

Rozpoznávanie obrazcov

šk.r. 2015-16

Štatistika 2

Zuzana Berger Haladova

Náhodná premenná

- Funkcia, ktorej hodnota je určená výsledkom náhodného pokusu
- Priraduje číselnú hodnotu každému elementárnemu javu
- Distribučná funkcia - opisuje rozdelenie pravdepodobnosti náhodnej premennej definovanej na pravdepodobnostnom priestore

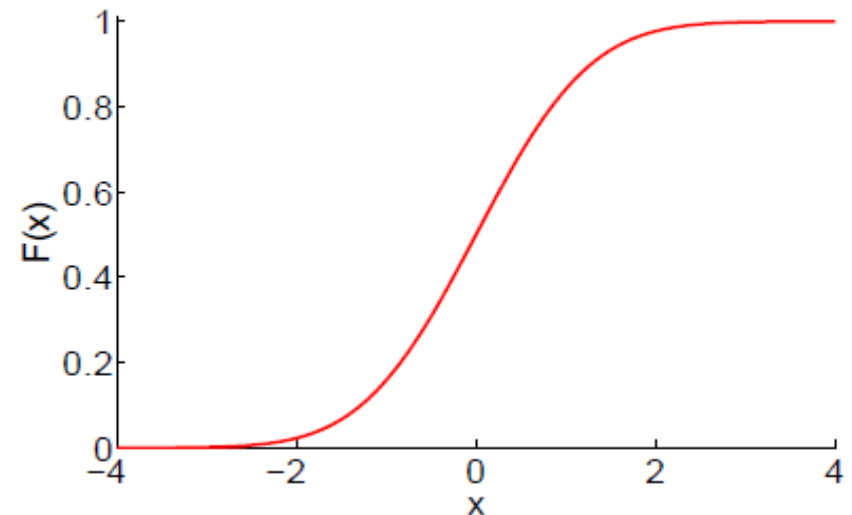
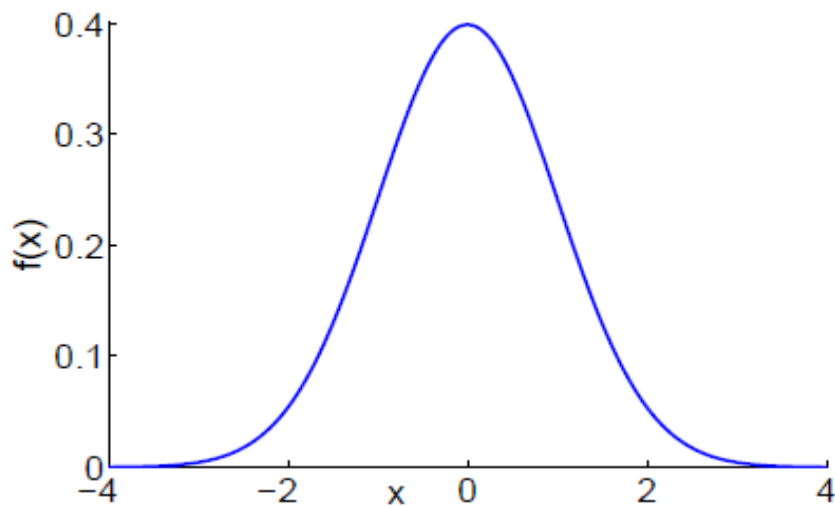
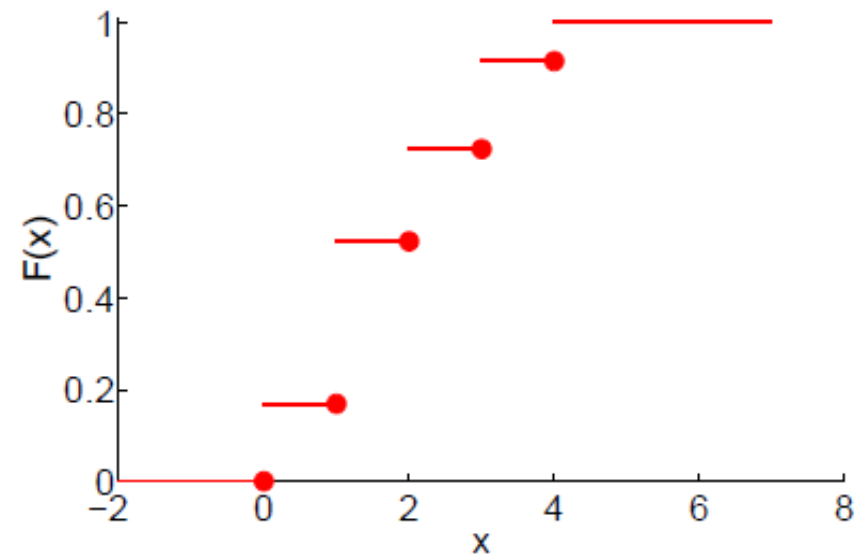
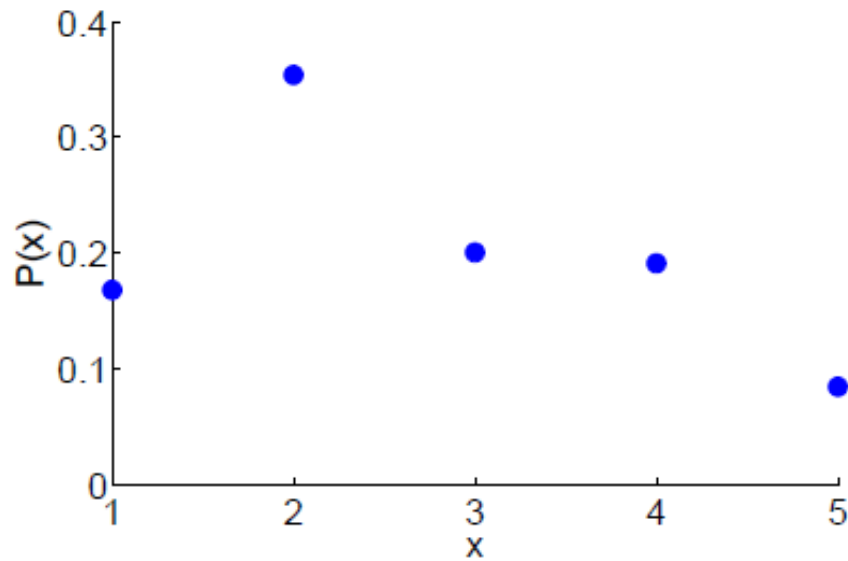
Príklad 11

- Máme 4 stroje (nezávislé na sebe)
 - 1. sa pokazí s pravdepodobnosťou 10% $P(S1)$
 - 2. s pravdepodobnosťou 15% $P(S2)$
 - 3. s pravdepodobnosťou 30% $P(S3)$
 - 4. s pravdepodobnosťou 50% $P(S4)$
- Náhodná premenná X označuje počet pokazených strojov.
- Popíšte túto náhodnú premennú

Príklad 11 pokračovanie

- $P(X=0) = P(S1 \text{ C} \cap S2 \text{ C} \cap S3 \text{ C} \cap S4 \text{ C}) =$
 $= 0.9 * 0.85 * 0.7 * 0.5 = 0.26775$
- $P(X = 1) = P(S1 \cap S2 \text{ C} \cap S3 \text{ C} \cap S4 \text{ C}) \cup$
 $P(S1 \text{ C} \cap S2 \cap S3 \text{ C} \cap S4 \text{ C}) \cup$
 $P(S1 \text{ C} \cap S2 \text{ C} \cap S3 \cap S4 \text{ C}) \cup$
 $P(S1 \text{ C} \cap S2 \text{ C} \cap S3 \text{ C} \cap S4) = 0.4595$
- $P(X=2) = 0.23$
- $P(X=3) = 0.0405$
- $P(X=4) = 0.00225$

Vztáhy mezi P a DF IV



Bernoulliho schéma

- Uvažujeme postupnosť nezávislých pokusov, ktorých celkový počet je n
- Pravdepodobnosť úspechu v každom z nich je p
- Potom

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Príklad 12

- Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aké sú pravdepodobnosti, že študent, ktorý látku vôbec nepozná a volí odpovede náhodne, zodpovie správne a) aspoň 5 otázok b) najviac 5 otázok

Príklad 12 – riešenie a)

- $n = 10, p = 0,25$
- $P(A) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$
- $P(A) = \binom{10}{5} 0,25^5 \cdot 0,75^5 + \binom{10}{6} 0,25^6 \cdot 0,75^4 + \dots + \binom{10}{10} 0,25^{10} \cdot 0,75^0$
=

Príklad 12 – riešenie b)

- $n = 10, p = 0,25$
- $P(A) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$
- $P(A) = \binom{10}{0} 0,25^0 \cdot 0,75^{10} + \binom{10}{1} 0,25^1 \cdot 0,75^9 + \dots + \binom{10}{5} 0,25^5 \cdot 0,75^5 =$
- $P(A) =$

Príklad 14

- Asi 75% zahraničným turistom chutia naše bryndzové halušky. Aká je pravdepodobnosť, že z 20 zahraničných hostí si
- a) aspoň 17 na haluškách pochutí,
- b) že halušky budú chutiť všetkým hosťom?

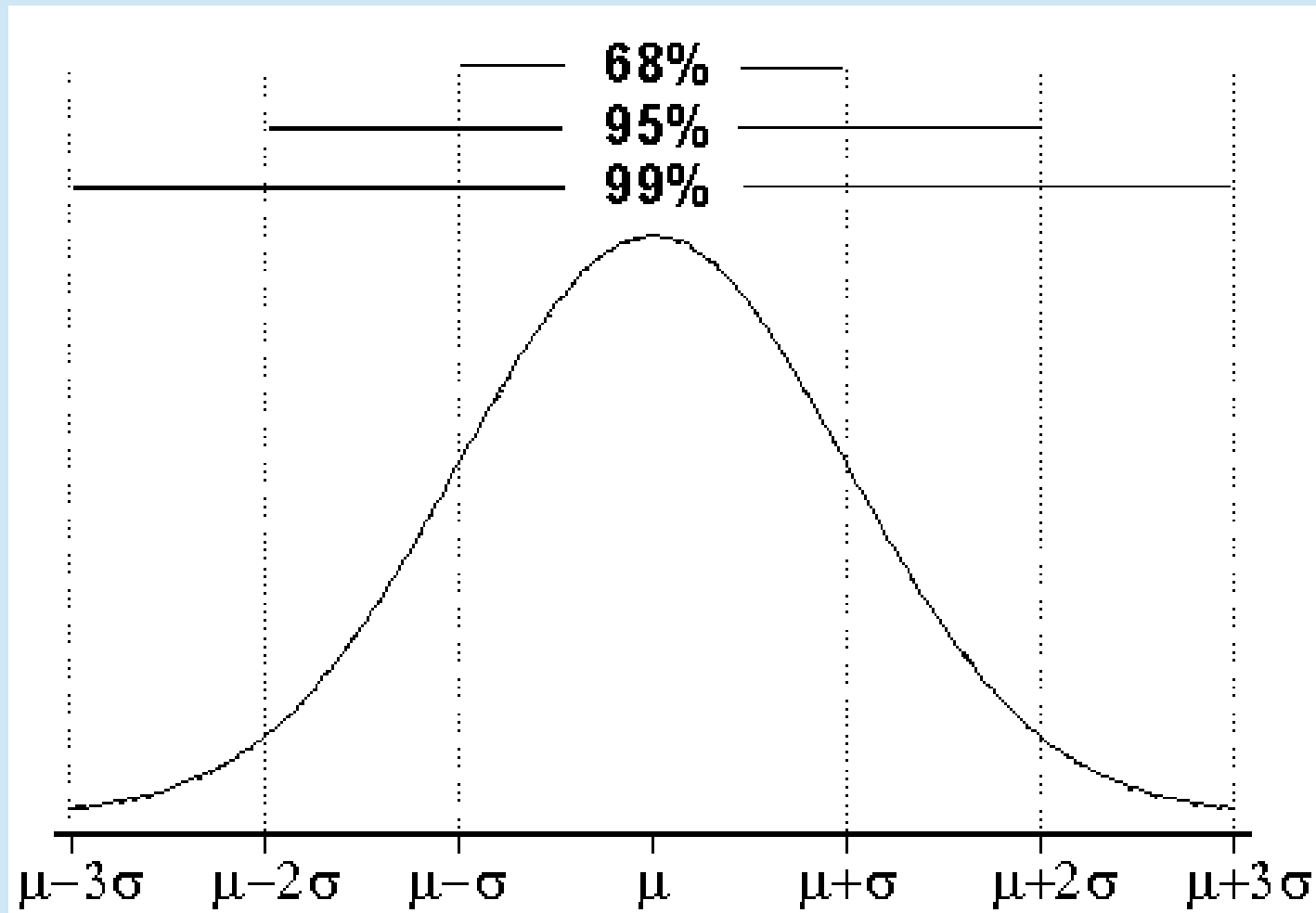
Príklad 14 – riešenie a)

- $n = 20, p = 0,75$
- $P(A) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)$
- $$P(A) = \binom{20}{17} 0,75^{17} \cdot 0,25^3 + \binom{20}{18} 0,75^{18} \cdot 0,25^2 + \dots + \binom{20}{20} 0,75^{20} \cdot 0,25^0$$
$$=$$
- $P(A) =$

Príklad 14 – riešenie b)

- $n = 20, p = 0,75$
- $P(A) = P(X = 20)$
- $P(A) = \binom{20}{20} 0,75^{20} \cdot 0,25^0 =$
- $P(A) =$

Pravidlo 3σ



Odhad parametrov rozdelení

Výberový priemer $\bar{X}$: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Výberový rozptyl S^2 : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

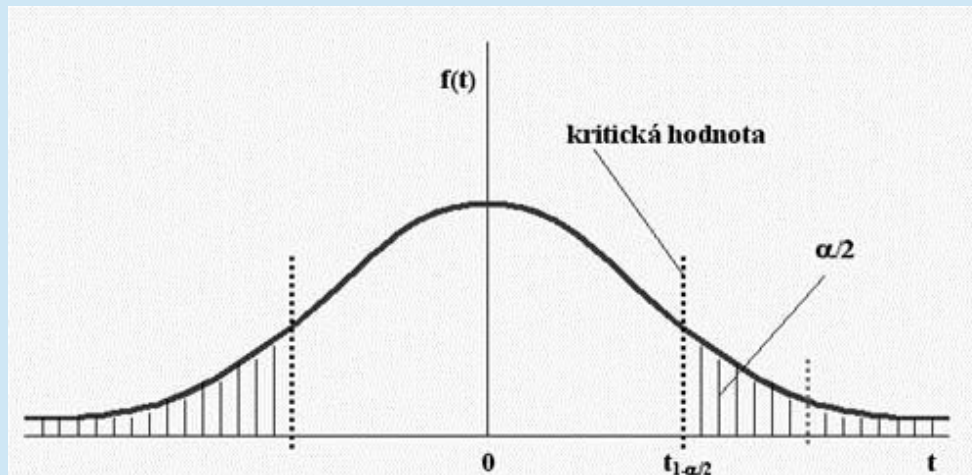
Výberová smerodajná odchýlka S : $S = \sqrt{S^2}$

Výberová kovariancia S_{XY} : $S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n(n-1)}$

Intervalový odhad spoľahlivosti

$$P(G_D < \theta < G_H) = 1 - \alpha$$

- Interval spoľahlivosti
- Koeficient spoľahlivosti
- Skutočná hodnota parametra



Kritické hodnoty rozdelenia $N(0,1)$

$$X \sim N(0,1), P(|X| > u_\alpha) = \alpha$$

α	0,01	0,02	0,05	0,1	0,2
u_α	2,5758	2,3263	1,9599	1,6448	1,299

$$\begin{aligned} 1 - \alpha &= P(-u_\alpha \leq U \leq u_\alpha) = P(-u_\alpha \leq \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \leq u_\alpha) = \\ &= P(\bar{X} - u_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_\alpha \cdot \frac{\sigma}{\sqrt{n}}). \end{aligned}$$

Kritické hodnoty rozdelenia $N(0,1)$

• Pre priemer

- 1. Ak poznáme σ
- 2. Ak nepoznáme σ a $n > 30$
- 3. Ak nepoznáme σ a $n < 30$

$$\begin{array}{ccc} u = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} & u = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}} & t = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}} \\ 1 & 2 & 3 \end{array}$$

Príklad 17

- Predpokladáme, že výška chlapcov vo veku 9-10 rokov má normálne rozdelenie s neznámou strednou hodnotou a odchýlkou $\sigma^2 = 39,112$. Zmerali sme výšku 15 chlapcov a určili výberovú strednú hodnotu 139,13. Určte 99% obojstranný interval spoľahlivosti pre strednú hodnotu.

Príklad 17 - riešenie

- $\sigma = 6,253$
- $n = 15$
- $\bar{X} = 139,13$
- $1 - \alpha = P\left(\bar{X} - \mu_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \mu_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$
- $139,13 - 2,5758 \cdot \frac{6,253}{\sqrt{15}} \leq \mu \leq 139,13 + 2,5758 \cdot \frac{6,253}{\sqrt{15}}$
- $134,97 \leq \mu \leq 143,28$

Príklad 18

- Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25. Nájdí 95%-ný obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ .

Príklad 18 - riešenie

- $\bar{X} = 112$
- $S = 5$
- $n = 20$
- $1 - \alpha = P\left(\bar{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}\right)$
- $112 - 2,093 \cdot \frac{5}{\sqrt{20}} \leq \mu \leq 112 + 2,093 \cdot \frac{5}{\sqrt{20}}$
- $109,65 \leq \mu \leq 114,34$

Príklad 19

- Náhodná premenná X má normálne rozdelenie, kde stredná hodnota aj odchýlka sú neznáme.
- Namerané realizácie X sú 27, 15, -3, -6, 12, 20, 13, 0, 7, 10.
- Nájdite 95% obojstranný interval spoľahlivosti pre strednú hodnotu.

Príklad 19 - riešenie

- $n = 10$
- $\bar{X} = 9,5$
- $S = 10,319$
- $1 - \alpha = P\left(\bar{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}\right)$
- $9,5 - 2,262 \cdot \frac{10,319}{\sqrt{10}} \leq \mu \leq 9,5 + 2,262 \cdot \frac{10,319}{\sqrt{10}}$
- $2,118 \leq \mu \leq 16,881$

Príklad 20

- Zo základného súboru s normálnym rozdelením, kde je známy rozptyl $\sigma^2 = 0,06$ sme urobili náhodný výber s prvkami 1,3 1,8 1,4 1,2 0,9 1,5 1,7
- Zistite 95%-ný obojstranný interval spoľahlivosti pre strednú hodnotu μ

Príklad 20

- $n = 7$
- $\bar{X} = 1,4$
- $\sigma = 0,245$
- $1 - \alpha = P\left(\bar{X} - \mu_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \mu_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$
- $1,4 - 1,9599 \cdot \frac{0,245}{\sqrt{7}} \leq \mu \leq 1,4 + 1,9599 \cdot \frac{0,245}{\sqrt{7}}$
- $1,218 \leq \mu \leq 1,581$

MATLAB vsuvka

- V MATLAB-e môžeme vypočítať kritické hodnoty pre rôzne hladiny významnosti pre normálne rozdelenie pomocou funkcie $\text{norminv}(\alpha)$
- Podobne pre Studentovo rozdelenie pomocou funkcie $\text{tinv}(\alpha, n)$, kde n je počet stupňov voľnosti rozdelenia a rovná sa počet prvkov výberu mínus jedna

MATLAB vsuvka II

- Iný spôsob riešenia tejto úlohy je použitie funkcie **`[h,p,ci] = ztest(x,m,sigma, aplha, tail)`**, kde **`ci`** je confidence interval (interval spoľahlivosti), **`x`** je výberový súbor, **`m`** je jeho stredná hodnota, **`sigma`** je odchýlka, **`alpha`** ja hladina významnosti (ak ju ne zadáme, tak je 0,05) a **`tail`** hovorí o tom, či interval je obojstranný alebo nie

MATLAB vsuvka III

- V prípade, že nepoznáme smerodajnú odchýlku použijeme funkciu **[h,p,ci] = ttest(x,m, aplha, tail)**, kde všetky parametre majú podobný význam ako v predošlom prípade a MATLAB dopočíta výberovú smerodajnú odchýlku a použije Studentovo rozdelenie na určenie kritických hodnôt pre interval spoľahlivosti

Príklad 21

- Tvrdíme, že guľôčky, vyrábané automatickým sústruhom majú strednú hodnotu polomeru 10 mm.
- Použitím obojstranného testu významnosti na hladine 0,05 otestujte túto hypotézu, ak vo výbere zo 16 guľôčok je priemer ich polomerov 10,3 mm a
 - a) disperzia je známa $\sigma^2 = 1$
 - b) odhad disperzie $S^2 = 1,21$

Príklad 21 – riešenie a)

- $n = 16$
- $\bar{X} = 10,3$ Hypotéza H_0 je, že $\mu = 10$
- $\sigma = 1$ Hypotéza H_1 je, že $\mu \neq 10$
- $1 - \alpha = P\left(\bar{X} - \mu_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \mu_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$
- $10,3 - 1,9599 \cdot \frac{1}{\sqrt{16}} \leq \mu \leq 10,3 + 1,9599 \cdot \frac{1}{\sqrt{16}}$
- $9,81 \leq \mu \leq 10,789$
- Hypotézu H_0 nezamietneme

Príklad 21 – riešenie b)

- $n = 16$
- $\bar{X} = 10,3$ Hypotéza H_0 je, že $\mu = 10$
- $S = 1,1$ Hypotéza H_1 je, že $\mu \neq 10$
- $1 - \alpha = P\left(\bar{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}\right)$
- $10,3 - 2,131 \cdot \frac{1,1}{\sqrt{16}} \leq \mu \leq 10,3 + 2,131 \cdot \frac{1,1}{\sqrt{16}}$
- $9,71 \leq \mu \leq 10,88$
- Hypotézu H_0 nezamietneme