ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М.А. БОНЧ- БРУЕВИЧА

С.В.Протасеня

Методические указания к лабораторным работам «Теория анализа биологических сигналов»

САНКТ-ПЕТЕРБУРГ
2016

Лабораторная работа №1 исследование статистических функций.

При статических исследованиях широко используются специальные функции закона нормального распределения, распределений хи-квадрат, Стьюдента и Фишера. Получим графики этих функций и исследуем их свойства.

Пример. Построим график функции плотности нормального распределения и исследуем влияние на него параметров *m* и σ

Решение. Запускаем программу EXCEL и задаем значения параметров *m*. Пусть, например,σ = 1 и *m*=3. Для этого в ячейки A1и A2 первого листа вводим подписи «m=» и «sig=» (кавычки здесь и далее вводить не надо), а в соседние B1 и B2 вводим значения 3 и 1.

Для построения графика протабулируем в столбцах C и D функцию плотности нормального распределения на отрезке (0;6) с шагом 0,2. Для этого вводим в C1 подпись «X=», а в D1 подпись «f=». Вводим в C2 значение 0,в C3 значение 0,2, обводим, выделяя, ячейки C2 и C3 и захватив за нижний правый угол рамки вокруг ячеек C2 и C3, перетягиваем его вниз до ячейки C32, что позволит автоматически занести в столбец значения от 0 до 6 с шагом 0,2. Ставим курсор в ячейку D2 и вызываем функцию плотности нормального распределения. Для этого нажимаем кнопку мастера функций *fx*, выбираем категорию «**Статистические**» и функцию **НОРМРАСП (NORMDIST)**.

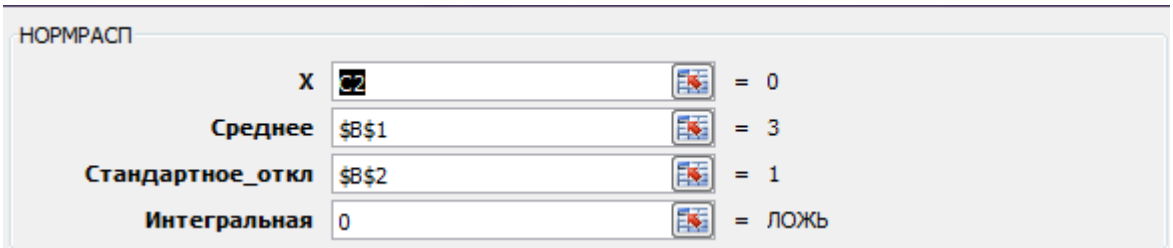

рис.1.1

Появляется окно, показанное на рисунке рис.1.1.

Вводим ссылкой на переменную X: «C2» (для ввода ссылки достаточно щелкнуть мышью по ячейке с данной адресацией), ссылкой на *m* - «B1» и «B2». Эти ссылки

абсолютные, т.к. ячейки со значениями σ и *m* и s всегда B1 и B2, поэтому пишется знак $ (чтобы быстро относительную ссылку сделать абсолютной в EXCEL нужно после ввода ссылки нажать F4(вCALC:SHIFT+F4).

Строим график по данным. Ставим курсор в любой свободной ячейке. Вызываем мастер диаграмм, выбрав пункты меню ВСТАВКА/ДИАГРАММА. Выбираем тип диаграммы «График» и вид– левый график в верхнем ряду, нажимаем «Далее». Ставим курсор в поле «Диапазон» и обводим курсором ячейки D2-D32, переходим на закладку «Ряд», ставим курсор в поле «Подписи оси X» и обводим диапазон данных C2-C32, нажимаем «Готово». Получаем график плотности нормального распределения.

Исследуем, как влияют параметры на вид графика. Для этого изменяем в ячейке B1 значение 3 на значение 4, нажимаем Enter. Видим, что график сместился вправо, изменяем на 2, график сместился влево. Возвращаем в B1 значение 3, и изменяем в B2 значение 1 на 2. График растянулся. Изменяем в B2 на 0,5 – график сжался.

Делаем вывод: параметр *m* влияет на ширину графика, с увеличением параметра график растягивается. σ изменяет положение графика, с увеличением параметра график смещается вправо. Параметр

Рассмотрим теперь другие виды законов распределений.

1. **Распределение хи – квадрат** определяется как сумма *k* независимых стандартных нормальных величин. Число *k* называется числом степеней свободы. Когда *k* =1 случайная величина равна квадрату стандартной нормальной величины. Хи – квадрат распределение имеет только один параметр – число степеней свободы *k,* являющийся целым положительным числом. Функция, возвращающая значение плотности распределения хи-квадрат находится в категории «**Статистические**» и называется «**ХИ2РАСП (CHIDIST)**».
2. **t - распределение Стьюдента** важно в тех случаях, когда рассматриваются оценки среднего, оценки коэффициентов регрессионного уравнения, оценки параметров временных рядов. Распределение Стьюдента с единственным параметром *k,* называемым степенью свободы сосредоточено на всей действительной оси, симметрично относительно начала координат(см. рис.) при $k \to \infty$ t-распределение приближается к нормальному. Функция, возвращающая значение плотности распределения Стьюдента находится в категории

«**Статистические**» и называется«**СТЬЮДРАСП(TDIST**)». Функция имеет дополнительный чисто вычислительный параметр «Хвосты», который не связан с распределением Стьюдента, а связан с выводом полученных результатов программой EXCEL. Его всегда задаем равным 1.

3. **F- распределение Фишера** возникает в регрессионном, дисперсионном, дискриминантном анализе, а также в других видах многомерного анализа данных. Случайная величина, имеющая F- распределение с парой степеней свободы *m* и *n***,** определяется как отношение двух независимых случайных величин, имеющих распределение хи – квадрат со степенями свободы *m* и *n* с умножением на нормированный сомножитель *n/m***.** F- распределение сосредоточено на положительной полуоси. Это распределение несимметрично. Функция, возвращающая значение плотности распределения Фишера находится в категории «**Статистические**» и называется «**FРАСП(FDIST**)». Она имеет два параметра *m* и *n*, называемых степенями свободы.

Задание к лабораторной работе №1.

1. Изучить материал п.1, п.2, п.7.
2. Создать рабочую книгу Excel «ФИО_ статистика».
3. Первый лист книги переименовать в «Графики функций» и выполнить все задания лабораторной работы на этом листе.
4. Построить график функции плотности нормального распределения (см. пример);
5. Построить графики функций плотности распределения (см. свой вариант задания в таблице):

1. хи-квадрат,
2. Стьюдента;
3. Фишера.

6. Проанализировать влияние параметров распределения на графики функций. Вывод записать под графиком соответствующей функции.

В таблице приведены варианты (верхняя строка),значения границ интервала *(a, b),* шага табуляции *h*и параметров *k, m, n* (первый столбец).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
a	0	0	0	0	0	0	0	0	0	0	0	0
b	10	12	9	8	10	14	11	12	14	10	14	8
h	0.2	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.2	0.1
k	5	4	5	4	6	5	6	4	5	6	4	5
m	4	5	1	5	4	6	4	5	6	4	5	4
n	5	15	6	30	7	7	6	10	7	6	12	6

Контрольные вопросы:

1. Зачем статистика врачу?
2. От каких параметров зависит функция плотности нормального распределения? Каков их смысл.
3. Как влияет изменение параметров m и σ на кривую Гаусса.
4. С помощью каких функций в Excel вычисляют значения функций плотности распределений Хи-квадрат, Фишера, t-Стьюдента.

Лабораторная работа №2 статистические методы обработки данных.

*Множество мыслимых объектов изучаемого явления, называется **генеральной совокупностью.** Часть объектов, отобранных из генеральной совокупности по определенным правилам, называется **выборкой,** или **выборочной совокупностью**.*

Дадим другое определение выборки: ***выборкой объема n** называются числах$_1$, x$_2$,…, x$_n$(называемые вариантами), получаемые на практике при n– кратном повторении эксперимента в неизменных условиях*. На практике выборку чаще всего представляют

статистическим рядом. ***Статистическим дискретным рядом*** *называется ранжированный в порядке возрастания (убывания) ряд вариантов с соответствующими им частотами (частостями).*

(10-20). Если же количество различных вариант существенно больше, то результаты представляют в виде ***статистического интервального ряда распределения***. Для построения такого ряда всю область наблюдаемый значений изучаемого признака X разбивают на *m* интервалов (см. п 1.5), вычисляют середины интервалов z_i, и считают число элементов выборки, попадающих в каждый интервал n_i.

Числа n_i, показывающие, сколько раз встречаются варианты из данного интервала в выборке, называются ***частотами,*** *а их отношение к общему числу наблюдений n_i/n –* ***частостями,*** *или* ***относительными частотами.***

Полигон *служит для отображения дискретного статистического ряда и представляет собой ломаную, в которой концы отрезков имеют координаты (z_i , n_i).*

Гистограмма *представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака $m_i=x_{i+1}-x_i$ и высотами равными частотам n_i.*Гистограмма служит для визуального представления только интервальных вариационных рядов.

Накопленные частоты *для каждого интервала находятся последовательным суммированием частот (частостей) всех предшествующих интервалов, включая данный.*

Кумулятивная кривая (кумулята) *– кривая накопленных частот.*

Рассмотрим решение задачи первичной обработки статистических данных в программе EXCEL на следующем примере.

ПРИМЕР. Дана выборка числа посетителей некоторого специалиста поликлиники за 25 дней.

14, 18, 16, 21, 12, 19, 27, 19, 15, 20, 27, 29, 22, 28, 19, 17, 18, 24,23, 22, 19, 20, 23, 21, 19.

Построим статистический ряд, полигон, гистограмму и кумулятивную кривую.

Решение.

Откроем книгу программы EXCEL. Введем в первый столбец (ячейки A1-A25) исходные данные. Определим область чисел, в которой лежат данные. Для этого найдем максимальный и минимальный элементы выборки. Введем в B1подпись «Максимум», а в B2 -подпись «Минимум». В соседних ячейках C1и C2 определим функции «MAX» и «MIN». Для этого ставим курсор в C1и вызываем мастер функций, нажав на кнопку *fx*, в открывшемся окне в поле «Категория» выбираем «Статистические», и ниже ищем функцию МАКС(MAX) и вызываем ее двойным щелчком мыши по названию. В качестве аргумента функции (в графе «Число 1») обведем область данных (ячейки A1-A25).Поле «Число 2» оставляем пустым. Нажимаем «ОК». Результатом будет число 29. Ставим курсор в ячейку C2 и аналогично вводим функцию МИН(MIN). Результат – число 12.

Все числа попадают в интервал [12,29], но для удобства нахождения шага интервала расширим размах, пусть X_{MAX}=30. Размах вариации R=X_{MAX}–X_{MIN}= 30-12=18. Согласно таблице (п.1.5.) разделим его шесть интервалов, найдем шаг: h=18/6=3, получим интервалы:

12-15, 15-18, 18-21, 21-24, 24-27, 27-30.

	A	B	C	D	E	F	G	H	I
1	14			нижняя гр	верхняя гр	середина интерв	частоты	накоплен част	частости
2	18	максимум	29	12	15				
3	16	минимум	12	15	18				
4	21			18	21				
5	12			21	24				
6	19			24	27				
7	27			27	30				

В ячейки D2-D9 вводим нижние границы интервалов группировки – числа 12, 15, 18, 21, 24, 27. В ячейки E2-E7 вводим верхние границы интервалов группировки – числа 15, 18, 21, 24, 27, 30. В столбце F вычислим середины интервалов: для этого в ячейкуF2 ставим курсор и вызываем функцию из категории статистические СРЗНАЧ(AVERAGE), в качестве аргументов функции задаем диапазон D2:E2. Выделяем ячейку F2 и мышкой, удерживая правый нижний угол, протаскиваем до ячейки F7, тем самым копируем введенную формулу на весь диапазон F2-F7.

Для вычисления частот *ni* используют функцию ЧАСТОТА(FREQUENCY), находящуюся в категории Статистические (Массив)». Введем ее в ячейку G2. В строке «Массив данных» введем диапазон выборки (ячейки A1-A25). В строке «Массив интервалов» введем диапазон верхних границ интервалов группировки (ячейки E2-E7). Результат функции является массивом и выводится в ячейках G2-G7. В Excel для полного вывода (не только первого числа в G2) нужно выделить ячейки G2-G7, обведя их мышью, и нажать F2, а далее одновременно CTRL+SHIFT+ENTER. Результат – частоты интервалов3,4,9,5,2,2.

Для построения гистограммы нужно выбрать ВСТАВКА/ДИАГРАММА или нажать на соответствующий значок на основной панели (при этом курсор должен стоять в свободной ячейке). Далее выбрать тип: ГИСТОГРАММА, вид по выбору, на основной панели ВЫБРАТЬ ДАННЫЕ, Появиться диалоговое окно ВЫБОР ИСТОЧНИКА ДАННЫХ, в строке ДИАПАЗОН обвести частоты G2-G7, перейти в окно ПОДПИСИ ГОРИЗОНТАЛЬНОЙ ОСИ, нажать ИЗМЕНИТЬ и обвести диапазон E2-E7 и ОК. Ввести название«ГИСТОГРАММА», подписи осей: ось X - «ИНТЕРВАЛЫ» и ось Y -«ЧАСТОТА», нажать ГОТОВО.

Для построения полигона перейти на пустую ячейку, выбрать ВСТАВКА/ДИАГРАММА /ГРАФИК.В строке ДИАПАЗОН обвести частоты G2-G7, перейти в окно ПОДПИСИ ГОРИЗОНТАЛЬНОЙ ОСИ и ввести середины интервалов F2-F7. Далее ввести название «ПОЛИГОН» и подписи по осям.

Для построения кумулятивной кривой нужно посчитать накопленные частоты. Для этого в ячейкуH2 вводим «**=G2**», в H3 – вводим «**=H2+G3**»

и автозаполнением перетаскиваем эту ячейку до F9. Далее строим график как и в случае полигона, но в строке ДИАПАЗОН вводим накопленные частоты, ссылаясь на H2-H7, а в окне ПОДПИСИ ГОРИЗОНТАЛЬНО ЙОСИ вводим диапазонF2-F7.

Для подсчета частостей (относительных частот) в ячейку I2 вводим «==G2/25», автозаполняем до I9.Для построения гистограммы и полигона в качестве ряда можно использовать частости.

Задание к лабораторной работе №2.

1. Изучить п.1, п.3, п.4, п.6.
2. По данным выборки построить интервальный ряд, полигон, гистограмму и кумулятивную кривую (первый столбец – номер варианта). В качестве ряда для построения гистограммы и полигона использовать частости.
3. По виду гистограммы (полигона) сделать предварительное предположение о виде распределения изучаемого признака.

	Выборка														
1.	46	64	61	63	58	59	61	63	71	72	49	59	57	59	56
	57	60	59	56	59	49	65	54	55	62	63	61	64	57	66
2.	13,4	14,7	15,2	15,1	13,0	8,8	14.0	17.9	15.1	16.5	16.6	14.2	16.3	14.6	11.7
	16.4	15.1	17.6	14.1	18.8	11.6	13.9	18.0	12.4	17.2	14.5	16.3	13.7	15.5	16.2

3.	119	110	134	119	120	118	115	121	123	118	117	112	112	106	104
	105	110	117	114	117	115	113	116	117	133	119	112	128	118	111
4.	23	27	56	45	27	38	64	55	49	34	44	58	61	69	70
	56	20	35	33	48	49	57	51	58	36	65	48	70	55	35
5.	23.7	4.4	41.4	19.2	23.2	24.5	31.0	33.3	22.2	34.0	33.5	17.5	29.7	25.1	21.1
	20.7	15.8	22.1	30.0	33.0	25.8	19.9	25.9	14.1	18.4	15.9	26.4	25.3	22.9	14.2
6.	16	17	19	18	28	30	24	19	17	17	18	21	23	25	22
	20	17	16	20	33	28	25	22	23	18	17	26	28	29	27
7.	10.8	7.0	13.0	7.6	15.0	14.1	11.2	13.9	8.6	12.3	12.8	16.7	11.7	11.3	11.7
	16.9	19.0	11.6	11.4	12.4	9.4	10.5	6.8	11.1	16.8	11.5	8.7	14.7	6.2	10.3
8.	10	12	17	22	33	18	15	17	17	25	20	27	19	25	30
	18	24	20	26	29	22	23	19	20	22	15	13	18	14	17
9.	74	64	82	60	68	72	71	71	69	70	69	72	61	66	80
	73	67	76	58	63	62	68	70	74	63	73	64	77	57	73
10.	1.7	4.1	4.3	2.6	0.9	0.8	1.2	2.1	3.2	2.9	1.1	3.2	4.5	2.1	3.1
	5.1	1.1	1.9	0.9	3.1	0.9	3.1	3.3	2.8	2.8	2.5	4.0	4.3	1.1	2.1
11.	32	23	35	35	36	34	41	33	31	27	36	47	31	33	30
	37	45	37	44	51	32	38	35	32	24	28	26	14	28	27
12.	112	103	101	98	100	97	98	100	98	107	108	99	98	92	98
	110	106	105	102	100	101	100	95	100	105	100	102	102	99	97

Лабораторная работа № 3 точечное и интервальное оценивание параметров распределений

Для исследования основных свойств явления или объекта, представленного выборкой вычисляют точечные и интервальные оценки.

Часть 1. Точечное оценивание.

Точечные оценки *параметров распределения это оценки одним числом*, *полученные по выборке и приближенно равные оцениваемым параметрам*.

Основными точечными оценками являются:

объем выборки n*– количество элементов в выборке.*

Выборочное среднее x*– оценка математического ожидания,*

Среднее арифметическое элементов выборки.

Выборочная дисперсия S^2 *– среднее квадратов отклонения элементов выборки от выборочного среднего, является оценкой дисперсии, характеризует разброс выборочных значений.*

Стандартное отклонение S*– корень из дисперсии.*

Медиана Me*– средний элемент вариационного ряда или полусумма двух средних элементов, если объем выборки четный.*

Мода Mo*– наиболее часто повторяющаяся варианта в выборке.*

Коэффициент эксцесса E *- характеризует «островерхость» гистограммы*

или полигона по сравнению с кривой Гаусса нормального

распределения.

Коэффициент асимметрии A *- характеризует степень симметричности гистограммы или полигона.*

Процентил*ь на уровне p- значение t p , меньше которого*

p ×100% элементов выборки.

ПРИМЕР. Из продукции произведенной фармацевтической фабрикой за месяц, случайным образом отобраны 25 коробочек некоторого препарата, количество таблеток в которых оказалось равным соответственно 50, 51, 48, 52, 50, 51, 49 50, 47, 50, 51, 49, 50, 48, 51, 50, 49, 50, 52, 49, 50, 48, 49, 50, 51.

Найти основные числовые характеристики выборки.

Решение.

Запускаем программу EXCEL, первый лист. Вводим исходные

данные в ячейки A1-A25. Находим числовые характеристики. Для ввода функций выделяем два столбца, например B и C, в первом вводим

название характеристики, во втором – функцию. В ячейки B1-B11 вводим подписи числовых характеристик, то есть вписываем в эти ячейки

первый столбец таблицы приведенной ниже. В C1 вводим текст «Функция» и ниже определяем функции, соответствующие названию(из второй колонки таблицы). Все функции вызываются нажатием на кнопку *fx*, находятся в категории «Статистические» и в качестве массива данных (поле «ЧИСЛО 1»), указывается ссылка на A1-A25. Например, для ввода первой из них ставим курсор в C2, нажимаем *fx*,

выбираем категорию «Статистические» и функцию «Счет»(Count), в открывшемся окне ставим курсор в поле «Число 1» и обводим курсором

ячейки A1-A25, нажимаем «ОК». Также поступаем и с другими функциями.

Характеристика	**Функция**
Объем выборки	СЧЁТ(массив данных) COUNT
Выборочное среднее	СРЗНАЧ(массив данных) AVERAGE
Дисперсия	ДИСПВ(массив данных) VAR
Стандартное отклонение	СТАНДОТКЛОН(массив данных) STDEV
Медиана	МЕДИАНА(массив данных)

	MEDIAN
Мода	МОДА(массив данных) MODE
Коэффициент эксцесса	ЭКСЦЕСС(массив данных) KURT
Коэффициент асимметрии	СКОС(массив данных) SKEW
Процентиль25%	ПЕРСЕНТИЛЬ(массив данных; 0,25) PERCENTILE
Процентиль75%	ПЕРСЕНТИЛЬ(массив данных; 0,75) PERCENTILE

Существует другой способ вычисления числовых характеристик выборки в программе EXCEL. Для этого ставим курсор в свободную ячейку (например, D1).Затем вызываем в меню «Сервис»(«Данные» Excel 2010) подменю «Анализ данных» (Data Analysis1). Если в меню «Сервис» отсутствует этот пункт, то в меню «Сервис» нужно выбрать пункт «Надстройки» и в нем поставить флажок напротив пункта «Пакет анализа» (AnalysisToolPak).После этого в меню «Сервис» появится «Анализ данных» (DataAnalysis).

В окне «Анализ данных» нужно выбрать пункт «Описательная статистика» (DescriptiveStatistics). В появившемся окне в поле «Входной интервал» (InputRange) делаем ссылку на выборку A1-A25, помещая курсор в поле и обводя эти ячейки. Оставляем группирование «По столбцам» (Columns). В разделе «Параметры вывода» (OutputOptions)

ставим флажок на «Выходной интервал» (OutputRange) и в соседнем поле задаем ссылку на верхнюю левую ячейку области вывода (например D1), ставим флажок напротив «Описательная статистика» (SummaryStatistics), нажимаем «ОК». Результат – основные характеристики выборки рис. (сделайте шире столбец D, переместив его границу в заголовке).

Столбец1	
Среднее	44,80
Стандартная ошибка	1,07
Медиана	45,00
Мода	38,00
Стандартное отклонени	5,35
Дисперсия выборки	28,58
Эксцесс	-0,62
Асимметричность	0,00
Интервал	21,00
Минимум	34,00
Максимум	55,00
Сумма	1120,00
Счет	25,00

Задание 1 к лабораторной работе №2.

1. Изучить п.п.1,
2. Откройте вашу рабочую книгу Excel. Второй лист документа переименуйте в «Точечные и интервальные оценки распределения». На этом листе выполните данное задание.
3. Вычислить основные числовые характеристики выборки двумя способами согласно своему варианту. В таблице даны выборки объема n=30 , первый столбец -номер по порядку, первая строка - номер варианта).
4. Можно ли считать, что выборка извлечена из совокупности с нормальным распределением? Обоснуйте ответ и запишите.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	6,7	29	18	10,0	65	111	46	56	109	45	25	5,4
2	5,2	40	12	9,1	64	137	36	57	142	23	23	9,7
3	13,5	32	20	6,3	63	133	37	58	107	36	13	4,4
4	3,9	44	13	7,2	65	112	40	54	101	39	24	6,9
5	9,8	32	24	9,3	75	130	59	51	104	28	25	7,4
6	6,3	42	22	10,7	70	127	42	46	97	47	30	6,2
7	6,3	31	18	2,6	80	127	40	60	119	39	29	5,8

8	2,6	41	18	4,9	79	128	42	49	101	39	25	3,6
9	5,1	37	14	8,3	62	139	47	58	116	35	26	5,9
10	5,5	34	14	3,7	73	132	36	48	124	41	26	4,1
11	7,4	26	17	7,1	78	121	31	52	105	34	20	3,8
12	3,9	36	16	7,9	59	112	44	55	108	30	24	5,7
13	6,4	25	19	5,3	70	124	46	53	122	39	23	5,7
14	4,7	44	12	5,8	65	122	43	51	121	47	31	6,8
15	4,7	29	12	6,7	74	111	40	48	109	40	28	4,3
16	5,6	41	22	3,2	74	109	43	53	109	42	26	4,3
17	7,6	27	11	4,1	69	118	36	47	108	37	17	3,3
18	11,8	33	20	4,9	66	116	37	48	126	33	19	7,5
19	9,7	36	17	7,0	62	140	47	50	115	43	22	3,7
20	11,5	38	27	5,2	71	110	31	48	113	29	23	7,8
21	0,5	48	16	6,8	70	118	44	51	116	40	23	2,7
22	3,0	39	14	4,5	77	126	43	44	120	31	34	5,3
23	6,1	33	17	9,9	78	115	37	46	113	41	19	2,4
24	2,8	32	15	10,6	74	116	41	56	102	39	35	7,6
25	7,0	34	11	2,0	70	137	40	53	117	41	18	8,1
26	2,9	29	16	6,1	74	126	38	41	110	43	13	6,2
27	7,8	38	5	4,1	66	124	57	44	92	34	31	8,5
28	15,7	41	21	5,3	65	115	40	47	94	41	23	7,2
29	2,7	34	8	8,7	71	113	55	50	125	36	10	6,4
30	1,5	32	16	5,7	59	120	37	45	108	41	34	5,7

Часть 2. Интервальное оценивание.

Точечные оценки являются лишь приближенным значением неизвестных параметров генеральной совокупности. Чтобы получить представление о точности и надежности оценки используют интервальную оценку параметра.

Интервальной оценкой параметра ***θ*** называется числовой интервал (a,b) который с заданной вероятностью ***p*** покрывает неизвестное значение параметра *θ.*

Такой интервал *(a,b)* называется ***доверительным***, а вероятность *p* ***доверительной вероятностью.*** Вместо нее часто задают величину ***α =1- p , называемую уровнем значимости.***

Рассмотрим теперь методы интервального оценивания.

Если выборка объема *n* представляет случайную величину, ***распределенную нормально***, то доверительный интервалы для математического ожидания:

$$m \approx \left(-\frac{S \cdot t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}}; +\frac{S \cdot t_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}} \right)$$

$$m \approx (-\Delta; +\Delta)$$

Для дисперсии:

$$\sigma^2 \approx \left(\frac{S^2 \cdot (n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}; \frac{S^2 \cdot (n-1)}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right)$$

где $t_p(n)$ и $\chi^2_p(n)$ - квантили распределения Стьюдента и хи-квадрат,

α =1- *p* .

Электронная таблица Excel (Calc), содержащая встроенные статистические функции позволяет **легко и быстро** найти доверительные интервалы для математического ожидания и дисперсии. Рассмотрим решение этой задачи.

Возвращаемся на лист 1 электронной таблицы с данными примера и для них вычислим доверительные интервалы при *p*=0,95.

Вводим данные согласно рисунку:

	F	G	H	I
1		Уровень значимости		0,05
2		Интервал	Левая граница	Правая граница
3		Матожидание		
4		Дисперсия		

Для вычисления величины Δ служит функция «ДОВЕРИТ»(«CONFIDENCE») категории «Статистические» с тремя параметрами «Альфа» - уровень значимости =1- *p* ,

«Станд_откл» - среднеквадратическое отклонение *S*,

«Размер» - объем выборки *n*.

Таким образом, вводим в H3 функцию:

=СРЗНАЧ(A1:A25)-ДОВЕРИТ(I1;СТАНДОТКЛОН(A1:A25);25)

(=AVERAGE(A1:A25)-CONFIDENCE(I1;STDEV(A1:A25);25))

а в ячейку I3 функцию:

=СРЗНАЧ(A1:A25)+ДОВЕРИТ(I1;СТАНДОТКЛОН(A1:A25);25)

(=AVERAGE(A1:A25)+CONFIDENCE(I1;STDEV(A1:A25);25))

Для вычисления доверительного интервала для дисперсии следует отметить, что функция вычисления квантили распределения хи-квадрат

(обратного распределения хи-квадрат) называется «ХИ2ОБР» («CHIINV»)

(категория «Статистические») и имеет два параметра:

первый «Вероятность»содержит доверительную вероятность *p*,

второй – степень свободы *n*-1.

Для вычисления левой границы доверительного интервала для дисперсии

в ячейку H4вводим запись:

=ДИСП(A1:A25)*24/ХИ2ОБР(0,025;24)

(=VAR(A1:A25)*24/CHIINV(0,025;24))

Для вычисления левой границы доверительного интервала для дисперсии

а в ячейку I4 запись:

=ДИСП(A1:A25)*24/ХИ2ОБР(0,975;24)

(=VAR(A1:A25)*24/CHIINV(0,975;24))

.

Получаем значения границ доверительных интервалов.

Задание 2 к лабораторной работе №2.

1. Для выборки вашего варианта (задание 1) вычислить доверительные интервалы для математического ожидания и дисперсии при = 0,01.
2. Изменяя значение уровня значимости сделать вывод о его влиянии на ширину интервала. Вывод записать.

Лабораторная работа № 3 проверка статистической гипотезы о виде распределения

Проверка статистических гипотез используется когда необходимо обосновать вывод о преимуществах того или иного метода лечения, обучения, о пользе лекарства, об уровне доходности ценных бумаг и т.д.

Статистической гипотезой называется любое предположение о виде и параметрах неизвестного закона распределения.

*Проверяемую гипотезу обычно называют **нулевой** (или основной) и обозначают H_0. Наряду с нулевой гипотезой рассматривают **альтернативную** или конкурирующую, гипотезу H_1, являющуюся логическим отрицанием H_0.*

По своему прикладному содержанию статистические гипотезы можно подразделить на несколько основных типов:

- о равенстве числовых характеристик генеральной совокупности;
- о числовых значениях параметров;

- об однородности выборок (т.е. принадлежности их одной и той же генеральной совокупности);
- о стохастической независимости элементов выборки.

Вероятность α отвергнуть гипотезу, когда она верна, называется уровнем значимости критерия.

Одной из важных задач статистики является установление теоретического закона распределения случайной величины, характеризующей изучаемый признак по опытным данным. Предположение о виде распределения может быть сделано, исходя из теоретических предпосылок (выполнение условий центральной предельной теоремы может свидетельствовать о нормальном законе распределения случайной величины), опыта аналогичных предшествующих исследований, на основании графического изображения эмпирического распределения. Параметры распределения, как правило, неизвестны, их заменяют наилучшими оценками по выборке.

Между эмпирическим и теоретическим распределениями неизбежны расхождения. Возникает вопрос: эти расхождения объясняются случайными обстоятельствами или они являются существенными и теоретический закон распределения подобран неудачно. Для ответа на этот вопрос служат критерии согласия.

Одним из наиболее мощных критериев согласия является критерий Пирсона, называемый еще ***критерием Хи-квадрат***. Его суть заключается в сравнении эмпирических частот элементов выборки ni (для дискретных распределений) с теоретическими частотами $n_i' = np_i$, где p_i- вероятность принять это значение, рассчитанное по исследуемому закону распределения. Если распределение непрерывное, то строится группированный статистический ряд из k интервалов и $p_i = F(b_i) - F(a_i)$ есть вероятность попасть в i-й интервал группировки (здесь $F(x)$ - функция распределения проверяемого закона функция Лапласа).

Статистикой критерия является величина $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}$. Эта величина является мерой расхождения эмпирических частот n_i и теоретических частот n_i'. Критическое значение критерия равно обратному распределению хи-квадрат со степенями свободы $(k-r-1)$:

$\chi^2_{kr} = \chi^2_{1-\varsigma}(k-r-1)$,где k–количество интервалов эмпирического распределения, r – число оцениваемых параметров закона распределения, **α**–заданный уровень значимости.

Если $\chi^2 > \chi^2_{kr}$, то гипотеза H_0 отвергается; если выполняется условие $\chi^2 < \chi^2_{kr}$, то распределение можно считать соответствующим теоретическому, другими словами гипотеза H_0не противоречит опытным данным.

ПРИМЕР 1. Имеется выборка измерения пульса у 40 больных, подвергнутых некоторой лечебной процедуре. Проверить гипотезу о том, что значение пульса у подобных больных распределено по *нормальному закону распределения*. Взять уровень значимости α= 0,05.

Выборка ЧСС у 40больных (уд/мин).

64 5669 78 78 83 47 65 77 57 61 52 50 58 60 48 62 63 68 64
64 64 79 66 65 62 85 75 88 61 82 52 72 75 84 66 62 73 64 74

РЕШЕНИЕ.

Для проверки гипотезы H_0 о принадлежности генеральной совокупности нормальному виду распределения необходимо строить интервальный вариационный ряд, т.к. нормальное распределение является непрерывным. Для этого нужно вычислить размах выборки, который равен разнице между максимальным и минимальным элементами выборки. Кроме того, нужно рассчитать точечные оценки математического ожидания и среднеквадратического отклонения (СКО).

Открываем электронную таблицу и вводим данные выборки в ячейкиA2-A41, делаем подписи для расчетных параметров в соответствии с рисунком:

	A	B	C	D	E	F	G
1	Выборка	Параметры	Интервалы	Частота	Вер-ть	Теор.част	Критерий
2	64	Объем	0				
3	56		50				
4	69	Максимум	55				
5	78		60				
6	78	Минимум	65				
7	83		70				
8	47	Среднее	75				
9	65		80				
10	77	СКО	85				
11	57		90				

сумма

Вычисляем параметры по выборке. Для этого вводим в ячейкуB3: **«=СЧЁТ(A2:A41)**(=COUNT(A2:A41))»

(здесь и далее кавычки вводить не надо, функции можно вводить с помощью мастера функций из категории «Статистические», как в лабораторной работе № 2, ссылки на ячейки можно ввести щелкнув мышью по ячейке).

ВВ5 вводим: «**=МАКС(A2:A41)(** =MAX(A2:A41))»,

вВ7: «**=МИН(A2:A41)(**=MIN(A2:A41))»,

вВ9: «**=СРЗНАЧ(A2:A41)(** =AVERAGE(A2:A41))»,

в В11:«**=СТАНДОТКЛОН(A2:A41)(** =STDEV(A2:A41))».

Видно, что весь диапазон значений элементов лежит на интервале от 47

до 88. Разобьем этот интервал на интервалы группировки:

[0; 50], (50; 55], (55; 60], (60; 65], (65; 70], (70; 75], (75; 80], (80; 85],(85; 90]. Для этого вводим в ячейки С2-С11 границы интервалов:

ячейка	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
число	0	50	55	60	65	70	75	80	85	90

Для вычисления частот *пі* используем функцию ЧАСТОТА

(FREQUENCY из категории «массив»).

Для этого в D3 вводим формулу

«=ЧАСТОТА(A2:A41;C3:C11)(FREQUENCY(A2:A41;C3:C11))».

В Calc значения частот появятся сразу для всех интервалов.

В Excel: обводим курсором ячейки D3-D11, выделяя их и нажимаем F2, а затем одновременно Ctrl+Shift+Enter. В результате в ячейках D3-D11 окажутся значения частот.

Для расчета теоретической вероятности pi = F(bi) - F(ai)

вводим в ячейку Е3 разницу между функциями нормального распределения (функция НОРМРАСП (NORMDIST)категории «Статистические»

с параметрами:

«X» – значение границы интервала, «Среднее» - ссылка на ячейкуВ9, «Стандартное_откл» - ссылка на В11, «Интегральная» - 1.

В результате в Е3 будет формула:

=НОРМРАСП(C3;B9;B11;1)-НОРМРАСП(C2;B9;B11;1)

(=NORMDIST(C3;B9;B11;1)-NORMDIST(C2;B9;B11;1))

Автозаполняем эту формулу на E3-E10, перемещая нижний правый

угол E3 до ячейки E10.

В последней ячейке столбца E11 для соблюдения условия нормировки вводим дополнение предыдущих вероятностей до единицы. Для этого вводим в E11: «=1-СУММ(E3:E10)» (можно без нормировки)

Для расчета теоретической частоты $n_i' = np_i$ вводим в F3 формулу: «=E3*B3», автозаполняем ее на F3-F11.

Для вычисления элементов суммы $\frac{(ni-ni')^2}{ni'}$ критерия Пирсона

вводим в G3 значение «=(D3-F3)*(D3-F3)/F3» и автозаполняем его на

диапазон G3-G11.

Находим значение критерия χ^2 и критическое значение χ^2_{kr}.

Для этого вводим в F12 подпись «Сумма», а в F13 подпись «Критич.».

Вводим в соседние ячейки формулы –

в G12: **«=СУММ(G3:G11)**(=SUM(G3:G11))»,

а вG13: «**=ХИ2.ОБР.ПХ(0,05;6)**(=CHIINV(0,05;6))»,

здесь параметр α= 0,05 взят из условия, а степень свободы (k-r-1)=(9-2-1)=6, так как k=9 – число интервалов группировки, а r=2, т.к. были оценены два параметра нормального распределения: математическое ожидание и СКО.

Видно, что, следовательно гипотеза H_0 принимается, то есть можно считать, что ЧСС у данной группы больных распределена по нормальному закону распределения.

Наглядно увидеть это можно, построив графики плотностей эмпирического и теоретического распределений.

Ставим курсор в любую свободную ячейку и вызываем мастер диаграмм (Вставка/Диаграмма). Выбираем тип диаграммы «График» и вид «График с маркерами» самый левый во второй строке, нажимаем «Далее».

Ставим курсор в поле «Диапазон» и удерживая кнопку CTRL обводим мышью область ячеек D3-D11 а затем F3-F11. Переходим на закладку «Ряд» и в поле «Подписи оси X»обводим область C3-C11. Нажимаем «Готово». Видно, что графики достаточно хорошо совпадают, что говорит о соответствии данных нормальному закону.

Задание. Проверить по критерию Пирсона на уровне значимости α = 0,02 статистическую гипотезу о том, что генеральная совокупность, представленная выборкой, имеет нормальный закон распределения.

Данные взять из задания 4 лабораторной работы № 1.

Вариант **Выборка**

1. 45 52 49 48 42 51 54 54 50 47 56 53 59 57 50
45 50 46 55 46 54 55 64 67 51 49 47 47 55 40
2. 48 43 52 42 38 57 47 47 51 52 55 53 50 46 53
50 49 58 53 44 51 49 53 51 51 48 45 46 49 54
3. 65 81 76 84 81 80 78 86 85 83 75 85 83 80 77
69 73 78 75 75 91 79 74 67 68 78 80 81 81 81
4. 75 82 79 78 72 81 84 84 80 77 86 83 89 87 80
75 80 76 85 76 84 85 94 97 81 79 77 77 85 70
5. 78 73 82 72 68 87 77 77 81 82 85 83 80 76 83
80 79 88 83 74 81 79 83 81 81 78 75 76 79 84
6. 70 59 57 62 49 63 59 60 57 66 64 57 59 58 59
56 62 56 57 63 59 55 58 62 61 60 59 59 61 63
7. 39 41 35 41 42 38 41 41 36 45 40 39 41 41 40
42 45 39 39 35 41 36 36 39 41 43 40 41 38 44
8. 15 31 26 34 31 30 28 36 35 33 25 35 33 30 27

19 23 28 25 25 41 29 24 17 18 28 30 31 31 31
9. 25 32 29 28 22 31 34 34 30 27 36 33 39 37 30
25 30 26 35 26 34 35 44 47 31 29 27 27 35 20
10. 59 60 65 50 55 64 66 63 55 62 60 58 67 58 65
63 59 57 65 56 66 59 59 60 61 65 59 50 64 63
11. 40 41 37 37 40 42 39 43 38 41 45 44 48 43 28
39 41 39 38 44 37 41 42 45 40 43 35 44 44 44
12. 54 59 55 57 44 42 52 55 49 53 51 50 61 59 53
46 47 44 52 49 48 56 40 52 46 46 45 52 59 57

Лабораторная работа № 4 основы регрессионного и корреляционного анализа

Цель: Освоить методы построения линейного уравнения парной регрессии , научиться получать и анализировать основные характеристики регрессионного уравнения.

Корреляционный анализ — метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя или более переменными. Корреляционный анализ тесно связан с регрессионным анализом (также часто встречается термин «корреляционно-регрессионный анализ», который является более общим статистическим понятием), с его помощью определяют необходимость включения тех или иных факторов в уравнение множественной регрессии, а также оценивают полученное уравнение регрессии на соответствие выявленным связям (используя коэффициент детерминации).

Рассмотрим следующую задачу. Была проведена серия измерений двух случайных величин X и Y, причем измерения проводились попарно: т.е. за одно измерение мы получали два значения - xi и yi . Имея выборку, состоящую из пар (xi , yi), мы хотим определить, имеется ли между этими двумя переменными зависимость.

Зависимость между случайными величинами может иметь функциональный характер, т.е. быть строгим функциональным отношением, связывающим их значения. Однако при обработке экспериментальных данных гораздо чаще встречаются статистические зависимости. Различие между двумя видами зависимостей состоит в том, что функциональная зависимость устанавливает строгую взаимосвязь между переменными, а статистическая зависимость лишь говорит о том, что распределение случайной величины Y зависит от того, какое значение принимает случайная величина X. Например, если X – количество вводимого объекту препарата, то его концентрация в крови Y в произвольный момент времени статистически зависит от величины X так ка определяется не только количеством вводимого препарата, но и многими факторами (масса тела больного, скорость выведения вещества из организма, количество других веществ в крови и т.д.)

Одной из мер статистической зависимости между двумя переменными является коэффициент корреляции. Он показывает, насколько ярко выражена тенденция к росту одной переменной при увеличении другой. Коэффициент корреляции находится в диапазоне [-1, 1]. Нулевое значение коэффициента обозначает отсутствие такой тенденции (но не обязательно отсутствие зависимости вообще). Если тенденция ярко выражена, то коэффициент корреляции близок к +1 или -1 (в зависимости от знака зависимости), причем строгое равенство единице обозначает крайний случай статистической зависимости - функциональную зависимость. Промежуточные значения коэффициента корреляции говорят, что хотя тенденция к росту одной переменной при увеличении другой не очень ярко выражена, но в какой-то мере она все же присутствует.

Практическая значимость коэффициента корреляции определяется его величиной, возведенной в квадрат, получившая название коэффициента детерминации. Например, если r = 0,8, то r2 = 0,64, т.е. 64% всех изменений одного признака связано с изменением другого.

Регрессии. Эмпирические формулы.

Задача о форме корреляционной связи решается с помощью регрессий.

Регрессией Y от X называется функциональная зависимость между значениями x и соответствующими условными средними **y**(x).Регрессии можно представить геометрически в виде ломаных линий, соединяющих точки(*x*;y(x)). Эти линии называются *эмпирическими (полученными из опыта) ломаными линиями регрессии.*

Регрессии, полученные в виде таблиц или ломаных линий, характеризуют форму корреляционной зависимости между X и Y лишь для выборочных совокупностей. Для генеральной же совокупности они дают приближенную картину этой зависимости. Очевидно, приближение будет тем точнее, чем больше объем выборки n и чем меньше частные интервалы Dx и Dy. При этом ломаная линия регрессии будет приближаться к некоторой плавной кривой. Правда, такую плавную кривую можно получить и иначе – если ломаную линию регрессии “сгладить” посредством какой-либо известной линии (прямой, параболы, гиперболы и т.п.).

Если показателей два, то регрессия называется парной. Если зависимость между показателями X и Y пропорциональная, то регрессия будет линейной и описывается уравнением вида $y = ax + b$.

ПРИМЕР.

Рассмотрим методику построения регрессионного уравнения на примере анализа веса щитовидной железы (Y) и соответствующей площади ее скенографического изображения (X).

X	11	17	25	32	33	44	46	52	73	78	89	95
Y	12	23	41	59	62	96	102	122	203	215	270	282

Введем эту таблицу в ячейки A1-M2 электронной книги Excel.

Просмотрим предварительно, как лежат точки на графике и какое уравнение регрессии лучше выбрать. Для этого строим график.

Вызвав мастер диаграмм, выбираем тип диаграммы «Точечная», нажимаем «Далее» и, поместив курсор в поле «Диапазон» обводим курсором данные Y (ячейки B2-M2). Переходим на закладку «Ряд» и в поле «Значения X»

делаем ссылку на ячейки B1-M1, обводя их курсором. Нажимаем «Готово».

Как видно из графика, точки хорошо укладываются на прямую линию, поэтому будем находить уравнение линейной регрессии вида $y = ax + b$.

Для нахождения коэффициентов a и b уравнения регрессии

служат функции НАКЛОН (SLOPE) и ОТРЕЗОК(INTERCEPT) категории «Статистические».

Вводим в A5 подпись «a=» а в соседнюю ячейку B5 вводим функцию

НАКЛОН(SLOPE), ставим курсор в поле «Изв_знач_y» задаем ссылку на ячейки B2-M2, обводя их мышью. Аналогично в поле «Изв_знач_x» даем

ссылку на B1-M1. Результат 3,36. Найдем теперь коэффициент b.

Вводим в A6 подпись «b=», а в B6 функцию ОТРЕЗОК(INTERCEPT) с теми же параметрами, что и у функции НАКЛОН (SLOPE),. Результат -42,6.

Следовательно, уравнение линейной регрессии есть y = 3,33x- 42,6 .

Построим график уравнения регрессии. Для этого в третью строчку таблицы введем значения функции регрессии в заданных точках X (первая строка) – $y(x_i)$. Для получения этих значений используется функция ТЕНДЕНЦИЯ (FORECAST) категории «Статистические». Вводим в

A3 подпись «Y(X)» и, поместив курсор в B3, вызываем функцию

ТЕНДЕНЦИЯ (FORECAST).

Для Excel: в полях «Изв_знач_y» и «Изв_знач_x» даем ссылку наB2-M2 и B1-M1. В поле «Нов_знач_x» вводим также ссылку на B1-M1.В поле «Константа» вводят 1, если уравнение регрессии имеет виду = ax + b , и 0, если y = ax . В нашем случае вводим единицу. Функция ТЕНДЕНЦИЯ (FORECAST)является массивом, поэтому для вывода всех ее значений выделяем область B3-M3 и нажимаем F2 и Ctrl+Shift+Enter. Результат – значения уравнения регрессии в заданных точках.

В Calc: в поле «Значение» вводим массив B1-M1, в поле «Данные Y» - ссылку на B2-M2, в поле «Данные X» -ссылку на B1-M1. В поле «Массив» поставим флажок , нажимаем ОК и массив B3:M3 заполнится значениями, вычисленными по линейной регрессии.

Строим график.

EXCEL: Ставим курсор в любую свободную клетку, вызываем мастер диаграмм, выбираем категорию «Точечная», вид графика – линия без

точек, нажимаем «Далее», в поле «Диапазон» вводим ссылку на B3-M3. Переходим на закладку «Ряд» и в поле «Значения X» вводим ссылку на B1-M1, нажимаем «Готово». Результат –прямая линия регрессии.

Посмотрим, как различаются графики опытных данных и уравнения регрессии. Для этого ставим курсор в любую свободную ячейку, вызываем мастер диаграмм, категория «График»,вид графика – ломаная линия с точками (или точечная с прямыми отрезками), нажимаем «Далее», в поле «Диапазон» вводим ссылку на вторую и третью строки B2-M3.

Переходим на закладку «Ряд» и в поле «Подписи оси X» вводим ссылку на B1-M1, нажимаем «Готово». Результат – две линии (Синяя – исходные данные, красная – уравнение регрессии). Видно, что линии мало различаются между собой.

CALC: В график с точками вставляем линию тренда (она же линия регрессии). Для этого в области построения диаграммы нажимаем правую кнопку мыши и выбираем « вставить линию тренда».

Линия регрессии позволяет с некоторой вероятностью предсказать в интервале от X=11 до X=89 любые значения функции Y при отсутствующих значениях фактора X, но и за пределами данного интервала. Так, например, чтобы вычислить вес щитовидной железы, соответствующий площади скеннографического изображения равной X=90 см^3, воспользуемся встроенной статистической функцией ПРЕДСКАЗ(CALC?). Расчет показывает, что вес щитовидной железы будет в этом случае равен Y=259,64.

Для вычисления коэффициента корреляции r_{xy} служит функция КОРРЕЛ (CORREL). Размещаем графики так, чтобы они располагались выше

25 строки, и в A25 делаем подпись «Корреляция», в B25 вызываем

функцию КОРРЕЛ (CORREL), в полях которой «Массив 1» и «Массив 2» вводим ссылки на исходные данные B1-M1 и B2-M2. Результат 0,9951. Коэффициент детерминации Rxy – это квадрат коэффициента корреляции .В A26 делаем подпись «Детерминация», а в B26 – формулу«=B25*B25». Результат 0,99, т.е. 99% всех изменений одного признака связано с изменением другого.

Однако существует одна функция, которая рассчитывает

все основные характеристики линейной регрессии. Это функция ЛИНЕЙН (LINEST). Ставим курсор в B28 и вызываем функцию ЛИНЕЙН(LINEST). категории «Статистические». В полях «Изв_знач_y» и «Изв_знач_x» даем ссылку на B2-M2 и B1-M1. Поле «Константа» имеет тот же смысл, что и в функции ТЕНДЕНЦИЯ, у нас она равна 1. Поле «Стат» должно содержать 1, если нужно вывести полную статистику о регрессии. В нашем случае ставим туда единицу. Функция возвращает массив размером 2 столбца и 5 строк. После ввода выделяем мышью ячейки B28-C32 и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица значений, числа в которой имеют следующий смысл:

Коэффициент *a*	Коэффициент b

Стандартная ошибка a ¿ $m_{¿}$	Стандартная ошибка m_b
Коэффициент детерминации *R xy*	Среднеквадратическое отклонение *y*
F – статистика	Степени свободы *n-2*
Регрессионная сумма квадратов S_b^2	Остаточная сумма квадратов

***Анализ результата*:**

в первой строчке – коэффициенты уравнения регрессии, сравните их с рассчитанными функциями НАКЛОН и ОТРЕЗОК.

Вторая строчка – стандартные ошибки коэффициентов.

Если одна из них по модулю больше чем сам коэффициент, то коэффициент считается нулевым. Коэффициент детерминации характеризует качество связи между факторами. Полученное значение 0,9902 говорит об очень хорошей связи факторов. F – статистика проверяет ***гипотезу об адекватности регрессионной модели***.

Данное число нужно сравнить с критическим значением. Для его получения вводим в E33 подпись «F-критическое», а в F33 функцию FРАСПОБР (FINV), аргументами которой вводим соответственно «0,05» (уровень значимости), «1»(число факторов X) и «10» (степени свободы). Видно, что F – статистика больше, чем F– критическое, значит регрессионная модель адекватна.

В последней строке приведены регрессионная сумма квадратов и остаточные суммы квадратов. Важно, чтобы регрессионная сумма (объясненная регрессией) намного больше остаточной (не объясненная регрессией, вызванная случайными факторами). В нашем случае это условие выполняется, что говорит о хорошей регрессии.

Задание. Даны выборки факторов xi и yi. По этим выборкам:

1. Построить эмпирическую линию регрессии (ломаную линию).
2. Найти уравнение линейной регрессии.
3. Найти коэффициент парной корреляции, коэффициент детерминации.
4. Проверить на уровне значимости a = 0,05 регрессионную модель на адекватность.

Вариант 1, 7

Средняя длина тела плода (см)	32	34	36	38	41	43	45	47	50
Возраст внутри утробного плода (нед)	24	26	28	30	32	34	36	38	40

Вариант 2, 8

Площадь поражения артерии таза (%)	22,3	3,1	48,3	17	7,5	40,2	23,1	16	32,5	29
Возраст (в годах)	55	35	75	50	45	65	55	45	60	65

Вариант 3, 9

Содержание андростерона в моче (мг/в сут)	0,82	0,9	0,98	1,06	1,2	1,29	1,48	1,42	1,4	1,08
Возраст (в годах)	82	82	75	65	55	45	25	25	35	65

Вариант 4, 10

Концентрация пролактина в крови (нг/мл)	25	120	75	50	185	125	70	145	170	80
Возраст (в годах)	1	5	4	2	9	6	3	7	8	4

Вариант 5, 11

Поверхность тела ($м^2$)	1,1	1,5	1,2	1,3	1,9	1,3	2	1,7	1,5	1,7
Вес(кг)	22	45	27	33	78	38	88	60	52	68

Вариант 6, 12

Объем циркулирующей крови (л)	4,83	5,08	3,81	5,34	4,06	5,34	4,32	5,59	4,57	5,8
Рост (см)	170	175	150	175	155	180	160	185	165	190

Замечание: При решении задачи выборку (X_i ;Y_i) целесообразно занести в электронную таблицу по возрастанию значений фактора X.