ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ
КАФЕДРА «ОБЩЕЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ»

# Курсовая работа по дисциплине «Теория вероятностей и математическая статистика»

Студент: Киселев А.А.
Группа: 5361
Преподаватель: Мартыненко А. И.
Вариант: 7

МОСКВА 2010

# Содержание

**Задание 1.** Для данных двух выборок (табл. 1) построить эмпирические функции распределения и гистограммы относительных частот. С помощью эмпирической функции распределения, нормальной и экспоненциальной вероятностных бумаг для каждой выборки определить вид закона распределения (нормальное, экспоненциальное, равномерное) и оценить его параметры.

*Решение.* Эмпирической функцией распределения случайной величины $\xi$ называется функция $F_э(x) = P_э(\xi < x) = \dfrac{n_x}{n}$, которая при каждом $x$ равна отношению $n_x$ — числа наблюдений, меньших $x$, к объему выборки $n$. Значения данных в каждой выборке упорядочиваются по возрастанию (табл. 2). Объем каждой выборки $n = 25$, поэтому $F_э$ — ступенчатая функция, увеличивающая свое значение на $\dfrac{1}{n}$ при каждом значении $x$, вошедшем в выборку. Построим графики $F_э(x)$ для обеих выборок (рис. 1 и 2).

Табл. 1. Неупорядоченные выборки 1 и 2.

| Выборка 1 | Выборка 2 |
|---|---|
| 1,374713 | 0,591799 |
| −4,440732 | 0,196073 |
| −0,745084 | 0,160135 |
| −0,111441 | 3,518885 |
| −2,801433 | 4,768065 |
| −2,530841 | 0,884235 |
| −3,666468 | 4,209723 |
| −5,581626 | 4,51201 |
| −3,719059 | 1,565542 |
| −0,078647 | 0,423245 |
| −0,438544 | 1,063247 |
| 1,393335 | 1,157221 |
| −2,970003 | 2,548808 |
| −3,443457 | 1,30273 |
| −0,546616 | 0,82157 |
| −4,46825 | 1,600444 |
| −3,621161 | 7,929842 |
| 0,932521 | 0,353165 |
| −3,513652 | 0,306291 |
| −0,875469 | 3,442844 |
| −4,283991 | 0,900104 |
| −2,251157 | 2,341499 |
| 0,252497 | 3,137087 |
| −0,62873 | 1,093911 |
| 1,00652 | 0,36036 |

Табл. 2. Отсортированные выборки 1 и 2.

| Выборка 1 | Выборка 2 |
|---|---|
| −5,581626 | 0,160135 |
| −4,46825 | 0,196073 |
| −4,440732 | 0,306291 |
| −4,283991 | 0,353165 |
| −3,719059 | 0,36036 |
| −3,666468 | 0,423245 |
| −3,621161 | 0,591799 |
| −3,513652 | 0,82157 |
| −3,443457 | 0,884235 |
| −2,970003 | 0,900104 |
| −2,801433 | 1,063247 |
| −2,530841 | 1,093911 |
| −2,251157 | 1,157221 |
| −0,875469 | 1,30273 |
| −0,745084 | 1,565542 |
| −0,62873 | 1,600444 |
| −0,546616 | 2,341499 |
| −0,438544 | 2,548808 |
| −0,111441 | 3,137087 |
| −0,078647 | 3,442844 |
| 0,252497 | 3,518885 |
| 0,932521 | 4,209723 |
| 1,00652 | 4,51201 |
| 1,374713 | 4,768065 |
| 1,393335 | 7,929842 |

Гистограмма относительных частот является ступенчатым приближением плотности распределения случайной величины $\xi$. Для ее построения разделим весь диапазон наблюдений на $s$ полуинтервалов вида $[a_{j-1}; a_j)$ и определим количество наблюдений $m_j$, попавших в $j$-й в полуинтервал. Относительная частота наблюдений, попавших в $j$-й полуинтервал, равна $P_j^* = \dfrac{m_j}{n}$. Так как $m_1 + \cdots + m_s = n$, то сумма всех частот

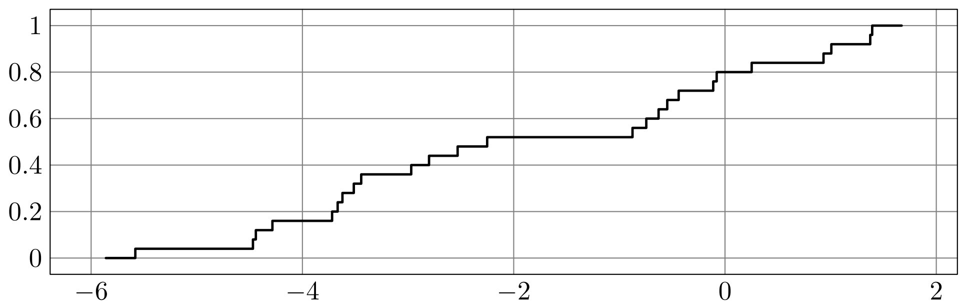


Рис. 1. Эмпирическая функция распределения выборки 1.

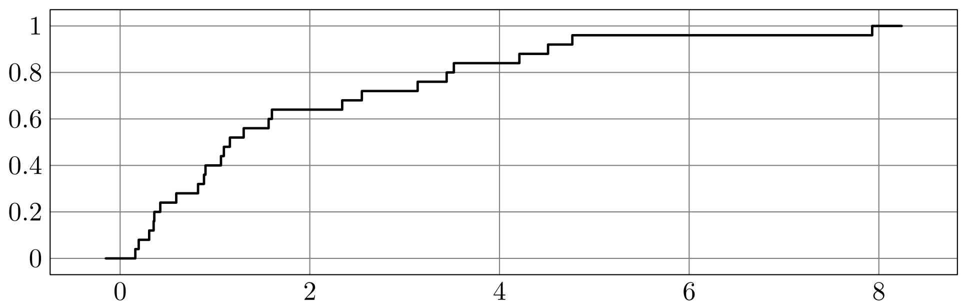


Рис. 2. Эмпирическая функция распределения выборки 2.

равна единице: $\sum\limits_{j=1}^{s} P_j^* = \sum\limits_{j=1}^{s} \dfrac{m_j}{n} = \dfrac{1}{n} \cdot \sum\limits_{j=1}^{s} m_j = \dfrac{1}{n} \cdot n = 1$. Над каждым полуинтервалом $j$ построим прямоугольник высоты $h_j = \dfrac{m_j}{n(a_j - a_{j-1})}$. Полученная фигура, состоящая из этих прямоугольников, называется гистограммой относительных частот. Ее площадь должна быть равна единице. Соответственно, площадь каждого прямоугольника равна относительной частоте наблюдений $P_j^*$, попавших в этот интервал.

Для построения гистограмм следует сгруппировать данные выборок. Для этого разобъем область данных на несколько интервалов равной длины. Для первой выборки длину интервала $[a_{j-1}; a_j)$ возьмем равной 1,4, за $a_0$ возьмем минус 6. Для второй выборки длину интервала $[a_{j-1}; a_j)$ возьмем равной 1,6, за $a_0$ возьмем 0,0. Вычислим относительные частоты попадания в эти интервалы (табл. 3 и 4) и над каждым из них построим столбец соответствующей высоты (рис. 3 и 4).

Табл. 3. Выборка 1, относительные частоты.

| Интервал | Отн. частота |
|---|---|
| $[-6; -4{,}6)$ | 0,0285714285714286 |
| $[-4{,}6; -3{,}2)$ | 0,228571428571429 |
| $[-3{,}2; -1{,}8)$ | 0,114285714285714 |
| $[-1{,}8; -0{,}4)$ | 0,142857142857143 |
| $[-0{,}4; 1{,}0)$ | 0,114285714285714 |
| $[1{,}0; 2{,}4)$ | 0,0857142857142857 |

Табл. 4. Выборка 2, относительные частоты.

| Интервал | Отн. частота |
|---|---|
| $[0{,}0; 1{,}6)$ | 0,375 |
| $[1{,}6; 3{,}2)$ | 0,1 |
| $[3{,}2; 4{,}8)$ | 0,125 |
| $[4{,}8; 6{,}4)$ | 0,0 |
| $[6{,}4; 8{,}0)$ | 0,025 |

На основании построенных гистограмм следует выдвинуть и проверить ряд статистических гипотез о виде и параметрах распределения обеих случайных величин.

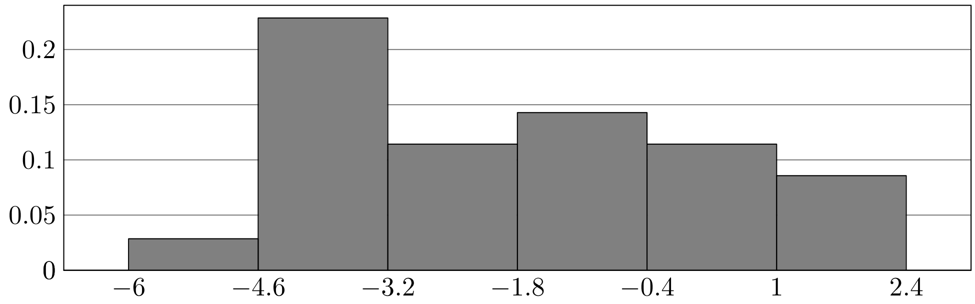


Рис. 3. Гистограмма относительных частот выборки 1.

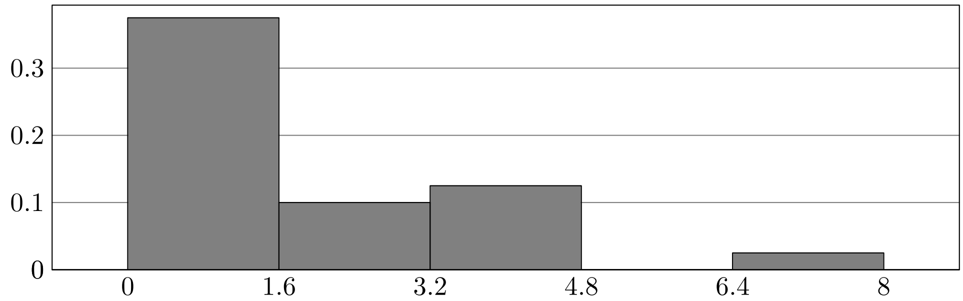


Рис. 4. Гистограмма относительных частот выборки 2.

Из построенных гистограмм видно, что из трех гипотез о нормальном, равномерном и экспоненциальном распределении для первой выборки наиболее правдоподобна гипотеза о равномерном, а для второй — экспоненциальном распределении. Дополнительные основания для принятия решения в каждом случае даст построение так называемых вероятностных бумаг.

Дадим определение вероятностной бумаге. Предположим, что закон теоретической функции распределения известен (например, нормальное, экспоненциальное или равномерное распределение), но неизвестны параметры этого распределения. Нормальное и равномерное распределения определяются двумя параметрами $\alpha$ и $\beta$ — математическим ожиданием и среднеквадратичным отклонением, а экспоненциальное распределение одним параметром $\lambda$ — математическим ожиданием. Тогда теоретическую функцию распределения можно рассматривать как зависящую от переменной $x$ и соответствующих параметров: $F = F(x, \alpha, \beta)$ или $F = F(x, \lambda)$. Выберем такую систему координат $u = u(x)$, $v = v(x)$, чтобы график теоретической функции распределения в новых координатах $(u, v)$ стал прямой линией $v = ku+b$ с коэффициентами $k$ и $b$, зависящими от неизвестных параметров $\alpha$ и $\beta$: $k = \varphi(\alpha, \beta)$, $b = \psi(\alpha, \beta)$. Для экспоненциального распределения существует такая система координат, в которой уравнение прямой имеет вид $v = ku$, где $k = \varphi(\lambda)$.

Поскольку эмпирическая функция распределения $y = F_э(x)$ при достаточно большом объеме выборки лежит вблизи от теоретической функции рапределения $y = F(x, \alpha, \beta)$ (или при экспоненциальном распределении $y = F(x, \lambda)$), то после замены переменных график эмпирической функции должен лежать вблизи искомой прямой. Новая система координат $(u, v)$ с нанесенными соответствующими значениями $(x, y)$, преобразованными в новые координаты, называется вероятностной бумагой. Построив в новых координатах график эмпирической функции распределения, подбирают и скомую прямую так, чтобы по обе стороны от нее находилось примерно одинаковое количество «ступенек» построенного графика (это можно делать произвольно, а можно воспользоваться, например методом наименьших квадратов). Затем определяют $k$ — величину тангенса угла, образованного этой прямой с осью $Ou$, и $b$ — координату пересечения с осью $Ov$. Приравняв полученные величины к $\varphi(\alpha,\beta)$ и $\psi(\alpha,\beta)$, находят оценки неизвестных параметров $\alpha$ и $\beta$ из системы уравнений:

$$\begin{cases} k = \varphi(\alpha,\beta), \\ b = \psi(\alpha,\beta). \end{cases}$$

Если неизвестный параметр один (в случае экспоненциального распределения), то уравнение будет единственным.

Для нормального закона распределения

$$F(x,m,\sigma) = \Phi\left(\frac{x-m}{\sigma}\right) + 0.5,$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{x^2}{2}}dx$ — функция Лапласа. Сделаем следующую замену координат:

$$u = x, v = \Phi^{-1}(y-0.5),$$

где $x = \Phi^{-1}(y)$ — функция, обратная к функции Лапласа. В новых координатах $(u;v)$ уравнение функции распределения $y = \Phi(\frac{x-m}{\sigma}) + 0.5$, примет следующий вид:

$$v = \Phi^{-1}(y-0.5) = \Phi^{-1}\left(\Phi\left(\frac{x-m}{\sigma}\right) + 0.5 - 0.5\right) =$$
$$= \Phi^{-1}\left(\Phi\left(\frac{x-m}{\sigma}\right)\right) = \frac{x-m}{\sigma} = \frac{u-m}{\sigma}.$$

Получившееся уравнение $v = \dfrac{u-m}{\sigma} = \dfrac{1}{\sigma}u + (-\dfrac{m}{\sigma})$ есть уравнение искомой прямой линии. Отсюда

$$\left\{\begin{array}{rcl} k & = & \dfrac{1}{\sigma} \\ b & = & -\dfrac{m}{\sigma}. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{rcl} \sigma & = & \dfrac{1}{k} \\ m & = & -\dfrac{b}{k}. \end{array}\right.$$

Теоритическая функция распределения экспоненциально распределенной случайной величины имеет вид:

$$y = F(x,\lambda) = \begin{cases} 1 - e^{-\lambda x}, & \text{если } x \geq 0; \\ 0, & \text{иначе.} \end{cases}$$

Здесь применяют следующую замену координат:

$$u = x, v = -\ln(1-y).$$

В новых координатах уравнение теоритической функции распределения $y = F(x,\lambda)$ примет такой вид:

$$v = -\ln(1-y) = -\ln(1-F(x,\lambda)) = -\ln(1-(1-e^{-\lambda x})) = -\ln(e^{-\lambda x}) = \lambda x = \lambda u \ (u \geq 0).$$

Заметим, что прямая $v = \lambda u$ должна проходить через начало координат.

Для равномерного распределения имеем следующую теоритическую функцию распределения случайной величины:

$$y = F(x, \alpha, \beta) = \begin{cases} 1, & \text{если } x > \beta; \\ \frac{x-\alpha}{\beta-\alpha}, & \text{если } x \in [\alpha, \beta]; \\ 0, & \text{иначе.} \end{cases}$$

Данная функция уже имеет линейный вид, поэтому можно обойтись тождественным преобразованием координат $u = x$, $v = y$. При этом для равномерного распределения известны следующие выражения для математического ожидания и среднеквадратичного отклонения:

$$\begin{cases} m &= \frac{\alpha+\beta}{2} \\ \sigma &= \frac{\beta-\alpha}{2\sqrt{3}}. \end{cases}$$

Имея на вероятностной бумаге линию $kx + b$, исходя из этого, параметры распределения можно оценить следующим образом:

$$\begin{cases} \alpha &= -\frac{b}{k} \\ \beta &= \frac{1-b}{k}. \end{cases} \iff \begin{cases} m &= \frac{1-2b}{2k} \\ \sigma &= \frac{1}{2k\sqrt{3}}. \end{cases}$$

Для каждой выборки построим вероятностные бумаги равномерного, нормального и экспоненциального распределений (для выборки 1 — рис. 5, 6 и 7, для выборки 2 — рис. 8, 9 и 10).

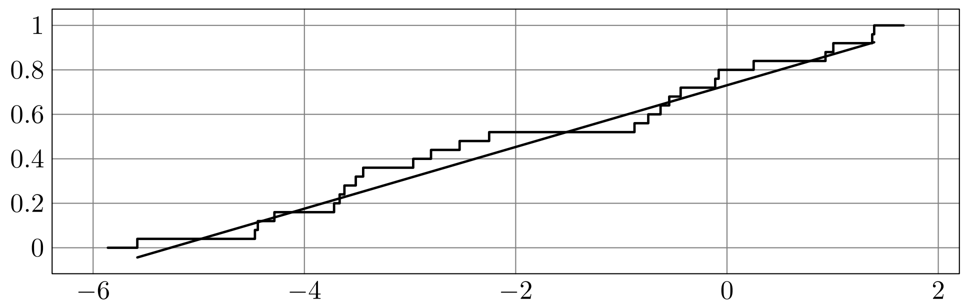


Рис. 5. Равномерная вероятностная бумага для выборки 1.

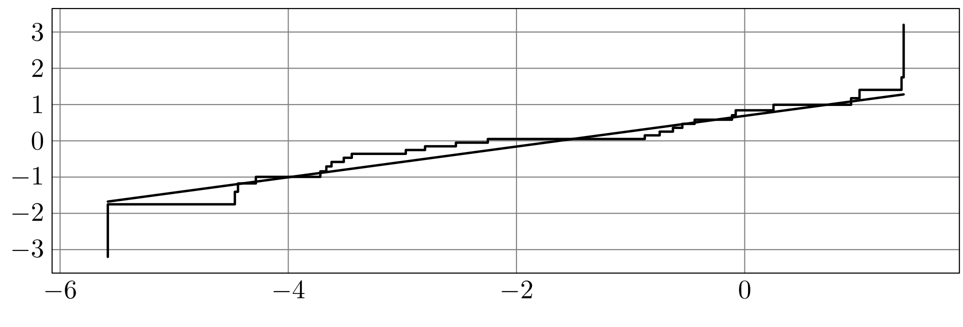


Рис. 6. Нормальная вероятностная бумага для выборки 1.

График эмпирической функции распределения выборки 1 наилучшим образом приближается прямой на равномерной вероятностной бумаге (рис. 5). Оценим параметры распределения выборки 1. Аппроксимирующая прямая $y = kx + b$ выбрана так, что имеет параметры $k = 0{,}1387$ и $b = 0{,}7307$. Откуда $\sigma = \dfrac{1}{2\sqrt{3}k} \approx 2{,}0807$ и $m = \dfrac{1-2b}{2k} \approx -1{,}6632$.

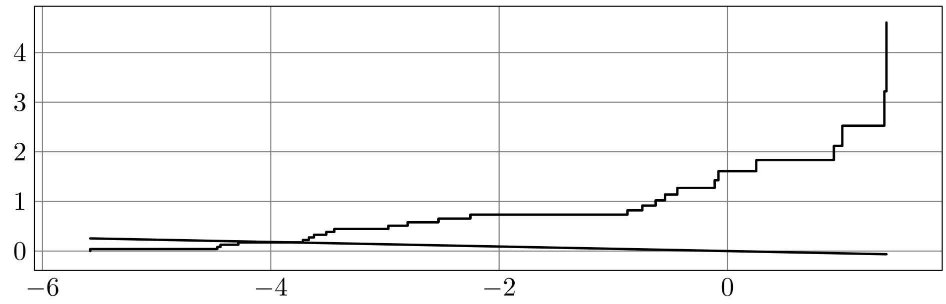


Рис. 7. Экспоненциальная вероятностная бумага для выборки 1.

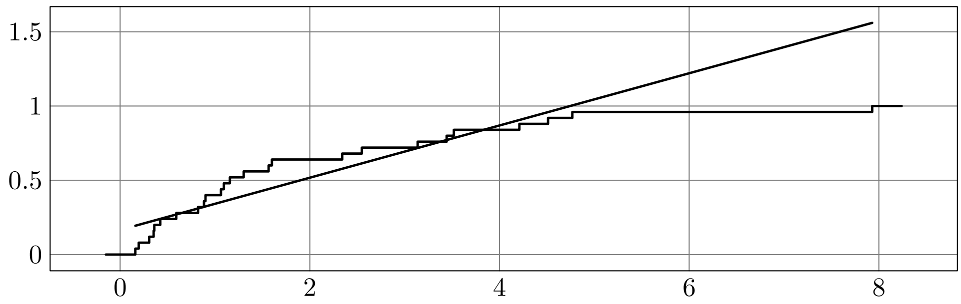


Рис. 8. Равномерная вероятностная бумага для выборки 2.

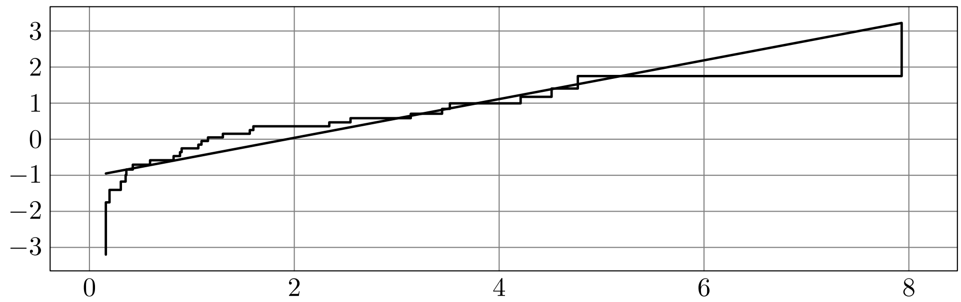


Рис. 9. Нормальная вероятностная бумага для выборки 2.

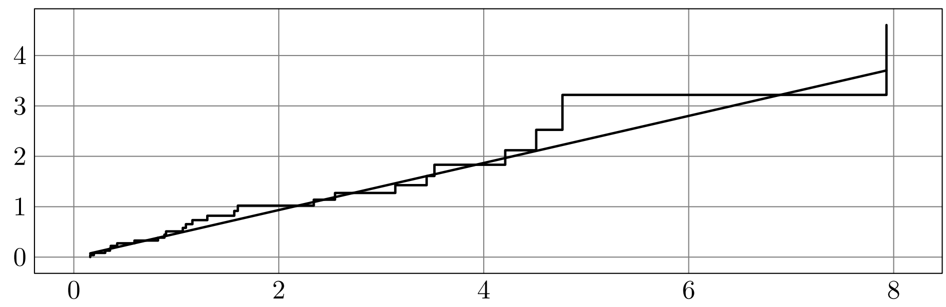


Рис. 10. Экспоненциальная вероятностная бумага для выборки 2.

Для выборки 2 удалось подобрать прямую для вероятностной бумаги экспоненциального распределения, проходящую через начало координат, которая приближает эмпирическую функцию распределения выборки 2 лучше, чем прямые на других вероятностных бумагах. Поэтому, предположим, что выборка 2 имеет экспоненциальное рапределение и оценим его параметры. Прямая на вероятностной бумаге (рис. 10) имеет параметры $k = 0{,}467$ и $b = 0$, соответственно параметр распределения $\lambda = 0{,}467$.

*Вывод.* На основании построенных гистограмм относительных частот и анализа равномерной, нормальной и экспоненциальной вероятностных бумаг можно выдвинуть гипотезы о равномерном распределении выборки 1 с параметрами $\sigma = 2{,}0807$ и $m = -1{,}6632$, и экспоненциальном распределении выборки 2 с параметрами $\lambda = 0{,}467$.

**Задание 3.** Для двух выборок, приведенных в задании 1, найти выборочные средние, выборочные дисперсии и средние квадратичные отклонения, выборочный коэффициент корреляции, написать уравнения выборочных прямых регрессии и построить их.

*Решение.* Простейшей характеристикой распределения является выборочное среднее, которое для простой статистической совокупности объема $n$ вычисляется по формуле:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \text{ где } x_i \text{ — очередной элемент выборки.}$$

Выборочное среднее является приближенной оценкой математического ожидания изучаемой случайной величины. Для характеристики разброса значений случайной величины относительно ее среднего значения используются выборочная дисперсия $S^2$, являющаяся оценкой теоритической дисперсии. Для простой совокупности выборочная дисперсия вычисляется по формуле:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = \overline{(x - \bar{x})^2}.$$

Выборочное среднеквадратичное отклонение $S = \sqrt{S^2}$ является оценкой теоретического среднеквадратического отклонения $\sigma$.

На практике значение $S^2$ вычисляется по более простой формуле:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2) = \overline{x^2} - \bar{x}^2.$$

Если объем выборки невелик, то $S^2$ является смещенной оценкой теоритической дисперсии: $M(S^2) \neq D(\xi)$. Чтобы получить несмещенную оценку, вводят поправочный коэффициент $\frac{n}{n-1}$ и получают так называемую исправленную поправочную дисперсию

$$S^{*2} = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{\sum\limits_{i=1}^{k} x_i^2 - n\bar{x}^2}{n-1}.$$

Вычислим требуемые параметры для обеих выборок по формулам для простой совокупности и для данных, сгруппированных при построения гистограммы в задании 1. Результаты привидены в табл. 7.

Табл. 5. Точечные оценки параметров распределения для выборок 1 и 2.

| | $\bar{x}$ | $S^2$ | $S$ | $S^{*2}$ | $S^*$ |
|---|---|---|---|---|---|
| Выборка 1 | -1,830271 | 4,247264 | 2,060889 | 4,424233 | 2,103386 |
| Выборка 2 | 1,967553 | 3,462236 | 1,860708 | 3,606495 | 1,899078 |

Далее вычисляется выборочной коэффициент корреляции и строятся выборочные прямые регрессии. Вместо теоритического коэффициента корреляции и теоритических среднеквадратических отклонений вычисляются выборочные их оценки по данным выборки случайной величины $(\xi; \zeta)$ объема $n$.

Вначале строится корреляционная таблица, каждая $i$-я строка таблицы соответствует значению $\xi_i$, а каждый $j$-й столбец — значению $\zeta_i$. На пересечении строки $i$

Табл. 6. Структура корреляционной таблицы.

| $\xi \backslash \zeta$ | $y_1$ | $y_2$ | $\cdots$ | $y_s$ | $n_{i\bullet} = \sum\limits_{j=1}^{s} n_{ij}$ |
|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1s}$ | $n_{1\bullet}$ |
| $x_2$ | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2s}$ | $n_{2\bullet}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $x_k$ | $n_{k1}$ | $n_{k2}$ | $\cdots$ | $n_{ks}$ | $n_{k\bullet}$ |
| $n_{\bullet j} = \sum\limits_{i=1}^{k} n_{ij}$ | $n_{\bullet 1}$ | $n_{\bullet 2}$ | $\cdots$ | $n_{\bullet s}$ | $\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{s} n_{ij}$ |

и столбца $j$ записывается число, показывающее, сколько раз в выборке встретилась пора $(\xi_i; \zeta_j)$. Если таких пар нет, клетка остается пустой. При обработке корреляционной таблицы в последнем столбце указывают сумму частот по строкам, а в последней строке — сумму частот по столбцам. Структура корреляционной таблицы представлена в талице 6.

Первый и последний столбцы корреляционной таблицы образуют статистическое распределение выборки случайной величины $\xi$, а первая и последняя строки образуют выборку случайной величины $\zeta$. Выборочные числовые характеристики по этим данным вычисляются по следующим формулам:

$$\bar{x} = \frac{\sum\limits_{i=1}^{k} n_{i\bullet} x_i}{n}, \; S_x^{*2} = \frac{\sum\limits_{i=1}^{k} n_{i\bullet} x_i^2 - n\bar{x}^2}{n},$$

$$\bar{y} = \frac{\sum\limits_{i=1}^{k} n_{\bullet j} y_i}{n}, \; S_y^{*2} = \frac{\sum\limits_{i=1}^{k} n_{\bullet j} y_i^2 - n\bar{y}^2}{n}.$$

Далее вычисляется выборочный коэффициент корреляции $r^*_{xy}$, являющийся статистической оценкой теоретического коэффициента корреляции:

$$r^*_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{S^*_x S^*_y}, \text{ где } \overline{xy} = \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{s} n_{ij} x_i y_j}{n}.$$

Выборочные уравнения прямых регрессии $\zeta$ на $\xi$ и $\xi$ на $\zeta$ имеют вид:

$$y = r^*_{xy}\frac{S^*_y}{S^*_x}(x - \bar{x}) + \bar{y} \text{ (уравнение прямой регрессии } \zeta \text{ на } \xi),$$

$$x = r^*_{xy}\frac{S^*_x}{S^*_y}(y - \bar{y}) + \bar{x} \text{ (уравнение прямой регрессии } \xi \text{ на } \zeta).$$

Они задают линейную зависимость условных выборочных математических ожиданий (вернее, условных средних) одной случайной величины (левые части уравнений) от значения другой случайной величины ($x$ или $y$ в правых частях уравнений).

В предположении что $i$-е значения выборок образуют $i$-ю пару $(\xi_i; \zeta_i)$ $(1 \leqslant i \leqslant 25)$, составим корреляционную таблицу (табл. **??**). Также схематически изобразим выборку случайной величины $(\xi; \zeta)$, отметив пары $(x_i, y_i)$ на плоскости $Oxy$ (рис. 11)

| $\xi \setminus \zeta$ | 0,160135 | 0,196073 | 0,306291 | 0,353165 | 0,36036 | 0,423245 | 0,591799 | 0,82157 | 0,884235 | 0,900104 | 1,063247 | 1,093911 | 1,157221 | 1,30273 | 1,565542 | 1,600444 | 2,341499 | 2,548808 | 3,137087 | 3,442844 | 3,518885 | 4,209723 | 4,51201 | 4,768065 | 7,929842 | $n_{i\bullet}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -5,581626 | | | | | | | | | | | | | | | | | | | | | | | 1 | | | 1 |
| -4,46825 | | | | | | | | | | | | | | | | 1 | | | | | | | | | | 1 |
| -4,440732 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | 1 |
| -4,283991 | | | | | | | | | | 1 | | | | | | | | | | | | | | | | 1 |
| -3,719059 | | | | | | | | | | | | | | | 1 | | | | | | | | | | | 1 |
| -3,666468 | | | | | | | | | | | | | | | | | | | | | | 1 | | | | 1 |
| -3,621161 | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 1 |
| -3,513652 | | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 |
| -3,443457 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | 1 |
| -2,970003 | | | | | | | | | | | | | | | | | | 1 | | | | | | | | 1 |
| -2,801433 | | | | | | | | | | | | | | | | | | | | | | | | 1 | | 1 |
| -2,530841 | | | | | | | | | 1 | | | | | | | | | | | | | | | | | 1 |
| -2,251157 | | | | | | | | | | | | | | | | | 1 | | | | | | | | | 1 |
| -0,875469 | | | | | | | | | | | | | | | | | | | | 1 | | | | | | 1 |
| -0,745084 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | 1 |
| -0,62873 | | | | | | | | | | | | 1 | | | | | | | | | | | | | | 1 |
| -0,546616 | | | | | | | | 1 | | | | | | | | | | | | | | | | | | 1 |
| -0,438544 | | | | | | | | | | | 1 | | | | | | | | | | | | | | | 1 |
| -0,111441 | | | | | | | | | | | | | | | | | | | | | 1 | | | | | 1 |
| -0,078647 | | | | | | 1 | | | | | | | | | | | | | | | | | | | | 1 |
| 0,252497 | | | | | | | | | | | | | | | | | | | 1 | | | | | | | 1 |
| 0,932521 | | | | 1 | | | | | | | | | | | | | | | | | | | | | | 1 |
| 1,00652 | | | | | 1 | | | | | | | | | | | | | | | | | | | | | 1 |
| 1,374713 | | | | | | | 1 | | | | | | | | | | | | | | | | | | | 1 |
| 1,393335 | | | | | | | | | | | | | 1 | | | | | | | | | | | | | 1 |
| $n_{\bullet j}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 25 |

Табл. 7. Корреляционная таблица выборок 1 и 2, $x_i$ соответствуют упорядоченным значениям из выборки 1, а $y_j$ — упорядоченным значениям из выборки 2, приведенным в табл. 2.

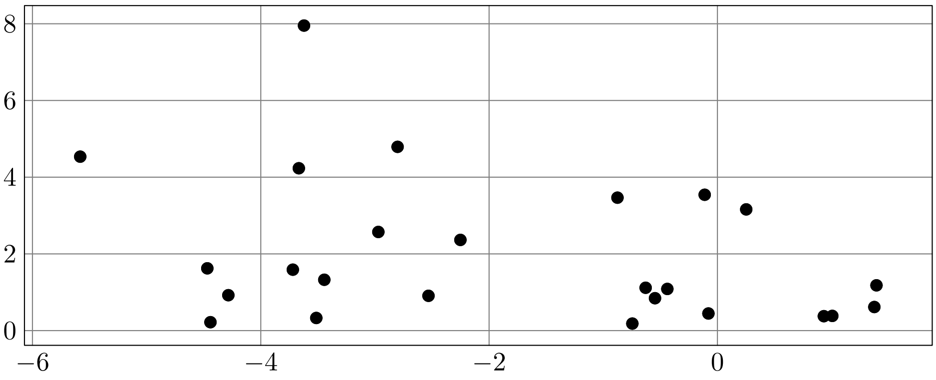


Рис. 11. Выборка случайной величины $(\xi;\zeta)$.

Коэффициент корреляции $r^*_{xy} = -0{,}323006$.
Уравнения прямых регрессии:

$$y_x = -0{,}2916x + 1{,}4338,$$

$$x_y = -0{,}3578y - 1{,}1264.$$

Графики прямых регрессии изображены на рисунке 12.

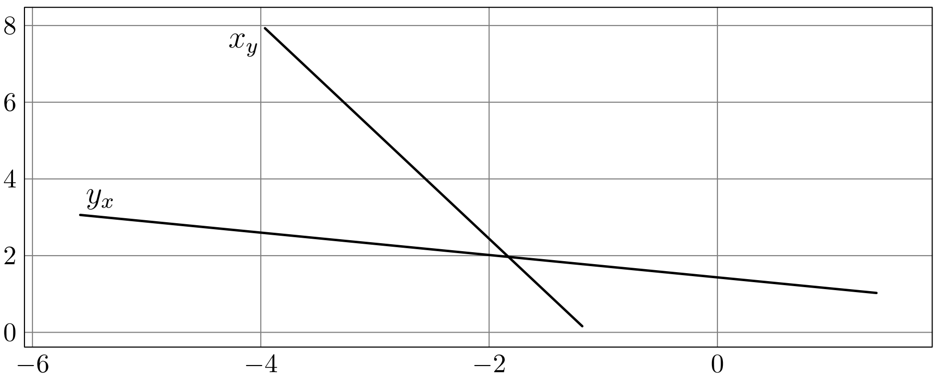


Рис. 12. Графики прямых регрессии.

**Задание 4.** По двум данным в задании 1 выборкам проверить гипотезу о некоррелированности соответствующих случайных величин с уровнем значимости 0,1.

*Решение.* Пусть на основании данных корреляционной таблицы найден выборочный коэффициент корреляции $r^*_{xy}$, который оказался отличным от нуля. Так как выборка отобрана случайно, то возникает вопрос о том, будет ли отличен от нуля теоретический коэффициент корреляции $r_{\xi\zeta}$. Необходимо при заданном уровне значимости $\alpha$ проверить гипотезу $H_0$ о том, что при конкурирующей гипотезе $H_1$: $r_{\xi\zeta} \neq 0$. Если $H_0$ отвергается, то выборочный коэффициент корреляции значимо отличается от нуля, а случайные величины $\xi$ и $\zeta$ коррелированны. Если же $H_0$ принимается, то делаем вывод, что выборочный коэффициент корреляции незначимо отличается от нуля, а случайные величины $\xi$ и $\zeta$ некоррелированны.

В качестве критерия для проверки $H_0$ выбирается случайная величина

$$T = r^*_{xy}\frac{\sqrt{n-2}}{\sqrt{1-{r^*_{xy}}^2}}.$$

При справедливости гипотезы $H_0$ величина $T$ имеет так называемое распределение Стьюдента с $n-2$ степенями свободы. Критическая область для рассматриваемой гипотезы будет двусторонней, $T_{\text{крит1}} = -T_{\text{крит2}}$. Критическое значение $T_{\text{крит2}}$ определяется по заданным уровню значимости $\alpha$ и по числу степеней свободы $n-2$. Если $|T_{\text{набл}}| \geqslant t_{\text{крит2}}$, то гипотеза $H_0$ отвергается с уровнем значимости $\alpha$; при $|T_{\text{набл}}| < t_{\text{крит2}}$ нет оснований отвергнуть $H_0$ с уровнем значимости $\alpha$.

Пусть $H_0$ — нулевая гипотеза, что $r_{xy} = 0$, то есть что случайные величины некоррелированны; и $H_1$ — альтернативная гипотеза: $r_{xy} \neq 0$. Вычислим критерий:

$$T_{\text{набл}} = r^*_{xy}\frac{\sqrt{n-2}}{\sqrt{1-{r^*_{xy}}^2}}.$$

Выборочный коэффициент корреляции вычислен в задании 3: $r^*_{xy} = -0{,}515753$, $n = 25$, откуда:

$$T_{\text{набл}} = -0{,}515753 \cdot \frac{\sqrt{23}}{\sqrt{1-(-0{,}515753)^2}} \approx -2{,}887075.$$

Имея в виду, что уровень значимости $\alpha$ и число степеней свободы для распределения Стьюдента равно $25-2 = 23$, найдем $T_{\text{крит2}} = 1{,}71387$.

$|T_{\text{набл}}| \geqslant t_{\text{крит2}}$, следовательно, гипотеза $H_0$ отвергается с уровнем значимости $\alpha = 0{,}1$.

*Вывод.* После проверки гипотезы о значимости выборочного коэффициента корреляции для выборки случайной величины $(\xi;\zeta)$, было сделано заключение о том, что гипотезу $H_0 : r_{\xi\zeta} = 0$ необходимо отвергнуть с уровнем значимости $\alpha = 0{,}1$. Это означает, что случайные величины коррелированны, выборочный коэффициент корреляции значительно отличается от нуля.