

Средняя квадратическая / стандартная ошибка выборки

Пояснение для таких, как я, жирафов

Зачем эта презентация?

- Во-первых, «средняя квадратическая / стандартная ошибка выборки» – длинное и сложное название, которое часто обрубают в задачах до «средней» или «стандартной» ошибки. То, что это одно и то же, в свое время было для меня настоящим открытием.
- Эта пресловутая ошибка бывает разная и записывается всегда по-разному, что здорово путает.
- Оказывается, эта штука много где попадает, но постоянно меняет обличья. Из-за этого мы зубрим целую кучу формул, когда можно обойтись одной-двумя.



Она бывает:

- Для выборочной средней \bar{x}
- Для выборочной доли (w , или в учебниках иногда p)



Как ее обозначают?

Как только не измывались над несчастной!

$$S_{\bar{x}}$$

$$\varepsilon_{\bar{x}}$$

$$\mu_{\bar{x}}$$

Это варианты написания стандартной ошибки для средней в лекциях и учебниках. Над ошибкой доли или издевались точно так же, или вообще забыли о ее существовании и записывали сразу формулой, что здорово путает несчастных студентов.



Здесь я обозначу ее через «ε», потому что это, хвала Богам, редкая буква, и ее не перепутать ни с моментом, ни с выборочным СКО.



Собственно, формула

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

(корень из дисперсии на число элементов в выборке или СКО разделить на корень из объема выборки)

Это основная формула, фундамент, основа основ.

Достаточно выучить только её, а дальше просто поработать головой! Как? Читай дальше!



Разновидности и откуда они взялись

I. Для доли.

У доли дисперсия считается необычно. Если долю изучаемого признака взять за p , а долю «всего остального» - за q , то дисперсия равна $p \cdot q$ или $p \cdot (1-p)$. Отсюда взялась формула:

$$\varepsilon_p = \sqrt{\frac{p(1-p)}{n}}$$



Разновидности и откуда они взялись (2)

2. Где взять генеральное СКО?

σ – это, вообще-то, **генеральное СКО**, которое вам в задаче фиг дадут.
Есть выход – выборочная дисперсия S^2 , которая, как всем известно, смещена.

Поэтому оцениваем генеральную так:
(чтобы и не думала смещаться), и подставляем.

$$\sigma^2 = S^2 \times \frac{n}{n-1}$$

А можно сразу так: $\varepsilon = \sqrt{\frac{S^2}{n-1}}$

Но есть такая фишка  **Если $n > 30$** , разница между S и σ крайне мала ©, поэтому можно схитрить и написать проще:

$$\varepsilon = \frac{S}{\sqrt{n}}$$



Разновидности и откуда они взялись (3)

□ «Откуда взялись еще какие-то скобки и энки???»

Есть 2 метода формирования выборки, помним? – повторный и бесповторный. Так вот, все предыдущие формулы годятся для **повторной** выборки **или** когда выборка n по отношению к генеральной совокупности N настолько мала, что **отношением n/N можно пренебречь**.

В случае, когда **прям принципиально, что выборка бесповторная**, или когда в задаче открытым текстом говорится, сколько единиц в генеральной совокупности, **обязательно использовать:**

$$\varepsilon_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \times \left(1 - \frac{n}{N}\right)}$$

$$\varepsilon_p = \sqrt{\frac{p(1-p)}{n} \times \left(1 - \frac{n}{N}\right)}$$



Где используют среднюю квадратическую (стандартную) ошибку?

□ Тесты Стьюдента – ошибка «сидит» в знаменателе

□ t-статистика для подтверждения гипотезы о средней.

□ t-статистика для подтверждения гипотезы о равенстве средних.

□ Z-тесты о долях

$$Z = \frac{w - w_0}{\sqrt{p(1-p)/n}}$$

□ Предельная ошибка средней

□ И еще где-то)

$$t = \frac{x - a}{\sqrt{\frac{S^2}{n}}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n}}}$$

$$Z = \frac{\bar{f}}{\frac{S_f}{\sqrt{n}}}$$



Спасибо за чтение)

что-то прояснилось?