**1.Основные понятия математической статистики**.

**Математическая статистика** - раздел прикладной математики, предметом которого является разработка рациональных приемов и методов получения, описания и обработки опытных данных в целях изучения закономерностей массовых случайных явлений.

**Основные задачи математической статистики**:

1.1 Определение по статистическим данным законов распределения случайных величин;

1.2 Определение по статистическим данным параметров распределения случайных величин;

1.3 Выработка рекомендаций по проведению эксперимента и обработки результатов наблюдений;

1.4 Проверка правдоподобия статистических гипотез.

 Множество всех объектов, надлежащих изучению, называется генеральной совокупностью.

 Множество случайно отобранных объектов называется выборочной совокупностью или выборкой.

 Число объектов генеральной совокупности или выборки называется объёмом.

 Статистическим рядом распределения называют упорядоченную совокупность значений признака и соответствующие им частоты (вариационный ряд).

 Отношение частоты nί значения признака Xί к объёму выборки n называют относительной частотой W=nί/n

Пример 1.1: Распределение частот выборки

|  |  |  |  |
| --- | --- | --- | --- |
| Xί | 1 | 3 | 5 |
| nί | 20 | 30 | 10 |

График распределения частот называют полигоном частот.

Пример 1.2: В результате испытания случайная величина приняла значения:

Х1=2, Х2=5, Х3=7, Х4=1, Х5=10, Х6=5, Х7=9, Х8=6, Х9=8, Х10=6, Х11=2, Х12=3, Х13=7, Х14=6, Х15=8, Х16=3, Х17=8, Х18=10, Х19=6, Х20=7, Х21=3, Х22=9, Х23=4, Х24=5, Х25=6.

 Составить таблицу, устанавливающую зависимость между значениями случайной величины и её частотами; построить статистическое распределение, изобразить полигон частот, полигон относительных частот, гистограмму и кумуляту.

Решение: Объём выборки n=25

Составим таблицу:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Хί | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| nί | 1 | 2 | 3 | 1 | 3 | 5 | 3 | 3 | 2 | 2 |

Таблица статистического распределения:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Xί | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Wί | 0,04 | 0,08 | 0,12 | 0,04 | 0,12 | 0,16 | 0,12 | 0,12 | 0,08 | 0,08 |

Полигон частот (рис. 1)



 Рис. 1.

Полигон относительных частот (рис. 2.)



 Рис. 2.

Гистограммой частот называют ступенчатую фигуру (для каждого **Xί** строят прямоугольник, высота которого равна nί )

Гистограмма частот (рис. 3.)



 Рис. 3.

Кумулятивная кривая (кумулята) – кривая накопленных частот.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Xί | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| nί | 1 | 3 | 6 | 7 | 10 | 15 | 18 | 21 | 23 | 25 |



**2.Числовые характеристики статистического распределения**

* 1. Выборочная средняя (выборочное математическое ожидание) (несмещенная оценка)

 $\overbar{Х\_{В}}=\frac{\sum\_{i=1}^{k}x\_{i}n\_{i}}{n}=M^{\*}\left(X\right)=m\_{x}^{\*}$

* 1. Выборочная дисперсия (смещенная оценка)

 $D\_{B}=\frac{\sum\_{i=1}^{k}\left(x\_{i}-\overbar{x\_{B}}\right)^{2}∙n\_{i}}{n}$

* 1. Среднеквадратичное или стандартное отклонение (стандарт)



* 1. Мода - значение варианты, имеющий наибольшую частоту (обозначается М0).
	2. Медиана (mе) – серединная варианта, если n=2k, то me=
	3. Выборочный начальный момент:



* 1. Выборочный центральный момент:



* 1. Асимметрия: 
	2. Эксцесс:

 

Статистические характеристики являются случайными величинами, но как числовые характеристики они постоянны.

Пример 2.1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Xί | 1 | 2 | 3 | 4 |
| nί | 20 | 15 | 10 | 5 |

Выборочная средняя 

Ð6==1

Мо=1 me=

Точечной называют оценку, которая определяется одним числом.

Интервальной называют оценку, определяемую двумя числами – концами интервала.

**Надёжностью** (доверительной вероятностью) оценки θ по θ\* называют вероятность γ, с которой осуществляется неравенство ׀θ-θ\*׀<δ

Надёжность 0,95 означает, что если произведено достаточно большое количество выборок, 95% из них определяют доверительный интервал, в котором параметр действительно содержится. Чаще всего задают надежность, равную 0,95; 0,99; 0,999.

Если из генеральной совокупности извлечена выборка и х1 наблюдалось n1 раз, х2-n2 раза, хk- nk раз и Σni=n, то наблюдаемые значения хi называют вариантами, последовательность вариант записанных в возрастающем порядке - вариационным рядом; ni - частоты.

Равноотстоящими называют варианты, которые образуют арифметическую прогрессию с разностью h.

Условными называют варианты, определяемые равенством:

Ui = ложный нуль, шаг.

В качестве ложного нуля можно принять любую варианту, но проще ту, которая расположена примерно в середине ряда и имеет большую частоту.

Пример 2.2:

Для изучения производительности труда Х (тыс. руб.) обследовано n предприятий данной отрасли. Результаты представлены выборкой.

 20 15 17 19  **23** 18 21 15 16 13

 20 16 19 20 14 20 16 14 20 19

 15 19 17 16 15 22 21 12  **10** 21

 18 14 18 14 13 18 19 18 20  **23**

 16 20 19 17 19 17 21 17 19 13

1. Построим интервальный статистический ряд из семи интервалов: Xmin=10, Xmax=23, n = 50.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10-11 | 12-13 | 14-15 | 16-17 | 18-19 | 20-21 | 22-23 |
| 1 | 4 | 8 | 10 | 13 | 11 | 3 |

2. От ряда 1 перейдём к точечному распределению, взяв в качестве вариант середины интервалов.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10,5 | 12,5 | 14,5 | 16,5 | 18,5 | 20,5 | 22,5 |
| 1 | 4 | 8 | 10 | 13 | 11 | 3 |

Составим расчётную таблицу 1.

За ложный нуль возьмём С=16,5 (середина интервала).

Таблица 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| xi | ni | ui | niui | niui2 | ni(ui+1)2 |
| 10,5 | 1 | -3 | -3 | 9 | 4 |
| 12,5 | 4 | -2 | -8 | 16 | 4 |
| 14,5 | 8 | -1 | -8 | 8 | 0 |
| 16,5 | 10 | 0 | 0 | 0 | 10 |
| 18,5 | 13 | 1 | 13 | 13 | 52 |
| 20,5 | 11 | 2 | 22 | 44 | 99 |
| 22,5 | 3 | 3 | 9 | 27 | 48 |
|  | n=50 |  | 25 | 117 | 217 |

Контроль: 

 217=117+2\*25+50 – верно

Вычислим условные моменты первого и второго порядков:

 $M\_{1}^{\*}=\frac{\sum\_{}^{}n\_{i}u\_{i}}{n}=\frac{25}{50}=0.5$

 $M\_{2}^{\*}=\frac{\sum\_{}^{}n\_{i}\left(u\_{i}\right)^{2}}{n}=\frac{117}{50}=2.34$

Шаг (разность между любыми двумя соседними вариантами): h=12,5-10,5=2

Вычислим искомые: 





- исправленная дисперсия



4. 

Доверительный интервал для математического ожидания:

  где ; $Ф\left(t\right)=\frac{γ}{2}$

$Ф\left(t\right)=\frac{0.95}{2}=0,475 \rightarrow $ t = 1.96 (приложение 2 [Гм]);

;

17,5 - 0,8 < a < 17,5 + 0,8

16,7 < a<18,3.

Доверительный интервал для $σ$ нормального распределения:

 $S\left(1-q\right)<σ<S\left(1+q\right),$

q ( γ, n ) = q (0,95; 50)= 0,21 ([Гм] приложение 4);

 $S=\sqrt{S^{2}}=\sqrt{8.53}=2.92$

 $2.92\left(1-0.21\right)<σ<2.92\left(1+0.21\right)$

 $2.31<σ<3.53$

5. Вычислим теоретические частоты по формуле:

$$n\_{i}=\frac{nh}{σ\_{b}}∙φ\left(u\_{i}\right)=\frac{50∙2}{2.89}φ\left(u\_{i}\right)=34.6∙φ\left(u\_{i}\right)$$

Составим расчётную таблицу 2.

Значение φ(ui) находим по приложению 1 ([Гм]).

 Таблица 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| i | xi | $$u\_{i}=\frac{x\_{i}-\overbar{x\_{b}}}{σ\_{b}}$$ | φ (ui) | ni|=34,6φ(ui) |
| 1 | 10.5 | - 2.42 | 0.0213 | 0.7 |
| 2 | 12.5 | - 1.73 | 0.0893 | 3.1 |
| 3 | 14.5 | - 1.04 | 0.2323 | 8.0 |
| 4 | 16.5 | - 0.35 | 0.3752 | 13.0 |
| 5 | 18.5 | 0.35 | 0.3752 | 13.0 |
| 6 | 20.5 | 1.04 | 0.2323 | 8.0 |
| 7 | 22.5 | 1.73 | 0.0893 | 3.1 |

Сравним эмпирические и теоретические частоты. Для этого составим расчётную таблицу 3.

Таблица 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | ni | ni| | ni - ni| | (ni - ni| )2 |  |
| 1 | 1 | 0.7 | 0.3 | 0.09 | 0.1 |
| 2 | 4 | 3.1 | 0.9 | 0.81 | 0.3 |
| 3 | 8 | 8.0 | 0 | 0 | 0 |
| 4 | 10 | 13.0 | -3 | 9 | 0.7 |
| 5 | 13 | 13.0 | 0 | 0 | 0 |
| 6 | 11 | 8.0 | 3 | 9 | 1.1 |
| 7 | 3 | 3.1 | -0.1 | 0.01 | 0 |
|  | 50 |  |  |  | $$χ^{2}=2,2$$ |

 Наблюдаемое

По таблице критических точек распределения χ2 (приложение 5 [Гм]), по уровню значимости α = 0,05 и числу степеней свободы k=S-3=7-3=4.

Находим критическую точку правосторонней критической области:

$$χ\_{критич}^{2}\left(0,05;4\right)=9,5$$

Так как $ χ\_{набл.}^{2}<χ\_{критич}^{2}$ - нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности, то есть расхождение частот незначимо (случайно).

**Пример (по аналогии с примером 2.2)**

Продолжительность работы электронных ламп одного типа (в часах)

13,4 14,7 15,2 15,1 13,0 8,8 14,0 17,9 15,1 16,5

14,2 16,3 14,6 11,7 16,4 15,1 17,6 14,1 18,8 11,6

18,0 12,4 17,2 14,5 16,3 13,7 15,5 16,2 8,4 14,7

11,3 10,7 16,9 15,8 16,1 12,3 14,0 17,7 14,7 16,2

10,1 15,8 18,3 17,5 12,7 20,7 13,5 14,0 15,7 21,9

17,7 15,4 10,9 18,2 17,3 15,2 16,7 17,3 12,1 19,2

16,6 13,9 15,4 17,1 14,3

Первый интервал взять [8.4 – 10.4)

Проверить гипотезу о нормальном распределении генеральной совокупности.

Таблица значений q = q($γ;n$) Приложение 4

|  |  |  |  |
| --- | --- | --- | --- |
| n | $$γ$$ | n | $$γ$$ |
| 0.95 | 0.99 | 0.999 | 0.95 | 0.99 | 0.999 |
| 5 | 1.37 | 2.67 | 5.64 | 20 | 0.37 | 0.58 | 0.88 |
| 6 | 1.09 | 2.01 | 3.88 | 25 | 0.32 | 0.49 | 0.73 |
| 7 | 0.92 | 1.62 | 2.98 | 30 | 0.28 | 0.43 | 0.63 |
| 8 | 0.80 | 1.38 | 2.42 | 35 | 0.26 | 0.38 | 0.56 |
| 9 | 0.71 | 1.20 | 2.06 | 40 | 0.24 | 0.35 | 0.50 |
| 10 | 0.65 | 1.08 | 1.80 | 45 | 0.22 | 0.32 | 0.46 |
| 11 | 0.59 | 0.98 | 1.60 | 50 | 0.21 | 0.30 | 0.43 |
| 12 | 0.55 | 0.90 | 1.45 | 60 | 0.188 | 0.269 | 0.38 |
| 13 | 0.52 | 0.83 | 1.33 | 70 | 0.174 | 0.245 | 0.34 |
| 14 | 0.48 | 0.78 | 1.23 | 80 | 0.161 | 0.226 | 0.31 |
| 15 | 0.46 | 0.73 | 1.15 | 90 | 0.151 | 0.211 | 0.29 |
| 16 | 0.44 | 0.70 | 1.07 | 100 | 0.143 | 0.198 | 0.27 |
| 17 | 0.42 | 0.66 | 1.01 | 150 | 0.115 | 0.160 | 0.211 |
| 18 | 0.40 | 0.63 | 0.96 | 200 | 0.099 | 0.136 | 0.185 |
| 19 | 0.39 | 0.60 | 0.92 | 250 | 0.089 | 0.120 | 0.162 |

65= $\frac{60+70}{2}\rightarrow q\left(γ,n\right)=q\left(0.95;65\right)= \frac{0.188+0.174}{2}=0.181$