Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и вычислительной физики, ФизМех

Направление подготовки
«01.03.02 Прикладная математика и информатика»

Отчет по лабораторной работе № 2
дисциплина "Математическая статистика"

Выполнил студент гр. 5030102/00201 Трусов Н. А.
Преподаватель: Баженов А.Н.

Санкт-Петербург

2023

Содержание

Список иллюстраций

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 ρ 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:

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

   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 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. Проделать 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 μ и σ 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, использзу критерий согласия χ^2. В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2.

   Исследовать точность (чувствительность) критерия χ^2- сгенерировать выборки равномерного распределения и распределения Лапласа малого объема (например, 20 элементов). Проверить их на нормальность.

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 χ^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

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

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

Параметр ρ называется коэффициентом корреляции.

2.2 Корреляционный момент (ковариация) и коэффициент корреляции

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

$$K = cov(X, Y) = M[(X - \overline{x})(Y - \overline{y})]. \tag{2}$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \tag{3}$$

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n}\sum(x_i - \overline{x})^2\frac{1}{n}\sum(y_i - \overline{y})^2}} = \frac{K}{s_X s_Y}, \tag{4}$$

где K, s_X^2, s_Y^2 - выборочные ковариация и дисперсии с.в. X и Y [1, с. 535]

2.3.2 Выборочный квадратный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

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

где $n_1, n_2, n_3 n_4$ - количества точек с координатами (x_i, y_i), попавшими соответственно в $I, II, IIIIV$ квадранты 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 коэффицент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной X, через u, а ранги, соответствующие значениям переменной Y, - через v.
Выборочный коэффициент ранговой корреляции Спирмена:

$$r_S = \frac{\frac{1}{n}\sum(u_i - \overline{u})(v_i - \overline{v})}{\sqrt{\frac{1}{n}(u_i - \overline{u})^2 \frac{1}{n}\sum(v_i - \overline{v})^2}}, \tag{6}$$

где $\overline{u} = \overline{v} = \frac{1+2+...+n}{n} = \frac{n+1}{2}$ - среднее значение рангов [1, с. 540-541].

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

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

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

$$tg\ 2\alpha = \frac{2\rho\ \sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{8}$$

2.5 Простая линейная регрессия

2.5.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, ..., n, \tag{9}$$

где $x_1, ..., x_n$ - заданные числа (значения фактора); $y_1, ..., y_n$ - наблюдаемые значения отклика; $\epsilon_1, ..., \epsilon_n$ - независимые, нормально распределенные $N(0, \sigma)$ с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины (ненаблюдаемые); β_0, β_1 - неизвестные параметры, подлежащие оцениванию.

2.5.2 Метод наименьших квадратов

Метод наименьших квадратов (МНК):

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

2.5.3 Расчетные формулы для МНК-оценок

МНК-оценки параметров β_0 и β_1:

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

$$\hat{\beta}_0 = \overline{y} - \overline{x}\hat{\beta}_1. \tag{12}$$

2.6 Робастные оценки коэффицентов линейной регрессии

Метод наименьших модулей:

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \to \min_{\beta_0, \beta_1}. \tag{13}$$

$$\hat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \tag{14}$$

$$\hat{\beta}_{0R} = med\ y - \hat{\beta}_{1R} med\ x, \tag{15}$$

$$r_Q = \frac{1}{n} \sum_{i=1} n sgn(x_i - med\ x) sgn(y_i - med\ y), \tag{16}$$

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

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

$$l = n - l + 1$$

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

Уравнение регрессии здесь имеет вид

$$y = \hat{\beta}_{0R} + \hat{\beta}_{1R} x. \tag{18}$$

2.7 Метод максимального правдоподобия

$L(x_1, ..., x_n, \theta)$ - функция правдоподобия (ФП), рассматриваемая как функция неизвестного параметра θ:

$$L(x_1, ..., x_n, \theta) = f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta) \tag{19}$$

Оценка максимального правдоподобия:

$$\hat{\theta} = ard \max_{\theta} L(x_1, ..., x_n, \theta). \tag{20}$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial L}{\partial \theta_k} = 0 \text{ или } \frac{\partial lnL}{\partial \theta_k} = 0,\ k = 1, ..., m. \tag{21}$$

2.8 Проверка гипотезы о законе распределения генеральной совокупности. Метод хм-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 χ^2:

1. Выбираем уровень значимости α
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 $p_i = P(X \in \Delta_i), i = 1, ..., k$.
4. Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, ..., k$.
5. Вычисляем выборочное значение статистики критерия χ^2:

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

6. Сравниваем χ^2_B и квантиль $\chi^2_{1-\alpha}(k-1)$.

   (a) Если $\chi^2_B < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.

   (b) Если $\chi^2_B \geq \chi^2_{1-\alpha}(k-1)$, 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 $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочное среднее $\overline{x}$ и выборочное среднее квадратическое отклонение s. Параметры m и σ нормального распределения неизвестны. Доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$:

$$P\left(\overline{x} - \frac{sx}{\sqrt{n-1}} < m < \overline{x} + \frac{sx}{\sqrt{n-1}}\right) = 2F_T(x) - 1 = 1 - \alpha, \tag{22}$$

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

2.9.2 Доверительный интервал для среднего квадратического отклонения σ нормального распределения

Дана выборка $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочную дисперсию s^2. Параметры m и σ нормального распределения неизвестны.
Задаёмся уровнем значимости α.
Доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$:

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

2.10 Доверительные интервалы для математического ожидания m и среднего квадратического отклонения σ произвольного распределения при большом объеье выборки. Асимптотический подход

При большом объёме 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 большом объеме выборки

Предполагаем, что исследуемое генеральное распределение имеет конечные математическое ожидание m и дисперсию σ^2.
$u_{1-\alpha/2}$ - квантиль нормального распределения $N(0,1)$ порядка $1-\alpha/2$.
Доверительный интервал для m с доверительной вероятностью $\gamma = 1-\alpha$:

$$P\left(\overline{x} - \frac{s u_{1-\alpha/2}}{\sqrt{n}} < m < \overline{x} + \frac{s u_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma \tag{24}$$

2.10.2 Довверительный интервал для среднего квадратического отклонения σ 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; $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4$ - четвёртый выборочный центральный момент.

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

или

$$s(1-0.5U) < \sigma < s(1+0.5U) \tag{26}$$

где $U = u_{1-\alpha/2}\sqrt{(e+2)/n}$
Формулы (25) или (26) дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1-\alpha$ [1, с. 461-462].
Замечание. Вычисления по формуле (25) дают более надёжный результат, так как в ней меньше грубых приближений.

3 Результаты

3.1 Выборочные коэффициенты корреляции

$\rho = 0$	$r(4)$	$r_S(6)$	$r_Q(5)$
E(z)	0.007	0.006	0.0
E(zˆ2)	0.025	0.026	0.04
D(z)	0.053	0.054	0.052
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.509	0.481	0.4
E(zˆ2)	0.259	0.232	0.16
D(z)	0.031	0.035	0.05
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.905	0.878	0.8
E(zˆ2)	0.819	0.771	0.64
D(z)	0.003	0.005	0.026

Таблица 1: Двумерное нормальное распределение, n = 20

$\rho = 0$	r	r_S	r_Q
E(z)	0.002	0.006	0.0
E(zˆ2)	0.008	0.008	0.004
D(z)	0.017	0.016	0.016
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.509	0.486	0.333
E(zˆ2)	0.26	0.236	0.111
D(z)	0.01	0.011	0.015
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.901	0.886	0.733
E(zˆ2)	0.812	0.785	0.538
D(z)	0.001	0.001	0.009

Таблица 2: Двумерное нормальное распределение, n = 60

$\rho = 0$	r	r_S	r_Q
E(z)	0.002	0.003	0.0
E(zˆ2)	0.005	0.005	0.006
D(z)	0.01	0.01	0.011
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.5	0.481	0.32
E(zˆ2)	0.25	0.232	0.102
D(z)	0.006	0.007	0.009
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.902	0.889	0.72
E(zˆ2)	0.813	0.791	0.518
D(z)	0.0	0.001	0.005

Таблица 3: Двумерное нормальное распределение, n = 100

$size = 20$	r	r_S	r_Q
E(z)	0.8	0.878	0.6
E(zˆ2)	0.639	0.771	0.36
D(z)	0.008	0.004	0.037
$size = 60$	r	r_S	r_Q
E(z)	0.796	0.886	0.6
E(zˆ2)	0.634	0.785	0.36
D(z)	0.002	0.001	0.011
$size = 100$	r	r_S	r_Q
E(z)	0.792	0.889	0.56
E(zˆ2)	0.627	0.791	0.314
D(z)	0.001	0.001	0.006

Таблица 4: Смесь нормальных распределений

3.2 Эллипсы рассеивания

Для уравнения эллипса выбиралась константа равная $const = 2 \cdot (2 \cdot \sigma)$
$\sigma_x = \sigma_y = 1$

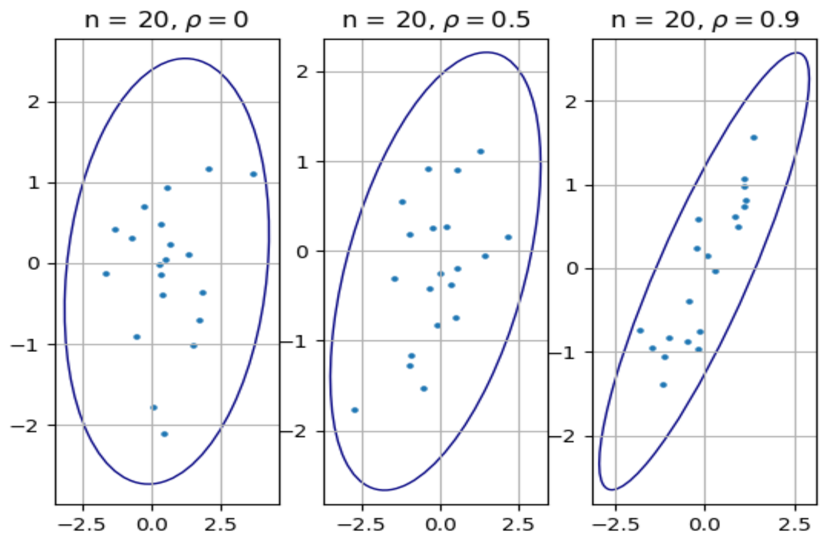

Рис. 1: Двумерное нормальное распределение, $n = 20$

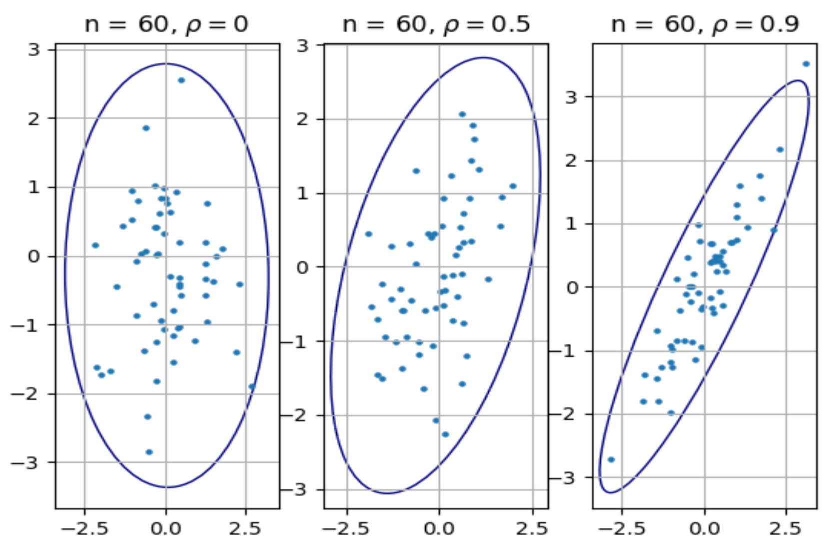

Рис. 2: Двумерное нормальное распределение, $n = 60$

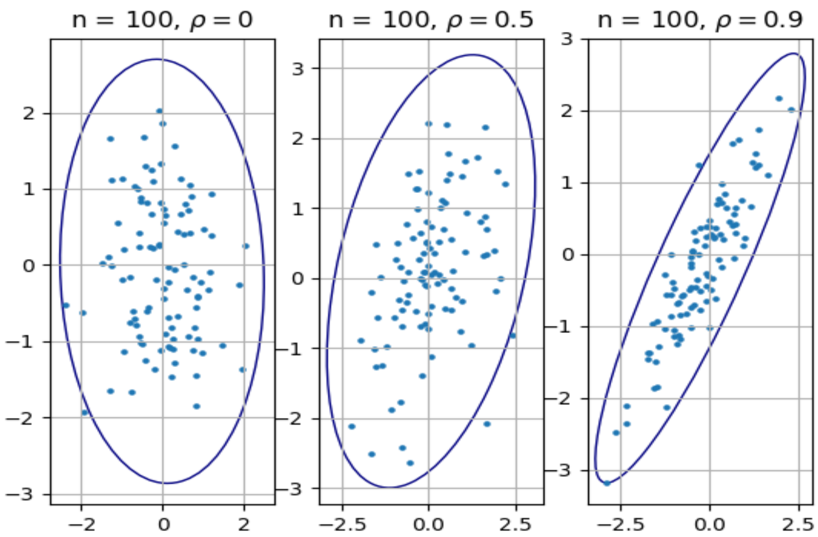

Рис. 3: Двумерное нормальное распределение, $n = 100$

3.3 Оценки коэффициентов линейной регрессии

Метрика удаленности:

$$\text{distance} = \sum_{i=0}^{n}(y_{model}[i] - y_{regr}[i])^2 \ (27)$$

3.3.1 Выборка без возмущений

1. Критерий наименьших квадратов: $\hat{a} \approx 2.11$, $\hat{b} \approx 2.31$
2. Критерий наименьших модулей: $\hat{a} \approx 1.90$, $\hat{b} \approx 2.27$
   МНК $distance = 16.18$
   МНМ $distance = 14.90$

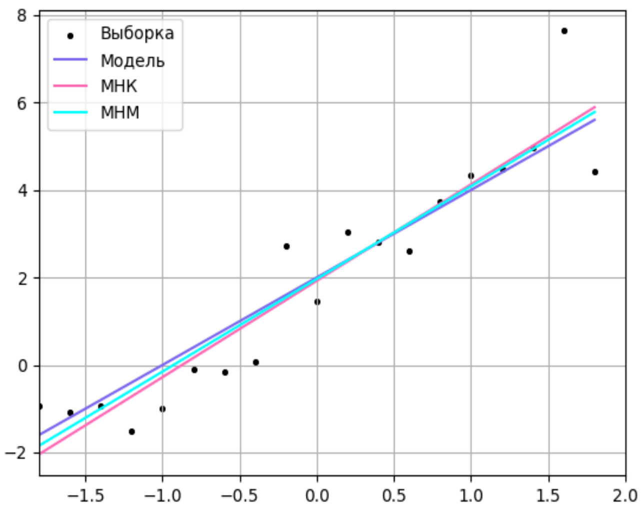

Рис. 4: Выборка без возмущений

3.3.2 Выборка с возмущениями

1. Критерий наименьших квадратов: $\hat{a} \approx 2.07$, $\hat{b} \approx 0.56$
2. Критерий наименьших модулей: $\hat{a} \approx 1.82$, $\hat{b} \approx 1.94$
   МНК $distance = 176.38$
   МНМ $distance = 30.19$

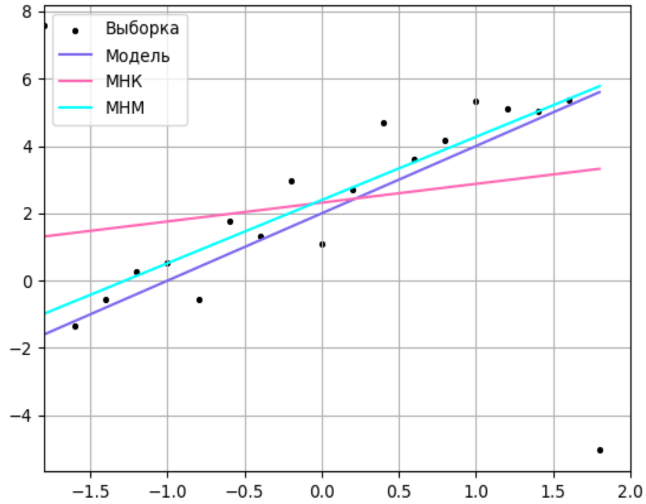

Рис. 5: Выборка с возмущениями

МНМ	$\hat{a}$	$\hat{a}'$	$\hat{b}$	$\hat{b}'$	$\dfrac{\hat{a}-\hat{a}'}{\hat{a}}$	$\dfrac{\hat{b}-\hat{b}'}{\hat{b}}$
	1.90	1.82	2.27	1.94	$\dfrac{1.90-1.82}{1.90} \approx 0.042$	$\dfrac{2.27-1.94}{2.27} \approx 0.15$
МНК	$\hat{a}$	$\hat{a}'$	$\hat{b}$	$\hat{b}'$	$\dfrac{\hat{a}-\hat{a}'}{\hat{a}}$	$\dfrac{\hat{b}-\hat{b}'}{\hat{b}}$
	2.11	2.07	2.31	0.56	$\dfrac{2.11-2.07}{2.11} \approx 0.019$	$\dfrac{2.31-0.56}{2.31} \approx 0.75$

Таблица 5: Оценка относительного отклонения параметров a и b

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

Метод максимального правдоподобия:

$$\hat{\mu} \approx -0.08, \hat{\sigma} \approx 1.07$$

Критерий согласия χ^2:
Количество промежутков $k = 8$

Уровень значимости α= 0.05
Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(7)$. Из таблицы [3, с. 358] $\chi^2_{0.95}(7) \approx 14.06$.

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.1]	12	0.1357	13.57	-1.57	0.18
2	[-1.1, -0.73]	16	0.096	9.6	6.4	4.26
3	[-0.73, -0.37]	12	0.1253	12.53	-0.53	0.02
4	[-0.37, 0.0]	13	0.1431	14.31	-1.31	0.12
5	[0.0, 0.37]	14	0.1431	14.31	-0.31	0.01
6	[0.37, 0.73]	8	0.1253	12.53	-4.53	1.64
7	[0.73, 1.1]	10	0.096	9.6	0.4	0.02
8	[1.1, ∞]	15	0.1357	13.57	1.43	0.15
$\sum$	-	100	1	100	-0	6.4

Таблица 6: Вычисление χ^2_B при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$

Сравнивая $\chi^2_B = 6.4$ и $\chi^2_{0.95}(7) \approx 14.06$, видим, что $\chi^2_B < \chi^2_{0.95}(7)$.

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

$$\hat{\mu} \approx -0.16, \hat{\sigma} \approx 0.93$$

Количество промежутков $k = 5$
Уровень значимости α= 0.05
$\chi^2_{0.95}(7) \approx 9.49$

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.1]	4	0.1357	2.71	1.29	0.61
2	[-1.1, -0.37]	4	0.2213	4.43	-0.43	0.04
3	[-0.37, 0.37]	5	0.2861	5.72	-0.72	0.09
4	[0.37, 1.1]	5	0.2213	4.43	0.57	0.07
5	[1.1, ∞]	2	0.1357	2.71	-0.71	0.19
$\sum$	-	20	1	20	-0	1

Таблица 7: Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $L(x, \hat{\mu}, \hat{\sigma})$, $n = 20$

Сравнивая $\chi^2_B = 1$ и $\chi^2_{0.95}(4) \approx 9.49$, видим, что $\chi^2_B < \chi^2_{0.95}(4)$.

Проведём аналогичный анализ для равномерного распределения

$$\hat{\mu} \approx -0.28, \hat{\sigma} \approx 1.0$$

Количество промежутков $k = 5$
Уровень значимости $\alpha = 0.05$
$\chi^2_{0.95}(7) \approx 9.49$

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[$-\infty'$, -1.1]	6	0.1357	2.71	3.29	3.98
2	[-1.1, -0.37]	4	0.2213	4.43	-0.43	0.04
3	[-0.37, 0.37]	5	0.2861	5.72	-0.72	0.09
4	[0.37, 1.1]	3	0.2213	4.43	-1.43	0.46
5	[1.1, ∞]	2	0.1357	2.71	-0.71	0.19
$\sum$	-	20	1	20	-0	4.76

Таблица 8: Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $U(x, \hat{\mu}, \hat{\sigma})$, $n = 20$

Сравнивая $\chi^2_B = 4.76$ и $\chi^2_{0.95}(4) \approx 9.49$, видим, что $\chi^2_B < \chi^2_{0.95}(4)$.

3.5 Доверительные интервалы для параметров нормального распределения

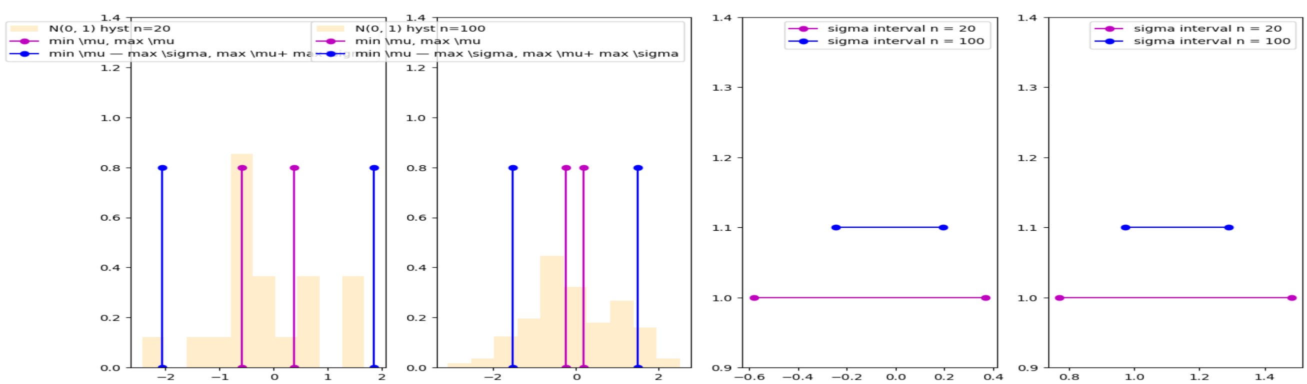

Рис. 6: Гистограммы нормальных распределений и доверительные интервалы их параметров

n = 20	m	σ	x twin 20
	-0.27$< m <$ 0.55	0.67 $< \sigma <$ 1.28	[[-0.27, 0.55][-1.55, 1.83]]
n = 100	m	σ	x twin 100
	-0.29 $< m <$ 0.08	0.81 $< \sigma <$ 1.07	[[-0.29, 0.08][-1.36, 1.15]]

Таблица 9: Доверительные интервалы для параметров нормального распределения

3.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

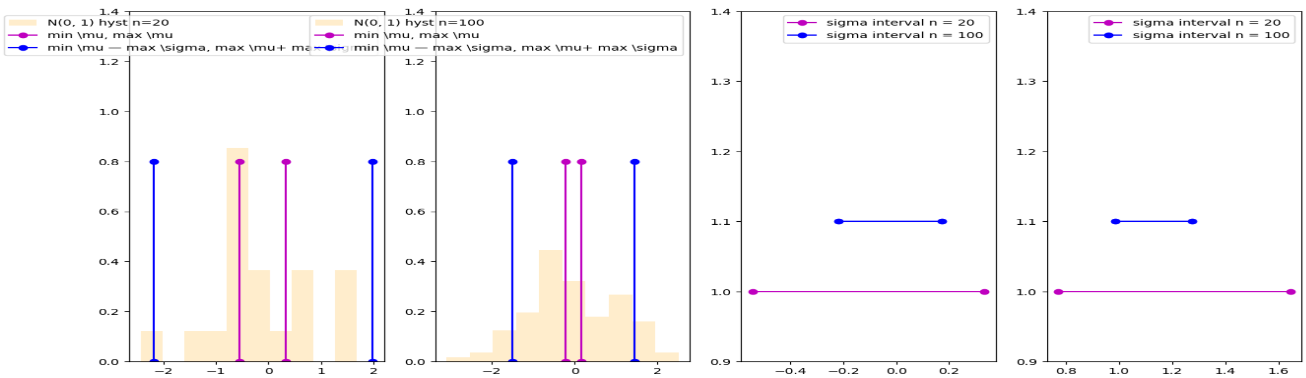

Рис. 7: Гистограммы нормальных распределений и доверительные интервалы их параметров. Асимптотический подход

n = 20	m	σ	x twin 20
	-0.30 $< m <$ 0.58	0.68 $< \sigma <$ 1.33	[[-0.30, 0.58][-1.63, 1.90]]
n = 100	m	σ	x twin 100
	-0.30$< m <$ 0.09	0.81 $< \sigma <$ 1.08	[[-0.30, 0.09][-1.38, 1.18]]

Таблица 10: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

4 Обсуждение

4.1 Выборочные коэффициенты корреляции

Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r < r_S < r_Q$; для смеси распределений получили обратную картину: $r_Q < r_S < r$.
Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению (95%).

4.2 Оценки коэффициентов линейной регрессии

Из полученных результатов(см. метрику удаленности модельной прямой от теоретической - *distance* и таблицу оценки относительного отклонения параметров a и b) можно сделать несколько выводов.

Критерий наименьших квадратов точнее оценивает коэффициенты линейной регрессии на выборке без возмущений. Относительное отклоение параметра $a = 0.019$, относительное отклонение параметра $b = 0.75$.

Критерий наименьших модулей точнее оценивает коэффициенты линейной регрессии на выборке с возмущениями. Относительное отклоение параметра $a = 0.042$, относительное отклонение параметра $b = 0.15$. Критерий наименьших модулей устойчив к редким выбросам, т.к. с возмущениями произошло незначительное изменение метрики удаленности

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

Заключаем, что гипотеза H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$ согласуется с выборкой для нормального распределения $N(x, 0, 1)$.

Также видно, что гипотеза H_0 оказалась принята для выборок сгенерированных по равномерному закону и закону Лапласа.

По исследованию на чувствительность видим, что при небольших объемах выборки уверенности в полученных результатах нет, критерий может ошибиться. Это обусловлено тем, что теорема Пирсона говорит про асимптотическое распределение, а при малых размерах выборки результат не будет получаться достоверным

4.4 Доверительные интервалы для параметров распределения

- Генеральные характеристики ($m = 0$ и $\sigma = 1$) накрываются построенными доверительными интервалами.
- Также можно сделать вывод, что для большей выборки доверительные интервалы являются соответственно более точными, т.е. меньшими по длине.
- Кроме того, при большом объеме выборки асимптотические и классические оценки практически совпадают.

5 Литература

Язык программирования *Python* 3.11
Подключенные библиотеки для *Python*: *numpy*, *math*, *seaborn*, *matplotlib*, *scipy*, *tabulate*
Ссылка на Github: `https://github.com/iMaanick/MathStatLab5-8`