



Организация выборочного наблюдения



2016-2017 уч.год

Рассматриваемые вопросы

- Сущность генеральной и выборочной совокупности
- Классификация методов выборки
- Основные типы задач, решаемых при проведении выборки
- Характеристика генеральной совокупности на основе данных, полученных по выборке

Виды наблюдения

- Сплошное наблюдение

- Несплошное наблюдение:

- Способ основного массива
- Выборочное наблюдение
- Монографическое наблюдение

К использованию выборочного метода (или выборки) прибегают в следующих случаях:

- если само наблюдение связано с порчей или уничтожением наблюдаемых единиц;
- если необходимо получить информацию о слишком большом объеме совокупности, а возможности привлечения большого штата сотрудников для сбора данных ограничены;
- если исследование больших совокупностей необходимо провести в сжатые сроки или при небольших затратах;
- если необходимо повысить точность наблюдения: уменьшение числа единиц наблюдения резко снижает ошибки регистрации.

Отбор единиц в выборочную совокупность может быть повторным и бесповторным.

■ **Повторный отбор**

При таком отборе вероятность попадания каждой отдельной единицы в выборку остается постоянной, так как отобранная единица после обследования снова возвращается в генеральную совокупность и снова может быть выбранной.

■ **Бесповторный отбор.**

При таком отборе каждая отобранная единица не возвращается в генеральную совокупность, а, следовательно, вероятность попадания в выборку оставшихся единиц все время меняется.

Классификация выборочных методов



Пример:

Для изучения платежеспособного спроса населения было решено провести опрос 2000 чел., причем обеспечить представительство жителей городов и поселков пропорционально численности проживающих в этих населенных пунктах. Какая выборка будет произведена?

Ответы:

- а) механическая;
- б) типическая;
- в) серийная;
- г) случайная.
- д) квотная
- е) удобная

Подходы к определению объема выборки

- Исходя из имеющихся в наличии средств
- Правило «большого пальца»
- Исходя из заранее оговоренных условий
- На основе статистических методов

- При индивидуальном повторном отборе:

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

- При индивидуальном бесповторном отборе:

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

t – коэффициент, связанный с вероятностью (P), гарантирующей результат.

При $P = 0.954$ $t = 2$;

При $P = 0.997$ $t = 3$;

σ^2 – общая дисперсия признака;

Δ - предел ошибки выборки;

N - объем генеральной совокупности.

Величина σ^2 зачастую бывает неизвестна, поэтому используют приближенные способы ее оценки:

- можно провести так называемое пробное маркетинговое исследование (для небольшого объема), на базе которого и определяется величина дисперсии признака :

$$\sigma^2 = \frac{\sum (X_i - X_{\text{проб.}})^2}{n_{\text{проб.}} - 1}$$

- можно использовать данные прошлых выборочных обследований. Если структура и условия развития явления достаточно стабильны, то $\sigma \approx 1/3 X$;
- если распределение признака в генеральной совокупности подчиняется нормальному закону, то размах вариации приблизительно равен 6σ (крайние значения отстоят в ту и другую сторону от средней на расстояние 3σ , т.е. $\sigma = 1/6 (X_{\max} - X_{\min})$);
- для относительной величины признака принимают максимальную величину дисперсии $\sigma^2 = 0,5 * 0,5 = 0,25$.

Пример расчета объема выборки

- Фирма- производитель бытовой техники изучала в одном из регионов степень обеспеченности населения товарами бытовой техники. В ходе предыдущих исследований было выявлено, что $\frac{1}{4}$ семей имеют моющие пылесосы. Каков должен быть объем выборки семей в предстоящем исследовании, чтобы гарантировать результат с вероятностью 95,4 % и ошибкой не более 5 % ?

Итак, имеются все исходные данные для расчета объема выборки: $t=2$ (для вероятности 95,4 %); $\sigma^2 = 0,25 \cdot 0,75 = 0,188$; $\Delta^2 = (0,05)^2 = 0,0025$.

Подставим исходную информацию в формулу расчета объема выборки (n):

$$n := \frac{2^2 \cdot (0.25 \cdot 0.75)}{(0.05)^2} = 300$$

Ответ: Объем выборки составляет 300 семей

Определение предела ошибки выборки

- **Предел ошибки выборки** – величина возможных отклонений показателей генеральной совокупности от показателей выборочной совокупности. Предельная ошибка (Δ) зависит от средней ошибки выборки (μ) и от величины вероятности, с которой гарантируется результат выборочного наблюдения. Обычно вероятность принимается равная 0,954 или 0,997, которой соответствуют коэффициенты (t), равные 2 или 3. Между названными показателями существует взаимосвязь: $\Delta = t\mu$.

Средняя ошибка выборки (μ) рассчитывается по формулам:

- для повторного отбора:

$$\mu = \sigma \sqrt{\frac{2}{n}}$$

- для бесповторного отбора

$$\mu = \sigma \sqrt{\frac{2}{n} \cdot \frac{N-n}{N}}$$

Нередко на практике задаются величиной не абсолютной, а относительной погрешности, выраженной в процентах к средней:

$$\Delta_{\text{отн.}} = \Delta_{\text{абс.}} / X * 100 \%$$

$$\Delta_{\text{абс.}} = \Delta_{\text{отн.}} * X / 100 \%$$

Пример расчета абсолютной погрешности и объема выборки.

Меховое акционерное общество «Белка» проводит исследование мнения потенциальных покупателей о приемлемой цене на норковые мужские шапки. В начале сезона средняя цена в магазинах на шапку- ушанку составляла 4500 руб., со средним квадратическим отклонением 1000 руб. Каков должен быть объем выборки, чтобы гарантировать результат с вероятностью 95,4 % и ошибкой не более 3 %?

Последовательность расчета:

$$\Delta_{\text{абс}} = 3 * 4500 : 100 \% = 135$$

$$n := \frac{2^2 \cdot 1000^2}{(135)^2} = 219.479$$

Ответ: Абсолютная погрешность равна 135 руб., а объем выборки – 220 чел. (округляем в сторону увеличения, т.к. 219 человек недостаточно для обеспечения репрезентативности выборки).

Характеристика генеральной совокупности на основе данных, полученных по выборке

- Выборочные характеристики распространяются на генеральную совокупность с учетом возможной средней ошибки выборки μ , либо предельной ошибки - $\Delta = t\mu$, т.е. устанавливается доверительный интервал, в который, как ожидается, попадут оценки для совокупности в целом.

Доверительный интервал

Под **доверительным интервалом** понимают диапазон, крайним точкам которого соответствует определенный процент ответов на какой-либо вопрос. Из свойств нормальной кривой распределения вытекает, что конечные точки доверительного интервала, для вероятности 95.4 %, определяются как $X \pm 2\mu$, а для вероятности 99.7 % - $X \pm 3\mu$. Имеются специальные таблицы, которые дают возможность определять доверительные интервалы с различной вероятностью.

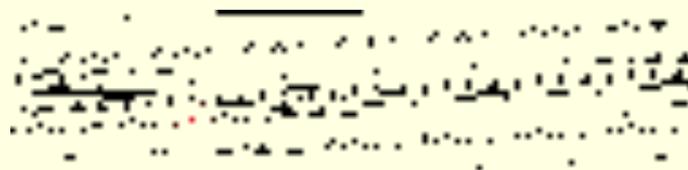
Пример:

- Допустим, что в выборочное обследование мнений потенциальных потребителей нового продукта попали 200 женщин и 300 мужчин. 70 % женщин и 80 % мужчин одобрили новый продукт. С вероятностью 95.4 % определим доверительный интервал доли мужчин и женщин в генеральной совокупности, которые одобрили бы продукт этот продукт

Результаты выборочного наблюдения

Группы лиц, попавших в выборку	Численность (чел.)	Доля лиц, одобряющих продукт (P)	Дисперсия $\sigma^2 = P * (1 - P)$
Женщины	200	0,7	0,21
Мужчины	300	0,8	0,16
Итого	500	0,76	0,18

- Средняя ошибка выборки равна:



При $t = 2$ $\Delta = 2 * 0,019 = 0,038$; следовательно, в генеральной совокупности доля лиц, которым понравится продукт будет находиться в доверительном интервале:

$$0,76 - 0,038 < P < 0,76 + 0,038$$

$$0,722 < P < 0,798$$

Таким образом, с вероятностью 95.4 % можно утверждать, что от 72 до 80 % населения одобряют данный продукт.

Вариационный ряд:

Использование результатов выборочного наблюдения

Пример: В результате выборочного наблюдения населения, ищущего работу, получен следующий ряд распределения .

С вероятностью 0,954 определите границы:

- а) среднего возраста незанятого населения;
- б) удельного веса лиц, моложе 25 лет, в общей численности

Возраст, лет	До 25	25-35	35-45	45-55	55 и более
Численно сть	15	37	71	45	22

Расчет среднего возраста незанятого населения и дисперсии

Возраст	Численность	Середина интервала	xf	X^2f
До 25	15	20	300	6000
25-35	37	30	1100	33300
35-45	71	40	2840	113600
45-55	45	50	2250	112500
55 и более	22	60	1320	79200
Итого	190	-	7820	344600

- а) средняя величина: $x = 7820/190=41,2$
- б) дисперсия: $\sigma^2 = 116,24$
- в) среднее квадратическое отклонение:
 $\sigma=10,78$
- Средняя ошибка выборки
 $\mu = \frac{10,78}{\sqrt{190}} = 0,8$ года

$$\Delta_{\text{Абс}} = 2 * 0,8 = 1,6 \text{ года}$$

$$41,2 - 1,6 < X < 41,2 + 1,6$$

$$39,6 < X < 42,8$$

2-ой вопрос: определение доли лиц до 25 лет в генеральной совокупности

- Доля лиц в возрасте до 25 лет:

- $15/190=0,079$

- Дисперсия: $0,079*0,921=0,073$

- Средняя ошибка выборки μ :

$$\sqrt{0,073/190} = 0,02$$

- Предел ошибка выборки Δ :

$$2*0,02=0,04$$

- Границы генеральной доли

$$0,079-0,04 < X < 0,079 + 0,04$$

$$0,039 < X < 0,119$$