Санкт-Петербургский Политехнический Университет
Петра Великого

Институт прикладной математики и механики

Кафедра "Прикладная математика"

Отчёт
Лабораторная работа №5 – 8
по дисциплине
"Математическая статистика"

Выполнил студент:
Салихов С.Р.
группа: 3630102/70401

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

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, ρ).
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 $y_2 0$ 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.

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

$$N(x,y,\overline{x},\overline{y},\delta_x,\delta_y,\rho) = \frac{1}{2\pi\delta_x\delta_y\sqrt{1-\rho^2}} * exp\left\{ -\tfrac{1}{2(1-\rho^2)}\left[\tfrac{(x-\overline{x})^2}{\delta_x^2} - 2\rho\tfrac{(x-\overline{x})(y-\overline{y})}{\delta_x\delta_y} + \tfrac{(y-\overline{y})^2}{\delta_y^2}\right] \right\}$$

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями x, y и средними квадратическими отклонениями δ_x, δ_y соответственно.
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.

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

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

$$\rho = \frac{K}{\delta_x \delta_y}$$

Коэффициент корреляции — это нормированная числовая характеристика, являющаяся мерой близости зависимости между случайными величинами к линейной.

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Пусть по выборке значений

$\{x_i, y_i\}_1^n$ 1 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, —

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

где K, s_X^2, s_Y^2 - выборочные ковариации и дисперсии с.в. X и Y.

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

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

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 $x' = x - medx, y' = y - medy$ и с центром в точке с координатами (med x, med y).

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 :

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

где $\overline{u} = \overline{\upsilon} = \frac{1+2+..+n}{n} = \frac{n+1}{2}$ - среднее значение рангов.

2.4 Простая линейная регрессия

2.4.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

, i = 1, ... , n, 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, δ) с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины (ненаблюдаемые); β_0, β_1 — неизвестные параметры, подлежащие оцениванию.
В модели 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.

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):

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

Задача минимизации квадратичного критерия (10) носит название задачи метода наименьших квадратов (МНК), а оценки $\hat{\beta}_0$, $\hat{\beta}_1$ параметров β_0, β_1, реализующие минимум критерия, называют МНК-оценками.

2.5.1 Расчётные формулы для МНК-оценок

МНК-оценки параметров $\hat{\beta}_0$ и $\hat{\beta}_1$ находятся из условия обращения функции Q(β_0, β_1) в минимум.
Для нахождения МНК-оценок $\hat{\beta}_0$ и $\hat{\beta}_1$ выпишем необходимые условия экстремума:

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

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

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

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

Тогда:

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

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

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

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

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

2.6 Метод наименьших модулей

Критерий наименьших модулей – заключается в минимизации следующей функции

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

2.7 Метод максимального правдоподобия

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

$$\hat{\theta}_{\text{МП}} = argmax\mathbf{L}(x_1, x_2, \ldots, x_n, \theta)$$

Где **L** это функция правдоподобия, которая представляет собой совместную плотность вероятности независимых случайных величин $X_1, x_2, \ldots, x_n$ и является функцией неизвестного параметра θ

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

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

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

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

2.8 Хи-квадрат

Разобьём генеральную совокупность на k непересекающихся 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 iый 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:

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

Тогда

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

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 распредеелени

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:

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

Функция плотности распределения χ^2:

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

Интервальная оценка математического ожидания:

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

где $t_{1-\frac{a}{2}}$ – квантиль распределения Стьюдента порядка $1 - \frac{a}{2}$.

Интервальная оценка дисперсии:

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

где $\chi^2_{1-\frac{a}{2}},\ \chi^2_{\frac{a}{2}}$ – квантили распределения Стьюдента порядков $1 - \frac{a}{2}$ и $\frac{a}{2}$ 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:

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

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

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

3 Реализация

Для генерации выборки был использован *Python* 3.7 и модуль numpy. Для отрисовки графиков использовался модуль matplotlib. scipy.stats для обработки функций распределений.

4 Результаты

4.1 Выборочные коэффициенты корреляции

Таблица 1: Результаты для двумерного нормального распределения при n = 20

$\rho = 0$	r	r_S	r_Q
E(z)	-0.005	-0.004	-0.002
E(z^2)	0.057	0.057	0.056
D(z)	0.057	0.057	0.056
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.481	0.455	0.315
E(z^2)	0.263	0.242	0.148
D(z)	0.031	0.034	0.048
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.897	0.866	0.699
E(z^2)	0.808	0.754	0.516
D(z)	0.002	0.004	0.027

Таблица 2: Результаты для двумерного нормального распределения при n = 60

$\rho = 0$	r	r_S	r_Q
E(z)	-0.002	-0.002	-0.001
E(z^2)	0.016	0.017	0.017
D(z)	0.016	0.017	0.017
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.499	0.477	0.328
E(z^2)	0.258	0.238	0.122
D(z)	0.009	0.010	0.014
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.898	0.881	0.704
E(z^2)	0.807	0.779	0.504
D(z)	0.0007	0.0011	0.008

Таблица 3: Результаты для двумерного нормального распределения при n = 100

$\rho = 0$	r	r_S	r_Q
E(z)	-0.004	-0.004	-0.002
E(z^2)	0.009	0.009	0.1
D(z)	0.009	0.009	0.010
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.496	0.477	0.330
E(z^2)	0.252	0.234	0.117
D(z)	0.006	0.006	0.009
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.899	0.887	0.710
E(z^2)	0.809	0.787	0.510
D(z)	0.0004	0.0006	0.005

Таблица 4: Смесь нормальных распределений

n = 20	r	r_S	r_Q
E(z)	0.384	0.364	0.251
E(z^2)	0.187	0.174	0.113
D(z)	0.039	0.042	0.051
n = 60	r	r_S	r_Q
E(z)	0.397	0.378	0.260
E(z^2)	0.170	0.156	0.084
D(z)	0.012	0.013	0.016
n = 100	r	r_S	r_Q
E(z)	0.397	0.380	0.259
E(z^2)	0.165	0.152	0.077
D(z)	0.007	0.008	0.009

4.2 Эллипсы рассеивания

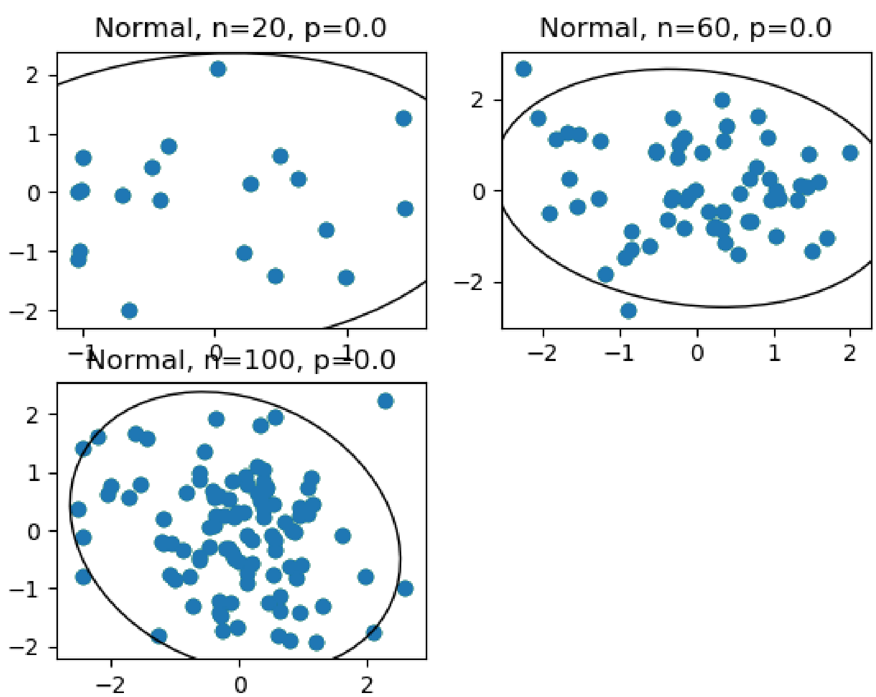

Рис. 1: Двумерное нормальное распределение, при $\rho = 0$

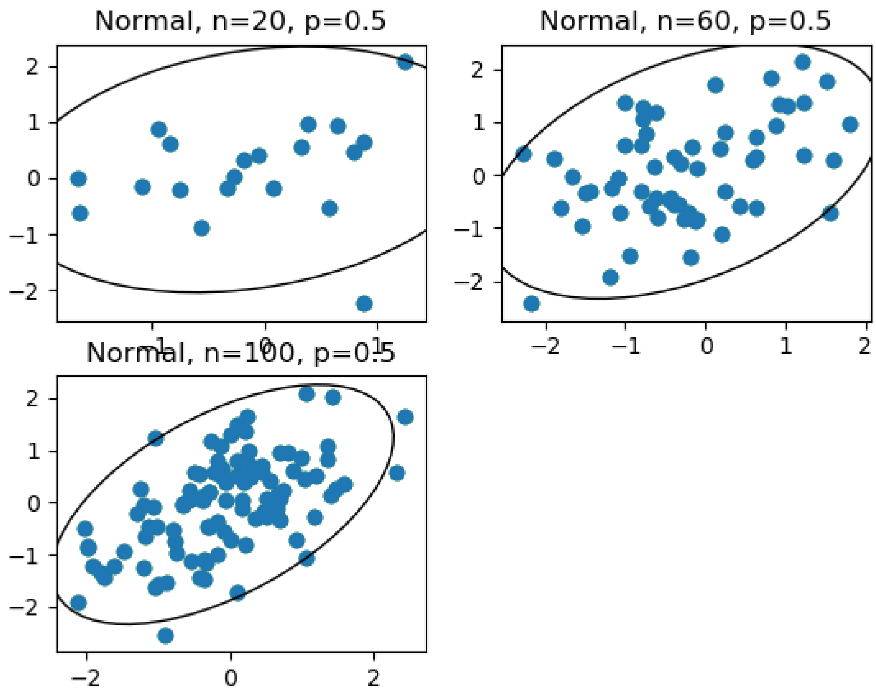

Рис. 2: Двумерное нормальное распределение, при $\rho = 0.5$

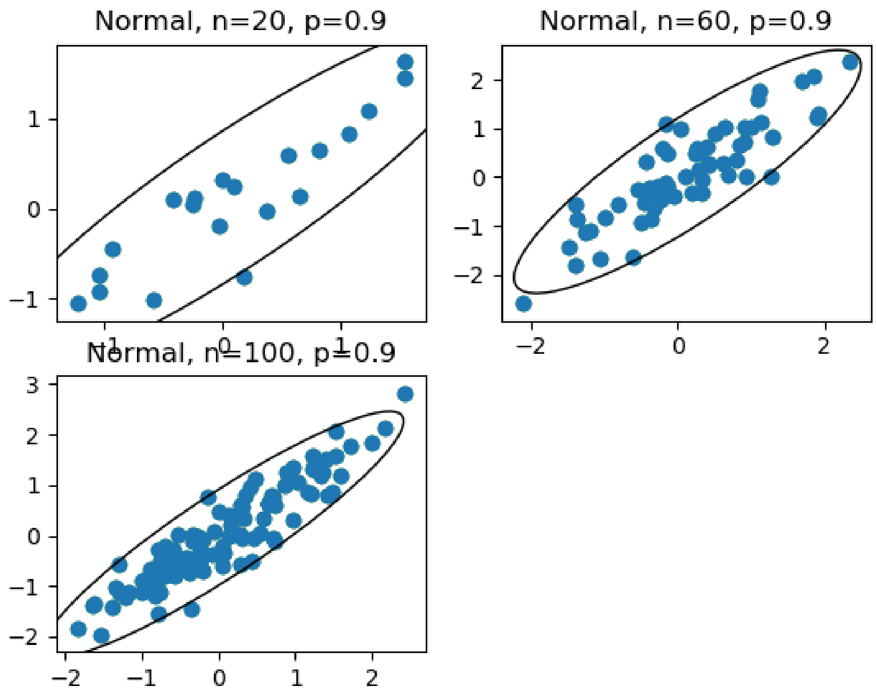

Рис. 3: Двумерное нормальное распределение, при $\rho = 0.9$

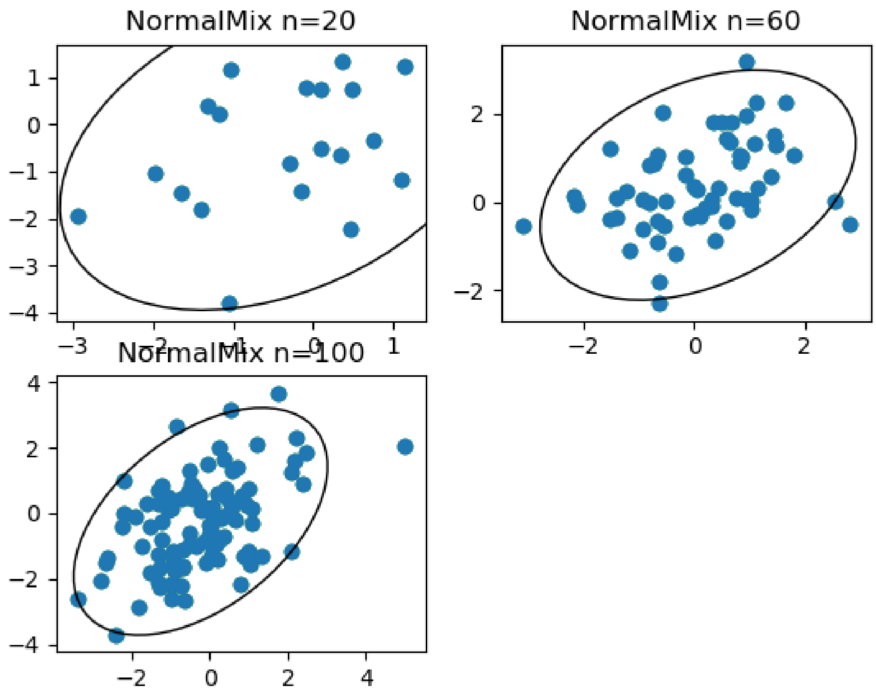

Рис. 4: Смесь нормальных распределений

4.3 Графики линейной регрессии

4.4 Выборка без возмущений

Критерий наименьших квадратов:

$$\hat{a} \approx 1.93, \hat{b} \approx 2.19$$

Критерий наименьших модулей:

$$\hat{a} \approx 2.24, \hat{b} \approx 1.77$$

4.5 Выборка с возмущниями

Критерий наименьших квадратов:

$$\hat{a} \approx 0.48, \hat{b} \approx 1.76$$

Критерий наименьших модулей:

$$\hat{a} \approx 1.85, \hat{b} \approx 1.39$$

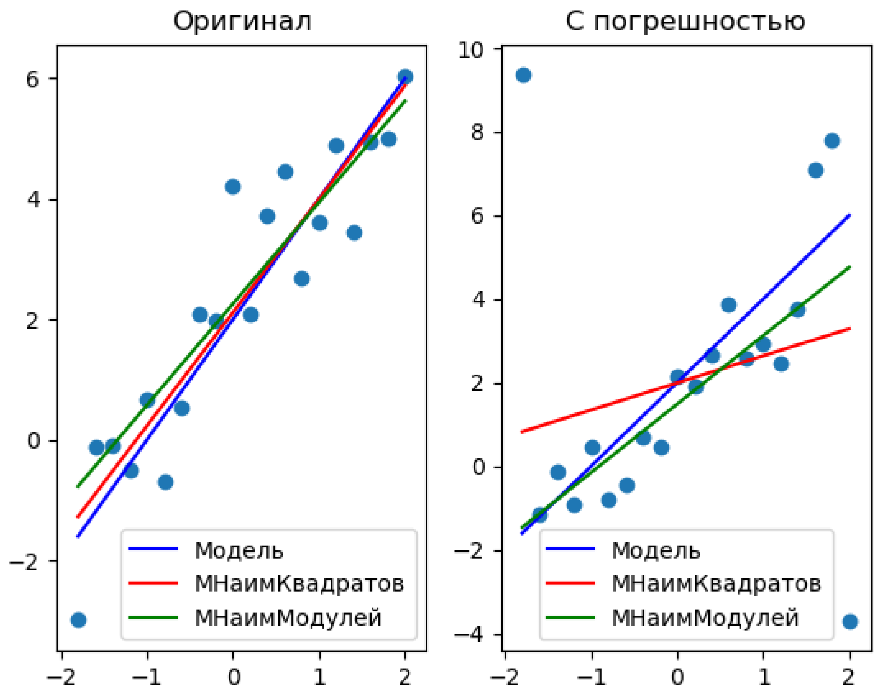

Рис. 5: Графики линейной регрессии при выборке с возмущением и без

4.6 Метод максимального правдоподобия и χ^2

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.03, \hat{\sigma} \approx 0.95$$

Критерий согласия χ^2:

Таблица 5: Вычисление χ^2_B при проверке гипотизы H_0 о нормальном законе распределения N$(x, \hat{\mu} \approx 0.16, \hat{\sigma} \approx 1.04)$

i	$Delta_i$, a_{i-1}, a_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.01]	14	0.1562	15.62	-1.62	0.17
2	[-1.01, -0.37]	19	0.2004	20.04	-1.04	0.05
3	[-0.37, 0.28]	23	0.2517	25.17	-2.17	0.19
4	[0.28, 0.92]	19	0.2122	21.22	-2.22	0.23
5	[0.92, 1.56]	14	0.1201	12.01	1.99	0.33
6	[1.56, ∞]	11	0.0594	5.94	5.06	4.32
$\sum$	-	100	1.0000	100.00	0.00	$5.29 = \chi^2_B$

Количество промежутков k = 6.
Уровень значимости $\alpha = 0.05$.

4.7 Доверительные интервалы

Таблица 6: Доверительные интервалы для параметров нормального распределения

n = 20	m	σ
	-0.48 < m < 0.54	0.83 < σ < 1.59
n = 100	m	σ
	-0.25 < m < 0.15	0.9 < σ < 1.19

Таблица 7: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

n = 20	m	σ
	-0.43 < m < 0.49	0.93 < σ < 1.25
n = 100	m	σ
	-0.25 < m < 0.15	0.95 < σ < 1.09

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

1)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.

2)Процент попавших элементов выборки в эллипс рассеивания (95%-я доверительная область) примерно равен его теоретическому значению (95%). При малом количестве событий $<\approx$ 12 Процент попавших элементов в эллипс рассеивания примерно равен 100%.

5.2 Сравнение с теоретическими данными

По таблицам видно, что при увеличении объёма выборки, подсчитанные коэффициенты стремятся к теоретическим.
По графикам видно, что при уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении растягивается.

5.3 МНК и МНМ

1)МНК оценивает коэффициенты линейной регрессии точнее, на выборке без возмущений.
Для доказательства этого введём метрику суммы квадратов разностей значений по оси y между МНК и модели и МНМ и модели. $\rho_1 = \sum(y_{MNK} - y_{etl})^2$, $\rho_2 = \sum(y_{MNM} - y_{etl})^2$ и увидим, что ρ_1 всегда меньше ρ_2.

Пример:
y_{etl} = [-1.60, -1.20, -0.80, -0.40, 0.00, 0.40, 0.80, 1.20, 1.60, 2.00, 2.40, 2.80, 3.20, 3.60, 4.00, 4.40, 4.80, 5.20, 5.60, 6.0]
$y_{MNK} \approx$ [-1.27, -0.89, -0.52, -0.14, 0.23, 0.61, 0.98, 1.36, 1.74, 2.11, 2.491, 2.86, 3.24, 3.62 3.99 4.37, 4.75, 5.12, 5.50 5.85]
$y_{MNM} \approx$ [-0.77, -0.43, -0.09, 0.23, 0.57, 0.911, 1.24, 1.58, 1.92, 2.25, 2.59, 2.93, 3.26, 3.60, 3.94, 4.27, 4.61, 4.95, 5.28, 5.62]
МНК : $\hat{a} = 1.88$, $\hat{b} = 2.11$
МНМ : $\hat{a} = 2.26$, $\hat{b} = 1.68$
$\rho_1 = 16.59$
$\rho_2 = 17.96$
Таким образом, $\rho_1 < \rho_2$.

2)На выборке с возмущениями эффективнее использовать МНМ. Таким образом, метод наименьших модулей устойчив к редким выбросам, в свою очередь МНМ обладает большей сложностью вычислений, чем МНК(т.к. ,в коде, МНК - оценки вычисляются из расчётных формул, а МНМ - оценки вычисляются, через решение задачи минимизации).

5.4 Критерий Пирсона

По результатам работы, значение критерия согласия Пирсона:
$\chi^2_B = 5.29$. Табличное значение квартиля $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(5) = 11.07$.

Таким образом, $\chi^2_B < \chi^2_{0.95}(5)$, из этого следует, что основная гипотеза H_0(о нормальном законе распределения $\mathrm{N}(x, \hat{\mu}, \hat{\sigma})$), на уровне зависимости $\alpha = 0.05$, соотносится с выборкой.

При распределении Лапласа(n = 30),

$$\hat{\mu} \approx 0.53, \hat{\sigma} \approx 1.52$$

По результатам работы, значение критерия согласия Пирсона:
$\chi^2_B = 12.94$.
Таким образом, $\chi^2_B > \chi^2_{0.95}(5)$, из этого следует, что основная гипотеза H_0(о нормальном законе распределения $\mathrm{N}(x, \hat{\mu}, \hat{\sigma})$), на уровне зависимости $\alpha = 0.05$, при распределении Лапласа не верна.

5.5 Доверительные интервалы

Генеральные характеристики m = 0, $\sigma = 0$ накрываются построенными доверительными интервалами.

По полученным результатам видно, что лучший результат достигается на выборках большого объема.

При сравнении результатов при объеме n = 20, видим, что интервал меньше для параметров произвольного распределения.

6 Литература

Модуль numpy

Модуль matplotlib

Модуль scipy

Многомерное нормальное распределение
Корреляция
Метод наименьших модулей

Метод наименьших квадратов

Таблица значений χ^2

Нормальное распределение

Доверительные интервалы

7 Приложения

Код 5-й лаборатрной
Код 5-о отчёта
Код 6-й лаборатрной
Код 6-о отчёта
Код 7-й лаборатрной
Код 7-о отчёта
Код 8-й лаборатрной
Код 8-о отчёта