

Эконометрика

Эконометрика

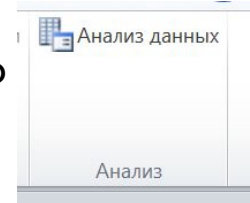
/Лабораторные работы в Microsoft Excel/

Лабораторная работа № 1

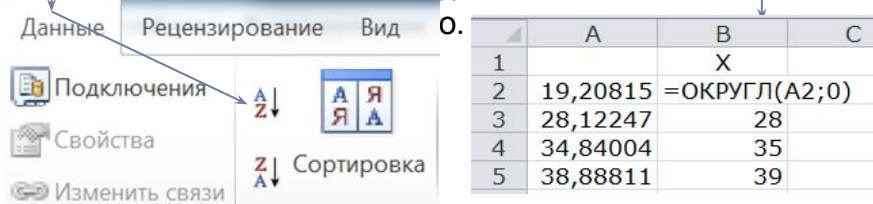
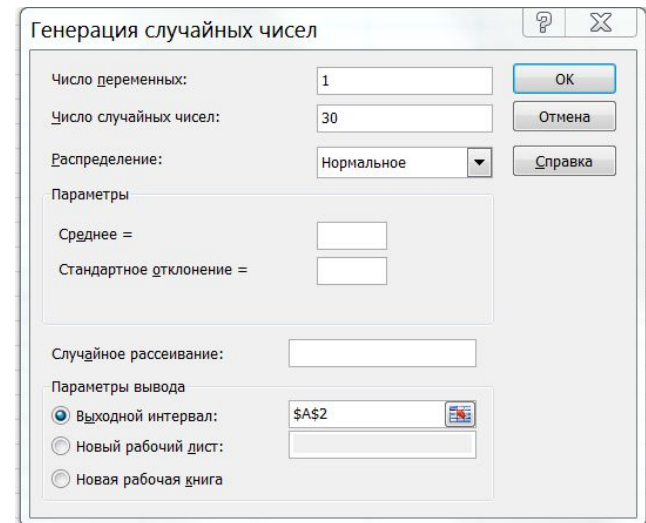
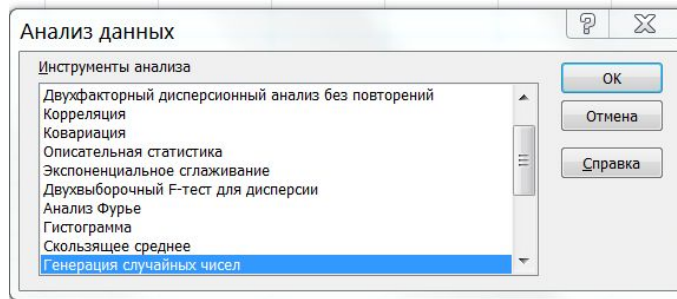
Основные описательные статистики

С помощью генератора случайных чисел для дискретной случайной величины X создать выборку объема $N=30$.

- Открыть меню «Анализ данных» на вкладке «Данные». (Если его нет, то установить: *Файл\Параметры\Надстройки*; выбрать *Пакет анализа* и нажать *Перейти*; поставить флажок *Пакет анализа* и нажать *ОК*).
- Выбрать Генерация случайных чисел.



- В появившемся окне задать параметры нормального распределения: среднее = $\{N \text{ варианта} * 10\}$; стандартное отклонение = $\{N \text{ варианта} * 2\}$.
- Отсортировать полученные значения в столбце А по возрастанию (вариационный ряд).
- Округлить значения (столбец В - функция ОКРУГЛ()). «Число_разрядов»



Для полученной выборки x^1, \dots, x^N объема $N=30$ вычислить

1. Оценку математического ожидания -выборочное среднее (функция СРЗНАЧ())

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

2. Оценку дисперсии (центральный момент второго порядка) - выборочную дисперсию s^2 , несмещенную оценку дисперсии s_x^2 (функция ДИСП.В())

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad s_x^2 = \frac{N}{N-1} s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad s_x^2 = \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i \right)^2}{N} \right]$$

$$\mu_2^* = s^2 \quad \mu_2 = s_x^2$$

3. Среднее квадратичное отклонение s_x (функция СТАНДОТКЛОН.В())

4. Стандартная ошибка среднего $\frac{s_x}{\sqrt{N}}$

5. Выборочный центральный момент 3-го порядка μ_3^* и несмещенную оценку μ_3

$$\mu_3^* = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3 \quad \mu_3 = \frac{N^2}{(N-1)(N-2)} \mu_3^*$$

6. Выборочный центральный момент 4-го порядка μ_4^*

$$\mu_4^* = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4$$

7. Коэффициент асимметрии (функция СКОС()) $\gamma = \frac{\mu_3}{\mu_2^{3/2}}$

8. Коэффициент эксцесса (функция ЭКСЦЕСС()) $\delta = \frac{\mu_4}{\mu_2^2} - 3 = \frac{N^2 - 1}{(N-2)(N-3)} \left[\frac{\mu_4^*}{\mu_2^{*2}} - 3 + \frac{6}{N+1} \right]$

9. Размах выборки $L = x^N - x^1$

10. Среднее отклонение (функция СРОТКΛ())

$$d = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

11. Относительное отклонение по модулю (линейный коэффициент вариации)

$$m = \frac{d}{\bar{x}}$$

12. Мера точности

$$h = \frac{1}{\sigma\sqrt{2}}$$

13. Вероятное отклонение

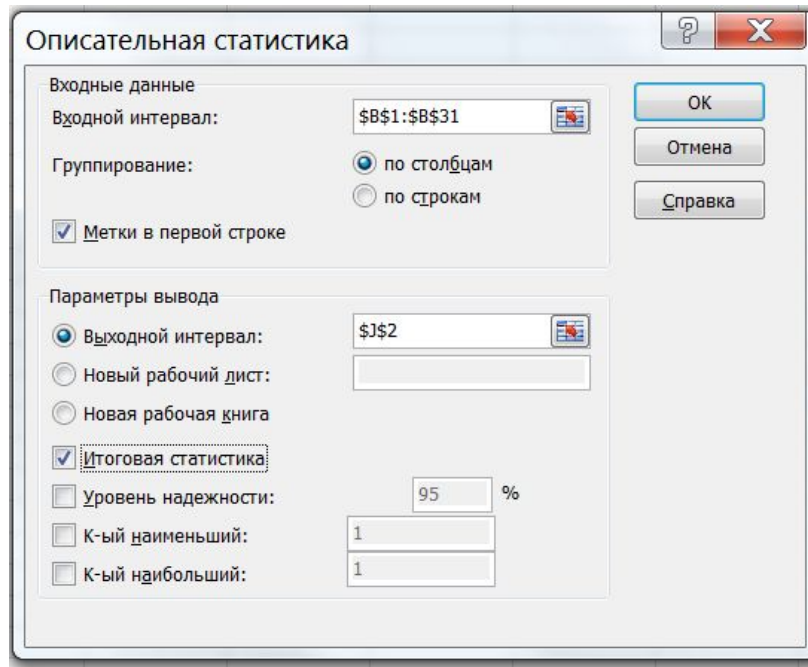
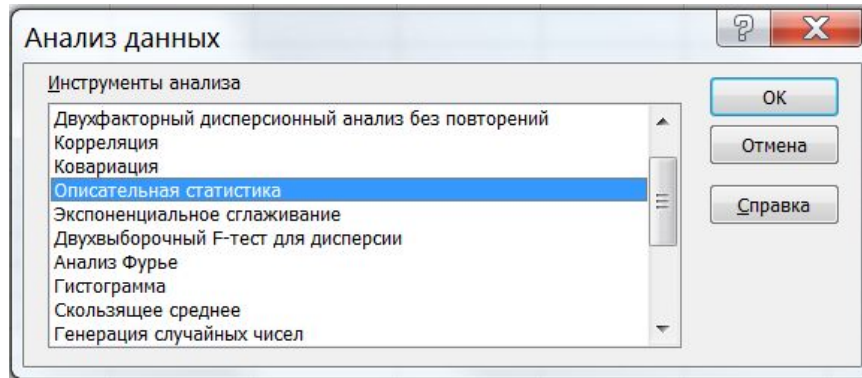
$$E = 0.674\sigma = \rho\sqrt{2}\sigma, \rho \approx 0.4769$$

14. Мера изменчивости (коэффициент вариации)

$$V = \frac{\sigma}{\bar{x}}$$



Вычислить параметры распределения с помощью Анализ данных\Описательная статистика и сравнить с рассчитанными по формулам



Пример таблицы с вычисленными описательными статистиками

	A	B
1		X
2	19,20815	19
3	28,12247	28
4	34,84004	35
5	38,88811	39
6	39,13438	39
7	41,74577	42
8	43,18121	43
9	44,00509	44
10	44,56359	45
11	45,53241	46
12	47,39712	47
13	47,54166	48
14	48,34275	48
15	49,24386	49
16	50,97806	51
17	51,6875	52
18	52,12221	52
19	52,95114	53
20	54,14445	54
21	56,11715	56
22	57,28376	57
23	58,7854	59
24	59,00814	59
25	59,63905	60
26	61,76827	62
27	63,15266	63
28	63,22933	63
29	69,11258	69
30	71,1865	71
31	79,74994	80

	D	E
1	N=	30

	F	G	H	I
1		Выборочная оценка	Несмещенная оценка	Функция
2	Среднее значение	51,1		51,1
3	Дисперсия	154,957	160,3	160,3
4	Ср.кв.отклонение	12,448	12,661	12,661
5	Размах выборки	61		
6	Среднее отклонение	9,567		9,567
7	Модуль	0,187		
8	Мера точности	0,057	0,056	
9	Вероятное отклонение	8,390	8,533	
10	Коэффициент вариации	0,2436	0,2478	
11	Станд.ошибка среднего		2,312	
12	μ_3	-359,788	-398,780	
13	Коэффициент асимметрии		-0,19649	-0,19649
14	μ_4	83333,98		
15	Коэффициент эксцесса		0,78975	0,78975
16				
17	Медиана	51,5		51,5
18	Мода			39

	J	K
	X	
Среднее		51,1
Стандартная ошибка		2,311565
Медиана		51,5
Мода		39
Стандартное отклонение		12,66096
Дисперсия выборки		160,3
Эксцесс		0,789746
Асимметричность		-0,19649
Интервал		61
Минимум		19
Максимум		80
Сумма		1533
Счет		30

Мода (функция МОДА.ОДН())

Медиана (функция МЕДИАНА())

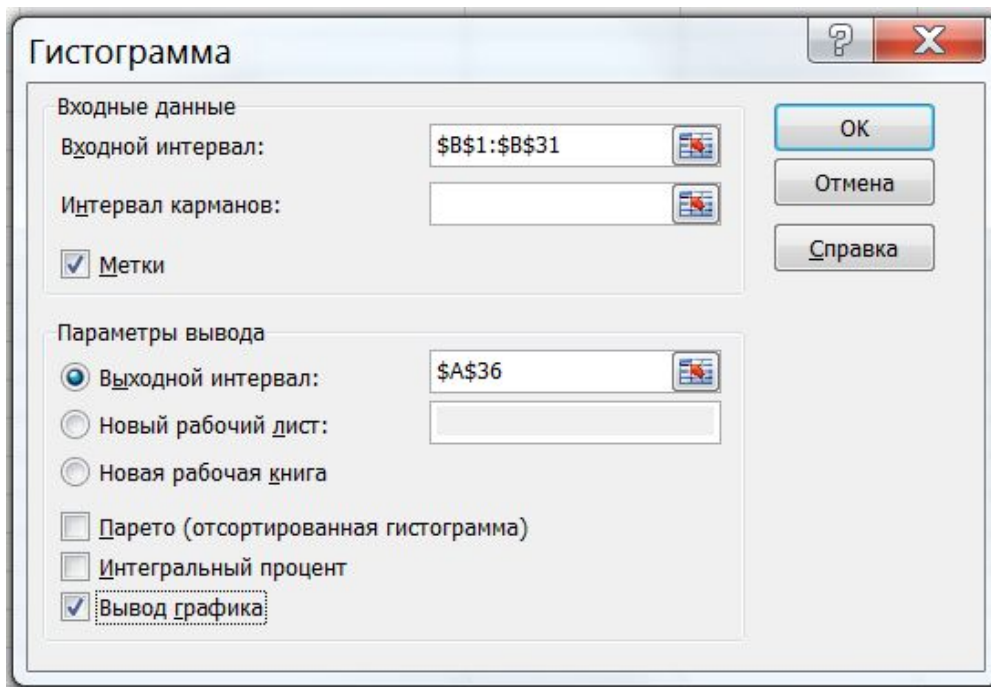
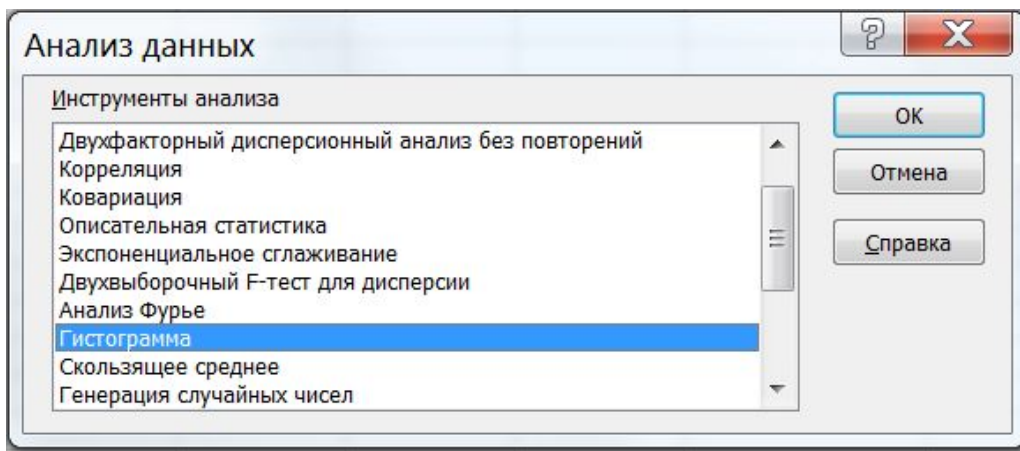
=

$$= \begin{cases} x_{N/2+1}, & N - \text{нечетное} \\ \frac{(x_{N/2} + x_{N/2+1})}{2}, & N - \text{четное} \end{cases}$$

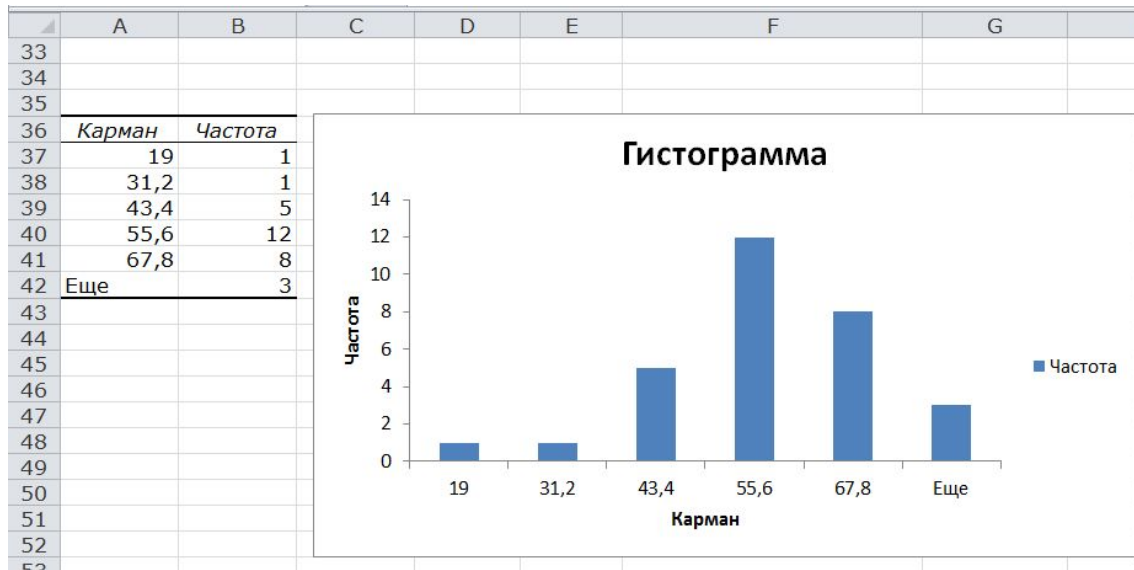
Примечание. Функция МОДА.ОДН () Возвращает наиболее часто встречающееся или повторяющееся значение в массиве или интервале данных.

Пример таблицы, полученной при выполнении «Анализа данных»

Построить гистограмму с помощью Анализ данных\Гистограмма



Пример гистограммы, построенной с помощью Анализ данных\Гистограмма (интервал Карманов выбирался автоматически)



Вычислить моду распределения по таблице частот.

1. Определить модальный интервал (по максимальной частоте). В примере: 43.4-55.6
2. Определить значение моды по формуле

$$M_0 = x_0 + h \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})},$$

где M_0 – значение моды (в примере = 51.16),

x_0 – нижняя граница модального интервала (43.4),

h – величина интервала (12.2),

f_m – частота модального интервала (12),

f_{m-1} – частота интервала, предшествующего модальному (5),

f_{m+1} – частота интервала, следующего за модальным (8).

Лабораторная работа № 2

Группировка данных

Сгруппировать полученные в лабораторной работе № 1 значения случайной величины X .

- Скопировать столбец X на новый рабочий лист (*Вставить\значения*), например, в A2:A31
- Выбрать число интервалов группировки. Рассчитать количество интервалов по формулам:

(Стерджесс) $K \approx 1 + 3.322 \lg N$

(Брукс и Каррузер) $K \leq 5 \sqrt[4]{N}$

(Хейнхольд) $K \approx \sqrt{N}$

(Максимов Ю. Д.) $K \approx 1.72 N^{1/3}$

$$h = \frac{(x_{\max} - x_{\min})}{K}$$

- Рассчитать длину интервала
- Округлить h , например, один знак до запятой – до десятых; два – до целых; три или более – кратно 50 или 100
- Рассчитать границы и середину интервалов

$$z_i = \frac{(x_{i-1} + x_i)}{2}$$

	B	C	D	E	F	G	H
1		N= 30			Интервал		Z
2					19	29	24
3	Стерджесс	≈	6		29	39	34
4	Брукс и Каррузер	≤	17		39	49	44
5	Хейнхольд	≈	5		49	59	54
6	Максимов Ю. Д.	≈	5		59	69	64
7					69	79	74
8			10,33				
9			h= 10				

Вычислить

- Абсолютные частоты n_j (функция ЧАСТОТА. Относится к диапазону ячеек: выделить диапазон для результата, ввести формулу, нажать *Ctrl+Shift+Enter*)

и $\frac{w_i}{h}$

$$w_i = \frac{n_i}{N}$$

- Накопленные частоты

$$\sum_{i=1}^j n_i = M_j$$

- Накопленные относительные частоты

$$\sum_{i=1}^j w_i = W_j$$

- Выборочную функцию распределения вероятности

$$F_N(x) = \sum_{z_j < x} w_j = \frac{1}{N} \sum_{z_j < x} n_j$$



Вычислить

- Среднее значение (для сгруппированных данных)

$$m_x^* = \sum_{j=1}^K w_j z_j = \frac{1}{N} \sum_{j=1}^K n_j z_j$$

- Дисперсию

$$s_x^2 = \frac{1}{N-1} \sum_{j=1}^K n_j (z_j - \bar{x})^2 - \frac{h^2}{12} = \frac{N}{N-1} \sum_{j=1}^K w_j (z_j - \bar{x})^2 - \frac{h^2}{12}$$

- Среднее квадратичное отклонение s_x
- Построить интегральную и дифференциальную функции для нормального закона распределения с полученными средним значением и дисперсией.
Вычислить с помощью функции НОРМ.РАСП(),
параметр *Интегральная* = 0 для дифференциального закона распределения
= 1 для интегрального закона распределения
- Вычислить дифференциальную функцию распределения по формуле

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - m_x)^2}{2\sigma_x^2}\right)$$

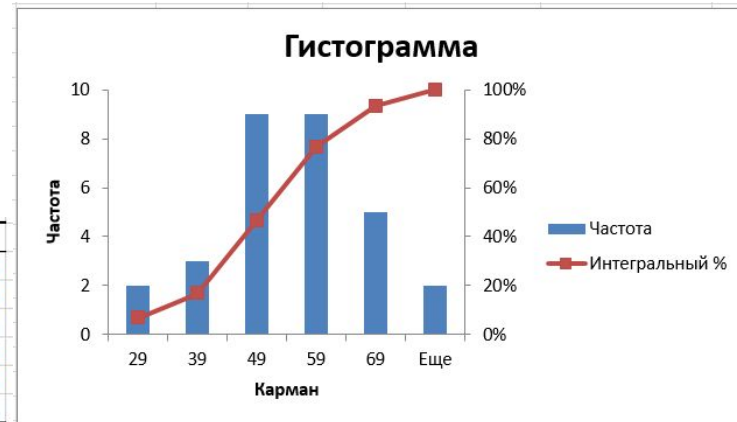
Примечание. У операции «унарный минус» выше приоритет, чем у \wedge , поэтому в формуле нужны скобки. Например, EXP(-((H2-\$H\$11)^2))...

Пример таблицы

	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Интервал		Z	Абсолютные частоты	Относительные частоты	Накоплен. частоты	Накоплен. относител. частоты	Эмпирич. диффер. функция распределения W_i/h	Эмпирич. интегральная функция распределения $F(z_i)$	Нормальное распрод. дифферен. функция	Нормальное распрод. ф-ция (по формуле)	Нормальное распрод. интегральная функция
2		19	29	24	2	0,067	2	0,067	0,007	0	0,0036	0,0036	0,0182
3		29	39	34	3	0,1	5	0,167	0,010	0,067	0,0140	0,0140	0,0990
4		39	49	44	9	0,3	14	0,467	0,030	0,167	0,0286	0,0286	0,3146
5		49	59	54	9	0,3	23	0,767	0,030	0,467	0,0305	0,0305	0,6262
6		59	69	64	5	0,167	28	0,933	0,017	0,767	0,0170	0,0170	0,8700
7		69	79	74	2	0,067	30	1	0,007	0,933	0,0050	0,0050	0,9733
8					30		1			1			
9													
10					=ЧАСТОТА(A2:A31;G2:G6)								
11	Средн. X	50		50									
12	Дисперсия	162,8	154,4										
13	Ср. кв. откл	12,8	12,4	10									

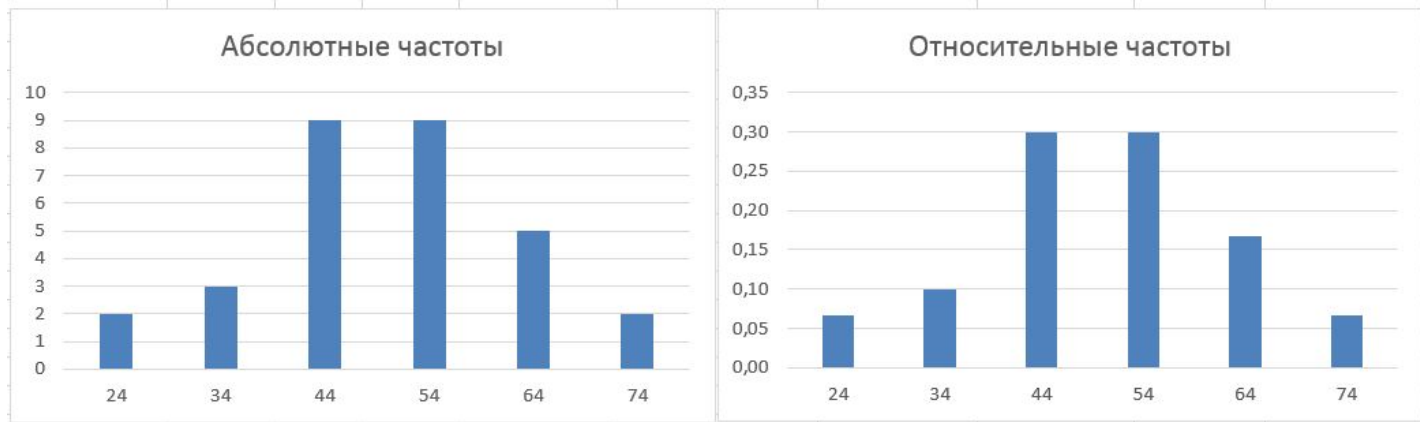
- Построить гистограмму с помощью Анализ данных\Гистограмма интервал карманов – как в функции ЧАСТОТА. (в примере: G2:G6); Установить флажок *Интегральный процент*

Карман	Частота	Интегральный %
29	2	6,67%
39	3	16,67%
49	9	46,67%
59	9	76,67%
69	5	93,33%
Еще	2	100,00%

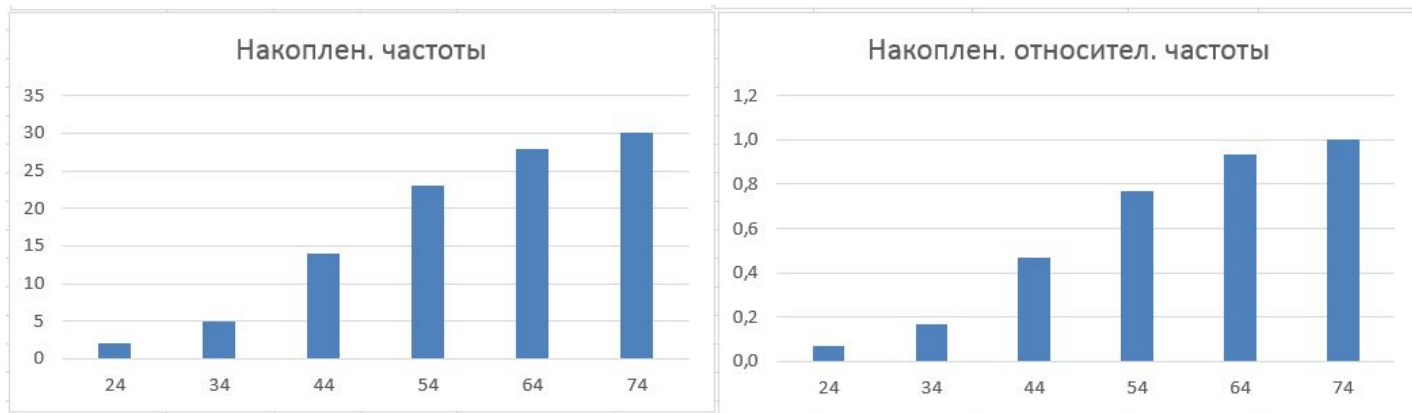


Построить гистограммы

- Абсолютных и относительных частот

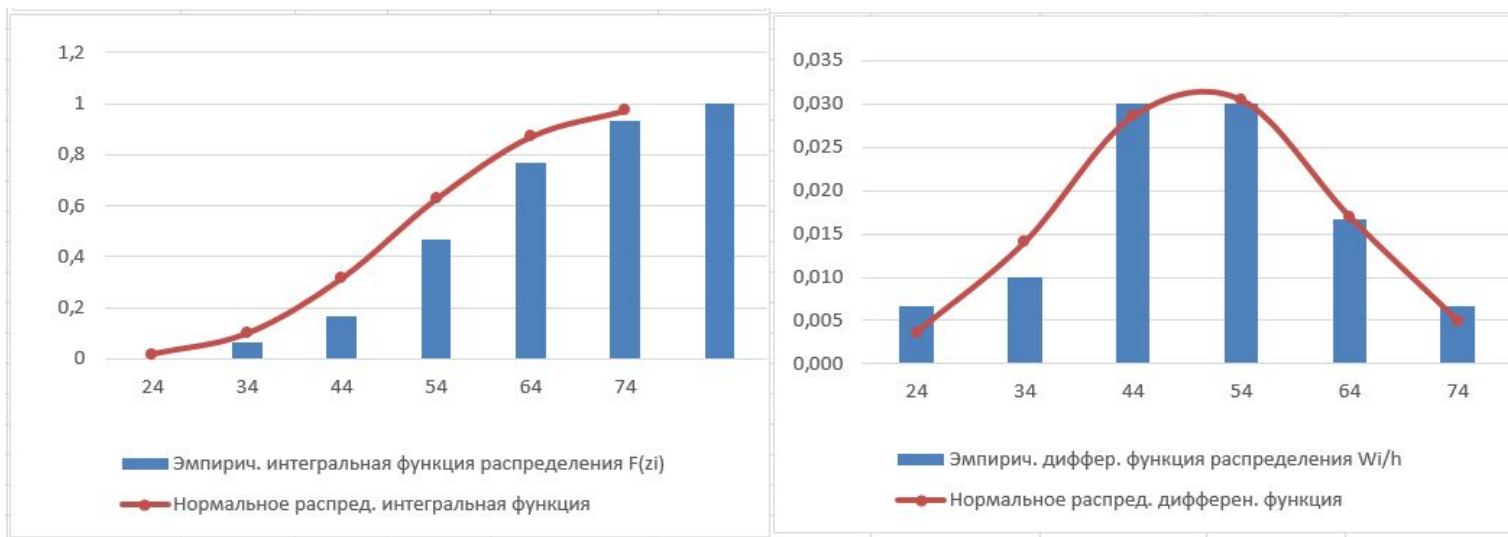


- Кумулятивных и относительных кумулятивных частот



Построить гистограммы

- Эмпирический дифференциальный и интегральный законы распределения
- Интегральную и дифференциальную функции для нормального закона распределения



Вычислить

❖ Основные описательные статистики (лабораторная работа №1)

При доверительной вероятности $\gamma = 0.95$ (уровне значимости $\alpha = 0.05$) найти

❖ Доверительный интервал для среднего значения

- по формуле

$$\Delta = \left(\bar{x} - t_{\alpha, N-1} \frac{s_x}{\sqrt{N}}, \bar{x} + t_{\alpha, N-1} \frac{s_x}{\sqrt{N}} \right)$$

$t_{\alpha, N-1}$ определяется с помощью функции **СТЮДЕНТ.ОБР.2Х()**

- с помощью функции **ДОВЕРИТ.СТЮДЕНТ()** – вычисляет $t_{\alpha, N-1} \frac{s_x}{\sqrt{N}}$
- «Анализ данных \Описательная статистика» – поставить флажок **Уровень надежности**

❖ Доверительный интервал для дисперсии

$$\Delta = \left(\frac{(N-1)s_x^2}{\chi_{1-\frac{\alpha}{2}, N-1}^2}, \frac{(N-1)s_x^2}{\chi_{\frac{\alpha}{2}, N-1}^2} \right)$$

$\chi_{\alpha, N-1}$ определяется с помощью функции **ХИ2.ОБР()**

❖ Доверительный интервал для среднего квадратичного отклонения

Лабораторная работа № 4

Линейная регрессия

Даны две выборки, состоящие из N значений случайных величин X и Y , x_1, x_2, \dots, x_N и y_1, y_2, \dots, y_N . X – независимая переменная, влияющая на Y . $y = (y_1, y_2, \dots, y_N)$ – отклик, $x = (x_1, x_2, \dots, x_N)$ – фактор, влияющий на отклик.

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

1. Построить уравнение регрессии

$$\tilde{y} = a_0 + a_1 x$$

Метод наименьших квадратов

$$I(a_0, a_1) = \sum_{i=1}^N (a_0 + a_1 x_i - y_i)^2 \rightarrow \min$$
$$\frac{\partial I}{\partial a_0} = 0, \quad \frac{\partial I}{\partial a_1} = 0$$

$$\begin{cases} \frac{\partial I}{\partial a_0} = 2 \sum_{i=1}^N (a_0 + a_1 x_i - y_i) = 0 \\ \frac{\partial I}{\partial a_1} = 2 \sum_{i=1}^N (a_0 + a_1 x_i - y_i) x_i = 0 \end{cases}$$

$$\begin{cases} a_0 \cdot \sum_{i=1}^N 1 + a_1 \cdot \sum_{i=1}^N x_i = \sum_{i=1}^N y_i \\ a_0 \cdot \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i \end{cases} \quad (*)$$



Метод наименьших квадратов

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{pmatrix}$$

$$\mathbf{C} \cdot \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \mathbf{D} \quad \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \mathbf{C}^{-1} \cdot \mathbf{D}$$



Вычислить коэффициенты регрессии

1. Методом наименьших квадратов, решив систему линейных уравнений (*)
2. С помощью функций НАКЛОН(), ОТРЕЗОК()
3. С помощью функции ЛИНЕЙН()
4. Построив линию тренда на диаграмме
5. С помощью «Анализ данных \ Регрессия»

Выполнение работы

- Ввести исходные данные на рабочий лист *Линейная регрессия*
- Для решения системы ввести матрицу коэффициентов и правую часть системы, найти значения коэффициентов регрессии – решить систему с помощью функций МУМНОЖ(), МОБР()

	A	B
1	x	y
2	22,8	73
3	14,4	46
4	23,7	74
5	23,1	72
6	18,6	58
7	18,1	63
8	15,6	51
9	19	61
10	22,7	80
11	18,8	68
12	18,1	63
13	14,8	55
14	23,1	80
15	21,8	62
16	23,8	79
17	23,1	69
18	14,9	42
19	18,2	55
20	14,5	51
21	17,5	61
22	15,4	45
23	16,1	58
24	19,4	68
25	23,6	73
26	22,8	74

	D	E	F	G	H	I	J	K
1	Решение системы линейных уравнений							
2	=СУММКВ(A2:A26)							
3								
4								
5								
6								
7								
8								
9								
10								
11								

а= 2,981987
b= 5,52066

=МУМНОЖ(МОБР(D5:E6);G5:G6)
(формула относится к двум ячейкам - операции с массивами)

выделить две ячейки, ввести формулу и нажать Ctrl+Shift+Enter

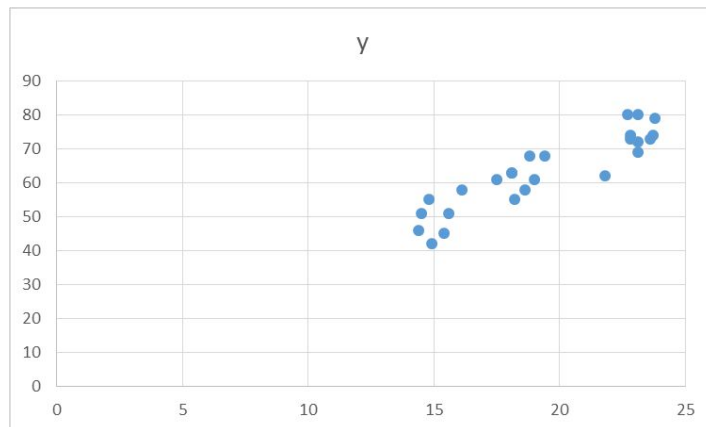
Функции рабочего листа НАКЛОН(), ОТРЕЗОК(), ЛИНЕЙН()

	D	E	F	G	H	I	J	K
11								
12								
13	НАКЛОН(), ОТРЕЗОК()							
14								
15		=НАКЛОН(B2:B26;A2:A26)			a=	2,981987		
16		=ОТРЕЗОК(B2:B26;A2:A26)			b=	5,52066		
17								
18	ЛИНЕЙН()							
19								
20								
21		=ЛИНЕЙН(B2:B26;A2:A26;;1)			a=	2,981987	5,52066	=b
22						0,280356	5,506402	
23						0,831049	4,671283	
24						113,1338	23	
25						2468,68	501,8803	

Выделить два столбца и пять строк, ввести формулу, нажать Ctrl+Shift+Enter

Построение диаграммы и линии тренда

- Выделить два столбца значений X и Y вместе с заголовками
- Вставить **точечную** диаграмму (Тип диаграммы – *Точечная*)



- Выделить на диаграмме ряд
- В контекстном меню выбрать *Добавить линию тренда*
- Изменить параметры линии тренда (см. след слайд): *Линейная* (установлено по умолчанию), *Показывать уравнение на диаграмме*, *Поместить на диаграмме величину достоверности аппроксимации R^2*

Построение диаграммы и линии тренда



Формат линии тренда

ПАРАМЕТРЫ ЛИНИИ ТRENDA

Экспоненциальная

Линейная

Логарифмическая

Полиномиальная Степень

Степенная

Линейная фильтрация Точки

Название аппроксимирующей (сглаженной) кривой

Автоматическое

Другое

Прогноз

Вперед на период

Назад на период

Пересечение кривой с осью Y в точке


показывать уравнение на диаграмме


поместить на диаграмму величину достоверности аппроксимации (R^2)

Анализ данных \ Регрессия

Регрессия

Входные данные


Входной интервал Y: 

Входной интервал X: 

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал: 

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK
Отмена
Справка

Анализ данных \ Регрессия

	M	N	O	P	Q	R	S	T	U
1	Анализ данных \ Регрессия								
2									
3	Вывод итогов								
4									
5	<i>Регрессионная статистика</i>								
6	Множественный R	0,911618661							
7	R-квадрат	0,831048583							
8	Нормированный R-квадрат	0,823702869							
9	Стандартная ошибка	4,671282854							
10	Наблюдения	25							
11									
12	Дисперсионный анализ								
13		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
14	Регрессия	1	2468,679679	2468,679679	113,1338096	2,35688E-10			
15	Остаток	23	501,8803206	21,8208835					
16	Итого	24	2970,56						
17									
18		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
19	Y-пересечение	5,520659776	5,506401949	1,002589318	0,326491069	-5,870200522	16,91152007	-5,870200522	16,91152007
20	x	2,981986992	0,280355796	10,63643782	2,35688E-10	2,402026841	3,561947143	2,402026841	3,561947143

**Проверить, чтобы значения
коэффициентов регрессии, найденные
разными способами, совпадали!**

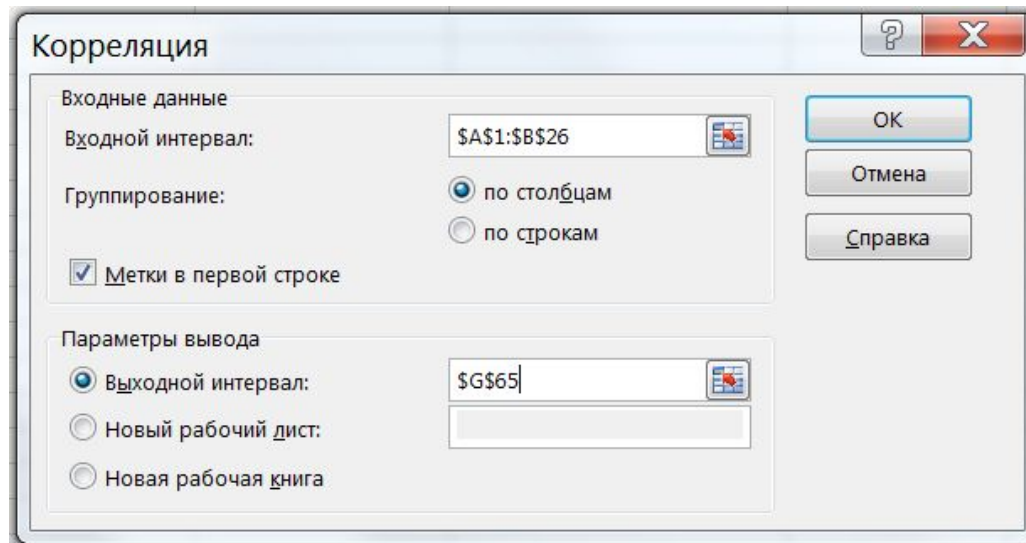
2. Вычислить коэффициент корреляции

- Выборочная ковариация (функция КОВАРИАЦИЯ.В())

$$\text{cov}_{YX}^* = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

- Коэффициент корреляции (функция КОРРЕЛ(), Анализ данных \ Корреляция)

$$r_{YX}^* = \frac{1}{N-1} \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{\text{cov}_{YX}^*}{s_x s_y}$$



Вычисление коэффициента корреляции

	D	E	F	G	H	I	J	K
51								
52								
53	Коэффициент корреляции							
54	N=	25						
55								
56	Средн. X	19,356	Станд.отклон X	3,401113				
57	Средн. Y	63,24	Станд.отклон Y	11,12535				
58								
59					Формула	Функция		
60			Ковариация	34,49433	34,49433			
61			Коэффициент корреляции	0,911619	0,911619			
62								
63			Анализ данных \ Корреляция					
64								
65				x	y			
66			x	1				
67			y	0,911619	1			
68								

3. Оценить уровень значимости уравнения регрессии

- Коэффициент детерминации
$$R^2 = \frac{\sum_{i=1}^N (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^N (y_i - \tilde{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

В случае одной независимой переменной X
$$R^2 = (r_{YX}^*)^2$$

- F-статистика
$$F = \frac{R^2 / (m - 1)}{(1 - R^2) / (N - m)}$$
 $F > F_{k_1, k_2}^\alpha$
 m – число параметров в уравнении регрессии, (a_0, a_1) , $m = 2$ $F > 4F_{k_1, k_2}^\alpha$

(сравниваем с F.ОБР.ПХ($\alpha = 0,05$; $k_1 = m - 1$; $k_2 = N - m = 23$) = 4,28

	D	E	F	G	H	I	J	K
77								
78								
79	Уровень значимости уравнения регрессии							
80								
81		0,831049	= r_{YX}^* ²					
82		0,831049	По формуле					
83								
84		113,1338	F-статистика					

Вычислить

- Сумма квадратов, обусловленная регрессией (RSS)

$$Q_R = \sum_{i=1}^N (\tilde{y}_i - \bar{y})^2$$

- Сумма квадратов, обусловленная ошибкой (ESS)

$$Q_e = \sum_{i=1}^N (y_i - \tilde{y}_i)^2$$

- Общая сумма квадратов (TSS)

$$Q = \sum_{i=1}^N (y_i - \bar{y})^2 \quad Q = Q_R + Q_e$$

- Фактическая дисперсия

$$D_{\text{факт}} = \frac{Q_R}{m-1}$$

- Остаточная дисперсия

$$D_{\text{ост}} = \frac{Q_e}{N-m}$$

$$R^2 = 1 - \frac{Q_e}{Q} = \frac{Q_R}{Q}$$

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)}{N/N - m}$$

$$F = \frac{Q_R}{Q_e} \cdot \frac{N-m}{m-1}$$

4. Оценить значимость коэффициентов уравнения регрессии

- Несмещенная оценка дисперсии остаточной компоненты

$$s_e^2 = \frac{\sum_{i=1}^N (y_i - \tilde{y}_i)^2}{N - m} = D_{\text{ост}}$$

- Стандартная ошибка оценки s_e
- Стандартные ошибки коэффициентов уравнения регрессии

$$s_1 = s_e \sqrt{d_{11}}$$

$$s_0 = s_e \sqrt{d_{22}}$$

d_{jj} - диагональные элементы матрицы C^{-1} (обратной матрицы системы уравнений (*))

- t-статистика $t_1 = \frac{a_1}{s_1}$ $t_0 = \frac{a_0}{s_0} > t_{\alpha, N-m}$
СТЮДЕНТ.ОБР.2Х($\alpha = 0,05$; $N - m = 23$) =
2,069

- Доверительный интервал
 $(a_1 - t_{\alpha, N-m} s_1; a_1 + t_{\alpha, N-m} s_1)$
 $(a_0 - t_{\alpha, N-m} s_0; a_0 + t_{\alpha, N-m} s_0)$

5. Оценить значимость коэффициента корреляции

$$t = R \sqrt{\frac{N - m}{1 - R^2}} > t_{\alpha, N - m}$$

5. Найти доверительные интервалы для среднего и дисперсии

Лабораторная работа № 5

Линейная регрессия для сгруппированных данных

Даны две выборки, состоящие из K значений случайной величины X, x_1, x_2, \dots, x_K и M значений случайной величины Y, y_1, y_2, \dots, y_M . Каждая пара значений (x_j, y_i) встречается n_{ji} раз.

$X \backslash Y$	y_1	y_2	...	y_i	...	y_M	P
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1M}	P_1
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2M}	P_2
...
x_j	n_{j1}	n_{j2}	...	n_{ji}	...	n_{jM}	P_j
...
x_K	n_{K1}	n_{K2}	...	n_{Kj}	...	n_{KM}	P_K
Q	Q_1	Q_2	...	Q_i	...	Q_M	N

1. Построить уравнение регрессии $\tilde{y} = a_0 + a_1x$

Метод наименьших квадратов

$$\begin{cases} a_0 \cdot N + a_1 \cdot \sum_{j=1}^K x_j P_j = \sum_{i=1}^M y_i Q_i \\ a_0 \cdot \sum_{j=1}^K x_j P_j + a_1 \cdot \sum_{j=1}^K x_j^2 P_j = \sum_{j=1}^K \sum_{i=1}^M x_j y_i n_{ji} \end{cases}$$

Выполнение работы

- Ввести исходные данные на рабочий лист *Линейная регрессия* – *сгруппированные данные*
- Для решения системы ввести матрицу коэффициентов и правую часть системы, найти значения коэффициентов регрессии – решить систему с помощью функций МУМНОЖ(), МОБР()

	A	B	C	D	E	F	G	H	I
1	X\Y	1,2	1,3	1,4	1,5	1,6	Q		
2	30		2		8		10	438	
3	40		4	2	9		15	860	
4	50	2	12	3	1		18	1185	
5	60	21	14				35	2604	
6	70	1					1	84	
7	P	24	32	5	18	0	79	5171	
8									
9							=СУММ(H2:H6)		
10									
11									=A2*СУММПРОИЗВ(\$B\$1:\$F\$1;B2:F2)

Вспомогательный столбец H для расчета

$$\left\{ \sum_{j=1}^K \sum_{i=1}^M x_j y_i n_{ji} \right.$$

Выполнение работы

- Сделать проверку $\tilde{y} = a_0 + a_1x$

	K	L	M	N	O	P
1						
2						
3	=G7	=СУММПРОИЗВ(A2:A6;G2:G6)				
4	↓	↓				
5			=СУММПРОИЗВ(B1:F1;B7:F7)			
6	↓	↓	↓			
7	79	3970	104,4	$a_0 =$	1,725	
8	3970	208900	5171	$a_1 =$	-0,008	
9	↑	↑	↑			
10	=L7		=H7			
11		=СУММПРОИЗВ(A2:A6^2;G2:G6)				
12	Проверка					
13	x	\hat{y}				
14	30	1,484				
15	40	1,404				
16	50	1,324				
17	60	1,243				
18	70	1,163				

2. Построить уравнение регрессии

$$\tilde{x} = b_0 + b_1 y$$

$$\begin{cases} b_0 \cdot N + b_1 \cdot \sum_{i=1}^M y_i Q_i = \sum_{j=1}^K x_j P_j \\ b_0 \cdot \sum_{i=1}^M y_i Q_i + b_1 \cdot \sum_{i=1}^M y_i^2 Q_i = \sum_{j=1}^K \sum_{i=1}^M x_j y_i n_{ji} \end{cases}$$

Сделать самостоятельно по аналогии с п. I



3. Вычислить коэффициент корреляции

- Выборочная ковариация

$$\text{cov}_{YX}^* = \frac{1}{N-1} \sum_{j=1}^K \sum_{i=1}^M (x_j - \bar{x})(y_i - \bar{y})n_{ij}$$

- Коэффициент корреляции

$$r_{YX}^* = \frac{1}{N-1} \frac{\sum_{j=1}^K \sum_{i=1}^M (x_j - \bar{x})(y_i - \bar{y})n_{ij}}{s_x s_y} = \frac{\text{cov}_{YX}^*}{s_x s_y}$$

Примечания

- При вычислении средних значений и средних квадратичных отклонений использовать формулы для сгруппированных данных (функции СРЗНАЧ() и СТАНДОТКЛОН.В() не подходят)
- Ковариацию вычислить, используя вспомогательный столбец

Вычислить

- Сумма квадратов, обусловленная регрессией (RSS)

$$RSS = \sum_{j=1}^K \sum_{i=1}^M (\tilde{y}_i - \bar{y})^2 n_{ij}$$

- Сумма квадратов, обусловленная ошибкой (ESS)

$$ESS = \sum_{j=1}^K \sum_{i=1}^M (y_i - \tilde{y}_i)^2 n_{ij}$$

- Общая сумма квадратов (TSS)

$$TSS = \sum_{j=1}^K \sum_{i=1}^M (y_i - \bar{y})^2 n_{ij} \quad TSS = RSS + ESS$$

- Фактическая дисперсия

$$D_{\text{факт}} = \frac{RSS}{m-1}$$

- Остаточная дисперсия

$$D_{\text{ост}} = \frac{ESS}{N-m}$$

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$$

$$F = \frac{RSS}{ESS} \cdot \frac{N-m}{m-1}$$

$$R^2_{\text{adj}} = 1 - \frac{(1-R^2)}{N/N-m}$$

$$t, s_e, s_0, s_1, t_0, t_1$$

	A	B	C	D	E	F
13	RSS	0,6056	=СУММ(B15:F19)			
14			=(L14-\$L\$23)^2*B2			
15		0	0,053	0	0,212	0
16		0	0,027	0,0136	0,061	0
17		8E-06	5E-05	1E-05	4E-06	0
18		0,1286	0,086	0	0	0
19		0,0251	0	0	0	0
20						
21						
22	ESS	0,3678				
23			=(B\$1-\$L14)^2*B2			
24		0	0,068	0	0,002	0
25		0	0,043	3E-05	0,083	0
26		0,0305	0,007	0,0175	0,031	0
27		0,0393	0,045	0	0	0
28		0,0014	0	0	0	0
29						
30						
31	TSS	0,9734		0,9734		
32		0	9E-04	0	0,255	0
33		0	0,002	0,0123	0,287	0
34		0,0295	0,006	0,0185	0,032	0
35		0,3101	0,006	0	0	0
36		0,0148	0	0	0	0

	K	L
12	Проверка	
13	x	y
14	30	1,484
15	40	1,404
16	50	1,324
17	60	1,243
18	70	1,163
19		
20		
21		
22	Средн. X	50,253
23	Средн. Y	1,3215

	A	B	C	D
37				
38	Dфакт	0,6056		
39	Дост	0,0048		
40	F	126,79		
41	R ²	0,6222		
42	R	0,7888	r _{xy}	
43	t	11,26		
44	s _e	0,0691		
45	s ₀	0,0367	s ₁	0,0007
46	t ₀	47,046	t ₁	-11,26

Лабораторная работа № 6

Множественная линейная регрессия

Даны три выборки, состоящие из N значений случайных величин X и Y и Z , x_1, x_2, \dots, x_N ; y_1, y_2, \dots, y_N и z_1, z_2, \dots, z_N .

1. Построить уравнение регрессии

$$\tilde{z} = a_0 + a_1x + a_2y$$

Метод наименьших квадратов

$$I(a_0, a_1, a_2) = \sum_{i=1}^N (a_0 + a_1x_i + a_2y_i - z_i)^2 \rightarrow \min$$

$$\frac{\partial I}{\partial a_0} = 0, \quad \frac{\partial I}{\partial a_1} = 0, \quad \frac{\partial I}{\partial a_2} = 0$$

$$\begin{cases} \frac{\partial I}{\partial a_0} = 2 \sum_{i=1}^N (a_0 + a_1x_i + a_2y_i - z_i) = 0 \\ \frac{\partial I}{\partial a_1} = 2 \sum_{i=1}^N (a_0 + a_1x_i + a_2y_i - z_i)x_i = 0 \\ \frac{\partial I}{\partial a_2} = 2 \sum_{i=1}^N (a_0 + a_1x_i + a_2y_i - z_i)y_i = 0 \end{cases} \quad \begin{cases} a_0 \cdot \sum_{i=1}^N 1 + a_1 \cdot \sum_{i=1}^N x_i + a_2 \cdot \sum_{i=1}^N y_i = \sum_{i=1}^N z_i \\ a_0 \cdot \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2 + a_2 \sum_{i=1}^N y_i x_i = \sum_{i=1}^N x_i z_i \\ a_0 \cdot \sum_{i=1}^N y_i + a_1 \sum_{i=1}^N x_i y_i + a_2 \sum_{i=1}^N y_i^2 = \sum_{i=1}^N y_i z_i \end{cases} \quad (*)$$

Выполнение работы

- Ввести исходные данные на рабочий лист *Множественная линейная регрессия*
- Для решения системы ввести матрицу коэффициентов и правую часть системы, найти значения коэффициентов регрессии – решить систему с помощью функций МУМНОЖ(), МОБР()

	F	G	H	I	J	K	L
1		N= 25		m= 3			
2							
3		Решение системы линейных уравнений					
4		25	483,9	166,3	2124	a ₀ =	10,63
5		483,9	9644	3232,5	41970	a ₁ =	2,967
6		166,3	3232	1133,7	14239	a ₂ =	2,541
7							

- Вычислить коэффициенты регрессии при помощи функции ЛИНЕЙН()

	F	G	H	I
7				
8		Функция ЛИНЕЙН		
9		2,541	2,967	10,634
10		0,896	0,282	7,4854
11		0,857	4,638	
12		65,66	22	
13		2826	473,3	
..				

Анализ данных \ Регрессия

	N	O	P	Q	R	S	T
4	ВЫВОД ИТОГОВ						
5							
6	<i>Регрессионная статистика</i>						
7	Множественный R	0,92548237					
8	R-квадрат	0,856517617					
9	Нормированный R-квадрат	0,843473764					
10	Стандартная ошибка	4,6384884					
11	Наблюдения	25					
12							
13	<i>Дисперсионный анализ</i>						
14		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
15	Регрессия	2	2825,617358	1412,808679	65,66446412	5,30621E-10	
16	Остаток	22	473,3426421	21,51557464			
17	Итого	24	3298,96				
18							
19		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
20	Y-пересечение	10,63367714	7,485367974	1,420595111	0,169453123	-4,890025828	26,15738012
21	x	2,966851342	0,281815353	10,5276427	4,69822E-10	2,382402075	3,55130061
22	y	2,540581521	0,896028928	2,835378904	0,009628482	0,682331267	4,398831775

Анализ данных \ Корреляция

	G	H	I	J
21				
22	Анализ данных \ Корреляция			
23				
24		x	y	z
25	x	1		
26	y	0,155	1	
27	z	0,897	0,3656	1

Вычислить

$Q_R(RSS), Q_E(ESS), Q(TSS),$

$D_{\text{факт}}, D_{\text{ост}}$

$F, R^2, R, R_{\text{adj}}$

$s_e, s_0, s_1, s_2, t_0, t_1, t_2$

	G	H	I	J	K	L	M
29							
30	Q_R	RSS	2825,6				
31	Q_E	ESS	473,34				
32	Q	TSS	3299	3299			
33		$D_{\text{факт}}$	1412,8				
34		$D_{\text{ост}}$	21,516				
35		F	65,664				
36		R^2	0,8565				
37		R	0,9255				
38		R^2_{adj}	0,8737				
39		s_e	4,6385				
40		s_0	7,4854	s_1	0,282	s_2	0,896
41		t_0	1,4206	t_1	10,53	t_2	2,835

Ф.ОБР.ПХ($\alpha = 0,05; k_1 = m - 1; k_2 = N - m = 22$) = 3,44

СТЮДЕНТ.ОБР.2Х($\alpha = 0,05; N - m = 22$) = 2,074