Лабораторная работа №5: Непрерывные распределения

Используемое ПО: MathCad.

Цель работы: Научиться с помощью программы MathCad находить основные характеристики непрерывного распределения.

Задание

1. Построить график плотности вероятности.
2. Проверить, выполняется ли условие нормировки.
3. Найти функцию распределения и построить ее график.
4. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины.
5. Найти вероятность попадания случайной величины в указанный интервал с помощью плотности вероятности и с помощью функции распределения.
6. Найти медиану и квантиль, соответствующую указанной вероятности.

Таблица 1 – Содержание вариантов к лабораторной работе №1

№	Содержание варианта	№	Содержание варианта
1	$\varphi(x)=\begin{cases}0, & x<0\\ \frac{1}{9}x^2, & 0\le x\le 3\\ 0, & x>3\end{cases}$ (1,2) p=0,1	2	$\varphi(x)=\begin{cases}0, & x<1\\ \frac{4}{x^5}, & x\ge 1\end{cases}$ (2,3) p=0,2
3	$\varphi(x)=\begin{cases}0, & x<1\\ \frac{9}{x^{10}}, & x\ge 1\end{cases}$ (3,4) p=0,3	4	$\varphi(x)=\begin{cases}0, & x<0\\ e^{-x}, & x\ge 0\end{cases}$ (1,2) p=0,4
5	$\varphi(x)=\begin{cases}0, & x<0\\ 2\cos 2x, & 0\le x\le \frac{\pi}{4}\\ 0, & x>\frac{\pi}{4}\end{cases}$ (4,5) p=0,6	6	$\varphi(x)=\begin{cases}0, & x<0\\ \frac{2}{9}(3x-x^2), & 0\le x\le 3\\ 0, & x>3\end{cases}$ (1,2) p=0,7

Окончание таблицы №1			
7	$\varphi(x)=0{,}5e^{-\lvert x\rvert}$ (1,5) p=0,8	8	$\varphi(x)=\begin{cases}0, & x<0\\ 0{,}5\sin x, & 0\le x\le\pi\\ 0, & x>\pi\end{cases}$ (1,2) p=0,1
9	$\varphi(x)=\begin{cases}0, & x<-1\\ \frac{2}{7}(2-\lvert x\rvert), & -1\le x\le 2\\ 0, & x>2\end{cases}$ (1,2) p=0,2	10	$\varphi(x)=\begin{cases}0, & x<0\\ 5e^{-5x}, & x\ge 0\end{cases}$ (3,6) p=0,3

Некоторые теоретические сведения

1-2. Функция плотности вероятности непрерывной случайной величины считается заданной корректно, если выполняются два условия:

а) $\varphi(x)\ge 0$;

б) $\int\limits_{-\infty}^{\infty}\varphi(x)dx=1$ (условие нормировки).

3. Функция распределения $F(x)$ случайной величины X определяется как

$$F(x)=\int\limits_{-\infty}^{x}\varphi(t)dt\,.$$

4. Если случайная величина X имеет непрерывное распределение, то для любой функции $g(x)$

$$M(g(X))=\int\limits_{-\infty}^{\infty}g(x)\varphi(x)dx$$

(при условии, что интеграл сходится абсолютно).

Таким образом,

$$M(X)=\int\limits_{-\infty}^{\infty}x\varphi(x)dx\,,$$

$$D(X)=M(X^2)-M^2(X)=\int\limits_{-\infty}^{\infty}x^2\varphi(x)dx-M^2(X)\,,$$

$$\sigma(X)=\sqrt{D(X)}\,.$$

5. Вероятность попадания случайной величины в интервал есть интеграл от плотности вероятности по этому промежутку:

$$p(a < X < b) = \int_a^b \varphi(x)dx,$$

или приращение функции распределения на этом промежутке:

$$p(a < X < b) = F(b) - F(a).$$

6. Квантиль – это функция, обратная функции распределения. То есть, если $p = F(x)$, то квантиль $x_p = F^{-1}(p)$. В частном случае, когда $p = 0{,}5$, квантиль называют медианой. Таким образом, для нахождения квантили, соответствующей вероятности *p*, следует решить уравнение

$$F(x) - p = 0.$$

Пояснения к работе с программой

1. Для выполнения работы понадобятся панели инструментов «*Графики*», «*Матанализ*», «*Арифметика*», «*Греческий алфавит*», «*Программирование*», «*Булево*». Вывести их на экран можно через пункт меню «*Вид / Панель инструментов*».

2. Программу желательно составлять так, чтобы она обладала универсальностью. В том случае, если придется изменить данные: функцию плотности, интервал или вероятность, то пусть в программе это придется сделать всего один раз – в самом ее начале.

3. MathCad различает большие и маленькие буквы.

4. Задать функцию кусочно можно с помощью команд панели программирования «*Add Line*» и «*If*».

5. Построить декартов график можно с помощью команды панели графиков «*Декартов график*». При этом следует указать в нижнем поле ввода имя независимой переменной и в левом – функцию.

6. Если известно, что корень уравнения $f(x) = 0$ находится на отрезке $[a,b]$, то его можно найти командой **root**. Первый ее параметр – функция $f(x)$, второй – имя независимой переменной, в нашем случае *x*, третий и четвертый параметры – границы отрезка *a* и *b*. Так, например, один из корней уравнения $x^2 = 3$ можно найти так: **root(*x*²–3,*x*,1,2)**.

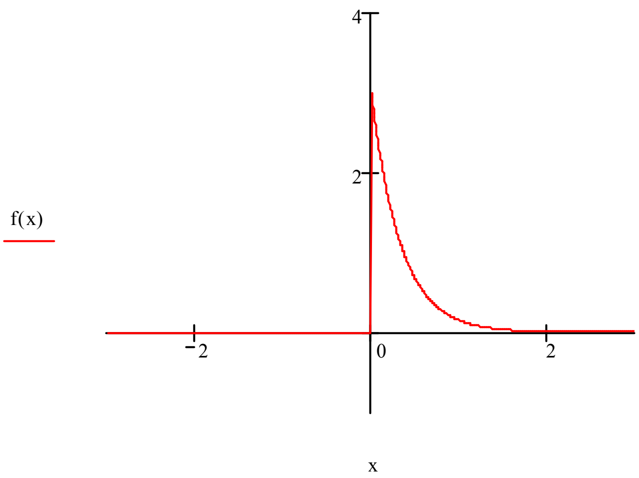
Программа MathCad
Разбор нулевого варианта

Пусть $\varphi(x)=\begin{cases}3e^{-3x}, x\geq 0\\0, \quad x<0\end{cases}$, интервал (1,2), вероятность p=0,1.

0. Условие задачи

$f(x) := \left| \begin{array}{l} 0 \text{ if } x<0 \\ 3\cdot e^{-3\cdot x} \text{ if } x\geq 0 \end{array}\right.$ $\qquad a := 1 \qquad b := 2 \qquad p := 0.1$

1. График функции плотности

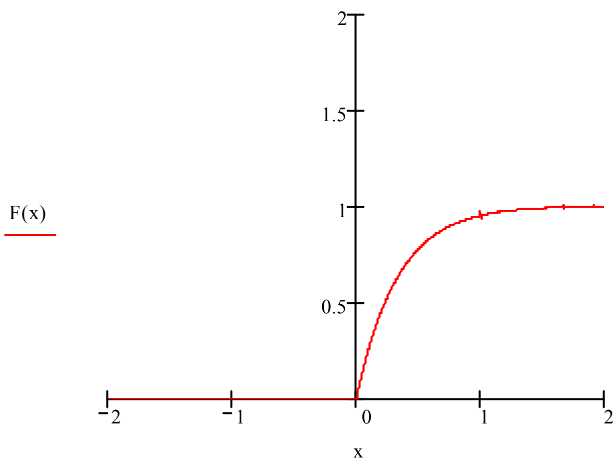


2. Условие нормировки

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

3. Функции распределения и ее график

$$F(x) := \int_{-\infty}^{x} f(t)\,dt$$



4. Математическое ожидание, дисперсия и стандартное отклонение

$M := \int_{-\infty}^{\infty} x\cdot f(x)\,dx \qquad M = 0.333$

$D := \int_{-\infty}^{\infty} x^2\cdot f(x)\,dx - M^2 \qquad D = 0.111$

$\sigma := \sqrt{D} \qquad \sigma = 0.333$

5. Вероятность попадания в интервал

$p1 := \int_{a}^{b} f(x)\,dx \qquad p1 = 0.047$

$p2 := F(b) - F(a) \qquad p2 = 0.047$

6. Медиана и квантиль

$\mathrm{root}(F(x) - 0.5, x, 0, 2) = 0.231 \qquad \mathrm{root}(F(x) - p, x, 0, 2) = 0.035$

Лабораторная работа №6: Сравнение двух выборок

Используемое ПО: MathCad, Excel, StatGraph, Stadia.

Цель работы: Научиться с помощью вышеуказанных программ находить основные характеристики случайных выборок и сравнивать их.

Задание

1. Найти средние арифметические и эмпирические стандарты для каждой из выборок.
2. Построить доверительные интервалы для математических ожиданий и стандартных отклонений.
3. Проверить гипотезу о равенстве дисперсий.
4. Если гипотеза о равенстве дисперсий принята, найти сводную оценку стандартного отклонения.
5. Проверить гипотезу о равенстве математических ожиданий.
6. Если гипотеза о равенстве математических ожиданий принята, найти сводную оценку математического ожидания и объединенную оценку стандартного отклонения.
7. Объединить две выборки в одну и проверить гипотезу о том, что экспериментальные данные имеют нормальный закон распределения. Рассматривать интервалы равной длины. Число интервалов равно L.
8. Построить гистограмму.

Указания:

1) Во всех пунктах брать уровень значимости $\alpha = 0{,}05$.
2) Выполнить пункты 1-7 в программе MathCad. В пункте 7 искать только теоретическую квантиль.
3) Выполнить пункты 1,2,3,5,7 в программе Excel. Доверительные интервалы для математических ожиданий искать с помощью встроенной команды и без нее. В пункте 7 искать только теоретическую квантиль.
4) Выполнить пункты 1,2,3,5,7,8 в программе StatGraph.
5) Выполнить пункты 1,2,3,5,7,8 в программе Stadia.

Содержание вариантов к лабораторной работе № 2

Вариант 1. L=7.

1 серия измерений. $n_1 = 28$.

7,2	7,1	3,7	5,3	6,4	5,2	9,7	8,8	6,3
5,9	6,9	4,5	9,0	5,5	6,1	7,5	8,9	3,7
6,3	6,1	6,3	7,2	6,2	3,5	9,0	6,4	7,5
9,8								

2 серия измерений. $n_2 = 31$.

6,7	5,1	7,4	5,9	9,8	6,6	8,8	9,3	7,9
5,6	7,2	6,2	6,8	5,4	6,8	8,2	9,3	8,0
6,0	6,0	7,6	7,5	8,9	4,9	5,8	8,5	8,9
8,7	6,4	6,6	5,7					

Вариант 2. L=7.

1 серия измерений $n_1 = 38$.

23,2	20,1	18,8	24,1	21,6	22,8	22,1	25,2	24,8
20,6	25,9	27,0	21,9	23,5	21,7	21,1	21,3	18,3
21,0	23,8	17,4	17,3	17,9	20,6	18,4	24,2	20,7
22,0	18,3	22,6	20,2	21,5	16,5	21,3	21,5	17,9
26,2	29,1							

2 серия измерений. $n_2 = 21$.

24,2	30,4	21,7	21,2	20,7	25,4	17,8	19,0	21,5
18,7	22,1	25,6	19,6	19,9	21,9	26,4	21,9	24,1
21,1	23,5	23,9						

Вариант 3. L=8.

1 серия измерений. $n_1 = 36$.

4,4	4,3	2,9	3,8	5,1	5,6	3,7	4,7	5,6
4,4	4,6	4,0	4,7	5,0	5,4	4,3	6,2	6,4
4,5	5,1	4,4	5,4	5,8	4,8	6,2	5,0	5,6
4,8	4,7	4,0	5,3	5,4	2,5	5,4	5,4	6,9

2 серия измерений. $n_2 = 22$.

5,0	4,3	3,9	6,0	3,4	4,1	4,7	3,4	4,2
5,1	3,5	3,1	4,3	3,7	3,7	6,2	4,8	3,5
4,3	6,2	2,7	7,1					

Вариант 4. L=7.

1 серия измерений $n_1 = 34$.

0,37	0,67	0,64	0,94	0,82	0,60	0,70	0,84	0,81
0,67	0,77	0,52	0,70	0,54	0,76	0,47	0,86	0,62
0,54	0,88	0,85	0,84	0,62	0,65	0,70	0,97	0,55
0,72	0,74	0,93	0,65	0,61	0,66	0,78		

2 серия измерений $n_2 = 23$.

0,84	0,85	0,84	0,64	0,97	0,78	0,81	0,73	0,93
0,60	0,79	0,69	0,85	0,73	0,72	0,84	0,54	0,95
0,42	0,90	0,90	0,87	0,78				

Вариант 5. L=8.

1 серия измерений. $n_1 = 32$.

23,5	26,5	22,8	26,0	23,7	25,8	22,7	18,9	24,6
26,1	26,2	20,9	23,0	31,9	23,2	21,3	21,8	24,1
26,7	19,3	27,1	26,6	29,5	20,5	21,8	26,8	23,9
20,3	21,5	24,8	25,8	20,5				

2 серия измерений. $n_2 = 24$.

24,6	20,4	20,1	25,7	24,0	24,2	19,5	18,3	21,2
18,8	15,4	22,1	24,2	20,4	19,2	15,2	18,0	21,5
22,2	21,4	17,0	17,8	25,4	21,5			

Вариант 6. L=6.

1 серия измерений. $n_1 = 30$.

4,9	4,8	3,6	4,8	6,5	8,6	5,5	7,7	6,1
7,0	7,5	7,9	5,7	6,2	6,6	5,0	7,2	5,9
7,2	4,0	6,3	6,1	5,0	6,7	3,6	6,4	3,9
3,5	5,4	6,5						

2 серия измерений. $n_2 = 25$.

4,5	5,8	5,0	7,1	5,7	6,2	6,2	6,6	5,6
4,9	5,5	5,7	5,3	6,7	5,5	7,1	6,2	3,0
6,3	7,9	5,9	9,5	6,7	7,3	5,5		

Вариант 7. L=7.

1 серия измерений. $n_1 = 28$.

0,22	0,75	0,77	0,59	0,85	0,99	0,79	1,10	1,04
0,51	0,83	0,72	0,65	1,03	0,92	0,53	0,63	0,85
0,91	0,72	0,62	0,58	0,81	0,91	0,81	1,02	1,15
0,35								

2 серия измерений. $n_2 = 26$.

0,65	0,96	0,85	0,98	0,69	1,01	0,79	0,62	0,71
0,84	1,13	0,81	1,02	0,65	0,54	0,78	0,69	0,65
0,61	0,61	0,72	1,01	0,54	0,58	0,70	0,82	

Вариант 8. L=6.

1 серия измерений. $n_1 = 26$.

29,8	29,5	30,4	30,4	28,5	35,6	29,3	28,0	26,4
24,2	32,3	26,2	22,9	25,9	27,5	20,2	28,4	22,7
21,3	23,3	23,2	29,7	24,0	26,5	28,5	24,7	

2 серия измерений. $n_2 = 27$.

27,9	32,9	29,1	31,1	28,9	34,1	35,7	23,7	33,9
25,2	25,3	29,3	29,7	29,9	36,7	24,7	32,5	30,4
26,4	31,1	28,8	34,7	32,9	36,1	30,5	39,7	30,8

Вариант 9. L=7.

1 серия измерений. $n_1 = 24$.

6,7	6,6	5,7	5,8	6,5	6,5	7,5	5,6	6,3
5,9	3,9	5,8	8,0	7,3	7,9	5,5	9,4	6,3
6,2	7,8	7,1	9,4	7,6	7,3			

2 серия измерений. $n_2 = 28$.

4,9	6,0	3,7	6,6	5,5	4,7	8,0	6,7	4,7
4,4	5,3	5,7	5,1	5,7	6,5	6,3	6,9	5,1
6,2	4,2	2,0	5,7	5,6	7,3	5,1	6,6	6,1
5,6								

Вариант 10. L=6.

1 серия измерений. $n_1 = 22$.

0,82	0,61	1,10	0,51	0,44	0,65	0,88	0,58	0,68
0,89	1,07	1,15	0,96	1,01	0,49	0,99	0,90	1,10
0,74	0,88	1,09	1,22					

2 серия измерений. $n_2 = 29$.

1,12	0,90	1,03	1,35	1,38	0,77	1,05	0,90	0,60
0,74	1,04	1,00	1,32	0,52	1,13	0,68	0,90	1,04
0,66	0,95	0,66	0,99	0,95	1,19	0,90	1,26	1,12
0,99	1,14							

Некоторые теоретические сведения

1. Несмещенными оценками математического ожидания и стандартного отклонения являются среднее арифметическое и эмпирический стандарт:

$$\overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}, \; s = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1}}.$$

2. Доверительные интервалы для математического ожидания и стандартного отклонения соответственно имеют вид:

$$\overline{y} - |t|_p \frac{s}{\sqrt{n}} < \beta < \overline{y} + |t|_p \frac{s}{\sqrt{n}}, \qquad (2.1)$$

$$s \cdot \sqrt{\frac{k}{a_2}} < \sigma < s \cdot \sqrt{\frac{k}{a_1}}.$$

Здесь $p = 1 - \alpha$, $k = n - 1$, a_1 и a_2 – квантили распределения Пирсона для вероятностей $\frac{1-p}{2}$ и $\frac{1+p}{2}$ соответственно. Видно, что доверительный интервал для математического ожидания симметричен относительно среднего арифметического, поэтому иногда считают, что интервал найден, если указаны среднее арифметическое и радиус интервала $R = |t|_p \frac{s}{\sqrt{n}}$.

3. Пусть $F_{Э1}$ – отношение большей эмпирической дисперсии к меньшей, $F_{Э2}$ – меньшей к большей. Пусть F_{m1} и F_{m2} – квантили распределения Фишера $F(p_1, k_1, k_2)$ и $F(p_2, k_2, k_1)$. Здесь $p_1 = 1 - \frac{\alpha}{2}$, $p_2 = \frac{\alpha}{2}$, k_1 – число степеней свободы большей эмпирической дисперсии k_2 – меньшей. Критерий Фишера

требует отвергнуть гипотезу о равенстве дисперсий, если $F_{Э1} > F_{m1}$ ($F_{Э2} < F_{m2}$). В противном случае гипотеза принимается.

4. Если гипотеза о равенстве дисперсий не противоречит экспериментальным данным, можно найти сводную оценку дисперсии

$$s_{св}^2 = \frac{s_1^2 k_1 + s_2^2 k_2}{k_1 + k_2},$$

где s_1^2 и s_2^2 – эмпирические дисперсии с числом степеней свободы k_1 и k_2 соответственно.

5. Пусть

$$T_Э = \frac{|\overline{y_1} - \overline{y_2}|}{s_{св}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

а $|t|_p$ – квантиль распределения модуля отношения Стьюдента для вероятности $p = 1 - \alpha$ и числа степеней свободы $k = k_1 + k_2$. Критерий Стьюдента требует отвергнуть гипотезу о равенстве математических ожиданий, если $T_Э > |t|_p$. В противном случае гипотеза принимается.

6. Если гипотеза о равенстве математических ожиданий не противоречит экспериментальным данным, можно найти сводную оценку математического ожидания и объединенную оценку дисперсии:

$$\overline{y_{св}} = \frac{\overline{y_1} n_1 + \overline{y_2} n_2}{n_1 + n_2},\ s_{об}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n_1} y_i^2 + \sum_{i=1}^{n_2} y_i^{/2} - n \cdot \overline{y_{св}}^2\right).$$

Здесь $n = n_1 + n_2$, $\{y_i\}$ и $\{y_i^/\}$ – данные выборки. Заметим, что $\overline{y_{св}}$ и $s_{об}$ есть обычное среднее арифметическое и обычный эмпирический стандарт для объединенной выборки.

7. Для проверки гипотезы о том, что экспериментальные данные распределены по нормальному закону, можно применить критерий согласия Пирсона. Для этого весь диапазон изменения случайной величины разбивают на L участков точками $x_1 = -\infty$, x_2, …, $x_{L+1} = +\infty$, подсчитывают частоты попадания в эти участки $\frac{N_1}{n}$, $\frac{N_2}{n}$, …, $\frac{N_L}{n}$, а также теоретические вероятности попадания в эти же участки:

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{x_i}^{x_{i+1}} e^{-\frac{U^2}{2}} dU.$$

Затем вычисляют значение величины

$$\chi^2 = \sum_{i=1}^{L} \frac{(N_i - np_i)^2}{np_i}$$

и сравнивают его с квантилью распределения Пирсона $\chi_p^2(k)$ для вероятности $p = 1 - \alpha$ и числа степеней свободы $k = L - 3$. Критерий Пирсона требует отвергнуть гипотезу о том, что экспериментальные данные распределены по нормальному закону, если $\chi^2 > \chi_p^2(k)$, в противном случае гипотеза принимается. На практике точки дробления x_i выбирают так, чтобы либо все интервалы имели одинаковую длину, либо все вероятности p_i равнялись бы между собой.

8. Если по оси абсцисс отметить точки x_i из пункта 7, только вместо x_1 и x_{L+1} взять наименьшее и наибольшее числа в выборке, и на каждом отрезке построить прямоугольник высотой $\frac{N_i}{n \cdot h_i}$, где h_i – длина *i*-го участка, то графическое изображение полученных данных даст гистограмму.

Пояснения к работе с программой MathCad

1. Для выполнения работы понадобятся панели инструментов «*Арифметика*», «*Греческий алфавит*», «*Программирование*», «*Матрицы*». Вывести их на экран можно через пункт меню «*Вид / Панель инструментов*».

2. Программу желательно составлять так, чтобы она обладала универсальностью. В том случае, если придется изменить выборки, то пусть в программе при этом ничего не придется менять.

3. Перед тем, как приступить к работе с MathCad’ом, следует подготовить два текстовых файла с выборками, например, с помощью программы Блокнот. Набирать данные следует либо в одну строку, разделяя их пробелами, либо в один столбец. В качестве десятичного разделителя следует использовать точку. Сохраните эти текстовые файлы в той же папке, где будет храниться файл, созданный в MathCad’е.

4. Укажем некоторые команды, которые нам понадобятся для выполнения работы.

READPRN(“FileName.txt”)

Считать данные из текстового файла.

length(x)

Длина массива x.

mean(x)

Среднее арифметическое массива x.

stdev(x)

Смещенная оценка стандартного отклонения $\tilde{s} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n}}$.

qchisq(p,k)

Квантиль распределения Пирсона для вероятности p и числа степеней свободы k.

qt(p,k)

Обычная квантиль (не модуля) отношения Стьюдента для вероятности p и числа степеней свободы k. Заметим, что из обычной квантили qt(p,k) можно получить нужную нам квантиль модуля так: $-qt\left(\frac{1-p}{2}, k\right)$.

qF(p,k_1,k_2)

Квантиль распределения Фишера для вероятности p и числа степеней свободы k_1 и k_2.

stack(x,y)

Объединение двух массивов-столбцов в один.

augment(x,y)

Объединение двух массивов-строк в один.

$\boldsymbol{x:=x^{\mathrm{T}}}$

Транспонирование массива x. Некоторые команды программы MathCad применимы к массивам-строкам, некоторые к массивам-столбцам. С помощью команды транспонирования (панель «*Матрицы*») можно преобразовывать рассматриваемый массив.

Пояснения к работе с программой Excel

1. Список всех доступных команд можно получить с помощью кнопки быстрого доступа «*Вставка функции*» или через меню «*Вставка / Функция…*».

2. Укажем некоторые команды, которые нам понадобятся для выполнения работы.

СРЗНАЧ(A_i:B_j)

Среднее арифметическое диапазона ячеек A_i:B_j.

СТАНДОТКЛОН(A_i:B_j)

Эмпирический стандарт диапазона ячеек A_i:B_j.

КОРЕНЬ(A_i)

Квадратный корень.

ХИ2ОБР(α,k)

Квантиль распределения Пирсона для уровня значимости α и числа степеней свободы k.

СТЬЮДРАСПОБР(α,k)

Квантиль распределения модуля отношения Стьюдента для уровня значимости α и числа степеней свободы k.

FРАСПОБР(α,k_1,k_2)

Квантиль распределения Фишера для уровня значимости α и числа степеней свободы k_1 и k_2.

ДОВЕРИТ(α,s,n)

Радиус доверительного интервала для математического ожидания. Здесь α – уровень значимости, s – эмпирический стандарт, n – число элементов в выборке. Заметим, что эта команда находит радиус, заменяя в формуле (2.1) квантиль модуля отношения Стьюдента квантилью модуля стандартного нормального распределения. Учитывая то, что при $n \to \infty$ распределение Стьюдента стремится к стандартному нормальному, при достаточно больших n эти квантили практически совпадают.

Пояснения к работе с программой StatGraph

1. Следует подготовить три столбца с экспериментальными данными: два с соответствующими выборками и третий – с объединенной выборкой.

2. В качестве десятичного разделителя следует использовать запятую.

3. Для выполнения пунктов 1,2,3,5 используйте пункт меню «*Compare / Two Samples / Two-sample Comparison…*». В окошке «*Two-sample Comparison*» используйте кнопку быстрого доступа «*Tabular Options*» для получения доступа к разделам «*Comparison Of Means*» и «*Comparison Of Standard Deviations*».

4. Для выполнения пунктов 7,8 используйте пункт меню «*Describe / Distribution Fitting / Uncensored Data…*». В окошке «*Uncensored Data*» используйте кнопки быстрого доступа «*Tabular Options*» и «*Graphical Options*» для получения доступа к разделам «*Test For Normality*» и «*Frequency Histogram*» соответственно.

Пояснения к работе с программой Stadia

1. Следует подготовить три столбца с экспериментальными данными: два с соответствующими выборками и третий – с объединенной выборкой.

2. В качестве десятичного разделителя следует использовать точку.

3. Для выполнения всех пунктов воспользуемся пунктом меню «*Статист-F9*».

4. Для выполнения пунктов 1,2 воспользуемся кнопкой «*Описательная статистика*».

5. Для выполнения пунктов 3,5 воспользуемся кнопкой «*Стьюдента и Фишера*».

6. Для выполнения пунктов 7,8 воспользуемся кнопкой «*Гистограмма/Нормальность*».

Разбор нулевого варианта

Вариант 0. L=8.

1 серия измерений. $n_1 = 26$.

0,44	1,29	1,25	1,06	1,24	0,87	0,96	1,25	0,88
1,09	0,90	0,75	1,09	0,73	1,04	1,07	1,20	1,15
0,94	0,83	1,19	0,88	0,77	0,84	0,79	0,80	

2 серия измерений. $n_2 = 32$.

0,89	0,97	0,33	0,93	0,84	1,43	1,34	0,88	0,75
0,96	1,09	0,83	0,95	0,33	0,56	1,20	1,12	0,92
0,73	1,30	0,70	1,27	0,82	0,86	1,30	1,00	1,12
0,69	1,03	0,58	1,26	1,16				

Программа MathCad
Разбор нулевого варианта

0. Подготовка

$X := READPRN("00Vib1.txt")$ $Y := READPRN("00Vib2.txt")$

$n1 := length(X)$ $n2 := length(Y)$

$k1 := n1 - 1$ $k2 := n2 - 1$

$a := 0.05$ $p := 1 - a$

X =

	0
0	0.44
1	1.29
2	1.25
3	1.06
4	1.24
5	0.87
6	0.96
7	1.25
8	0.88
9	1.09
10	0.9
11	0.75
12	1.09
13	0.73
14	1.04
15	1.07

Y =

	0
0	0.89
1	0.97
2	0.33
3	0.93
4	0.84
5	1.43
6	1.34
7	0.88
8	0.75
9	0.96
10	1.09
11	0.83
12	0.95
13	0.33
14	0.56
15	1.2

1. Средние и стандарты

$x1 := mean(X)$ $x1 = 0.973$

$y1 := mean(Y)$ $y1 = 0.942$

$s1 := stdev(X) \cdot \sqrt{\frac{n1}{k1}}$ $s1 = 0.207$

$s2 := stdev(Y) \cdot \sqrt{\frac{n2}{k2}}$ $s2 = 0.276$

2. Доверительные интервалы

Для сигма 1

$p1 := \frac{1-p}{2}$ $p2 := \frac{1+p}{2}$ $a1 := qchisq(p1, k1)$ $a2 := qchisq(p2, k1)$

$sig11 := s1 \cdot \sqrt{\frac{k1}{a2}}$ $sig12 := s1 \cdot \sqrt{\frac{k1}{a1}}$ $sig11 = 0.163$ $sig12 = 0.286$

Для сигма 2

$a1 := qchisq(p1, k2)$ $a2 := qchisq(p2, k2)$

$sig21 := s2 \cdot \sqrt{\frac{k2}{a2}}$ $sig22 := s2 \cdot \sqrt{\frac{k2}{a1}}$ $sig21 = 0.221$ $sig22 = 0.366$

Для бета 1

$tp := -qt\left(\frac{1-p}{2}, k1\right)$ $r1 := tp \cdot \frac{s1}{\sqrt{n1}}$ $beta11 := x1 - r1$ $beta11 = 0.889$

$beta12 := x1 + r1$ $beta12 = 1.057$

Для бета 2

$tp := -qt\left(\frac{1-p}{2}, k2\right)$ $r2 := tp \cdot \frac{s2}{\sqrt{n2}}$ $beta21 := y1 - r2$ $beta21 = 0.843$

$beta22 := y1 + r2$ $beta22 = 1.041$

3. Гипотеза о равенстве дисперсий

$$F1 := \begin{cases} \dfrac{s1^2}{s2^2} & \text{if } (s1 > s2) \\ \dfrac{s2^2}{s1^2} & \text{if } (s2 \ge s1) \end{cases} \qquad K1 := \begin{cases} k1 & \text{if } (s1 > s2) \\ k2 & \text{if } (s2 \ge s1) \end{cases} \qquad K2 := \begin{cases} k1 & \text{if } (s2 > s1) \\ k2 & \text{if } (s1 \ge s2) \end{cases}$$

$F1 = 1.768$ $\qquad F2 := qF\left(1 - \dfrac{a}{2}, K1, K2\right)$ $\qquad F2 = 2.174$

$$Flag := \begin{cases} \text{"No"} & \text{if } (F1 > F2) \\ \text{"Yes"} & \text{if } (F2 \ge F1) \end{cases} \qquad Flag = \text{"Yes"}$$

4. Сводная оценка стандартного отклонения

$$Ssv := \frac{s1^2 \cdot k1 + s2^2 \cdot k2}{k1 + k2} \qquad Ssv := \sqrt{Ssv} \qquad Ssv = 0.247$$

5. Гипотеза о равенстве математических ожиданий

$$T1 := \frac{|x1 - y1|}{Ssv \cdot \sqrt{\dfrac{1}{n1} + \dfrac{1}{n2}}} \qquad T1 = 0.478 \qquad T2 := -qt\left(\frac{1-p}{2}, k1 + k2\right) \qquad T2 = 2.003$$

$$Flag := \begin{cases} \text{"No"} & \text{if } (T1 > T2) \\ \text{"Yes"} & \text{if } (T2 \ge T1) \end{cases} \qquad Flag = \text{"Yes"}$$

6. Сводная оценка математического ожидания, объединенный стандарт

$$Ysv := \frac{x1 \cdot n1 + y1 \cdot n2}{n1 + n2} \qquad Ysv = 0.956 \qquad z := stack(X, Y)$$

$$Sob := stdev(z) \cdot \sqrt{\frac{length(z)}{length(z) - 1}} \qquad Sob = 0.246$$

7. Квантиль распределения Пирсона для критерия согласия

$$\chi := qchisq(p, 5) \qquad \chi = 11.07$$

z =

	0
0	0.44
1	1.29
2	1.25
3	1.06
4	1.24
5	0.87
6	0.96
7	1.25
8	0.88
9	1.09
10	0.9
11	0.75
12	1.09
13	0.73
14	1.04
15	1.07

Программа Excel
Разбор нулевого варианта

0,44	0,89	26	**x1-сред.**	**y1-сред.**	**S1**	**S2**
1,29	0,97	32	0,973077	0,941875	0,207283	0,275616
1,25	0,33	8	**Сигма1-1**	**Сигма1-2**	**Сигма2-1**	**Сигма2-2**
1,06	0,93		0,162563	0,286135	0,220962	0,366426
1,24	0,84		**Квантили ст. норм. распр.**			
0,87	1,43		**Бета1-1**	**Бета1-2**	**Бета2-1**	**Бета2-2**
0,96	1,34		0,893402	1,052752	0,846381	1,037369
1,25	0,88		**Квантили распр. Стьюд.**			
0,88	0,75		**Бета1-1**	**Бета1-2**	**Бета2-1**	**Бета2-2**
1,09	0,96		0,889354	1,0568	0,842505	1,041245
0,9	1,09		**F-эксп.**	**F-квант.**	**F-эксп.**	**F-квант.**
0,75	0,83		0,565611	0,460064	1,767999	2,173607
1,09	0,95		**Т-эксп.**	**Т-квант.**		
0,73	0,33		0,47757	2,003239		
1,04	0,56		**Х2-квант.**			
1,07	1,2		11,07048			
1,2	1,12					
1,15	0,92					
0,94	0,73					
0,83	1,3					
1,19	0,7					
0,88	1,27					
0,77	0,82					
0,84	0,86					
0,79	1,3					
0,8	1					
	1,12					
	0,69					
	1,03					
	0,58					
	1,26					
	1,16					

Лабораторная работа №7: Регрессионный анализ

Используемое ПО: MathCad, Excel, StatGraph, Stadia.

Цель работы: Научиться с помощью вышеуказанных программ строить регрессионные модели, проверять их адекватность и значимость коэффициентов.

Задание

1. Найти линейное, квадратичное и кубическое регрессионные уравнения, уравнение вида $y = e^{ax+b}$. Изобразить на одном графике найденную экспоненциальную функцию и результаты эксперимента. Построить график остатков для экспоненциальной модели. Данные считать из файла. В качестве значений y брать средние арифметические. Все веса считать равными 1 (Программа MathCad).

2. Найти линейное уравнение регрессии. В качестве значений y брать средние арифметические. Все веса считать равными 1. Проверить адекватность построенной модели и значимость коэффициентов (Программа Excel).

3. Найти линейное и квадратичное уравнения регрессии с учетом весов. Для каждой регрессионной модели построить график найденной функции и результатов эксперимента. Проверить адекватность каждой построенной модели и значимость всех коэффициентов (Программа StatGraph).

4. Найти линейное, квадратичное и кубическое регрессионные уравнения, уравнение вида $y = e^{ax+b}$. Изобразить на одном графике найденную квадратичную функцию и результаты эксперимента. В качестве значений y брать средние арифметические. Все веса считать равными 1. Проверить адекватность каждой построенной модели и значимость всех коэффициентов (Программа Stadia).

5. Напишите программу в MathCad для построения полноценных линейной и квадратичной регрессионных моделей с учетом весов, проверки адекватности построенных моделей и значимости всех коэффициентов.

Таблица 2 – Содержание вариантов к лабораторной работе № 3

Вариант 1					
X	*Y*				
0,00	4,96	4,94	4,96	4,94	
0,10	5,42	5,46	5,42	5,39	
0,15	5,63	5,64	5,63	5,64	
0,25	6,00	6,03	6,02		
0,30	6,19	6,25	6,20		
0,55	7,04	7,04	7,04		
0,70	7,46	7,41	7,43		
0,75	7,54	7,54	7,53	7,60	
0,80	7,66	7,65	7,64	7,64	
0,90	7,87	7,79	7,81	7,86	
Вариант 2					
X	*Y*				
5	53	52	53		
10	63	63	62	64	
25	91	91	91	90	
30	96	94	97	97	
40	108	108	107		
55	117	118	116		
70	117	116	117	117	
75	114	114	114	115	
90	103	102	101	106	
100	89	89	91		
Вариант 3					
X	*Y*				
50	48,8	48,4	48,6	48,7	48,2
80	46,3	46,3	46,2	46,2	
90	45,6	45,3	45,8		
110	44,9	44,8	44,9	44,7	
130	43,3	43,4	43,4		
160	41,8	41,5	42,0		
170	41,5	41,5	41,7	41,4	
180	40,5	40,6	40,6		
200	39,9	39,8	40,2	40,4	
230	38,9	38,8	38,7	39,0	38,6

Продолжение таблицы №2

Вариант 4					
X	*Y*				
0,10	5,33	5,32	5,31		
0,20	5,70	5,69	5,73	5,73	5,71
0,25	5,93	5,91	5,87	5,94	
0,30	6,07	6,08	6,12	6,09	6,09
0,40	6,38	6,38	6,41		
0,65	7,13	7,15	7,09		
0,80	7,43	7,44	7,44	7,45	7,44
0,85	7,56	7,52	7,53	7,58	
0,90	7,63	7,66	7,62	7,59	7,61
1,05	7,82	7,86	7,86		
Вариант 5					
X	*Y*				
0	37	37	37	39	37
10	60	63	63		
20	81	83	82		
30	94	97	97	97	96
50	115	115	112		
60	121	118	118		
70	117	118	118	121	118
75	115	117	114		
85	110	111	111		
100	90	90	89	93	88
Вариант 6					
X	*Y*				
10	50,6	50,7	50,7		
30	49,2	48,8	49,4	49,1	
50	47,2	47,8	47,9	48,1	47,7
70	46,8	46,2	46,4	46,1	46,6
100	44,4	44,5	44,8		
110	43,9	43,9	43,9		
150	42,1	41,8	41,8	41,7	41,6
170	40,7	41,0	40,7	41,0	40,5
190	40,3	39,7	39,8	39,9	
220	38,4	38,7	38,8		

Продолжение таблицы №2

Вариант 7					
X	*Y*				
0,00	4,96	4,95	5,00	4,94	
0,05	5,21	5,23	5,23		
0,20	5,83	5,81	5,82	5,88	5,86
0,30	6,28	6,20	6,22		
0,40	6,59	6,60	6,55	6,54	6,54
0,50	6,86	6,89	6,90	6,88	6,87
0,65	7,32	7,34	7,26		
0,70	7,42	7,39	7,36	7,35	7,44
0,80	7,65	7,64	7,63		
0,90	7,81	7,77	7,85	7,90	
Вариант 8					
X	*Y*				
5	52	51	52	52	52
25	90	90	88		
35	102	104	102	105	103
40	110	109	106	106	
45	111	113	112		
50	114	115	112		
60	119	119	117	117	
65	117	118	118	120	116
80	112	115	110		
95	98	97	97	100	97
Вариант 9					
X	*Y*				
10	51,4	51,4	50,9		
40	49,7	49,5	49,2	49,1	49,0
50	48,2	48,7	48,7		
60	47,6	47,8	47,7		
100	45,0	45,3	45,3	45,1	44,9
120	43,6	43,7	43,8	43,9	44,1
150	42,2	42,1	42,3		
180	40,9	40,9	41,0	40,6	41,1
190	40,8	40,6	40,7		
200	40,1	40,1	40,2		

Окончание таблицы №2					
Вариант 10					
X	*Y*				
0,00	4,89	4,96	4,88	4,89	4,89
0,05	5,13	5,15	5,13	5,15	
0,20	5,70	5,76	5,71		
0,30	6,06	6,06	6,05	6,07	
0,40	6,40	6,40	6,38		
0,55	6,87	6,86	6,86		
0,60	7,00	7,01	6,96	7,02	
0,65	7,10	7,10	7,11		
0,85	7,55	7,49	7,56	7,54	
0,90	7,68	7,64	7,62	7,63	7,63

Некоторые теоретические сведения

1. Для построения линейной регрессионной модели $y = ax + b$ по методу наименьших квадратов коэффициенты a и b ищут исходя из условия минимизации взвешенной суммы квадратов отклонений результатов эксперимента от точек на прямой: $MinZ = \sum_{i=1}^{n}(ax_i + b - y_i)^2 w_i$. Нахождение минимума этой функции после нахождения частных производных по параметрам a и b сведется к решению системы линейных уравнений:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 w_i + b\sum_{i=1}^{n} x_i w_i = \sum_{i=1}^{n} x_i y_i w_i \\ a\sum_{i=1}^{n} x_i w_i + b\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} y_i w_i \end{cases}.$$

2. Для построения квадратичной регрессионной модели $y = ax^2 + bx + c$ по методу наименьших квадратов коэффициенты a, b и c ищут исходя из условия минимизации взвешенной суммы квадратов отклонений результатов эксперимента от точек на параболе: $MinZ = \sum_{i=1}^{n}(ax_i^2 + bx_i + c - y_i)^2 w_i$. Нахождение минимума этой функции после нахождения частных производных по параметрам a, b и c сведется к решению системы линейных уравнений:

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 w_i + b\sum_{i=1}^{n} x_i^3 w_i + c\sum_{i=1}^{n} x_i^2 w_i = \sum_{i=1}^{n} x_i^2 y_i w_i \\ a\sum_{i=1}^{n} x_i^3 w_i + b\sum_{i=1}^{n} x_i^2 w_i + c\sum_{i=1}^{n} x_i w_i = \sum_{i=1}^{n} x_i y_i w_i \\ a\sum_{i=1}^{n} x_i^2 w_i + b\sum_{i=1}^{n} x_i w_i + c\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} y_i w_i \end{cases}.$$

3. Аналогично строятся и другие регрессионные модели.

4. Заметим, что из рассматриваемых программ только StatGraph позволяет сразу искать регрессионные модели с учетом весов.

5. Для проверки адекватности построенных моделей в том случае, если имел место повторный эксперимент, используют следующую схему. Вычисляют $s_{ад}^2 = \sum_{i=1}^{n} \frac{(\Delta Y_i)^2 w_i}{k_{ад}}$, где $\Delta Y_i = (ax_i + b - y_i)$ для линейной модели (a и b – найденные коэффициенты регрессионной модели), $\Delta Y_i = ax_i^2 + bx_i + c - y_i$ для квадратичной модели (a, b и c – найденные коэффициенты регрессионной модели), $k_{ад} = n - l$, где l – число оцениваемых параметров, то есть $l = 2$ для линейной модели, $l = 3$ – для квадратичной. Величина $s_{ад}^2$ (дисперсия адекватности) иногда обозначается $s_{ост}^2$ (остаточная дисперсия). Затем вычисляют $s_{св}^2 = \frac{\sum_{i=1}^{n} s_i^2 k_i}{\sum_{i=1}^{n} k_i}$, где $s_i^2 = \sum_{j=1}^{w_i} \frac{(y_{ij} - \overline{y_i})^2}{k_i}$, $k_i = w_i - 1$. Если $s_{ад}^2 < s_{св}^2$, то модель является адекватной. Если нет, то находят $F_э = \frac{s_{ад}^2}{s_{св}^2}$ и сравнивают с квантилью распределения Фишера $F_m = F(p, k_{ад}, k_{св})$, где $k_{св} = \sum_{i=1}^{n} k_i$, $p = 1 - \alpha$, а α – уровень значимости. Если $F_э < F_m$, то модель является адекватной на уровне значимости α, в противном случае – нет.

6. Для проверки адекватности построенных моделей в том случае, если не имел место повторный эксперимент, используют следующую схему. Сначала вычисляют $s_{ад}^2$ (см. пункт 5), все веса принимаются равными 1. Затем вычисляют $s_{мод}^2 = \sum_{i=1}^{n} \frac{(\Delta \overline{Y}_i)^2}{l - 1}$, где $\Delta \overline{Y}_i = ax_i + b - \overline{y}$ для линейной модели,

$\Delta \overline{Y}_i = ax_i^2 + bx_i + c - \overline{y}$ – для квадратичной, а $\overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}$. Величина $s_{мод}^2$ (дисперсия модели) иногда обозначается $s_{рег}^2$ (регрессионная дисперсия). После этого сравнивают величину $F_э = \frac{s_{мод}^2}{s_{ад}^2}$ с квантилью распределения Фишера $F_m = F(p, l-1, n-l)$. Если окажется, что $F_э < F_m$, то гипотеза об адекватности модели экспериментальным данным отвергается (на уровне значимости α), в противном случае гипотеза принимается.

7. Коэффициент уравнения регрессии называется незначимым, если его математическое ожидание (истинное значение) равно нулю. Для проверки значимости найденных коэффициентов регрессионной модели находят отношения этих коэффициентов к их стандартным ошибкам. Если через A обозначить матрицу системы для нахождения коэффициентов регрессионной модели, то стандартную ошибку *i*-го коэффициента d_i можно найти так:

$$s(d_i) = \sqrt{s_{ад}^2 \cdot b_{ii}},$$

где b_{ii} – соответствующий элемент матрицы $B = A^{-1}$.

Если модуль этого отношения больше квантили модуля отношения Стьюдента для числа степеней свободы $k_{ад}$, то есть $\left|\frac{d_i}{s(d_i)}\right| > |t|_p \ (k = k_{ад})$, то коэффициент считается значимым, в противном случае – нет.

Пояснения к работе с программой MathCad (задание 1)

1. Следует подготовить два текстовых файла, например, в «Блокноте» со значениями *x* и средних арифметических *y*.

2. Укажем некоторые команды, которые нам понадобятся для выполнения работы.

slope(*x*,*y*)

Возвращает коэффициент при *x* в линейной регрессионной модели.

intercept(*x*,*y*)

Возвращает свободный член в линейной регрессионной модели.

regress(*x*,*y*,*l*–1)

Возвращает вектор, последние *l* координат которого – коэффициенты степенной регрессионной модели степени *l*–1.

3. MathCad позволяет искать практически любые регрессионные модели, даже не линейные относительно параметров. Для этого необходимо задать вектор, координатами которого являются приближающая функция и ее частные производные по параметрам (в примере – это вектор *F*). Затем следует задать начальное приближение для параметров (в примере – это вектор *vb*). Начальное приближение можно брать наугад, но затем всегда следует проверять правильность построенной модели графически, так как в случае неудачно подобранного начального приближения программа может выдать неправильный результат. Далее надо воспользоваться командой **genfit(*x*,*y*,*vb*,*F*)**, которая возвратит коэффициенты регрессионной модели.

4. Индексация в массивах в MathCad'е начинается с нуля.

Пояснения к работе с программой Excel

1. Для построения линейной регрессионной модели в программе Excel следует воспользоваться командой **ЛИНЕЙН(*Y*,*X*)**. И так как результатом этой команды является массив (из двух чисел), то, указав диапазон ячеек для *x* и для *y*, следует нажать комбинацию клавиш Ctrl+Shift+Enter.

2. Если же указать диапазон ячеек из двух столбцов и пяти строк, то программа выдаст дополнительную статистику, а именно:

Коэффициент *a* линейной модели	Коэффициент *b* линейной модели
Стандартная ошибка коэф-та *a*	Стандартная ошибка коэф-та *b*
Коэффициент детерминации	Стандартная ошибка для *y* ($s_{ад}$)
F-наблюдаемое ($F_э$)	Степени свободы ($k_{ад} = n - l$)
Регрессионная сумма квадратов ($(\Delta \overline{Y}_i)^2$)	Остаточная сумма квадратов ($(\Delta Y_i)^2$)

Пояснения к работе с программой StatGraph

1. Для построения линейной регрессионной модели воспользуйтесь пунктом меню «*Relate / Multiple Regression…*», для построения квадратичной модели – «*Relate / Polynomial Regression…*».

2. Столбец «p-Value» из первой таблицы показывает уровни значимости коэффициентов регрессионной модели. Если уровень значимости некоторого коэффициента не превосходит заданного уровня значимости α, то этот коэффициент считается значимым, в противном случае – незначимым. Столбец «p-Value» из второй таблицы показывает уровень значимости $F_Э$. Если уровень

значимости $F_Э$ не превосходит заданного уровня значимости α, то модель считается адекватной, в противном случае – не адекватной.

Пояснения к работе с программой Stadia

1. Следует подготовить два столбца с экспериментальными данными, один – со значениями *x*, второй – со средними арифметическими для *y*.
2. В качестве десятичного разделителя следует использовать точку.
3. Для выполнения задания воспользуемся пунктом меню «*Статист-F9*». В появившемся окне следует нажать на кнопку «*Простая регрессия/ тренд*», затем выбрать вид регрессионной модели.
4. Строка «Значим.» из первой таблицы показывает уровни значимости коэффициентов регрессионной модели. Вывод относительно адекватности построенной модели выдается текстовым сообщением.

Пояснения к работе с программой MathCad (задание 5)

1. Для выполнения работы понадобятся панели инструментов «*Арифметика*», «*Матанализ*», «*Программирование*», «*Матрицы*», «*Булево*». Вывести их на экран можно через пункт меню «*Вид / Панель инструментов*».
2. Вновь будем составлять по возможности универсальную программу для своего задания.
3. Перед тем, как приступить к работе с MathCad'ом, следует подготовить два текстовых файла, например, с помощью программы Блокнот. Один – со значениями *x* (в нашем примере эти значения записаны в столбец), второй – со значениями *y* (в нашем примере для фиксированного *x* соответствующие значения *y* записаны в строки). В качестве десятичного разделителя следует использовать точку. Сохраните эти текстовые файлы в той же папке, где будет хранится файл, созданный в MathCad'е.
4. Для выполнения работы нам понадобятся кнопки «*Add Line*», «*Локальное присвоение*», «*Цикл For*», «*Оператор If*» (панель «*Программирование*»); «*Суммирование*» (панель «*Матанализ*»); «*Булево равенство*» (панель «*Булево*»); «*Создать матрицу или вектор*», «*Нижний индекс*» (панель «*Матрицы*»).
5. Команда **lsolve(*a*,*b*)** возвращает вектор, являющийся решением системы линейных уравнений с матрицей системы *a* и вектором свободных членов *b*.
6. Напомним, что индексация в массивах в MathCad'е начинается с нуля.

Разбор нулевого варианта

Таблица 3 – Содержание нулевого варианта к лабораторной работе № 3

Нулевой вариант					
X	*Y*				
10	50,7	50,6	50,6		
20	49,9	50,0	50,1	50,2	
40	48,3	48,4	48,4		
60	47,4	47,4	46,8	47,4	
80	46,0	45,9	46,2	46,1	45,5
120	43,2	43,3	43,1	43,5	43,0
140	41,9	42,2	41,9	42,2	
150	41,7	41,7	42,0		
180	40,2	40,4	40,5	40,1	
200	39,5	39,9	39,7		

Программа MathCad (задание 1)
Разбор нулевого варианта

Подготовка

X := READPRN("Var0-x.txt")   Y := READPRN("Var0-ys.txt")

X =

	0
0	10
1	20
2	40
3	60
4	80
5	120
6	140
7	150
8	180
9	200

Y =

	0
0	50.633
1	50.05
2	48.367
3	47.25
4	45.94
5	43.22
6	42.05
7	41.8
8	40.3
9	39.7

Линейное уравнение регрессии

$\mathrm{slope}(X,Y) = -0.060$ $\mathrm{intercept}(X,Y) = 50.886$

Ответ y=-0.060x+50.886

Квадратичное и кубическое уравнения

$Z1 := \mathrm{regress}(X,Y,2)$ $Z2 := \mathrm{regress}(X,Y,3)$

$$Z1 = \begin{pmatrix} 3 \\ 3 \\ 2 \\ 51.566 \\ -0.081 \\ 1.073\times 10^{-4} \end{pmatrix} \qquad Z2 = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 51.316 \\ -0.067 \\ -6.525\times 10^{-5} \\ 5.453\times 10^{-7} \end{pmatrix}$$

Ответ y=0.0001073x^2-0.081x+51.566

Ответ y=0.0000005453x^3-0.00006525x^2-0.067x+51.316

Экспоненциальное уравнение регрессии, график

$$F(w,u) := \begin{pmatrix} e^{u_0\cdot w+u_1} \\ w\cdot e^{u_0\cdot w+u_1} \\ e^{u_0\cdot w+u_1} \end{pmatrix} \qquad vb := \begin{pmatrix} 0.1 \\ 1 \end{pmatrix} \qquad p := \mathrm{genfit}(X,Y,vb,F) \qquad p = \begin{pmatrix} -1.336\times 10^{-3} \\ 3.935 \end{pmatrix}$$

$g(r) := F(r,p)_0$ $i := 0..9$

Ответ y=e^(-0.001336x+3.935)

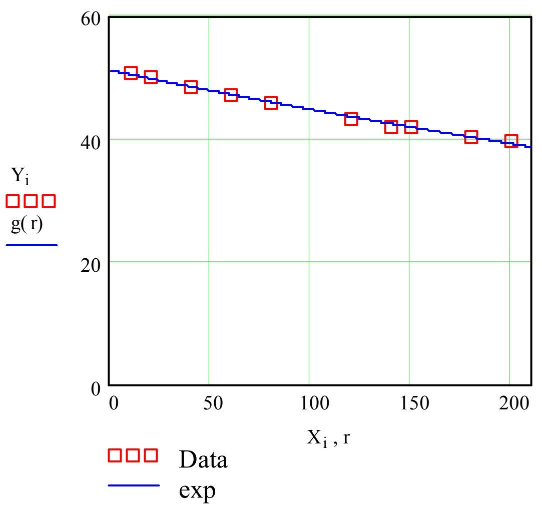
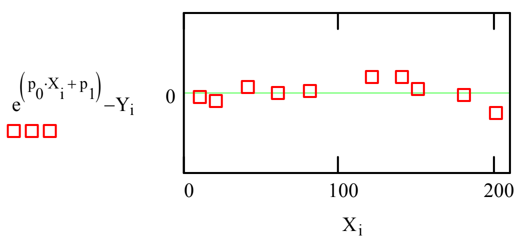


График остатков



Программа Excel
Разбор нулевого варианта

X	Y				
10	50,633333				
20	50,05				
40	48,366667				
60	47,25				
80	45,94				
120	43,22				
140	42,05				
150	41,8				
180	40,3				
200	39,7				
Линейная модель:					
-0,0595463	50,8856341	**Коэффициенты**	-28,984815	208,593654	**Отношения коэф-в к их ошибкам**
0,0020544	0,24394622	**Их стандартные ошибки**	2,30600563	**|t|(k=8)**	**Оба коэф. знач., т.к. 29,0>2,3 и 208,6>2,3**
0,99056737	0,41598385	**Коэф. детерм. / Станд. ош. для у**			
840,119521	8	**F-наблюдаемое / Степени свободы**	5,31764499	**F(1;8)**	**Модель адекватна, т.к. 5,3<840,1**
145,376437	1,38434052	**Регрес. / Остаточ. суммы квадратов**			

Программа Mathcad (задание 5)
Разбор нулевого варианта

Подготовка

x := READPRN("Var0-x.txt")

y := READPRN("Var0-y.txt")

n := 10   p := 0.95

$$ysr := \begin{array}{|l} \text{for } i \in 0..n-1 \\ \quad ysr_i \leftarrow \sum_{j=0}^{w_i-1} \frac{y_{i,j}}{w_i} \\ ysr \end{array}$$

$$w := \begin{pmatrix} 3 \\ 4 \\ 3 \\ 4 \\ 5 \\ 5 \\ 4 \\ 3 \\ 4 \\ 3 \end{pmatrix}$$

x =

	0
0	10
1	20
2	40
3	60
4	80
5	120
6	140
7	150
8	180
9	200

y =

	0	1	2	3	4
0	50.7	50.6	50.6	0	0
1	49.9	50	50.1	50.2	0
2	48.3	48.4	48.4	0	0
3	47.4	47.4	46.8	47.4	0
4	46	45.9	46.2	46.1	45.5
5	43.2	43.3	43.1	43.5	43
6	41.9	42.2	41.9	42.2	0
7	41.7	41.7	42	0	0
8	40.2	40.4	40.5	40.1	0
9	39.5	39.9	39.7	0	0

Построение линейной модели

$$a1 := \begin{bmatrix} \sum_{i=0}^{n-1} (x_i)^2 \cdot w_i & \sum_{i=0}^{n-1} x_i \cdot w_i \\ \sum_{i=0}^{n-1} x_i \cdot w_i & \sum_{i=0}^{n-1} w_i \end{bmatrix} \qquad a2 := \begin{pmatrix} \sum_{i=0}^{n-1} x_i \cdot w_i \cdot ysr_i \\ \sum_{i=0}^{n-1} ysr_i \cdot w_i \end{pmatrix}$$

a := lsolve(a1, a2)

$$a = \begin{pmatrix} -0.060 \\ 50.881 \end{pmatrix}$$

$$b := a1^{-1}$$

Ответ: y=-0.060x+50.881

Построение квадратичной модели

$$A1 := \begin{bmatrix} \sum_{i=0}^{n-1} (x_i)^4 \cdot w_i & \sum_{i=0}^{n-1} (x_i)^3 \cdot w_i & \sum_{i=0}^{n-1} (x_i)^2 \cdot w_i \\ \sum_{i=0}^{n-1} (x_i)^3 \cdot w_i & \sum_{i=0}^{n-1} (x_i)^2 \cdot w_i & \sum_{i=0}^{n-1} (x_i) \cdot w_i \\ \sum_{i=0}^{n-1} (x_i)^2 \cdot w_i & \sum_{i=0}^{n-1} (x_i) \cdot w_i & \sum_{i=0}^{n-1} w_i \end{bmatrix} \qquad A2 := \begin{bmatrix} \sum_{i=0}^{n-1} (x_i)^2 \cdot w_i \cdot ysr_i \\ \sum_{i=0}^{n-1} x_i \cdot w_i \cdot ysr_i \\ \sum_{i=0}^{n-1} w_i \cdot ysr_i \end{bmatrix}$$

A := lsolve(A1, A2)

$$A = \begin{pmatrix} 1.062 \times 10^{-4} \\ -0.082 \\ 51.591 \end{pmatrix}$$

$$B := A1^{-1}$$

Ответ: y=0.0001062x^2-0.082x+51.591

Проверка адекватности обеих моделей

$$Sad1 := \sum_{i=0}^{n-1} \left(a_0 \cdot x_i + a_1 - ysr_i\right)^2 \cdot w_i \qquad Sad1 := \frac{1}{n-2} \cdot Sad1 \qquad Sad1 = 0.6353$$

$$Sad2 := \sum_{i=0}^{n-1} \left[A_0 \cdot \left(x_i\right)^2 + A_1 \cdot x_i + A_2 - ysr_i\right]^2 \cdot w_i \qquad Sad2 := \frac{1}{n-3} \cdot Sad2 \qquad Sad2 = 0.1107$$

$$Ssv := \frac{\displaystyle\sum_{i=0}^{n-1} \left[\sum_{j=0}^{w_i-1} \frac{\left(y_{i,j} - ysr_i\right)^2}{w_i - 1}\right] \cdot \left(w_i - 1\right)}{\displaystyle\sum_{i=0}^{n-1} \left(w_i - 1\right)} \qquad Ssv = 0.0394$$

$$F1 := \frac{Sad1}{Ssv} \qquad F1 = 16.1234 \qquad F2 := \frac{Sad2}{Ssv} \qquad F2 = 2.8089$$

$$Fkv1 := qF\left[p, n-2, \sum_{i=0}^{n-1} \left(w_i - 1\right)\right] \qquad Fkv2 := qF\left[p, n-3, \sum_{i=0}^{n-1} \left(w_i - 1\right)\right]$$

$$Fkv1 = 2.291 \qquad Fkv2 = 2.359$$

$$FlagLi1 := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } (Sad1 < Ssv) \\ \text{"No"} \ \text{ if } (Sad1 > Ssv) \end{array}\right. \qquad FlagSq1 := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } (Sad2 < Ssv) \\ \text{"No"} \ \text{ if } (Sad2 > Ssv) \end{array}\right.$$

$$FlagLi2 := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } F1 < Fkv1 \\ \text{"No"} \ \text{ if } F1 > Fkv1 \end{array}\right. \qquad FlagSq2 := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } F2 < Fkv2 \\ \text{"No"} \ \text{ if } F2 > Fkv2 \end{array}\right.$$

$$FlagLi := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } (FlagLi1 = \text{"Yes"}) \\ FlagLi2 \ \text{ if } FlagLi1 \ne \text{"Yes"} \end{array}\right. \qquad FlagSq := \left|\begin{array}{l} \text{"Yes"} \ \text{ if } (FlagSq1 = \text{"Yes"}) \\ FlagSq2 \ \text{ if } FlagSq1 \ne \text{"Yes"} \end{array}\right.$$

$$FlagLi = \text{"No"} \qquad FlagSq = \text{"No"}$$

Ответ: обе модели не адекватны

Проверка значимости коэффициентов

Стандартные ошибки коэффициентов

$s11 := \sqrt{Sad1} \cdot \sqrt{b_{0,0}}$ $\quad s11 = 2.126\times 10^{-3}$ $\quad s12 := \sqrt{Sad1} \cdot \sqrt{b_{1,1}}$ $\quad s12 = 0.249$

$s21 := \sqrt{Sad2} \cdot \sqrt{B_{0,0}}$ $\quad s21 = 1.702\times 10^{-5}$ $\quad s22 := \sqrt{Sad2} \cdot \sqrt{B_{1,1}}$ $\quad s22 = 3.582\times 10^{-3}$

$s23 := \sqrt{Sad2} \cdot \sqrt{B_{2,2}}$ $\quad s23 = 0.154$

Отношения коэффициентов к их ошибкам

$t11 := \frac{a_0}{s11} \qquad t12 := \frac{a_1}{s12} \qquad t21 := \frac{A_0}{s21} \qquad t22 := \frac{A_1}{s22} \qquad t23 := \frac{A_2}{s23}$

Проверка значимости

$Tkv1 := -qt\left(\frac{1-p}{2}, n-2\right)$ $\quad Tkv1 = 2.306$ $\quad Tkv2 := -qt\left(\frac{1-p}{2}, n-3\right)$ $\quad Tkv2 = 2.365$

$$Flag11 := \left| \begin{array}{l} \text{"a1\_Znachim"} \ \text{ if } (|t11| > Tkv1) \\ \text{"a1\_Ne\_Znachim"} \ \text{ if } (|t11| < Tkv1) \end{array} \right. \qquad Flag12 := \left| \begin{array}{l} \text{"a2\_Znachim"} \ \text{ if } (|t12| > Tkv1) \\ \text{"a2\_Ne\_Znachim"} \ \text{ if } (|t12| < Tkv1) \end{array} \right.$$

$$Flag21 := \left| \begin{array}{l} \text{"A1\_Znachim"} \ \text{ if } (|t21| > Tkv2) \\ \text{"A1\_Ne\_Znachim"} \ \text{ if } (|t21| < Tkv2) \end{array} \right. \qquad Flag22 := \left| \begin{array}{l} \text{"A2\_Znachim"} \ \text{ if } (|t22| > Tkv2) \\ \text{"A2\_Ne\_Znachim"} \ \text{ if } (|t22| < Tkv2) \end{array} \right.$$

$$Flag23 := \left| \begin{array}{l} \text{"A3\_Znachim"} \ \text{ if } (|t23| > Tkv2) \\ \text{"A3\_Ne\_Znachim"} \ \text{ if } (|t23| < Tkv2) \end{array} \right.$$

Flag11 = "a1_Znachim"

Flag12 = "a2_Znachim"

Flag21 = "A1_Znachim"

Flag22 = "A2_Znachim"

Flag23 = "A3_Znachim"

Ответ: все коэффициенты значимы

Лабораторная работа №8 Исследование линейной корреляции

Используемое ПО: MathCad.

Цель работы: Научиться с помощью программы MathCad делать выводы о силе и характере связи между двумя величинами.

Задание

1. Найти эмпирический коэффициент корреляции. Найти уравнения эмпирических прямых регрессии y на x и x на y. На одном чертеже построить прямые регрессии.
2. Проверить гипотезу о незначимости коэффициента корреляции.
3. Построить доверительный интервал для коэффициента корреляции и сделать вывод о силе и характере связи между величинами.

Таблица 4 – Содержание вариантов к лабораторной работе №4

Вариант 1									
X	– 3	1	2	– 5	0	3	–1	5	7
Y	0,3	0,7	0,9	0,1	0,5	1,0	0,4	1,3	1,8
Вариант 2									
X	– 3	1	2	– 5	0	3	–1	5	7
Y	1,6	1,2	1,3	2,0	1,7	1,0	0,4	0,5	0,2
Вариант 3									
X	1,5	1,0	2,5	4,0	7,5	5,5	6,0	7,0	9,0
Y	0,27	0,15	0,34	0,42	0,91	0,58	0,72	0,84	1,12
Вариант 4									
X	1,5	1,0	2,5	4,0	7,5	5,5	6,0	7,0	9,0
Y	1,98	2,03	1,54	1,33	0,51	1,20	1,09	0,33	0,31
Вариант 5									
X	2,1	1,4	2,7	3,9	1,9	5,2	4,3	3,2	6,0
Y	3,6	0,1	3,5	5,1	2,0	9,1	7,4	4,3	9,8
Вариант 6									
X	2,1	1,4	2,7	3,9	1,9	5,2	4,3	3,2	6,0
Y	8,3	9,9	7,2	6,2	8,0	4,0	5,4	6,9	2,8
Вариант 7									
X	– 3	1	2	– 5	– 4	9	–1	5	7
Y	0,4	1,7	1,9	0,3	0,5	3,6	0,9	2,3	2,8

Окончание таблицы №4									
Вариант 8									
X	-3	1	2	-5	-4	9	-1	5	7
Y	5,3	3,7	3,9	6,1	5,5	2,0	4,4	3,3	2,8
Вариант 9									
X	-3	1	-2	-5	6	3	-1	5	7
Y	0,23	0,47	0,39	0,12	2,35	1,04	0,42	2,23	3,02
Вариант 10									
X	-3	1	-2	-5	6	3	-1	5	7
Y	7,63	5,78	6,92	8,11	0,5	4,60	6,40	3,93	3,08

Некоторые теоретические сведения

1. Напомним, что эмпирические дисперсии рассчитываются по следующим формулам:

$$s_x^2 = \tilde{K}_{11} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}, \; s_y^2 = \tilde{K}_{22} = \sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1}.$$

Оценкой ковариации $\operatorname{cov}(x, y) = K_{12}$ является величина

$$\tilde{K}_{12} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})(y_i - \overline{y})}{n-1}.$$

Эмпирический коэффициент корреляции находится по формуле:

$$R = \frac{\tilde{K}_{12}}{\sqrt{\tilde{K}_{11}\tilde{K}_{22}}} = \frac{\tilde{K}_{12}}{s_x s_y}.$$

Уравнение эмпирической прямой регрессии Y на X задается так:

$$\frac{y - \overline{y}}{s_y} = R\frac{x - \overline{x}}{s_x};$$

уравнение прямой X на Y так:

$$\frac{y - \overline{y}}{s_y} = \frac{1}{R}\frac{x - \overline{x}}{s_x}.$$

2. Эмпирический коэффициент корреляции называется незначимым, если его истинное значение $\rho = 0$, то есть линейная зависимость между величинами X и Y отсутствует. Для проверки гипотезы о незначимости коэффициента корреляции сравним экспериментальное значение

$$T_Э = \left| R\sqrt{\frac{n-2}{1-R^2}} \right|$$

с квантилью модуля отношения Стьюдента $|t|_p$ $(k = n-2)$. Если $T_Э > |t|_p$, то гипотеза отклоняется, в противном случае гипотеза принимается.

3. Доверительный интервал для коэффициента корреляции имеет вид:

$$th\left(Z - \frac{|u|_p}{\sqrt{n-3}}\right) < \rho < th\left(Z + \frac{|u|_p}{\sqrt{n-3}}\right),$$

где $Z = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)$, $|u|_p$ – квантиль модуля стандартного нормального распределения. Если $\rho = 0$ не принадлежит найденному доверительному интервалу, то гипотеза о существовании линейной зависимости принимается, как не противоречащая экспериментальным данным; в противном случае гипотеза о существовании линейной зависимости отвергается.

Пояснения к работе с программой MathCad

1. Для выполнения работы понадобятся панели инструментов «*Арифметика*», «*Программирование*», «*Матрицы*», «*Булево*», «*Графики*». Вывести их на экран можно через пункт меню «*Вид / Панель инструментов*».

2. Программу желательно составлять так, чтобы она обладала универсальностью. В том случае, если придется изменить выборки, то пусть в программе при этом практически ничего не придется менять.

3. Укажем некоторые новые команды, которые нам понадобятся для выполнения работы.

corr(*x*,*y*)

Эмпирический коэффициент корреляции.

qnorm(*p*,*α*,*σ*)

Обычная квантиль (не модуля) нормального распределения с параметрами (α,σ) для вероятности *p*. Заметим, что из обычной квантили можно получить нужную нам квантиль модуля так:

$$qnorm\left(\frac{1+p}{2}, \alpha, \sigma\right).$$

tanh(*x*)

Гиперболический тангенс числа *x*.

Разбор нулевого варианта

Таблица 5 – Содержание нулевого варианта к лабораторной работе № 4

Нулевой вариант									
X	– 2,0	0,5	– 1,5	– 5,5	3,5	– 1,0	2,0	0,0	1,5
Y	0,11	0,09	0,13	0,11	0,06	0,12	0,08	0,11	0,07

Программа MathCad

Разбор нулевого варианта

0 Подготовка

$X := (\,-2\ \ 0.5\ \ -1.5\ \ -5.5\ \ 3.5\ \ -1\ \ 2\ \ 0\ \ 1.5\,)$

$Y := (\,0.11\ \ 0.09\ \ 0.13\ \ 0.11\ \ 0.06\ \ 0.12\ \ 0.08\ \ 0.11\ \ 0.07\,)$

$n := 9$

1 Эмпирический коэффициент корреляции, прямые регрессии

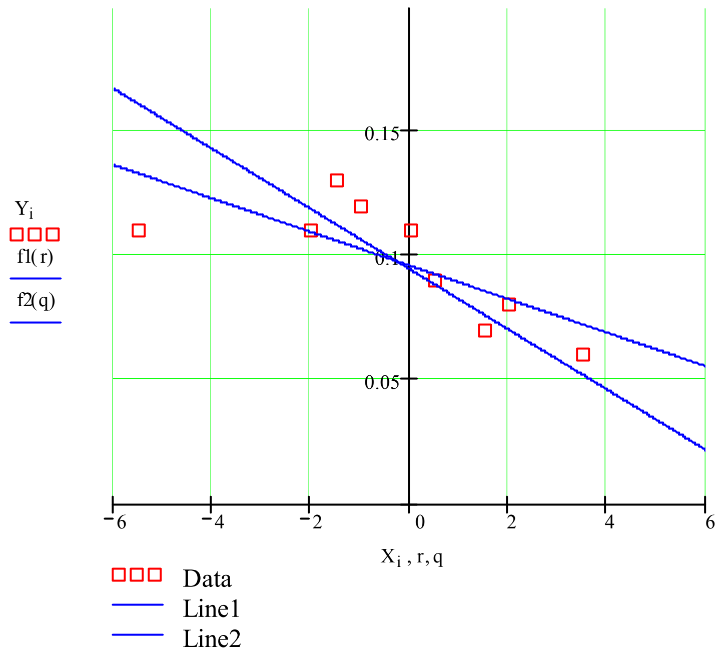
$x1 := \mathrm{mean}(X)$ $\qquad$ $y1 := \mathrm{mean}(Y)$

$sx := \mathrm{stdev}(X)\cdot\sqrt{\dfrac{n}{n-1}}$ $\qquad$ $sy := \mathrm{stdev}(Y)\cdot\sqrt{\dfrac{n}{n-1}}$

$R := \mathrm{corr}(X, Y)$

$f1(x) := y1 + R\cdot\dfrac{sy}{sx}\cdot(x - x1)$ $\qquad$ $f2(x) := y1 + \dfrac{1}{R}\cdot\dfrac{sy}{sx}\cdot(x - x1)$

$X := X^{T}$ $\qquad$ $Y := Y^{T}$ $\qquad$ $i := 0..8$



2 Проверка гипотезы о незначимости коэффициента корреляции

$t1 := \left| R \cdot \sqrt{\frac{n-2}{1-R^2}} \right|$ $t1 = 2.968$ $t2 := -qt(0.025, n-2)$ $t2 = 2.365$

$Flag := \left| \begin{array}{l} \text{"Znachim"} \quad \text{if } (t1 > t2) \\ \text{"Ne\_Znachim"} \quad \text{if } (t2 \ge t1) \end{array} \right.$ $Flag = \text{"Znachim"}$

3 Доверительный интервал для R

$Z := 0.5 \cdot \ln\left(\frac{1+R}{1-R}\right)$ $Z = -0.965$ $Sz := \sqrt{\frac{1}{n-3}}$ $Sz = 0.408$

$u := qnorm(0.975, 0, 1)$ $u = 1.96$

$r1 := \tanh(Z - u \cdot Sz)$ $r1 = -0.943$ $r2 := \tanh(Z + u \cdot Sz)$ $r2 = -0.163$

$Flag := \left| \begin{array}{l} \text{"Linei"} \quad \text{if } (r1 > 0) \vee (r2 < 0) \\ \text{"Ne\_Linei"} \quad \text{if } (r1 < 0) \wedge (r2 > 0) \end{array} \right.$ $Flag = \text{"Linei"}$