Введение в экономико-математическое моделирование

Лекция 16. Выборочный метод

канд. физ.-матем. наук, доцент Д. В. Чупраков

usr10381@vyatsu.ru

Генеральная совокупность

Совокупность всех подлежащих изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом, называется генеральной совокупностью.

Определение

Генеральная совокупность — это случайная величина X, заданная на пространстве элементарных событий с выделенным в нем классом событий, для которых указаны их вероятности.

N — объем генеральной совокупности

Выборка

Выборка — совокупность объектов, отобранных случайным образом из генеральной совокупности. n — объем выборки.

Определение

Выборка — это последовательность $X_1, X_2, \ldots, X_n$ независимых одинаково распределенных с.в., распределение каждой из которых совпадает с распределением генеральной случайной величины.

Метод статистического исследования, состоящий в том, что на основе изучения выборочной совокупности делается заключение о всей генеральной совокупности, называется выборочным.

Выборочный метод

- Существенно экономит затраты ресурсов.
- Дает возможность проведения углубленного исследования за счет расширения программы исследования.
- Является единственно возможным в случаях, когда:
  - генеральная совокупность бесконечна;
  - исследование связано с уничтожением наблюдаемых объектов.
- Позволяет снизить ошибки регистрации — расхождения между истинным и зарегистрированным значениями признака.

Основное свойство выборки

Репрезентативность — свойство выборки достаточно полно представлять изучаемые признаки генеральной совокупности.

Условие обеспечения репрезентативности

Все объекты генеральной совокупности должны иметь равные вероятности попасть в выборку.

Типы выборки

- Повторная выборка — отобранный объект возвращается в генеральную совокупность перед извлечением следующего;
- Бесповторная выборка — объект не возвращается.

Если объем выборки значительно меньше объема генеральной совокупности, то различие между повторной и бесповторной выборками очень мало.

На практике чаще используется бесповторная выборка.

Способы отбора

- простой — из генеральной совокупности извлекают последовательно по одному объекту;
- типический — генеральную совокупность делят на „типы“ и отбор осуществляется из каждой части
- механический — каждый t–й объект генеральной совокупности.
- серийный — объекты из генеральной совокупности отбираются „сериями“, которые подвергаются сплошному обследованию.

На практике пользуются сочетанием вышеупомянутых способов отбора.

Алгоритм

1. Формирование выборки: $X_1, X_2, \ldots, X_n$.
2. Исследование выборки: Получение значений признака

$$X_1 = x_1, \quad X_2 = x_2, \quad \ldots, \quad X_n = x_n$$

   Получается набор $(x_1, x_2, \ldots, x_n)$ — реализация выборки.
   x_i — значение признака.
3. Ранжирование вариант — расположение значений $x_1, \ldots, x_n$ по неубыванию.
4. Группировка — объединение значений опытных данных в группы так, чтобы в каждой группе признак принимал одно и то же значение.

Связанные понятия

- Конкретные значения $x_1, x_2, \ldots, x_k$ элементов выборки выборки, полученные в результате наблюдений (испытаний) называются вариантами.
- Вариационный ряд — последовательность вариант $x_1, \ldots, x_k$, расположенных в возрастающем порядке

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_k.$$

- Статистический ряд — упорядоченная последовательность групп вариант с соответствующими частотами или относительными частотами принадлежности значений группе.

Дискретный статистический ряд

Дискретный статистический ряд — упорядоченная последовательность вариант с соответствующими частотами n_i или относительными частотами v_i.

Дискретный ряд частот

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

где $n_1 + n_2 + \ldots + n_k = n$

Дискретный ряд частостей (относительных частот) $v_i = \dfrac{n_i}{n}$

x_i	x_1	x_2	$\cdots$	x_k
ω_i	v_1	v_2	$\cdots$	v_k

где $v_1 + v_2 + \ldots + v_k = 1$

Пример

В результате тестирования группа студентов в тестированию по математике абитуриентов набрала баллы:

$$10, 8, 5, 6, 9, 7, 10, 9, 6, 10.$$

Записать полученную выборку в виде статистического ряда.

- Проранжируем выборку: $5, 6, 6, 7, 8, 9, 9, 10, 10, 10$
- Подсчитаем частоту и частость вариант

x_i	5	6	7	8	9	10	Σ
n_i	1	2	1	1	2	3	10
ω_i	0.1	0.2	0.1	0.1	0.2	0.3	1

Изображение статистического ряда

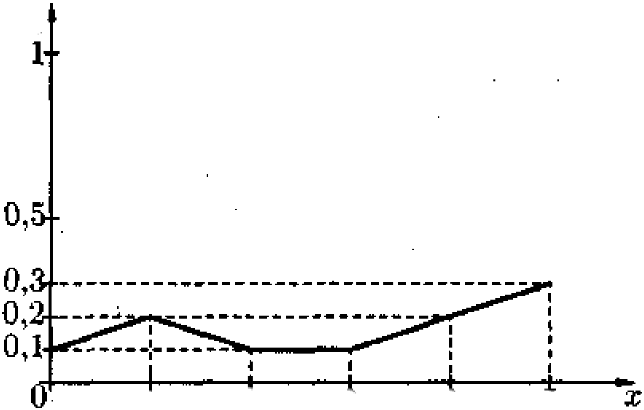
Дискретный Вариационный ряд изображается в виде полигона частот (или частостей)
Полигоном частот называют ломаную, отрезки которой соединяют точки с координатами (x_i, n_i),

Полигоном частостей — ломаную, отрезки которой соединяют точки с координатами с координатами (x_i, ω_i)

Варианты x_i откладываются на оси абсцисс, а частоты n_i (частости ω_i) — на оси ординат.

Пример

x_i	5	6	7	8	9	10	Σ
n_i	1	2	1	1	2	3	10
ω_i	0.1	0.2	0.1	0.1	0.2	0.3	1



Статистическое распределение

Статистический ряд частостей задает статистическое распределение — оценку неизвестного распределения X на генеральной совокупности.

При больших объемах выборки n статистическое распределение мало отличается от истинного распределения.

Интервальный статистический ряд — упорядоченная совокупность интервалов значений случайной величины с соответствующими частотами или относительными частотами попаданий значений величины в каждый из них

x_i	$(x_0, x_1]$	$(x_1, x_2]$	$\cdots$	$(x_{k-1}, x_k]$
n_i	n_1	n_2	$\cdots$	n_k

где $n_1+n_2+\ldots+n_k = n$

x_i	$(x_0, x_1]$	$(x_1, x_2]$	$\cdots$	$(x_{k-1}, x_k]$
ω_i	ω_1	ω_2	$\cdots$	ω_k

где $\omega_1+\omega_2+\ldots+\omega_k = 1$

Характеристики интервалов

- Число интервалов определяется по формуле Стерджеса

$$k = 1 + [3.322 \lg n]$$

- Длина интервала: $h = x_1 - x_0 = x_2 - x_1 = \ldots = x_k - x_{k-1}$ вычисляется как

$$h = \left[\frac{x_{\max} - x_{\min}}{3.322 \lg n}\right]$$

- Левая граница первого интервала:

$$x_0 = x_{\min} - \frac{h}{2}$$

Пример I

Измерили рост (с точностью до см) 30 наудачу отобранных студентов:
153, 154, 155, 155, 156, 157, 158, 159, 160, 163, 164, 165, 166, 167, 167, 169, 170, 171, 171, 172, 173, 173, 175, 175, 178, 179, 179, 182, 183, 186.
Построить интервальный статистический ряд.

- X —рост студента — непрерывная с. в.
- Данные ранжированы,

$$n = 30, \qquad x_{\min} = 153, \qquad x_{\max} = 186$$

Пример II

- Вычислим количество интервалов:

$$k = 1 + [3.322 \lg 30] = 6$$

- Находим длину частичного интервала

$$h = \left[\frac{x_{\max} - x_{\min}}{3.322 \lg n}\right] = \left[\frac{186 - 153}{5}\right] = 6$$

- Находим левую границу:

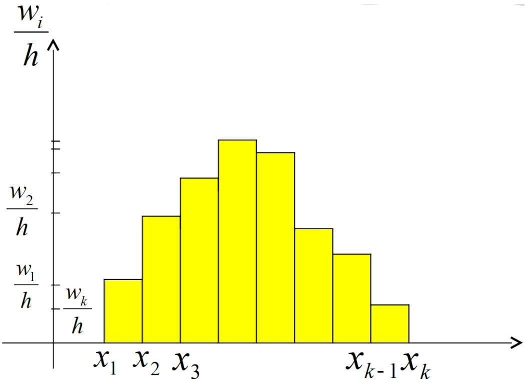
$$x_0 = x_{\min} - \frac{h}{2} = 153 - \frac{6}{2} = 150$$

- Строим интервальный ряд:

x_i	$[150; 156)$	$[156; 162)$	$[162; 168)$	$[168; 174)$	$[174; 180)$	$[180; 186)$	Σ
x_i^*	153	159	165	171	177	183	
n_i	4	5	6	7	5	3	30
ω_i	0,13	0,17	0,20	0,23	0,17	0,10	1

Определение

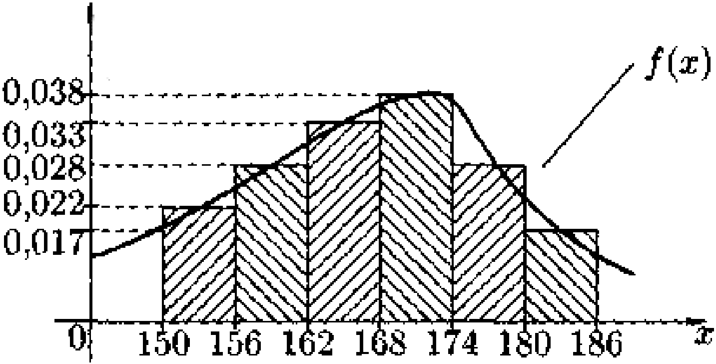
Гистограмма частостей — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы $(x_{i-1}; x_i]$ длиною h, а высоты равны отношению $\frac{\omega_i}{h}$ — плотности частостей.



$$S = 1$$

Пример

x_i	[150; 156)	[156; 162)	[162; 168)	[168; 174)	[174; 180)	[180; 186)	Σ
x_i^*	153	159	165	171	177	183	
n_i	4	5	6	7	5	3	30
ω_i	0.13	0.17	0.20	0.23	0.17	0.10	1
$\frac{\omega_i}{h}$	0.022	0.033	0.038	0, 23	0.028	0.017	



Числовые характеристики

- Средние величины
  - Среднее выборочное $\bar{x}_{\text{в}} = \sum\limits_{i=1}^{k} \omega_i x_i$
  - Мода $Mo_{\text{в}}$ — значение с наибольшей частотой.
  - Медиана $Me_{\text{в}}$ — значение признака, приходящееся на середину вариационного ряда.
- Показатели вариации
  - Размах вариации называется число $R = x_{\max} - x_{\min}$
  - Среднее линейное отклонение: $\bar{x}_{\text{в}} = \sum\limits_{i=1}^{k} \omega_i |x_i - \bar{x}_{\text{в}}|$
  - Выборочная дисперсия $s_{\text{в}}^2 = \sum\limits_{i=1}^{k} \omega_i (x_i - \bar{x}_{\text{в}})^2$
  - Среднее квадратическое отклонение $s_{\text{в}} = \sqrt{s_{\text{в}}^2}$

Дополнительне характеристики

- Начальный момент k-порядка: $\tilde{\nu}_k = \sum\limits_{i=1}^{k} \omega_i x_i^k$
- Центральный момент k-порядка: $\tilde{\mu}_k = \sum\limits_{i=1}^{k} \omega_i (x_i - \bar{x})^k$
- Коэффициент ассимметрии: $\tilde{A} = \dfrac{\mu_3}{s^3}$
  Если $\tilde{A} = 0$ то распределение симметрично.
- Эксцесс: $\tilde{\gamma} = \dfrac{\mu_4}{s^4} - 3$ показывает насколько крут вариационный ряд по сравнению с нормальным распределением.
  Для нормально распределенной величины эксцесс равен 0.

Чтобы вычислить характеристики выборочного распределения, заданного интервальным рядом берут представителя каждого интервала —– значеине, расположенное в его середине.