Генеральная и выборочная совокупность. Несмещенная оценка. Выборочная средняя. Условные варианты. Генеральная, выборочная, исправленная дисперсия. Асимметрия и эксцесс.

Выполнили

Студентки 3 курса МГПУ ИЦО

Группы МАТ-181

Алешина Анастасия и Кирдяшкина Ирина

Генеральная и выборочная совокупность

- **Генеральная совокупность** – это совокупность объектов, из которой производится выборка.
- Объем совокупности – это число объектов этой совокупности. Объем генеральной совокупности обозначается N
- **Выборочная совокупность** – это совокупность случайно отобранных объектов.
- Объем выборочной совокупности обозначается n

Пример

- С завода на склад поступило 10 тыс. деталей. Необходимо исследовать их на наличие дефектов. Все детали исследовать нет возможности. Поэтому выбирают случайным образом 100 деталей и тщательно обследуют их. По выбранным деталям делают вывод обо всех деталях.
- Что является генеральной совокупностью и выборочной совокупностью?
- Генеральная совокупность – это все 10 тыс. деталей, а выборочная совокупность – это 100 обследованных деталей.

Точечные оценки

- Необходимо определить значение неизвестного параметра θ распределения случайной величины X по выборке x1, x2, . . . , xn.
- **Определение**. Функцию $\theta * = \theta * $ (x1, x2, . . . , xn) называют **точечной оценкой** (статистикой) **параметра $\boldsymbol{\theta}$**, если мы принимаем $\theta \approx \theta *$.
-
- **Точечной** называют оценку, определяющуюся одним числом

Интервальные оценки

- **Интервальной оценкой** называют оценку, определяющуюся двумя концами интервала.

- При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. По этой причине при небольшом объеме выборки следует пользоваться другими оценками. Интервальные оценки позволяют определить точность и надежность оценок.

Несмещенная оценка

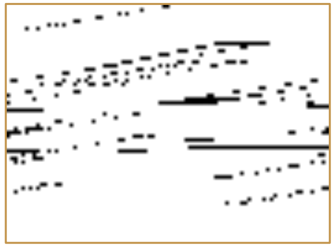
- Несмещенная оценка – это точечная оценка математического ожидания, которая равна оценивающему параметру.
- Оценка θ* называется несмещенной, если M(θ*) = θ. В противном случае оценка называется смещенной.
- Разность d(θ*) = M(θ*) – θ называется смещением оценки θ *.

Генеральная средняя

$$\bar{x}_Г = \frac{\sum_{i=1}^{k} N_i \cdot x_i}{N}$$

Несмещенной оценкой генеральной средней [математического ожидания] служит выборочная средняя

Выборочная средняя



- Из генеральной совокупности извлечена выборка объема *n*, заданная вариантами *Xi* и соответствующими им частотами. Найти несмещенную оценку генеральной средней.

Варианта *Xi*	2	5	7	10
Частота *ni*	16	12	8	14

Объем данной выборки равен

Далее по формуле вычисляем **несмещенную оценку** генеральной средней:

Условные варианты

- Предположим, что варианты выборки расположены в возрастающем порядке, т. е. в виде вариационного ряда.
- *Равноотстоящими* называют варианты, которые образуют арифметическую прогрессию с разностью h.
- *Условными* называют варианты, определяемые равенством
  - $u_i=(x_i-C)/h$,
- где C—ложный нуль (новое начало отсчета); h — шаг, т. е. разность между любыми двумя соседними первоначальными.

- *Условными* называют варианты, определяемые равенством

$$u_i=(x_i-C)/h,$$

- где C—ложный нуль (новое начало отсчета); h — шаг, т. е. разность между любыми двумя соседними первоначальными.

- Замечание 1. В качестве ложного нуля можно принять любую варианту. Максимальная простота вычислений достигается, если выбрать в качестве ложного нуля варианту, которая расположена примерно в середине вариационного ряда (часто такая варианта имеет наибольшую частоту).
- Замечание 2. Варианте, которая принята в качестве ложного нуля, соответствует условная варианта, равная нулю.

- **Пример.** Найти условные варианты статистического распределения: варианты . . . 23,6 28,6 33,6 38,6 43,6 частоты ... 5 20 50 15 10

- Решение. Выберем в качестве ложного нуля варианту 33,6 (эта варианта расположена в середине вариационного ряда).
- Найдем шаг:
  - h = 28,6 —23,6 = 5.
- Найдем условную варианту:
  - $u_1=(x_i-C)/h=(23{,}6 - 33{,}6)/5=-2$.
- Аналогично получим: $u_2=-1$, $u_3=0$, $u_4=1$, $u_5=2$.

Генеральная дисперсия

- *Генеральной дисперсией* $D_г$ называют среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения.

Если все значения признака генеральной совокупности объема N различны, то

$$D_г = \left(\sum_{i=1}^{N}(x_i - \bar{x}_г)^2\right)\Big/N.$$

$$\bar{x}_Г = \frac{\sum_{i=1}^{k} N_i \cdot x_i}{N}$$

Если же значения признака имеют соответственно частоты N_1, N_2, …, N_k, *где* $N_1 + N_2 + \ldots + N_k = N$, *то*

$$D_г = \left(\sum_{i=1}^{k} N_i(x_i - \bar{x}_г)^2\right)\Big/N,$$

Генеральным средним квадратическим отклонением (стандартом) называют квадратный корень из генеральной дисперсии:

$$\sigma_г = \sqrt{D_г}.$$

Пример 1. Генеральная совокупность задана таблицей распределения:

x_i	2	4	5	6
N_i	8	9	10	3

Найти генеральную дисперсию.

Решение: Найдем генеральную среднюю:

$$\bar{x}_Г = \frac{\sum_{i=1}^{k} N_i \cdot x_i}{N} = \frac{8 \cdot 2 + 9 \cdot 4 + 10 \cdot 5 + 3 \cdot 6}{30} = 4.$$

Найдем генеральную дисперсию:

$$D_Г = \frac{\sum_{i=1}^{k} N_i \cdot (x_i - \bar{x}_Г)^2}{N} = \frac{8 \cdot (2-4)^2 + 9 \cdot (4-4)^2 + 10 \cdot (5-4)^2 + 3 \cdot (6-4)^2}{30} = 1{,}8$$

Выборочная дисперсия

- **Выборочная дисперсия** – это среднее арифметическое значений вариантов части отобранных объектов генеральной совокупности (выборки).

- **Выборочная дисперсия** при различии всех значений варианта выборки находится по ф

$$\widehat{D}_{\text{В}} = \frac{\sum\limits_{i-1}^{n} (x_i - \overline{x}_{\text{В}})^2}{n}$$

- Для значений признаков выборочной совокупности с частотами n1, n2,…,nk формула выглядит следующим образом:

$$\widehat{D}_{\text{В}} = \frac{\sum\limits_{i-1}^{k} n_i (x_i - \overline{x}_{\text{В}})^2}{n}$$

- Квадратный корень из выборочной дисперсии характеризует рассеивание значений вариантов выборки вокруг своего среднего значения. Данная характеристика называется выборочным **средним квадратическим отклонением** и имеет вид:

$$\widehat{\sigma}_{\text{В}} = \sqrt{\widehat{D}_{\text{В}}}$$

- **Пример 1**
- Найти выборочную дисперсию выборки со значениями:
- x_i: 1, 2, 3, 4;
- n_i: 20, 15, 10, 5.
- **Решение:**
- Для начала необходимо определить выборочную среднюю:

$$\overline{x}_\text{В} = \frac{1}{50}(1 \cdot 20 + 2 \cdot 15 + 3 \cdot 10 + 4 \cdot 5) = \frac{1}{50} \cdot 100 = 2$$

- Затем найдем выборочную дисперсию:

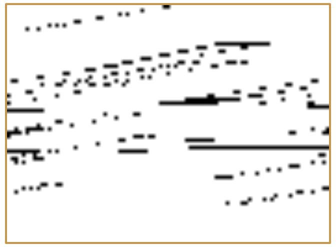
$$D_\text{В} = \frac{1}{50}((1-2)^2 \cdot 20 + (2-2)^2 \cdot 15 + (3-2)^2 \cdot 10 + (4-2)^2 \cdot 5) = 1$$

- Чему будет равно среднее квадратическое отклонение?
- 1

$$\widehat{D}_\text{В} = \frac{\sum\limits_{i-1}^{n} (x_i - \overline{x}_\text{В})^2}{n}$$

$$\widehat{D}_\text{В} = \frac{\sum\limits_{i-1}^{k} n_i(x_i - \overline{x}_\text{В})^2}{n}$$

$$\widehat{\sigma}_\text{В} = \sqrt{\widehat{D}_\text{В}}$$



Исправленная дисперсия

- **Исправленная дисперсия** используется для несмещенной оценки генеральной дисперсии и обозн

$$S^2 = \frac{n}{n-1} \cdot D_{\text{В}} = \frac{\sum\limits_{i=1}^{k} n_i (x_i - \overline{x}_{\text{В}})^2}{n-1}$$

- Среднеквадратическая генеральная совокупность оценивается при помощи исправленного **среднеквадратического отклонения**, которое вычисляется по формуле

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

- **Математическое ожидание** выборочной дисперсии вычисляется так:

$$M\left[D_B\right] = \frac{n-1}{n} D_{\Gamma}$$

- DГ – это истинное значение дисперсии генеральной совокупности.

- **Пример**
- Длину стержня измерили одним и тем же прибором пять раз. В результате получили следующие величины: 92 мм, 94 мм, 103 мм, 105 мм, 106 мм. Задача найти выборочную среднюю длину предмета и выборочную исправленную дисперсию ошибок измерительного прибора.
- **Решение**
- Сначала вычислим выборочную среднюю: $\overline{x}_{\text{В}} = \frac{92+94+103+105+106}{5} = 100$
- Затем найдем выборочную дисперсию:

$$D_{\text{В}} = \frac{\sum_{i=1}^{k}(x_i - \overline{x}_{\text{В}})^2}{n} = \frac{(92-100)^2+(94-100)^2+(103-100)^2+(105-100)^2+(106-100)^2}{5} = 34$$

- Теперь рассчитаем исправленную дисперсию: $S^2 = \frac{5}{5-1} \cdot 34 = 42,5$

- Выборочная совокупность задана следующей таблицей распред

x_i	5	10	15	20	25
n_i	1	5	8	4	2

- Найдем для нее выборочную дисперсию, выборочное среднее квадратическое отклонение, исправленную дисперсию и исправленное среднее квадратическое отклонение.

- Р… а сделаем расчетную таблицу:

x_i	n_i	$x_i n_i$	$\overline{x_в}$	$x_i - \overline{x_в}$	$(x_i - \overline{x_в})^2 n_i$
5	1	5	-	-10,25	105,0625
10	5	50	-	-5,25	137,8125
15	8	120	-	-0,25	0,5
20	4	80	-	4,75	90,25
25	2	50	-	9,75	190,125
Итого:	20	305	15,25		523,75

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{305}{20} = 15,25$$

Найдем выборочную дисперсию по формуле:

$$D_в = \frac{\sum_{i=1}^{k} (x_i - \overline{x_в})^2 n_i}{n} = \frac{523,75}{20} = 26,1875$$

Выборочное среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в} \approx 5,12$$

Исправленная дисперсия:

$$S^2 = \frac{n}{n-1} D_в = \frac{20}{19} \cdot 26,1875 \approx 27,57$$

Исправленное среднее квадратическое отклонение:

$$S = \sqrt{S^2} \approx 5,25$$

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

$$\widehat{D}_В = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x}_В)^2}{n}$$

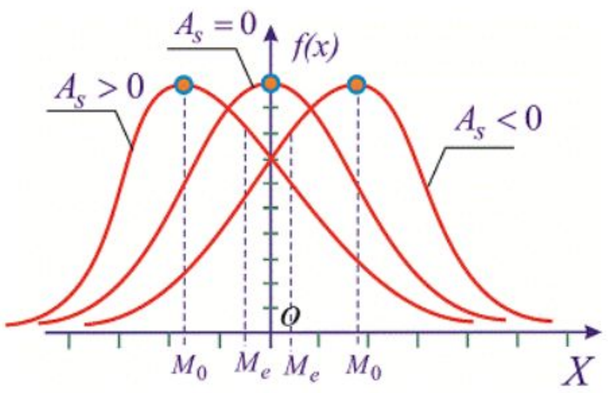
$$\widehat{\sigma}_В = \sqrt{\widehat{D}_В}$$

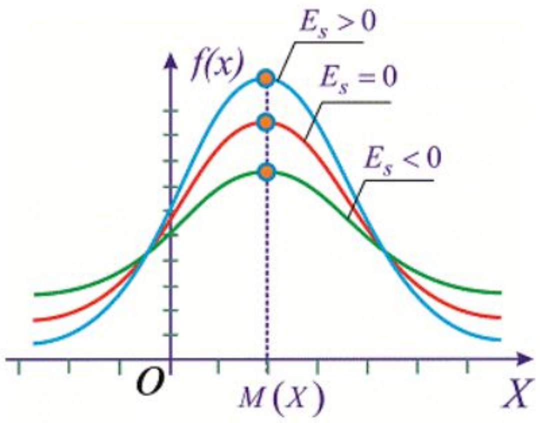
$$S^2 = \frac{n}{n-1} \cdot D_В = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x}_В)^2}{n-1}$$

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

Асимметрия. Эксцесс.

- Вычисление асимметрии и эксцесса позволяет установить симметричность распределения случайной величины X относительно M(X)=1.
- Для этого находят центральный элеме $\mu_3 = 0$, характеризующий асимметрию закона распределения. Если он равен нулю, то величина симметрично распределена относительно мат. ожидания.
- Коэффициент асимметри $As = \frac{\mu_3}{\sigma^3}$.
- Эксцесс: $Es = \frac{\mu_4}{\sigma^4} - 3.$





Пример

Дана плотность вероятностей:

$$f(x)=\begin{cases}0, & x\le -2;\\ \frac{2}{9}(x+2)(1-x), & -2<x\le 1;\\ 0, & x>1.\end{cases}$$

Решение. Вычисляем математическое ожидание

$$M(X)=\int_{-2}^{1} x\,f(x)\,dx=\int_{-2}^{1} x\frac{2}{9}(x+2)(1-x)\,dx=$$

$$=\frac{2}{9}\int_{-2}^{1}(2x-x^2-x^3)dx=-\frac{1}{2};$$

после этого - третий момент инерции

$$\mu_3=\int_{-2}^{1}(x-M(X))^3 f(x)dx=\int_{-2}^{1}(x+\frac{1}{2})^3\frac{2}{9}(x+2)(1-x)dx=$$

$$=\int_{-2}^{1}(\frac{1}{18}+\frac{11}{36}x+\frac{17}{36}x^2-\frac{1}{18}x^3-\frac{5}{9}x^4-\frac{2}{9}x^5)dx=0$$

Поскольку момент нулевой $\mu_3=0$, то и асимметрия равна нулю $As=0$. .Следовательно, возможные значения случайной величины X симметрично распределены относительно единицы $M(X)=1$.. Для вычисления эксцесса Es необходимо найти четвертый момент μ_4 и среднее квадратическое отклонение.

$$\mu_4=\int_{-2}^{1}(x-M(X))^4 f(x)dx=\int_{-2}^{1}(x+\frac{1}{2})^4\frac{2}{9}(x+2)(1-x)dx=$$

$$=\int_{-2}^{1}(\frac{1}{36}+\frac{5}{24}x+\frac{13}{24}x^2+\frac{4}{9}x^3-\frac{1}{3}x^4-\frac{2}{3}x^5-\frac{2}{9}x^6)dx=\frac{243}{560};$$

$$M(X^2)=\int_{-2}^{1} x^2 f(x)\,dx=\int_{-2}^{1} x^2 \frac{2}{9}(x+2)(1-x)\,dx=$$

$$=\frac{2}{9}\int_{0}^{2}\left(2x^2-x^3-x^4\right)dx=\frac{7}{10};$$

Окончательно получим

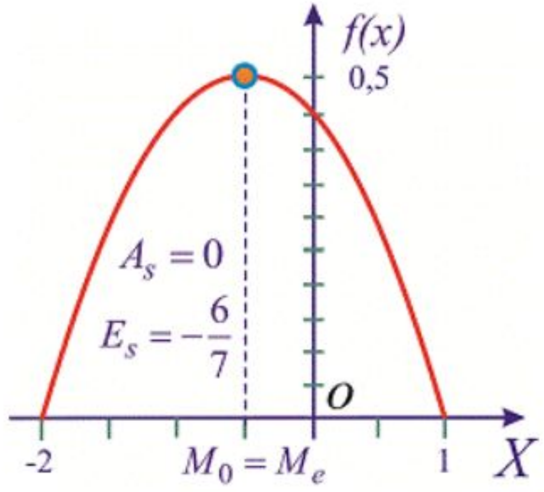
$$Es=\frac{\mu_4}{\sigma^4}-3=\frac{243}{560}\left(\frac{20}{9}\right)^2-3=\frac{15}{7}-3=-\frac{6}{7}.$$

По найденным значениям вычисляем дисперсию

$$D(X)=M(X^2)-M^2(X)=\frac{7}{10}-\left(-\frac{1}{2}\right)^2=\frac{9}{20};$$

$$\sigma(X)=\sqrt{D(X)}=\frac{3}{2\sqrt{5}};$$

отрицательный эксцесс, что указывает на пологость функции распределения. Сам график функции с найденными величинами приведен на рисунку ниже



Спасибо за внимание!