

Выборочное наблюдение

Определение выборочного наблюдения.

Сплошное наблюдение предусматривает обследование всех единиц изучаемой совокупности. Сплошное наблюдение зачастую невозможно по различным причинам, например высокая стоимость.

Выборочное наблюдение - это такое несплошное наблюдение, при котором отбор подлежащих обследованию единиц осуществляется в случайном порядке, отобранная часть изучается, а результаты распространяются на всю исходную совокупность.

Задачи выборочного метода исследования

- Задачей выборочного метода исследования является правильная оценка показателей, которыми характеризуется генеральная совокупность, по данным, полученным при изучении выборочной совокупности.
- Выяснение степени надежности найденных показателей генеральной совокупности с помощью статистических методов исследования.

Ошибки выборочного наблюдения

Ошибкой выборочного наблюдения, или ошибкой репрезентативности, называется разница между значениями показателя, полученного по выборке и по генеральной совокупности.

Основные понятия выборочного метода наблюдения

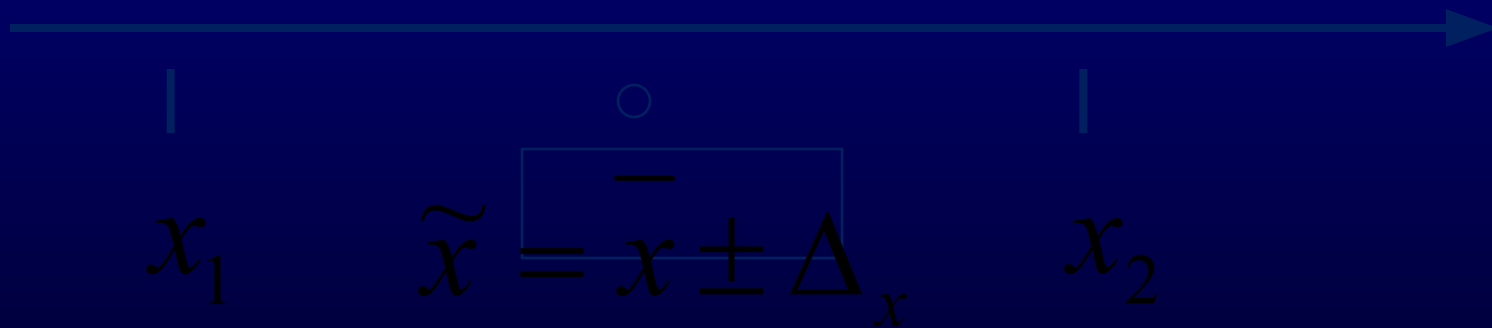
Совокупность, из которой производится отбор, называется *генеральной*, совокупность отобранных единиц - *выборочной*; все показатели, которые характеризуют выборочную совокупность, называются *выборочными* показателями.

Чтобы различать генеральные и выборочные показатели, будем снабжать показатели, рассчитанные для генеральной совокупности дополнительным индексом 0 внизу символа показателя, например \tilde{x}, σ_0^2 .

Для выборочной совокупности сохраним
—
прежнее обозначение (\bar{x}, σ^2)

Интервальная оценка параметра X состоит в определении границ числового интервала (x_1, x_2) , с вероятностью γ накрывающего неизвестное значение параметра.

Для интервальной оценки генеральной средней мы должны построить *доверительный интервал* с центром в точке $x = \bar{x}$ шириной Δ .



В зависимости от метода отбора используют разные способы расчета среднеквадратичной ошибки.

При *повторном* отборе

$$s_x = \sqrt{\frac{\sigma^2}{n}}$$

При *бесповторном* отборе из генеральной совокупности объема N

$$s_x = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

При повторном отборе отобранная единица после обследования возвращается в генеральную совокупность.

При бесповторном отборе единица совокупности, попавшая в выборку, в генеральную совокупность не возвращается.

Значения коэффициента доверия t рассчитаны для разных вероятностей и имеются в специальных таблицах (интеграл Лапласа), из которых в статистике широко применяются следующие сочетания:

Вероятность	0,683	0,866	0,950	0,954	0,988	0,990	0,997	0,999
t	1	1,5	1,96	2	2,5	2,58	3	3,5

При различных способах отбора единиц генеральной совокупности дисперсию выборки вычисляют по-разному.

При *собственно-случайном* и *механическом* отборе дисперсия выборки рассчитывается по обычной формуле.

При *типическом* отборе в качестве дисперсии выборки берется *среднегрупповая* дисперсия

При *серийном* отборе в качестве дисперсии выборки берется *межгрупповая* дисперсия. Кроме того при серийном отборе в формулах объем выборки заменяется на число серий в выборке, а объем генеральной совокупности - на число серий в генеральной совокупности ($n > 30, n \rightarrow r, N \rightarrow R$)

Виды выборки

Различают следующие виды выборки:

- ✓ Собственно–случайную выборку, образованную случайным выбором элементов из генеральной совокупности;
- ✓ Механическую выборку, при которой выбирается, например, каждый десятый элемент из генеральной совокупности;

- ✓ Серийную. выборку, в которую случайным методом отбираются не элементы, а целые группы элементов, которые подвергаются сплошному наблюдению. Например из партии конфет выбираются отдельные коробки, в каждой из которой обследуются все изделия.
- ✓ Типическую выборку, при которой случайным образом отбираются элементы из типических групп, на которые предварительно делится генеральная совокупность по некоторым признакам;

Коэффициенты доверия (t) для расчета предельных ошибок

Вероятность	0,683	0,866	0,950	0,954	0,988	0,990	0,997	0,999
t	1	1,5	1,96	2	2,5	2,58	3	3,5

$$\Delta = S_x * t$$

Задача

Проведено выборочное повторное исследование жилищных условий жителей города, результаты которого представлены в таблице:

Общая площадь в расчете на одного чел, м. кв	До 5	5 - 10	10- 15	15- 20	20-2 5	25-3 0	30- и выше
Число жителей	8	95	204	270	210	130	83

Вычислить чему равно среднее значение жилой площади, приходящейся на одного человека с доверительной вероятностью 0,95.

Составим расчетную таблицу

Решение

Середина интервала X	Частоты f	$X*f$	X^2*f
2,5	8	20	50
7,5	95	712,5	5343,8
12,5	204	2550	31875
17,5	270	4725	82688
22,5	210	4725	106313
27,5	130	3575	98313
32,5	83	2698	87669
Всего	1000	19005	412250

Продолжение вычислений

- Вычислим дисперсию, среднее квадратическое отклонение, среднее значение и интервал, в котором с вероятностью 0,95 будет заключено среднее значение жилой площади, приходящейся на одного человека, в генеральной совокупности

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 f_i}{\sum_{i=1}^n f_i} - \left(\frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \right)^2, \quad \bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}.$$

Результаты и выводы:

$$= \frac{412250}{1000} \left(\frac{19005}{1000} \right)^2 = 51,05;$$

$$= \sqrt{51,05} = 7,14; \quad s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = 0,23;$$

$$= t \cdot s_{\bar{x}} = 1,96 \cdot 0,23 = 0,45; \quad \bar{x} = \frac{19005}{1000} = 19,01$$

$$18,56 \leq \bar{x}_0 < 19,01 + 0,45 \text{ (м}^2\text{)}.$$

$$18,56 \leq X \leq 19,46$$

Задача

При обследовании выработки 1000 рабочих цеха в отчетном году по сравнению с предыдущим по схеме собственно - случайной выборки было отобрано 100 рабочих (полученные данные изображены на след. слайде).

Определить:

а) вероятность того, что средняя выработка рабочих цеха отличается от средней выборочной не более чем на 1%;

б) границы в которых с вероятностью 0,954 заключена средняя выработка рабочих цеха.

Данные о выработке рабочих в отчетном году.

Выработка в отчетном году в % к предыдущему	Число рабочих
94,0 - 100,0	3
100,0 - 106,0	7
106,0 - 112,0	11
112,0 - 118,0	20
118,0 - 124,0	28
124,0 - 130,0	19
130,0 - 136,0	10
136,0 - 142,0	2
Всего	100

Найдем вначале среднее
и дисперсию используя электронные таблицы.

Интервалы	Середина интервалов X	Часто- ты f	$X \cdot f$	$(X - X_{\text{ср}})^2$
94,0 - 100,0	97	3	291	1478,5
100,0 - 106,0	103	7	721	1837,1
106,0 - 112,0	109	11	1199	1144,4
112,0 - 118,0	115	20	2300	352,8
118,0 - 124,0	121	28	3388	90,72
124,0 - 130,0	127	19	2413	1156
130,0 - 136,0	133	10	1330	1904,4
136,0 - 142,0	139	2	278	784,08
			11920	8748
Срзнач=119,2%	Дисперсия =87,48		119,2	87,48

Найдем среднеквадратическую ошибку выборки :

$$S_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}} =$$
$$= \sqrt{\frac{87,48}{100}} = 0,919.$$

Искомую доверительную вероятность найдем из условия ($\Delta = 1\%$)

$$\begin{aligned} P(|\bar{X} - \bar{X}_0| < \Delta) &= 2\Phi(\Delta\sqrt{n-1}/\sigma) = \\ &= 2\Phi\left(\frac{1}{0,94}\right) = 2\Phi(1,06) = 0,710 \end{aligned}$$

Таким образом, вероятность того, что выборочная средняя отличается от генеральной не более чем на 1% равна 0,71. Можно сказать, что в 71 случаях из 100 произведенное выборочное исследование даст ошибку определения средней производительности труда для всего цеха не более чем 1%.

Пример. Оздоровительный центр рекламируя свои услуги, предлагает клиентам за короткий срок снижение своего веса до 10 кг. По результатам выборочного обследования 15 женщин, воспользовавшихся услугами центра, были получены следующие данные о снижении их веса:

№	ΔP	№	ΔP	№	ΔP
1	10,2	6	5,7	11	5,0
2	7,6	7	13,7	12	3,7
3	6,1	8	6,9	13	4,7
4	8,4	9	5,2	14	3,6
5	6,0	10	6,1	15	3,2

Требуется определить на основании полученных данных степень достоверности рекламной информации.

Решение

Выборочная средняя и дисперсия равны

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_{15}}{15} = \frac{96,1}{15} = 6,61 \text{ кг};$$

$$\sigma^2 = \frac{\sum_{i=1}^{15} (x_i - \bar{x})^2}{15} = 7,061.$$

Среднюю квадратическую ошибку выборочного среднего найдем по формуле

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n-1}} = \sqrt{\frac{7,061}{14}} = 0,71 \text{ кг.}$$

Оценим с вероятностью 0,99 предел возможных расхождений выборочной и генеральной средней.

Число степеней свободы в данном случае равно 14, поэтому при уровне значимости 0,01 находим предельное значение параметра $t = 2,97$, которое можно объяснить случайными факторами.

Теперь можно найти и предельную ошибку выборочной средней.

$$t = \frac{\bar{X} - \bar{X}_0}{S_{\bar{X}}}; \quad \Delta = t \cdot S_{\bar{X}} = \\ = 2,97 \cdot 0,71 = 2,11 \text{ кг.}$$

Таким образом с вероятностью 0,99 отклонение генеральной средней не должно превышать 2,11 кг от выборочного среднего значения, равного 6,41 кг. Генеральная средняя будет с вероятностью 0,99 находиться внутри интервала от 4,3 кг до 8,52 кг и следовательно утверждения рекламы не являются обоснованными.

Определение численности выборки

Для повторных наблюдений:

$$n = \frac{\sigma^2 t^2}{\Delta^2}$$

Для бесповторных наблюдений:

$$n = \frac{\sigma^2 t^2}{\Delta^2 + \frac{\sigma^2 t^2}{N}}$$

Дисперсию принимаем приблизительно одним из следующих способов

- Берется из предыдущих выборочных наблюдений;
- Используется правило, согласно которому в размахе вариации укладывается примерно шесть стандартных отклонений;

$$\sigma^2 = \frac{R^2}{36}$$

- Используется правило «трех сигм», согласно которому в средней величине укладывается примерно 3 стандартных отклонения.

$$\sigma^2 = \frac{\overline{X}^2}{9}$$

Для расчета долей

Дисперсия равна $p \cdot q$

Принимаем максимальную дисперсию
 $= 0,25$