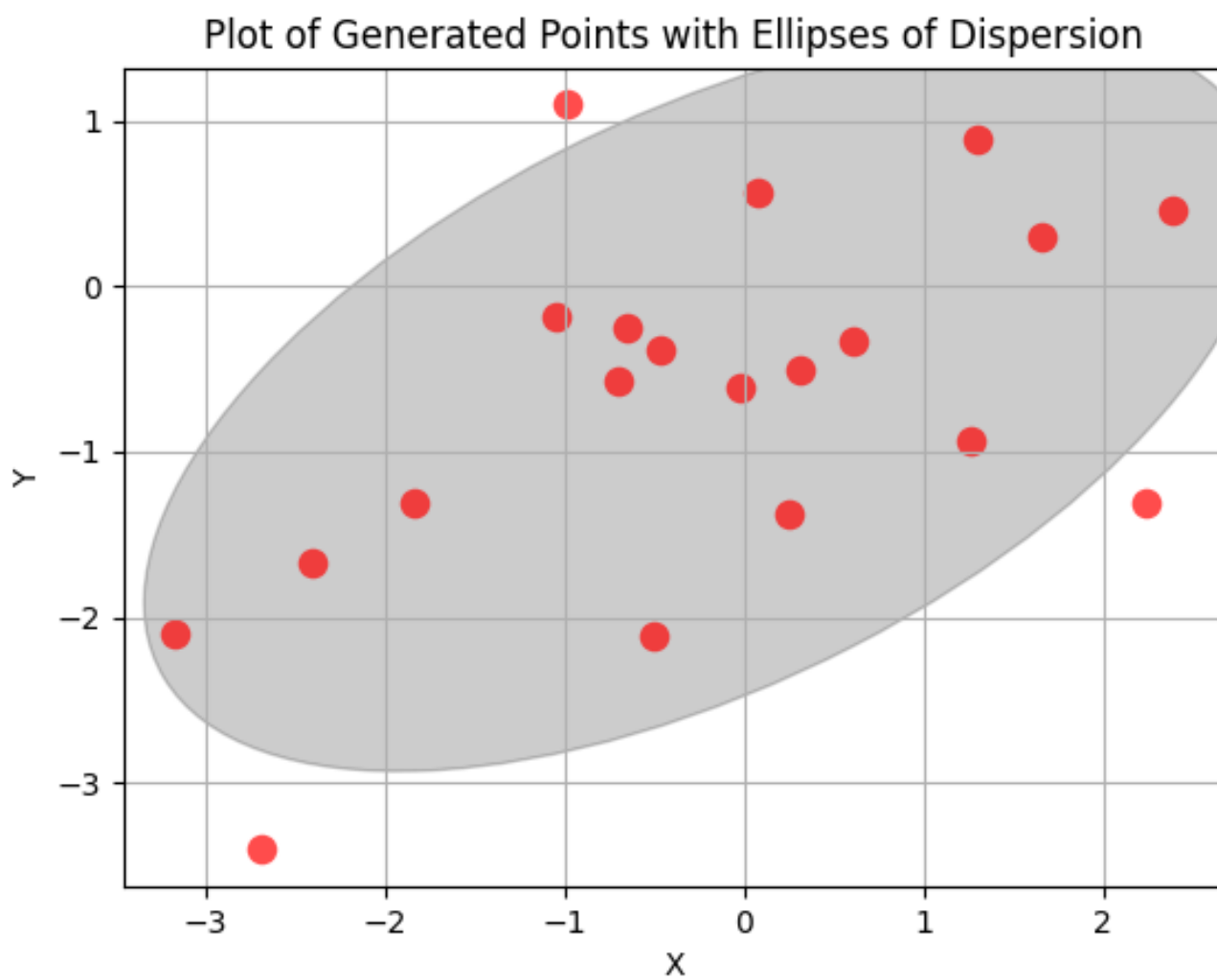


Рис. 10: Эллипсы равновероятности для выборки размером 100,  $\rho = 0$ , смешанное распределение

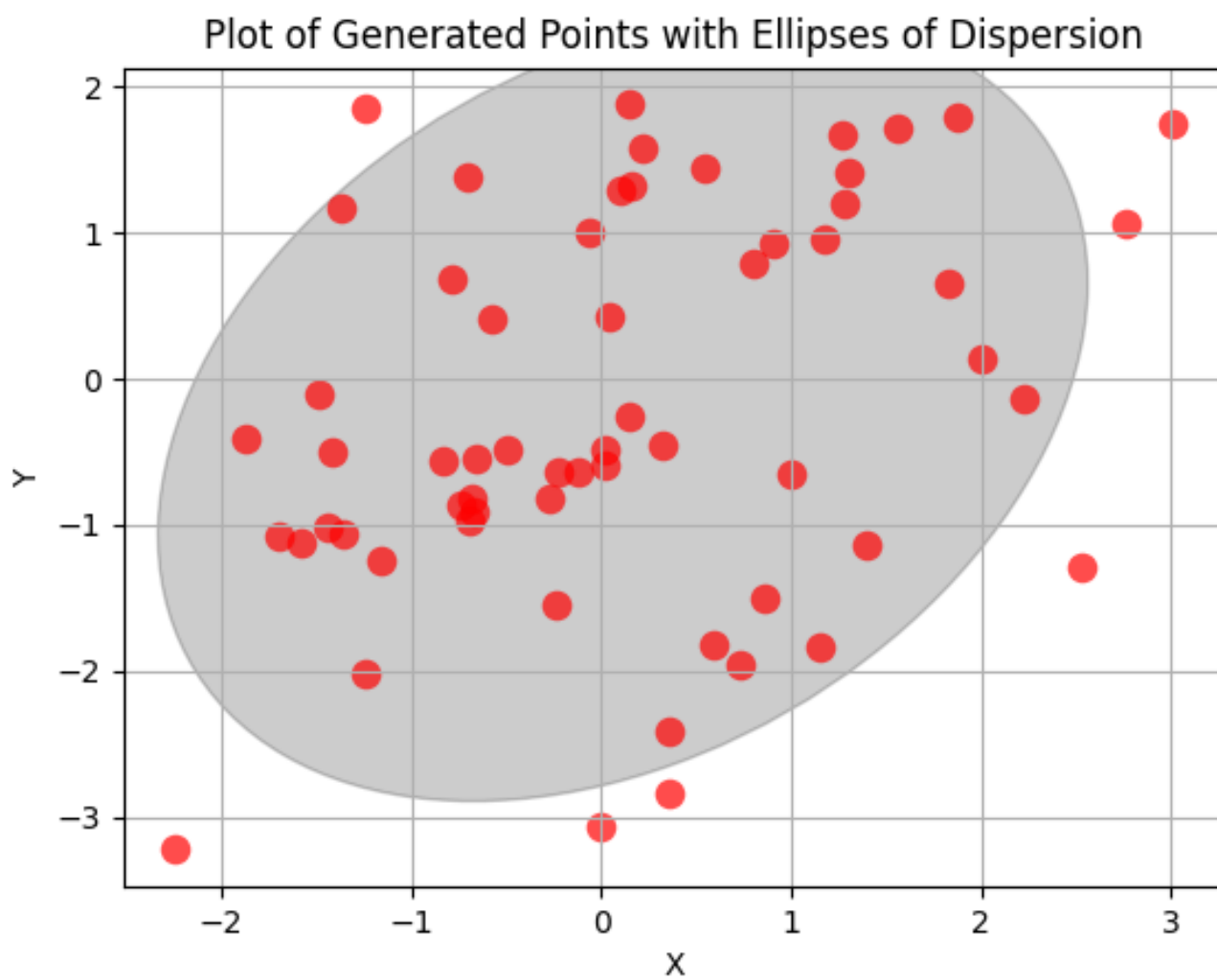


Рис. 11: Эллипсы равновероятности для выборки размером 100,  $\rho = 0,5$ , смешанное распределение

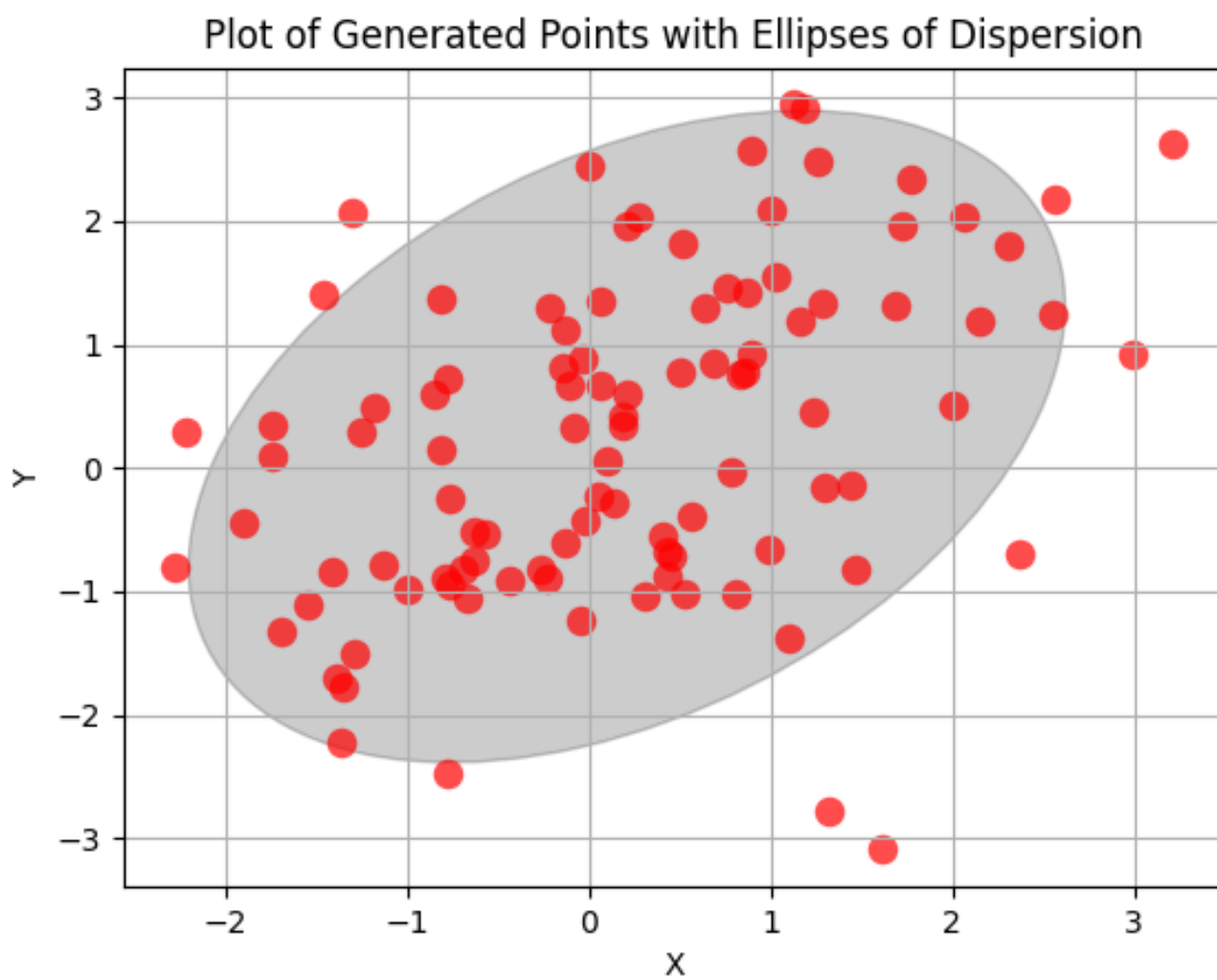


Рис. 12: Эллипсы равновероятности для выборки размером 100,  $\rho = 0,9$ , смешанное распределение

### 3.3.2 Простая линейная регрессия

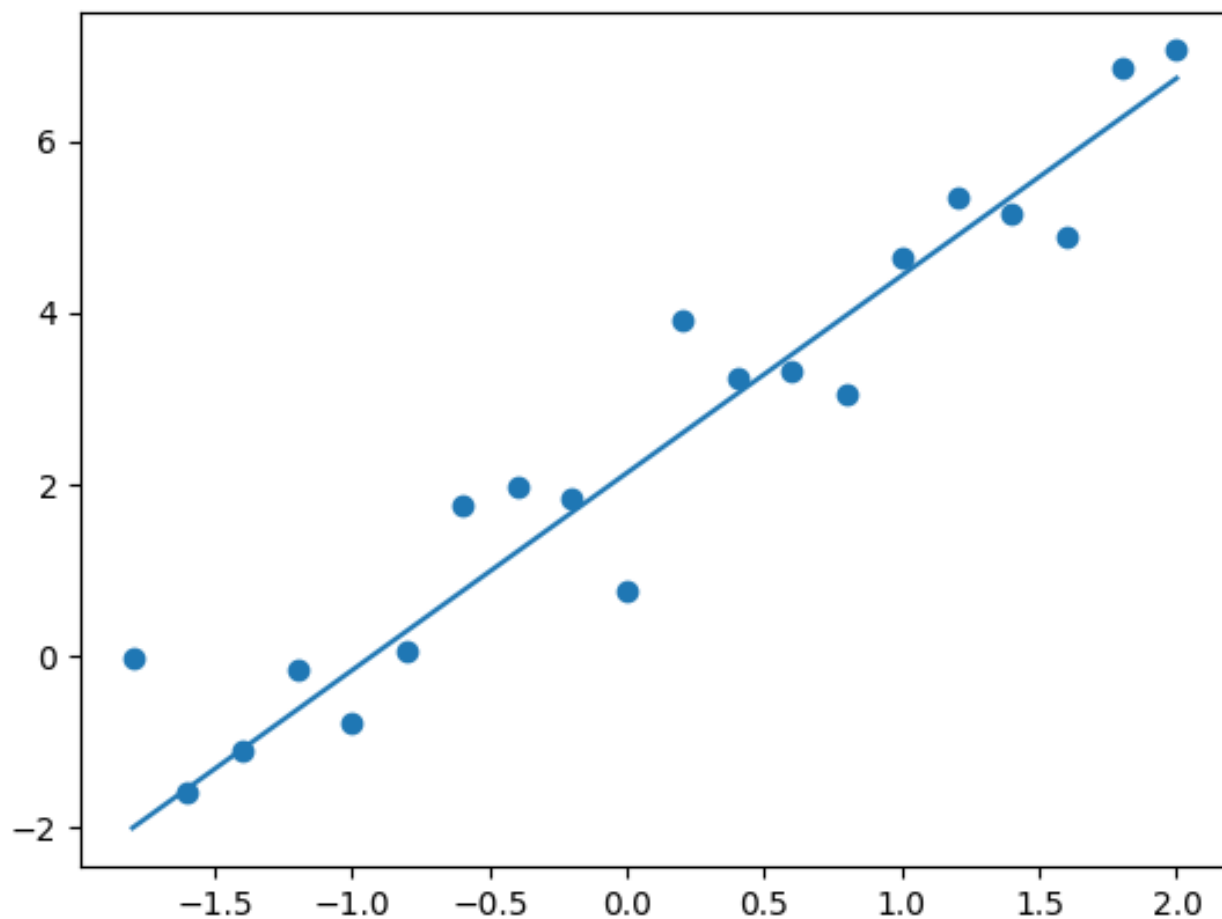


Рис. 13: Метод наименьших квадратов  
 $y = kx + b; k = 1.906, b = 2.011$

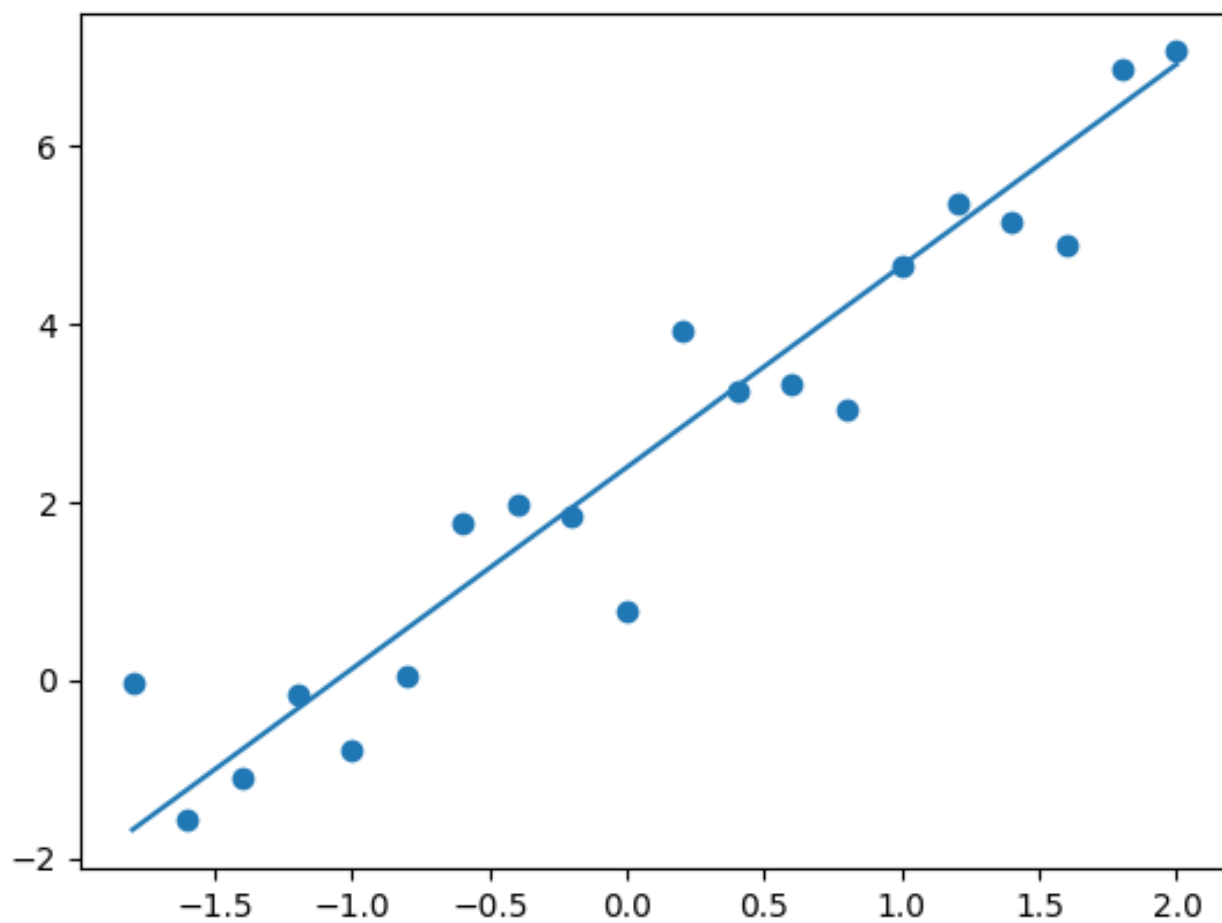


Рис. 14: Метод наименьших модулей  
 $y = kx + b; k = 2.027, b = 2.132$



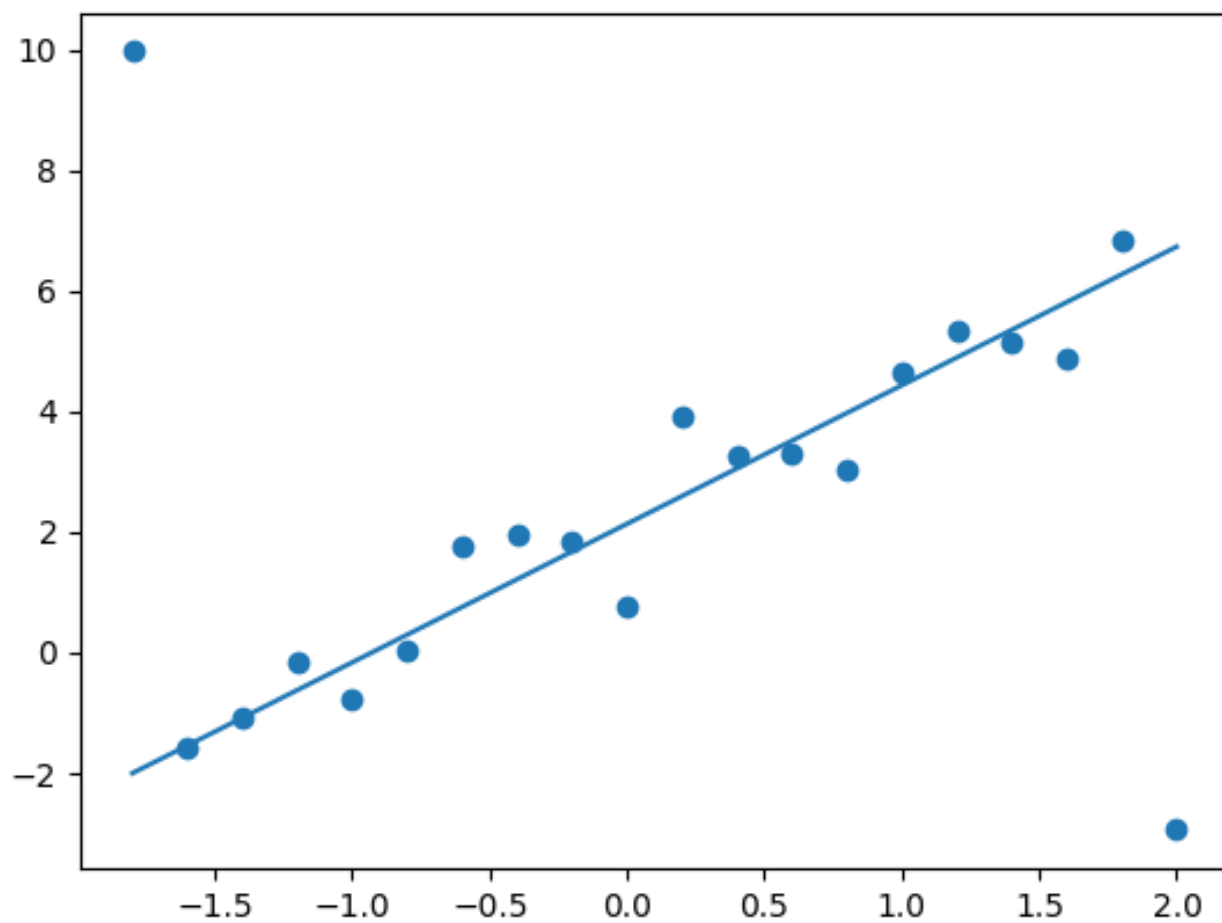


Рис. 15: Метод наименьших квадратов с возмущениями  
 $y = kx + b; k = 2.049, b = 0.582$

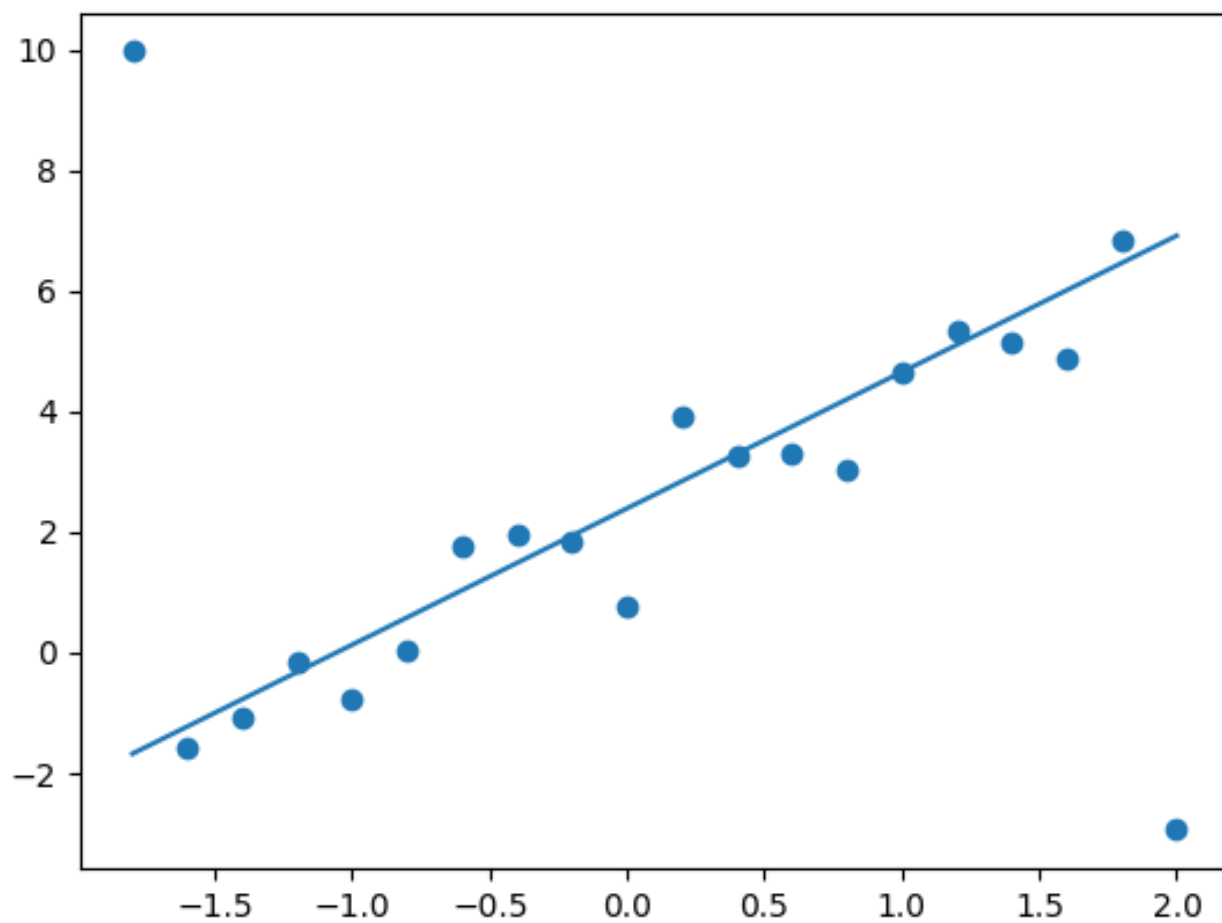


Рис. 16: Метод наименьших модулей с возмущениями  
 $y = kx + b; k = 1.867, b = 1.660$

Таблица 5: Оценки коэффициентов для выборки с выбросами

	МНК	МНМ
$\hat{a}$	1.906	2.027
$\frac{ a-\hat{a} }{a} * 100\%$	12	16.5
$\hat{b}$	2.011	2.132
$\frac{ b-\hat{b} }{b} * 100\%$	6.5	6.5

Таблица 6: Оценки коэффициентов для выборки без выбросов

	МНК	МНМ
$\hat{a}$	2.049	1.867
$\frac{ a-\hat{a} }{a} * 100\%$	19	19
$\hat{b}$	0.582	1.660
$\frac{ b-\hat{b} }{b} * 100\%$	78	30

### 3.4 Выводы

При увеличении размера выборки наблюдается повышение точности оценок, что проявляется в уменьшении дисперсий коэффициентов корреляции. Этот эффект соответствует основным принципам центральной предельной теоремы и закона больших чисел. Увеличение коэффициента корреляции  $\rho$  сопровождается ростом средних значений коэффициентов Пирсона, Спирмена и квадратичного коэффициента корреляции вследствие прямой зависимости между  $\rho$  и другими коэффициентами корреляции.

Метод наименьших квадратов проявляет эффективность в условиях, когда данные не содержат значительных выбросов, в то время как метод наименьших модулей демонстрирует превосходство в случае наличия значительных возмущений. При выборе метода следует учитывать особенности данных: при наличии выбросов предпочтительнее использовать метод наименьших модулей из-за его устойчивости к выбросам.

## 4 Лабораторные 7-8

### 4.1 Теоретическое обоснование

#### 4.1.1 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат.

Исчерпывающей характеристикой изучаемой случайной величины является её закон распределения. Поэтому естественно стремление исследователей построить этот закон приближённо на основе статистических данных.

Сначала выдвигается гипотеза о виде закона распределения.

После того как выбран вид закона, возникает задача оценивания его параметров и проверки (тестирования) закона в целом.

Для проверки гипотезы о законе распределения применяются критерии согласия. Таких критериев существует много. Мы рассмотрим наиболее обоснованный и наиболее часто используемый в практике — критерий  $\chi^2$  (хи-квадрат), введённый К.Пирсоном (1900 г.) для случая, когда параметры распределения известны. Этот критерий был существенно уточнён Р.Фишером (1924 г.), когда параметры распределения оцениваются по выборке, используемой для проверки.

Мы ограничимся рассмотрением случая одномерного распределения.

Итак, выдвинута гипотеза  $H_0$  о генеральном законе распределения с функцией распределения  $F(x)$ .

Рассматриваем случай, когда гипотетическая функция распределения  $F(x)$  не содержит неизвестных параметров.

Разобьём генеральную совокупность, т.е. множество значений изучаемой случайной величины  $X$  на  $k$  непересекающихся подмножеств  $\Delta_1, \Delta_2, \dots, \Delta_k$ .

Пусть  $p_i = P(X \in \Delta_i), i = 1, \dots, k$ .

Если генеральная совокупность — вся вещественная ось, то подмножества  $\Delta_i = (a_{i-1}, a_i]$  — полуоткрытые промежутки ( $i = 2, \dots, k-1$ ). Крайние промежутки будут полубесконечными:  $\Delta_1 = (-\infty, a_1], \Delta_k = (a_{k-1}, +\infty)$ . В этом случае  $p_i = F(a_i) - F(a_{i-1})$ ;  $a_0 = -\infty, a_k = +\infty (i = 1, \dots, k)$ .

Отметим, что  $\sum_{i=1}^k p_i = 1$ . Будем предполагать, что все  $p_i > 0 (i = 1, \dots, k)$ .

Пусть, далее,  $n_1, n_2, \dots, n_k$  — частоты попадания выборочных элементов в подмножества  $\Delta_1, \Delta_2, \dots, \Delta_k$  соответственно.

В случае справедливости гипотезы  $H_0$  относительные частоты  $n_i/n$  при большом  $n$  должны быть близки к вероятностям  $p_i (i = 1, \dots, k)$ , поэтому за меру отклонения выборочного распределения от гипотетического с функцией  $F(x)$  естественно выбрать величину

$$Z = \sum_{i=1}^k c_i \left( \frac{n_i}{n} - p_i \right)^2 \quad (1.1)$$

где  $c_i$  — какие-нибудь положительные числа (веса). К.Пирсоном в качестве весов выбраны числа  $c_i = n/p_i (i = 1, \dots, k)$ . Тогда получается статистика критерия хи-квадрат К.Пирсона

$$\chi^2 = \sum_{i=1}^k \frac{n}{p_i} \left( \frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (1.2)$$

которая обозначена тем же символом, что и закон распределения хи-квадрат.

К.Пирсоном доказана теорема об асимптотическом поведении статистики  $\chi^2$ , указывающая путь её применения.

**Теорема К.Пирсона.** Статистика критерия  $\chi^2$  асимптотически распределена по закону  $\chi^2$  с  $h-1$  степенями свободы. Это означает, что независимо от вида проверяемого распределения, т.е. функции  $F(x)$ , выборочная функция распределения статистики  $\chi^2$  при  $n \rightarrow \infty$  стремится к функции распределения случайной величины с плотностью вероятности

$$f_{k-1}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2^{\frac{k-1}{2}} \Gamma(\frac{k-1}{2})} & x \geq 0 \end{cases} \quad (1.3)$$

Для пояснения сущности метода  $\chi^2$  сделаем ряд замечаний.

**Замечание 1.** Выбор подмножеств  $\Delta_1, \Delta_2, \dots, \Delta_k$  и их числа  $k$  в принципе ничем не регламентируется, так как  $n \rightarrow \infty$ . Но так как число  $n$  хотя и очень большое, но конечное, то  $k$  должно быть с ним согласовано. Обычно его берут таким же, как и для построения гистограммы, т.е. можно руководствоваться формулой

$$k \approx 1.72 \sqrt[3]{n} \quad (1.4)$$

или формулой Старджесса

$$k \approx 1 + 3.3 \lg(n) \quad (1.5)$$

При этом, если  $\Delta_1, \Delta_2, \dots, \Delta_k$  — промежутки, то их длины удобно сделать равными, за исключением крайних — полубесконечных.

**Замечание 2.** (о числе степеней свободы). Числом степеней свободы функции (по старой терминологии) называется число её независимых аргументов. Аргументами статистики  $\chi^2$  являются частоты  $n_1, n_2, \dots, n_k$ . Эти частоты связаны одним равенством  $n_1 + n_2 + \dots + n_k = n$ , а в остальном независимы в силу независимости элементов выборки. Таким образом, функция  $\chi^2$  имеет  $k - 1$  независимых аргументов: число частот минус одна связь. В силу теоремы Пирсона число степеней свободы статистики  $\chi^2$  отражается на виде асимптотической плотности  $f_{k-1}(x)$ .

На основе общей схемы проверки статистических гипотез сформулируем следующее правило. Правило проверки гипотезы о законе распределения по методу  $\chi^2$

1. Выбираем уровень значимости  $\alpha$ .
2. По таблице находим квантиль  $\chi^2_{1-\alpha}(k-1)$  распределения хи-квадрат с  $k-1$  степенями свободы порядка  $1 - \alpha$ .
3. С помощью гипотетической функции распределения  $F(x)$  вычисляем вероятности  $p_i = P(X \in \Delta_i), i = 1, \dots, k$ .
4. Находим частоты  $n$  попадания элементов выборки в подмножества  $\Delta_i, i = 1, \dots, k$ .
5. Вычисляем выборочное значение статистики критерия  $\chi^2$ :

$$\chi_B^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (1.7)$$

6. Сравниваем  $\chi_B^2$  и квантиль  $\chi^2_{1-\alpha}(k-1)$ 
  - Если  $\chi_B^2 < \chi^2_{1-\alpha}(k-1)$ , то гипотеза  $H_0$  на данном этапе проверки принимается.
  - Если  $\chi_B^2 \geq \chi^2_{1-\alpha}(k-1)$ , то гипотеза  $H_0$  отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

**Замечание 3.** Из формулы (1.2) видим, что веса  $c_i = n/p_i$  пропорциональны  $n$ , т.е. с ростом  $n$  увеличиваются. Отсюда следует, что если выдвинутая гипотеза неверна, то относительные частоты  $n_i/n$  не будут близки к вероятностям  $p_i$ , и с ростом  $n$  величина  $\chi_B^2$  будет увеличиваться. При фиксированном уровне значимости  $\alpha$  будет фиксировано пороговое число — квантиль  $\chi^2_{1-\alpha}(k-1)$ , поэтому, увеличивая  $n$ , мы придём к неравенству  $\chi_B^2 > \chi^2_{1-\alpha}(k-1)$ , т.е. с увеличением объёма выборки неверная гипотеза будет отвергнута. Отсюда следует, что при сомнительной ситуации, когда  $\chi_B^2 \approx \chi^2_{1-\alpha}(k-1)$ , можно попытаться увеличить объём выборки (например, в 2 раза), чтобы требуемое неравенство было более чётким.

**Замечание 4.** Теория и практика применения критерия  $\chi^2$  указывают, что если для каких-либо подмножеств  $\Delta_i (i = 1, \dots, k)$  условие  $np_i \geq 5$  не выполняется, то следует объединить соседние подмножества (промежутки).

Это условие выдвигается требованием близости величин

$$(n_i - np_i) / \sqrt{np_i}$$

квадраты которых являются слагаемыми  $\chi^2$  к нормальным  $N(0, 1)$ . Тогда случайная величина в формуле (1.2) будет распределена по закону, близкому к хи-квадрат. Такая близость обеспечивается достаточной численностью элементов в подмножествах  $\Delta_i$  [2, с. 481 — 485].

#### 4.1.2 Проверка гипотезы о равенстве дисперсий двух нормальных генеральных совокупностей.

Эта гипотеза возникает при сравнении точности двух одинаковых измерительных приборов, при выяснении стабильности работы одного и того же прибора в разное время, при выяснении однородности выборки.

Имеются две генеральные совокупности  $X$  и  $Y$ , распределённые нормально. Пусть  $x_i^m$  — выборка объема  $m$  из первой, а  $y_i^n$  — выборка объема  $n$  из второй генеральной совокупности, полученной независимо от первой. Генеральные математические ожидания неизвестны. Проверяется гипотеза  $H_0$  о равенстве дисперсий:  $\sigma_X^2 = \sigma_Y^2$  против гипотезы  $H_1 : \sigma_X^2 \neq \sigma_Y^2$ . Для этого вычисляются несмещенные оценки дисперсий

$$s_X^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2 \quad (1.8)$$

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (1.9)$$

Здесь  $\bar{x}$  и  $\bar{y}$  — выборочные средние, найденные заранее.

Статистикой критерия является отношение Фишера

$$F = \frac{s_X^2}{\sigma_X^2} : \frac{s_Y^2}{\sigma_Y^2} \quad (1.10)$$

В предположении справедливости гипотезы  $H_0$  это отношение упрощается, так как  $\sigma_X^2 = \sigma_Y^2$ , и записывается в виде

$$F = s_X^2 / s_Y^2 \quad (1.11)$$

Статистика критерия Фишера (1.11) имеет **распределение Фишера** с  $k_1 = m - 1$  и  $k_2 = n - 1$  степенями свободы  $F(k_1, k_2)$ . Обычно в числителе ставится большая из двух сравниваемых дисперсий.

При отношении дисперсий  $\sigma_X^2 / \sigma_Y^2 = 1$  отношение оценок (1.11) должно быть близким к 1, поэтому гипотеза  $H_0$  должна отвергаться при слишком малых или слишком больших значениях  $F$ . Тест Фишера.

1. Вычисляем несмещенные оценки дисперсий (1.8) и (1.9).
2. Выбираем статистику критерия (1.11).
3. Выбираем уровень значимости  $\alpha$ .
4. По таблице квантиль  $F_{1-\frac{\alpha}{2}}(k_1, k_2)$ . (1.12)
5. Вычисляем выборочное значение  $F_B$  (1.13) статистики критерия.
6. Сравниваем  $F_B$  и  $F_{1-\frac{\alpha}{2}}(k_1, k_2)$ . Если  $F_B < F_{1-\frac{\alpha}{2}}(k_1, k_2)$ , то гипотеза  $H_0$  на выбранном уровне значимости  $\alpha$  принимается. В противном случае — отвергается.

## 4.2 Постановка задачи

Мощность распределения  $n = 20, 100$ .

Распределения: нормальное, Стьюдента ( $k = 3$ ), равномерное.

Провести исследование по методике 1.1.1.

Результаты вычислений оформить в виде таблицы.

## 4.3 Результат

### 4.3.1 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Промежуточные вычисления  $\chi_B^2$   $n = 20$

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.026998	0.000729	0.026998
$[-3, 3]$	20	19.946004	0.002916	0.000146
$[3, \infty]$	0	0.026998	0.000729	0.026998

Таблица 7: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (норм - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.576689	0.33257	0.576689
$[-3, 3]$	20	18.846622	1.33028	0.070585
$[3, \infty]$	0	0.576689	0.33257	0.576689

Таблица 8: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (норм - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	4.0	16.0	4.0
$[-3, 3]$	20	12.0	64.0	5.333333
$[3, \infty]$	0	4.0	16.0	4.0

Таблица 9: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (норм - равномерное)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.026998	0.000729	0.026998
$[-3, 3]$	19	19.946004	0.894924	0.044867
$[3, \infty]$	1	0.026998	0.946733	35.066833

Таблица 10: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (студент - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.576689	0.33257	0.576689
$[-3, 3]$	20	18.846622	1.33028	0.070585
$[3, \infty]$	0	0.576689	0.33257	0.576689

Таблица 11: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (студент - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	1	4.0	9.0	2.25
$[-3, 3]$	19	12.0	49.0	4.083333
$[3, \infty]$	0	4.0	16.0	4.0

Таблица 12: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (студент - равномерное)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	1	0.026998	0.946733	35.066833
$[-3, 3]$	14	19.946004	35.354965	1.772534
$[3, \infty]$	5	0.026998	24.730749	916.022866

Таблица 13: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (равномерное - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	6	0.576689	29.412304	51.002032
$[-3, 3]$	10	18.846622	78.262726	4.152613
$[3, \infty]$	4	0.576689	11.719059	20.321286

Таблица 14: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (равномерное - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	6	4.0	4.0	1.0
$[-3, 3]$	12	12.0	0.0	0.0
$[3, \infty]$	2	4.0	4.0	1.0

Таблица 15: Промежуточные вычисления  $\chi_B^2$  для  $n = 20$  (равномерное - равномерное)



Промежуточные вычисления  $\chi_B^2$   $n = 100$

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.13499	0.018222	0.13499
$[-3, -1.5]$	5	6.54573	2.389282	0.365014
$[-1.5, 0]$	39	43.31928	18.656179	0.430667
$[0, 1.5]$	51	43.31928	58.993462	1.361829
$[1.5, 3]$	5	6.54573	2.389282	0.365014
$[3, \infty]$	0	0.13499	0.018222	0.13499

Таблица 16: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (норм - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	1	20.0	361.0	18.05
$[-3, -1.5]$	6	15.0	81.0	5.4
$[-1.5, 0]$	41	15.0	676.0	45.066667
$[0, 1.5]$	43	15.0	784.0	52.266667
$[1.5, 3]$	8	15.0	49.0	3.266667
$[3, \infty]$	1	20.0	361.0	18.05

Таблица 18: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (норм - равномерное)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	2	2.883444	0.780474	0.270674
$[-3, -1.5]$	8	8.645749	0.416992	0.048231
$[-1.5, 0]$	42	38.470807	12.455205	0.323757
$[0, 1.5]$	36	38.470807	6.104886	0.158689
$[1.5, 3]$	6	8.645749	6.999988	0.809645
$[3, \infty]$	6	2.883444	9.71292	3.368513

Таблица 20: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (студент - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	25	0.13499	618.268732	4580.114332
$[-3, -1.5]$	16	6.54573	89.383215	13.655194
$[-1.5, 0]$	12	43.31928	980.897292	22.643435
$[0, 1.5]$	16	43.31928	746.343053	17.228889
$[1.5, 3]$	13	6.54573	41.657597	6.364087
$[3, \infty]$	18	0.13499	319.158589	2364.316281

Таблица 22: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (равномерное - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	2.883444	8.314251	2.883444
$[-3, -1.5]$	9	8.645749	0.125494	0.014515
$[-1.5, 0]$	47	38.470807	72.747138	1.89097
$[0, 1.5]$	39	38.470807	0.280046	0.007279
$[1.5, 3]$	5	8.645749	13.291486	1.537343
$[3, \infty]$	0	2.883444	8.314251	2.883444

Таблица 17: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (норм - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	0	0.13499	0.018222	0.13499
$[-3, -1.5]$	6	6.54573	0.297822	0.045499
$[-1.5, 0]$	33	43.31928	106.487537	2.458202
$[0, 1.5]$	45	43.31928	2.82482	0.065209
$[1.5, 3]$	11	6.54573	19.840518	3.031063
$[3, \infty]$	5	0.13499	23.668324	175.334163

Таблица 19: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (студент - норм)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	5	20.0	225.0	11.25
$[-3, -1.5]$	6	15.0	81.0	5.4
$[-1.5, 0]$	47	15.0	1024.0	68.266667
$[0, 1.5]$	34	15.0	361.0	24.066667
$[1.5, 3]$	6	15.0	81.0	5.4
$[3, \infty]$	2	20.0	324.0	16.2

Таблица 21: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (студент - равномерное)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	22	2.883444	365.442703	126.738257
$[-3, -1.5]$	8	8.645749	0.416992	0.048231
$[-1.5, 0]$	24	38.470807	209.404248	5.443199
$[0, 1.5]$	18	38.470807	419.053928	10.892777
$[1.5, 3]$	13	8.645749	18.959502	2.192928
$[3, \infty]$	15	2.883444	146.810922	50.915124

Таблица 23: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$  (равномерное - студент)

$i$	$n_i$	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
$[-\infty, -3]$	20	20.0	0.0	0.0
$[-3, -1.5]$	12	15.0	9.0	0.6
$[-1.5, 0]$	14	15.0	1.0	0.066667
$[0, 1.5]$	19	15.0	16.0	1.066667
$[1.5, 3]$	13	15.0	4.0	0.266667
$[3, \infty]$	22	20.0	4.0	0.2

Таблица 24: Промежуточные вычисления  $\chi_B^2$  для  $n = 100$   
(равномерное - равномерное)

Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Таблица 25:  $\chi_B^2$  для  $n = 20$   $\chi_{1-\alpha}^2(k-1) = 7.814728$

$F_{real} \setminus F_{hip}$	Нормальное	Стьюдента	Равномерное
Нормальное	0.054142	1.223962	13.333333
Стьюдента	35.138698	1.223962	7.5
Равномерное	472.798306	57.515779	3.833333

Таблица 26: Проверка гипотез для выборки  $n = 20$

$F_{real} \setminus F_{hip}$	Нормальное	Стьюдента	Равномерное
Нормальное	верна	верна	не верна
Стьюдента	верна	верна	не верна
Равномерное	не верна	не верна	не верна

Таблица 27:  $\chi_B^2$  для  $n = 100$   $\chi_{1-\alpha}^2(k-1) = 12.591587$

$F_{real} \setminus F_{hip}$	Нормальное	Стьюдента	Равномерное
Нормальное	7.546754	7.691214	191.066667
Стьюдента	95.717563	5.176739	153.783333
Равномерное	3561.497349	259.374862	7.833333

Таблица 28: Проверка гипотез для выборки  $n = 100$

$F_{real} \setminus F_{hip}$	Нормальное	Стьюдента	Равномерное
Нормальное	верна	верна	не верна
Стьюдента	не верна	верна	не верна
Равномерное	не верна	не верна	верна

#### 4.3.2 Проверка гипотезы о равенстве дисперсий двух нормальных генеральных совокупностей

Случай 1: выбрать две выборки мощностью  $m = 20$  и  $n = 40$ . Обозначим за  $(N_1)$

Случай 2: Выбрать две выборки мощностью  $m = 20$  и  $n = 100$ . Обозначим за  $(N_2)$

	$s_X^2$	$s_Y^2$	$F_B$	$F_{1-\alpha 2}(k_1, k_2)$
$N_1$	0.740	0.975	1.317	2.096
$N_2$	0.740	1.045	1.412	1.870

Таблица 29: Выборочное значение  $F_B$  статистики критерия для разных для случаев  $N_1$  и  $N_2$

- $N_1$ :  
 $F_B < F_{1-\alpha 2}(k_1, k_2) \Rightarrow$  гипотеза  $H_0$  на данном этапе проверки принимается.
- $N_2$ :  
 $F_B < F_{1-\alpha 2}(k_1, k_2) \Rightarrow$  гипотеза  $H_0$  на данном этапе проверки принимается.

## 4.4 Выводы

В ходе лабораторной работы были измерены значения статистики  $\chi^2$  для выборок разного размера и сравнены с соответствующими квантилями распределения хи-квадрат. Результаты, представленные в таблице, показывают, что метод  $\chi^2$  позволяет проверить соответствие рассматриваемых распределений (нормального, Стьюдента с параметром  $k = 3$  и равномерного) гипотетическому закону.

Также были вычислены значения статистики  $F_B$  и сравнены с соответствующими квантилями распределения Фишера  $F_{1-\frac{\alpha}{2}}(k_1, k_2)$ . Из таблицы видно, что при  $\alpha = 0.05$  гипотеза о равенстве дисперсий для нормального распределения справедлива для выборок размером  $n = 100$  из генеральной совокупности.

## 5 Ссылка на репозиторий GitHub

<https://github.com/11AgReS1SoR11/MatStat>