***Лабораторная работа № 5. Первичная статистическая обработка выборочных данных.***

**Цель**: получить навыки статистической обработки выборочных данных, ознакомиться с функциями и графическими средствами ЭТ Excel, применяемыми для статистической обработки данных.

**1. Краткие теоретические сведения**

Математическая статистика - это наука, разрабатывающая методы обработки опытных данных, полученных в результате наблюдений над случайными явлениями.

Основу статистического исследования составляет множество данных, полученных по результатам измерения одного или нескольких признаков. Результаты наблюдений или измерений при этом интерпретируются совокупностью значений, принятых в результате *n* опытов некоторой случайной величиной *Х*. Наблюдаемые значения  называют выборкой наблюдений объема *n* из генеральной совокупности.

Статистическая обработка данных выполняется поэтапно. Основные этапы статистической обработки данных приводятся ниже.

**1. Выборочные данные представляются в виде вариационного ряда**, (*x*1=*xmin*, *x*2, *x*3,…, *xn*=*xmax*) т.е. упорядочиваются в порядке возрастания. Далее определяются значения минимальной и максимальной вариант *xmin* и *xmax* и вычисляется размах вариации *R*= *xmax*- *xmin*.

**2. По упорядоченной статистической совокупности строится статистическая функция** **распределения.** Статистической функцией распределении называется функция , определяющая для каждого выборочного значения *х* относительную частоту события . Тогда по определению

,

где - число наблюдений в выборке, меньших *х*. Если *х* изменяется, то изменяется и относительная частота, т.е. относительная частота  есть функция от *х*. Функция  имеет число скачков равное числу различных выборочных значений. Величина скачка равна «кратности» *li* значения *xi*в выборке, деленной на *n*. Функция  записывается в аналитической форме

 (1)

**3. Построение группированного статистического ряда.** Для этого необходимо весь участок оси абсцисс, на котором расположены выборочные значения случайной величины *Х*, разделить на интервалы или разряды. Число интервалов *k* обычно определяется эмпирическим путем. В частности, , если объем выборки *n*≤100. При этом длина одного интервала вычисляется по формуле: .

Группированный статистический ряд представляется в таблице вида

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *X* |  |  | . . . |  | … |  |
| *Р* |  |  | . . . |  | . . . |  |

В таблице *хi* , *xi*-1 – границы *i*–го интервала, *i*=1,*k*. Пусть  - число выборочных значений, попавших в *i*-й разряд, тогда

, (2)

 - статистическая относительная частота,  - статистическая абсолютная частота.

Статистическая плотность определяется по формуле

. (3)

Графическое изображение , *i*=1,*k* называют гистограммой.

**4.Точечные оценки параметров выборки.**

Оценка математического ожидания случайной величины *Х* вычисляется как среднее арифметическое ее выборочных значений

. (4)

Для получения несмещенной оценки дисперсии случайной величины *Х* используется формула

. (5)

**5. Интервальные оценки** параметров выборки. ***Доверительным*** называется интервал *Iγ*, в который с заданной вероятностью (надежностью) *γ* попадают значения параметра Θ. Вероятность *γ* выбирается близкой к 1: 0,9; 0,95; 0,975; 0,99. При определении интервальных оценок параметров выборки делают предположение на основании центральной предельной теоремы, что выборка имеет нормальное распределение, если объем ее *n*≥20.

Доверительный интервал с надежностью *γ* для математического ожидания имеет вид

 (6)

*tγ,n-*1 – табличное критическое значение распределения Стьюдента *t*-критерий Стьюдента.

Пример. Значение *t*-критерия Стьюдента для выборки объемом *n*=9 и *γ*=0,95 в ЭТ Excel:



Здесь вероятность α=1-*γ* .

Интервал *Iγ*, для дисперсии случайной величины *X* с надежностью *γ* имеет вид

, (7)

Где  и  - значения, взятые из таблицы распределения χ2.

Пример. Считается, что количественный признак X генеральной совокупности распределен нормально. По выборке объема n=25 определена несмещенная дисперсия s2=0,64. Найти доверительный интервал, покрывающий генеральную дисперсию с вероятностью 0,9.

В соответствии с (7) вычислим













**6. Проверка гипотезы о согласии выборочного распределения теоретическому на основе критерия Пирсона.**

Критерий согласия Пирсона (χ2 ) является одним из наиболее часто применяемых критериев.

Алгоритм:

1. Построить интервальный статистический ряд и гистограмму.
2. По виду гистограммы выдвинуть гипотезу:

Н0 – СВ *X* распределена по такому-то закону: f(x)=f0(x),

Н1 - СВ *X* не распределена по такому-то закону: f(x) ≠f0(x), f0(x) - плотность гипотетического закона распределения.

1. Определить оценки неизвестных параметров ** гипотетического закона распределения.
2. Вычислить значение критерия по формуле

, (8)

*pi* - теоретическая вероятность попадания случайной величины в i-й интервал при условии, что гипотеза *H*0 верна:



При расчете p1 и рk в качестве крайних границ первого и последнего интервалов *x*min и *x*max следует использовать теоретические границы гипотетического закона распределения. Например, для нормального закона *x*min=-, *x*max=. После вычисления всех вероятностей *pi* проверить, выполняется ли контрольное соотношение:

.

5. Из таблицы распределения *χ*2 выбирается значение χ2(α,r), где где α – уровень значимости, *r* – число степеней свободы, рассчитанное по формуле:

*r*=*k*-1-*m*.

Здесь *m* - число неизвестных параметров гипотетического закона распределения, значения которых были определены в п. 3.

6. Если значение, вычисленное по формуле (8), больше, чем критическое значение, т.е.χ2> χ2(α,r), то гипотезаН0 отклоняется, в противном случае нет оснований ее отклонять.

**Замечания.**

***1.*** ***Определение теоретических частот для нормального закона распределения N*(*mx,σx*)*.***

*pi*=НОРМРАСП(*xi*+1; ; *s*; ИСТИНА)- НОРМРАСП(*xi*; ; *s*; ИСТИНА), *i*=1,*k.*

Для 1-го интервала: *p*1=НОРМРАСП(*x*2; ; *s*; ИСТИНА)-0;

Для *k*-го интервала *pk*=1- НОРМРАСП(*xk*; ; *s*; ИСТИНА).

Число степеней свободы: *r*=*k*-1-*m=*5-1-2=2 (*m*=2: *mx*, *σx*).

***2.*** ***Определение теоретических частот для равномерного закона распределения U*(*a,b*)*.***

* 1. Определяются границы интервала (*a*\*,*b*\*):



Определяются вероятности для i-го интервала:



Для 1-го интервала: 

Для *k*-го интервала .

Число степеней свободы: *r*=*k*-1-*m=*5-1-2=2 (*m*=2: *a\**, *b\**).

***3. Определение теоретических частот для экспоненциального распределения Exp*(*λ*)*.***

Определим параметр распределения - выборочное .

Определим теоретические частоты для *i*-го интервала:

*pi*=ЭКСПРАСП(*xi*+1; *λ\**; ИСТИНА)- ЭКСПРАСП(*xi*; *λ\**; ИСТИНА)

Для 1-го интервала: *p*1=ЭКСПРАСП(*xi*+1; *λ\**; ИСТИНА)

Для k-го интервала: *pk*=1- ЭКСПРАСП(*xk*; *λ\**; ИСТИНА)

Число степеней свободы: *r*=*k*-1-*m=*5-1-1=3 (*m*=1: *λ*)

**2. Задание для выполнения лабораторной работы.**

Индивидуальные варианты для выполнения лабораторной работы приведены в Приложении А.

*Задание* 1. По выборочным данным:

* построить интервальный статистический ряд;
* графически представить эмпирическую функцию распределения, полигон частот, гистограмму.

*Задание* 2. Вычислить точечные и интервальные оценки математического ожидания и дисперсии с вероятностью *γ*=0,90 для нечетных вариантов и *γ*=0,95 – для четных.

*Задание* 3. Проверить гипотезу о согласии выборочного распределения теоретическому, используя критерий Пирсона. В качестве гипотезы принять распределение из Приложения А.

**3. Содержание отчета.**

1. Цель работы.

2. Постановка задачи.

3. Необходимые расчеты (формулы и таблицы с результатами) и графики.

4. Анализ результатов и выводы.

**ПРИЛОЖЕНИЕ А**

### Индивидуальные варианты для выполнения лабораторной работы

|  |  |  |
| --- | --- | --- |
| Номер  варианта | Выборочные значения | Распределение |
| 1 | 95,6,21, 77, 62, 46,3,71 ,93 ,68 ,84 ,55 ,64 ,89 ,66 ,26 ,18, 91,7,50 | Равномерное |
| 2 | 68, 11, 65, 92, 94, 27, 76,1,65 ,67, 92, 93, 10, 52, 25, 60,9,33 ,66 ,62 | Экспоненциальное |
| 3 | 70, 15, 8, 25, 42, 43, 35, 27, 45, 18, 9, 40, 34, 61, 32, 5, 38,31, 28, 87 | Нормальное |
| 4 | 18, 85, 81, 60, 53, 27, 54, 50, 40, 7, 68, 34, 9, 64, 65, 5, 11, 31, 61, 62 | Равномерное |
| 5 | 22, 87, 27, 14, 65, 6, 92, 73, 9, 62, 51, 75, 27, 56, 95, 47, 8, 50, 38, 98 | Экспоненциальное |
| 6 | 81, 86, 89, 75, 70,0,73, 86, 52, 43, 38,0,29, 81 ,75, 70, 87, 36 ,46 ,41 | Нормальное |
| 7 | 18, 47,2,46, 48, 50, 23, 94, 59, 23, 22, 98,7,26, 99, 18, 77, 37, 27, 95 | Равномерное |
| 8 | 47, 79,2,5,23, 69, 12, 75, 77, 43, 81, 20, 16, 40, 6,9,41, 46, 95, 37 | Экспоненциальное |
| 9 | 86, 30, 63, 33, 45, 22, 57, 43, 23, 25,6,2,80, 47, 28, 35, 34,1,28, 67 | Нормальное |
| 10 | 61, 14, 76, 88,3,33, 50, 78, 64, 30, 81,9,63, 36, 34, 51, 74, 55,9,74 | Равномерное |
| 11 | 68, 44, 56, 17, 37, 62, 29, 46, 74, 11, 28, 56, 21, 35, 7, 65, 29,48, 85,8 | Экспоненциальное |
| 12 | 79, 97, 8 ,36, 64, 4, 20, 70, 25, 66, 91, 81, 95, 10, 84, 69,4,47, 57, 86 | Нормальное |
| 13 | 26, 14, 29, 19, 87, 50,10,20, 65, 41, 27, 6, 50, 48, 63, 41, 7, 34, 23, 80 | Равномерное |
| 14 | 79, 52, 31, 78 ,18, 6, 63,9,72, 32,4,1,55, 10, 65,6,42, 61, 32, 65 | Экспоненциальное |
| 15 | 12,0,55,22,39,92,93,41,9,72, 5 40, 71, 67, 40, 25, 74, 99, 29,8 | Нормальное |
| 16 | 28, 56,0,30, 98, 81, 80, 37, 20, 26, 24, 57, 97, 47, 14, 96, 14, 48,1,49 | Равномерное |
| 17 | 80,6, 48, 79, 86, 27, 25, 29, 39, 16, 85, 55, 67, 66, 16 ,10, 71, 90, 37,9 | Экспоненциальное |
| 18 | 8,63,2,0,72, 36, 10, 13, 42, 41, 53, 74, 31, 18, 41, 97, 59,1,76, 39 | Нормальное |
| 19 | 74, 63, 3, 63, 80, 15, 59, 71, 66, 86,9,56, 42, 47, 76, 42, 19, 52. 16, 10 | Равномерное |
| 20 | 92,4,52, 19, 16, 71, 91, 64, 80, 18, 77, 52, 27, 20. 83, 63, 62, 67, 7, 40 | Экспоненциальное |
| 21 | 51, 9, 84, 35, 13, 10, 48, 7, 44, 17, 67, 63, 35, 9, 40, 82, 66, 12, 71 21 | Нормальное |
| 22 | 88, 66, 7, 10, 34, 9, 48,6,11, 90, 57, 38, 72, 41, 66, 38, 59, 44,9,43 | Равномерное |
| 23 | 78, 82, 63, 55, 43, 10, 9, 59, 9, 70, 63, 79, 75, 88, 96, 8, 65, 37, 45, 65 | Экспоненциальное |
| 24 | 62, 26, 81, 50,8,72, 36, 75, 96, 38, 28, 83, 33, 69, 5, 90, 59, 46, 39,65 | Нормальное |
| 25 | 75, 57, 67, 18, 84, 87, 24, 50, 18, 19, 5, 32, 2,70, 47, 83, 84, 8, 31, 70 | Равномерное |
| 26 | 36, 10, 13, 42, 41, 53, 68, 34, 9, 64, 65, 5, 11, 31, 61, 62, 26 | Экспоненциальное |
| 27 | 81, 60, 53, 27, 54, 50, 40, 7, 68, 34, 9, 37, 62, 9, 46, 45, 22, 57, 43, 23 | Нормальное |
| 28 | 27, 56, 95, 47, 41, 66, 38, 59, 44,9, 70, 25, 66, 91, 69, 5, 90, 59 | Равномерное |
| 29 | 10, 13, 42, 12, 75, 7,41, 53, 74, 31, 18, 41, 97, 93 ,68 ,4 ,55 ,64 ,89,66 | Экспоненциальное |
| 30 | 9, 48,6,11, 90, 57, 38, 34, 61, 32, 87, 24, 50, 18, 19, 5, 32, 6, 63,7 | Нормальное |