

Санкт-Петербургский политехнический университет  
Петра Великого  
Физико-Механический институт  
Кафедра “Прикладная математика”

Отсчет  
По лабораторным работам №5–8  
По дисциплине  
“Математическая статистика”

Выполнил студент:  
Золотухин Илья Сергеевич  
Группа:  
5030102/90101  
  
Проверил:  
К.ф.-м.н., доцент  
Баженов Александр Николаевич

Санкт-Петербург  
2022 г.

ТАБЛИЦА 1 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, $n = 20$ .....	9
ТАБЛИЦА 2 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, $n = 60$ .....	9
ТАБЛИЦА 3 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, $n = 100$ .....	10
ТАБЛИЦА 4 СМЕСЬ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ.....	10
ТАБЛИЦА 5 ВЫЧИСЛЕНИЕ ХИ-КВАДРАТ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ О НОРМАЛЬНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ $N(x, \mu, \sigma)$ .....	14
ТАБЛИЦА 6 ВЫЧИСЛЕНИЕ ХИ-КВАДРАТ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ РАСПРЕДЕЛЕНИЯ ЛАПЛАСА .....	15
ТАБЛИЦА 7 ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ПАРАМЕТРОВ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ. ....	15
ТАБЛИЦА 8 ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛЫ ДЛЯ ПАРАМЕТРОВ ПРОИЗВОЛЬНОГО РАСПРЕДЕЛЕНИЯ. АСИМПТОТИЧЕСКИЙ ПОДХОД .....	16
РИСУНОК 1 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ $n = 20$ .....	11
РИСУНОК 2 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, $n = 60$ .....	11
РИСУНОК 3 ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, $n = 100$ .....	11
РИСУНОК 4 ВЫБОРКА БЕЗ ВОЗМУЩЕНИЙ.....	12
РИСУНОК 5 ВЫБОРКА С ВОЗМУЩЕНИЯМИ .....	13



# 1 Постановка задачи

1. Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения  $N(x, y, 0, 0, 1, 1, \rho)$ .  
Коэффициент корреляции  $\rho$  взять равны 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:  
$$f(x, y) = 0.9N(x, y, 0, 0, 1, 1, 0.9) + 0.1N(x, y, 0, 0, 10, 10, -0.9).$$
  
Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.
2. Найти оценки коэффициентов линейной регрессии  $y_i = a + bx_i + e_i$ , используя 20 точек на отрезке  $[-1.8; 2]$  с равномерным шагом равны 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. Сгенерировать выборку объёмом 100 элементов для нормального распределения  $N(x, 0, 1)$ . По сгенерированной выборке оценить параметры  $\mu$  и  $\sigma$  нормального закона методом максимального правдоподобия. В качестве основной гипотезы  $H_0$  будем считать, что сгенерированное распределение имеет вид  $N(x, \hat{\mu}, \hat{\sigma})$ . Проверить основную гипотезу, используя критерий согласия  $\chi^2$ . В качестве уровня значимости взять  $\alpha = 0.05$ . Привести таблицу вычислений  $\chi^2$ .
4. Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону  $N(x, 0, 1)$ , для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик  $\chi^2$  и Стьюдента. В качестве параметра надёжности взять  $\gamma = 0.95$ .

## 2 Теория

### 2.1 Двумерное нормальное распределение

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

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

Компоненты  $X, Y$  двумерной нормальной случайной величины также распределены с математическими ожиданиями  $\bar{x}, \bar{y}$  и средними квадратическими отклонениями  $\sigma_x, \sigma_y$  соответственно. [1, с.133-134]

Параметр  $\rho$  называется коэффициентом корреляции.

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

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

$$K = cov(X, Y) = M[(X - \bar{x})(Y - \bar{y})] \quad (2)$$

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

$$\rho = \frac{K}{\sigma_x\sigma_y} \quad (3)$$

### 2.3 Выборочные коэффициенты корреляции

#### 2.3.1 Выборочный коэффициент корреляции Пирсона:

*Выборочный коэффициент корреляции Пирсона:*

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x}) \sum (y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{K}{s_x s_y} \quad (4)$$

Где  $K, s_x^2, s_y^2$  – выборочные ковариация и дисперсия с.в.  $X$  и  $Y$  [1, с.535]

#### 2.3.2 выборочный квадратный коэффициент корреляции

*Выборочный квадратный коэффициент корреляции*

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \quad (5)$$

Где  $n_1, n_2, n_3$  и  $n_4$  - количества точек с координатами  $(x_i, y_i)$ , попавшими соответственно в  $I, II, III, IV$  квадраты декартовой системы с осями  $x' = x - medx, y' = y - medy$  и с центром в точке с координатами  $(medx, medy)$  [1, с.539]

### 2.3.3 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:

$$r_s = \frac{\frac{1}{n} \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum (u_i - \bar{u})^2 \frac{1}{n} \sum (v_i - \bar{v})^2}} \quad (7)$$

Где  $\bar{u} = \bar{v} = \frac{1+2+\dots+n}{n} = \frac{n+1}{2}$  – среднее значение рангов [1,с.540-541].

### 2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость  $xOy$ :

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2\rho \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = const \quad (7)$$

Центр эллипса (7) находится в точке с координатами  $(\bar{x}, \bar{y})$ ; оси симметрии эллипса составляют с осью  $Ox$  углы, определяемые уравнением:

$$tg 2\alpha = \frac{2\rho \sigma_x \sigma_y}{\sigma_x^2 - \sigma_y^2} \quad (8)$$

### 2.5 Простая линейная регрессия

#### 2.5.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, n \quad (9)$$

Где  $x_1, \dots, x_n$  – заданные числа (значение фактора);  $y_1, \dots, y_n$  – наблюдаемые значения отклика;  $\epsilon_1, \dots, \epsilon_n$  – независимые, нормально распределённые  $N(0, \sigma)$  с нулевым математическим ожиданием и одинаковой(неизвестной) дисперсией случайной величины (ненаблюдаемые);  $\beta_0, \beta_1$  – неизвестные параметры, подлежащие оцениванию.

#### 2.5.2 Метод наименьших квадратов

Метод наименьших квадратов (МНК):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \rightarrow \min_{\beta_0, \beta_1} \quad (10)$$

#### 2.5.3 Расчётные формулы для МНК-оценок

МНК-оценки параметров  $\beta_0$  и  $\beta_1$ :

$$\widehat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad (11)$$

$$\overline{\beta}_0 = \bar{y} - \bar{x}\widehat{\beta}_1. \quad (12)$$

## 2.6 Робастные оценки коэффициентов линейной регрессии

Метод наименьших модулей:

$$\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i| \rightarrow \min_{\beta_0, \beta_1} \quad (13)$$

$$\widehat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \quad (14)$$

$$\widehat{\beta}_{0R} = medy - \widehat{\beta}_{1R} medx, \quad (15)$$

$$r_Q = \frac{1}{n} \sum_{i=1}^n \operatorname{sgn}(x_i - medx) \operatorname{sgn}(y_i - medy), \quad (16)$$

$$q_y^* = \frac{y_{(j)} - y_{(l)}}{k_q(n)}, q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)} \quad (17)$$

$$l = \begin{cases} [n/4] + 1 & \text{при } n/4 \text{ дробном} \\ n/4 & \text{при } n/4 \text{ целом.} \end{cases}$$

$$j = n - l + 1$$

$$\operatorname{sgn} z = \begin{cases} 1 & \text{при } z > 0 \\ 0 & \text{при } z = 0 \\ -1 & \text{при } z < 0 \end{cases}$$

Уравнение регрессии здесь имеет вид:

$$y = \widehat{\beta}_{0R} + \widehat{\beta}_{1R}x \quad (18)$$

## 2.7. Метод максимального правдоподобия

$L(x_1, \dots, x_n, \theta)$  – функция правдоподобия (ФП), рассматриваемая как функция неизвестного параметра  $\theta$ :

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta)f(x_2, \theta), \dots, f(x_n, \theta) \quad (19)$$

Оценка максимального правдоподобия:

$$\widehat{\theta}_{\text{МП}} = \operatorname{argmax}_{\theta} L(x_1, \dots, x_n, \theta) \quad (20)$$

Система уравнений правдоподобия (в случае дифференцируемости правдоподобия):

$$\frac{\partial L}{\partial \theta_k} = 0 \quad \text{или} \quad \frac{\partial \ln L}{\partial \theta_k} = 0, k = 1, \dots, m \quad (21)$$

## 2.8 Проверка гипотезы о законе распределения генеральной совокупности.

Метод хи-квадрат

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.

### Правило проверки гипотезы о законе распределения по методу $\chi^2$ :

1. Выбираем уровень значимости  $\alpha$
2. По таблице [3, с.358] 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  $n_i$  попадания элементов выборки в подмножества  $\Delta_i, i = 1, \dots, k$
5. Вычисляем выборочное значение статистики критерия  $\chi^2$ :

$$\chi^2_B = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

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

## 2.9. Доверительные интервалы для параметров нормального распределения

### 2.9.1 Доверительный интервал для математического ожидания $m$ нормального распределения

Дана выборка  $(x_1, x_2, \dots, x_n)$  объёма  $n$  из нормальной генеральной совокупности. На её основе строим выборочное среднее  $\bar{x}$  и выборочное среднее квадратическое отклонение  $s$ . Параметры  $m$  и  $\sigma$  нормального распределения неизвестны.

Доверительный интервал для  $m$  и  $s$  с доверительной вероятностью  $\gamma = 1 - \alpha$

$$P\left(\bar{x} - \frac{s\bar{x}}{\sqrt{n-1}} < m < \bar{x} + \frac{s\bar{x}}{\sqrt{n-1}}\right) = 2F_T(x) - 1 = 1 - \alpha.$$

$$P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha \quad (22)$$

### 2.9.2 Доверительный интервал для среднего квадратического отклонения $\sigma$ нормального распределения

Дана выборка  $(x_1, x_2, \dots, x_n)$  объёма  $n$  из нормальной генеральной совокупности. На её основе строим выборочную дисперсию  $s^2$ . Параметры  $m$  и  $\sigma$  нормального распределения неизвестны.

Задаёмся уровнем значимости  $\alpha$ .

Доверительный интервал для  $\sigma$  с доверительной вероятностью  $\gamma = 1 - \alpha$

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1 - \alpha, \quad (23)$$



## 2.10 Доверительные интервалы для математического ожидания $m$ и среднего квадратического отклонения $\sigma$ произвольного распределения при большом объеме выборки. Асимптотический подход

При большом объеме выборки для построения доверительных интервалов может быть использован асимптотический метод на основе центральной предельной теоремы.

### 2.10.1 Доверительный интервал для математического ожидания $m$ произвольной генеральной совокупности при большой объеме выборки

Предполагаем, что исследуемое генеральное распределение имеет конечные математическое ожидание  $m$  и дисперсию  $\sigma^2$ .

$u_{1-\alpha/2}$  – квантиль нормального распределения  $N(0,1)$  порядка  $1 - \alpha/2$ .

Доверительный интервал для  $m$  с доверительной вероятностью  $\gamma = 1 - \alpha$ :

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} + \frac{su_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma \quad (24)$$

### 2.10.2 Доверительный интервал для среднего квадратического отклонения $\sigma$ произвольной генеральной совокупности при большом объеме выборки

Предполагаем, что исследуемая генеральная совокупность имеет конечные первые четыре момента.

$u_{1-\alpha/2}$  – квантиль нормального распределения  $N(0,1)$  порядка  $1 - \alpha/2$

$E = \frac{\mu_4}{\sigma^4} - 3$  – эксцесс генерального распределения,  $e = \frac{m_4}{s^4} - 3$  – выборочный эксцесс;

$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$  – четвертый выборочный центральный момент.

$$s(1 + U)^{-1/2} < \sigma < s(1 - U)^{-1/2} \quad (25)$$

Или

$$s(1 - 0.5U) < \sigma < s(1 + 0.5U) \quad (26)$$

Где  $U = u_{1-\alpha/2} \sqrt{(e + 2)/n}$

Формулы (25) или (26) дают доверительный интервал для  $\sigma$  с доверительной вероятностью  $\gamma = 1 - \alpha$  [1, с.461 – 462]

**Замечание:** Вычисление по формуле (25) дают более надёжный результат так как в ней меньше грубых приближений.

## 3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки jupyter notebook. Исходный код лабораторной работы приведен в приложении

## 4 Результаты

### 4.1 Выборочные коэффициенты корреляции

$\rho = 0$	$r$	$r_s$	$r_Q$
$E(z)$	0.00	0.00	-0.01
$E(z^2)$	0.05	0.05	0.05
$D(z)$	0.05	0.05	0.05
$\rho = 0.5$	$r$	$r_s$	$r_Q$
$E(z)$	0.50	0.50	0.33
$E(z^2)$	0.28	0.26	0.16
$D(z)$	0.03	0.04	0.05
$\rho = 0.9$	$r$	$r_s$	$r_Q$
$E(z)$	0.90	0.87	0.70
$E(z^2)$	0.81	0.76	0.51
$D(z)$	0.00	0.00	0.03

Таблица 1 Двумерное нормальное распределение,  $n = 20$

$\rho = 0$	$r$	$r_s$	$r_Q$
$E(z)$	0.00	-0.01	0.00
$E(z^2)$	0.02	0.02	0.02
$D(z)$	0.02	0.02	0.02
$\rho = 0.5$	$r$	$r_s$	$r_Q$
$E(z)$	0.50	0.48	0.33
$E(z^2)$	0.26	0.24	0.12
$D(z)$	0.01	0.01	0.01
$\rho = 0.9$	$r$	$r_s$	$r_Q$
$E(z)$	0.90	0.88	0.71
$E(z^2)$	0.81	0.78	0.51
$D(z)$	0.00	0.00	0.01

Таблица 2 Двумерное нормальное распределение,  $n = 60$

$\rho = 0$	$r$	$r_S$	$r_Q$
$E(z)$	0.00	0.00	0.00
$E(z^2)$	0.01	0.01	0.01
$D(z)$	0.01	0.01	0.01
$\rho = 0.5$	$r$	$r_S$	$r_Q$
$E(z)$	0.50	0.48	0.34
$E(z^2)$	0.26	0.24	0.12
$D(z)$	0.01	0.01	0.01
$\rho = 0.9$	$r$	$r_S$	$r_Q$
$E(z)$	0.90	0.89	0.71
$E(z^2)$	0.81	0.79	0.51
$D(z)$	0.00	0.00	0.00

Таблица 3 Двумерное нормальное распределение,  $n = 100$

$n = 20$	$r$	$r_S$	$r_Q$
$E(z)$	-0.08	-0.08	-0.06
$E(z^2)$	0.06	0.06	0.06
$D(z)$	0.06	0.06	0.05
$n = 60$	$r$	$r_S$	$r_Q$
$E(z)$	-0.10	-0.09	-0.06
$E(z^2)$	0.26	0.03	0.02
$D(z)$	0.02	0.02	0.02
$n = 100$	$r$	$r_S$	$r_Q$
$E(z)$	-0.10	-0.09	-0.06
$E(z^2)$	0.02	0.02	0.01
$D(z)$	0.01	0.01	0.01

Таблица 4 Смесь нормальных распределений

## 4.2 Эллипсы рассеивания

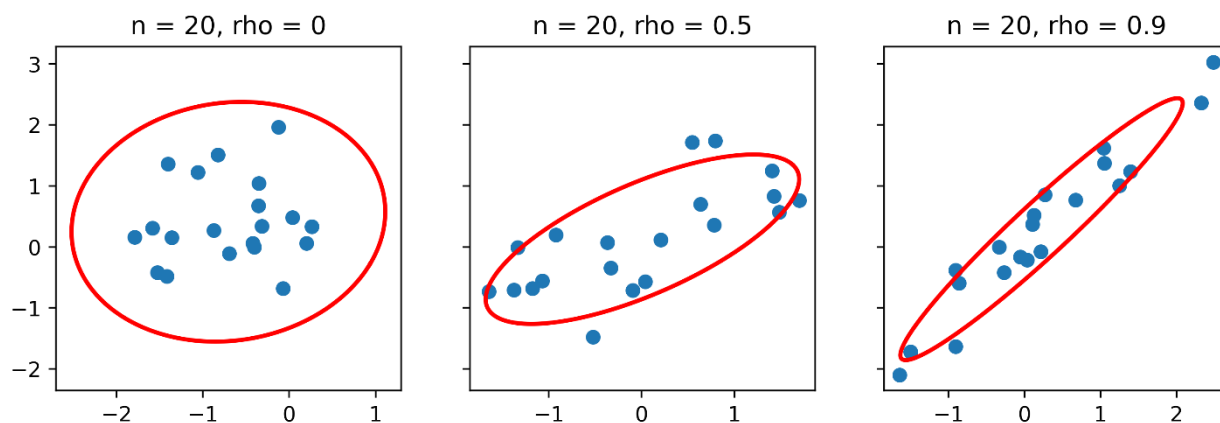


Рисунок 1 Двумерное нормальное распределение  $n = 20$

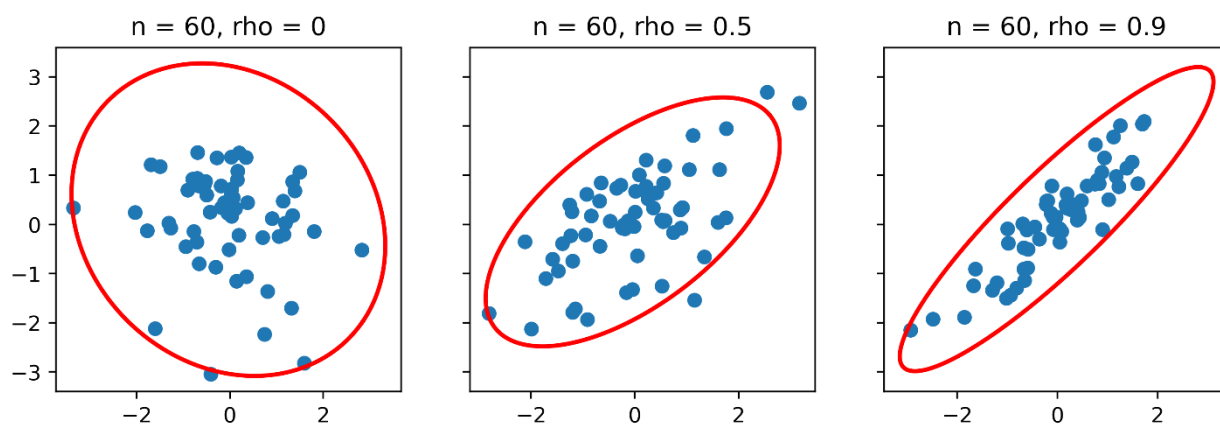


Рисунок 2 Двумерное нормальное распределение,  $n = 60$

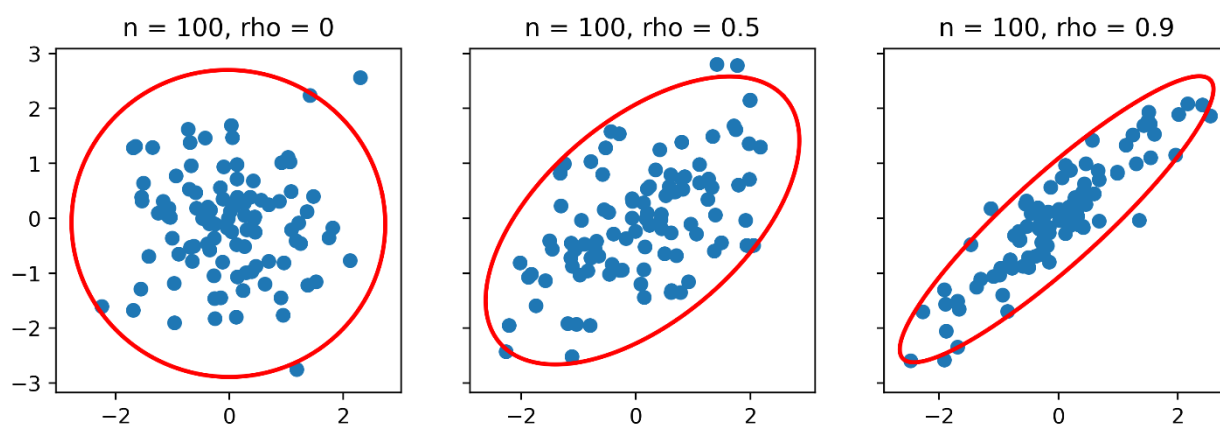


Рисунок 3 Двумерное нормальное распределение,  $n = 100$

## 4.3 Оценка коэффициентов линейной регрессии

### 4.3.1 Выборка без возмущений

- Критерий наименьших квадратов:

$$\hat{a} \approx 2.10, \hat{b} \approx 1.96$$

- Критерий наименьших модулей:

$$\hat{a} \approx 2.52, \hat{b} \approx 1.74$$

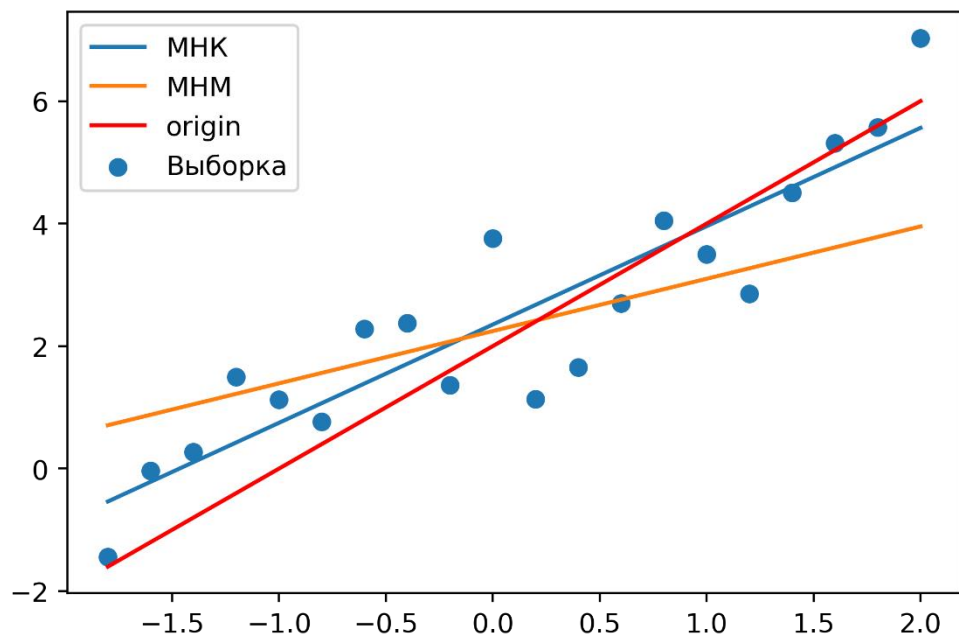


Рисунок 4 Выборка без возмущений

Критерий оптимальности:

$$\text{dist МНК} = 57.5$$

$$\text{dist МНМ} = 322.8$$

#### 4.3.2 выборка с возмущениями

- Критерий наименьших квадратов:

$$\hat{a} \approx 1.76, \hat{b} \approx 0.67$$

- Критерий наименьших модулей:

$$\hat{a} \approx 1.70, \hat{b} \approx 1.04$$

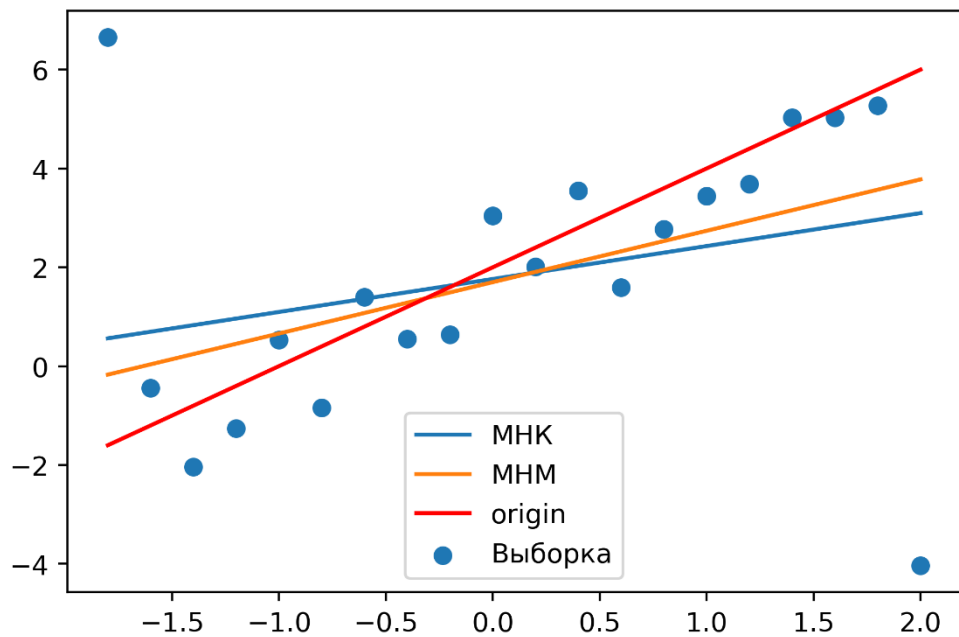


Рисунок 5 Выборка с возмущениями

Критерий оптимальности:

$$dist \text{ МНК} = 459.57$$

$$dist \text{ МНМ} = 255.80$$

#### 4.4 Проверка гипотезы о законе распределения генеральной совокупности.

##### Метод хи-квадрат

Метод максимального правдоподобия:

$$\hat{\mu} \approx -0.018, \hat{\sigma} \approx 1.01$$

Критерий согласия  $\chi^2$ :

- Критерий промежутков  $k = 6$
- Уровень значимости  $\alpha = 0.05$ .
- Тогда квантиль из таблицы [3, с.358]  $\chi_{1-\alpha}^2(k-1) = \chi_{0.95}^2$

i	Границы, $\Delta_i$ $a_{i-1},$ $a_i$	$n_i$	$p_i$	$np_i$	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$-\infty, -2.45$	1	0.01	0.82	-0.81	0.04
2	$-2.45, -1.30$	11	0.09	9.21	-9.12	0.35
3	$-1.30, -0.15$	34	0.34	33.70	-33.37	0.00
4	$-0.15, 1.00$	34	0.40	39.53	-39.13	0.77
5	$1.00, 2.14$	20	0.15	14.90	-14.75	1.75
6	$2.14, +\infty$	0	0.02	1.85	-1.83	1.85
$\Sigma$	-	100	1.00	100.00	0	4.76

Таблица 5 Вычисление хи-квадрат при проверке гипотезы  $H_0$  о нормальном законе распределения  $N(x, \hat{\mu}, \hat{\sigma})$

Сравним  $\chi_{0.95}^2(5) \approx 11.07$  и найденное  $\chi_B^2 \approx 4.76 : 11.07 > 4.76$ . Следовательно, гипотезу  $H_0^*$

Можно принять на данном этапе

##### Исследование на чувствительность

Рассмотрим гипотезу  $H_0^*$ , что выборка распределена согласно закону  $Laplace\left(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}}\right)$

Используем критерий согласия  $\chi^2$ :

- $\alpha = 0.05$  – уровень значимости,
- $n = 20$  – размер выборки,
- $k := [1 + 3.3lg20] = [5.3] = 5$  – количество промежутков
- Квантиль  $\chi_{1-\alpha}^2 = \chi_{0.95}^2(4) \approx 9.49$  из таблицы [3, с.358]

i	Границы, $\Delta_i$ $a_{i-1},$ $a_i$	$n_i$	$p_i$	$np_i$	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$-\infty, -2.45$	1	0.05	4.53	-4.48	2.76
2	$-2.45, -0.92$	15	0.16	15.68	-15.53	0.03
3	$-0.92, 0.61$	54	0.52	52.08	-51.56	0.07
4	$0.61, 2.15$	30	0.22	21.50	-21.29	3.36
5	$2.15 + \infty$	0	0.06	6.21	-6.14	6.21
$\Sigma$	-	100	1.00	100.00	-99	12.41

Таблица 6 Вычисление хи-квадрат при проверке гипотезы  $H_0^*$  о  $Laplace\left(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}}\right)$

Сравним  $\chi_{0.95}^2(4) \approx 9.49$  и найденное  $\chi_B^2 \approx 12.41$ :  $9.49 < 12.41$ . Следовательно гипотеза  $H_0^*$  на данном этапе проверки отвергается

Для равномерного распределения:  $U(0,1)$

Рассмотрим гипотезу  $H_0^*$ , что выборка распределена согласно закону  $U(0,1)$

При  $k = 5$

$\chi_{0.95}^2 \approx 515037.54$ . Следовательно, гипотеза  $H_0^*$  на данном этапе проверки отвергается

#### 4.5 Доверительные интервалы для параметров нормального распределения

n = 20	m	$\sigma$
	$-0.87 < m < 0.39$	$0.74 < \sigma < 2.44$
n = 100	$-0.07 < m < 0.33$	$0.90 < \sigma < 3.47$

Таблица 7 Доверительные интервалы для параметров нормального распределения.



#### 4.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход.

n = 20	m	$\sigma$
	$-0.50 < m < 0.15$	$0.75 < \sigma < 1.25$
n = 100	$-0.26 < m < 0.12$	$0.84 < \sigma < 1.11$

Таблица 8 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

## 5 Обсуждение

### 5.1 Выборочные коэффициенты корреляции и эллипсы рассеяния

Сравним дисперсии выборочных коэффициентов корреляции

- Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом:  $r < r_S < r_Q$
- Для смеси нормальных распределений дисперсии выборочных коэффициентов корреляции упорядочены следующим образом:  $r_Q < r_S < r$

Процент попавших элементов выборки в эллипс рассеивания (95-% доверительная область) примерно равен его теоретическому значению

### 5.2 Оценки коэффициентов линейной регрессии

Критерий квадратов точнее оценивает коэффициенты линейной регрессии на выборке без возмущений

Критерий наименьших модулей точнее оценивает коэффициенты линейной регрессии на выборке с возмущениями.

Критерий наименьших модулей устойчив к редким выбросам

### 5.3 Проверка гипотезы о законе распределения генеральной совокупности.

#### Метод Хи-квадрат

- По результатам проверки на близость с помощью критерия хи-квадрат можно принять гипотезу  $H_0^*$  о нормальном распределении  $N(x, \hat{\mu}, \hat{\sigma})$  на уровне значимости  $\alpha = 0.05$  для выборки, сгенерированной согласно  $N(x, 0, 1)$ .
- Видим так же, что критерий не принял гипотезу о том, что 20-элементная выборка, сгенерированная согласно  $N(x, 0, 1)$  описывается законом распределения  $Laplace\left(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}}\right)$
- По исследованию на чувствительность видим, что при небольших объемах выборки уверенности в полученных результатах нет, критерий может ошибаться. Это обусловлено тем, что теорема Пирсона говорит про асимптотическое распределение, а при малых размерах выборки результат не будет получаться достоверным.

## **Список литературы**

*Вентцель Е.С. Теория вероятностей: Учеб. для вузов. — 6-е изд. стер. . (1999).*

*Вероятностные разделы математики. Учебник для бакалавров технических направлений.//Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров»,. (2001.).*

*Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей : учеб. пособие /. (2009).*