Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторным работам №5-8
по дисциплине
«Математическая статистика»

Выполнил студент:

Кондратьев Д. А.
группа: 3630102/70301

Проверил:

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

Санкт-Петербург
2020 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:

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

   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, используя критерий согласия χ^2. В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2.

   Также проверить данную гипотезу на равномерном распределении $U(x, -2, 2)$ при размере выборки ровной 20 элементам.

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:

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

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

Корреляционным моментом, иначе *ковариацией*, двух случайных величин X и Y называется математическое ожидание произведения отклонений этих случайных величин от их математических ожиданий [1, с. 141].

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

Коэффициентом корреляции ρ двух случайных величин X и Y называется отношение их корреляционного момента к произведению их средних квадратических отклонений:

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

Коэффициент корреляции — это нормированная числовая характеристика, являющаяся мерой близости зависимости между случайными величинами к линейной [1, с. 150].

2.3. Выборочные коэффициенты корреляции

2.3.1. Выборочный коэффициент корреляции Пирсона

Пусть по выборке значений $\{x_i, y_i\}_1^n$ двумерной с.в. (X, Y) требуется оценить коэффициент корреляции $\rho = \frac{cov(X,Y)}{\sqrt{DXDY}}$. Естественной оценкой для ρ служит его статистический аналог в виде выборочного коэффициента корреляции, предложенного К.Пирсоном, —

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

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

2.3.2. Выборочный квадрантный коэффициент корреляции

Кроме выборочного коэффициента корреляции Пирсона, существуют и другие оценки степени взаимосвязи между случайными величинами. К ним относится *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 точке с координатами (x_i, y_i), попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - med\, x, y' = y - med\, y$ и с центром в точке с координатами $(med\, x, med\, y)$ [1, с. 539].

2.3.3. Выборочный коэффициент ранговой корреляции Спирмена

На практике нередко требуется оценить степень взаимодействия между качественными признаками изучаемого объекта. Качественным называется признак, который нельзя измерить точно, но который позволяет сравнивать изучаемые объекты между собой и располагать их в порядке убывания или возрастания их качества. Для этого объекты выстраиваются в определённом порядке в соответствии с рассматриваемым признаком. Процесс упорядочения называется *ранжированием*, и каждому члену упорядоченной последовательности объектов присваивается ранг, или порядковый номер. Например, объекту с наименьшим значением признака присваивается ранг 1, следующему за ним объекту — ранг 2, и т.д. Таким образом, происходит сравнение каждого объекта со всеми объектами изучаемой выборки.

Если объект обладает не одним, а двумя качественными признаками — переменными X и Y , то для исследования их взаимосвязи используют выборочный коэффициент корреляции между двумя последовательностями рангов этих признаков.

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

Выборочный коэффициент ранговой корреляции Спирмена определяется как выборочный коэффициент корреляции Пирсона между рангами u, υ переменных X, Y :

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

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

2.4. Простая линейная регрессия

2.4.1. Модель простой линейной регрессии

Регрессионную модель описания данных называют *простой линейной регрессией*, если

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, ..., n, \tag{7}$$

где x_1, ... , x_n — заданные числа (значения фактора);
y_1, ... , y_n — наблюдаемые значения отклика;
ε_1, ... , ε_n — независимые, нормально распределённые N(0, σ) с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины (ненаблюдаемые);
β_0, β_1 — неизвестные параметры, подлежащие оцениванию.

В модели (7) отклик y зависит зависит от одного фактора x, и весь разброс экспериментальных точек объясняется только погрешностями наблюдений (результатов измерений) отклика y. Погрешности результатов измерений x в этой модели полагают существенно меньшими погрешностей результатов измерений y, так что ими можно пренебречь [1, с. 507].

2.5. Метод наименьших квадратов

При оценивании параметров регрессионной модели используют различные методы. Один из наиболее распрстранённых 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):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \to \min_{\beta_0, \beta_1}. \tag{8}$$

Задача минимизации квадратичного критерия (8) носит название задачи *метода наименьших квадратов* (МНК), а оценки $\hat{\beta}_0$, $\hat{\beta}_1$ параметров β_0, β_1, реализующие минимум критерия (8), называют МНК-*оценками* [1, с. 508].

2.5.1. Расчётные формулы для МНК-оценок

МНК-оценки параметров $\hat{\beta}_0$ и $\hat{\beta}_1$ находятся из условия обращения функции $Q(\beta_0, \beta_1)$ в минимум.

Для нахождения МНК-оценок $\hat{\beta}_0$ и $\hat{\beta}_1$ выпишем необходимые условия экстремума:

$$\begin{cases} \frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0, \\ \frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{cases} \tag{9}$$

Далее для упрощения записи сумм будем опускать индекс суммирования. Из системы (9) получим

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i \end{cases} \tag{10}$$

Разделим оба уравнения на n и используя известные статистические обозначения для выборочных первых и вторых начальных моментов

$$\overline{x} = \frac{1}{n} \sum x_i,\ \overline{y} = \frac{1}{n} \sum y_i,\ \overline{x^2} = \frac{1}{n} \sum x_i^2,\ \overline{xy} = \frac{1}{n} \sum x_i y_i,$$

получим

$$\begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \overline{x} = \overline{y} \\ \hat{\beta}_0 \overline{x} + \hat{\beta}_1 \overline{x^2} = \overline{xy} \end{cases} \tag{11}$$

откуда МНК-оценку $\hat{\beta}_1$ наклона прямой регрессии находим по формуле Крамера

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

а МНК-оценку $\hat{\beta}_0$ определяем непосредственно из первого уравнения системы (11):

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

2.6. Метод наименьших модулей

Критерий наименьших модулей — заключается в минимизации следующей функции:

$$M(a,b) = \sum_{i=1}^{n} |y_i - ax_i - b| \to \min \tag{14}$$

2.7. Метод максимального правдоподобия

Метод максимального правдоподобия — метод оценивания неизвестного параметра путём максимимзации функции правдоподобия.

$$\hat{\theta}_{\text{МП}} = \arg\max_{\theta} L(x_1, x_2, \ldots, x_n, \theta) \tag{15}$$

где L — функция правдоподобия, которая представляет собой совместную плотность вероятности независимых случайных величин $x_1, x_2, \ldots, x_n$ и является функцией неизвестного параметра θ.

$$L = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \cdots \cdot f(x_n, \theta) \tag{16}$$

Оценкой максимального правдоподобия будем называть такое значение $\hat{\theta}_{\text{МП}}$ из множества допустимых значений параметра θ, для которого функция правдоподобия принимает при заданных $x_1, x_2, \ldots, x_n$ максимальное значение.

Тогда при оценивании математического ожидания m и дисперсии σ^2 нормального распределения $N(m, \sigma)$ получим:

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 \tag{17}$$

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

Разобьём генеральную совокупность на k неперсекающихся подмножеств $\Delta_1, \Delta_2, \ldots, \Delta_k$, $\Delta_i = (a_i, a_{i+1}]$, $p_i = P(X \in \Delta_i)$, $i = 1, 2, \ldots, k$ — вероятность того, что точка попала в i-ый промежуток.

Так как генеральная совокупность — это $\mathbb{R}$, то крайние промежутки будут бесконечными: $\Delta_1 = (-\infty, a_1],\ \Delta_k = (a_k, \infty),\ p_i = F(a_i) - F(a_{i-1})$.

n_i — частота попадания выборочных элементов в $\Delta_i,\ i = 1, 2, \ldots, k$.

В случае справедливости гипотезы H_0 относительно частоты $\frac{n_i}{n}$ при больших n должны быть близки к p_i, значит в качестве меры имеет смысл взять:

$$Z = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 \tag{18}$$

Тогда:

$$\chi^2_B = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{19}$$

Теорема К.Пирсона. Статистика критерия χ^2 асимптотически распределена по закону χ^2 с $k-1$ степенями свободы.

2.9. Доверительные интервалы для параметров нормального распределения

Оценкой максимального правдоподобия для математического ожидания является среднее арифметическое: $\mu = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$.

Оценка максимального правдоподобия для дисперсии вычисляется по формуле: $\sigma^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^2$.

Доверительным интервалом или интервальной оценкой числовой характеристики или параметра распределения θ с доверительной вероятностью γ называется интервал со случайными границами (θ_1, θ_2), содержащий параметр θ с вероятностью γ.

Функция распределения Стьюдента:

$$T = \sqrt{n-1} \frac{\overline{x} - \mu}{\delta} \tag{20}$$

Функция плотности распределения χ^2:

$$f(x) = \begin{cases} 0, & x \le 0 \\ \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \end{cases} \tag{21}$$

Интервальная оценка математического ожидания:

$$P = \left(\overline{x} - \frac{\sigma t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n-1}} < \mu < \overline{x} + \frac{\sigma t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n-1}} \right) = \gamma, \tag{22}$$

где $t_{1-\frac{\alpha}{2}}$ — квантиль распределения Стьюдента порядка $1-\frac{\alpha}{2}$.

Интервальная оценка дисперсии:

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

где $\chi^2_{1-\frac{\alpha}{2}}$, $\chi^2_{\frac{\alpha}{2}}$ — квантили распределения Стьюдента порядков $1-\frac{\alpha}{2}$ и $\frac{\alpha}{2}$ соответственно.

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

Асимптотическая интервальная оценка математического ожидания:

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

где $u_{1-\frac{\alpha}{2}}$ — квантиль нормального распределения $N(x,0,1)$ порядка $1-\frac{\alpha}{2}$.

$$\sigma(1 - 0.5u_{1-\alpha/2}\sqrt{e+2}/\sqrt{n}) < \sigma < \sigma(1 + 0.5u_{1-\alpha/2}\sqrt{e+2}/\sqrt{n}) \tag{25}$$

3. Реализация

Лабораторная работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для генерации выборки и работы с массивами;
- *matplotlib.pyplot* и *seaborn* — для построения эллипсов рассеивания и графиков;
- *tabulate* — для построения таблиц;
- *scipy.stats* — содержит необходимые распределения и критерий χ^2;
- *scipy.optimize* — для решения задач оптимизации.

Ссылка на исходный код лабораторной работы приведена в приложении.

4. Результаты

4.1. Выборочные коэффициенты корреляции

$\rho = 0.0$ (3)	r (4)	r_S (6)	r_Q (5)
$E(z)$	0.01	0.00	0.00
$E(z^2)$	0.05	0.05	0.05
$D(z)$	0.0531	0.0520	0.0515

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.49	0.46	0.32
$E(z^2)$	0.27	0.25	0.15
$D(z)$	0.0307	0.0348	0.0480

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.896	0.866	0.70
$E(z^2)$	0.805	0.755	0.51
$D(z)$	0.0022	0.0048	0.0291

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

$\rho = 0.0$	r	r_S	r_Q
$E(z)$	0.00	0.00	-0.00
$E(z^2)$	0.02	0.02	0.02
$D(z)$	0.0172	0.0175	0.0171

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.494	0.47	0.32
$E(z^2)$	0.254	0.23	0.12
$D(z)$	0.0097	0.0109	0.0147

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8988	0.883	0.706
$E(z^2)$	0.8086	0.781	0.507
$D(z)$	0.0007	0.0013	0.0089

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

$\rho = 0.0$	r	r_S	r_Q
$E(z)$	-0.001	0.000	-0.00
$E(z^2)$	0.010	0.010	0.01
$D(z)$	0.0098	0.0098	0.0105

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.496	0.476	0.333
$E(z^2)$	0.251	0.233	0.120
$D(z)$	0.0056	0.0064	0.0089

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8995	0.8868	0.712
$E(z^2)$	0.8094	0.7871	0.511
$D(z)$	0.0004	0.0006	0.0051

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

$n = 20$	r	r_S	r_Q
$E(z)$	0.783	0.75	0.56
$E(z^2)$	0.622	0.57	0.35
$D(z)$	0.0085	0.0128	0.0396

$n = 60$	r	r_S	r_Q
$E(z)$	0.791	0.768	0.57
$E(z^2)$	0.628	0.594	0.34
$D(z)$	0.0024	0.0033	0.0111

$n = 100$	r	r_S	r_Q
$E(z)$	0.789	0.771	0.575
$E(z^2)$	0.624	0.596	0.337
$D(z)$	0.0015	0.0019	0.0063

Таблица 4. Смесь нормальных распределений

4.2. Эллипсы рассеивания

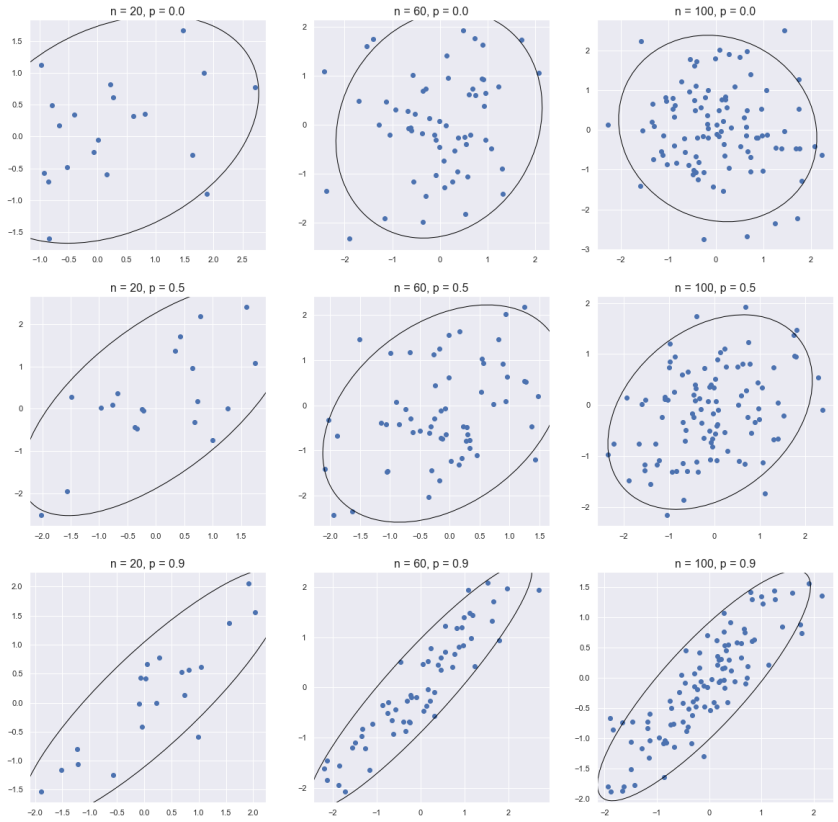

Рис. 1. Двумерное нормальное распределение

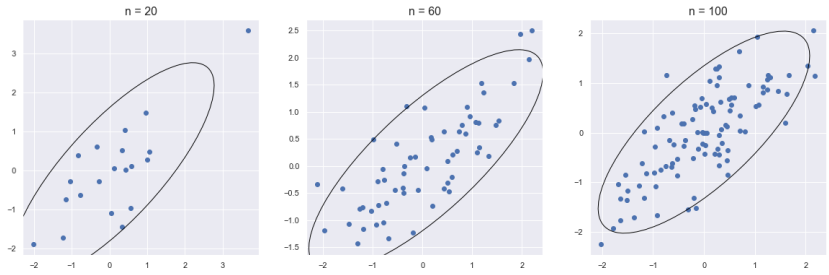

Рис. 2. Смесь нормальных распределений

4.3. Оценки коэффициентов линейной регрессии

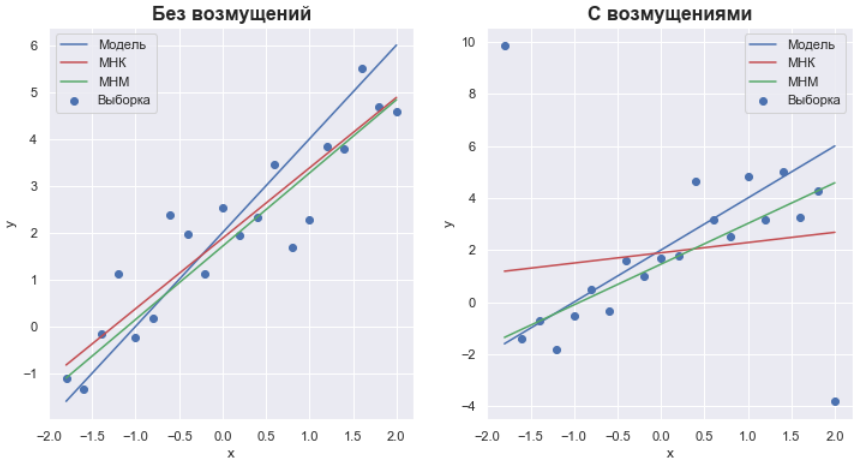

Рис. 3. Линейная регрессия

Выборка без возмущений

Критерий	a	b
МНК	1.502	1.878
МНМ	1.709	1.563

Выборка с возмущениями

Критерий	a	b
МНК	0.394	1.895
МНМ	1.457	1.564

Таблица 5. Оценки коэффициентов линейной регрессии

Введем следующую метрику:

$$\rho = \sum (y_i^* - y_i)^2$$

где y_i — значение эталонной функции в точке x_i, y_i^* — значение функции в точке x_i, полученное путем оценки.

Теперь посчитаем данные метрики для МНК и МНМ:

- выборка без возмущений: $\rho_{\text{МНК}} = 7.18,\ \rho_{\text{МНМ}} = 7.33$;
- выборка с возмущениями:$\rho_{\text{МНК}} = 70.06,\ \rho_{\text{МНМ}} = 11.94$.

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

4.4.1. Стандартное нормальное распределение

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.01, \quad \hat{\sigma} \approx 1.02.$$

Критерий согласия χ^2:
Количество промежутков $k = 6$.
Уровень значимости $\alpha = 0.05$.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.01]$	16	0.1562	15.62	0.38	0.01
2	$(-1.01, -0.37]$	17	0.2004	20.04	-3.04	0.46
3	$(-0.37, 0.28]$	31	0.2517	25.17	5.83	1.35
4	$(0.28, 0.92]$	18	0.2122	21.22	-3.22	0.49
5	$(0.92, 1.56]$	10	0.1201	12.01	-2.01	0.34
6	$(1.56, \infty]$	8	0.0594	5.94	2.06	0.72
$\sum$	—	100	1.0000	100	0.00	$3.36 = \chi^2_B$

Таблица 6. Вычисление χ^2_B при проверке гипотизы H_0 о нормальном законе распределения $\mathrm{N}(x, \hat{\mu}, \hat{\sigma})$. $N(x, 0, 1)$

4.4.2. Равномерное распределение

Метод максимального правдоподобия:

$$\hat{\mu} \approx -0.07, \quad \hat{\sigma} \approx 1.2.$$

Критерий согласия χ^2:
Количество промежутков $k = 5$.
Уровень значимости $\alpha = 0.05$.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.01]$	5	0.1562	3.12	1.88	1.13
2	$(-1.01, -0.37]$	5	0.2828	5.66	-0.66	0.08
3	$(-0.37, 0.28]$	3	0.3200	6.40	-3.40	1.81
4	$(0.92, 1.56]$	4	0.1815	3.63	0.37	0.04
5	$(1.56, \infty]$	3	0.0594	1.19	1.81	2.77
$\sum$	—	20	1.0000	20	0.00	$5.81 = \chi^2_B$

Таблица 7. Вычисление χ^2_B при проверке гипотизы H_0 о нормальном законе распределения $\mathrm{N}(x, \hat{\mu}, \hat{\sigma})$. $U(x, -2, 2)$

4.5. Доверительные интервалы для параметров нормального распределения

n	m	σ
20	$(-0.62; 0.28)$	$(0.73; 1.40)$
100	$(-0.24; 0.12)$	$(0.81; 1.07)$

Таблица 8. Доверительные интервалы для параметров нормального распределения

4.6. Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

n	m	σ
20	$(-0.58; 0.24)$	$(0.83; 1.09)$
100	$(-0.24; 0.12)$	$(0.86; 1.01)$

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

5. Обсуждение

5.1. Выборочные коэффициенты корреляции и эллипсы рассеивания

Исходя из полученных результатов можно сделать следующие выводы:

- Верны следующие соотношения для дисперсий выборочных коэффициентов корреляции:
  - для двумерного нормального распределения: $r < r_S < r_Q$;
  - для смеси нормальных распределений: $r < r_S < r_Q$.
- Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению (95%).
- При уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении — растягивается.

5.2. Оценки коэффициентов линейной регрессии

Исходя из полученных результатов можно сделать следующие выводы:

- Критерий наименьших квадратов точнее оценивает коэффициенты линейной регрессии на выборке без возмущений, так как $\rho_{\text{МНК}} < \rho_{\text{МНМ}}$.
- Критерий наименьших модулей точнее оценивает коэффициенты линейной регрессии на выборке с возмущениями, так как $\rho_{\text{МНМ}} < \rho_{\text{МНК}}$.
- Критерий наименьших модулей устойчив к редким выбросам по сравнению с критерием наименьших квадратов. Но при этом обладает большей вычислительной сложностью из-за необходимости решения задачи минимизации.

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

5.3.1. Стандартное нормальное распределение

Табличное значение квартиля: $\chi^2_{0.95}(5) = 11.07$.

Так как $\chi^2_B < \chi^2_{0.95}(5)$, то можно заключить, что гипотеза H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$, согласуется с выборкой.

5.3.2. Равномерное распределение

Табличное значение квартиля: $\chi^2_{0.95}(4) = 9.49$.

Так как $\chi^2_B < \chi^2_{0.95}(4)$, то можно заключить, что гипотезу H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$, нельзя опровергнуть. Такое несоответствие действительности можно объяснить довольно малым размером выборки.

5.4. Доверительные интервалы

Исходя из полученных результатов можно сделать следующие выводы:

- Генеральные характеристики ($m = 0$ и $\sigma = 1$) накрываются построенными доверительными интервалами.
- Лучший результат достигается на выборках большого объема, так как получаемые интервалы получаются меньшей длины.
- Доверительные интервалы для параметров нормального распределения более надёжны, так как основаны на точном, а не асимптотическом распределении.

6. Литература

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

2) Correlation and dependence. URL: `https://en.wikipedia.org/wiki/Correlation_and_dependence`;

3) Least squares. URL: `https://en.wikipedia.org/wiki/Least_squares`;

4) Least absolute deviations. URL: `https://en.wikipedia.org/wiki/Least_absolute_deviations`;

5) Maximum likelihood estimation. URL: `https://en.wikipedia.org/wiki/Maximum_likelihood_estimation`;

6) Chi-squared test. URL: `https://en.wikipedia.org/wiki/Chi-squared_test`;

7) Таблица значений χ^2. URL: `http://statsoft.ru/home/textbook/modules/sttable.html#chi`;

8) Confidence interval. URL: `https://en.wikipedia.org/wiki/Confidence_interval`

7. Приложение

1) Код лабораторной №5. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_5/Lab_5.ipynb`;

2) Код лабораторной №6. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_6/Lab_6.ipynb`;

3) Код лабораторной №7. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_7/Lab_7.ipynb`;

4) Код лабораторной №8. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_8/Lab_8.ipynb`;

5) Код общего отчёта №5-8. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_5-8/Lab_report_5-8.tex`