(10) Testy o strednej hodnote

1. Priebežná písomka (normálne rozdelené dáta + otázka, či je stredná hodnota rovná danej konštante)

Pred písomkou vznikli po otázke o očakávanom priemere hypotézy, že to bude 10 bodov, iný tip bol, že to bude 20 bodov.

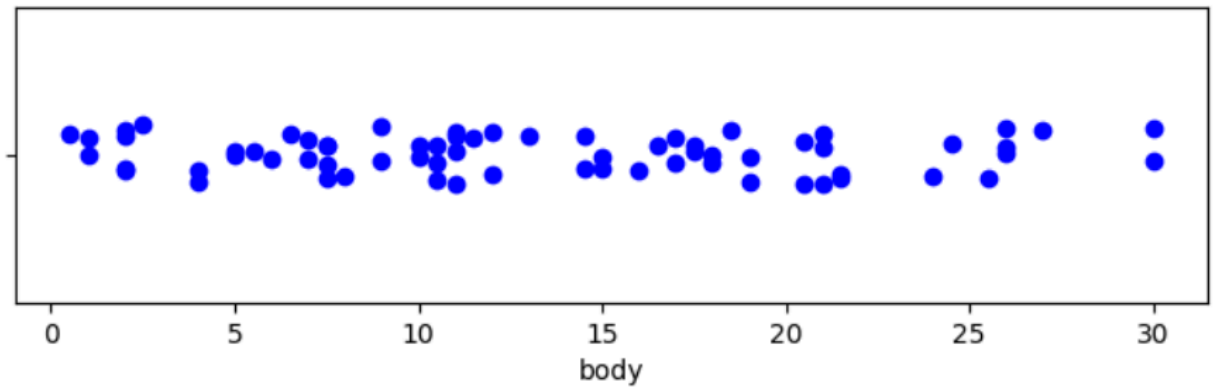
Anonymné výsledky:

body	0.5	1.0	2.0	2.5	4.0	5.0	5.5	6.0	6.5	7.0	7.5	8.0	\
count	1	2	4	1	2	2	1	1	1	2	4	1	

body	9.0	10.0	10.5	11.0	11.5	12.0	13.0	14.5	15.0	16.0	16.5	17.0	\
count	2	2	3	4	1	2	1	2	2	1	1	2	

body	17.5	18.0	18.5	19.0	20.5	21.0	21.5	24.0	24.5	25.5	26.0	27.0	30.0
count	2	2	1	2	2	3	2	1	1	1	3	1	2

Vizualizácia (`sns.stripplot`):



Uvidíme, že bude užitočný výsledok testu normality - normálne rozdelenie nezamietame:

```
from statsmodels.stats.diagnostic import lilliefors

statistika, p_hodnota = lilliefors(body)
print(f"štatistika: {statistika:.5f}")
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
štatistika: 0.09818
p-hodnota: 0.14754
```

Náhodnú premennú vyjadrujúcu počet bodov z písomky teda môžeme modelovať normálnym rozdelením. **Chceme testovať hypotézu o strednej hodnote tohto normálneho rozdelenia.**

- **Nulová hypotéza: stredná hodnota sa rovná 10**
- **Alternatívna hypotéza: nerovná sa 10**

Pripomeňme si princíp testovania hypotéz na príklade, v ktorom sme testovali schopnosť človeka rozlíšiť dva druhy cukríkov:

- *Človek tvrdí, že dokáže na základe chute rozlíšiť Lentilky a M&M's. Nie stopercentne, ale tvrdí, že s pravdepodobnosťou 0,85 to dokáže. My pochybujeme, že to dokáže až tak presne. Chceme preto zrealizovať experiment, na základe ktorého budeme vedieť (s určitou presnosťou) rozhodnúť, či jeho tvrdenie o schopnosti rozlišovať cukríky s uvedenou pravdepodobnosťou 0,85 zamietnuť alebo nie.*
- *Intuícia: Necháme ho určovať druh cukríkov (so zaviazanými očami, cukríky podávame na lyžičke, aby nemohol hmatom zistiť ich veľkosť, možnosť napiť sa medzi pokusmi kvôli neutralizácii chute) viackrát a ak bude mať správnych odpovedí príliš málo, jeho tvrdenie o presnosti 0,85 zamietneme. Povedzme, že spravíme 20 pokusov. Kam dať hranicu - koľko správnych odpovedí bude "príliš málo"?*

Čo sme nakoniec spravili:

- Hypotézu o pravdepodobnosti 0,85 (v prospech alternatívy) sme zamietli, ak bol počet správnych odpovedí príliš malý.
- To "príliš malý" sme určili tak, aby pravdepodobnosť zamietnutia pravdivej hypotézy neprekročila 5 percent (alebo inú stanovenú hranicu).
- Pritom sa snažíme čo najviac priblížiť k tým 5 percentám, aby sme na druhej strane maximalizovali pravdepodobnosť odhalenia neplatnej hypotézy.

Čo by sme mohli spraviť analogicky v príklade s písomkou:

- Hypotézu o strednej hodnote 10 zamietneme, ak sa bude priemer "príliš líšiť" od 10.
- **Ak by sme poznali disperziu tohto rozdelenia,** tak by sme vedeli určiť, čo to znamená "príliš líšiť" tak, aby pravdepodobnosť zamietnutia pravdivej hypotézy bola 5 percent:

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$, pričom σ^2 poznáme

$H_0: \mu = 10$

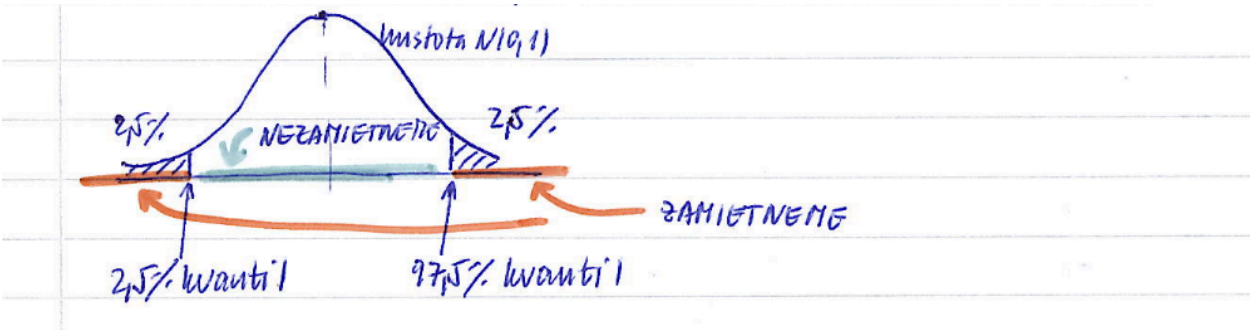
Hypotézu H_0 zamietneme, ak $|\bar{X} - 10| > c$

Teraz vieme povedať:

Ak platí H_0, tak $\bar{X} \sim N\left(10, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X} - 10}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

Myšlienka zamietnuť H_0 v prípade, že sa $\bar{X}$ veľmi líši od 10 znamená, že štatistika $\frac{\bar{X}-10}{\sqrt{\frac{\sigma^2}{n}}}$ je veľmi kladná alebo veľmi záporná. Nakoľko - to určíme tak, aby pravdepodobnosť zamietnutia pravdivej hypotézy bola 0,05 (resp. iná stanovená hodnota)

Keďže už máme rozdelenie N(0,1), v ktorom nie sú neznáme parametre, tak môžeme použiť kvantily tohto rozdelenia:



Číselné hodnoty:

```
from scipy.stats import norm

print(f"2.5% kvantil: {norm.ppf(0.025):.5f}")
print(f"97.5% kvantil: {norm.ppf(0.975):.5f}")
```

```
2.5% kvantil: -1.95996
97.5% kvantil: 1.95996
```

Postup – **v prípade známej disperzie –** teda je:

① Vypočítame priemer $\bar{X}$

② Vypočítame štatistiku $\frac{\bar{X}-10}{\sqrt{\frac{\sigma^2}{n}}}$. Táto štatistika má za platnosti H_0 rozdelenie N(0,1)

③ H_0 zamietneme, ak je štatistika väčšia ako 97,5% kvantil alebo menšia ako 2,5% kvantil N(0,1)

<u>Prakticky ale disperziu nepoznáme</u> (v prípade našich písomiek a tiež vo väčšine iných situácií). Prirodzenou myšlienkou je **nahradiť disperziu výberovou disperziou z dát**. Normálne rozdelenie sa zrejme nezachová, keďže nedelíme konštantou, ale náhodnou premennou. To ale nemusí byť problém, ak sa nám podarí zistiť, aké rozdelenie má podiel, ktorý takto vznikne.

Pozrime sa najskôr na **rozdelenie súčtu druhých mocnín dát od ich priemeru:**

- Najskôr náhodné premenné aj ich priemer vyjadríme pomocou normalizovaných náhodných premenných:

$X_i \sim N(\mu,\sigma^2) \Rightarrow \frac{X_i-\mu}{\sigma} \sim N(0,1)$, a teda X_i môžeme zapísať ako $X_i = \mu+\sigma Z_i$, kde Z_i sú nezávislé náhodné premenné (lebo X_i sú nezávislé) s rozdelením $N(0,1)$. Potom $\bar{X} = \frac{1}{n}\sum X_i = \mu+\sigma\frac{1}{n}\sum Z_i = \mu+\sigma\bar{Z}$

- Teraz budeme analyzovať súčet druhých mocnín odchýliek najskôr pre **n=2:**

Pre $\boxed{n=2}$ dostaneme:

$$\sum_{i=1}^{2}(X_i-\bar{X})^2 = (X_1-\bar{X})^2+(X_2-\bar{X})^2 = ((\mu+\sigma Z_1)-(\mu+\sigma\bar{Z}))^2 + ((\mu+\sigma Z_2)-(\mu+\sigma\bar{Z}))^2 =$$

$$= (\sigma Z_1-\sigma\bar{Z})^2+(\sigma Z_2-\sigma\bar{Z})^2 = \sigma^2\left[(Z_1-\bar{Z})^2+(Z_2-\bar{Z})^2\right]$$

Výraz v zátvorke:

$$(Z_1-\bar{Z})^2+(Z_2-\bar{Z})^2 = \left(Z_1-\frac{Z_1+Z_2}{2}\right)^2+\left(Z_2-\frac{Z_1+Z_2}{2}\right)^2 = \left(\frac{Z_1-Z_2}{2}\right)^2+\left(\frac{Z_2-Z_1}{2}\right)^2 =$$

$$= \frac{1}{4}(Z_1-Z_2)^2+\frac{1}{4}(Z_2-Z_1)^2 = \frac{1}{2}(Z_1-Z_2)^2 = \left(\frac{Z_1-Z_2}{\sqrt{2}}\right)^2$$

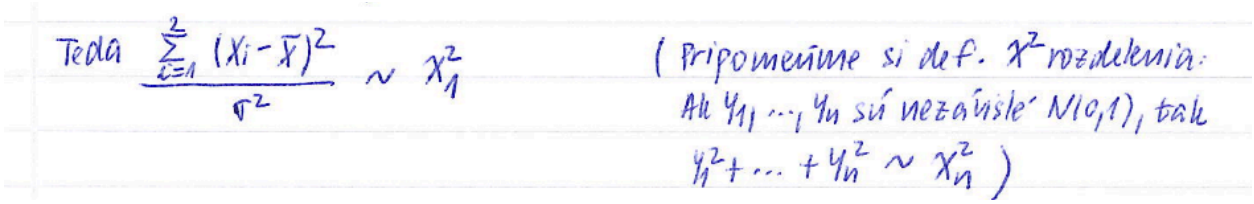
Je to teda druhá mocnina náhodnej premennej

$\frac{Z_1-Z_2}{\sqrt{2}} \sim N(0,1)$,

keďže $E(Z_1-Z_2) = E(Z_1)-E(Z_2) = 0-0 = 0$

$D(Z_1-Z_2) = D(Z_1+(-Z_2)) \overset{\text{nezávislosť}}{=} D(Z_1)+D(-Z_2) = 1+1 = 2$

$D\left(\frac{1}{\sqrt{2}}(Z_1-Z_2)\right) = \frac{1}{2}\cdot D(Z_1-Z_2) = \frac{1}{2}\cdot 2 = 1$



- Vo všeobecnosti je v sume **n** sčítancov, ale nie sú nezávislé. Ak poznáme **n-1** odchýliek od priemeru, posledná je už jednoznačne daná (súčet odchýliek od priemeru je nulový). V prípade **n=2** sme videli, že sa počet stupňov voľnosti chí-kvadrát rozdelenie zmenšil o jednotku v porovnaní so sčitovaním nezávislých náhodných premenných. Vyskúšajme simulačne, či to nebude platiť aj všeobecne:

```
import numpy as np
from scipy.stats import chi2, kstest

def suma_stvorcov_odchyliek(n):
    x = np.random.randn(n)          # IID N(0,1)
    x_priemer = np.mean(x)
    return np.sum((x - x_priemer)**2)

n = 10      # pocet Xi
N = 10**4  # pocet simulacii

simulacie = [suma_stvorcov_odchyliek(n) for _ in range(N)]
```

```
ks_stat, p_hodnota = kstest(simulacie, chi2(n).cdf) # chi^2(n)
print(f"štatistika: {ks_stat:.4f}")
print(f"p-hodnota: {p_hodnota:.4f}")
```

```
štatistika: 0.0963
p-hodnota: 0.0000
```

```
ks_stat, p_hodnota = kstest(simulacie, chi2(n-1).cdf) # chi^2(n-1)
print(f"štatistika: {ks_stat:.4f}")
print(f"p-hodnota: {p_hodnota:.4f}")
```

```
štatistika: 0.0085
p-hodnota: 0.4636
```

- Simulácia bola pre **n=10**, ale dá sa dokázať, že to platí aj všeobecne – takýmto postupom vznikne **chí-kvadrát rozdelenie s n-1 stupňami voľnosti.**

Teda máme:

Za platnosti $H_0 : \mu = \mu_0$ je:

$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$

Pri použití stratégie, že namiesto štandardnej odchýlky použijeme výberovú štandardnú odchýlku, dostaneme:

$$\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}}{\frac{S}{\sigma}} = \frac{\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{S^2}{\sigma^2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$$

Navyše sa dá dokázať, že priemer a výberová disperzia sú nezávislé náhodné premenné. Nie je to zrejmé – pri výpočte výberovej disperzie sa priemer priamo používa. Dokazovať to nebudeme, ukážeme si ale simuláciu:

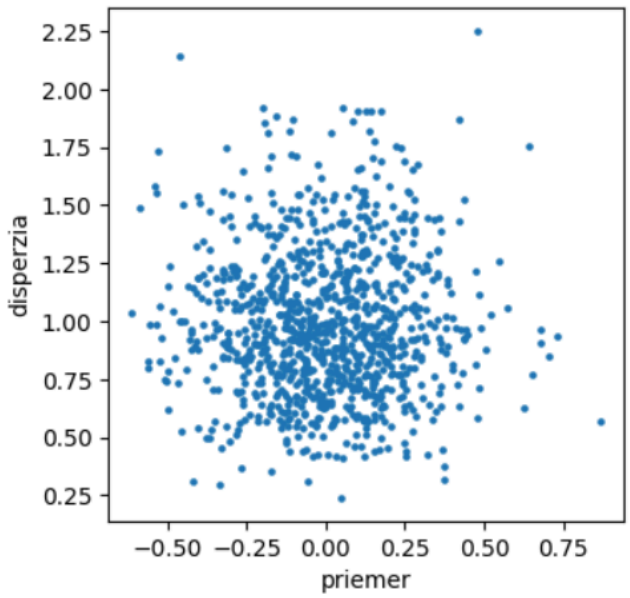
```
import numpy as np
import matplotlib.pyplot as plt

n = 20
N = 1000

priemer = []
disperzia = []

for _ in range(N):
    x = np.random.randn(n)
    priemer.append(np.mean(x))
    disperzia.append(np.var(x, ddof=1))

plt.figure(figsize=(4, 4))
plt.scatter(priemer, disperzia, s=5)
plt.xlabel("priemer")
plt.ylabel("disperzia")
plt.show()
```



Tým pádom dostávame náhodnú premennú so známym rozdelením – **Studentovým (resp. t) rozdelením:**

Nech X je náhodná premenná s rozdelením $N(0,1)$ a Y je náhodná premenná s χ^2 rozdelením s k stupňami voľnosti, pričom X a Y sú nezávislé. Potom náhodná premenná

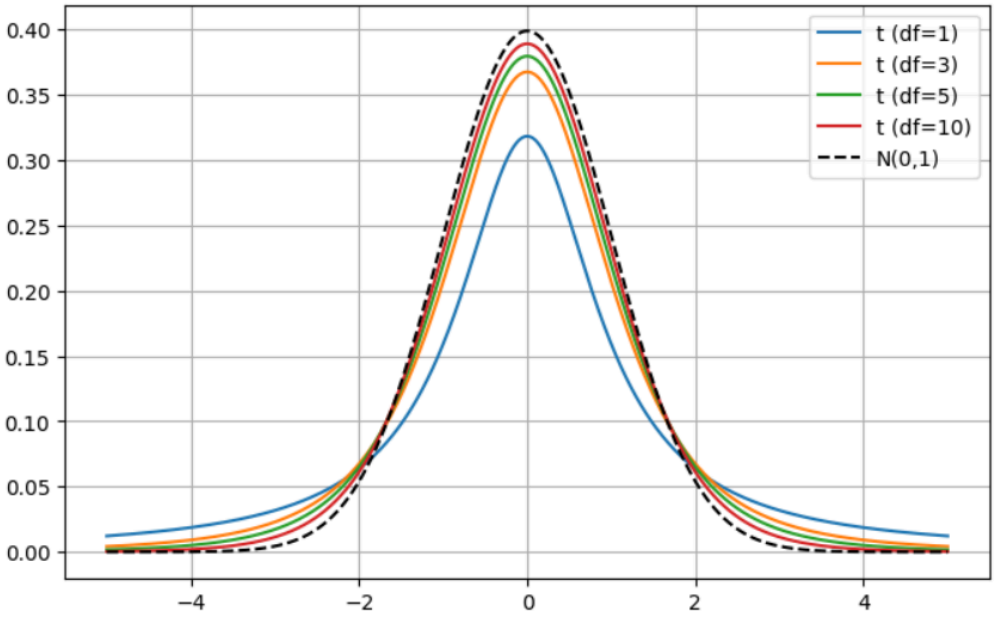
$$T = \frac{X}{\sqrt{\frac{Y}{k}}}$$

má **Studentovo rozdelenie** s k stupňami voľnosti, čo zapisujeme ako $T \sim t_k$. Schematicky:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2_k}{k}}} \sim t_k$$

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html

Hustota sa podobá na hustotu N(0, 1) rozdelenia, ale má “ťažšie chvosty” – väčšiu pravdepodobnosť výraznejších odchýliek od strednej hodnoty (ktorá je nulová – okrem prípadu jedného stupňa voľnosti, kedy príslušný integrál nekonverguje).



Pre momenty platí:

$$\mathbb{E}(t_k) = 0 \text{ pre } k > 1, \ \mathbb{D}(t_k) = \frac{k}{k-2} \text{ pre } k > 2$$

Záver – ako testovať hypotézu o strednej hodnote bez znalosti disperzie:

- Predpoklad: **Xi** sú nezávislé náhodné premenné s rovnakým normálnym rozdelením, parametre nepoznáme
- Nulová hypotéza: stredná hodnota sa rovná zadanej konštane.
- Alternatívna hypotéza: stredná hodnota sa danej konštante nerovná
- Testovacia štatistika:

$$t = \frac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{\overline{X} - \mu_0}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}}\sqrt{n}$$

- Ak platí nulová hypotéza, táto štatistika má Studentovo rozdelenie s **n-1** stupňami voľnosti.
- Na základe tohto rozdelenia vypočítame:
  - pomocou vypočítanej štatistiky p-hodnotu
  - porovnáme štatistiku s kvantilom Studentovho rozdelenia

  Obidva postupy si ukážeme na príklade s bodmi z písomky.

Vrátime sa k výsledkom písomky a testovaniu hypotézy, že stredná hodnota je rovná 10.

<u>Postup 1 - postup, ktorý sme odvodili:</u>

```
import numpy as np
from scipy.stats import t

mu_0 = 10.0

n = len(body)
priemer = np.mean(body)
odhad_sigma2 = np.var(body, ddof=1)
```

Otázka teda je, či je priemer signifikantne rôzny od testovanej hodnoty 10:

```
print(f"priemer bodov: {priemer:.5f}")
```

```
priemer bodov: 13.26471
```

Odbočka - odhadovanie disperzie, parameter **ddof:**

```
# ako skuska
scitance = [(x - priemer)**2 for x in body]
odhad_sigma2_rucne = sum(scitance)/(n - 1)
odhad_sigma2_bez_ddof = np.var(body)

print(f"odhad sigma^2: {odhad_sigma2_rucne:.5f}")
print(f"odhad sigma^2 s parametrom ddof=1: {odhad_sigma2:.5f}")
print(f"odhad sigma^2 bez ddof: {odhad_sigma2_bez_ddof:.5f}")
```

```
odhad sigma^2: 62.33187
odhad sigma^2 s parametrom ddof=1: 62.33187
odhad sigma^2 bez ddof: 61.41522
```

Pokračujeme v teste:

```
t_stat = (priemer - mu_0)*np.sqrt(n)/np.sqrt(odhad_sigma2)
print(f"t-statistika: {t_stat:.5f}")

kriticka_hodnota_zaporna, kriticka_hodnota_kladna = t.ppf(0.025, df=n-1), t.ppf(0.975, df=n-1)
print(f"kritické hodnoty: {kriticka_hodnota_zaporna:.5f}, {kriticka_hodnota_kladna:.5f}")
```

```
t-statistika: 3.40991
kritické hodnoty: -1.99601, 1.99601
```

Nulovú hypotézu zamietame, ak je

- štatistika menšia ako záporná kritická hodnota
- alebo väčšia ako kladná kritická hodnota

V tomto prípade nastáva druhý prípad, a teda hypotézu zamietame.

Výpočet p-hodnoty:

```
p_hodnota = t.cdf(-t_stat, df=n-1) + (1 - t.cdf(t_stat, df=n-1))
print(f"p-hodnota: {p_hodnota:.5f}")
```

```
p-hodnota: 0.00110
```

```
p_hodnota_univerzalne = 2*(1 - t.cdf(abs(t_stat), df=n-1))
p_hodnota_univerzalne_inak = 2*t.sf(abs(t_stat), df=n-1) # sf = survival function = 1 - cdf

print(f"p-hodnota (2. vypocet): {p_hodnota_univerzalne:.5f}")
print(f"p-hodnota (3. vypocet): {p_hodnota_univerzalne_inak:.5f}")
```

```
p-hodnota (2. vypocet): 0.00110
p-hodnota (3. vypocet): 0.00110
```

Postup 2: zabudovaná funkcia v Pythone

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_1samp.html

```
from scipy.stats import ttest_1samp

mu_0 = 10.0
statistika, p_hodnota = ttest_1samp(body, popmean=mu_0, alternative='two-sided')
print(f"SciPy ttest - štatistika: {statistika:.5f}")
print(f"SciPy ttest - p-hodnota: {p_hodnota:.5f}")
```

```
SciPy ttest - štatistika: 3.40991
SciPy ttest - p-hodnota: 0.00110
```

- **t-test** – názov pochádza z toho, že štatistika má pri platnosti nulovej hypotézy Studentovo rozdelenie, t-testov je viac, ako budeme o chvíľu vidieť.
- Toto je **jednovýberový t-test (one sample t-test)**, pretože máme len jeden výber (v našom prípade body z písomky) a testujeme hypotézu o strednej hodnote rozdelenia, z ktorého pochádza
- Parameter **popmean** je skratka pre “population mean”, teda pre strednú hodnotu rozdelenia. Píšeme sem hodnotu, ktorú chceme testovať.
- **Obojstranná alternatíva (two-sided alternative)** znamená, že alternatívna hypotéza má tvar “stredná hodnota sa nerovná danej konštante”, a nie napríklad “je od nej menšia”.

Druhý tip pred písomkou bol, že priemer bude 20. Otestujme, či je stedná hodnota rovná 20:

```
mu_0 = 20.0
statistika, p_hodnota = ttest_1samp(body, popmean=mu_0, alternative='two-sided')
print(f"SciPy ttest - štatistika: {statistika:.5f}")
print(f"SciPy ttest - p-hodnota: {p_hodnota:.5f}")
```

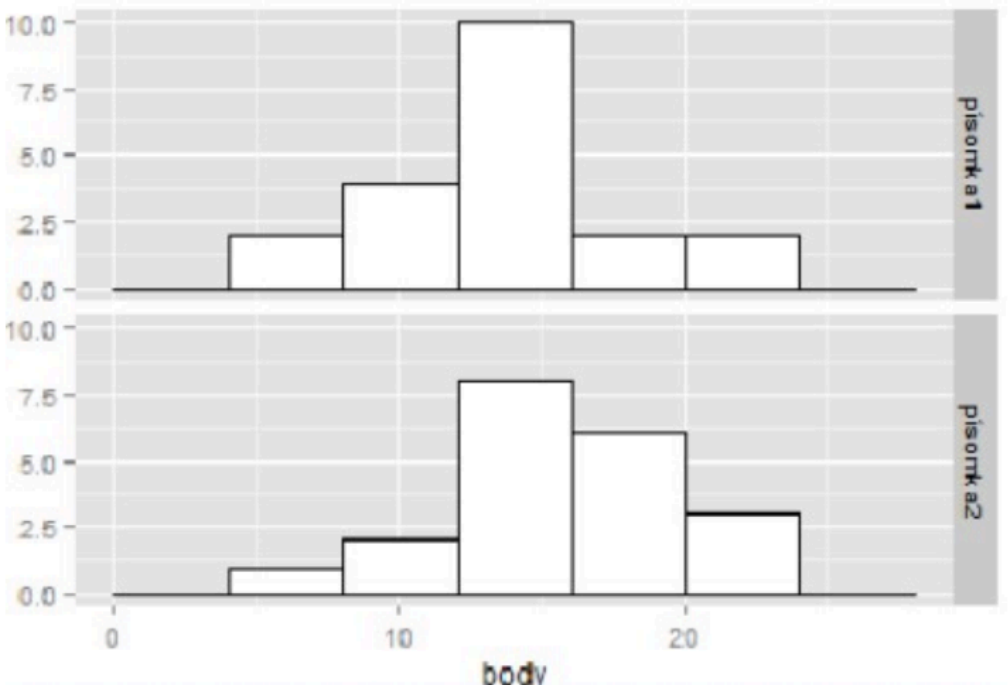
```
SciPy ttest - štatistika: -7.03487
SciPy ttest - p-hodnota: 0.00000
```

Túto hypotézu tiež zamietame.

2. Porovnanie dvoch písomiek v tej istej skupine (normálne rozdelené rozdiely – párový t-test)

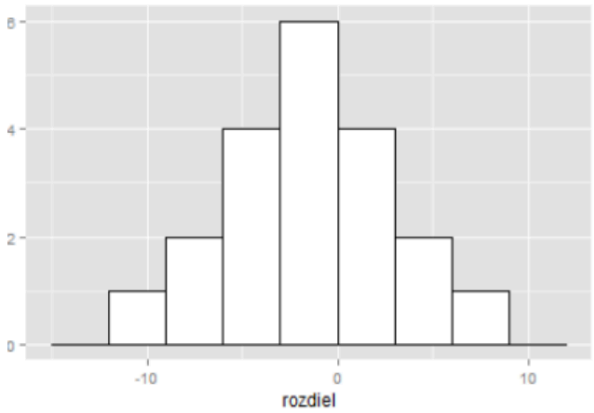
Príklad:

- Cvičenia z matematickej analýzy pre 2. ročník EFM, ZS 2005/2006, dve písomky počas semestra
- Počet bodov z písomky je náhodná premenná
- Chceme testovať hypotézu, že stredná hodnota je rovnaká pre obidve písomky



Pre každého študenta vypočítame rozdiel medzi bodmi z jednotlivých písomiek:

Poradové číslo	Písomka 1	Písomka 2	Rozdiel
1	14,5	12	2,5
2	7	18,5	-11,5
3	13,5	13	0,5
4	12,5	17	-4,5
5	10	13	-3
6	12	21	-9
7	12,5	12	0,5
8	13	12,5	0,5
9	17	18	-1
10	21,5	22	-0,5
11	11	7	4
12	12,5	16	-3,5
13	10,5	13	-2,5
14	16	9,5	6,5
15	13	15	-2
16	12,5	17	-4,5
17	10	16	-6
18	21,5	22	-0,5
19	5	12,5	-7,5
20	14,5	11,5	3
priemer	13	14,93	-1,93
štandardná odchýlka			4.47



Čo v tejto chvíli máme:

- Máme náhodný výber z dvojrozmerného rozdelenia (X, Y)
- Testujeme hypotézu, že E(X) = E(Y)
- Vypočítame rozdiely Z = X – Y
- Hypotéza E(X) = E(Y) potom znamená, že E(Z) = 0
- Ak majú rozdiely Z normálne rozdelenie (toto vieme testovať), hypotézu o nulovej strednej hodnote už vieme testovať jednovýberovým t-testom

Podmienka pre použitie tohto postupu je, že máme **párové dáta**, aby malo zmysel počítať rozdiely X – Y.

Párový t test – príklady

- Očakávaný a skutočný počet bodov z písomky – vie študent dobre odhadnúť, ako písomku napísal?
- Výsledky jazykového testu na začiatku a na konci opakovania lekcie - bolo opakovanie užitočné?
    - Rozumná alternatívna hypotéza je, že stredná hodnota na konci je väčšia
    - Kritická hodnota a p-hodnota pri takejto alternatíve je analogická ako jednostranná alternatíva pri Z-teste z testovania hypotézy o parametre **p** binomického rozdlenia
- Počet bodov z dvoch písomiek – dá sa povedať, že písomky dopadli rovnako?

Na čo sa nedá použiť párový t test – príklady:

- Počet bodov z písomky v dvoch krúžkoch
- Výsledky jazykového testu v dvoch skupinách, ktoré používali rôzne metódy pri opakovaní

Dôvod: nie sú to párové dáta

V Pythone:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html

3. Porovnanie písomiek v dvoch skupinách (normálne rozdelenie, nepárový t-test)

- Máme nezávislé realizácie dvoch náhodných premenných premenných s normálnym rozdelením a chceme testovať hypotézu, že majú rovnakú strednú hodnotu.
- Základná myšlenka je jednoduchá – vypočítajú sa priemery a zaujíma nás, či je ich rozdiel signifikantne rôzny od nuly.
- Vec sa komplikuje tým, že potrebujeme odhadnúť disperziu. Nebudeme uvádzať vzorce pre jej odhadovanie v jednotlivých prípadoch, na konkrétnych príkladoch sa na cvičení naučíme používať pythonovskú funkciu a interpretovať výsledky.

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

equal_var : *bool, optional*

If True (default), perform a standard independent 2 sample test that assumes equal population variances [1]. If False, perform Welch's t-test, which does not assume equal population variance [2].

4. Porovnanie písomiek vo viacerých skupinách (normálne rozdelenie, nezávislé skupiny dát − ANOVA)

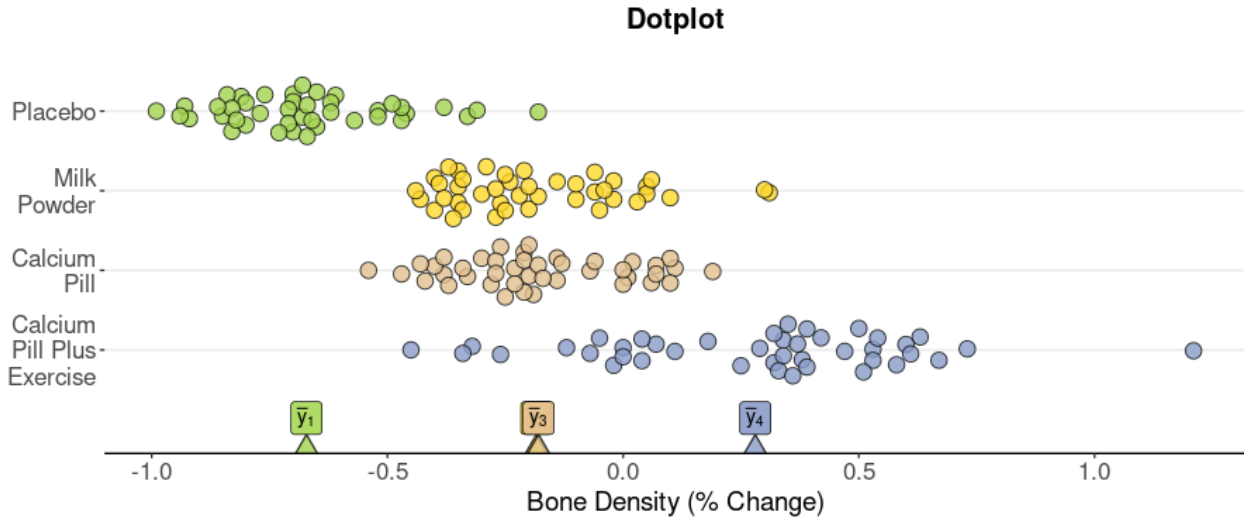
- Základná myšlienka:
  - Najskôr otestujeme hypotézu, či sú všetky stredné hodnoty rovnaké – na to slúži ANOVA (analysis of variance).
  - Ak sa táto hypotéza zamietne, tak pomocou t-testov zistíme, ktoré dvojice majú rôznu strednú hodotu.
- Tiež budeme robiť na cvičenia v Pythone.

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html

Príklad: https://istats.shinyapps.io/ANOVA/ (vizualizácie k učebnici *Agresti, Franklin, Klingenberg: Statistics: The Art and Science of Learning from Data*), dáta o osteoporóze

Descriptive Statistics:

Group	Sample Size	Mean	Standard Deviation	Standard Error
Placebo	42	-0.672	0.181	0.0280
Milk Powder	42	-0.184	0.189	0.0292
Calcium Pill	42	-0.179	0.181	0.0279
Calcium Pill Plus Exercise	42	0.280	0.330	0.0509



Z výstupu nás bude zaujímať len p-hodnota, aby sme vedeli povedať, či hypotézu o zhode všetkých stredných hodnôt zamietnuť alebo nie:

ANOVA Table:

Source	df	Sum of Squares	Mean Square	F Statistic	P-value
Group	3	19.04	6.346	120.7	<0.0001
Error	164	8.624	0.05259		
Total	167	27.664			

Hypotéza sa zamieta.

Na cvičení sa spraví aj porovnanie dvojíc pomocou t-testov, aby sme zistili, ktoré dvojice majú rôzne stredné hodnoty.

5. Čo ak dáta nemajú normálne rozdelenie

V takom prípade netestujeme strednú hodnotu, ale **medián.** Pri mediáne nepotrebujeme konkrétne pravdepodobnostné rozdelenie dát, dôležité je iba ich poradie.

Normálne rozdelenie	*Analógia pre nie normálne rozdelenie*
t-test pre strednú hodnotu, párový t-test	**Wilcoxonov test** Python: scipy.stats.wilcoxon **znamienkový test** Python: statsmodels.stats.descriptivestats.sign_test
Nepárový t-test	**Mann-Whitneyho test** Python: scipy.stats.mannwhitneyu
ANOVA	**Kruskal-Wallisov test** Python: scipy.stats.kruskal

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html

https://www.statsmodels.org/stable/generated/statsmodels.stats.descriptivestats.sign_test.html