Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики Кафедра «Прикладная математика»

Отчет по лабораторным работам 5-8
по дисциплине
«Математическая статистика»

Выполнил студент:
Файзрахманов А. Р.
группа: 5030102/90201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург, 2022

Содержание

1 Постановка задачи **4**

2 Теория **5**

2.1 Двумерное нормальное распределение 5
2.2 Корреляционный момент (ковариация) и коэффициент корреляции 5
2.3 Выборочные коэффициенты корреляции 5
2.3.1 Выборочный коэффициент корреляции Пирсона 5
2.3.2 Выборочный квадратный коэффициент корреляции 5
2.3.3 Выборочный коэффицент ранговой корреляции Спирмена 6
2.4 Эллипсы рассеивания 6
2.5 Простая линейная регрессия 6
2.5.1 Модель простой линейной регрессии 6
2.5.2 Метод наименьших квадратов 6
2.5.3 Расчетные формулы для МНК-оценок 6
2.6 Робастные оценки коэффицентов линейной регрессии 7
2.7 Метод максимального правдоподобия 7
2.8 Проверка гипотезы о законе распределения генеральной совокупности. Метод хм-квадрат 7
2.9 Доверительные интервалы для параметров нормального распределения 8
2.9.1 Доверительный интервал для математического ожидания m нормального распределения 8
2.9.2 Доверительный интервал для среднего квадратического отклонения σ нормального распределения 8
2.10 Доверительные интервалы для математического ожидания m и среднего квадратического отклонения σ произвольного распределения при большом объеье выборки. Асимптотический подход 8
2.10.1 Доверительный интервал для математического ожидания m произвольной генеральной совокупности при большом объеме выборки 9
2.10.2 Довверительный интервал для среднего квадратического отклонения σ произвольной генеральной совокупности при большом объеме выборки 9

3 Реализация **10**

4 Результаты **10**

4.1 Выборочные коэффициенты корреляции 10
4.2 Эллипсы рассеивания 11
4.3 Оценки коэффициентов линейной регрессии 12
4.3.1 Выборка без возмущений 12
4.3.2 Выборка с возмущениями 13
4.4 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат 14
4.5 Доверительные интервалы для параметров нормального распределения 15
4.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход 15

5 Обсуждение **16**

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания 16
5.2 Оценки коэффициентов линейной регрессии 16
5.3 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат 16
5.4 Доверительные интервалы для параметров распределения 16

Литература **17**

Список иллюстраций

1 Двумерное нормальное распределение, n=20 11
2 Двумерное нормальное распределение, n=60 12
3 Двумерное нормальное распределение, n=100 12
4 Выборка без возмущений 13
5 Выборка с возмущениями 13

Список таблиц

1 Двумерное нормальное распределение, $n = 20$ 10
2 Двумерное нормальное распределение, $n = 60$ 10
3 Двумерное нормальное распределение, $n = 100$ 11
4 Смесь нормальных распределений 11
5 Вычисление χ^2_B при проверке гипотезы H_0 о нормальном законе $N(x, \hat{\mu}, \hat{\sigma})$ . . 14
6 Вычисление χ^2_B при проверке гипотезы H_0^* о $Laplace(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$ 14
7 Вычисление χ^2_B при проверке гипотезы H_0^* о равномерном распределении . . . 15
8 Доверительные интервалы для параметров нормального распределения 15
9 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход 15

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 регресии $y_i = a + bx_i + e_i$, используя 20 точек на отрезке $[-1.8; 2]$ с равномерным шагом равным 0.2. Ошибку e_i считать нормально распределенной 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 значении 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, используя 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):

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

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:

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

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

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

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

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

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

$$l = n - l + 1$$

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

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 $\gamma = 1 - \alpha$:

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

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

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 $\gamma = 1 - \alpha$:

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

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 $\gamma = 1-\alpha$:

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

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 Реализация

Лабораторная работа выполнена на языке программирования Python с помощью библиотек numpy, matplotlib, scipy, statsmodels. Отчет написан в среде разработки TexWorks с помощью pdfLaTeX

4 Результаты

4.1 Выборочные коэффициенты корреляции

p = 0	r	r_S	r_Q
$E(z)$	-0.009	-0.007	0.002
$E(z^2)$	0.051	0.053	0.054
$D(z)$	0.051	0.053	0.054
p = 0.5	r	r_S	r_Q
$E(z)$	0.486	0.456	0.326
$E(z^2)$	0.269	0.243	0.154
$D(z)$	0.032	0.036	0.048
p = 0.9	r	r_S	r_Q
$E(z)$	0.895	0.867	0.694
$E(z^2)$	0.803	0.756	0.509
$D(z)$	0.002	0.005	0.027

Таблица 1: Двумерное нормальное распределение, $n = 20$

p = 0	r	r_S	r_Q
$E(z)$	0.0	0.0	0.001
$E(z^2)$	0.016	0.016	0.017
$D(z)$	0.016	0.016	0.017
p = 0.5	r	r_S	r_Q
$E(z)$	0.496	0.475	0.331
$E(z^2)$	0.255	0.236	0.124
$D(z)$	0.009	0.01	0.014
p = 0.9	r	r_S	r_Q
$E(z)$	0.898	0.882	0.704
$E(z^2)$	0.807	0.779	0.505
$D(z)$	0.001	0.001	0.009

Таблица 2: Двумерное нормальное распределение, $n = 60$

p = 0	r	r_S	r_Q
$E(z)$	-0.003	-0.002	0.002
$E(z^2)$	0.01	0.01	0.01
$D(z)$	0.01	0.01	0.01
p = 0.5	r	r_S	r_Q
$E(z)$	0.494	0.474	0.328
$E(z^2)$	0.25	0.231	0.117
$D(z)$	0.006	0.006	0.009
p = 0.9	r	r_S	r_Q
$E(z)$	0.898	0.885	0.706
$E(z^2)$	0.807	0.784	0.503
$D(z)$	0.0	0.001	0.005

Таблица 3: Двумерное нормальное распределение, $n = 100$

n = 20	r	r_S	r_Q
$E(z)$	0.874	0.843	0.659
$E(z^2)$	0.767	0.717	0.464
$D(z)$	0.003	0.006	0.03
n = 60	r	r_S	r_Q
$E(z)$	0.876	0.858	0.675
$E(z^2)$	0.768	0.738	0.466
$D(z)$	0.001	0.002	0.01
n = 100	r	r_S	r_Q
$E(z)$	0.877	0.862	0.676
$E(z^2)$	0.769	0.744	0.463
$D(z)$	0.001	0.001	0.006

Таблица 4: Смесь нормальных распределений

4.2 Эллипсы рассеивания

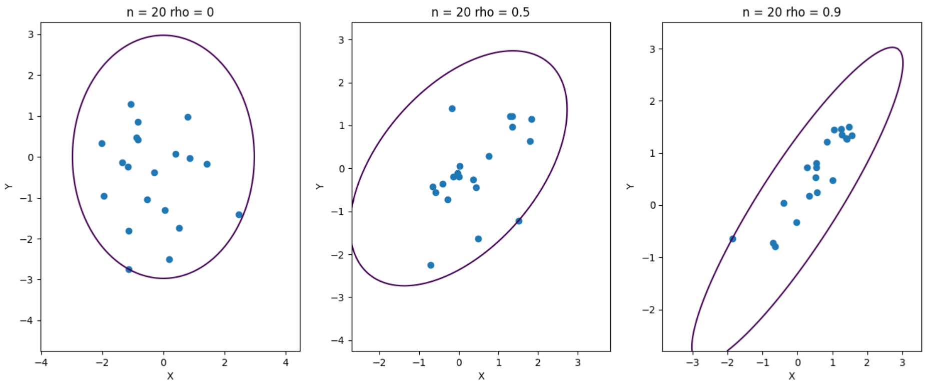

Рис. 1: Двумерное нормальное распределение, n=20

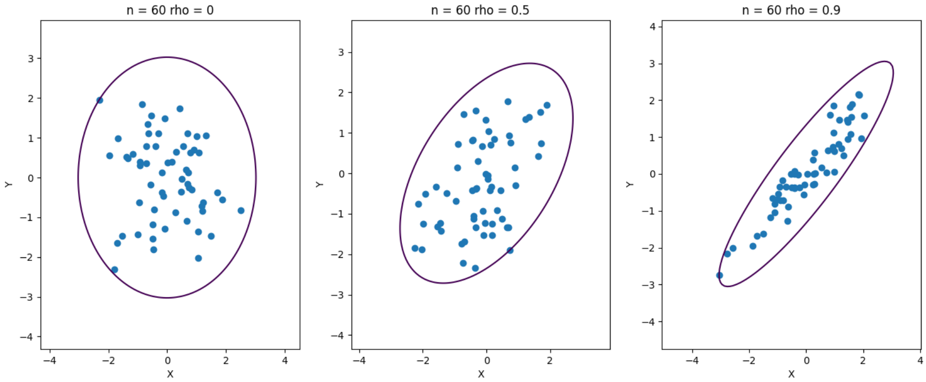

Рис. 2: Двумерное нормальное распределение, n=60

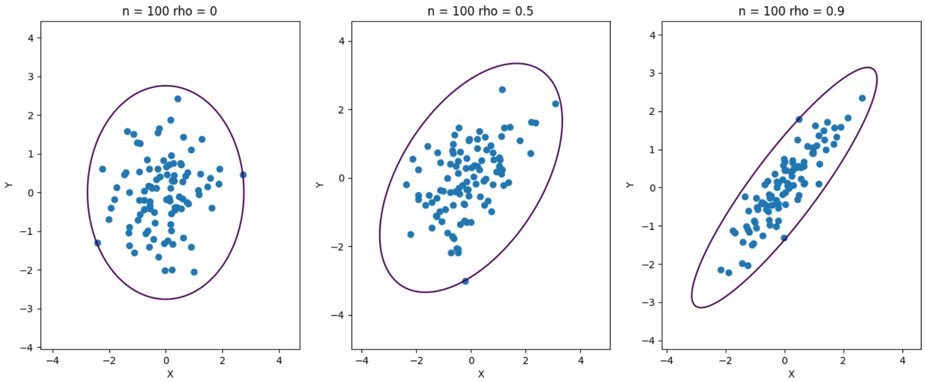

Рис. 3: Двумерное нормальное распределение, n=100

4.3 Оценки коэффициентов линейной регрессии

4.3.1 Выборка без возмущений

- Критерий наимеших квадратов:

$$\hat{a} \approx 1.868,\ \hat{b} \approx 1.738$$

- Критерий наименьших модулей:

$$\hat{a} \approx 2.311,\ \hat{b} \approx 1.614$$

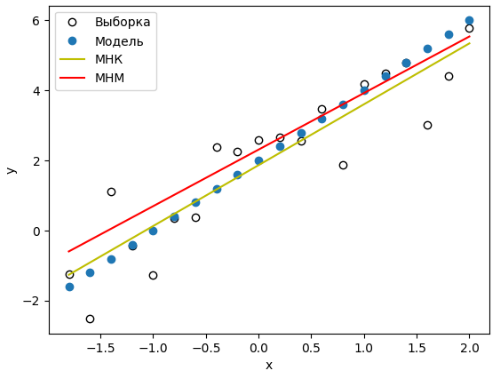

Рис. 4: Выборка без возмущений

4.3.2 Выборка с возмущениями

- Критерий наимеших квадратов:

$$\hat{a} \approx 1.786,\ \hat{b} \approx 0.052$$

- Критерий наименьших модулей:

$$\hat{a} \approx 1.346,\ \hat{b} \approx 1.445$$

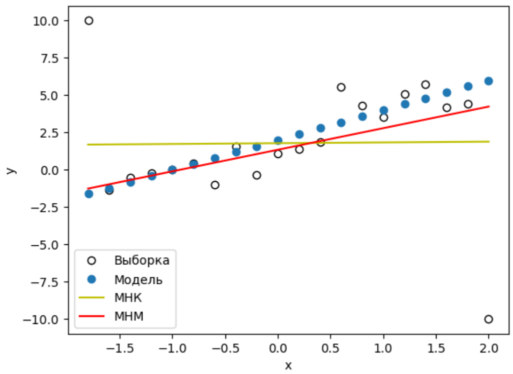

Рис. 5: Выборка с возмущениями

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

$$\hat{\mu} \approx 0.044,\ \hat{\sigma} \approx 0.975$$

- Количество промежутков $k = 8$
- Уровень значимости $\alpha = 0.05$
- $\chi^2_{0.95} = 14.067$

i	Границы	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-inf, -1.0]	14	0.159	15.866	-1.866	0.219
2	[-1.0, -0.667]	10	0.094	9.384	0.616	0.04
3	[-0.667, -0.333]	17	0.117	11.695	5.305	2.407
4	[-0.333, 0.0]	10	0.131	13.056	-3.056	0.715
5	[0.0, 0.333]	12	0.131	13.056	-1.056	0.085
6	[0.333, 0.667]	13	0.117	11.695	1.305	0.146
7	[0.667, 1.0]	6	0.094	9.384	-3.384	1.22
8	[1.0, inf]	18	0.159	15.866	2.134	0.287
$\sum$	-	100.0	1.0	100.0	0.0	5.12

Таблица 5: Вычисление χ^2_B при проверке гипотезы H_0 о нормальном законе $N(x, \hat{\mu}, \hat{\sigma})$

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

$$\hat{\mu} \approx 0.077,\ \hat{\sigma} \approx 2.029$$

- Количество промежутков $k = 5$
- Уровень значимости $\alpha = 0.05$
- $\chi^2_0.95 = 9.488$

i	Границы	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-inf, -1.0]	4	0.159	3.173	0.827	0.215
2	[-1.0, -0.333]	2	0.211	4.216	-2.216	1.165
3	[-0.333, 0.333]	7	0.261	5.222	1.778	0.605
4	[0.333, 1.0]	3	0.211	4.216	-1.216	0.351
5	[1.0, inf]	4	0.159	3.173	0.827	0.215
$\sum$	-	20.0	1.0	20.0	0.0	2.551

Таблица 6: Вычисление χ^2_B при проверке гипотезы H_0^* о $Laplace(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$

Видим, что $\chi^2_B < \chi^2_{0.95}$, следовательно гипотеза о нормальности принимается.
Рассмотрим гипотезу H_0^*, что выборка распределена согласно равномерному распределению$(x, \hat{\mu}, \sigma)$

$$\hat{\mu} \approx -0.222,\ \hat{\sigma} \approx 0.969$$

- Количество промежутков $k = 5$
- Уровень значимости $\alpha = 0.05$
- $\chi^2_{0.95} = 9.488$

i	Границы	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-inf, -1.0]	6	0.159	3.173	2.827	2.518
2	[-1.0, -0.333]	5	0.211	4.216	0.784	0.146
3	[-0.333, 0.333]	2	0.261	5.222	-3.222	1.988
4	[0.333, 1.0]	4	0.211	4.216	-0.216	0.011
5	[1.0, inf]	3	0.159	3.173	-0.173	0.009
$\sum$	-	20.0	1.0	20.0	0.0	4.673

Таблица 7: Вычисление χ^2_B при проверке гипотезы H_0^* о равномерном распределении

Видим, что $\chi^2_B < \chi^2_{0.95}$, следовательно гипотеза о нормальности принимается.

4.5 Доверительные интервалы для параметров нормального распределения

n = 20	m	σ
	$-0.334 < m < 0.412$	$0.606 < \sigma < 1.163$
n = 100	m	σ
	$-0.089 < m < 0.308$	$0.88 < \sigma < 1.164$

Таблица 8: Доверительные интервалы для параметров нормального распределения

4.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

n = 20	m	σ
	$-0.301 < m < 0.379$	$0.557 < \sigma < 3.282$
n = 100	m	σ
	$-0.086 < m < 0.305$	$0.832 < \sigma < 1.327$

Таблица 9: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

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 примерно равен его теоретическому значению (95%)

5.2 Оценки коэффициентов линейной регрессии

Критерий наименьших 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 также, что критерий принял гипотезу H_0 для выборок, сгенерированных по равномерному закону и закону распределения 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

5.4 Доверительные интервалы для параметров распределения

Генеральные характеристики ($m = 0, \sigma = 1$) накрываются построенными доверительными интервалами

Доверительные интервалы, полученные по большей выборке, являются соответственно более точными, т.е. меньшими по длине

Доверительные интервалы для параметров нормального распределения более надёжны, так как основаны на точном, а не асимптотическом распределении

Литература‘

[1] Вероятностные разделы математики. Учебник для бакалавров технических направлений.//Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.

[2] Вентцель Е.С. Теория вероятностей: Учеб. для вузов. — 6-е изд. стер. — М.: Высш. шк., 1999.— 576 с.

[3] Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей : учеб. пособие / Ю.Д. Максимов; под ред. В.И. Антонова. — СПб. : Изд-во Политехн. ун-та, 2009. — 395 с. (Математика в политехническом университете).

Ссылка на репозиторий: `https://github.com/Hembos/Mathematical_statistics/tree/main/Lab2`