Понятие и задачи мат. Статистики.

Определение: Математической статистикой называется математическая наука, изучающая закономерности массовых, случайных явлений на основе анализа методами теории вероятности, результатов наблюдений их экспериментов.

Задачи мат. статистики:

1) Сбор и группировка данных.
2) Анализ данных
2.1 оценка неизвестной вероятности,
2.2 оценка неизвестных параметров заданного закона распределения,
2.3 оценка гипотезы о неизвестном законе распределения,
2.4 определение связи между случайными величинами.

Виды экспериментов

1. пассивный, заключающийся в обработке уже имеющихся данных, влиять на которые невозможно (например: контроль размеров изделия)
2. активный, заключающийся в планировании эксперимента с целью получить такие данные, которые можно достаточно легко обрабатывать (например: создание новых материалов с заданными свойствами)

Генеральная и выборочная совокупности.

Определение: Генеральной называется совокупность всех исследуемых объектов.
Определение: Выборочной совокупностью (выборкой) называется часть генеральной совокупности, по элементам которой можно судить о характеристиках генеральной совокупности.
Определение: Количество элементов входящих в совокупность называется объемом.
Обозначения: N - объем генеральной совокупности, n - объем выборки, причем n не должна быть меньше 100.

Виды выборок.

1. Повторная – элементы которой после испытаний возвращаются в генеральную совокупность (например: измерение размеров детали).
2. Безповторная – элементы которой не возвращаются в генеральную совокупность (например: определение прочности материала под прессом).
3. Репрезентативная (представительная) – элементы которой выбираются случайным образом и имеют одинаковую вероятность попадания в выборку (для этого используются или генератор случайных чисел ЭВМ, или карточки с инвентарными номерами).

Способы организации выборки.

1. Простой случайный отбор – при котором из всей генеральной совокупности случайным образом выбираются элементы выборки.
2. Частичный типовой отбор – когда генеральная совокупность разбивается на типовые части, и из каждой части случайным образом выбирается объект.
3. Частичный механический отбор – из генеральной совокупности выбирается каждый k - тый объект.
4. Частичный серийный отбор – генеральная совокупность разбивается на части, и все объекты случайно выбранной части подвергаются сплошному исследованию.

Распределение выборки. Вариационный ряд. Эмпирическая функция распределения.

Обозначения: $X(x_1,x_2,\ldots,x_n)$ – выборка, x_i – элементы выборки (варианты), каждая варианта может встречаться в результате испытаний с частотой n_i (для проверки $\sum_{i=1}^{m} n_i = n$ - объём выборки), относительная частота: $W_i = n_i / n$

Определение: Распределением выборки называется соответствие между вариантами и частотами или относительными частотами.

Определение: Вариационным рядом называется выборка, элементы которой расположены в порядке возрастания значений, соответствующих этим элементам.

Определение: Эмпирической функцией распределения называется функция $F^*(x)$, которая определяет относительные частоту появления события $X<x_m$:

$$F^*(x_m) = \sum_{i=1}^{m} W_i$$

Полигон и гистограмма.

Определение: Дискретной случайной величиной называется величина, которая принимает отдельные, независимые друг от друга значения.
Определение: Полигоном называется ломанная линия, соединяющая точки с координатами (x_i,n_i) или (x_i,W_i).
Замечание: полигон строится для дискретной случайной величины.

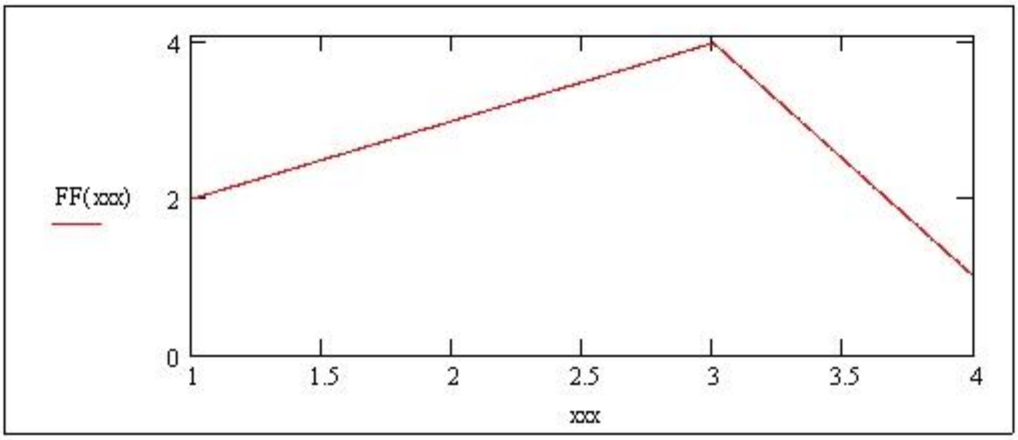
Пример: X(1, 2, 3, 4) n_i(2, 3, 4, 1)

$m := 4$

$$\boxed{X := \begin{pmatrix}1\\2\\3\\4\end{pmatrix}} \quad \boxed{ni := \begin{pmatrix}2\\3\\4\\1\end{pmatrix}} \quad n := \sum_{i=0}^{m-1} ni_i \quad \boxed{n = 10} \quad i := 0..m-1 \quad Wi_i := \frac{ni_i}{n} \quad Wi = \begin{pmatrix}0.2\\0.3\\0.4\\0.1\end{pmatrix} \quad max1 := max(ni)$$

$FF(xxx) := linterp(X, ni, xxx)$

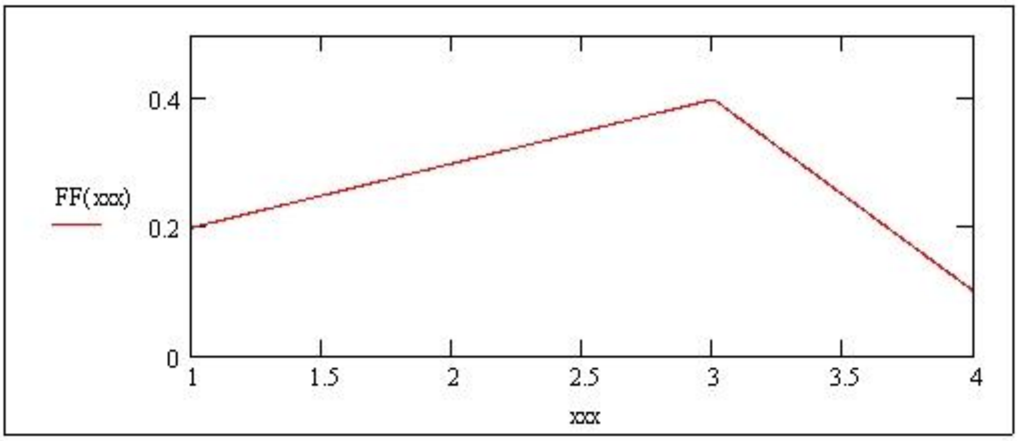
Полигон частот



max1 := max(Wi)

Полигон относительных частот

FF(xxx) := linterp(X, Wi, xxx)



Определение: Гистограммой называется ступенчатая фигура, образованная прямоугольниками с основанием h и высотой h_i/h или w_i/h.

Определение: Непрерывной называется случайная величина, которая принимает бесчисленное множество значений из заданного интервала.

Замечание: Гистограмма строится для непрерывной случайной величины на основе вариационного ряда.

Порядок построения гистограммы.

1. Выбирается границы интервала (a,b), на котором расположены все варианты данной выборки: $a \le \min\{x_i\}$ $b \ge \max\{x_i\}$ $\to$ $a \le x_1$ $b \ge x_n$
2. Выбирается количество участков m, на которое разбивается (a,b).
3. Вычисляется длинна каждого участка: h = (b - a) / m
4. На каждом участке подсчитывается количество вариант. лежащих на этом участке. Это количество будет определять частоту появления вариант на этом участке n_i .
5. Строится гистограмма.

Пример.

ГИСТОГРАММА

Объем выборки $n := 20$

$i := 0 .. (n - 1)$

Число участков $m := 6$

$$x := \begin{pmatrix} 10.2 \\ 5.2 \\ 0.8 \\ 7.8 \\ 9.5 \\ 10.4 \\ 8.4 \\ 0.7 \\ 5.7 \\ 3.5 \\ 11.5 \\ 6.5 \\ 1.2 \\ 4.2 \\ 0.6 \\ 11.5 \\ 7.5 \\ 11.8 \\ 4.8 \\ 5.6 \end{pmatrix}$$

1.Вариационный ряд

$xb := sort(x)$ $\qquad$ $xp(i) := xb_i$

2.Границы интервала

$a := xb_0$ $\qquad$ $a = 0.6$ $\qquad$ $a := 0$

$b := xb_{n-1}$ $\qquad$ $b = 11.8$ $\qquad$ $b := 12$

3.Длина учатков

$$h := \frac{b - a}{m}$$

xb =

	0
0	0.6
1	0.7
2	0.8
3	1.2
4	3.5
5	4.2
6	4.8
7	5.2
8	5.6
9	5.7
10	6.5
11	7.5
12	7.8
13	8.4
14	9.5
15	10.2
16	10.4
17	11.5
18	11.5
19	11.8

$$ni := \begin{pmatrix} 4 \\ 1 \\ 5 \\ 3 \\ 2 \\ 5 \end{pmatrix}$$

$k := 1..m \qquad xi_{k-1} := a + (k-1)\cdot h \qquad xpi(k) := xi_{k-1} \qquad xp_i := xp(i) \qquad np(k) := ni_{k-1}$

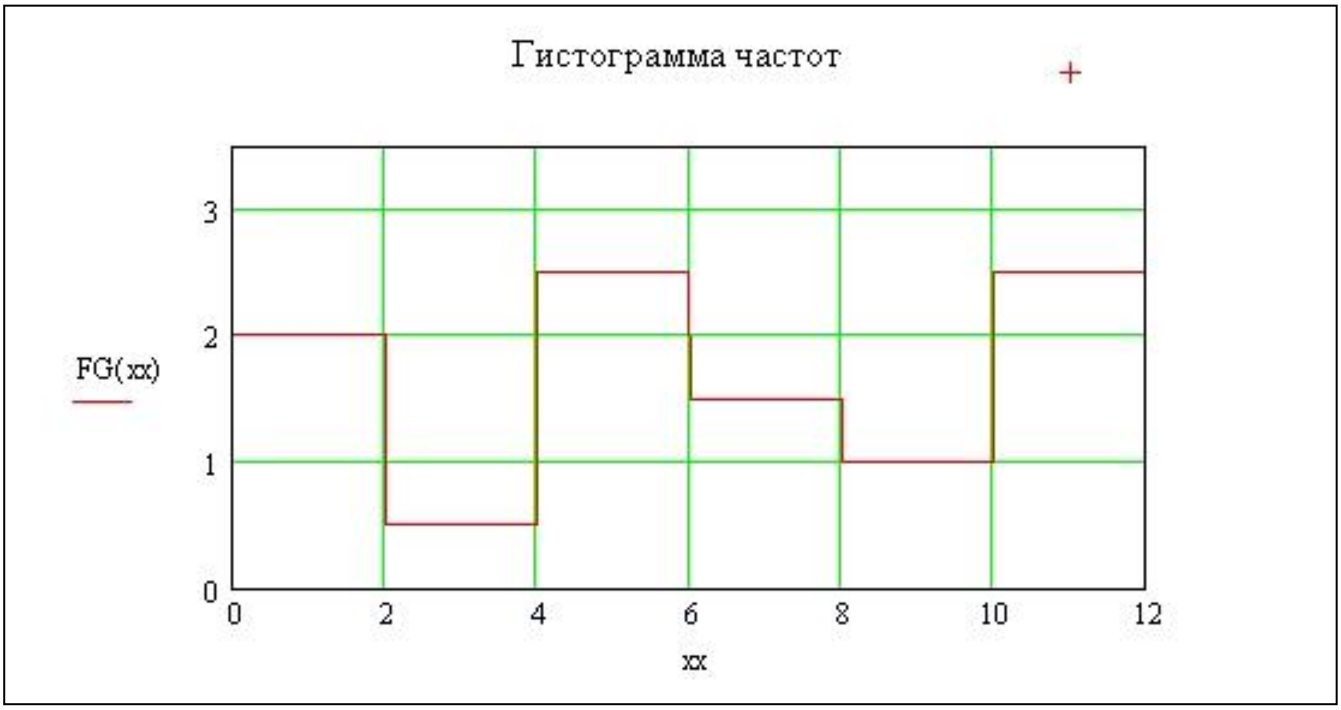
$xi1_{k-1} := a + k\cdot h \qquad xpi1(k) := xi1_{k-1} \qquad maxnp := max\left(\frac{ni}{h}\right) \qquad n1 := \sum_{ii=0}^{m-1} ni_{ii} \qquad n1 = 20$

$W(k) := \frac{np(k)}{n} \qquad Wh(k) := \frac{W(k)}{h} \qquad nph(k) := \frac{np(k)}{h} \qquad xisr_{k-1} := \frac{xi1_{k-1} + xi_{k-1}}{2}$

Таблица расчётов (m+1 - строк, 8 столбцов)

k =	xpi(k) =	xpi1(k) =	np(k) =	W(k) =	nph(k) =	Wh(k) =
1	0	2	4	0.2	2	0.1
2	2	4	1	0.05	0.5	0.025
3	4	6	5	0.25	2.5	0.125
4	6	8	3	0.15	1.5	0.075
5	8	10	2	0.1	1	0.05
6	10	12	5	0.25	2.5	0.125

$xisr = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix}$



Точечная оценка.

Определение: Точечной оценкой теоретического параметра θ называется величина θ*, которая выражается одним числом.

Свойства точечных оценок.

1. Несмещенность. Точечной оценка θ* является несмещенной оценкой теоретического параметра θ, если эти параметры связаны следующим отношением: M(θ*)=θ
2. Эффективность. θ* является эффективной оценкой θ, если она имеет минимальную дисперсию: D(θ)=0
3. Состоятельность. θ* является состоятельной оценкой θ, если $\underset{n\to\infty}{Lim}D(\theta^*)$

Характеристики выборки и виды точечных оценок.

1. Выборочная средняя: $\overline{X_B} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$

Замечание: Если частоты не заданы, то это значит, что все частоты равны единице.

Теорема: Выборочная средняя является несмещенной и состоятельной оценкой математического ожидания генеральной совокупности.

Доказательство:

$$M\left(\overline{X_B}\right) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}M\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M\left(x_i\right)$$

Каждый элемент выборки имеет одинаковый закон распределения с генеральными совокупностями. $M(x_1) = M(x_2) = \ldots = M$

$$\to M\left(\overline{X_B}\right) = \frac{1}{n}\sum_{i=1}^{n} M = \frac{1}{n}nM = M$$

Определение: Отклонением варианты от его среднего значения называется разность между ними.

Теорема: Сумма произведений отклонений на соответствующие частоты равна нулю.

Доказательство:

$$\sum_{i=1}^{m} n_i\left(x_i - \overline{x_в}\right) = \sum_{i=1}^{m} n_i x_i - \sum_{i=1}^{m} n_i\overline{x_B} = n\frac{1}{n}\sum_{i=1}^{m} n_i x_i - \overline{x_B}\sum_{i=1}^{m} n_i = n\overline{x_B} - \overline{x_B}n = 0$$

Вывод: отклонение не может являться точечной оценкой.

2. Выборочная дисперсия: $D_B = \frac{1}{n}\sum_{i=1}^{m} n_i\left[x_i - \overline{x_B}\right]^2$

3. Выборочное средне квадратическое отклонение: $\sigma = \sqrt{D_B}$

4. Выборочное среднее квадратов: $\left(\overline{x^2}\right)_B = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2$

Вычисление дисперсии

$$D_B = \frac{1}{n}\sum_{i=1}^{m} n_i\left(x_i - \overline{x_B}\right)^2 = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \frac{2}{n}\sum_{i=1}^{m} n_i x_i \overline{x_B} + \frac{1}{n}\sum_{i=1}^{m} n_i \overline{x_B^2}$$

С учётом 1 и 4 оценок:

$$D_B = \left(\overline{x^2}\right)_B - 2\overline{x_в}\,\frac{1}{n}\sum_{i=1}^{m} n_i x_i + \frac{\overline{x_в}^2}{n}\sum_{i=1}^{m} n_i = \left(\overline{x^2}\right)_B - 2\overline{x_B^2} + \overline{x_B^2} = \left(\overline{x^2}\right)_B - \left(\overline{x_B}\right)^2$$

5. Исправленное выборочное среднее: $S^2 = \frac{n}{n-1} D_B$

Замечание: Исправленное выборочное среднее является несмещенной оценкой дисперсии генеральной совокупности.

6. Исправленное выборочное средне квадратическое отклонение: $S = \sqrt{S^2}$

Замечание: Исправленное выборочное средне квадратическое отклонение является несмещенной оценкой средне квадратического отклонение генеральной совокупности.

Пример: X(1,3,5,7,9,11); n_i(4,1,5,3,2,5)

Точечная оценка мат.ожидания

Выборочная средняя $xb := \frac{1}{n}\cdot\sum_{ii=0}^{m-1} ni_{ii}\cdot xisr_{ii}$ $\quad xb = 6.3$

Выборочная средняя квадратов $xb2 := \frac{1}{n}\cdot\sum_{ii=0}^{m-1} ni_{ii}\cdot\left(xisr_{ii}\right)^2$ $\quad xb2 = 52.6$

Выборочная диспесрсия $Db := xb2 - (xb)^2$ $\quad Db = 12.91$

Точечная оценка дисперсии $S2 := \frac{n}{n-1}\cdot Db$ $\quad S2 = 13.589$

Точечная оценка среднего квадратического отклонения $S := \sqrt{S2}$ $\quad S = 3.686$

$$\sigma_В = \sqrt{D_В} = 3{,}593$$

Интервальные оценки.

Определение: Интервальной называется оценка, которая выражается двумя числами (концами интервала).
Определение: Доверительной вероятностью или надежностью называется число ω, которая показывает с какой вероятностью выполняется следующее неравенство: $|\theta - \theta^*| < \delta$, где θ - теоретический параметр, а θ* - эмпирический параметр: $\omega = P(|\theta - \theta^*| < \delta) = 0{,}95 \div 0{,}999$

Выразим: $|\theta - \theta^*| < \delta \rightarrow -\delta < \theta - \theta^* < \delta \rightarrow \theta^* - \delta < \theta < \theta^* + \delta$

Определение: Доверительным интервалом называется интервал $(\theta^*-\delta\ ;\ \theta^*+\delta)$, которой с заданной степенью надежности ω покрывает теоретический параметр θ.

Интервальная оценка для нормального распределения.

Интервальная оценка для математического ожидания.

$$M \in \left(\overline{x}_B - \delta, \overline{x}_B + \delta\right)$$

1 случай: Среднеквадратическое отклонение известно

Замечание: т.к. каждый элемент выборки имеет одно и тоже распределение, то из теории вероятности: $\sigma_B = \frac{\sigma}{\sqrt{n}}$, где n – объем выборки.

Для формулы надежности $\omega = P\left(\left|\theta - \theta^*\right| < \delta\right)$ теоретический параметр θ = M, эмпирический параметр θ* = $X_в$. С учётом этого получим: $\omega = P\left(\left|M - \overline{x_B}\right| \langle \delta\right)$

Для нормального распределения вероятность того, что отклонение не превысит заданной точности вычисляется через функцию Лапласа:

$$\omega = P\left(\left|M - \overline{x_B}\right| \langle \delta\right) = 2\Phi\left(\frac{\delta}{\sigma_B}\right)$$

Обозначим $t = \frac{\delta}{\sigma_B} \Rightarrow \delta = t\sigma_B$. По таблице приложения 2 по известной функции Лапласа F(t)= ω /2 можно найти аргумент t этой функции. Тогда интервальная оценка математического ожидания:

$$M \in \left(\overline{x}_B - t\sigma_B, \overline{x}_B + t\sigma_B\right) \text{ или } M \in \left(\overline{x}_B - \frac{t\sigma}{\sqrt{n}}; \overline{x_B} + \frac{t\sigma}{\sqrt{n}}\right)$$

2 случай: Среднеквадратическое отклонение неизвестно

Для этого случая было составлено дополнительное распределение со случайной величиной: $T = \dfrac{M - \overline{x}_B}{S/\sqrt{n}}$, которое носит название «распределение Стьюдента», где S – исправленное, среднеквадратичное отклонение.

Для этой случайной величины была вычислена надежность: $\omega = P\left(|T| \langle t_\upsilon\right)$, где t_υ – протабулирован в приложении 3 в зависимости от n и ω.

Тогда $|T|\langle t_\upsilon \to \left|\dfrac{M - \overline{x}_B}{S/\sqrt{n}}\right| \langle t_\upsilon \to -t_\upsilon < \dfrac{M - \overline{x}_B}{S/\sqrt{n}} < t_\upsilon \to \dfrac{St_\upsilon}{\sqrt{n}} < M - \overline{x}_B < \dfrac{St_\upsilon}{\sqrt{n}}$

$\to M \in \left(\overline{x}_B - \dfrac{t_\upsilon S}{\sqrt{n}}; \overline{x}_B + \dfrac{t_\upsilon S}{\sqrt{n}}\right)$

Интервальная оценка для среднеквадратического отклонения.

Для формулы надежности $\omega = P\left(\left|\theta - \theta^*\right| < \delta\right)$ теоретический параметр θ = σ, эмпирический параметр θ* = S. С учётом этого получим: $\sigma \in (S - \delta; S + \delta)$ или $\sigma \in \left(S\left(1 - \dfrac{\delta}{S}\right); S\left(1 - \dfrac{\delta}{S}\right)\right)$.

Обозначим: $q = \dfrac{\delta}{S}$. Тогда окончательно: $\sigma \in \left(S(1-q); S(1+q)\right)$.

Параметр q протабулирован в приложении 4 в зависимости от n и ω.
Замечание: По своей сути среднеквадратического отклонение величина неотрицательная, поэтому при получении отрицательного значения её нижней границе, она задаётся равной нулю.

Пример:

$\sigma := 1$ Надежность $w := 0.999$

$x := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ $ni := \begin{pmatrix} 2 \\ 8 \\ 6 \\ 4 \end{pmatrix}$ Число участков $m := 4$ $i := 0..(m-1)$ $n := \sum_{ii=0}^{m-1} ni_{ii}$ Объем выборки $\boxed{n = 20}$

Выборочная средняя $xb := \frac{1}{n} \cdot \sum_{ii=0}^{m-1} ni_{ii} \cdot x_{ii}$ $\boxed{xb = 2.6}$

Выборочная средняя квадратов $xb2 := \frac{1}{n} \cdot \sum_{ii=0}^{m-1} ni_{ii} \cdot (x_{ii})^2$ $\boxed{xb2 = 7.6}$

Выборочная дисперсия $Db := xb2 - (xb)^2$ $\boxed{Db = 0.84}$

Точечная оценка дисперсии $S2 := \frac{n}{n-1} \cdot Db$ $\boxed{S2 = 0.884}$

Точечная оценка среднеквадратического отклонения $S := \sqrt{S2}$ $\boxed{S = 0.94}$

1. мат.ожидание

$F := \frac{w}{2}$ $F = 0.4995$ По приложению 2 $\boxed{t := 3.4}$ $\Delta := \frac{t \cdot \sigma}{\sqrt{n}}$ $\Delta = 0.76$ $M1 := xb - \Delta$ $M2 := xb + \Delta$

$M \in ($ $M1 = 1.84$; $M2 = 3.36$ $)$

2. мат.ожидание

По приложению 3 $\boxed{t\nu := 3.883}$ $\Delta := \frac{t\nu \cdot S}{\sqrt{n}}$ $\Delta = 0.816$ $M1 := xb - \Delta$ $M2 := xb + \Delta$

$M \in ($ $M1 = 1.784$; $M2 = 3.416$ $)$

3. Среднеквадратическое отклонение

По приложению 4 $\boxed{q := 0.88}$ $\sigma1 := S \cdot (1 - q)$ $\sigma2 := S \cdot (1 + q)$

$\sigma \in ($ $\sigma1 = 0.113$; $\sigma2 = 1.768$ $)$

4. Дисперсия $D \in ($ $\sigma1^2 = 0.013$; $\sigma2^2 = 3.125$ $)$

Выравнивание эмпирических распределений.

Определение: Эмпирическим распределением называется распределение наблюдаемых n_i.

Определение: Теоретическим или выравнивающим распределением называется распределение частот $n_i^{'}$, вычисленных в предположении, что данная случайная величина распределена по одному из известных типовых распределений.

Теоретическая частота вычисляется по формуле: $n_i^{'} = nP_i$, где P_i - вероятность, вычисленная для соответствующего распределения.

Определение теоретических частот для Пуассоновского распределения.

Для этого распределения вероятность вычисляется по формуле: $P_n(k) = \frac{(nP)^k}{k!} e^{-np}$. По свойствам математического ожидания $M = nP$. Заменим теоретический параметр M на точную оценку $\overline{x}_B$.

Тогда формула теоретической частоты для Пуассоновского распределения будет иметь вид: $n_i^{'} = n\frac{\left(\overline{x}_B\right)^k}{k!} e^{-\overline{x}_B}$

Замечание: т.к. частота может выражаться только данными числами, то $n_i^{'}$ округляется до целого после вычислений.

Замечание: т.к. сумма частот должна быть равна объему выборки, то для достижения этого некоторые теоретические частоты подправляются на единицу.

Пример:

РАСЧЁТ ТЕОРЕТИЧЕСКИХ ЧАСТОТ – Пуассоновское распределение

$X := \begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix}$ $ni := \begin{pmatrix} 11 \\ 6 \\ 2 \\ 1 \end{pmatrix}$ $nk := 4$ $n := \sum_{i=0}^{nk-1} ni_i$ Объём выборки $\boxed{n = 20}$ $i := 0..(nk-1)$

Выборочное среднее $Xb := \frac{1}{n} \cdot \sum_{i=0}^{nk-1} ni_i X_i$ $\boxed{Xb = 2.3}$

$niT_i := \frac{n \cdot Xb^{i+1}}{(i+1)!} \cdot e^{-Xb}$ $niTO_i := \mathrm{round}(niT_i)$ $nT := \sum_{i=0}^{nk-1} niTO_i$ $\boxed{nT = 16}$ $k := n - nT$ $\boxed{k = 4}$

$p := \mathrm{if}(k > 0, 1, \mathrm{if}(k < 0, -1, 0))$ $niTOP_i := niTO_i$

$kk := \mathrm{if}(k = 0, 1, |k|)$ $i := 0..(kk-1)$ $niTOP_i := niTO_i + p$

Теоретическая частота	Теоретическая частота округленная	Теоретическая частота округленная и подправленная
$niT = \begin{pmatrix} 4.612 \\ 5.304 \\ 4.066 \\ 2.338 \end{pmatrix}$	$niTO = \begin{pmatrix} 5 \\ 5 \\ 4 \\ 2 \end{pmatrix}$	$niTOP = \begin{pmatrix} 6 \\ 6 \\ 5 \\ 3 \end{pmatrix}$

Эмпирическая функции (сплошная линия)

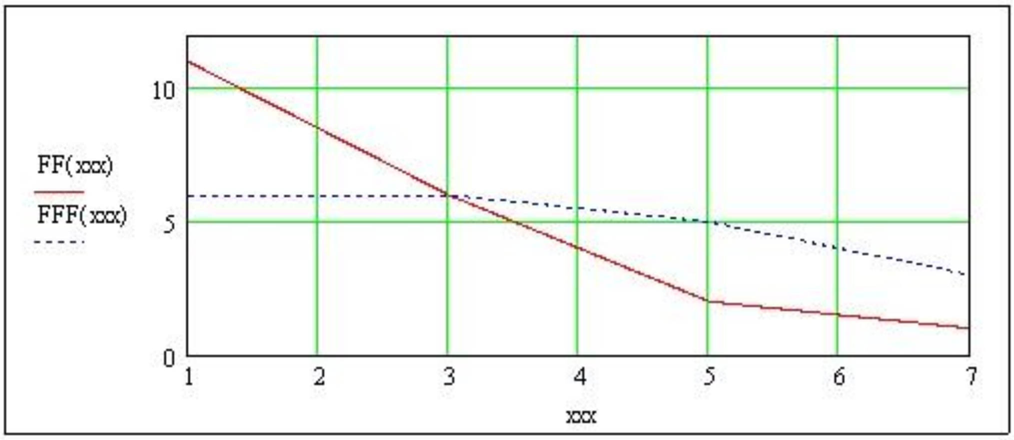
$FF(xxx) := \mathrm{linterp}(X, ni, xxx)$

$maxnp := \max(ni)$

График эмпирических и теоретических частот

Теоретическая функции (пунктирная линия)

$FFF(xxx) := \mathrm{linterp}(X, niTOP, xxx)$



Определение теоретических частот для нормального распределения.

Для нормального распределения известна функция плотности распределения вероятностей: $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-M)^2}{2\sigma^2}}$. Обозначим: $\varphi(x)=\frac{1}{\sqrt{2\pi}}-e^{-\frac{x^2}{2}}$, $u=\frac{x-M}{\sigma}\to\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\to f(x)=\frac{1}{\sigma}\varphi(u)$. Теоретический параметр М заменим на точную оценку $\overline{x_B}$, а σ на $\sigma_в$ → $u=\frac{x-\overline{x}_B}{\sigma_B}; f(x)=\frac{1}{\sigma_B}\varphi(u)$. Разобьем отрезок [*а,в*] на *m* частей и в середине каждой части [x_i; x_{i+1}] найдем значение x_i*. Длина каждого участка $h=\frac{b-a}{m}$. Для каждой части будут известны:

$$u_i=\frac{x_i^*-\overline{x}_B}{\sigma_B};\ f(x_i^*)=\frac{1}{\sigma_B}\varphi(u_i).$$

Для нормального распределения можно вычислить вероятность попадания случайной величины на каждый участок:

$$P_i=P(x_i<x<x_{i+1})=\int_{x_i}^{x_{i+1}}f(x_i^*)dx=f(x_i^*)\int_{x_i}^{x_{i+1}}dx=f(x_i^*)x\Big|_{x_i}^{x_{i+1}}$$

$$=f(x_i^*)(x_{i+1}-x_i)=\frac{1}{\sigma_B}\varphi(u_i)h$$

Окончательно получаем формулу вычисления теоретической частоты:

$$n_i'=nP_i=\frac{nh}{\sigma_B}\varphi(u_i)$$

Пример:

Теоретические частоты для нормального распределения

$$x := \begin{pmatrix} -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ 1 \end{pmatrix} \quad ni := \begin{pmatrix} 2 \\ 5 \\ 7 \\ 8 \\ 4 \\ 4 \end{pmatrix}$$

$$\phi(z) := \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} \quad h := 1 \quad m := 6$$

$$n := \sum_{i=0}^{m-1} ni_i \quad xb := \frac{1}{n} \cdot \sum_{i=0}^{m-1} x_i \cdot ni_i \quad xb2 := \frac{1}{n} \cdot \sum_{i=0}^{m-1} ni_i (x_i)^2 \quad Db := xb2 - (xb)^2$$

$$n = 30 \quad xb = -1.367 \quad xb2 = 3.9 \quad Db = 2.032$$

$$\sigma b := \sqrt{Db} \quad \sigma b = 1.426$$

$$i := 0..m-1$$

$$c1_i := x_i - xb \quad u_i := \frac{c1_i}{\sigma b} \quad c2_i := \phi(u_i) \quad nil_i := \frac{n \cdot h}{\sigma b} \cdot c2_i$$

$$nilo_i := \mathrm{round}(nil_i) \quad nilp_i := nilo_i \quad nilo_i := \mathrm{if}(nilo_i = 0, 1, nilo_i) \quad nk := \sum_{i=0}^{m-1} nilo_i \quad nk = 29$$

$$k := n - nk \quad k = 1 \quad p := \mathrm{if}(k > 0, 1, \mathrm{if}(k < 0, -1, 0)) \quad kk := \mathrm{if}(k = 0, 1, |k|) \quad j := 0..kk-1 \quad nilp_j := nilo_j + p$$

$$\begin{array}{cccccccc} X^* & n_i & x^* - xb & u_i & \varphi(u_i) & n_I & \text{округленные частоты} & \text{подправленные} \\ x = \begin{pmatrix} -4 \\ -3 \\ -2 \\ -1 \\ 0 \\ 1 \end{pmatrix} & ni = \begin{pmatrix} 2 \\ 5 \\ 7 \\ 8 \\ 4 \\ 4 \end{pmatrix} & c1 = \begin{pmatrix} -2.633 \\ -1.633 \\ -0.633 \\ 0.367 \\ 1.367 \\ 2.367 \end{pmatrix} & u = \begin{pmatrix} -1.847 \\ -1.146 \\ -0.444 \\ 0.257 \\ 0.959 \\ 1.66 \end{pmatrix} & c2 = \begin{pmatrix} 0.072 \\ 0.207 \\ 0.361 \\ 0.386 \\ 0.252 \\ 0.101 \end{pmatrix} & nil = \begin{pmatrix} 1.524 \\ 4.355 \\ 7.607 \\ 8.122 \\ 5.302 \\ 2.116 \end{pmatrix} & nilo = \begin{pmatrix} 2 \\ 4 \\ 8 \\ 8 \\ 5 \\ 2 \end{pmatrix} & nilp = \begin{pmatrix} 3 \\ 4 \\ 8 \\ 8 \\ 5 \\ 2 \end{pmatrix} \end{array}$$

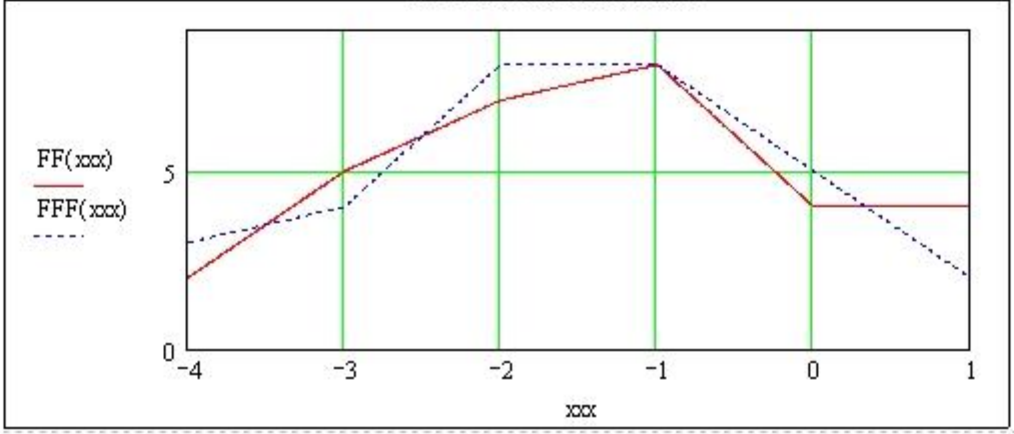
Эмпирическая функции (сплошная линия)

Теоретическая функции (пунктирная линия)

$$maxnp := \max(ni)$$

$$FF(xxx) := \mathrm{linterp}(x, ni, xxx) \qquad FFF(xxx) := \mathrm{linterp}(x, nilp, xxx)$$

График эмпирических и теоретических частот



Статистическая проверка гипотез. Основные понятия.

Определение: Статистической гипотезой называется предположение о виде неизвестного закона распределения или о значениях параметров известного закона распределения.

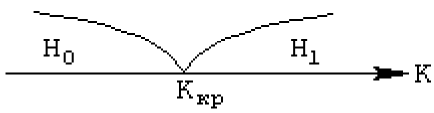
Виды статистических гипотез.

1. Нулевая или основная – гипотеза, которая выдвигается (H_0).
2 .Альтернативная – противоположная (H_0) гипотеза (H_1).
3. Простая – состоит из одного предположения (о виде какого-то распределения).
4. Сложная – состоит из двух и более предположений (о значениях характеристик нормального распределения).

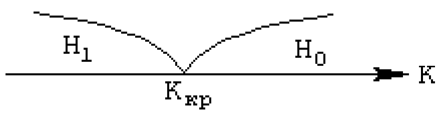
Определение: Статистическим критерием называется случайная величина *К*, при одних значениях которой гипотеза H_0 принимается, а при других отвергается.
Определение: Множество значений величины *К*, при которой гипотеза H_0 принимается называется областью принятия решения .

Определение: Множество значений, при котором гипотеза H_0 отвергается называется критической областью.
Определение: Точка, отделяющая область принятия решения от критической области называется критической точкой.
Виды критических областей

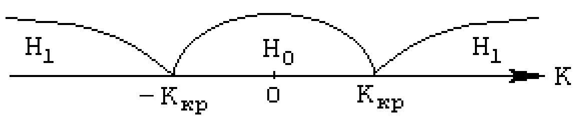
1.Правосторонняя



2.Левостороняя

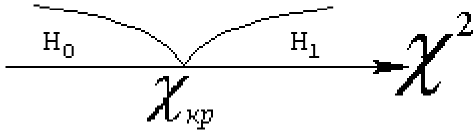


3.Двухстороняя



Критерий согласия Пирсона для нормального распределения

Определение: Критерием согласия называется статистический критерий, который определяет степень соответствия эмпирических данных предполагаемому закону распределения. Для нормального распределения этот критерий согласия является правосторонним:



и вычисляется: $\chi^2 = \sum_{i=1}^{m} \frac{\left(\bar{n}'_i - n_i\right)^2}{\bar{n}'_i}$

Если $\chi^2 \le \chi^2_{кр}$, то гипотеза о нормальном распределении принимается, в противном случае отвергается.

Критическое значение критерия $\chi_{кр}$ определяется в зависимости от уровня значимости $\alpha \le 1-\omega$ и степени свободы: $K = m - 1 - R$, где : m - число участков, на которые разбиваются отрезок $[a,b]$.

R - число характеристик предполагаемого закон распределения. Для нормального распределения таких характеристик две: математическое ожидание и среднеквадратичное отклонение: $R = 2 \to K = m - 3$

Пример:

Теоретические частоты и критерий Пирсона для нормального распределения

$$x := \begin{pmatrix} 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{pmatrix} \quad \boxed{ni := \begin{pmatrix} 1 \\ 5 \\ 9 \\ 9 \\ 5 \\ 1 \end{pmatrix}}$$

$$\phi(z) := \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} \quad h := 1 \quad m := 6 \quad \chi 2kr := 7.8 \quad \alpha := 0.05 \quad k := m - 3 \quad k = 3$$

$$n := \sum_{i=0}^{m-1} ni_i \quad xb := \frac{1}{n} \cdot \sum_{i=0}^{m-1} x_i \cdot ni_i \quad xb2 := \frac{1}{n} \cdot \sum_{i=0}^{m-1} ni_i (x_i)^2 \quad Db := xb2 - (xb)^2$$

$$n = 30 \quad xb = 12.5 \quad xb2 = 157.567 \quad Db = 1.317$$

$$\sigma b := \sqrt{Db} \quad \sigma b = 1.147$$

$$i := 0..m-1$$

$$c1_i := x_i - xb \quad u_i := \frac{c1_i}{\sigma b} \quad c2_i := \phi(u_i) \quad ni1_i := \frac{n \cdot h}{\sigma b} \cdot c2_i \quad ni1o_i := \mathrm{round}(ni1_i) \quad ni1p_i := ni1o_i \quad nk := \sum_{i=0}^{m-1} ni1o_i$$

$$nk = 28 \quad k := n - nk \quad k = 2 \quad p := \mathrm{if}(k > 0, 1, \mathrm{if}(k < 0, -1, 0)) \quad kk := \mathrm{if}(k = 0, 1, |k|) \quad j := 0..kk - 1 \quad ni1p_j := ni1o_j + p$$

X^*	n_i	$x^* -\ xb$	u_i	$\varphi(u_i)$	n'_I	округленные частоты	подправленные

$$x = \begin{pmatrix} 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{pmatrix} \quad ni = \begin{pmatrix} 1 \\ 5 \\ 9 \\ 9 \\ 5 \\ 1 \end{pmatrix} \quad c1 = \begin{pmatrix} -2.5 \\ -1.5 \\ -0.5 \\ 0.5 \\ 1.5 \\ 2.5 \end{pmatrix} \quad u = \begin{pmatrix} -2.179 \\ -1.307 \\ -0.436 \\ 0.436 \\ 1.307 \\ 2.179 \end{pmatrix} \quad c2 = \begin{pmatrix} 0.037 \\ 0.17 \\ 0.363 \\ 0.363 \\ 0.17 \\ 0.037 \end{pmatrix} \quad ni1 = \begin{pmatrix} 0.972 \\ 4.438 \\ 9.486 \\ 9.486 \\ 4.438 \\ 0.972 \end{pmatrix} \quad ni1o = \begin{pmatrix} 1 \\ 4 \\ 9 \\ 9 \\ 4 \\ 1 \end{pmatrix} \quad ni1p = \begin{pmatrix} 2 \\ 5 \\ 9 \\ 9 \\ 4 \\ 1 \end{pmatrix}$$

$$c3_i := ni1p_i - ni_i \quad c4_i := (c3_i)^2 \quad c5_i := \frac{c4_i}{ni1p_i} \quad \chi 2 := \sum_{k=0}^{m-1} c5_k$$

$$\chi 2 = 0.75$$

$$ni = \begin{pmatrix} 1 \\ 5 \\ 9 \\ 9 \\ 5 \\ 1 \end{pmatrix} \quad ni1p = \begin{pmatrix} 2 \\ 5 \\ 9 \\ 9 \\ 4 \\ 1 \end{pmatrix} \quad c3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} \quad c4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad c5 = \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0 \\ 0.25 \\ 0 \end{pmatrix}$$

$$o := \mathrm{if}(\chi 2 < \chi 2kr, \text{"YES"}, \text{"NO"})$$

$$o = \text{"YES"}$$

n_i	$\overline{n'_i}$	$\overline{n'_i} - n_i$	$\left(\overline{n'_i} - n_i\right)^2$	$\dfrac{\left(\overline{n'_i} - n_i\right)^2}{\overline{n'_i}}$

Функциональные статистические корреляционные зависимости. Понятие регрессии.

Определение: Функциональной называется зависимость между двумя случайными величинами x и y ,при которых каждому значению x соответствует одно значение y.

Определение: Статистической называется зависимость, при которой изменение величины х приводит к изменению закона распределения случайной величины у.

Определение: Корреляционной называется зависимость, при которой изменение случайной величины x приводит к изменению среднего значения случайной величины y.

Понятие регрессии.

Регрессией случайной величины X на случайную величину Y называется форма корреляционной зависимости между этими случайными величинами.
$\overline{y} = g(x)$ - регрессия X на Y; $\overline{x} = q(y)$ - регрессия Y на X.

Линейная регрессия и её основное свойство.

Линейная регрессия в терминах теории вероятностей имеет вид:

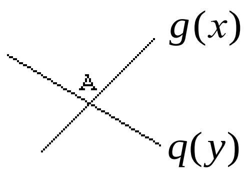
$$\overline{y} = g(x) = m_y + r\frac{\sigma_y}{\sigma_x}\left(x + m_x\right)$$

где $m_x = M(X); m_y = M(Y)$ – математические ожидания, σ_x, σ_y - средне-квадратичное отклонение для соответствующих случайных величин X,Y,
r - коэффициент корреляции $r = M((x - m_x, y - m_y))$.

Замечание: в дальнейшем черточки на переменными, обозначающие средние значение, для простоты записи опускаем. Тогда:

Регрессия X на Y имеет вид: $y = m_y + r\frac{\sigma_y}{\sigma_x}\left(x - m_x\right)$ (1)

Регрессия Y на X вычисляется аналогично $x = m_x + r\frac{\sigma_x}{\sigma_y}\left(y - m_y\right)$ (2)



Решаем уравнения (1,2) для нахождения координат т.А пересечения линий регрессии:

$$y = m_y + r\frac{\sigma_y}{\sigma_x}\left(m_x + r\frac{\sigma_x}{\sigma_y}\left(x - m_y\right) - m_x\right) \quad \to y = m_y + r^2 y - r^2 m_y \qquad \to y(1-r^2) = m_y(1-r^2)$$

$\to y = m_y$. Поставляем найденной значение в (2) $\to x = m_x \to A(m_x;\, m_y)$

Вывод: для точки A можно составить функцию между математическими ожиданиями двух случайных величин.

Определим при каком случае уравнения (1) и (2) совпадают. Для этого зададим: $r = \pm 1$

$$y = m_y \pm \frac{\sigma_y}{\sigma_x}\left(x - m_x\right) \ (3) \quad x = m_x \pm r\frac{\sigma_x}{\sigma_y}\left(y - m_y\right) \ (4)$$

Выразим y из (4) $\pm\frac{\sigma_x}{\sigma_y}\left(y - m_y\right) = x - m_x \to \left(y - m_y\right) = \pm\frac{\sigma_y}{\sigma_x}\left(x - m_x\right)$

$$y = m_y \pm \frac{\sigma_y}{\sigma_x}\left(x - m_x\right) \equiv (3)$$

Вывод:

Основное свойство регрессии:

Если коэффициент корреляции $r = \pm 1$, то между двумя случайными величинами существует функциональная зависимость.

Выборочное уравнение линейной регрессии.

Дано: две выборки одинакового объема n для случайных величин X и Y

X	x_1	x_2	…	x_n
X	y_1	y_2	…	y_n

Уравнение линейной регрессии имеет вид: $g(x) = a_0 + a_1x$, где a_0, a_1 - неизвестные коэффициенты.

Для нахождения a_0,a_1 используется метод наименьших квадратов, согласно которому определяется сумма квадратов отклонений искомой функции $g(x)$ и значение y_i (I = 1,2,3): $S = \sum_{i=1}^{n}\left[g(x_i) - y_i\right]^2 \to S(a_0, a_1) = \sum_{i=1}^{n}\left[a_0 + a_1x_i - y_i\right]^2$

Отклонения функции $g(x)$ от экспериментальных данных должны быть минимальными → необходимо найти: $min\{S(a_0,a_1)\}$

Для этого необходимо приравнять к нулю частные производные от функции S по a_0 и a_1:

$$\to \frac{\partial S}{\partial a_0} = 2\sum_{i=1}^{n}\left[a_0 + a_1x_i - y_i\right] = 0 \to \sum_{i=1}^{n}\left[a_0\right] + \sum_{i=1}^{n}\left[a_1x_i\right] - \sum_{i=1}^{n}\left[y_i\right] = 0 \to na_0 + a_1\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}y_i \quad (1)$$

$$\to \frac{\partial S}{\partial a_1} = 2\sum_{i=1}^{n}\left[a_0 + a_1x_i - y_i\right]x_i = 0 \to \sum_{i=1}^{n}\left[a_0\right]x_i + \sum_{i=1}^{n}\left[a_1x_i\right]x_i - \sum_{i=1}^{n}\left[y_i\right]x_i = 0$$

$$\to a_0\sum_{i=1}^{n}x_i + a_1\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}x_iy_i \quad (2)$$

Решаем систему уравнений (1) и (2) методом Крамара:

$$D = \begin{vmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{vmatrix} = n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2; \quad D_{a_0} = \begin{vmatrix} \sum_{i=1}^{n}y_i & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_iy_i & \sum_{i=1}^{n}x_i^2 \end{vmatrix} = \sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i;$$

$$D_{a_1} = \begin{vmatrix} \mathrm{n} & \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_iy_i \end{vmatrix} = n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i$$

$$a_0 = \frac{\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i}{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}, \quad a_1 = \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}$$

Подставляя найденные коэффициенты a_0 и a_1 в функцию $g(x)=a_0+a_1x$, получим выборочное уравнение линейной регрессии.
Для определения точности представления зависимости между случайными величинами виде линейной функции необходимо вычислить относительную погрешность $\Delta_i = \frac{g(x_i)-y_i}{y_i}100\%$.

Между двумя случайными величинами существует функциональная зависимость, если все погрешности $\Delta_i=0$, если $\Delta_i \neq 0$, то линейная корреляционная зависимость, если $\Delta_i > 5\%$ - нелинейная корреляционная зависимость.

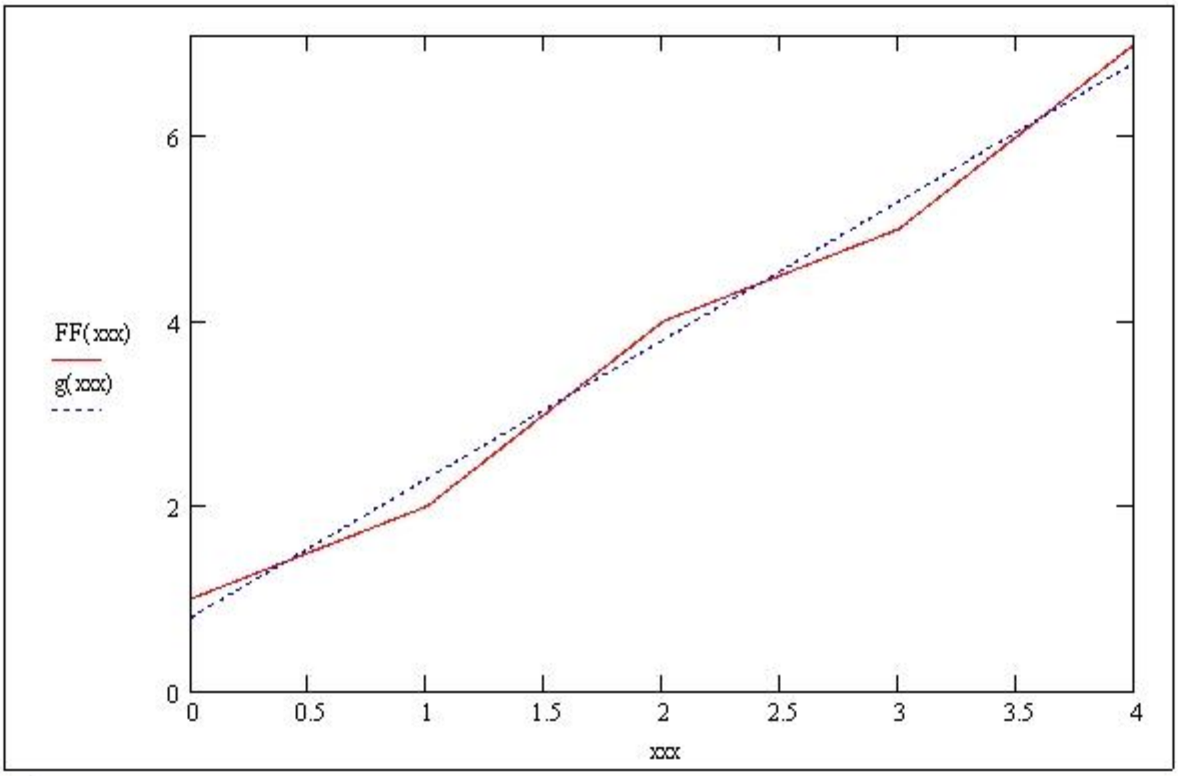
Пример:

$X := \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ $Y := \begin{pmatrix} 1 \\ 2 \\ 4 \\ 5 \\ 7 \end{pmatrix}$

$\mathrm{max1} := \max(Y)$

$FF(xxx) := \mathrm{linterp}(X, Y, xxx)$ $g(xxx) := 0.8 + 1.5 \cdot xxx$

Эмпирическая (сплошная) и теоретическая
(пунктирная) зависимости
для линейной регрессии



Корреляционная таблица и выборочный коэффициент корреляции.

Условные обозначения:

n_{x_i} - частоты появления случайной величины x_i

n_{y_i} - частоты появления случайной величины y_i

$n_{x_i y_i}$ - частоты совместного появления x_i, y_i

Корреляционная таблица.
(таблица вариант со своими частотами)

X_i	x_1	x_2	…	x_m	n_{y_i}
y_1	$n_{x_1y_1}$	$n_{x_2y_1}$	…	$n_{x_my_1}$	n_{y_1}
y_2	$n_{x_1y_2}$	$n_{x_2y_2}$	…	$n_{x_my_2}$	n_{y_2}
….	….	….	….	….	….
y_k	$n_{x_1y_k}$	$n_{x_2y_k}$	….	$n_{x_my_k}$	n_{y_k}
n_{x_j}	n_{x_1}	n_{x_2}	….	n_{x_m}	n

Рассмотрим систему уравнений (1) и (2), полученную для выборочного уравнения регрессии: $na_0 + a_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ (1) $\quad a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$ (2)

При «отсутствии» частот (когда все они равны единице) вычислим выборочные средние для каждой из случайных величин:

$$\overline{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y}_B = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \left(\overline{x^2}\right)_B = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad S_{xy} = \sum_{i=1}^{n} x_i y_j$$

Заменим на них соответствующие суммы в уравнениях (1), (2):

$$\sum_{i=1}^{n} x_i = n\overline{x}_B, \quad \sum_{i=1}^{n} y_i = n\overline{y}_B, \quad \sum_{i=1}^{n} x_i^2 = n\left(\overline{x^2}\right)_B, \quad \sum_{i=1}^{n} x_i y_j = S_{xy}$$

В результате получим уравнения (3), (4), которые решаем также методом Крамара:

$$a_0 + a_1\overline{x}_B = \overline{y}_B \ (3), \quad a_0 n\overline{x}_B + a_1 n\left(\overline{x^2}\right)_B = S_{xy} \ (4)$$

$$D = \begin{vmatrix} 1 & \overline{x}_B \\ n\overline{x}_B & n\left(\overline{x^2}\right)_B \end{vmatrix} = n\left[\left(\overline{x^2}\right)_B - \left(\overline{x}_B\right)^2\right] = nD_{Bx} = n\left(\sigma_{Bx}\right)^2; D_{a_1} = \begin{vmatrix} 1 & \overline{y}_B \\ n\,\overline{x}_B & S_{xy} \end{vmatrix} = S_{xy} - n\,\overline{x}_B\,\overline{y}_B$$

$a_1 = \dfrac{S_{xy} - n\,\overline{x}_B\,\overline{y}_B}{n\left(\sigma_{Bx}\right)^2}$. Умножаем обе части полученной формулы на $\dfrac{\sigma_{Bx}}{\sigma_{By}}$:

$$\frac{\sigma_{Bx}}{\sigma_{By}} a_1 = \frac{S_{xy} - n\,\overline{x}_B\,\overline{y}_B}{n\left(\sigma_{Bx}\right)^2}\frac{\sigma_{Bx}}{\sigma_{By}} \rightarrow \frac{\sigma_{Bx}}{\sigma_{By}} a_1 = \frac{S_{xy} - n\,\overline{x}_B\,\overline{y}_B}{n\sigma_{Bx}\sigma_{By}}$$

Обозначим полученную правую часть как выборочный коэффициент корреляции: $r_B = \dfrac{S_{XY} - n\overline{x}_B\,\overline{y}_B}{n\sigma_{BX}\sigma_{BY}}$, тогда: $\dfrac{\sigma_{Bx}}{\sigma_{By}}a_1 = r_B \rightarrow a_1 = r_B\dfrac{\sigma_{By}}{\sigma_{Bx}}$.

Подставляем найденный коэффициент в (3): $a_0 + r_B\dfrac{\sigma_{By}}{\sigma_{Bx}}\overline{x}_B = \overline{y}_B \rightarrow a_0 = \overline{y}_B - r_B\dfrac{\sigma_{By}}{\sigma_{Bx}}\overline{x}_B$

Подставляем найденные коэффициенты в уравнение линейной регрессии:

$g(x) = a_0 + a_1 x \rightarrow g(x) = \overline{y}_B - r_B\dfrac{\sigma_{By}}{\sigma_{Bx}}\overline{x}_B + r_B\dfrac{\sigma_{By}}{\sigma_{Bx}}x$

Окончательно имеем выборочное уравнение линейной регрессии (проверка):

$$g\left(x\right) = \overline{y}_B + r_B\frac{\sigma_{By}}{\sigma_{Bx}}\left(x - \overline{x}_B\right).$$

Сравниваем это уравнения с уравнением линейной регрессии в терминах теории вероятности: $y = m_y + r\dfrac{\sigma_y}{\sigma_x}\left(x - m_x\right)$. Структура этих уравнений одинаковая, значит, между двумя случайными величинами существует функциональная зависимость, если $r_B = 1$, в противном случае корреляционная зависимость.

При наличии частот:

$\overline{x}_B = \dfrac{1}{n}\sum\limits_{i=1}^{m} n_{x_i}x_i$, $\quad \overline{y}_B = \dfrac{1}{n}\sum\limits_{j=1}^{k} n_{y_j}y_j$, $\quad \left(\overline{x^2}\right)_B = \dfrac{1}{n}\sum\limits_{i=1}^{m} n_{x_i}x_i^2$, $\quad S_{xy} = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{k} n_{x_i y_j}x_i y_j$

$\mathrm{D_{Bx}} = \left(\mathrm{x}^2\right)_\mathrm{B} - \left(\mathrm{x_B}\right)^2; \mathrm{D_{By}} = \left(\mathrm{y}^2\right)_\mathrm{B} - \left(\mathrm{y_B}\right)^2; \sigma_\mathrm{Bx} = \sqrt{\mathrm{D_{Bx}}}; \sigma_\mathrm{By} = \sqrt{\mathrm{D_{By}}}$.

Понятие о корреляционном отношении и его свойства.

$$\eta = \frac{\sigma_{yx}}{\sigma_{By}},$$

где $\sigma_{yx} = \sqrt{D_{yx}}$, $D_{yx} = \frac{1}{n}\sum_{i=1}^{m} n_{x_i}\left(g(x_i) - \overline{y}_B\right)^2$,

$g(x_i)$ - вычисляется из выборочного уравнения линейной регрессии: $g\left(x\right) = \overline{y}_B + r_B \frac{\sigma_{By}}{\sigma_{Bx}}\left(x - \overline{x}_B\right)$ подстановкой в него каждого значения случайной величины x_i

Свойства корреляционного отношения

1. 0≤η≤1
2. Между двумя случайными величинами существует функциональная зависимость, если η=1,
3. между случайными величинами не существует никакой зависимости, если η=0,
4. между случайными величинами существует линейная зависимость, если η= |rв |.

Пример:

X	0	1	2	3	4
Y	1	2	4	5	7
$g(x)$	*0.8*	*2.3*	*3.8*	*5.3*	*6.8*
$\Delta\%$	*20*	*15*	*5*	*6*	*3*

i	x_i	$(x_i)^2$	y_i	$x_i\,y_i$
1	0	0	1	0
2	1	1	2	2
3	2	4	4	8
4	3	9	5	15
5	4	16	7	28
$\sum$	10	30	19	53

$$x := \begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix} \quad y := \begin{pmatrix}1\\2\\4\\5\\7\end{pmatrix} \quad nx := \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} \quad ny := \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}$$

$$m := 5 \qquad k := 5 \qquad i := 0..(m-1) \qquad j := 0..(k-1)$$

$$n := \sum_{j=0}^{m-1} nx_j \qquad n = 5 \qquad nxy_{i,j} := 0 \qquad nxy_{i,i} := 1$$

$$nxy = \begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix} \quad sx := \sum_{j=0}^{k-1} x_j \quad sy := \sum_{j=0}^{m-1} y_j \quad sx2 := \sum_{j=0}^{k-1} (x_j)^2 \quad sxy := \sum_{j=0}^{m-1}\sum_{i=0}^{k-1} nxy_{i,j}\cdot x_i \cdot y_j$$

$$\boxed{sx = 10} \quad \boxed{sy = 19} \quad \boxed{sx2 = 30} \quad \boxed{sxy = 53}$$

$$D := n\cdot sx2 - (sx)^2 \quad Da0 := sy\cdot sx2 - sx\cdot sxy \quad Da1 := n\cdot sxy - sx\cdot sy \quad a0 := \frac{Da0}{D} \quad a1 := \frac{Da1}{D}$$

$$\boxed{D = 50} \quad \boxed{Da0 = 40} \quad \boxed{Da1 = 75} \quad \boxed{a0 = 0.8} \quad \boxed{a1 = 1.5}$$

$g(xx) := a0 + a1\cdot xx$ Выборочное уравнение линейной регрессии $\boxed{g(xx) \rightarrow \frac{4}{5} + \frac{3}{2}\cdot xx}$

$g1 := g(x_0)$ $g1 = 0.8$ $g2 := g(x_1)$ $g2 = 2.3$

$g3 := g(x_2)$ $g3 = 3.8$ $g4 := g(x_3)$ $g4 = 5.3$

$g5 := g(x_4)$ $g5 = 6.8$

$\Delta_i := \text{if}\left[y_i > 0, \frac{(g(x_i) - y_i)}{y_i}, 0\right]$

$\Delta_i := |\Delta_i| \cdot 100$

$$\Delta = \begin{pmatrix} -0.2 \\ 0.15 \\ -0.05 \\ 0.06 \\ -0.029 \end{pmatrix}$$

$xb := \frac{1}{n} \cdot \sum_{j=0}^{k-1} nx_j x_j$ $yb := \frac{1}{n} \cdot \sum_{j=0}^{m-1} ny_j y_j$ $xb2 := \frac{1}{n} \cdot \sum_{j=0}^{k-1} nx_j (x_j)^2$ $yb2 := \frac{1}{n} \cdot \sum_{j=0}^{m-1} ny_j (y_j)^2$

$\boxed{xb = 2}$ $\boxed{yb = 3.8}$ $\boxed{xb2 = 6}$ $\boxed{yb2 = 19}$

$Dx := xb2 - (xb)^2$ $\sigma x := \sqrt{Dx}$ $Dy := yb2 - (yb)^2$ $\sigma y := \sqrt{Dy}$

$\boxed{Dx = 2}$ $\boxed{\sigma x = 1.414}$ $\boxed{Dy = 4.56}$ $\boxed{\sigma y = 2.135}$

$g(xx) := yb + Rb \cdot \frac{\sigma y}{\sigma x} \cdot (xx - xb)$ $g(xx) \rightarrow \frac{4}{5} + \frac{3}{2} \cdot xx$ $Rb := \frac{sxy - n \cdot xb \cdot yb}{n \cdot \sigma x \cdot \sigma y}$

Выборочный коэффициент корреляции $\boxed{Rb = 0.993}$

$Dyx := \frac{1}{n} \cdot \sum_{j=0}^{k-1} nx_j \cdot (g(x_j) - yb)^2$ $\sigma yx := \sqrt{Dyx}$

$\boxed{Dyx = 4.5}$ $\boxed{\sigma yx = 2.121}$ $\eta := \frac{\sigma yx}{\sigma y}$

Корреляционное отношение $\boxed{\eta = 0.993}$