Санкт-Петербургский политехнический университет имени Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики и физики

Математическая статистика
Отчёт по лабораторным работам №5-8

Выполнил:

Студент: Габдушев Рушан
Группа: 5030102/90201

Принял:

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

2022 г.

Содержание

Список иллюстраций

Список таблиц

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:

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

2. Для выборок из предыдущего пункта изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

3. 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.

4. 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.

5. 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

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\bar{x}$,$\bar{y}$ и средними квадратическими отклонениями σ_x, σ_y соответственно. Параметр ρ называется коэффициентом корреляции.

2.2. Корреляционный момент (ковариация) и коэффициент корреляции

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

$$K = cov(X,Y) = M[(X-\bar{x})(Y-\bar{y})] \tag{3}$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \tag{4}$$

2.3. Выборочные коэффициенты корреляции

2.3.1. Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum(x_i-\bar{x})^2\frac{1}{n}\sum(y_i-\bar{y})^2}} = \frac{K}{s_X s_Y}, \tag{5}$$

где K, s_X^2, s_Y^2 — выборочные ковариация и дисперсии с.в. X и Y.

2.3.2. Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1+n_3)-(n_2+n_4)}{n}, \tag{6}$$

где n_1, n_2, n_3, n_4– количества точек с координатами x_i, y_i, попавшими соответственно в I, II, III, IV квадранты декартовой системы с осями $x' = x - medx$, $y' = y - medy$ и с центром в точке с координатами $(medx,\ medy)$

2.3.3. Выборочный коэффициент ранговой корреляции Спирмена

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

Выборочный коэффициент ранговой корреляции Спирмена:

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

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

2.4. Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

$$\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} = const \tag{8}$$

Центр эллипса (8) находится в точке с координатами $(\bar{x}, \bar{y})$; оси симметрии эллипса составляют с осью Ox углы, определяемые уравнением

$$tg(2\alpha) = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{9}$$

2.5. Простая линейная регрессия

2.5.1. Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1..n \tag{10}$$

где $x_1, ..., x_n$— заданные числа (значения фактора); $y_1, ...y_n$— наблюдаемые значения отклика; $\varepsilon_1, ..., \varepsilon_n$— 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):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \to \min_{\beta_0, \beta_1} \tag{11}$$

2.5.3. Расчётные формулы для МНК-оценок

МНК-оценки параметров $\hat{\beta}_0, \hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} \tag{12}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 \tag{13}$$

2.6. Робастные оценки коэффициентов линейной регрессии

Метода наименьших модулей:

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \to \min_{\beta_0, \beta_1} \tag{14}$$

$$\hat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \tag{15}$$

$$\hat{\beta}_{0R} = medy - \hat{\beta}_{1R} medx, \tag{16}$$

$$r_Q = \frac{1}{n} \sum_{i=1}^{n} sgn(x_i - medx) sgn(y_i - medy), \tag{17}$$

$$q_y^* = \frac{y_{(j)} - y_{(l)}}{k_q(n)}, \quad q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)},$$
$$\begin{cases} [\frac{n}{4}] + 1 \text{ при } \frac{n}{4} \text{ дробном}, \\ \frac{n}{4} \text{ при } \frac{n}{4} \text{ целом}. \end{cases}$$
$$j = n - l + 1$$
$$sgn(z) = \begin{cases} 1 \text{ при } z > 0 \\ 0 \text{ при } z = 0 \\ -1 \text{ при } z < 0 \end{cases} \tag{18}$$

Уравнение регрессии здесь имеет вид

$$y = \hat{\beta}_{0R} + \hat{\beta}_{1R} x \tag{19}$$

2.7. Метод максимального правдоподобия

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

$$L(x_1, ..., x_n, \theta) = f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta) \tag{20}$$

Оценка максимального правдоподобия (о.м.п):

$$\hat{\theta_{\text{мп}}} = \arg\max_{\theta} L(x_1, ..., x_n, \theta) \tag{21}$$

система уравнений правдоподобия (в случае дифференцируемости ФП):

$$\frac{\partial L}{\partial \theta_k} = 0 \text{ или } \frac{\partial \ln L}{\partial \theta_k} = 0, k = 1, ..m \tag{22}$$

χ^2 асимптотически распределена по закону χ^2 с $k-1$ степенями свободы.

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 Δ_i, $i = 1, ..., k$.

5. Вычисляем выборочное значение статистики критерия χ^2:

$$\chi^2_B = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}. \tag{23}$$

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

а) Если $\chi^2_B < \chi^2_{1-\alpha}(\mathrm{k}-1)$, 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.

2.8. Доверительные интервалы для параметров нормального распределения

2.8.1. 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 $\bar{x}$ и выборочное среднее квадратическое отклонение s. Параметры m и σ нормального распределения неизвестны.

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

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

$$P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha \tag{24}$$

2.8.2. 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 $\gamma = 1 - \alpha$:

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1 - \alpha, \tag{25}$$

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

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

2.9.1. Доверительный интервал для математического ожидания m произвольной генеральной совокупности при большом объёме 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 $\gamma = 1-\alpha$:

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma, \tag{26}$$

2.9.2. Доверительный интервал для среднего квадратического отклонения σ произвольной генеральной совокупности при большом объёме выборки

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

$u_{1-\alpha/2}$ — корень этого уравнения — квантиль нормального распределения $N(0,1)$ порядка $1-\alpha/2$

$E = \frac{\mu_4}{\sigma^4} - 3$ — эксцесс генерального распределения, е $= \frac{m_4}{s^4} - 3$ — выборочный эксцесс; $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$ - четвёртый выборочный центральный момент.

$$s(1+U)^{-1/2} < \sigma < s(1-U)^{-1/2} \tag{27}$$

или

$$s(1-0.5U) < \sigma < s(1+0.5U) \tag{28}$$

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

Формулы (27) или 28 дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1-\alpha$.

Замечание. Вычисления по формуле (27) дают более надёжный результат, так как в ней меньше грубых приближений.

3. Реализация

Данная лабораторная работа была выполнена с использованием языка программирования Python 3.9 в среде разработки PyCharm 2021.3.2 с использованием следующих библиотек:

- scipy версии 1.8.0
- numpy версии 1.22.3
- matplotlib версии 3.5.1
- tabulate версии 0.8.9

4. Результаты

4.1. Выборочные коэффициенты корреляции

$\rho = 0$	r	r_S	r_Q
$E(z)$	-0.0043	-0.0037	-0.0038
$E(z^2)$	0.0491	0.0503	0.0541
$D(z)$	0.0491	0.0503	0.0541
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.4886	0.4583	0.3200
$E(z^2)$	0.2713	0.2461	0.1506
$D(z)$	0.0326	0.0360	0.0482
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8945	0.8671	0.7008
$E(z^2)$	0.8025	0.7565	0.5181
$D(z)$	0.0024	0.0045	0.0270

Таблица 1. Двумерное нормальное распределение, n = 20

$\rho = 0$	r	r_S	r_Q
$E(z)$	0.0020	0.0023	0.0018
$E(z^2)$	0.0164	0.0167	0.0165
$D(z)$	0.0164	0.0167	0.0165
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.4958	0.4751	0.3339
$E(z^2)$	0.2562	0.2368	0.1258
$D(z)$	0.0104	0.0111	0.0143
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8999	0.8838	0.7089
$E(z^2)$	0.8104	0.7822	0.5111
$D(z)$	0.0006	0.0010	0.0085

Таблица 2. Двумерное нормальное распределение, n = 60

$\rho = 0$	r	r_S	r_Q
$E(z)$	-0.0020	-0.0018	0.0009
$E(z^2)$	0.0099	0.0099	0.0105
$D(z)$	0.0099	0.0099	0.0105
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.5008	0.4792	0.3286
$E(z^2)$	0.2565	0.2360	0.1168
$D(z)$	0.0057	0.0064	0.0088
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8994	0.8863	0.7099
$E(z^2)$	0.8092	0.7862	0.5093
$D(z)$	0.0004	0.0007	0.0053

Таблица 3. Двумерное нормальное распределение, n = 100

$n = 20$	r	r_S	r_Q
$E(z)$	0.7828	0.7518	0.5646
$E(z^2)$	0.6220	0.5775	0.3554
$D(z)$	0.0093	0.0124	0.0366
$n = 60$	r	r_S	r_Q
$E(z)$	0.7865	0.7598	0.5696
$E(z^2)$	0.6246	0.5855	0.3489
$D(z)$	0.0059	0.0081	0.0245
$n = 100$	r	r_S	r_Q
$E(z)$	0.7877	0.7637	0.5725
$E(z^2)$	0.6250	0.5894	0.3463
$D(z)$	0.0045	0.0061	0.0186

Таблица 4. Смесь нормальных распределений

4.2. Эллипсы рассеивания

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

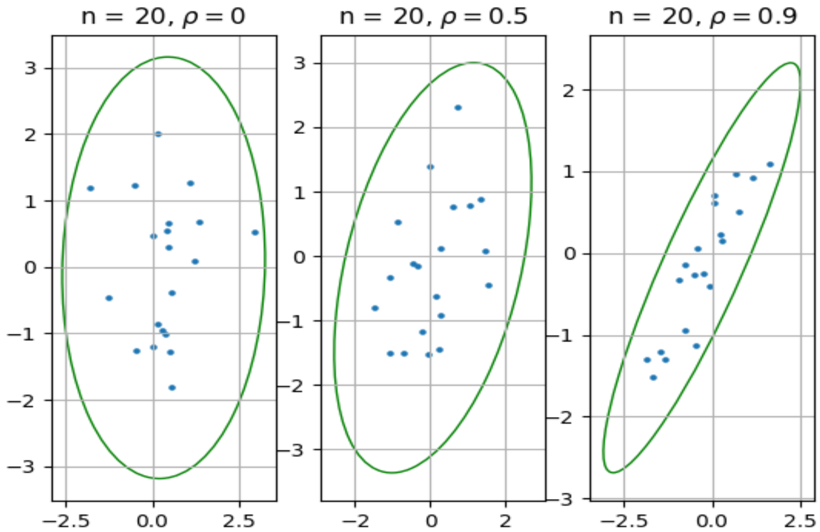

Рис. 1. Двумерное нормальное распределение, $n = 20$

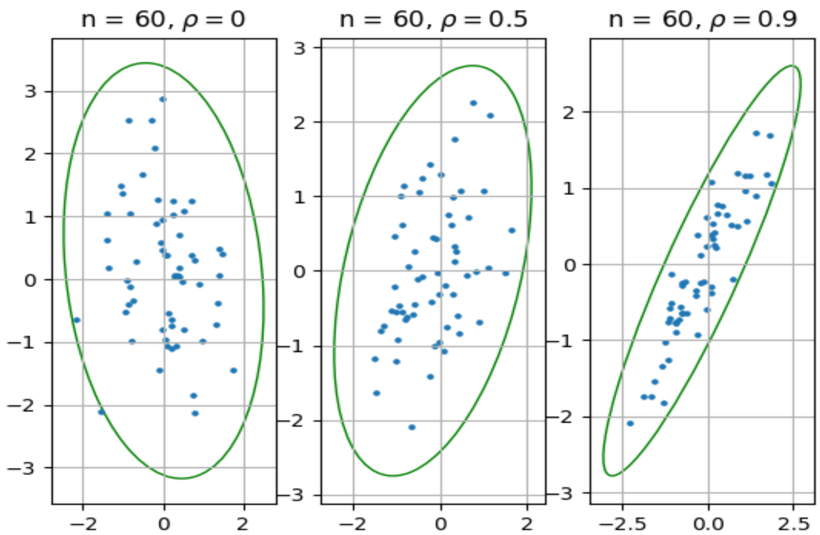

Рис. 2. Двумерное нормальное распределение, $n = 60$

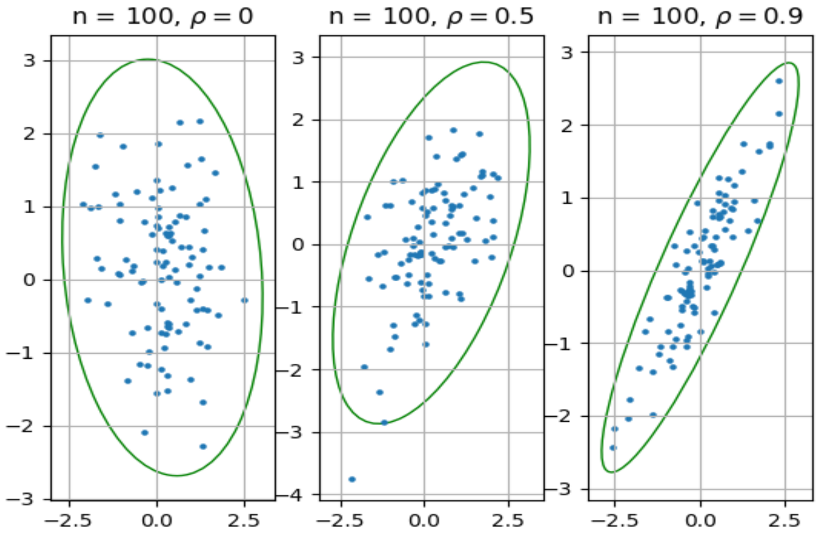

Рис. 3. Двумерное нормальное распределение, $n = 100$

4.3. Оценки коэффициентов линейной регрессии

4.3.1. Выборка без возмущений

1. Критерий наименьших квадратов: $\hat{a} \approx 1.76$, $\hat{b} \approx 1.89$
2. Критерий наименьших модулей: $\hat{a} \approx 1.76$, $\hat{b} \approx 1.89$

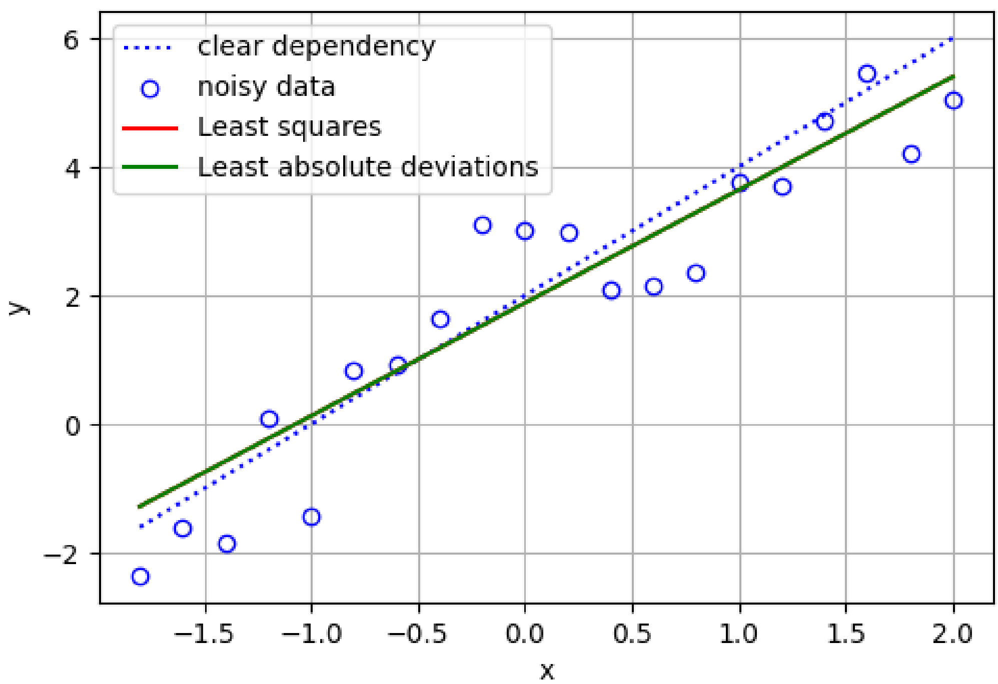

Рис. 4. Выборка без возмущений

4.3.2. Выборка с возмущениями

1. Критерий наименьших квадратов: $\hat{a} \approx 1.99$, $\hat{b} \approx 0.48$

2. Критерий наименьших модулей: $\hat{a} \approx 2.03$, $\hat{b} \approx 1.11$

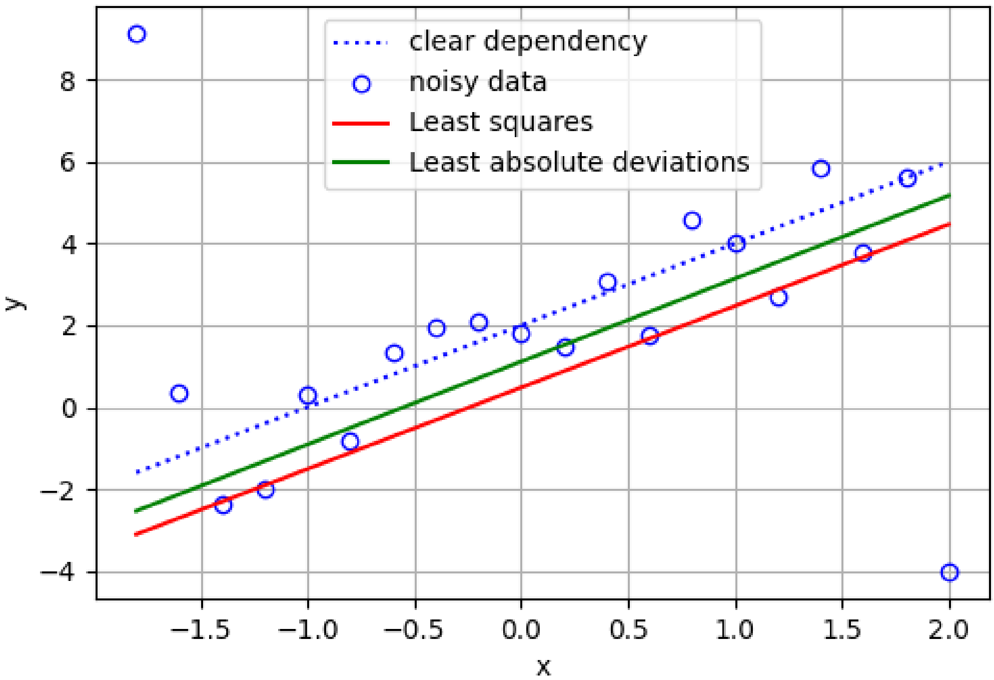

Рис. 5. Выборка с возмущениями

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

4.4.1. Метод максимального правдоподобия

$\hat{\mu} \approx 0.11, \hat{\sigma} \approx 0.99$
Критерий согласия χ^2:
Количество промежутков $k = 8$
Уровень значимости $\alpha = 0.05$
Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(7)$. Из таблицы [3, с. 358] $\chi^2_{0.95}(7) \approx 14.07$.

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	['-inf', -1.1]	9	0.1357	13.57	-4.57	1.54
2	[-1.1, -0.73]	12	0.096	9.6	2.4	0.6
3	[-0.73, -0.37]	9	0.1253	12.53	-3.53	0.99
4	[-0.37, 0.0]	13	0.1431	14.31	-1.31	0.12
5	[0.0, 0.37]	15	0.1431	14.31	0.69	0.03
6	[0.37, 0.73]	14	0.1253	12.53	1.47	0.17
7	[0.73, 1.1]	16	0.096	9.6	6.4	4.26
8	[1.1, 'inf']	12	0.1357	13.57	-1.57	0.18
$\sum$	-	100	1	100	-0	7.9

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

Сравнивая $\chi^2_B = 7.9$ и $\chi^2_{0.95}(7) \approx 14.07$, видим, что $\chi^2_B < \chi^2_{0.95}(7)$. Следовательно, на данном этапе гипотезу H_0 можно принять.

4.4.2. Исследование на чувствительность

Рассмотрим выборки, распределённые по законам $L(x, 0, \frac{1}{\sqrt{2}})$ и $U(x, -\sqrt{3}, \sqrt{3})$. Проверим гипотезу о том, что они распределены по нормальному закону.

$\hat{\mu} \approx -0.03, \hat{\sigma} \approx 0.81$

Критерий согласия χ^2:

Количество промежутков $k = 5$

Уровень значимости α= 0.05

Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(4)$. Из таблицы [3, с. 358] $\chi^2_{0.95}(4) \approx 9.49$.

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	['-inf', -1.1]	2	0.1357	2.71	-0.71	0.19
2	[-1.1, -0.37]	5	0.2213	4.43	0.57	0.07
3	[-0.37, 0.37]	8	0.2861	5.72	2.28	0.91
4	[0.37, 1.1]	4	0.2213	4.43	-0.43	0.04
5	[1.1, 'inf']	1	0.1357	2.71	-1.71	1.08
$\sum$	-	20	1	20	-0	2.29

Таблица 6. Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $L(x, \hat{\mu}, \hat{\sigma})$, $n = 20$

Сравнивая $\chi^2_B = 2.29$ и $\chi^2_{0.95}(4) \approx 9.49$, видим, что $\chi^2_B < \chi^2_{0.95}(4)$. Следовательно, на данном этапе гипотезу H_0 можно принять.

$\hat{\mu} \approx -0.34, \hat{\sigma} \approx 1.03$

Критерий согласия χ^2:

Количество промежутков $k = 5$
Уровень значимости $\alpha = 0.05$
Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(4)$. Из таблицы [3, с. 358] $\chi^2_{0.95}(4) \approx 9.49$.

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	['-inf', -1.1]	6	0.1357	2.71	3.29	3.98
2	[-1.1, -0.37]	7	0.2213	4.43	2.57	1.5
3	[-0.37, 0.37]	3	0.2861	5.72	-2.72	1.3
4	[0.37, 1.1]	1	0.2213	4.43	-3.43	2.65
5	[1.1, 'inf']	3	0.1357	2.71	0.29	0.03
$\sum$	-	20	1	20	-0	9.46

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

Сравнивая $\chi^2_B = 9.46$ и $\chi^2_{0.95}(4) \approx 9.49$, видим, что $\chi^2_B < \chi^2_{0.95}(4)$. Из-за близости значений следует провести дополнительные исследования.

4.5. Доверительные интервалы для параметров нормального распределения. Асимптотический подход

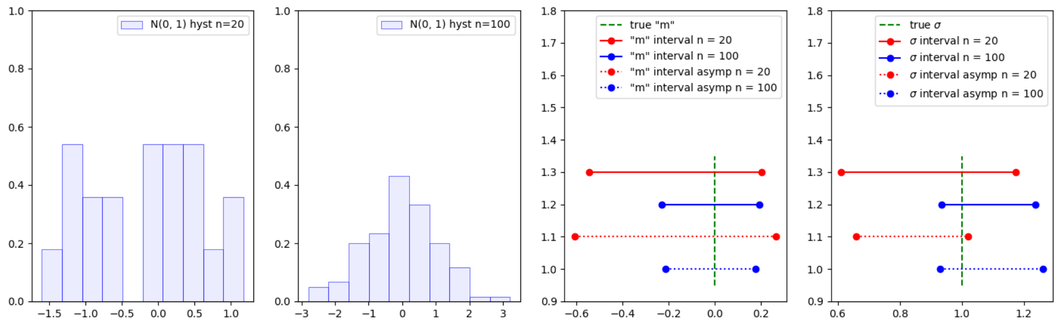

Рис. 6. Гистограммы нормальных распределений и доверительные интервалы их параметров

n = 20	m	σ
	$-0.55 < m < 0.20$	$0.61 < \sigma < 1.17$
n = 100	m	σ
	$-0.23 < m < 0.19$	$0.93 < \sigma < 1.24$

Таблица 8. Доверительные интервалы для параметров нормального распределения

n = 20	m	σ
	-0.61 $< m <$ 0.27	0.66 $< \sigma <$ 1.02
n = 100	m	σ
	-0.21 $< m <$ 0.18	0.93 $< \sigma <$ 1.26

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

5. Обсуждение

5.1. 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: $r_Q < r_S < r$.

5.2. Эллипсы рассеивания

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

5.3. Оценки коэффициентов линейной регрессии

По полученным результатам можно сказать, что критерий наименьших квадратов точнее оценивает коэффициенты линейной регрессии на выборки без возмущений. Если же редкие возмущения присутствуют, тогда лучше использовать критерий наименьших модулей, поскольку он более устойчив.

5.4. 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 для выборок сгенерированных по равномерному закону и закону Лапласа гипотеза H_0 оказалась принята.

5.5. 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.

6. Ссылки на библиотеки

1. `https://scipy.org/` - SciPy
2. `https://numpy.org/` - NumPy
3. `https://numpy.org/` - Matplotlib
4. `https://pypi.org/project/tabulate/` - Tabulate

7. Ссылки на репозиторий

`https://github.com/maloxit/matstat` - GitHub репозиторий