

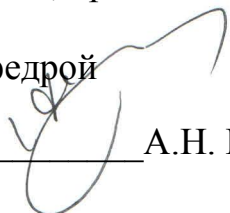
МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук
Кафедра «Вычислительная техника»

Утверждено на заседании кафедры
«Вычислительная техника»
«28 » января 2021 г., протокол № 8

Заведующий кафедрой


_____ А.Н. Ивутин

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
по выполнению лабораторных работ
по дисциплине (модулю)
«Математическая статистика»**

**основной профессиональной образовательной программы
высшего образования – программы бакалавриата**

по направлению подготовки
09.03.01 Информатика и вычислительная техника

с направленностью (профилем)
Электронно-вычислительные машины, комплексы, системы и сети

Форма обучения: очная, заочная

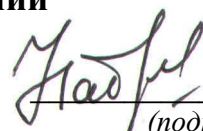
Идентификационный номер образовательной программы: 090301-02-21

Тула 2021 год

Разработчик(и) методических указаний

Набродова И.Н., доцент, к.т.н.

(ФИО, должность, ученая степень, ученое звание)



(подпись)

Оглавление

Лабораторная работа №1	4
СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ	4
Лабораторная работа №2	11
ОСНОВНЫЕ РАСПРЕДЕЛЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ..	11

Лабораторная работа №1

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ

1. Цель и задачи работы

Научиться основным методам обработки данных, представленных выборкой, путем построения гистограммы, определения выборочного среднего, выборочной дисперсии, выборочной медианы и моды.

2. Порядок выполнения работы

- ознакомиться с теоретическими сведениями;
- выполнить задание;
- оформить отчет;
- ответить на контрольные вопросы, заданные преподавателем.

3. Оформление отчета

Отчет должен содержать: титульный лист, цель работы, описание пунктов выполнения лабораторной работы в соответствии с заданием, ответы на контрольные вопросы и выводы по работе.

4. Теоретические сведения

Вероятностная модель ставит в соответствие результатам наблюдений

$$x_1, x_2, \dots, x_n \quad (1)$$

последовательность случайных величин

$$X_1, X_2, \dots, X_n. \quad (2)$$

Предполагается, что случайные величины $X_1, X_2, \dots, X_n$ независимы и имеют одно и то же распределение с функцией распределения $F(x)$.

Полагают, что наблюдения (1) являются значениями величин (2) при осуществлении вероятностного эксперимента. Несмотря на различие объектов (1) и (2), в математической статистике принято называть и то и другое выборкой из генеральной совокупности.

Количество наблюдений n называется объемом выборки.

Произвольная случайная величина X характеризуется своей функцией распределения вероятностей $F(x)$. Если эта функция неизвестна, но известна выборка (1), числовые данные которой являются значениями случайной величины X , то возможно построить эмпирическую функцию распределения вероятностей $F_n(x)$, которая служит оценкой теоретической функции распределения вероятностей $F(x)$. Если обозначить через $\mu_n(x)$ число тех значений $x_1, x_2, \dots, x_n$, которые меньше или равны x , то

$$F_n(x) = \frac{\mu_n(x)}{n}. \quad (3)$$

Если объем выборки n большой, то для представления о виде ее распределения строится гистограмма.

Вводим в первый столбец (ячейки A1...) исходные данные. Для элементов выборки находим минимальный и максимальный элементы, которые

ограничивают интервал, содержащий все элементы выборки. Для этого запишем в первую строку второго столбца (B1) слово Максимум, а во вторую строку второго столбца (B2) слово Минимум. В соседних ячейках C1 и C2 определим функции МАХ и MIN. Для этого ставим курсор в C1 и вызываем мастер функций, нажав на кнопку fx, в открывшемся окне в поле «Категория» выбираем СТАТИСТИЧЕСКИЕ, и ниже ищем функцию МАКС и вызываем ее двойным щелчком по названию. В качестве аргумента функции (в графе «Число 1») обведем область данных (ячейки A1...). Поле «Число 2» оставляем пустым. Нажимаем «ОК». Ставим курсор в ячейку C2 и аналогично вводим функцию МИН. В некоторых случаях для удобства обработки интервал расширяется, но не существенно.

Следующим шагом является разбиение построенного интервала на 5-10 более мелких интервалов. Если разбиение построено удачно, то гистограмма будет напоминать график плотности (если она существует) распределения вероятностей случайной величины, значениями которой являются элементы выборки. Если разбиение мелкое, то гистограмма не дает представления о плотности распределения вероятностей из-за случайных флуктуаций. Если разбиение крупное, то гистограмма также не дает представления о плотности распределения вероятностей из-за того, что теряется много информации.

Чтобы построить интервалы разбиения (группировки), нужно от максимального значения выборки вычесть минимальное значение и полученный результат разделить на число интервалов. Полученное значение называется шагом разбиения. Чтобы получить верхние границы интервалов группировки, нужно последовательно прибавлять шаг разбиения, начиная от минимального значения выборки.

В ячейки D1... вводим верхние границы интервалов группировки. Для вычисления частот n_i используется функция ЧАСТОТА, находящаяся в категории СТАТИСТИЧЕСКИЕ. Введем ее в ячейку E1. В строке «Массив данных» введем диапазон выборки (ячейки A1...). В строке «Массив интервалов» введем диапазон верхних границ интервалов группировки (ячейки D1...). Результат функции является массивом и выводится в ячейках E1... Для полного вывода (не только первого числа в E1) нужно выделить ячейки E1..., обведя их мышью, и нажать F2, а далее одновременно CTRL+SHIFT+ENTER. Результат – частоты n_i , которые показывают, сколько элементов выборки попало в каждый из интервалов разбиения.

При использовании EXCEL 2007 для создания диаграммы необходимо выделить блок данных, на основании которых строится диаграмма. В выделяемый блок данных включить не только числовые данные, но и заголовки строк (столбцов), в которых они расположены.

Заголовки будут использованы в качестве подписей по осям (меток) и для формирования условных обозначений (легенды). При выделении блоков с данными для построения диаграмм необходимо соблюдать два правила:

1. Выделенный фрагмент должен состоять из равновеликих столбцов.
2. В выделенном фрагменте не должно быть объединенных ячеек.

Для построения гистограммы необходимо перейти на вкладку ВСТАВКА, открыть список ГИСТОГРАММА выбрать нужную гистограмму. Гистограмма строится сразу. Иногда необходимо выделить построенную диаграмму и провести изменение размера шрифта или растянуть диаграмму для лучшего чтения данных в поле диаграммы. Если вызвать контекстное меню в поле всей диаграммы, то меню предлагает три отдельных шага в построении диаграммы (в предыдущих версиях было четыре шага): Изменить тип диаграммы; выбрать данные; переместить диаграмму.

В мастере функций fx существуют специальные функции, позволяющие вычислять выборочные характеристики.

Функция СРЗНАЧ вычисляет выборочное среднее (оценку теоретического математического ожидания)

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

Функция ДИСП вычисляет выборочную дисперсию (оценку теоретической дисперсии)

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Функция СТАНДОТКЛОН вычисляет квадратный корень из выборочной дисперсии.

Функция МЕДИАНА вычисляет выборочную медиану (оценку медианы) заданной выборки. Медианой случайной величины называется то ее значение, которое делит распределение на две равновероятные половины. В качестве выборочной медианы $\tilde{m}$ в выборке объема $2n+1$ берут значение $x_{(n+1)}$ в вариационном ряде. Если объем выборки равен $2n$, то в качестве выборочной

медианы $\tilde{m}$ берут $\frac{1}{2}(x_{(n)} + x_{(n+1)})$.

Функция МОДА вычисляет выборочную моду (оценку моды). Модой случайной величины называется ее наиболее вероятное значение. В Excel можно генерировать случайные числа, имеющие разные законы распределения. Для этого можно использовать надстройку АНАЛИЗ ДАННЫХ и пункт ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ЧИСЕЛ.

Если вы хотите сгенерировать, например, 100 случайных чисел из нормального распределения, то в поле ЧИСЛО ПЕРЕМЕННЫХ введите 1; в поле ЧИСЛО СЛУЧАЙНЫХ ЧИСЕЛ введите 100; в списке РАСПРЕДЕЛЕНИЕ выберите НОРМАЛЬНОЕ; введите параметры нормального распределения – СРЕДНЕЕ и СТАНДАРТНОЕ ОТКЛОНЕНИЕ. В качестве типа распределения можно выбрать, например, РАВНОМЕРНОЕ, БИНОМИАЛЬНОЕ или РАСПРЕДЕЛЕНИЕ ПУАССОНА. Введя для каждого распределения соответствующие параметры, получим сгенерированные случайные числа.

5. Оборудование

Персональный компьютер с установленной операционной системой Windows XP/7/8, браузер (Например, Internet Explorer, Google Chrome, Opera), OoO Writer (MS Word), OoO Calc (MS Excel) пакет офисных приложений «Мой офис».

6. Задание на работу

Выборка состоит из 50 значений некоторой случайной величины. Используя MSeXcel (OoO Calc) построить гистограмму, вычислить выборочное среднее, выборочную дисперсию (исправленную), выборочные медиану и моду.

Вариант 1.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	0.865	0.932	2.303	1.51	0.605	3.181	2.547	0.773	2.982	1.64	3.248

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	1.711	4.801	4.085	1.802	1.592	3.519	2.284	2.514	3.276	3.999	2.859	2.182	0.056

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	2.717	0.029	1.132	1.816	3.004	1.464	1.656	4.096	1.81	2.349	3.015	0.878	2.741

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	4.867	4.43	1.916	2.415	2.407	4.284	0.706	3.098	0.283	0.616	3.594	2.088	0.641

Вариант 2.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	0.122	- 0.359	0.053	- 0.903	- 2.371	1.087	0.759	2.113	5.384	2.617	2.97

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	2.724	2.831	2.346	- 1.089	1.138	- 0.511	2.393	0.636	- 0.289	- 0.446	- 0.033	2.116	0.51

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	1.179	3.524	- 0.411	1.004	3.215	2.785	- 4.802	- 3.314	1.412	- 0.232	- 1.395	1.198	2.542

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	1.612	1.023	- 0.529	0.182	- 0.348	0.736	3.036	1.361	2.027	2.48	0.967	1.558	1.324

Вариант 3.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	- 1.338	0.122	- 2.625	- 1.733	0.682	- 2.196	0.223	- 0.831	- 4.894	- 2.761	0.11

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	- 0.527	- 0.979	0.068	0.391	- 1.359	- 1.921	0.794	1.937	- 1.249	1.354	0.054	- 1.774	- 1.149

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	1.343	1.939	1.278	- 0.372	- 0.721	- 0.731	- 0.679	- 1.327	3.126	- 0.854	- 0.937	- 2.778	0.337

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	- 2.081	0.282	3.309	0.862	0.342	1.774	3.885	1.307	0.905	0.85	0.046	- 0.691	1.918

Вариант 4.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	- 0.263	- 1.213	- 2.124	- 0.824	- 1.833	1.022	- 0.107	- 0.018	- 0.664	3.053	1.798

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	- 0.672	2.504	1.565	2.29	1.544	0.726	- 1.365	1.687	- 0.571	0.656	- 0.852	0.448	- 4.465

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	-0.33	- 2.262	- 1.523	- 0.758	- 2.043	0.66	- 1.859	- 2.738	0.02	- 2.133	- 0.143	- 0.384	1.611

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	2.055	- 2.704	- 1.776	- 0.471	- 0.488	0.679	4.905	- 3.749	0.98	-2.49	- 0.751	1.8	- 2.027

Вариант 5.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	0.253	3.468	- 2.413	3.27	0.683	- 1.509	- 0.294	- 0.682	- 0.648	-2.21	2.707

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	-2.01	2.848	0.168	- 2.001	- 1.058	- 0.927	- 1.063	0.527	- 0.563	- 2.016	- 0.886	- 0.658	1.427

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	- 0.911	- 0.806	1.243	- 1.039	3.053	- 0.205	- 1.037	- 0.107	- 2.193	- 1.681	- 2.199	2.263	2.131

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	- 0.239	- 0.241	- 1.711	0.065	- 0.102	0.576	2.813	- 0.128	3.4	1.69	- 2.676	3.568	0.129

Вариант 6.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	-5.65	2.492	- 1.634	4.298	-2.39	3.629	2.128	3.072	- 0.626	3.484	- 0.011

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	3.968	- 1.269	- 1.432	-3.21	- 4.115	2.53	0.829	2.146	0.891	0.368	- 9.371	- 4.943	- 0.417

№	25	26	27	28	29	30	31	32	33	34	35	36	37
---	----	----	----	----	----	----	----	----	----	----	----	----	----

X	- 4.854	- 1.304	- 0.948	- 2.528	- 5.092	1.429	2.047	0.366	- 4.127	- 0.101	1.016	4.364	0.802
---	------------	------------	------------	------------	------------	-------	-------	-------	------------	------------	-------	-------	-------

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	- 1.595	0.583	2.488	1.578	- 4.117	1.013	-1.65	-0.89	0.21	- 3.219	- 0.576	-0.91	1.773

Вариант 7.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	4.848	2.896	2.26	- 0.985	1.096	5.162	2.655	2.437	5.242	3.461	3.066

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	1.883	3.825	1.658	1.06	4.475	- 0.217	2.094	1.834	2.429	4.285	4.333	1.378	0.577

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	1.759	0.816	4.409	3.157	4.032	5.137	2.924	4.108	6.047	- 0.296	2.047	3.438	2.516

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	3.412	4.349	3.462	2.136	2.181	5.722	1.031	2.254	0.554	2.202	- 1.089	5.441	0.93

Вариант 8.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	- 4.267	- 2.319	- 3.054	- 1.187	- 2.417	- 1.588	1.494	1.209	- 1.066	- 3.996	0.949

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	- 5.813	- 0.886	- 3.126	- 1.994	- 1.615	- 0.888	0.259	1.868	- 1.616	- 1.624	- 2.581	- 1.453	- 1.014

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	- 0.696	- 2.353	- 0.969	- 1.113	- 2.688	- 0.591	- 0.934	- 3.471	0.142	0.428	- 5.595	- 2.793	- 2.209

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	1.241	1.726	- 2.003	- 1.004	- 0.658	- 0.042	- 0.458	- 1.338	- 0.895	-2.11	0.215	- 2.095	-1.4

Вариант 9.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	2.316	2.249	1.829	1.547	2.294	2.151	1.066	2.793	2.716	3.044	2.811

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	2.617	1.755	1.914	2.685	2.791	2.358	1.789	1.844	2.136	2.864	3.958	2.069	2.565

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	2.557	3.891	3.479	- 0.294	2.662	1.559	1.526	2.089	1.301	2.347	1.125	3.9	- 0.726

№	38	39	40	41	42	43	44	45	46	47	48	49	50
---	----	----	----	----	----	----	----	----	----	----	----	----	----

X	2.87	4.179	3.085	3.176	1.029	0.424	2.464	3.474	3.165	1.237	2.509	2.496	1.198
---	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Вариант 10.

№ наблюдения	1	2	3	4	5	6	7	8	9	10	11
Значение X	0.643	- 0.713	- 1.043	4.572	-0.43	-1.92	1.139	- 0.589	- 1.805	- 1.555	- 0.029

№	12	13	14	15	16	17	18	19	20	21	22	23	24
X	0.048	0.397	0.737	0.357	- 1.027	- 2.422	2.117	1.457	- 1.734	0.216	2.546	-0.48	- 3.047

№	25	26	27	28	29	30	31	32	33	34	35	36	37
X	- 2.117	- 1.988	- 0.776	0.093	- 0.894	- 0.496	0.809	0.328	0.608	- 1.734	- 1.658	- 1.655	- 0.611

№	38	39	40	41	42	43	44	45	46	47	48	49	50
X	-1.97	- 4.261	1.914	1.187	0.214	- 1.509	-1.39	- 1.317	0.417	- 0.603	1.953	- 2.067	0.57

7. Контрольные вопросы

1. Что называется объемом выборки?
2. Чем характеризуется произвольная случайная величина X?
3. Для чего строится эмпирическая функция распределения?
4. Что такое гистограмма и для чего она строится?

Лабораторная работа №2

ОСНОВНЫЕ РАСПРЕДЕЛЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1. Цель и задачи работы

Исследовать основные распределения, используемые в математической статистике: нормальное распределение, распределение хи-квадрат, распределения Стьюдента и Фишера.

2. Порядок выполнения работы

- ознакомиться с теоретическими сведениями;
- выполнить задание;
- оформить отчет;
- ответить на контрольные вопросы, заданные преподавателем.

3. Оформление отчета

Отчет должен содержать: титульный лист, цель работы, описание пунктов выполнения лабораторной работы в соответствии с заданием, ответы на контрольные вопросы и выводы по работе.

4. Теоретические сведения

При статистических исследованиях широко используются случайные величины, имеющие нормальное распределение, распределение χ^2 (хи-квадрат), распределения Стьюдента и Фишера.

Случайная величина X имеет нормальное распределение с параметрами m (математическое ожидание) и σ^2 (дисперсия), если плотность распределения имеет вид:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right].$$

В качестве примера на рис. 1 изображены два графика плотности нормального распределения с одинаковым математическим ожиданием и разными дисперсиями.

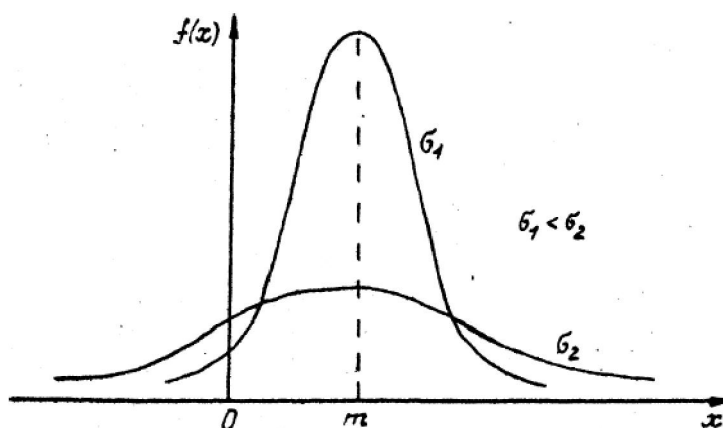


Рисунок 1 – Графики плотностей нормального распределения с одинаковым математическим ожиданием m и разными дисперсиями σ_1^2, σ_2^2

Нормальное распределение с параметрами 0 и 1 называют стандартным нормальным распределением и обозначают $N(0,1)$.

Нормально распределенная случайная величина с большой вероятностью принимает значения, близкие к своему математическому ожиданию, что выражается правилом сигм:

$$P(|X - m| < k\sigma) = \begin{cases} 0,6827, & k = 1 \\ 0,9545, & k = 2 \\ 0,9973, & k = 3 \end{cases}$$

Чаще всего используется правило трех сигм, т.е. $k = 3$.

Построим график плотности нормального распределения и исследуем влияние параметров m (математическое распределение) и σ (среднеквадратическое отклонение) на него.

Запускаем программу Excel и задаем значения параметров m и σ .

Для этого в ячейки первой строки первого столбца (A1) и второй строки первого столбца (A2) вводим подписи $m=$ и $\text{sig} =$, а в первую строку второго столбца (B1) и во вторую строку второго столбца (B2) вводим их числовые значения. Для построения графика протабулируем в третьем и четвертом столбцах (соответственно C и D) функцию плотности нормального распределения на интервале (a,b) с некоторым выбранным шагом h (интервал (a,b) лучше выбрать так, чтобы его серединой было значение m). Для этого вводим в C1 надпись $X=$, а в D1 надпись $f=$. Вводим в C2 значение a , а в C3 значение $a+h$. После этого обводим, выделяя, ячейки C2 и C3 и захватив за нижний правый угол рамки вокруг ячеек C2 и C3, перетягиваем его вниз до ячейки $C(2+(b-a)/h)$, что позволит автоматически занести в столбец значения от a до b с шагом h . Ставим курсор в ячейку D2 и вызываем функцию плотности нормального распределения. Для этого нажимаем кнопку мастера функций f_x и выбираем категорию СТАТИСТИЧЕСКИЕ и функцию НОРМРАСП.

Вводим ссылкой на переменную X : «C2» (для ввода ссылки достаточно щелкнуть мышью по ячейке с данной адресацией), ссылкой на m и σ - «\$B\$1» и «\$B\$2». Эти ссылки абсолютные, так как ячейки со значениями m и σ всегда B1 и B2, поэтому пишется знак \$ (чтобы быстро относительную ссылку сделать абсолютной нужно после ввода ссылки нажать F4). В поле «Интегральное» ставим 0 или «ложь», нажимаем «ОК».

В ячейке D2 появляется результат $f(a)$ – значение плотности нормального распределения, а в строке формул – запись $=\text{НОРМРАСП}(C2;\$B\$1;\$B\$2;\text{ложь})$. За нижний правый угол ячейки D2 автозаполняем результат на ячейки $D2-D(2+(b-a)/h)$. Если требуется построить функцию распределения вероятностей нормального распределения, то в поле «Интегральное» нужно поставить а или «истина».

Строим график плотности нормального распределения по данным. Ставим курсор в любой свободной ячейке. При использовании EXCEL 2007 установите курсор на ячейку, где хотите расположить график и вверху в меню переключитесь на вкладку ВСТАВКА. Затем нажмите на кнопку ГРАФИК,

выпадет несколько их видов. Выбрать можно любой, например, первый – классический график. На листе появится новый объект – чистый график. Когда он выделен, то верхняя панель с иконками действий имеет другой вид, специально для работы с графиками. Чтобы заполнить график, нажмите на кнопку ВЫБРАТЬ ДАННЫЕ. Отобразится окно выбора данных для графика. В нем имеется поле ВЫБОР ДАННЫХ ДЛЯ ДИАГРАММЫ. В конце поля необходимо нажать на кнопку выбора диапазона. Далее следует выделить мышкой таблицу с данными и подписями строк и столбцов и снова кликнуть на кнопку выбора диапазона данных. Нажимаем «Ок». График построен.

Следующим шагом является исследование влияния параметров m и σ на вид графика. Для этого увеличиваем значение математического ожидания m в ячейке B1 и нажимаем «Enter». График плотности нормального распределения должен сместиться вправо. Теперь уменьшаем значение математического ожидания m в ячейке B1 и нажимаем «Enter». График плотности нормального распределения должен сместиться влево. Возвращаем в B1 первоначальное значение m и начинаем изучать влияние среднеквадратического отклонения σ (или дисперсии σ^2) на график плотности. Увеличивая значение σ в ячейке B2, можно наблюдать растяжение графика. Уменьшая значение σ , можно наблюдать сжатие графика.

После такого исследования можно сделать соответствующие выводы о влиянии параметров m и σ на вид графика плотности нормального распределения.

Пусть $X_1, X_2, \dots, X_n$ – независимые случайные величины с общим распределением $N(0,1)$ (нормальное распределение с нулевым математическим ожиданием и единичной дисперсией). Тогда случайная величина

$$\chi^2 = \sum_{i=1}^n X_i^2$$

имеет распределение χ^2 с n степенями свободы или распределение χ_n^2 .

Плотность распределения вероятностей χ_n^2 имеет вид:

$$f_{\chi_n^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

где

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx.$$

Графики плотности распределения χ^2 с n степенями свободы асимметричны и, начиная с $n = 2$, имеют по одному максимуму в точке $x=n-2$ (рис. 2). При этом с ростом n кривая плотности приближается к симметричной функции.

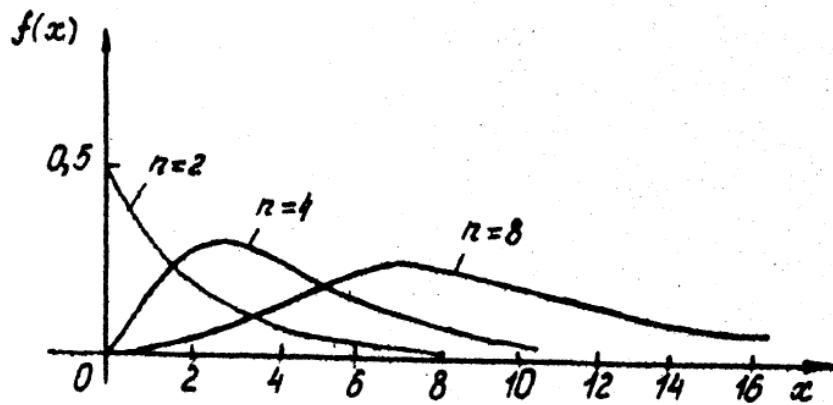


Рисунок 2 – Графики плотности распределения χ^2 с n степенями свободы

Интервал для построения необходимо выбрать с нулевой левой границей и произвести его разбиение. Пусть точки разбиения занимают, например, ячейки B1-B15. Для того, чтобы найти значение функции распределения вероятностей χ^2 с n степенями свободы, нужно нажать кнопку мастера функций fx, выбрать категорию СТАТИСТИЧЕСКИЕ и функцию ХИ2РАСП. Вызванная функция находит значения вероятностей $P(X > x)$, которые заполняют, например, ячейки C1-C15. Поскольку нам нужна функция распределения вероятностей $F(x) = P(X \leq x) = 1 - P(X > x)$, то ее значения мы получим в ячейках D1-D15 следующим образом: $D1 = 1 - C1, \dots, D15 = 1 - C15$. Для определения значений плотности распределения χ^2 с n степенями свободы необходимо воспользоваться формулой для приближенного вычисления первой производной от функции распределения. Если шаг разбиения равен h , а функция распределения вероятностей в соседних точках разбиения имеет значения соответственно $g((i-1)h)$ и $g(ih)$, то значение плотности распределения вероятностей в точке разбиения ih равна $(g(ih) - g((i-1)h))/h$. Значения плотности распределения вероятностей получаем в ячейках, например, начиная с E2 по формуле $E2 = (D2 - D1)/h$. Остальные ячейки автозаполняем. Таким образом, можно найти значения плотности χ^2 с n степенями свободы на выбранном отрезке, кроме нуля (левой границы отрезка).

Действуя аналогично построению графика плотности нормального распределения, строим график плотности распределения χ^2 с n степенями свободы и изучаем влияние степени свободы на вид графика, уменьшая или увеличивая n .

После исследования делаем соответствующие выводы.

Пусть случайная величина Y имеет распределение $N(0,1)$, а независимая от Y случайная величина Z принадлежит χ_n^2 (имеет распределение χ^2 с n степенями свободы). Тогда случайная величина

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}}$$

имеет распределение Стьюдента с n степенями свободы (t_n – распределение) и плотностью распределения вероятностей:

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in (-\infty, +\infty).$$

Графики плотности случайной величины, имеющей распределение Стьюдента, при любом $n=1,2,\dots$ симметричны относительно оси ординат (рис.3), поэтому при любом $n=1,2,\dots$ математическое ожидание равно нулю.

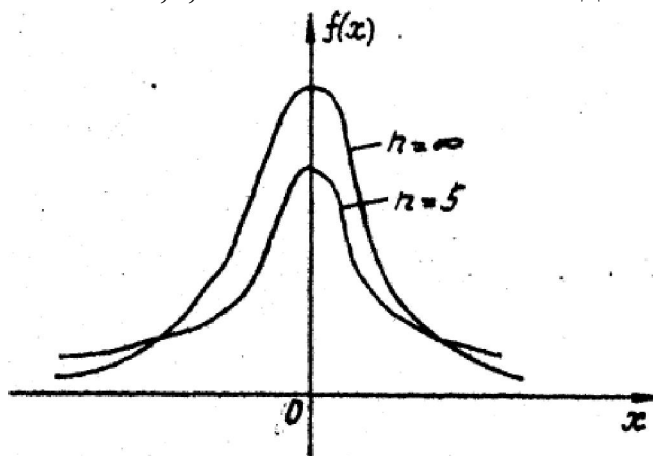


Рисунок 3 – Графики плотности распределения Стьюдента

С ростом n распределение Стьюдента приближается к $N(0,1)$.

Интервал для построения функции распределения Стьюдента t_n необходимо выбрать симметричным относительно нуля. Разбиение построить так, чтобы среди точек разбиения был нуль. Пусть точки разбиения занимают, например, ячейки B1-B15, среди них ячейки B1-B7 содержат отрицательные значения, ячейка B8 содержит нуль и ячейки B9-B15 содержат положительные значения.

Для того, чтобы найти значения функции распределения Стьюдента n t с n степенями свободы, нужно нажать кнопку мастера функций fx , выбрать категорию СТАТИСТИЧЕСКИЕ и функцию СТЫЮДРАСП. 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. В этом случае вызванная функция находит значения вероятностей $P(X > x)$ при неотрицательном x , которые заполняют, например, соответственно ячейки C9-C15. В ячейке C9 обязательно должно быть значение 0.5, поскольку это вероятность $P(X > 0) = 0.5$. Поскольку нам нужна функция распределения вероятностей $F(x) = P(X \leq x) = 1 - P(X > x)$, то ее значения мы получим в ячейках D9-D15 следующим образом: $D9=1-C9, \dots, D15=1-C15$. Значения функции распределения в ячейках D1-D7 получим по формуле: $D1=C15, D2=C14, \dots, D7=C9$, поскольку функция распределения Стьюдента из-за симметричности распределения обладает свойством $P(X \leq -x) = P(X > x)$.

Для нахождения значений плотности распределения Стьюдента t_n необходимо воспользоваться формулой для приближенного вычисления значений первой производной (по аналогии с распределением χ^2).

Далее строим плотность распределения вероятностей t_n с n степенями свободы (значение берется произвольно) и изучаем влияние степени свободы на вид графика, уменьшая или увеличивая значение n .

После исследования делаем соответствующие выводы.

Пусть U и V - независимые случайные величины, распределенные по закону χ^2 с n_1 и n_2 степенями свободы соответственно. Тогда случайная величина

$$X = \frac{\frac{U}{n_1}}{\frac{V}{n_2}}$$

имеет распределение Фишера с n_1 и n_2 степенями свободы (F_{n_1, n_2} – распределение) и плотностью распределения вероятностей

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right) \cdot n_1^{\frac{n_1}{2}} \cdot n_2^{\frac{n_2}{2}}}{\Gamma\left(\frac{n_1}{2}\right) \cdot \Gamma\left(\frac{n_2}{2}\right)} \cdot x^{\frac{n_1}{2} - 1} \cdot (n_2 + n_1 x)^{-\frac{n_1 + n_2}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Графики плотности распределения случайной величины асимметричны, имеют длинные «хвосты» и достигают максимума вблизи точки $x=1$ (рис.4).

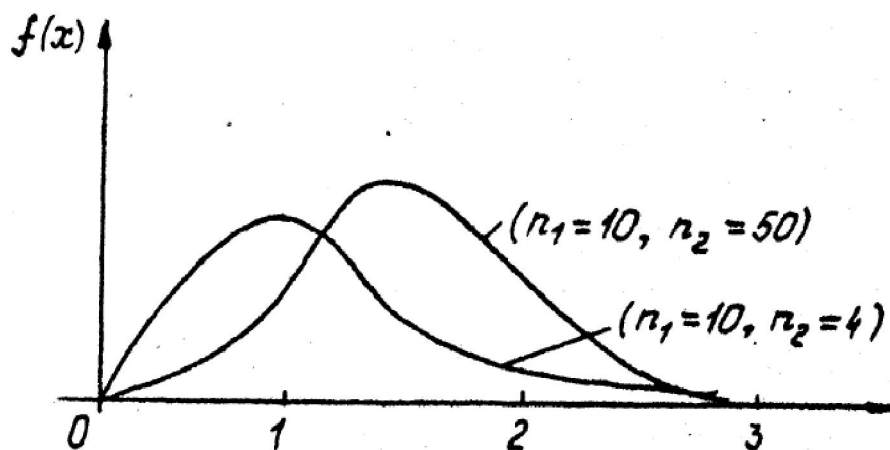


Рисунок 4 – Графики плотности распределения Фишера

Интервал для построения необходимо выбрать с нулевой левой границей. Пусть точки разбиения занимают, например, ячейки B1-B15. Для того, чтобы найти значения функции распределения Фишера F_{n_1, n_2} с n_1 и n_2 степенями свободы, нужно нажать кнопку мастера функций fx , выбрать категорию СТАТИСТИЧЕСКИЕ и функцию ФРАСП. Вызванная функция находит значения вероятностей $P(X > x)$, которые заполняют, например, ячейки C1-C15. Поскольку нам нужна функция распределения вероятностей

$F(x)=P(X\leq x)=1-P(X>x)$, то ее значения мы получим в ячейках D1-D15 следующим образом: D1=1-C1,...,D15=1-C15.

Для определения значений плотности распределения вероятностей необходимо воспользоваться формулой для приближенного вычисления значений первой производной.

Далее строим график плотности распределения F_{n_1, n_2} и изучаем влияние степеней свободы n_1 и n_2 на вид графика. Для этого фиксируем значение n_1 и уменьшаем или увеличиваем значение n_2 . После этого фиксируем значение n_2 и уменьшаем или увеличиваем значение n_1 . После исследования делаем соответствующие выводы.

5. Оборудование

Персональный компьютер с установленной операционной системой Windows XP/7/8, браузер (Например, Internet Explorer, Google Chrome, Opera), OoO Writer (MS Word), OoO Calc (MS Excel) пакет офисных приложений «Мой офис».

6. Задание на работу

Используя MSExcel (OoO Calc) выполнить задания:

1. Предприятие изготавливает трубы, средний внешний диаметр которых равен 20,20 мм, а стандартное отклонение равно 0,25 мм, Согласно ТУ трубы признаются годными, если диаметр находится в пределах 20,00+/-0,40 мм. Какая доля изготовленных труб соответствует ТУ? Обработать данные, построить график плотности нормального распределения. Сделать выводы.

2. Директор школы хочет узнать, действительно ли то, что учителя более предвзято относятся к мальчикам, чем к девочкам, т.е. более склонны хвалить девочек. Для этого им были проанализированы характеристики учеников, написанные учителями, на предмет частоты встречаемости трех слов: «пассивный», «активный», «старательный», «дисциплинированный», синонимы слов так же подсчитывались. Данные о частоте встречаемости слов были занесены в таблицу:

	Пассивный	Активный	Старательный	Дисциплинированный
Мальчики	10	5	6	9
Девочки	6	12	8	7

Обработать данные с использованием критерия хи-квадрат. Сделать выводы.

3. Необходимо сравнить между собой результаты выполнения логических задач до и после курса обучения дисциплины «Математическая статистика». Чтобы узнать различаются ли результаты до курса обучения и после необходимо вычислить критерий Стьюдента, построить график и сделать выводы по этой задаче

№	Результаты выполнения логических задач до курса (мин.)	Результаты выполнения логических задач после курса (мин.)
1	10	9

2	7	8
3	8	7
4	10	9
5	11	12
6	9	9
7	9	8
8	10	7
9	7	7

4. В двух третьих классах проводилось тестирование умственного развития по тесту ТУРМШ десяти учащихся. Полученные значения величин средних достоверно не различались, однако психолога интересует вопрос – есть ли различия в степени однородности показателей умственного развития между классами. Обработать данные с использованием критерия Фишера. Сделать выводы.

№ учащегося	Первый класс А	Второй класс Б
1	90	41
2	29	49
3	39	56
4	79	64
5	88	72
6	53	65
7	34	63
8	40	87
9	75	77
10	79	62

7. Контрольные вопросы

1. Что такое плотность нормального распределения и в каких случаях она используется?
2. Критерий хи-квадрат: дать определение и рассказать в каких случаях он используется.
3. Критерий Стьюдента: дать определение и рассказать в каких случаях он используется.
4. Критерий Фишера: дать определение и рассказать в каких случаях он используется.

Используемая литература

Основная литература

1. Энатская Н. Ю. Теория вероятностей и математическая статистика для инженерно-технических направлений: Учебник и практикум / Энатская Н. Ю., Хакимуллин Е. Р. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 399.
2. Гмурман В. Е. Теория вероятностей и математическая статистика: Учебник / Гмурман В. Е. – 12-е изд. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 479. – (Высшее образование). – Режим доступа: <https://www.biblio-online.ru>.
3. Севастьянов Б. А. Курс теории вероятностей и математической статистики / Б. А. Севастьянов. – Москва, Ижевск: Институт компьютерных исследований, 2019. – 272 с.
4. Махова Н. Б. Теория вероятностей и основы математической статистики: курс лекций / Н. Б. Махова. – Москва: Московская государственная академия водного транспорта, 2019. – 87 с.
5. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Саратов: Ай Пи Ар Букс, 2015. – 158 с.

Дополнительная литература

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для вузов / В.Е. Гмурман. – 11-е изд., перераб. – М.: Высш. образование, 2008. – 404с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 12-е изд., перераб. – М.: Высш. образование, 2008. – 479с.
3. Симогин А. А. Специальные разделы высшей математики. Теория вероятностей и математическая статистика для инженера-исследователя: учебное пособие / А. А. Симогин. – Макеевка: Донбасская национальная академия строительства и архитектуры, ЭБС АСВ, 2019. – 169 с.
4. Энатская Н. Ю. Теория вероятностей и математическая статистика для инженерно-технических направлений: Учебник и практикум / Энатская Н. Ю., Хакимуллин Е. Р. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 399.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

1. Электронный читальный зал “БИБЛИОТЕХ”: учебники авторов ТулГУ по всем дисциплинам.- Режим доступа: <https://tsutula.bibliotech.ru/>
2. Единое окно доступа к образовательным ресурсам: портал [Электронный ресурс]. - Режим доступа : <http://window.edu.ru>.
3. ЭБС IPRBooks универсальная базовая коллекция изданий.- Режим доступа: <http://www.iprbookshop.ru/>

4. Научная Электронная Библиотека eLibrary – библиотека электронной периодики, режим доступа: <http://elibrary.ru/>