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»  «29» января 2019г., протокол № 6 |
| 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 профилем

**«Программное обеспечение средств вычислительной техники и автоматизированных систем»**

Форма обучения: очная

Идентификационный номер образовательной программы: 090301-01-18

Тула 2019 год**Разработчик(и) методических указаний**

\_\_\_\_Набродова И.Н., доцент, к.т.н.\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО, должность, ученая степень, ученое звание) (подпись)

**Оглавление**

[Лабораторная работа №1 4](#_Toc48151987)

[СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ 4](#_Toc48151988)

[Лабораторная работа №2 11](#_Toc48151989)

[ОСНОВНЫЕ РАСПРЕДЕЛЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ 11](#_Toc48151990)

[Лабораторная работа №3 19](#_Toc48151991)

[ОПРЕДЕЛЕНИЕ ИНТЕРВАЛЬНЫХ ОЦЕНОК НЕИЗВЕСТНЫХ](#_Toc48151992) [ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ 19](#_Toc48151993)

[Лабораторная работа №4 22](#_Toc48151994)

[КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ 22](#_Toc48151995)

[Лабораторная работа №5 31](#_Toc48151996)

[ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ 31](#_Toc48151997)

**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**

- ознакомится 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

x1, x2, ..., xn (1)

последовательность случайных величин

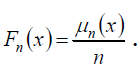
X1, X2, ..., Xn. (2)

Предполагается, что случайные величины X1, X2, ..., Xn 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 Fn (x), которая служит оценкой теоретической функции распределения вероятностей F(x). Если обозначить через μn(x) число тех значений x1, x2, ..., xn, которые меньше или равны x, то

 (3)

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

Вводим в первый столбец (ячейки А1…) 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 (В1) слово Максимум, а во вторую строку второго столбца (В2) слово Минимум. В соседних ячейках С1 и С2 определим функции МАХ и MIN. Для этого ставим курсор в С1 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 А1…). Поле «Число 2» оставляем пустым. Нажимаем «ОК». Ставим курсор в ячейку С2 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 ni используется функция ЧАСТОТА, находящаяся в категории СТАТИСТИЧЕСКИЕ. Введем ее в ячейку Е1. В строке «Массив данных» введем диапазон выборки (ячейки А1…). В строке «Массив интервалов» введем диапазон верхних границ интервалов группировки (ячейки D1…). Результат функции является массивом и выводится в ячейках Е1... Для полного вывода (не только первого числа в Е1) нужно выделить ячейки Е1…, обведя их мышью, и нажать F2, а далее одновременно CTRL+SHIFT+ENTER. Результат – частоты ni, 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:

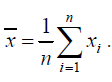
1. Выделенный фрагмент должен состоять из равновеликих столбцов.

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)



Функция ДИСП вычисляет выборочную дисперсию (оценку теоретической дисперсии)



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 в выборке объема 2n+1 берут значение x(n+1) в вариационном ряде. Если объем выборки равен 2n, то в качестве выборочной медианы  берут 

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), Ооо Calc (MS Excel) пакет офисных приложений «Мой офис».

**6. Задание на работу**

Выборка состоит из 50 значений некоторой случайной величины. 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 Х | 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 |
| Х | 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 |
| Х | 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 |
| Х | 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 Х | 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 |
| Х | 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 |
| Х | 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 |
| Х | 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 Х | -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 |
| Х | -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 |
| Х | 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 |
| Х | -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 Х | -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 |
| Х | -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 |
| Х | -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 |
| Х | 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 Х | 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 |
| Х | -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 |
| Х | -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 |
| Х | -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 Х | -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 |
| Х | 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 |
| Х | -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 |
| Х | -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 Х | 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 |
| Х | 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 |
| Х | 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 |
| Х | 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 Х | -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 |
| Х | -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 |
| Х | -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 |
| Х | 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 Х | 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 |
| Х | 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 |
| Х | 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 |
| Х | 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 Х | 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 |
| Х | 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 |
| Х | -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 |
| Х | -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**

- ознакомится с теоретическими сведениями;

- выполнить задание;

- оформить отчет;

- ответить на контрольные вопросы, заданные преподавателем.

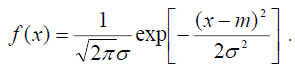
**3. Оформление отчета**

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 статических исследованиях широко используются случайные величины, имеющие нормальное распределение, распределение χ2 (хи-квадрат), распределения Стьюдента и Фишера.

Случайная величина X имеет нормальное распределение с параметрами m (математическое ожидание) и σ2 (дисперсия), если плотность распределения имеет вид:



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

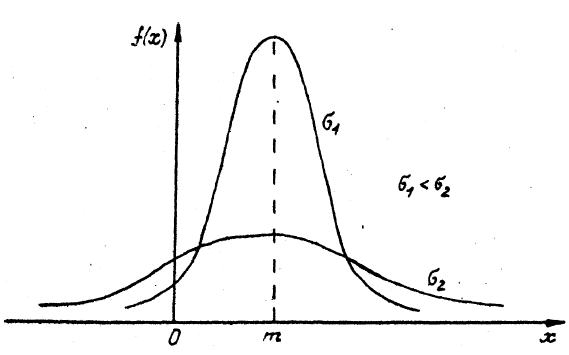
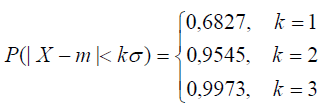


Рисунок 1 – Графики плотностей нормального распределения с одинаковым математическим ожиданием m и разными дисперсиями σ12, σ22

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:



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 (А1) и второй строки первого столбца (А2) вводим подписи m= и sig=, а в первую строку второго столбца (В1) и во вторую строку второго столбца (В2) вводим их числовые значения. Для построения графика протабулируем в третьем и четвертом столбцах (соответственно С 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 С1 надпись Х=, а в D1 надпись f=. Вводим в С2 значение *a*, а в С3 значение a+h. После этого обводим, выделяя, ячейки С2 и С3 и захватив за нижний правый угол рамки вокруг ячеек С2 и С3, перетягиваем его вниз до ячейки С(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 fx и выбираем категорию СТАТИСТИЧЕСКИЕ и функцию НОРМРАСП.

Вводим ссылкой на переменную Х: «С2» (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 В1 и В2, поэтому пишется знак $ (чтобы быстро относительную ссылку сделать абсолютной нужно после ввода ссылки нажать F4). В поле «Интегральное» ставим 0 или «ложь», нажимаем «ОК».

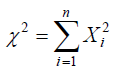
В ячейке D2 появляется результат f (a) – значение плотности нормального распределения, а в строке формул – запись =НОРМРАСП(C2;$B$1;$B$2;ложь). За нижний правый угол ячейки D2 автозаполняем результат на ячейки D2-D(2+(b−a)/h). Если требуется построить функцию распределения вероятностей нормального распределения, то в поле «Интегральное» нужно поставить a 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 В1 и нажимаем «Enter». График плотности нормального распределения должен сместиться вправо. Теперь уменьшаем значение математического ожидания m в ячейке В1 и нажимаем «Enter». График плотности нормального распределения должен сместиться влево. Возвращаем в В1 первоначальное значение m и начинаем изучать влияние среднеквадратического отклонения σ (или дисперсии σ2) на график плотности. Увеличивая значение σ в ячейке В2, можно наблюдать растяжение графика. Уменьшая значение σ, можно наблюдать сжатие графика.

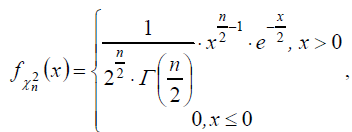
После такого исследования можно сделать соответствующие выводы о влиянии параметров m и σ на вид графика плотности нормального распределения.

Пусть X1, X2,..., Xn – независимые случайные величины с общим распределением N(0,1) (нормальное распределение с нулевым математическим ожиданием и единичной дисперсией). Тогда случайная величина

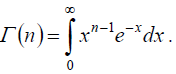


имеет распределение χ2 с n степенями свободы или распределение χn2.

Плотность распределения вероятностей χn2 имеет вид:



где



Графики плотности распределения χ2 с n степенями свободы асимметричны и, начиная с n = 2, имеют по одному максимуму в точке x=n−2 (рис. 2). Причем с ростом n кривая плотности приближается к симметричной функции.

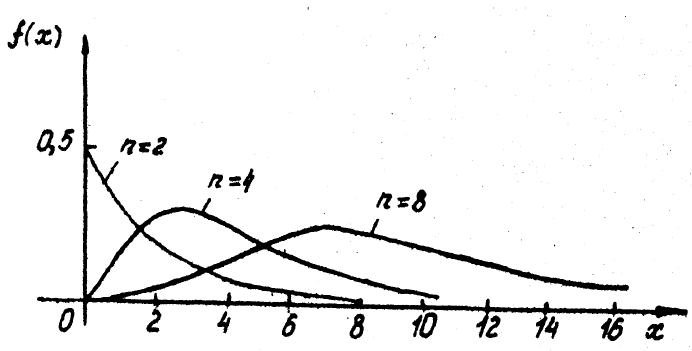


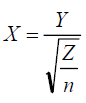
Рисунок 2 – Графики плотности распределения χ 2 с n степенями свободы

Интервал для построения необходимо выбрать с нулевой левой границей и произвести его разбиение. Пусть точки разбиения занимают, например, ячейки В1-В15. Для того, чтобы найти значение функции распределения вероятностей χ2 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 С1-С15. Поскольку нам нужна функция распределения вероятностей F(x)=P(X≤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 χ2 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 Е2 по формуле Е2=(D2-D1)/h. Остальные ячейки автозаполняем. Таким образом, можно найти значения плотности χ2 с n степенями свободы на выбранном отрезке, кроме нуля (левой границы отрезка).

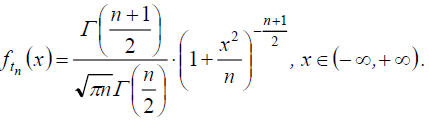
Действуя аналогично построению графика плотности нормального распределения, строим график плотности распределения χ 2 с n степенями свободы и изучаем влияние степени свободы на вид графика, уменьшая или увеличивая n.

После исследования делаем соответствующие выводы.

Пусть случайная величина Y имеет распределение N(0,1), а независимая от Y случайная величина Z принадлежит χn2 (имеет распределение χ 2 с n степенями свободы). Тогда случайная величина



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



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

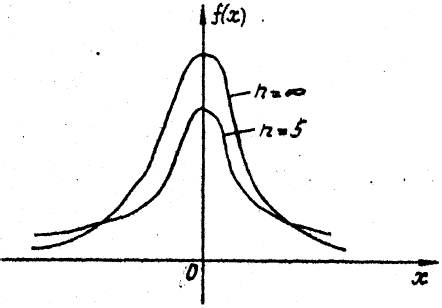


Рисунок 3 – Графики плотности распределения Стьюдента

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

Интервал для построения функции распределения Стьюдента tn необходимо выбрать симметричным относительно нуля. Разбиение построить так, чтобы среди точек разбиения был нуль. Пусть точки разбиения занимают, например, ячейки В1-В15, среди них ячейки В1-В7 содержат отрицательные значения, ячейка В8 содержит нуль и ячейки В9-В15 содержат положительные значения.

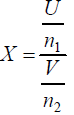
Для того, чтобы найти значения функции распределения Стьюдента n t с n степенями свободы, нужно нажать кнопку мастера функций fx, выбрать категорию СТАТИСТИЧЕСКИЕ и функцию СТЬЮДРАСП. Функция имеет дополнительный чисто вычислительный параметр «Хвосты», который не связан с распределением Стьюдента, а 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 С9-С15. В ячейке С9 обязательно должно быть значение 0.5, поскольку это вероятность P(X>0) = 0.5. Поскольку нам нужна функция распределения вероятностей F(x) = P(X ≤ x) =1− P(X > x), то ее значения мы получим в ячейках D9-D15 следующим образом: D9=1-C9,…,D15=1-C15. Значения функции распределения в ячейках D1-D7 получим по формуле: D1=C15, D2=C14,….D7=C9, поскольку функция распределения Стьюдента из-за симметричности распределения обладает свойством P(X ≤ −x) = P(X > x).

Для нахождения значений плотности распределения Стьюдента tn необходимо воспользоваться формулой для приближенного вычисления значений первой производной (по аналогии с распределением χ2 ).

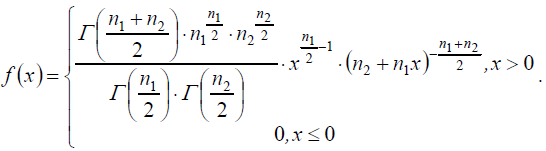
Далее строим плотность распределения вероятностей tn с n степенями свободы (значение берется произвольно) и изучаем влияние степени свободы на вид графика, уменьшая или увеличивая значение n.

После исследования делаем соответствующие выводы.

Пусть U и V - независимые случайные величины, распределенные по закону χ2 с n1 и n2 степенями свободы соответственно. Тогда случайная величина



имеет распределение Фишера с n1 и n2 степенями свободы (Fn1,n2 – распределение) и плотностью распределения вероятностей



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

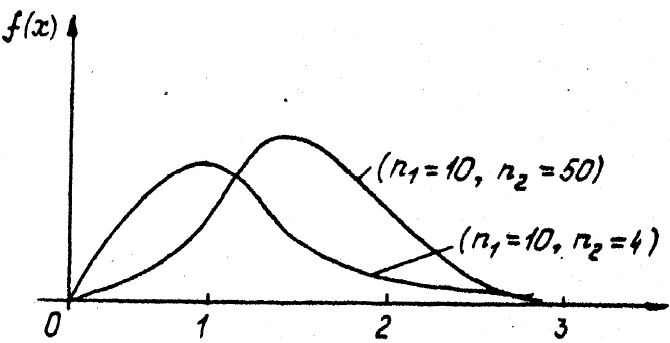


Рисунок 4 – Графики плотности распределения Фишера

Интервал для построения необходимо выбрать с нулевой левой границей. Пусть точки разбиения занимают, например, ячейки В1-В15. Для того, чтобы найти значения функции распределения Фишера Fn1,n2 с n1и n2 степенями свободы, нужно нажать кнопку мастера функций fx , выбрать категорию СТАТИСТИЧЕСКИЕ и функцию FРАСП. Вызванная функция находит значения вероятностей P(X>x), которые заполняют, например, ячейки С1-С15. Поскольку нам нужна функция распределения вероятностей F(x)=P(X≤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 Fn1,n2 и изучаем влияние степеней свободы n1и n2 на вид графика. Для этого фиксируем значение n1 и уменьшаем или увеличиваем значение n2. После этого фиксируем значение n2 и уменьшаем или увеличиваем значение n1. 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), Ооо Calc (MS Excel) пакет офисных приложений «Мой офис».

**6. Задание на работу**

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.

**Лабораторная работа №3**

**ОПРЕДЕЛЕНИЕ ИНТЕРВАЛЬНЫХ ОЦЕНОК НЕИЗВЕСТНЫХ**

**ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ**

**1. Цель и задачи работы**

Научиться определять интервальные оценки с заданным уровнем доверия (доверительные интервалы) для неизвестных параметров распределений.

**2. Порядок выполнения работы**

- ознакомится 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**

В ряде задач требуется не только найти для параметра θ подходящую оценку, но и указать, к каким ошибкам может привести замена параметра его оценкой. Другими словами, требуется оценить точность и надежность оценки. Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка в значительной мере случайна и приближенная замена этой оценкой истинного значения параметра θ может привести к серьезным ошибкам.

Для определения точности оценки θ в математической статистике пользуются доверительными интервалами, а для определения надежности – доверительными вероятностями.

Интервал (θ1,θ2) называется α – доверительным интервалом или доверительным интервалом с доверительной вероятностью 1−α, если



Рассмотрим задачу построения доверительных интервалов для неизвестных математического ожидания и дисперсии. Как известно, α – доверительный интервал для математического ожидания a выглядит следующим образом:



где s1 – среднеквадратическое отклонение (корень квадратный из исправленной выборочной дисперсии), Δn−1,α находится из таблицы для вероятностей P(tn−1>Δn−1,α)=α распределения tn−1 (распределение Стьюдента с n-1 степенью свободы).

Вводим в первый столбец, например, ячейки А1…А25 исходные данные. Задаем уровень значимости α=0.05. Далее для получения результатов подписываем ячейки, как на рис. 1.



Рисунок 1 – Пример подписи ячеек для получения результатов

Для вычисления величины служит функция ДОВЕРИТ категории СТАТИСТИЧЕСКИЕ с тремя параметрами «Альфа» – уровень значимости α, «Станд. откл» – s1 – среднеквадратическое отклонение (корень квадратный из исправленной выборочной дисперсии), «Размер» – объем выборки n.

Таким образом, вводим в ячейку Н3 функцию:

=СРЗНАЧ(А1:А25)-ДОВЕРИТ(I1;СТАНДОТКЛОН(А1:А25);25)

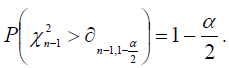
а в ячейку I3 функцию:

=СРЗНАЧ(А1:А25)+ДОВЕРИТ(I1;СТАНДОТКЛОН(А1:А25);25).

Как известно, α – доверительный интервал для дисперсии σ2 выглядит следующим образом:



Нам потребуется функция ХИ2ОБР (категория СТАТИСТИЧЕСКИЕ), которая вычисляет обратное значение xn−1,p односторонней вероятности распределения хи-квадрат P( χ2n−1>xn−1,p)=p.

В данном конкретном случае 

ХИ2ОБР имеет два параметра: первый «Вероятность» содержит доверительную вероятность соответственно α/2 и 1−α/2, второй – степень свободы n-1.

Вводим в соответствии с формулой для доверительного интервала σ2 в ячейку Н4 запись:

=ДИСП(A1:A25)\*24/ХИ2ОБР(0,025;24),

а в ячейку I4 запись:

=ДИСП(A1:A25)\*24/ХИ2ОБР(0,975;24).

Получаем значения границ доверительных интервалов для σ2.

**5. Оборудование**

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

**6. Задание на работу**

1. Станок производит детали, измерения которых приведено ниже. С доверительной вероятностью 0.95 построить доверительные интервалы для математического ожидания и дисперсии размера деталей.

**Вариант 1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 43.8 | 43.9 | 46.3 | 44.6 | 47.5 | 42.0 | 44.5 | 45.0 | 46.8 | 45.3 | 41.8 | 42.3 | 47.9 | 45.5 | 44.4 | 43.1 | 42.8 |
| 41.9 | 42.8 | 46.0 | 45.3 | 41.8 | 42.3 | 47.9 | 45.5 | 46.3 | 44.6 | 47.5 | 42.0 | 44.5 | 43.8 | 43.9 | 46.3 | 44.6 |

**Вариант 2**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 49.0 | 48.8 | 49.2 | 50.2 | 49.5 | 49.8 | 49.9 | 49.3 | 49.6 | 49.5 | 49.7 | 49.0 | 48.8 | 51.8 | 49.1 | 48.3 | 50.0 |
| 49.0 | 48.4 | 48.5 | 49.6 | 49.5 | 49.7 | 49.0 | 48.8 | 51.8 | 49.5 | 49.8 | 49.9 | 49.3 | 49.6 | 48.8 | 49.2 | 50.2 |

**Вариант 3**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 42.1 | 41.9 | 42.3 | 43.1 | 42.5 | 42.7 | 42.9 | 42.3 | 42.6 | 42.5 | 42.7 | 42.0 | 41.9 | 44.5 | 42.2 | 41.5 | 42.9 |
| 42.1 | 41.5 | 41.6 | 41.9 | 42.3 | 43.1 | 42.5 | 42.7 | 42.9 | 42.3 | 42.6 | 42.5 | 42.7 | 42.0 | 42.3 | 43.1 | 42.5 |

**Вариант 4**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 25.6 | 25.5 | 25.5 | 24.7 | 25.0 | 25.8 | 25.2 | 25.0 | 25.0 | 25.3 | 25.4 | 25.1 | 25.2 | 25.3 | 24.8 | 25.1 | 24.6 |
| 25.2 | 25.8 | 24.8 | 25.5 | 24.7 | 25.0 | 25.8 | 25.2 | 25.0 | 25.0 | 25.3 | 25.4 | 25.1 | 25.2 | 25.0 | 25.0 | 25.3 |

**Вариант 5**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 51.2 | 49.9 | 50.7 | 52.3 | 51.3 | 51.2 | 52.4 | 51.6 | 51.5 | 51.0 | 51.8 | 50.9 | 50.7 | 52.0 | 50.2 | 51.1 | 51.0 |
| 51.2 | 52.0 | 49.9 | 50.7 | 52.3 | 51.3 | 51.2 | 52.4 | 51.6 | 51.5 | 51.0 | 51.8 | 50.9 | 50.7 | 52.4 | 51.6 | 51.5 |

**Вариант 6**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 21.8 | 21.6 | 21.7 | 20.8 | 20.8 | 20.8 | 21.2 | 20.7 | 19.8 | 22.2 | 21.9 | 21.5 | 20.9 | 20.9 | 20.9 | 21.6 | 21.2 |
| 22.2 | 20.6 | 21.7 | 20.8 | 20.8 | 20.8 | 21.2 | 20.7 | 19.8 | 22.2 | 21.9 | 21.5 | 20.9 | 20.9 | 20.9 | 21.8 | 21.6 |

**Вариант 7**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 22.7 | 25.7 | 25.5 | 25.2 | 22.7 | 23.6 | 26.4 | 24.2 | 23.7 | 23.5 | 24.7 | 24.8 | 24.0 | 24.3 | 24.7 | 23.1 | 24.0 |
| 22.1 | 24.4 | 26.5 | 25.7 | 25.5 | 25.2 | 22.7 | 23.6 | 26.4 | 24.2 | 23.7 | 23.5 | 22.7 | 25.7 | 25.5 | 25.2 | 22.7 |

**Вариант 8**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 49.5 | 49.5 | 46.7 | 47.2 | 49.1 | 48.7 | 51.2 | 46.1 | 50.5 | 48.9 | 49.3 | 50.4 | 47.2 | 48.5 | 49.4 | 48.8 | 50.1 |
| 46.5 | 51.2 | 46.2 | 46.1 | 50.5 | 49.5 | 48.7 | 51.2 | 47.2 | 49.1 | 49.5 | 46.7 | 47.2 | 49.1 | 48.7 | 47.2 | 49.1 |

**Вариант 9**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 31.6 | 31.4 | 31 | 31.4 | 33.1 | 33.0 | 32.9 | 31.4 | 31.9 | 33.5 | 32.3 | 32.0 | 31.9 | 32.6 | 32.7 | 32.2 | 32.3 |
| 32.6 | 31.6 | 32.1 | 32.9 | 31.4 | 31.9 | 33.5 | 32.3 | 32.0 | 31.6 | 31.4 | 31 | 31.4 | 33.1 | 33.0 | 32.9 | 31.4 |

**Вариант 10**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 36.2 | 35.9 | 36.1 | 34.5 | 34.5 | 34.6 | 35.2 | 34.3 | 32.8 | 36.9 | 36.4 | 35.8 | 34.6 | 34.7 | 34.8 | 35.9 | 35.3 |
| 37 | 34.3 | 36.0 | 36.2 | 35.9 | 36.1 | 34.5 | 34.5 | 34.6 | 35.2 | 34.3 | 32.8 | 36.9 | 36.4 | 34.5 | 34.6 | 35.2 |

**7. Контрольные вопросы**

1. Дать определение интервальной оценки, что позволяют установить интервальные оценки?

2. Дать определение доверительной вероятности, каковы наиболее часто задаваемые ее значения?

3. Дать определение доверительного интервала. Кто разработал метод доверительных интервалов?

**Лабораторная работа №4**

КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

**1. Цель и задачи работы**

Научиться аппроксимировать набор наблюдений линейной функцией.

**2. Порядок выполнения работы**

- ознакомится с теоретическими сведениями;

- выполнить задание;

- оформить отчет;

- ответить на контрольные вопросы, заданные преподавателем.

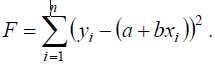
**3. Оформление отчета**

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

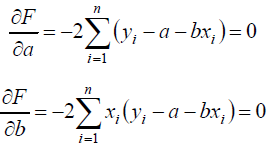
**4. Теоретические сведения**

Цель регрессионного анализа состоит в определении общего вида уравнения регрессии, построении оценок неизвестных параметров, входящих в уравнение регрессии, и проверке статистических гипотез о регрессии. В зависимости от формы связи между переменными различают линейную и нелинейную регрессию. Наиболее простым является случай, когда регрессия линейна.

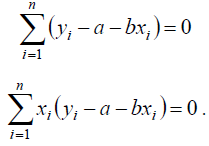
Рассмотрим задачу наилучшей аппроксимации набора наблюдений  линейной функцией *f(X)=a+bX* в смысле минимизации функционала



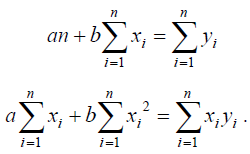
Запишем необходимые условия экстремума:



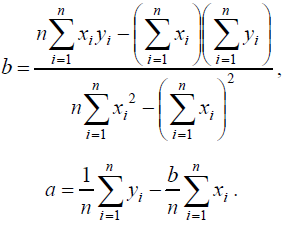
или



Раскрыв скобки, получим:



Решая систему уравнений, находим неизвестные *a* и *b*:



Добавим к постановке задачи некоторые статистические данные и запишем линейное регрессионное уравнение в виде:



где *X*i – неслучайная (детерминированная) величина, Yi, εi – случайные величины, *εi* – ошибки регрессии.

Основные гипотезы:

1.  – спецификация модели.

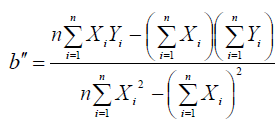
2. *Xi* – детерминированная величина; вектор *(X1, X2, ..., Xn)* не коллинеарен вектору (1,1,...,1).

3. *M(εi)=0, D(εi)=σ2,  M(εi ,εj)=0, i ≠ j*.

Часто добавляется условие:

4. *εi* – нормально распределенная случайная величина, *M(εi) = 0, D(εi)=σ2*.

Как утверждает теорема Гаусса-Маркова, в этих предположениях оценки неизвестных параметров модели

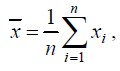
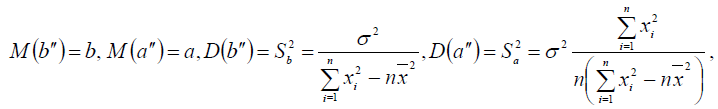


и

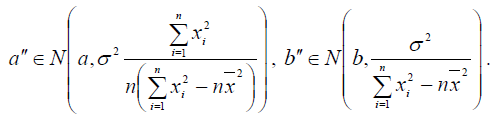


полученные по МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Непосредственно из 1) - 4) следует, что *Yi* – нормально распределенная случайная величина, *M(Yi)=a+bXi, D(Yi)=σ2*.

Нетрудно проверить, что

поэтому



Обозначим через  разницу между действительным значением переменной *Y* и модельным значением этой переменной, то есть

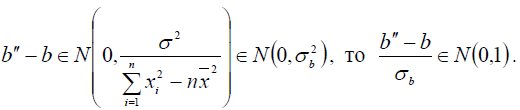
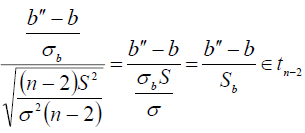


Несмещенной оценкой дисперсии ошибок σ2 является:



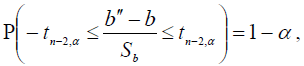
Нетрудно показать, что *S2* независима с *a′′ и b′′*, 

Построим статистику для проверки гипотезы *H0: b=b0* против альтернативной гипотезы *H1: b ≠ b0*.

Поскольку Из условия , следует что  (распределению Стьюдента с *n−2* степенями свободы).

Таким образом, для проверки гипотезы *H0: b=b0* против альтернативной гипотезы *H1: b ≠ b0* будет использоваться статистика 

Построим доверительный интервал для b, используя распределение *tn−2* и его двусторонние квантили *tn−2,α*, которые находятся из таблицы для вероятностей *Ρ(|tn−2|≤ tn−2,α) =1−α*, или Ρ(|tn−2|> tn−2,α)=α:



откуда следует 

Если b0 принадлежит отрезку *[b′′−tn−2,α Sb, b′′+ tn−2,α Sb]*, то принимается гипотеза H0 , в противном случае принимается гипотеза H1.

Если требуется проверить наличие связи между переменными *X* и *Y*, то используется статистика *b′′/Sb*, тем самым проверяется равенство нулю коэффициента *b*. Если в границы построенного при этом доверительного интервала попадает ноль, (то есть нижняя граница доверительного интервала отрицательна, а верхняя положительна), то коэффициент *b* принимается равным нулю и делается вывод об отсутствии связи между переменными *X* и *Y*. Другими словами, при делается вывод о достоверной связи между переменными X и Y, при  делается вывод об ее отсутствии.

Можно показать, что

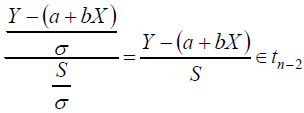


и использовать эту статистику для проверки аналогичных гипотез относительно коэффициента *a*.

Рассмотрим статистику , которая принадлежит стандартному нормальному распределению – *N(0,1)*. При известной *σ2* (дисперсии ошибок) можно было бы использовать *N(0,1)* для прогнозирования значений *Y* в виде доверительных интервалов.

Поскольку σ2 неизвестно, то будем использовать ее оценку S2, для которой известно, что 

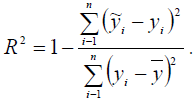
Таким образом,



и используется для построения доверительных интервалов с целью прогнозирования значений *Y*:



Количественным показателем качества построенной линейной модели является коэффициент детерминации



Коэффициент детерминации показывает, какая доля общей дисперсии *Y* объясняется уравнением регрессии:

*0 ≤ R2 ≤1.*

Чем ближе R2 к 1, тем лучше построенная регрессионная модель согласуется с исходными данными.

Для построения регрессии в Excel, создаем файл исходных входных и выходных данных и начинаем с построения корреляционного поля, позволяющего визуализировать наличие связи между этими данными.

Выбираем меню ВСТАВКА/ДИАГРАММА, тип диаграммы: ТОЧЕЧНАЯ вид: ТОЧЕЧНАЯ ДИАГРАММА. Нажимаем кнопку ДАЛЕЕ. В появившемся диалоговом окне указываем диапазон значений и расположение данных: В СТОЛБЦАХ. Нажимаем кнопку ДАЛЕЕ. В следующем диалоговом окне указываем название диаграммы, наименование осей. Нажимаем ДАЛЕЕ и ГОТОВО. Построенная таким образом диаграмма рассеяния представляет собой совокупность пар точек, абсциссами которых являются значения переменной *X*, а ординатами значения переменной *Y*.

В меню СЕРВИС выбираем АНАЛИЗ ДАННЫХ и РЕГРЕССИЯ. Указываем входной интервал *Y* (для примера А2: А26) и входной интервал *X* (для примера B2: B26), а также параметры вывода, остатки, нормальную вероятность как показано на рис. 1.

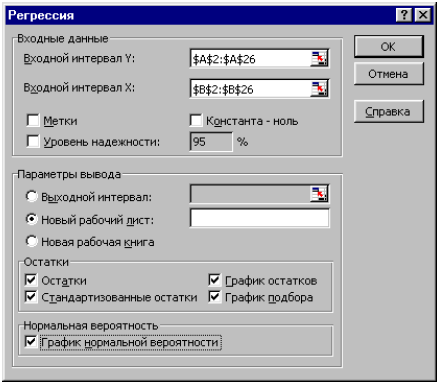


Рисунок 1 – Диалоговое окно Регрессия

В диалоговом окне задаются следующие параметры:

Входной интервал Y – диапазон ячеек, содержащий данные результативного признака;

Входной интервал Х – диапазон ячеек, содержащий данные факторного признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа-ноль – данный флажок необходимо установить, чтобы линия регрессии прошла через начало координат;

Уровень надежности – этот флажок необходимо использовать, если требуется уровень надежности отличный от 95%, принятый по умолчанию;

Выходной интервал – верхняя левая ячейка интервала, в который будут помещаться результаты вычислений.

Excel автоматически сгенерирует результаты по регрессионной статистике. Ниже в качестве примеров приведены возможные результаты и их расшифровки.

Регрессионная статистика

Множественный R 0,969525973

R‐квадрат 0,939980612

Нормированный R‐квадрат 0,935363736

Стандартная ошибка 14,22893673

Наблюдения 15.

Полученное значение коэффициента детерминации говорит об очень хорошей согласованности построенной регрессионной модели и исходных данных (соответственно об очень хорошей связи исследуемых факторов *X* и *Y*).

Результаты дисперсионного анализа будут представлены в виде:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ | | | | | |
|  | df | SS | MS | F | Значимость |
| Регрессия | 1 | 41220,72106 | 41220,72106 | 203,5966782 | 2,55346E‐09 |
| Остаток | 13 | 2632,014326 | 202,4626405 |  |  |
| Итого | 14 | 43852,73538 |  |  |  |

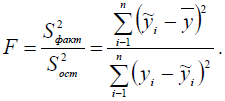
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Коэффи-  циенты | Стандартная  ошибка | t-  статис-  тика | P-  Значение | Нижние  95% | Верхние  95% |
| *Y* – пересечение | 4,746 | 7,003 | 0,678 | 0,510 | ‐10,384 | 19,876 |
| Переменная  *X 1* | 9,595 | 0,672 | 14,269 | 0,000 | 8,142 | 11,048 |

df ⁺! степени свободы (degree of freedom);

SS ⁺! сумма квадратов отклонений (Sum of squares);

MS ⁺! средний квадрат отклонения (Mean square);

F ⁺! отношение дисперсий (факторной к остаточной).



Значимость F – критическое значение квантиля распределения Фишера, которое используется для проверки нулевой гипотезы, состоящей в том, что факторная и остаточная дисперсии равны. По сути дела, нулевая гипотеза означает, что на результативный признак *Y* в равной степени влияют и независимая (факторная) переменная *X* и необъясненные факторы. В таком случае уравнение регрессии не значимо. Чтобы уравнение регрессии было значимым необходимо, чтобы факторная дисперсия превышала остаточную дисперсию в несколько раз.

В примере, приведенном выше, *F* больше, чем Значимость *F* (критическое значение), значит регрессионная модель адекватна.

Регрессионная сумма SS=41220,72106 (объясненная регрессией) намного больше остаточной SS=2632,014326 (не объясненной регрессией, вызванной случайными факторами), что тоже говорит о хорошей регрессии.

Коэффициенты – значения коэффициентов;

Стандартная ошибка – стандартная ошибка коэффициентов;

t-статистика – значение статистики критерия;

Р-значение – уровень значимости отклонения гипотезы равенства коэффициента нулю (вероятность принять равенство коэффициента нулю);

Нижние 95% – нижняя граница доверительного интервала, в котором находится значение коэффициента;

Верхние 95% – верхняя граница доверительного интервала, в котором находится значение коэффициента.

Приведенные в качестве примера результаты позволяют проверить значимость коэффициентов регрессии: свободного члена и коэффициента при переменной *X*. Значение коэффициента при *X* 9,595 больше, чем его стандартная ошибка. К тому же этот коэффициент является значимым, о чем можно судить по значениям показателя Р-значение в таблице, которые меньше заданного уровня значимости *α*=0,05. Для свободного члена ситуация диаметрально противоположная. В построенный для него доверительный интервал попадает ноль, что говорит о том, что он незначим и может быть принят равным нулю.

Есть возможность вывести таблицу стандартных и простых остатков, где для каждого значения ряда выводится предсказанное значение, с которым сопоставляется остаток, представляющий разность между прогнозным и реальным значением.

Простым и наглядным способом проверки удовлетворительности регрессионной модели является графическое представление отклонений, которое Excel представляет в виде графика остатков. Если регрессионная модель близка к реальной зависимости, то отклонения будут носить случайный характер и их сумма будет близка к нулю. Если необходимо получить дополнительную информацию и графики остатков, установите соответствующие флажки в диалоговом окне.

**5. Оборудование**

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

**6. Задание на работу**

Построить уравнение регрессии *Y=a+bx*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вариант 1 | | | Вариант 2 | | | Вариант 3 | | |
| № | X | Y | № | X | Y | № | X | Y |
| 1 | -1.132 | 1.554 | 1 | -0.132 | 1.791 | 1 | -0.332 | -1.65 |
| 2 | -0.204 | 4.601 | 2 | 0.796 | 1.51 | 2 | 0.596 | 2.31 |
| 3 | 0.858 | 2.943 | 3 | 1.858 | 4.17 | 3 | 1.658 | 4.953 |
| 4 | 1.715 | -1.157 | 4 | 2.715 | 3.007 | 4 | 2.515 | 8.62 |
| 5 | 2.494 | -6.048 | 5 | 3.494 | 5.875 | 5 | 3.294 | 7.49 |
| 6 | 4.013 | -1.194 | 6 | 5.013 | 7.187 | 6 | 4.813 | 6.09 |
| 7 | 4.964 | 11.465 | 7 | 5.964 | 9.005 | 7 | 5.764 | 11.958 |
| 8 | 6.167 | -6.257 | 8 | 7.167 | 14.865 | 8 | 6.967 | 14.975 |
| 9 | 7.658 | -11.07 | 9 | 8.658 | 12.008 | 9 | 8.458 | 18.518 |
| 10 | 8.243 | 10.243 | 10 | 9.243 | 11.718 | 10 | 9.043 | 21.794 |
| 11 | 9.296 | 10.995 | 11 | 10.296 | 16.744 | 11 | 10.096 | 17.245 |
| 12 | 10.259 | -11.17 | 12 | 11.259 | 18.789 | 12 | 11.059 | 20.881 |
| 13 | 11.275 | -10.84 | 13 | 12.275 | 12.863 | 13 | 12.075 | 18.787 |
| 14 | 12.202 | -9.78 | 14 | 13.202 | 20.862 | 14 | 13.002 | 16.334 |
| 15 | 12.687 | 15.066 | 15 | 13.687 | 15.309 | 15 | 13.487 | 28.613 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вариант 4 | | | Вариант 5 | | | Вариант 6 | | |
| № | X | Y | № | Х | Y | № | X | Y |
| 1 | -3.132 | 0.412 | 1 | 1.868 | 12.669 | 1 | -4.132 | -4.539 |
| 2 | -2.204 | 2.204 | 2 | 2.796 | -9.23 | 2 | -3.204 | -3.306 |
| 3 | -1.142 | -1.282 | 3 | 3.858 | 0.753 | 3 | -2.142 | -2.018 |
| 4 | -0.285 | -2.529 | 4 | 4.715 | -2.106 | 4 | -1.285 | -1.223 |
| 5 | 0.494 | -2.995 | 5 | 5.494 | -4.044 | 5 | -0.506 | -0.739 |
| 6 | 2.013 | -0.325 | 6 | 7.013 | 1.903 | 6 | 1.013 | -0.823 |
| 7 | 2.964 | -5.018 | 7 | 7.964 | 5.065 | 7 | 1.964 | -1.873 |
| 8 | 4.167 | -6.295 | 8 | 9.167 | 3.807 | 8 | 3.167 | 1.055 |
| 9 | 5.658 | -3.773 | 9 | 10.658 | 3.374 | 9 | 4.658 | 0.989 |
| 10 | 6.243 | -6.772 | 10 | 11.243 | 0.563 | 10 | 5.243 | 0.116 |
| 11 | 7.296 | -5.079 | 11 | 12.296 | 2.511 | 11 | 6.296 | 6.213 |
| 12 | 8.259 | -5.519 | 12 | 13.259 | 1.934 | 12 | 7.259 | 0.856 |
| 13 | 9.275 | -6.505 | 13 | 14.275 | 4.609 | 13 | 8.275 | 0.743 |
| 14 | 10.202 | -7.864 | 14 | 15.202 | 9.673 | 14 | 9.202 | 5.137 |
| 15 | 10.687 | -8.578 | 15 | 15.687 | 6.565 | 15 | 9.687 | 2.209 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вариант 7 | | | Вариант 8 | | | Вариант 9 | | |
| № | X | Y | № | X | Y | № | X | Y |
| 1 | 1.868 | 4.322 | 1 | -1.118 | 7.252 | 1 | 1.132 | 2.573 |
| 2 | 2.796 | 11.126 | 2 | -1.062 | 10.683 | 2 | 3.204 | 14.963 |
| 3 | 3.858 | 4.536 | 3 | 0.054 | 4.075 | 3 | 5.142 | 4.317 |
| 4 | 4.715 | 11.52 | 4 | 2.359 | 5.463 | 4 | 7.285 | 17.39 |
| 5 | 5.494 | 15.855 | 5 | 3.561 | 11.377 | 5 | 9.506 | 26.456 |
| 6 | 7.013 | 19.365 | 6 | 3.56 | 13.401 | 6 | 10.987 | 30.205 |
| 7 | 7.964 | 20.027 | 7 | 6.348 | 12.638 | 7 | 13.036 | 32.238 |
| 8 | 9.167 | 15.986 | 8 | 6.617 | 18.804 | 8 | 14.833 | 25.086 |
| 9 | 10.658 | 23.784 | 9 | 8.108 | 8.975 | 9 | 16.342 | 36.133 |
| 10 | 11.243 | 22.948 | 10 | 8.538 | 26.612 | 10 | 18.757 | 37.62 |
| 11 | 12.296 | 18.02 | 11 | 8.312 | 23.302 | 11 | 20.704 | 29.791 |
| 12 | 13.259 | 31.75 | 12 | 10.895 | 21.718 | 12 | 22.741 | 53.538 |
| 13 | 14.275 | 29.538 | 13 | 10.381 | 20.979 | 13 | 24.725 | 50.432 |
| 14 | 15.202 | 24.758 | 14 | 12.18 | 17.243 | 14 | 26.798 | 43.52 |
| 15 | 15.687 | 27.018 | 15 | 12.973 | 19.821 | 15 | 29.313 | 50.701 |

|  |  |  |
| --- | --- | --- |
| Вариант 10 | | |
| № | X | Y |
| 1 | -1.121 | -4.575 |
| 2 | 0.391 | 3.839 |
| 3 | 0.587 | 1.864 |
| 4 | 2.114 | 2.832 |
| 5 | 3.131 | 17.286 |
| 6 | 4.528 | 7.376 |
| 7 | 4.806 | 6.239 |
| 8 | 6.165 | 17.959 |
| 9 | 7.464 | 13.944 |
| 10 | 7.454 | 17.99 |
| 11 | 9.392 | 27.978 |
| 12 | 9.685 | 22.938 |
| 13 | 11.138 | 25.206 |
| 14 | 11.684 | 26.74 |
| 15 | 12.627 | 36.957 |

**7. Контрольные вопросы**

1. Какова цель регрессионного анализа?

2. Какова формула для вычисления коэффициента детерминации.

3. Каким способом производится проверка удовлетворительности регрессионной модели?

**Лабораторная работа №5**

**ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ**

**1. Цель и задачи работы**

Научиться производить проверку статистических гипотез.

**2. Порядок выполнения работы**

- ознакомится 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**

Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Нулевой (основной) называют выдвинутую гипотезу *Н0*. Конкурирующей (альтернативной) называют гипотезу Н1, которая противоречит нулевой.

Пример. Пусть *Н0* заключается в том, что математическое ожидание генеральной совокупности а=3. Тогда возможные варианты Н1: а) а ≠ 3; б) а > 3; в) а < 3.

Проверка гипотез

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближенное распределение которой известно. Эту величину называют статистическим критерием K. Для проверки гипотезы по данным выборок вычисляют точные значения входящих в критерий величин и таким образом получают наблюдаемое значение критерия Kнабл.

После выбора определенного критерия множество всех его возможных значений разбивают два непересекающихся подмножества: W, содержащее значения критерия, при которых *H0* принимается – область принятия гипотезы, и *W* – при которых нулевая гипотеза *H0* отвергается, – критическая область. При использовании любого критерия возможны ошибки следующих видов:

1) ошибка первого рода – принять гипотезу *H1*, когда верна *H0*;

2) ошибка второго рода – принять гипотезу *H0*, когда верна H1.

Для каждой из них могут быть заданы соответствующие вероятности:





Вероятность ошибки первого рода называется критерием (или уровнем) значимости. Нулевую гипотезу отвергают, если для нее значение *p* ниже уровня значимости, т. е., если *p*<*α*. Обычно назначают условное значение 0,05, тогда шанс допустить ошибку 1-го рода никогда не превысит выбранного уровня значимости, скажем *α*=0,05, так как нулевую гипотезу отвергают только тогда, когда *p*<0,05. Если обнаружено, что *p*>0,05, то нулевую гипотезу не отвергнут и, следовательно, не допустят ошибки 1-го рода. Величина *(1-β)* называется мощностью критерия.

Односторонние и двусторонние критические области.

Точки, отделяющие критическую область от области принятия гипотезы называют критическими точками *kкр*. Различают одностороннюю (правостороннюю и левостороннюю) и двустороннюю критические области:

1. Правосторонняя критическая область определяется неравенством *K*>*kкр*, где *kкр*>0. Например, когда проверяется гипотеза о том, что среднее значение равно 10.

2. Левосторонняя критическая область *K*<*kкр*, где *kкр*<0 .

3. Двусторонняя критическая область *K< k1, K>k2,* где *k2>k1*. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется как *|K |> kкр*.

Для нахождения критической точки *kкр* правосторонней критической области при условии, что справедлива гипотеза *H0*:

1) задаются достаточно малой вероятностью — уровнем значимости α (связанной с доверительной вероятностью β соотношением *α=1− β*);

2) ищут *kкр*, исходя из требования: *P (K >kкр=α)*;

3) по таблицам соответствующих критериев находят k кр и сравнивают с наблюдаемым значением критерия *Kнабл*;

4) если *Kнабл* > *kкр*, то *H0* отвергают; если *Kнабл < kкр* , нет оснований отвергать *H0*.

Отыскание левосторонней и двусторонней критических областей сводится к нахождению соответствующих критических точек:

*P(K< kкр)= α* – требование для левосторонней критической области;

*P(K<k1)+P(K>k2)=α* – требование для двусторонней критической области.

Если выбрать симметричные относительно нуля точки: −*kкр* и *kкр* (*kкр*>0), то получим следующее соотношение для отыскания критических точек двусторонней области:



Для пересчета одностороннего значения уровня значимости в двустороннее значение можно воспользоваться простой формулой: *αдвустор=2⋅αодностор.*

Существенным моментом является то, что статистические методы не позволяют подтвердить гипотезу. Корректное утверждение, которое можно сформулировать в результате анализа, звучит, например: «проведенный анализ не позволяет с заданным уровнем значимости отвергнуть гипотезу».

Проверить гипотезу о равенстве среднего значения заданному можно двумя способами.

Первый способ:

Пусть генеральная совокупность *Х* имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу *m0*. Рассмотрим две возможности.

1) Известна дисперсия *σ2* генеральной совокупности. Тогда по выборке объема *n* найдем выборочное среднее x и проверим нулевую гипотезу *Н0*: *М(Х*)=*m0.* Критерием может служить величина



где *t* – квантиль нормального распределения. Если же дисперсия неизвестна, то используется статистика



имеющая распределение Стъюдента с (n–1) степенью свободы, *t* – квантиль распределения Стьюдента, *s* – среднее квадратическое отклонение выборки.

Найденное значение *t* сравнивается с критическим *tкр*, которое определяется по таблице квантилей в зависимости от уровня значимости или вычисляется встроенными функциями Excel. При попадании выборочного значения в критическую область гипотеза отвергается.

Второй способ:

Для проверки гипотезы о равенстве среднего заданному значению может быть использована функция Z.ТЕСТ, которая вычисляет двухстороннюю вероятность значений z-теста при стандартном распределении.

Синтаксис:

Z.ТЕСТ (Данные; Число; Сигма)

Данные: массив данных.

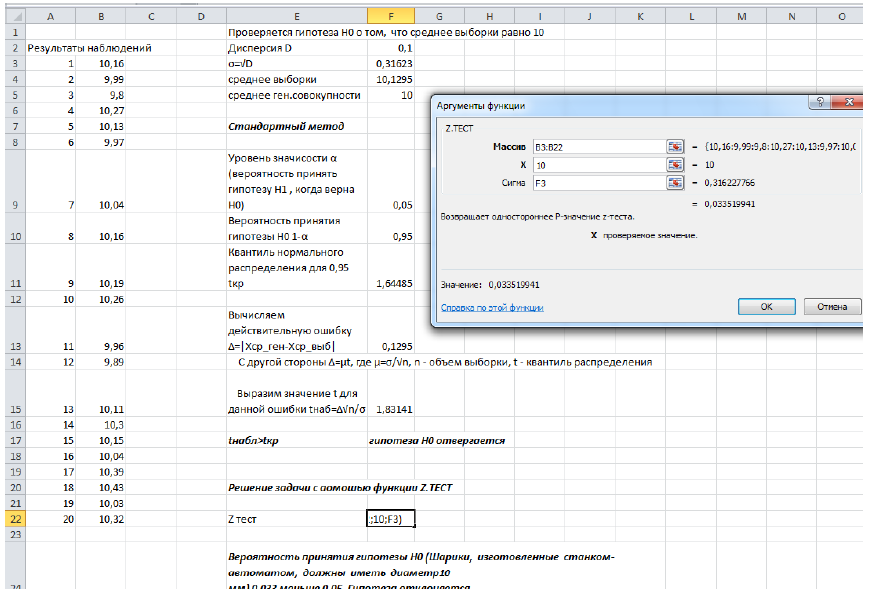
Число: значение для теста.

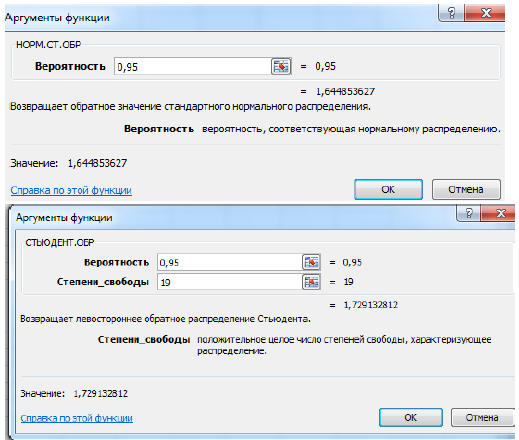
Сигма: необязательное стандартное отклонение генеральной совокупности. Если аргумент не указан, используется стандартное отклонение выборки.

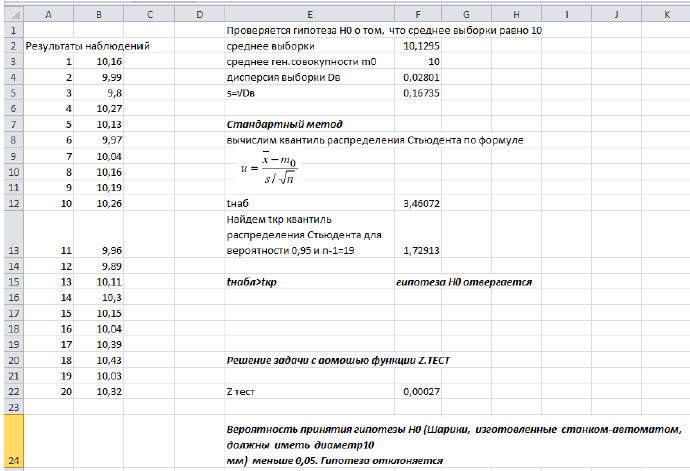
Пример:

=Z.ТЕСТ (A1:A50; 12) вычисляет вероятность того, что значение12 относится к стандартному распределению генеральной совокупности данных в A1:A50.

При использовании этой функции вычисляется вероятность того, что для генеральной совокупности справедлива гипотеза *H0: m=m0*. Используется двухсторонний критерий, то есть альтернативная гипотеза *H1: m≠m0*. Если эта вероятность меньше заданного уровня значимости, гипотеза отклоняется.







Проверка гипотез о равенстве дисперсий

При проверке гипотеза о равенстве дисперсий двух нормально распределенных совокупностей *Н0: σ12=σ22* при неизвестных математических ожиданиях *m1* и *m2* используется статистика

 (1)

которая имеет *F*-распределение Фишера с числом степеней свободы (*n1–1*) и (*n2–1*); здесь *n1* и *n2* – объемы выборок, *s12* и *s22* – соответствующие несмещенные дисперсии; при этом предполагается, что *s12* > *s22*.

Для проверки этой гипотезы в Excel есть функция F.ТЕСТ, которая возвращает результат *F*-теста. Синтаксис функции:

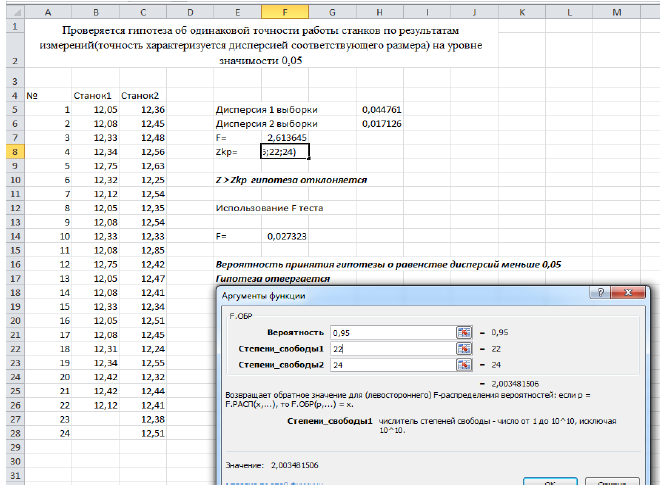
F.ТЕСТ (Данные1; Данные2)

Данные1: первый массив записей.

Данные2: второй массив записей.

Пример:

=F.ТЕСТ (A1:A30; B1:B12) вычисляет различие дисперсий для двух множеств данных и возвращает вероятность того, что оба множества представляют собой выборку из общей совокупности.



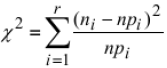
Проверка гипотез о виде распределения

Другой группой статистических гипотез являются гипотезы о проверке вида распределения: неизвестен вид распределения генеральной совокупности, и в частности, неизвестна функция распределения *F(x)*.

Пусть *x1,x2,…,xn* – выборка наблюдений случайной величины *X*. Проверяется гипотеза *Н0* о том, что случайная величина *X* имеет функцию распределения *F(x)*. Разобьем область возможных значений *X* на *r* интервалов Δ1, Δ2, … Δr. Пусть *ni* – число элементов выборки, принадлежащих интервалу Δi(i=1, …, r); при малых значениях *ni* интервалы объединяют таким образом, чтобы в каждом из них было *ni*≥5. Используя предполагаемый закон распределения – с функцией *F(x)*, c учетом оценок параметров этого закона, найденных по выборке, находят вероятности того, что значения *X* принадлежат интервалу Δi, то есть



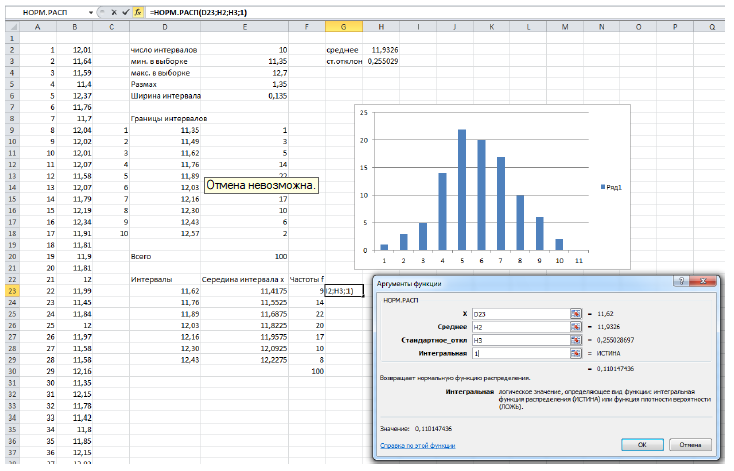
Статистика

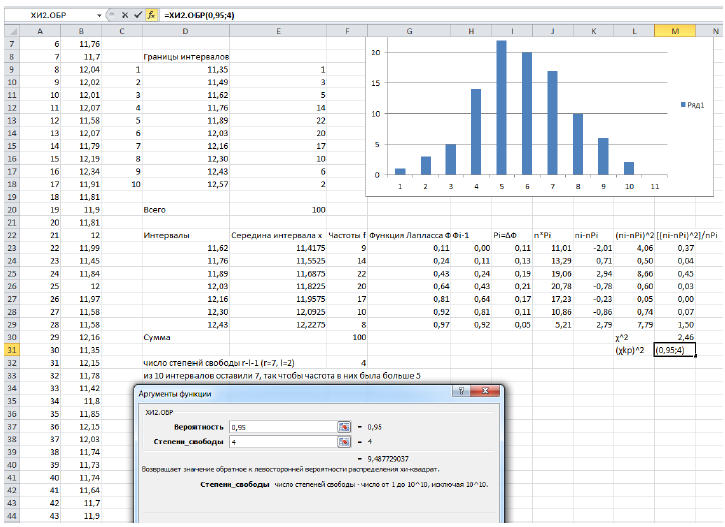


имеет распределение *χ2* с числом степеней свободы (*r–l–1*), где *r* – число интервалов, *l* – число неизвестных параметров распределения. Например, для нормального распределения *l*=2 (неизвестные параметры *m* и *σ*). Считается, что гипотеза *Н0* согласуется с опытом, если

,

где *χ2* – выборочное значение статистики,  – квантиль порядка (*1–α*) распределения *χ2* c числом степеней свободы (*r–l–1*). Рассмотренный метод проверки гипотезы вида распределения называется критерием хи-квадрат или критерием согласия Пирсона.





**5. Оборудование**

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

**6. Задание на работу**

1. Шарики, изготовленные станком-автоматом, должны иметь диаметр 10 мм; проверить эту гипотезу по заданной выборке на уровне значимости 0.05, если:

A) Дисперсия известна и равна 0,1 мм2

B) Дисперсия неизвестна.

Результаты наблюдений приведены в таблице.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 10,16 | 9,99 | 9,8 | 10,27 | 10,13 | 9,97 | 10,04 | 10,16 | 10,19 | 10,26 |
| 9,96 | 9,89 | 10,11 | 10,3 | 10,15 | 10,04 | 10,39 | 10,43 | 10,03 | 10,32 |

Воспользоваться стандартным алгоритмом проверки гипотез и встроенной функцией Z.ТЕСТ.

2. Проверить гипотезу об одинаковой точности работы станков по результатам измерений (точность характеризуется дисперсией соответствующего размера) на уровне значимости 0,05 с использованием формулы (1) и функции F.ТЕСТ. Результаты измерений контролируемого параметра на двух станках приведены в таблице.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Станок1 | Станок2 | № | Станок1 | Станок2 |
| 1 | 12,05 | 12,36 | 13 | 12,05 | 12,47 |
| 2 | 12,08 | 12,45 | 14 | 12,08 | 12,41 |
| 3 | 12,33 | 12,48 | 15 | 12,33 | 12,34 |
| 4 | 12,34 | 12,56 | 16 | 12,05 | 12,51 |
| 5 | 12,75 | 12,63 | 17 | 12,08 | 12,45 |
| 6 | 12,32 | 12,25 | 18 | 12,31 | 12,24 |
| 7 | 12,12 | 12,54 | 19 | 12,34 | 12,55 |
| 8 | 12,05 | 12,35 | 20 | 12,42 | 12,32 |
| 9 | 12,08 | 12,54 | 21 | 12,42 | 12,44 |
| 10 | 12,33 | 12,33 | 22 | 12,12 | 12,41 |
| 11 | 12,08 | 12,85 | 23 |  | 12,38 |
| 12 | 12,75 | 12,42 | 24 |  | 12,51 |

3. Дана выборка из 100 наблюдений; определить числовые характеристики, построить гистограмму частот, проверить нормальность распределения по критерию хи-квадрат.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 12,01 | 12 | 11,64 | 12,09 | 11,79 |
| 11,64 | 11,99 | 11,7 | 11,79 | 12,16 |
| 11,59 | 11,45 | 11,9 | 11,86 | 12,25 |
| 11,4 | 11,84 | 12,15 | 11,79 | 11,92 |
| 12,37 | 12 | 11,93 | 11,98 | 11,72 |
| 11,76 | 11,97 | 12,25 | 11,76 | 11,9 |
| 11,7 | 11,58 | 12,1 | 12,39 | 11,74 |
| 12,04 | 11,58 | 11,81 | 12,13 | 12,09 |
| 12,02 | 12,16 | 11,94 | 12,2 | 11,66 |
| 12,01 | 11,35 | 12,2 | 11,84 | 11,84 |
| 12,07 | 12,15 | 12,1 | 11,52 | 11,84 |
| 11,58 | 11,78 | 11,79 | 11,78 | 11,83 |
| 12,07 | 11,42 | 12,08 | 12,03 | 12,03 |
| 11,79 | 11,8 | 12,7 | 11,65 | 11,96 |
| 12,19 | 11,85 | 12,42 | 11,72 | 12,4 |
| 12,34 | 12,15 | 11,65 | 12,27 | 11,81 |
| 11,91 | 12,03 | 12,16 | 12,11 | 11,92 |

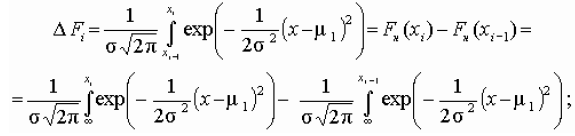
Для выполнения задания необходимо рассчитать и внести в таблицу следующие данные:

*ni* – частота попаданий элементов выборки в i-й интервал;

*xi* – верхняя граница i-го интервала;

*F(xi)* – значение функции нормального распределения;

Δ *Fi* – теоретическое значение вероятности попадания случайной величины в i-й интервал



*Fi*=Δ*Fi*\**n* – теоретическая частота попадания случайной величины в *i*-й интервал; (*ni* – *Fi*)2/*Fi* – взвешенный квадрат отклонения.

**7. Контрольные вопросы**

1. Что называется критерием, уровнем значимости, критической областью и областью допустимых значений критерия?

2. Что такое ошибки первого и второго рода?

3. Что называется мощностью критерия?

4. Сформулируйте этапы проверки статистических гипотез.

5. Проверку какой гипотезы осуществляет функция Excel Z.TEST?