

Стандартная формула расчёта доверительного интервала

#статистика, #доверительный-интервал, #гипотезы

Стандартная формула расчёта доверительного интервала выглядит примерно так:

$$\text{СИГМА} \times \text{SQRT}(\text{КОНВЕРСИЯ} \times (1 - \text{КОНВЕРСИЯ}) / \text{ВЫБОРКА})$$

СИГМА — это та самая точка на нормальном распределении, о котором мы говорили в прошлый раз. К примеру для 68% точности нам нужна 1 сигма, а для 95% точности нам нужно 1,96 сигмы. Для 99% точности нам нужно 2,58 сигмы.

1 сигма — это 68% точности. 1.28 сигмы — это 80% точность. 1.96 сигмы — это 95% точности. 3 сигмы — это 99,72% точности. В интернет-маркетинге такая точность не нужна.

Пример:

Баннер А — CTR=5% Баннер Б — CTR=5.5%

На обоих баннерах выборка (показы): 10 000. Какой из этих баннеров лучше с 80% точностью?

Считаем:

$$\text{СИГМА для } 80\% = 1.28$$

$$\text{Доверительный интервал для Баннера А } 1.28 \times \text{SQRT}(5\% \times (1 - 5\%) / 10\,000) = 0.28\%$$

$$\text{Доверительный интервал для Баннера Б } 1.28 \times \text{SQRT}(5.5\% \times (1 - 5.5\%) / 10\,000) = 0.29\%$$

Для «Баннера А» доверительный интервал будет:

$$\text{CTR}(\text{min}) = 5\% - 0.28\% = 4.72\% \quad \text{CTR}(\text{max}) = 5\% + 0.28\% = 5.28\%$$

Для «Баннера Б» доверительный интервал будет:

$$\text{CTR}(\text{min}) = 5.5\% - 0.29\% = 5.21\% \quad \text{CTR}(\text{max}) = 5.5\% + 0.29\% = 5.79\%$$

Таким образом баннеры с 80% вероятностью не различимы. Они пересекаются на диапазоне между 5.21% и 5.28%. Нужно больше выборки, чтоб доверительный интервал для каждого баннера сузить и выявить победителя.