Упарени t тест

- $(X_1,\dots,X_n)$ и $(Y_1,\dots,Y_n)$ — два зависна узорка, али су узорци $(X_1,Y_1),\dots,(X_n,Y_n)$ независни

Пример: X_i – време за које i-ти такмичар претрчи 100 m без тренинга, а Y_i - време за које он претрчи 100 m након тренинга.
- X_i и Y_i су зависне али можемо претпоставити да су времена различитих такмичара независна

- Овај тест се користи када претпостављамо да:

$$D_i := X_i - Y_i \sim \mathcal{N}(m,\sigma^2)$$

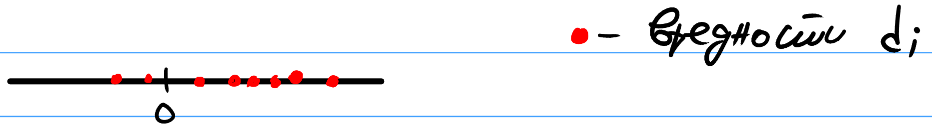
и њиме тестирамо $H_0: m=m_0$ vs $H_1: m \underset{>}{\overset{<}{\neq}} m_0$

Дакле, овај тест је користан за испитивање понашања разлике између 2 обележја, али само под датим претпоставкама.

Погледајмо 41. задатак:

41. Нека $X_1,\dots,X_{20}$ одговарају ценама у 1. супермаркету, а $Y_1,\dots,Y_{20}$ у другом. Дакле X_i и Y_i ће бити цене истог артикла изабраног у i-том извлачењу у 2 различита супермаркета.
Ми желимо да проверимо да ли су цене у 1. супермаркету веће него у другом. Како бисмо то учинили и када кажемо да су цене у једном супермаркету веће?

Простим речима, када је просечна разлика у ценама из два 2 супермаркета позитивна. У тако нешто ћемо поверовати ако видимо да се $d_1, d_2, \dots, d_{20}$ (реализоване вредности горе дефинисаних $D_1, \dots, D_{20}$) већином налазе десно од нуле, тј. ако за већи број артикала важи $d_i = x_i - y_i > 0$.



Претпоставка спареног t теста је да је $(D_1, \dots, D_{20})$ ПСУ из $\mathcal{N}(m, \sigma^2)$, па овакве вредности d_i могу сугерисати на то да је $m>0$.
Зато ћемо тестирати хипотезу: $m>0$ против неке њој супротне, коју ћемо покушати да оповргнемо. Нека то буде нул. хипотеза: $m=0$.
Дакле, тестираћемо:

$$H_0: m=0 \quad vs \quad H_1: m>0$$

Видимо да је ово стандардни случај теста:

$$H_0: m=m_0 \quad vs \quad H_1: m>m_0, \quad \text{где је } m_0=0.$$

Тест статистика која се користи је $\boxed{T=\frac{\bar{D}_n - m_0}{\tilde{S}_D}\sqrt{n}}$ и важи:

ако је H_0 тачна, $T \sim t_{n-1}$.

$\bar{D}_n$ – узор. средина $(D_1, \dots, D_n)$
$\tilde{S}_D$ – попр. узор. дисп. $(D_1, \dots, D_n)$

рекли смо да је корисно знати расподелу тест статистике под нултом хипотезом.

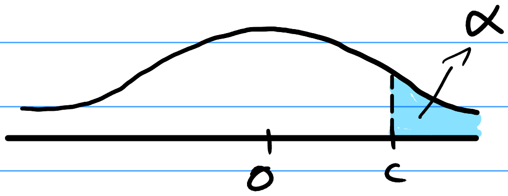
Критична област: $\boxed{W=\{T>c\}}$

Ако је $H_1: m<m_0 \to W=\{T<c\}$, а ако је $H_1: m \neq m_0 \to W=\{|T|>c\}$

Дакле, у овом задатку је $T = \frac{\bar{D}_n}{\tilde{S}_d}\sqrt{n}$.

$W = \{T > c\}$, c је т.д:

$$P_{H_0}\{T > c\} = \alpha \Rightarrow c = F^{-1}_{t_{n-1}}(1-\alpha)$$



- Густина t_{n-1} расподеле

У R-у: $c = qt(1-\alpha, n-1) = qt(0{,}95, 19) = 1{,}729133$

код нас: $\alpha = 0{,}05$, $n = 20$

Преостаје нам још да израчунамо реализовану вредност T и да видимо да ли је она упала у W. Ако јесте, одбацујемо H_0, ако није, не одбацујемо H_0.