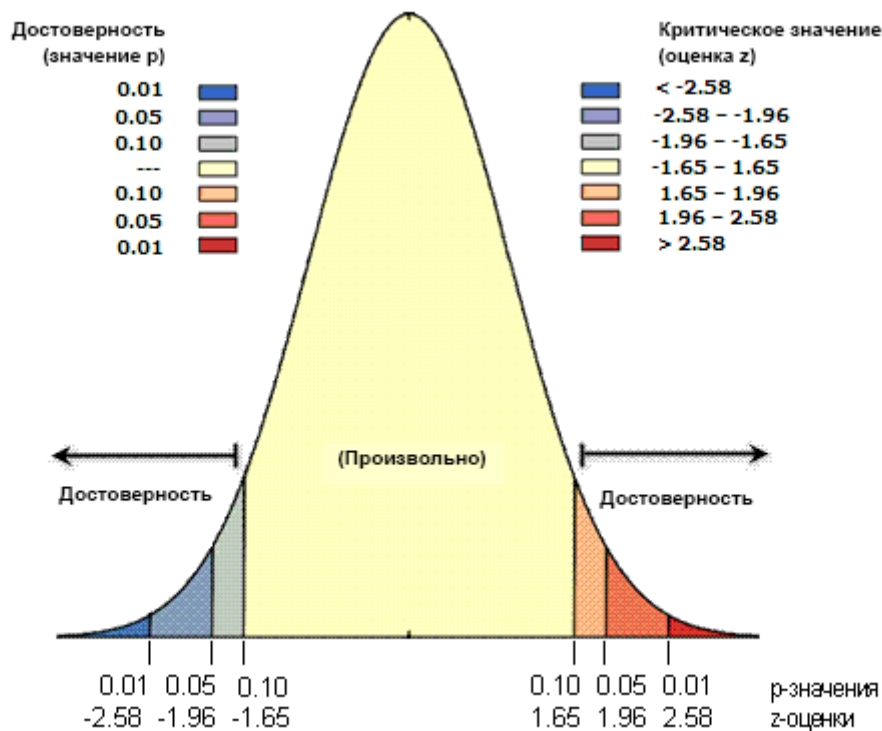


Лекция 5

Z – value

Определение:

Z – value - количество стандартных отклонений на которые вы отступили от матожидания.



$$X = \{X_1 \dots X_n\}$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Где \bar{X} - среднее.

Так как σ и μ не всегда (почти никогда) известны, то мы можем оценить $\sigma \sim S$, а μ взять из гипотезы.

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

При больших n t_{n-1} (Количество **степеней свободы** равно $n - 1$) будет сходиться к нормальному. Тест для проверки гипотез про среднее t -test. Самый простой t -test - проверить, что среднее равно μ .

Условия применения:

- $\bar{X} \sim N$ (стремится к такому по ЦПТ)
- Независимость

Работа с несколькими генеральными выборками

S_p - средняя двух генеральных совокупностей, в случае если мы хотим сравнить похожие вещи и их дисперсии равны.

$$S_p = \sqrt{\frac{S_{X_1}^2 + S_{X_2}^2}{2}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}} \cdot \sqrt{2}$$

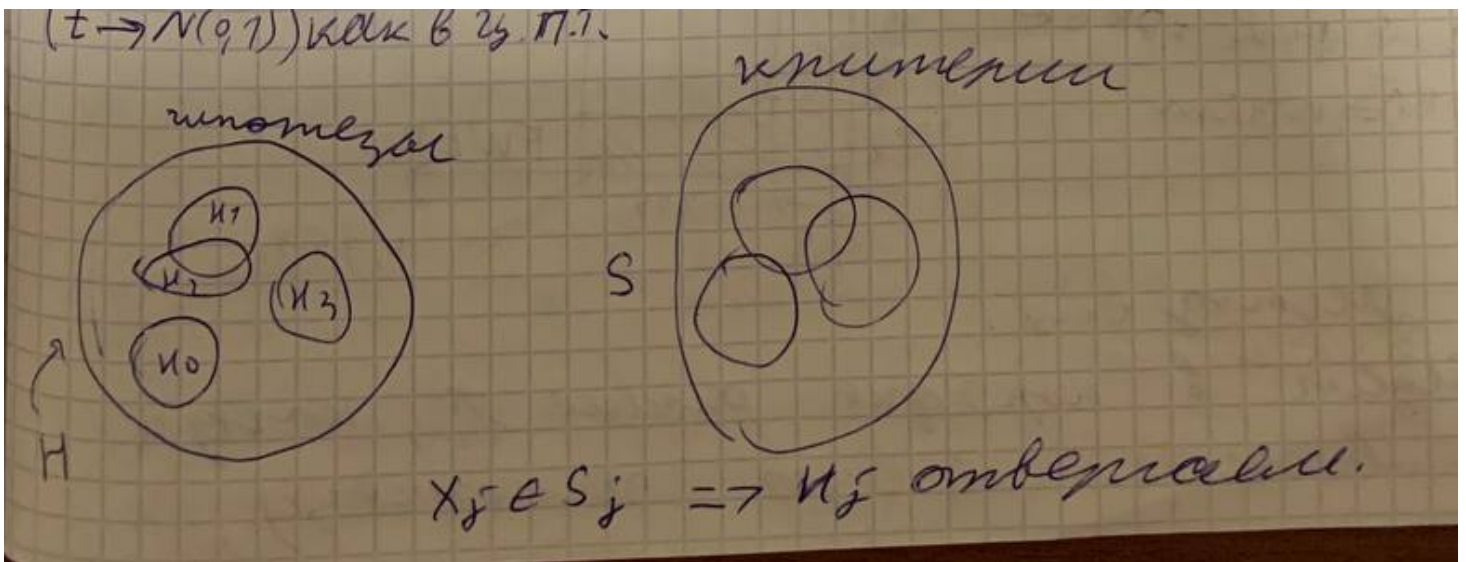
Это позволяет увеличить количество элементов, а как следствие уменьшить ошибку.

Множественные сравнения

□ у вас есть 2 генеральные совокупности, отличающиеся очень не сильно, H_0 - у них одинаковое распределение. В какой-то момент мы получили маленькое p — *value*, из-за чего H_0 возможно должна быть отвергнута.

Варианты:

1. Пихакинг (берём только нужные результаты)
2. Взять множество гипотез и определить уровень статистической значимости по другому



Определение:

$V_{s,p}(X)$ - множество ошибок I рода, где s - критерии, p - вероятности событий, X - выборка.

Определение:

$FWER = P(V_s(X) > 0)$ - вероятность ошибки I рода.

$1 - (1 - \alpha)^n$ - вероятность совершить ошибку I рода при одинаковых экспериментах.

Как видно, если взять большую α вероятность совершить ошибку растёт, поэтому стоит делать поправку на количество экспериментов, так, чтобы $FWER < \alpha$.

Поправки

Поправка Бонферони

m - число экспериментов

α_i - статистическая значимость для i - ой гипотезы.

$$\alpha_i = \frac{\alpha}{m}$$

Метод Холла

Пронумеруем гипотезы упорядочив p - value по возрастанию.

$$\alpha_i = \frac{\alpha}{m-i+1}$$

$$\alpha_1 = \frac{\alpha}{m}, \alpha_m = \alpha$$

Извините за шакалов ☹️



FDR - вероятности ошибки II рода
 $FDR = E\left(\frac{V}{R}\right)$, где R - количество гипотез.

Лемма:

$$FDR \leq FWER$$

Доказательство:

$$FDR = E\left(\frac{V}{R} \cdot I\{V > 0\}\right) \leq E(I(V > 0)) = P(V > 0) = FWER$$