ЗАДАЧИ ЭКОМЕТРИИ

1 задача: Указать способы сбора и группировки стат. сведений, полученных в результате наблюдений или некоторых поставленных экспериментов в области экологии и природопользования

2 задача: Разработать методы анализа стат.данных в зависимости от целей исследования:

а).оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения; оценка зависимости случайной величины от других случайных величин

б).проверка стат.гипотез о виде неизвестного распределения

Гл.1 Случайные величины

§1. Основные понятия

Случайная величина

Дискретная (ДСВ)

Непрерывная (НСВ)

ДСВ обозначаем: X,Y,Z…, а их значения x,y,z…

ДСВ имеет конечное число значений,
НСВ- имеет бесконечное число значений.

§2. ВЫБОРОЧНЫЙ МЕТОД

п.1.Генеральная и выборочная совокупность

СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ

↓

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ
(ОБЪЕМ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ)

↓

ВЫБОРКА (ОБЪЕМ ВЫБОРКИ)

ПОВТОРНЫЕ

БЕСПОВТОРНЫЕ

РЕПРЕЗЕНТАТИВНАЯ

п.2. Статистическое распределение выборки.

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

↓

ВЫБОРОЧНАЯ СОВОКУПНОСТЬ

Значение x_1 наблюдалось n_1 раз; x_2-n_2 раза;…; x_k- n_k раз.

$$\boxed{n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i}$$

$x_1;x_2;\ldots x_k$-варианты

$n_1;n_2;\ldots n_k$ – частоты

$$\frac{n_1}{n} = \omega_1;\ldots\frac{n_i}{n} = \omega_i$$

относительные частоты

ПРИМЕР

Задано распределение частот выборки объема *n*=20

Написать распределение относительных частот

x_i	2	6	12
n_i	3	10	7

$$\sum n_i = 20$$

$$\omega_1 = \frac{3}{20} = 0{,}15$$

$$\omega_2 = \frac{10}{20} = 0{,}5$$

$$\omega_3 = \frac{7}{20} = 0{,}35$$

x_i	2	6	12
ω_i	0,15	0,5	0,35

Проверка: 0,15+0,5+0,35=1

п.3. Полигон и гистограмма

Графическое изображение статистического распределения

Полигон частот: ломаная $(x_1;n_1)\dots(x_k;n_k)$

Полигон относительных частот: $(x_1;\omega_1)\dots(x_k;\omega_k)$

Графическое изображение непрерывного распределения

Гистограмма- ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высоты равны отношению $\dfrac{n_i}{h}$ -плотности частот

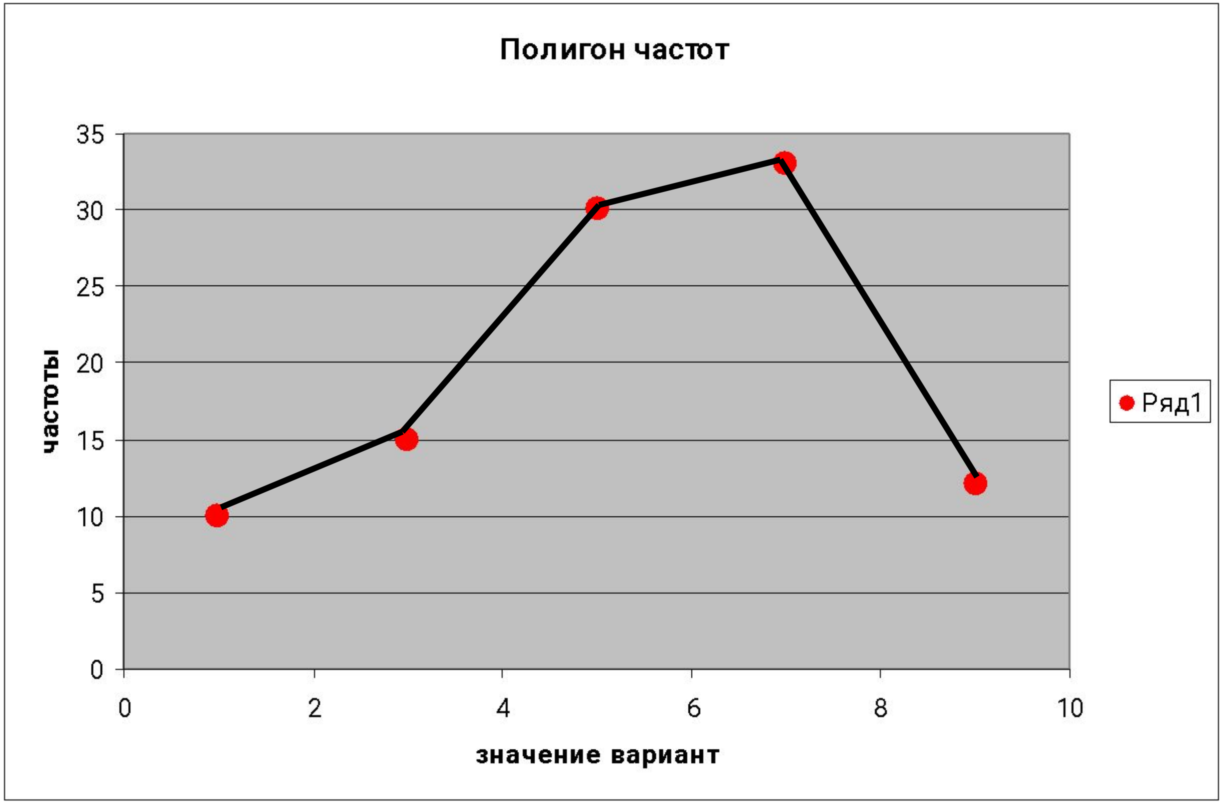
Пример 1.

$$\sum n_i = 100$$ (Рис.1.)

x_i	1	3	5	7	9
n_i	10	15	30	33	12

x_i	1	3	5	7	9
ω_i	0,1	0,15	0,3	0,33	0,12

$$\sum \omega_i = 1$$ (Рис.2.)



Полигон частот
35
30
25
20
15
10
5
0
частоты
0
2
4
6
8
10
значение вариант
Ряд1

Рис.1

Рис.1

САМОСТОЯТЕЛЬНО!

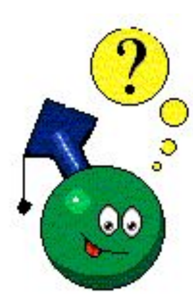
2). Построить гистограмму частот и относительных частот распределения.

Частичный интервал	Сумма частот вариант	Плотность частот
2-5	9	9/3=3
5-8	10	10/3=3,3(3)
8-11	25	25/3=8,3(3)
11-14	6	6/3=2

$$\sum n_i = 50$$

Длина частичного интервала равна 3

Найдем плотность частоты $\dfrac{n_i}{h}$



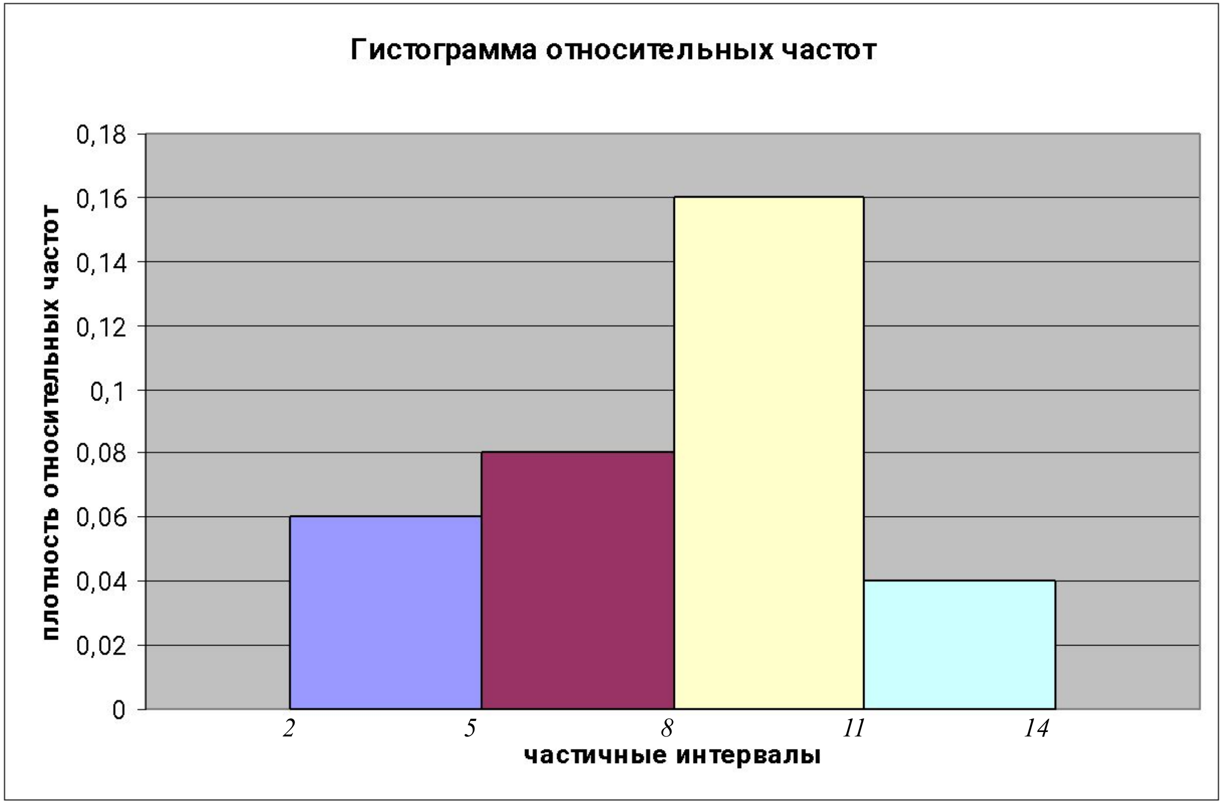
Строим гистограмму частот

(2-5;3); (5-8;3,3);

(8-11;8,3); (11-14;2)

Частичный интервал	$\frac{n_i}{n} = \omega_i$	Плотность отн. частоты $\frac{w_i}{h}$
2-5	9/50 =0,18	0,18/3=0,06
5-8	10/50=0,2	0,2/3=0,08
8-11	25/50 =0,5	0,5/3=0,16
11-14	6/50=0,12	0,12/3=0,04

Построить гистограмму


Гистограмма относительных частот
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
плотность относительных частот
2
5
8
11
14
частичные интервалы

П. 4. Эмпирическая функция распределения.

Пусть n_x - *число наблюдений*

n - *объем выборки*

$\frac{n_x}{n}$ - относительная частота

$\frac{n_x}{n}$ - функция, зависящая от x

эмпирическая - установленная опытным путем

ОПРЕДЕЛЕНИЕ

$$F^*(x) = \frac{n_x}{n}$$ функция распределения выборки

n_x-число вариант, меньших x

n-объем выборки

* Функцию распределения генеральной совокупности называют теоретической функцией распределения F(x)

- Теоретическая функция F(x) определяет вероятность события X<x, а эмпирическая функция $F^{*}(x)$ определяет относительную частоту этого же события .
- Если объем выборки *n* число большое, то функции F(x) и $F^{*}(x)$ мало отличаются друг от друга, т.е.

$$\lim_{n\to\infty} P(\left|F(x)-F^{*}(x)\right|<\varepsilon)=1 \qquad (\varepsilon>0)$$

Это равенство (теорема Чебышева) является теоретической основой выборочного метода.

ПРИМЕР:

Дано распределение выборки

x-варианты	1	5	9
n-частоты	6	12	22

n=40- объем выборки

Построить эмпирическую функцию.

1). Наименьшая варианта равна 1, по свойству функции распределения $F^*(x)=0$ при $x \le 1$

2). Значение $X<5$ наблюдалось 6 раз,т.е.

$$F^*(x) = \frac{6}{40} \qquad 1 < x \le 5$$

3). Значение X<9 наблюдалось 6+12=18 раз, т.е.

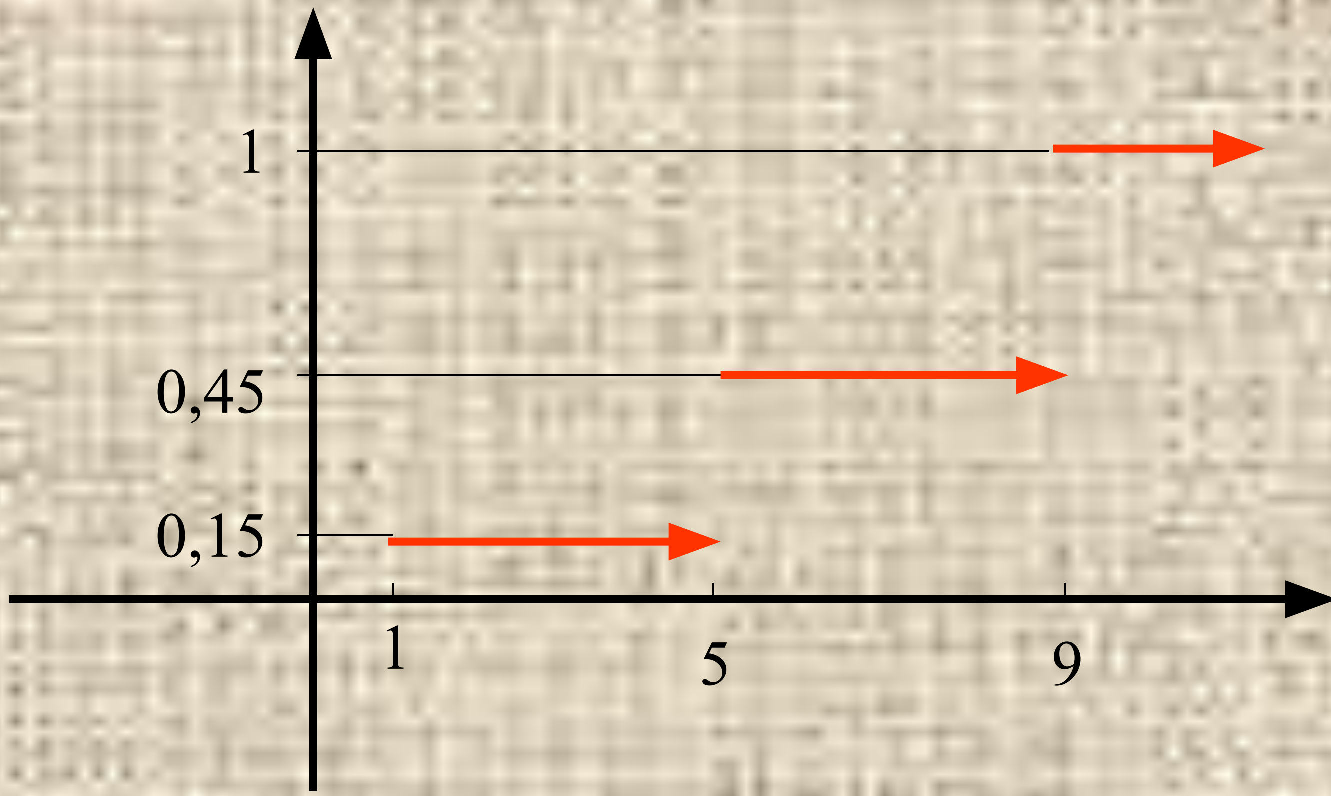
$$F^*(x) = \frac{18}{40} \qquad 5 < x \le 9$$

4) Наибольшая варианта x=9, тогда

$F^*(x)=1$ при $x>9$

$$\boxed{F^*(x) = \begin{cases} 0; x \le 1 \\ 0{,}15; 1 < x \le 5 \\ 0{,}45; 5 < x \le 9 \\ 1; x > 9 \end{cases}}$$

ПОСТРОИМ ГРАФИК ЭМПИРИЧЕСКОЙ ФУНКЦИИ:



§3. Оценки параметров генеральной совокупности по ее выборке.

П.1.

X-количественный признак, *x*-значения этого признака.

⇓

X-случайная величина, *x*-одно из возможных ее значений

$x_1;x_2;\ldots x_n$- значения количественного признака, полученные в результате *n*-независимых испытаний

Найти оценку неизвестного параметра-значит найти функцию от наблюдаемых СВ $X_1;X_2;\ldots X_n$, которая дает приближенное значение оцениваемого параметра

П.2. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ СРЕДНИЕ

Пусть изучается дискретная генеральная совокупность объема N относительно количественного признака X

ОПРЕДЕЛЕНИЕ

Генеральной средней $\overline{x_Г}$ называется среднее арифметическое значений признака генеральной совокупности, т.е.

$$\overline{x_Г} = \frac{x_1 + x_2 + ... + x_N}{N}$$

Если значения $x_1; x_2; \ldots x_k$ имеют соответственно частоты $N_1; N_2; \ldots; N_k,$ причем $N_1 + N_2 + \ldots + N_k = N$

$$\overline{x_Г} = \frac{x_1 N_1 + x_2 N_2 + ... + x_k N_k}{N} = \frac{1}{N}\sum_{i=1}^{k} x_i N_i$$

Так как каждый объект может быть извлечен с одной и той же вероятностью 1/N, то тогда генеральная средняя

$$\overline{x_Г} = M(X)$$

ОПРЕДЕЛЕНИЕ

Выборочной средней называется среднее арифметическое значений признака выборочной совокупности.

Если все значения $x_1; x_2; \ldots; x_n$ признака выборки объема n различны, то средняя выборочная равна

$$\overline{x_B} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Если же значения $x_1;x_2;\ldots x_k$ имеют частоты $n_1;n_2;\ldots n_k$, причем $n_1+n_2+\ldots+n_k=n$, то

$$\overline{x_B}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i$$

Выборочная средняя для различных выборок того же объема из той же генеральной совокупности может получаться различной.

Всевозможные, получающиеся выборочные средние есть возможные значения случайной величины, которая называется выборочной средней СВ

$\overline{X}$ -выборочная средняя

Если варианты x_i –большие числа, то для облегчения вычисления выборочной средней применяют, так называемый, «ложный нуль»

Пусть C-const, т.к. $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} (x_i - C) + nC$

тогда выборочная средняя вычисляется по формуле

$$\boxed{\overline{x_B} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - C) + nC\right) = C + \frac{1}{n}\sum_{i=1}^{n} (x_i - C)}$$

C-const-«ложный нуль», постоянную берут такой, чтобы $x_i - C$ были небольшими и C по возможности было числом круглым

Имеется выборка

$x_1=71{,}88$ $x_4=72{,}07$ $x_7=71{,}93$ $x_{10}=71{,}96$

$x_2=71{,}93$ $x_5=71{,}90$ $x_8=71{,}77$

$x_3=72{,}05$ $x_6=72{,}02$ $x_9=72{,}71$

C=72, найдем разность $x_i - C = \alpha_i$

$\alpha_1= -\ 0{,}12$ $\alpha_4=0{,}07$ $\alpha_7= -\ 0{,}07$ $\alpha_{10}= -\ 0{,}04$

$\alpha_2= -\ 0{,}07$ $\alpha_5= -\ 0{,}1$ $\alpha_8=-0{,}23$

$\alpha_3=0{,}05$ $\alpha_6=0{,}02$ $\alpha_9=0{,}71$

Их сумма равна 0,22

$$\frac{\sum \alpha_i}{10} \approx 0{,}02$$ среднее арифметическое

Тогда выборочная средняя равна 72+0,02=72,02

П.3. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ ДИСПЕРСИЯ

ОПРЕДЕЛЕНИЕ

Генеральной дисперсией $D_Г$ называют среднее арифметическое квадратов отклонений значений признака X генеральной совокупности от генеральной средней $\overline{x}_Г$

$$D_Г = \frac{1}{N}\sum_{i=1}^{n}(x_i - \overline{x_Г})^2$$

Генеральным средним квадратическим отклонением (стандартом) называется

$$\sigma_Г = \sqrt{D_Г}$$

$\sigma(\overline{X})$ - средняя квадратическая ошибка

ОПРЕДЕЛЕНИЕ

Выборочной дисперсией называется среднее арифметическое квадратов отклонений наблюдаемых значений признака X от выборочной средней

$$D_B = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x_B}\right)^2$$

Выборочный стандарт $$\sigma_B = \sqrt{D_B}$$

Выборочную дисперсию, рассматриваемую как случайную величину, будем обозначать

$$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$\overline{X}$ -выборочная средняя случайная величина

ТЕОРЕМА

$$M(\widetilde{S}^2) = \frac{n-1}{n}\cdot D_{\Gamma}$$

- Если варианты – большие числа,то для вычисления используем «ложный нуль» С

$$D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - C)^2 - (\overline{x_B - C})^2$$

§4.Точность оценки, доверительная вероятность(надежность), доверительный интервал

- Точечной называют оценку, которая определяется одним числом
- При выборке малого объема точечная оценка значительно отличается от оцениваемого параметра

- Интервальной называют оценку, которая определяется двумя числами, концами интервала
- Интервальные оценки позволяют установить точность и надежность оценки

Интервальные оценки

- Статистическая характеристика θ*, найденная по данным выборки, служит оценкой неизвестного параметра θ
- Пусть θ –постоянное число или случайная величина
- Чем меньше модуль разности $|\theta - \theta^*|$, тем точнее θ * определяет θ
- Пусть $|\theta - \theta^*| < \delta$,при δ>0. Чем меньше δ, тем оценка точнее, т.е. δ характеризует точность оценки

Интервальные оценки

- Доверительной вероятностью или надежностью оценки θ по θ * называют вероятность γ, с которой осуществляется неравенство

$$\left|\theta - \theta^*\right| < \delta$$

- (θ *-δ; θ *+δ) интервал доверительный
- Он покрывает неизвестный параметр с заданной надежностью γ

§5.Характеристики вариационного ряда

- Модой M_0 называют варианту, которая имеет наибольшую частоту.

Например имеется ряд вида:

варианта	1	4	7	9
частота	5	1	20	6

$\boxed{M_0=\mathbf{7}}$

Характеристики вариационного ряда

- Медианой m_e называют варианту, которая делит вариационный ряд на две части , равные по числу вариант. Если число вариант нечетно, т.е. $n=2k+1$, то $m_e=x_{k+1}$, при четном $n=2k$ медиана

$$m_e = \frac{x_k + x_{k+1}}{2}$$

- Размахом варьирования R называют разность между наибольшей и наименьшей вариантами $R=x_{max}-x_{min}$. Размах является простейшей характеристикой вариационного ряда.

Характеристики вариационного ряда

- Средним абсолютным отклонением

θ (тэта) называют среднее арифметическое абсолютных отклонений

$$\theta = \frac{\sum n_i \left| x_i - \overline{x_в} \right|}{\sum n_i}$$

Среднее абсолютное отклонение служит для характеристики рассеяния вариационного ряда

Пусть дан вариационный ряд

x_i	1	3	6	16
n_i	4	10	5	1

$$\overline{x_i} = \frac{4\cdot 1 + 10\cdot 3 + 5\cdot 6 + 1\cdot 16}{4+10+5+1} = \frac{80}{20} = 4$$

$$\theta = \frac{4\cdot\left|1-4\right| + 10\cdot\left|3-4\right| + 5\cdot\left|6-4\right| + 1\cdot\left|16-4\right|}{4+10+5+1} = 2{,}2$$

Среднее абсолютное отклонение

Характеристики вариационного ряда

- Коэффициентом вариации V называют выраженное в процентах отношение выборочного среднего квадратического отношения к выборочной средней

$$V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\%$$

- Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого коэффициент вариации больше.
- Коэффициент вариации- безразмерная величина, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность, например, если варианты одного ряда выражены в сантиметрах, а другого- в граммах.

- Если вариационный ряд составлен по данным выборки, то все описанные характеристики называют выборочными.
- Если вариационный ряд составлен по данным генеральной совокупности, то характеристики называют генеральными

$u_i = \frac{x_i - C}{h}$

§6.Методы расчета сводных характеристик выборки

П.1. Условные варианты

Пусть варианты выборки расположены в возрастающем порядке, т.е. в виде вариационного ряда.

Равноотстоящими называют варианты, которые образуют арифметическую прогрессию с разность *h*

Условными называют варианты, определяемые равенством

$$u_i = \frac{x_i - C}{h}$$

где С- ложный нуль, *h*- шаг, т.е. разность между любыми двумя соседними первоначальными вариантами.

методы расчета

- Упрощенные методы расчета сводных характеристик выборки основаны на замене первоначальных вариант условными.
- Если вариационный ряд состоит из *равноотстоящих* вариант с h- шагом, то условные варианты есть целые числа
- Выберем в качестве ложного нуля произвольную варианту, например x_m, тогда условная варианта

$$u_i = \frac{x_i - x_m}{h} = \frac{x_1 + (i-1)h - \left[x_1 + (m-1)h\right]}{h} = i - m$$

- т.к. i и m целые числа, то и их разность есть целое число

методы расчета

Замечание 1. В качестве ложного нуля можно взять любую варианту. Максимальная простота вычислений достигается, если в качестве ложного нуля выбрать варианту, которая расположена приблизительно в середине вариационного ряда (часто такая варианта имеет наибольшую частоту)

Замечание 2. Варианте, которая принята в качестве ложного нуля, соответствует условная варианта равная нулю.

П.2. Обычные начальные и центральные эмпирические моменты

- Обычным эмпирическим моментом порядка k называют среднее значение k-х степеней разностей x_i-C

$$\boxed{M_k^* = \frac{\sum n_i (x_i - C)^k}{n}}$$

x_i-наблюдаемая варианта, С- ложный нуль, n_i-частота варианты, n- объем выборки

эмпирические моменты

- Начальным эмпирическим моментом порядка *k* называется обычный момент порядка *k* при *C=0*

$$M_k = \frac{\sum n_i (x_i)^k}{n}$$

- *Начальный эмпирический момент первого порядка равен выборочной средней*

$$M_1 = \frac{\sum n_i x_i}{n} = \overline{x}_в$$

эмпирические моменты

- Центральным эмпирическим моментом порядка *k* называется обычный момент порядка *k* при *C*= $\overline{x}_в$

$$m_k = \frac{\sum n_i (x_i - \overline{x}_в)^k}{n}$$

- Центральный эмпирический момент второго порядка равен выборочной дисперсии

$$m_2 = \frac{\sum n_i (x_i - \overline{x}_в)^2}{n} = D_в$$

- Выразим центральные моменты через обычные

$$m_2 = M_2^* - (M_1^*)^2$$

$$m_3 = M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3$$

$$m_4 = M_4^* - 4M_3^* M_1^* + 6M_2^* (M_1^*)^2 - 3(M_1^*)^4$$

П.3. Условные эмпирические моменты. Отыскание центральных моментов по условным

- Для упрощения расчетов первоначальные варианты заменяем условными
- Условным эмпирическим моментом порядка *k* называется начальный момент порядка *k*, вычисленный для условных вариант

$$M_k^* = \frac{\sum n_i \left(\frac{x_i - C}{h} \right)^k}{n}$$

при $k = 1$

$$M_1^* = \frac{\sum n_i \left(\frac{x_i - C}{h} \right)}{n} = \frac{1}{h} \left[\frac{\sum n_i x_i}{n} - C \frac{\sum n_i}{n} \right] = \frac{1}{h} \left(\overline{x_в} - C \right)$$

Для того, чтобы найти выборочную среднюю, необходимо условный момент первого порядка умножить на шаги к результату прибавить ложный нуль

$$\overline{x_в} = M_1^* h + C$$

Найдя таким образом обычные моменты можно получить центральные, в итоге получаем удобные для вычислений формулы, выражающие центральные моменты через условные

$$\boxed{m_2 = (M_2^* - (M_1^*)^2)h^2}$$

$$\boxed{m_3 = (M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3)h^3}$$

$$\boxed{m_4 = (M_4^* - 4M_3^* M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4)h^4}$$

выборочная дисперсия

- В соответствии с предыдущими формулами получим формулу для вычисления выборочной дисперсии по условным моментам первого и второго порядков

$$D_{в} = \left(M_2^* - (M_1^*)^2 \right) h^2$$

П. 4

Метод произведений для вычисления выборочной средней и выборочной дисперсии

- Метод произведений –это удобный способ для вычисления условных моментов вариационного ряда с равноотстоящими вариантами. Зная условные моменты, найдем начальные и центральные моменты и соответственно выборочную среднюю и выборочную дисперсию
- Этот метод удобнее оформлять в таблицу

Выборочные варианты в возрастающем порядке	Частоты вариант	Условные варианты $u_i = \frac{x_i - C}{h}$	Частоты умножают на варианты $n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
…	…	…	…	…	…
	$\sum n_i = n$		$\sum n_i u_i$	$\sum n_i u_i^2$	$\sum n_i (u_i + 1)^2$

метод произведений

- Заполняя третий столбец, варианту с большей частотой или варианту, находящуюся примерно в середине вариационного ряда берут за 0, в клетках над ним берут -1,-2,-3…, под ним 1,2,3…и т. д.
- После заполнения расчетной таблицы вычисляются условные моменты и затем - выборочные средние и выборочная дисперсия:

1 $$M_1 = \frac{\sum n_i x_i}{n}$$

2 $$M_2^* = \frac{\sum n_i x_i^2}{n}$$

3 $$\overline{x_в} = M_1^* h + C$$

4 $$D_в = \left(M_2^* - (M_1^*)^2\right) h^2$$

П.5. Построение нормальной кривой по опытным данным

- Находим $\overline{x}_в ; \sigma_в$, например по методу произведений
- Находим ординаты (выравнивающие частоты) теоретической кривой по формуле

$$y_i = \frac{nh}{\sigma_в} \cdot \varphi(u_i)$$

где *n*- сумма наблюдаемых частот, *h*- разность между двумя соседними вариантами, значения выборочных средних равны

$$\boxed{u_i = \frac{x_i - \overline{x}_в}{\sigma_в}}$$

$$\boxed{\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}}$$

- Строим точки с координатами $(x_i;y_i)$ в прямоугольной системе координат и соединяем их плавной кривой
- Близость выравнивающих частот к наблюдаемым подтверждает правильность допущения о том, что обследуемый признак распределен нормально

пример построения нормальной кривой

Построить нормальную кривую по данному распределения

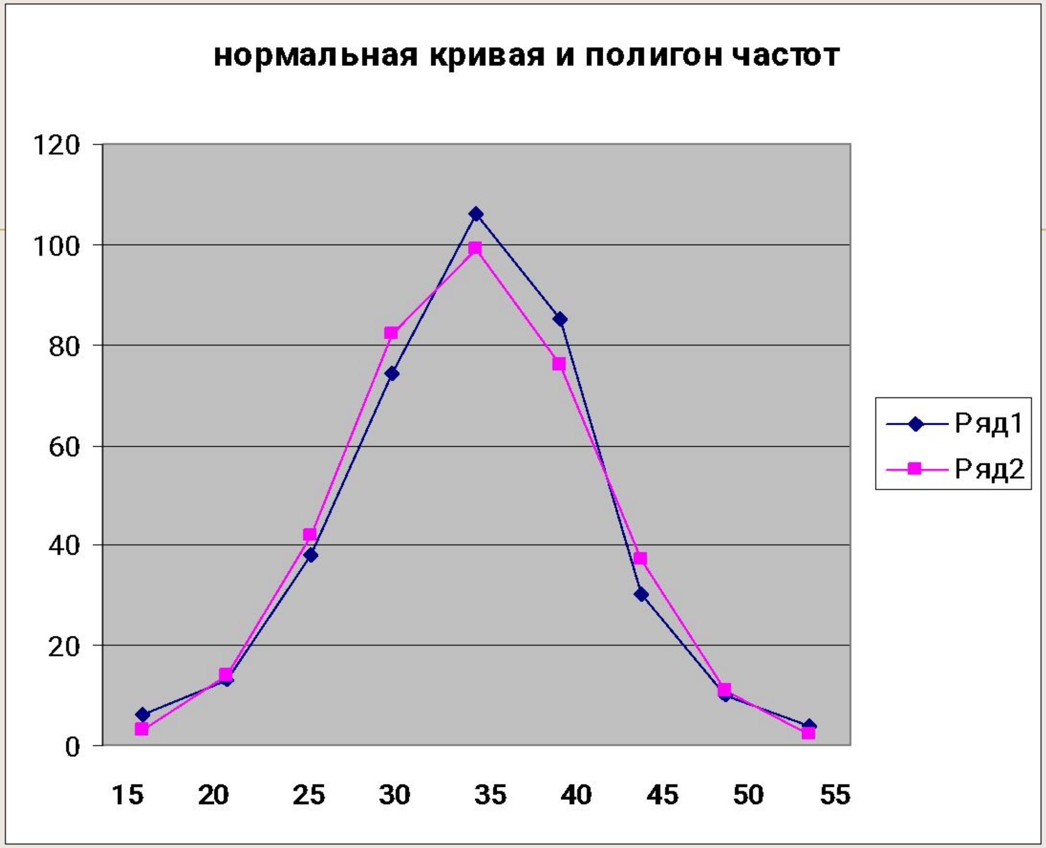
x_i	15	20	25	30	35	40	45	50	55
n_i	6	13	38	74	106	85	30	10	4

Пользуясь методом произведений получим $\overline{x}_в = 34.7$

$\sigma_в = 7.38$

Найдем выравнивающие частоты

x_i	n_i	$x_i - \overline{x}_в$	$u_i = \dfrac{x_i - \overline{x}_в}{\sigma_в}$	$\varphi(u_i)$	$y_i = \dfrac{nh}{\sigma_в} \cdot \varphi(u_i)$
15	6	-19.7	-2.67	0.0113	3
20	13	-14.7	-1.99	0.0551	14
25	38	-9.7	-1.31	0.1691	42
30	74	-4.7	-0.63	0.3271	82
35	106	0.3	0.05	0.3984	99
40	85	5.3	0.73	0.3056	76
45	30	10.3	1.41	0.1476	37
50	10	15.3	2.09	0.0449	11
55	4	20.3	2.77	0.0086	2
	366				366



Для того, чтобы более уверенно считать, что данные наблюдений свидетельствуют о нормальном распределении признака, пользуются специальными правилами – ***критериями согласия*** (рассмотреть самостоятельно!)