

## 6. Parametrické úlohy o jednom náhodném výběru z normálního rozložení

**6.1. Motivace:** Mnoho náhodných veličin, s nimiž se setkáváme ve výzkumu i praxi, se řídí normálním rozložením. Za jistých předpokladů obsažených v centrální limitní větě se dá rozložení jiných náhodných veličin aproximovat normálním rozložením. Proto je zapotřebí věnovat velkou pozornost právě náhodným výběrům z normálního rozložení.

**6.2. Věta:** Rozložení statistik odvozených z výběrového průměru a výběrového rozptylu

Nechť  $X_1, \dots, X_n$  je náhodný výběr z rozložení  $N(\mu, \sigma^2)$ . Pak platí

a) Výběrový průměr  $M$  a výběrový rozptyl  $S^2$  jsou stochasticky nezávislé.

b)  $M \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ , tedy  $U = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$ .

(Pivotová statistika  $U$  slouží k řešení úloh o  $\mu$ , když  $\sigma^2$  známe.)

c)  $K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ .

(Pivotová statistika  $K$  slouží k řešení úloh o  $\sigma^2$ , když  $\mu$  neznáme.)

d)  $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$ .

(Tato pivotová statistika slouží k řešení úloh o  $\sigma^2$ , když  $\mu$  známe.)

e)  $T = \frac{M - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$ .

(Pivotová statistika  $T$  slouží k řešení úloh o  $\mu$ , když  $\sigma^2$  neznáme.)

### Důkaz:

ad a) Nebudeme provádět, viz Jiří Anděl: Matematická statistika, SNTL/Alfa, Praha 1978, str. 82

ad b) Výběrový průměr  $M$  je lineární kombinace náhodných veličin s normálním rozložením, má tedy normální rozložení s parametry  $E(M) = \mu$ ,  $D(M) = \sigma^2/n$ . Statistika  $U$  se získá standardizací  $M$ .

ad c) Vhodnou úpravou výběrového rozptylu  $S^2$ , kde použijeme obrat  $X_i - M = (X_i - \mu) - (M - \mu)$ , lze statistiku  $K$  vyjádřit jako součet kvadrátů  $n - 1$  stochasticky nezávislých náhodných veličin se standardizovaným normálním rozložením. Tento součet se řídí rozložením  $\chi^2(n-1)$ .

ad d) Tato statistika je součet kvadrátů  $n$  stochasticky nezávislých náhodných veličin se standardizovaným normálním rozložením, řídí se tedy rozložením  $\chi^2(n)$ .

ad e)  $U \sim N(0, 1)$ ,  $K \sim \chi^2(n-1)$  jsou stochasticky nezávislé, protože  $M$  a  $S^2$  jsou stochasticky nezávislé, tudíž statistika  $T = \frac{U}{\sqrt{\frac{K}{n-1}}} = \frac{M - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$ .

**6.3. Příklad:** Hmotnost balíčku krystalového cukru baleného na automatické lince se řídí normálním rozložením se střední hodnotou 1002 g a směrodatnou odchylkou 8 g. Kontrolor náhodně vybírá 9 balíčků z jedné série a zjišťuje, zda jejich průměrná hmotnost je alespoň 999 g. Pokud ne, podnik musí zaplatit pokutu 20 000 Kč. Jaká je pravděpodobnost, že podnik bude muset zaplatit pokutu?

Řešení:

$$X \sim N(1002, 64), M \sim N\left(1002, \frac{64}{9}\right)$$

$$\begin{aligned} P\{M \leq 999\} &= P\left\{\frac{M - 1002}{\sqrt{\frac{64}{9}}} \leq \frac{999 - 1002}{\sqrt{\frac{64}{9}}}\right\} = P\left\{U \leq -\frac{3}{8}\right\} = \Phi\left(-\frac{3}{8}\right) = 1 - \Phi\left(\frac{3}{8}\right) = 1 - \Phi(0,375) \\ &= 1 - 0,6493 = 0,3507 = 35,07\% \end{aligned}$$

Pravděpodobnost, že podnik bude platit pokutu, je asi 35,07%.

### Řešení pomocí systému STATISTICA:

Využijeme toho, že STATISTICA pomocí funkce INormal(x;mu;sigma) umí vypočítat hodnotu distribuční funkce normálního rozložení se střední hodnotou  $\mu$  a směrodatnou odchylkou  $\sigma$ . Tedy  $P\{M \leq 999\} = \Phi(999)$ , kde  $\Phi$  je distribuční funkce rozložení  $N(1002, 64/9)$ .

Otevřeme nový datový soubor o jedné proměnné a jednom případě. Dvakrát klikneme na název proměnné Prom1. Do Dlouhého jména této proměnné napíšeme = INormal(999;1002;8/3).

V proměnné Prom1 se objeví hodnota 0,130295.

**6.4. Věta:** Vzorce pro meze  $100(1-\alpha)\%$  empirických intervalů spolehlivosti pro  $\mu$  a  $\sigma^2$

a) Interval spolehlivosti pro  $\mu$ , když  $\sigma^2$  známe (využití pivotové statistiky  $U$ )

$$\text{Oboustranný: } (d, h) = \left(m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right)$$

$$\text{Levostranný: } (d, \infty) = \left( m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \infty \right)$$

$$\text{Pravostranný: } (-\infty, h) = \left( -\infty, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \right)$$

b) Interval spolehlivosti pro  $\mu$ , když  $\sigma^2$  neznáme (využití pivotové statistiky T)

$$\text{Oboustranný: } (d, h) = \left( m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right)$$

$$\text{Levostranný: } (d, \infty) = \left( m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1), \infty \right)$$

$$\text{Pravostranný: } (-\infty, h) = \left( -\infty, m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) \right)$$

c) Interval spolehlivosti pro  $\sigma^2$ , když  $\mu$  neznáme (využití pivotové statistiky K)

$$\text{Oboustranný: } (d, h) = \left( \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \right)$$

$$\text{Levostranný: } (d, \infty) = \left( \frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)}, \infty \right)$$

$$\text{Pravostranný: } (-\infty, h) = \left( -\infty, \frac{(n-1)s^2}{\chi^2_{\alpha}(n-1)} \right)$$

d) Interval spolehlivosti pro  $\sigma^2$ , když  $\mu$  známe (využití pivotové statistiky

$$\left( \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma} \right)$$

$$\text{Oboustranný: } (d, h) = \left( \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{1-\alpha/2}(n)}, \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{\alpha/2}(n)} \right)$$

$$\text{Levostranný: } (d, \infty) = \left( \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{1-\alpha}(n)}, \infty \right)$$

$$\text{Pravostranný: } (-\infty, h) = \left( -\infty, \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{\alpha}(n)} \right)$$

**6.5. Příklad:** 10 krát nezávisle na sobě byla změřena jistá konstanta  $\mu$ . Výsledky měření byly: 2 1,8 2,1 2,4 1,9 2,1 2 1,8 2,3 2,2. Tyto výsledky považujeme

za číselné realizace náhodného výběru  $X_1, \dots, X_{10}$  z rozložení  $N(\mu, \sigma^2)$ , kde parametry  $\mu, \sigma^2$  neznáme. Najděte 95% empirický interval spolehlivosti jak pro  $\mu$ , tak pro  $\sigma^2$  a to

- oboustranný,
- levostranný,
- pravostranný.

**Řešení:**  $m = 2,06$ ,  $s^2 = 0,0404$ ,  $s = 0,2011$ ,  $\alpha = 0,05$ ,  $t_{0,975}(9) = 2,2622$ ,  $t_{0,95}(9) = 1,8331$ ,  $\chi^2_{0,975}(9) = 19,023$ ,  $\chi^2_{0,025}(9) = 2,7$ ,  $\chi^2_{0,95}(9) = 16,919$ ,  $\chi^2_{0,05}(9) = 3,325$

ad a) **Oboustranný interval spolehlivosti pro střední hodnotu  $\mu$**

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} 2,2622 = 1,92$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} 2,2622 = 2,20$$

$1,92 < \mu < 2,20$  s pravděpodobností aspoň 0,95.

**Oboustranný interval spolehlivosti pro rozptyl  $\sigma^2$**

$$d = \frac{(n-1) \cdot s^2}{\chi^2_{1-\alpha/2}(n-1)} = \frac{9 \cdot 0,0404}{19,023} = 0,0191$$

$$h = \frac{(n-1) \cdot s^2}{\chi^2_{\alpha/2}(n-1)} = \frac{9 \cdot 0,0404}{2,7} = 0,1347$$

$0,0191 < \sigma^2 < 0,1347$  s pravděpodobností aspoň 0,95.

ad b) **Levostranný interval spolehlivosti pro střední hodnotu  $\mu$**

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} 1,8331 = 1,94$$

$1,94 < \mu$  s pravděpodobností aspoň 0,95.

**Levostranný interval spolehlivosti pro rozptyl  $\sigma^2$**

$$d = \frac{(n-1) \cdot s^2}{\chi^2_{1-\alpha}(n-1)} = \frac{9 \cdot 0,0404}{16,919} = 0,0215$$

$\sigma^2 > 0,0215$  s pravděpodobností aspoň 0,95.

ad c) **Pravostranný interval spolehlivosti pro střední hodnotu  $\mu$**

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} 1,8331 = 2,18$$

$\mu < 2,18$  s pravděpodobností aspoň 0,95.

**Pravostranný interval spolehlivosti pro rozptyl  $\sigma^2$**

$$h = \frac{9 \cdot 0,0404}{3,325} = 0,1094$$

$\sigma^2 < 0,1094$  s pravděpodobností aspoň 0,95.

### Řešení pomocí systému STATISTICA:

Vytvoříme nový datový soubor o jedné proměnné X a 10 případech. Do proměnné X napíšeme dané hodnoty.

Statistika – Základní statistiky a tabulky – Popisné statistiky – OK – Proměnné X – OK – Detailní výsledky – zaškrtneme Meze spolehl. prům. a Meze sp. směr. odch. (ostatní volby zrušíme) – pro oboustranný 95% interval spolehlivosti ponecháme implicitní hodnotu pro Interval 95,00, pro jednostranné intervaly změníme hodnotu na 90,00.

Výsledky pro oboustranné 95% intervaly spolehlivosti pro střední hodnotu  $\mu$ , pro směrodatnou odchylku  $\sigma$  a rozptyl  $\sigma^2$ :

Proměnná	Int. spolehl.	Int. spolehl.	Spolehlivost	Spolehlivost	NProm1	NProm2
	-95,000%	95,000	Sm.Odch.	Sm.Odch.	= $\sqrt{3}$	= $\sqrt{4}$
X	1,916136	2,203864	-95,000%	+95,000%	0,019135	0,134795

Vidíme, že

$1,92 < \mu < 2,20$  s pravděpodobností aspoň 0,95,

$0,1383 < \sigma < 0,3671$  s pravděpodobností aspoň 0,95.

$0,0191 < \sigma^2 < 0,1348$  s pravděpodobností aspoň 0,95.

Výsledky pro jednostranné 95% intervaly spolehlivosti pro střední hodnotu  $\mu$ , pro směrodatnou odchylku  $\sigma$  a rozptyl  $\sigma^2$ :

Proměnná	Int. spolehl.	Int. spolehl.	Spolehlivost	Spolehlivost	NProm1	NProm2
	-90,000%	90,000	Sm.Odch.	Sm.Odch.	= $\sqrt{3}$	= $\sqrt{4}$
X	1,943421	2,176579	-90,000%	+90,000%	0,021514	0,10947

Vidíme, že

$\mu > 1,94$  s pravděpodobností aspoň 0,95,

$\mu < 2,20$  s pravděpodobností aspoň 0,95,

$\sigma > 0,1467$  s pravděpodobností aspoň 0,95,

$\sigma < 0,3309$  s pravděpodobností aspoň 0,95,

$\sigma^2 > 0,0215$  s pravděpodobností aspoň 0,95,

$\sigma^2 < 0,1095$  s pravděpodobností aspoň 0,95,

**6.6. Definice:** Jednotlivé typy testů pro parametry normálního rozložení

- a) Necht'  $X_1, \dots, X_n$  je náhodný výběr  $N(\mu, \sigma^2)$ , kde  $\sigma^2$  známe. Necht'  $n \geq 2$  a  $c$  je konstanta. Test  $H_0: \mu = c$  proti  $H_1: \mu \neq c$  se nazývá **jednovýběrový z-test**.
- b) Necht'  $X_1, \dots, X_n$  je náhodný výběr  $N(\mu, \sigma^2)$ , kde  $\sigma^2$  neznáme. Necht'  $n \geq 2$  a  $c$  je konstanta. Test  $H_0: \mu = c$  proti  $H_1: \mu \neq c$  se nazývá **jednovýběrový t-test**.
- c) Necht'  $X_1, \dots, X_n$  je náhodný výběr  $N(\mu, \sigma^2)$ , kde  $\mu$  neznáme. Necht'  $n \geq 2$  a  $c$  je konstanta. Test  $H_0: \sigma^2 = c$  proti  $H_1: \sigma^2 \neq c$  se nazývá **test o rozptylu**.

## Provedení testů o parametrech $\mu, \sigma^2$ pomocí kritického oboru

### a) Provedení jednovýběrového z-testu

Vypočteme realizaci testového kritéria  $t_0 = \frac{\bar{x} - c}{\frac{\sigma}{\sqrt{n}}}$ . Stanovíme kritický obor  $W$ .

Pokud  $t_0 \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha$  a přijímáme  $H_1$ .

**Oboustranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu \neq c$ . Kritický obor má tvar:

$$W = \left( -\infty, -u_{1-\alpha/2} \right) \cup \left( u_{1-\alpha/2}, \infty \right)$$

**Levostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu < c$ . Kritický obor má tvar:

$$W = \left( -\infty, -u_{1-\alpha} \right)$$

**Pravostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu > c$ . Kritický obor má tvar:

$$W = \left( u_{1-\alpha}, \infty \right)$$

### b) Provedení jednovýběrového t-testu

Vypočteme realizaci testového kritéria  $t_0 = \frac{\bar{x} - c}{\frac{s}{\sqrt{n}}}$ . Stanovíme kritický obor  $W$ .

Pokud  $t_0 \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha$  a přijímáme  $H_1$ .

**Oboustranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu \neq c$ . Kritický obor má tvar:

$$W = \left( -\infty, -t_{1-\alpha/2, n-1} \right) \cup \left( t_{1-\alpha/2, n-1}, \infty \right)$$

**Levostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu < c$ . Kritický obor má tvar:

$$W = \left( -\infty, -t_{1-\alpha, n-1} \right)$$

**Pravostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu > c$ . Kritický obor má tvar:

$$W = \left( t_{1-\alpha, n-1}, \infty \right)$$

### c) Provedení testu o rozptylu

Vypočteme realizaci testového kritéria  $t_0 = \frac{(n-1)s^2}{c}$ . Stanovíme kritický obor

$W$ . Pokud  $t_0 \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha$  a přijímáme  $H_1$ .

**Oboustranný test:** Testujeme  $H_0: \sigma^2 = c$  proti  $H_1: \sigma^2 \neq c$ . Kritický obor má tvar:

$$W = \left( 0, \chi^2_{\alpha/2} \right) \cup \left( \chi^2_{1-\alpha/2}, \infty \right)$$

**Levostranný test:** Testujeme  $H_0: \sigma^2 = c$  proti  $H_1: \sigma^2 < c$ . Kritický obor má tvar:

$$W = \left( 0, \chi^2_{\alpha} \right)$$

**Pravostranný test:** Testujeme  $H_0: \sigma^2 = c$  proti  $H_1: \sigma^2 > c$ . Kritický obor má tvar:

$$W = \left( \chi^2_{1-\alpha}, \infty \right)$$

**6.7. Příklad:** Podle údajů na obalu čokolády by její čistá hmotnost měla být 125 g. Výrobce dostal několik stížností od kupujících, ve kterých tvrdili, že hmotnost čokolád je nižší než deklarovaných 125 g. Z tohoto důvodu oddělení kontroly náhodně vybralo 50 čokolád a zjistilo, že jejich průměrná hmotnost je 122 g a směrodatná odchylka 8,6 g. Za předpokladu, že hmotnost čokolád se řídí normálním rozložením, můžeme na hladině významnosti 0,01 považovat stížnosti kupujících za oprávněné?

**Řešení:**  $X_1, \dots, X_{50}$  je náhodný výběr z  $N(\mu, \sigma^2)$ . Testujeme hypotézu  $H_0: \mu = 125$  proti levostranné alternativě  $H_1: \mu < 125$ . Protože neznáme rozptyl  $\sigma^2$ , použijeme jednovýběrový t-test.

$$\text{Testové kritérium } \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{122 - 125}{\frac{8,6}{\sqrt{50}}} = -2,4667$$

$$\text{Kritický obor } W = \left( -\infty, -t_{1-\alpha, n-1} \right) = \left( -\infty, -t_{0,99, 49} \right) = \left( -\infty, -2,4049 \right)$$

Jelikož testové kritérium se realizuje v kritickém oboru, zamítáme nulovou hypotézu na hladině významnosti 0,01. Stížnosti kupujících tedy lze považovat za oprávněné.

### Výpočet pomocí systému STATISTICA:

Statistiky – Základní statistiky a tabulky – Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – zaškrtneme Výběrový průměr vs. Střední hodnota a zvolíme jednostr. – do políčka Pr1 napíšeme 122, do políčka SmOd1 napíšeme 8,6, do políčka N1 napíšeme 50, do políčka Pr2 napíšeme 125 - Výpočet. Dostaneme p-hodnotu 0,0086, tedy zamítáme nulovou hypotézu na hladině významnosti 0,01

**6.8. Definice:** Definice rozdílového náhodného výběru na základě náhodného výběru z dvourozměrného normálního rozložení.

Nechť  $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$  je náhodný výběr z rozložení  $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$ , přičemž

$n \geq 2$ . Označíme  $\mu = \mu_1 - \mu_2$  a zavedeme **rozdílový náhodný výběr**  $Z_1 = X_1 - Y_1, \dots, Z_n = X_n - Y_n$ . Vypočteme  $M = \frac{1}{n} \sum_{i=1}^n Z_i, S^2 = \frac{1}{n} \sum_{i=1}^n Z_i^2 - M^2$ .

**6.9. Věta:** Vzorec pro meze  $100(1-\alpha)\%$  empirického intervalu spolehlivosti pro střední hodnotu rozdílového náhodného výběru

$$\text{Oboustranný: } (d, h) = \left(m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right)$$

$$\text{Levostranný: } (d, \infty) = \left(m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1), \infty\right)$$

$$\text{Pravostranný: } (-\infty, h) = \left(-\infty, m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1)\right)$$

**6.10. Příklad:** Dvěma rozdílnými laboratorními metodami se zjišťoval obsah chemické látky v roztoku (v procentech). Bylo vybráno 5 vzorků a proměřeno oběma metodami. Výsledky měření jsou obsaženy v tabulce:

číslo vzorku	1	2	3	4	5
1. metoda	2,3	1,9	2,1	2,4	2,6
2. metoda	2,4	2,0	2,0	2,3	2,5



Za předpokladu, že data mají normální rozložení, sestrojte 90% empirický interval spolehlivosti pro rozdíl středních hodnot výsledků obou metod.

### Řešení:

Přejdeme k rozdílovému náhodnému výběru, jehož realizace jsou: -0,1 -0,1 0,1 0,1 0,1. Vypočteme  $m = 0,02$ ,  $s^2 = 0,012$ ,  $s = 0,109545$ . Předpokládáme, že tato data pocházejí z normálního rozložení  $N(\mu, \sigma^2)$ . Vypočteme meze 90% oboustranného intervalu spolehlivosti pro  $\mu$  při neznámém  $\sigma$ :

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2, n-1} = 0,02 - \frac{0,109545}{\sqrt{5}} t_{0,95, 4} = 0,02 - \frac{0,109545}{\sqrt{5}} 2,1318 = -0,0844$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2, n-1} = 0,02 + \frac{0,109545}{\sqrt{5}} t_{0,95, 4} = 0,02 + \frac{0,109545}{\sqrt{5}} 2,1318 = 0,1244$$

$-0,0844 < \mu < 0,1244$  s pravděpodobností aspoň 0,9.

### Výpočet pomocí systému STATISTICA:

Vytvoříme nový datový soubor o 3 proměnných a 5 případech. Do 1. proměnné X napíšeme hodnoty pro 1. metodu, do 2. proměnné Y hodnoty pro 2. metodu a do 3. proměnné Z rozdíly mezi X a Y.

Statistiky – Základní statistiky a tabulky – Popisné statistiky, OK - Proměnné Z, Detailní výsledky – zaškrtneme Meze spolehl. Prům. – Interval 90% - Výpočet. Dostaneme tabulku:

Proměnná	Popisné statistiky (chemická látka)	
	Int. spolehl.	Int. spolehl.
Z	-90,000%	90,000
	-0,084439	0,124439

Vidíme tedy, že  $-0,0844 < \mu < 0,1244$  s pravděpodobností aspoň 0,9.

### 6.11. Definice: Definice párového t-testu

Nechť  $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$  je náhodný výběr z rozložení  $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$ , přičemž

$n \geq 2$ . Testujeme  $H_0: \mu_1 - \mu_2 = c$  (tj.  $\mu = c$ ) proti  $H_1: \mu_1 - \mu_2 \neq c$  (tj.  $\mu \neq c$ ) nebo testujeme nulovou hypotézu proti jedné z jednostranných alternativ. Tento test se nazývá **párový t-test**.

### Provedení párového t-testu

Vypočteme realizaci testového kritéria  $t_0 = \frac{\bar{m} - c}{\frac{s}{\sqrt{n}}}$ . Stanovíme kritický obor  $W$ .

Pokud  $t_0 \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha$  a přijímáme  $H_1$ .

**Oboustranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu \neq c$ . Kritický obor má tvar:

$$W = \left\{ t_0 < -t_{1-\alpha/2, n-1} \right\} \cup \left\{ t_0 > t_{1-\alpha/2, n-1} \right\}$$

**Levostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu < c$ . Kritický obor má tvar:

$$W = \left\{ t_{1-\alpha} \leq \dots \right\}$$

**Pravostranný test:** Testujeme  $H_0: \mu = c$  proti  $H_1: \mu > c$ . Kritický obor má tvar:

$$W = \left\{ t_{1-\alpha} \leq \dots \right\}$$

**6.12. Příklad:** V následující tabulce jsou údaje o výnosnosti dosažené 12 náhodně vybranými firmami při investování do mezinárodního podnikání (veličina X) a do domácího podnikání (veličina Y):

č.firmy	1	2	3	4	5	6	7	8	9	10	11	12
X	10	12	14	12	12	17	9	15	9	11	7	15
Y	11	14	15	11	13	16	10	13	11	17	9	19

(Výnosnost je vyjádřena v procentech a představuje podíl na zisku vložených investic za rok.)

Za předpokladu, že data pocházejí z dvourozměrného normálního rozložení, na hladině významnosti 0,1 testujte hypotézu, že neexistuje rozdíl mezi střední hodnotou výnosnosti investic do mezinárodního a domácího podnikání proti oboustranné alternativě.

Testování proveďte

- pomocí intervalu spolehlivosti
- pomocí kritického oboru.

(Pro úsporu času máte uvedeny realizace výběrového průměru  $m = -1,3$  a výběrového rozptylu  $s^2 = 4,78$  rozdílového náhodného výběru  $Z_i = X_i - Y_i$ ,  $i = 1, \dots, 12$ .)

**Řešení:**

Testujeme  $H_0: \mu = 0$  proti  $H_1: \mu \neq 0$

ad a)

90% interval spolehlivosti pro střední hodnotu  $\mu$  při neznámém rozptylu  $\sigma^2$  má meze:

$$d = m - \frac{s}{\sqrt{n}} t_{0,95} = -1,3 - \frac{\sqrt{4,78}}{\sqrt{12}} 1,7959 = -2,4677$$

$$h = m + \frac{s}{\sqrt{n}} t_{0,95} = -1,3 + \frac{\sqrt{4,78}}{\sqrt{12}} 1,7959 = -0,1989$$

Protože číslo  $c = 0$  neleží v intervalu  $(-2,4677; -0,1989)$ ,  $H_0$  zamítáme na hladině významnosti 0,1.

ad b)

$$\text{Vypočítáme realizaci testové statistiky } t_0 = \frac{m - c}{\frac{s}{\sqrt{n}}} = \frac{-1,3}{\frac{\sqrt{4,78}}{\sqrt{12}}} = -1,11085$$

Stanovíme kritický obor

$$W = \left( -\infty, -t_{0,95} \right) \cup \left( t_{0,95}, \infty \right) = \left( -\infty, -1,7959 \right) \cup \left( 1,7959, \infty \right)$$

Protože testová statistika se realizuje v kritickém oboru,  $H_0$  zamítáme na hladině významnosti 0,1.

### Výpočet pomocí systému STATISTICA:

Vytvoříme nový datový soubor o 2 proměnných a 12 případech. Do 1. proměnné X napíšeme hodnoty pro mezinárodní podnikání, do 2. proměnné hodnoty pro domácí podnikání.

Statistiky – Základní statistiky a tabulky – t-test pro závislé vzorky, OK - Proměnné X, Y – OK – Výpočet. Dostaneme tabulku:

t-test pro závislé vzorky (investovani)								
Označ. rozdíly jsou významné na hlad. $p < ,05000$								
Proměnná	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
X	11,91667	2,937480						
Y	13,25000	3,048845	12	-1,33333	2,188122	-2,11085	11	0,058490

Vypočtenou p-hodnotu 0,05849 porovnáme se zvolenou hladinou významnosti  $\alpha = 0,1$ . Protože  $p \leq \alpha$ , zamítáme nulovou hypotézu na hladině významnosti 0,1.