Cvičení 6.: Výpočet střední hodnoty a rozptylu

Příklad 1.: Postupně se zkouší spolehlivost čtyř přístrojů. Další se zkouší jen tehdy, když předchozí je spolehlivý. Každý z přístrojů vydrží zkoušku s pravděpodobností 0,8. Náhodná veličina X udává počet zkoušených přístrojů. Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.

Řešení:

X nabývá hodnot 1, 2, 3, 4 a její pravděpodobnostní funkce je:

$\pi(1) = 0,2,$

$\pi(2) = 0,8*0,2 = 0,16,$

$\pi(3) = 0,82*0,2 = 0,128,$

$\pi(4) = 0,83*0,2 + 0,84 = 0,512,$

$\pi(0) = 0$ jinak

$E(X) = 1*0,2 + 2*0,16 + 3*0,128 + 4*0,512 = 2,952$

$D(X) = 1^2*0,2 + 2^2*0,16 + 3^2*0,128 + 4^2*0,512 – 2,952^2 = 1,4697$

Postup ve STATISTICE:

Otevřeme nový datový soubor o dvou proměnných X a cetnost a čtyřech případech. Do proměnné X napíšeme 1, 2, 3, 4, do proměnné cetnost napíšeme 200, 160, 128, 512. Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – zavedeme proměnnou vah cetnost – OK - Proměnné X – OK – Detailní výsledky - zaškrtneme Průměr, Rozptyl – Výpočet.

	Popisné statistiky (Tabulka1)		
Proměnná	N platných	Průměr	Rozptyl
X	1000	2,952000	1,471167

Rozptyl však musíme upravit, musíme ho přenásobit číslem 999/1000. Do výstupní tabulky tedy přidáme za proměnnou Rozptyl novou proměnnou a do jejího Dlouhého jména napíšeme =v3*999/1000

	Popisné statistiky (Tabulka1)			
Proměnná	N platných	Průměr	Rozptyl	NProm
X	1000	2,952000	1,471167	1,469696

Příklad 2. (k samostatnému řešení): Náhodná veličina X udává počet ok při hodu kostkou. Pomocí systému STATISTICA vypočtěte její střední hodnotu a rozptyl.

Výsledek: E(X) = 3,5, D(X) = 2,9167

Příklad 3.: Ve 12 náhodně vybraných prodejnách ve městě byly zjištěny následující ceny určitého výrobku (v Kč): 102, 99, 106, 103, 96, 98, 100, 105, 103, 98, 104, 107. Těchto 12 hodnot považujeme za realizace náhodného výběru $X_1, ..., X_{12}$ z rozložení, které má střední hodnotu μ a rozptyl σ^2. Určete nestranné bodové odhady neznámé střední hodnoty μ a neznámého rozptylu σ^2.

Výpočet pomocí systému STATISTICA:

Otevřeme nový datový soubor o jedné proměnné (nazveme ji X) a 12 případech. Do proměnné X napíšeme zjištěné ceny.

Výpočet realizace výběrového průměru a výběrového rozptylu:

Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X – OK – Detailní výsledky – vybereme Průměr a Rozptyl – Výpočet. Dostaneme tabulku:

	Popisné statistiky (Tabulka15)	
Proměnná	Průměr	Rozptyl
X	101,7500	12,38636

Příklad 4.: Bylo zkoumáno 9 vzorků půdy s různým obsahem fosforu (veličina X). Hodnoty veličiny Y označují obsah fosforu v obilných klíčcích (po 38 dnech), jež vyrostly na těchto vzorcích půdy.

číslo vzorku	1	2	3	4	5	6	7	8	9
X	1	4	5	9	11	13	23	23	28
Y	64	71	54	81	76	93	77	95	109

Těchto 9 dvojic hodnot považujeme za realizace náhodného výběru (X_1,Y_1), ..., (X_9,Y_9) z dvourozměrného rozložení s kovariancí σ_{12} a koeficientem korelace ρ. Najděte bodové odhady výběrové kovariance σ_{12} a výběrového koeficientu korelace ρ.

Výpočet pomocí systému STATISTICA:

Otevřeme nový datový soubor o dvou proměnných X a Y 9 případech. Do proměnných X a Y zapíšeme zjištěné hodnoty obsafu fosforu v půdě a v obilných klíčcích.
Výpočet výběrové kovariance: Statistiky – Vícerozměrná regrese – Proměnné – Závisle proměnná Y, nezávisle proměnná X – OK – OK – Residua/předpoklady/předpovědi – Popisné statistiky – Další statistiky – Kovariance. Dostaneme tabulku:

	Kovariance (Tabulka18)	
Proměnná	X	Y
X	91,7500	130,0000
Y	130,0000	284,2500

Vidíme, že výběrová kovariance veličn X, Y se realizuje hodnotou 130. (Výběrový rozptyl proměnné X resp. Y nabyl hodnoty 91,75 resp. 284,25.)
Výpočet výběrového koeficientu korelace: V menu Další statistiky vybereme Korelace.

	Korelace (Tabulka18)	
Proměnná	X	Y
X	1,000000	0,804989
Y	0,804989	1,000000

Výběrový koeficient korelace veličin X, Y nabyl hodnoty 0,805, tedy mezi veličinami x, Y existuje silná přímá lineární závislost.
Upozornění: Výběrový koeficient korelace lze pomocí systému STATISTICA vypočítat i jiným způsobem: Statistika – Základní statistiky/tabulky – Korelační matice – OK – 1 seznam proměnných – X, Y – OK – Výpočet. Ve výsledné tabulce máme též realizace výběrových průměrů a směrodatných odchylek.

	Korelace (Tabulka18) Označ. korelace jsou významné na hlad. p < ,05000 N=9 (Celé případy vynechány u ChD)			
Proměnná	Průměry	Sm.odch.	X	Y
X	13,00000	9,57862	1,000000	0,804989
Y	80,00000	16,85972	0,804989	1,000000

Vzorce pro meze 100(1-α)% empirického intervalu spolehlivosti pro střední hodnotu μ normálního rozložení při známém rozptylu σ^2:

Oboustranný: $d = m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}$, $h = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}$.

Levostranný: $d = m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}$.

Pravostranný: $h = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}$.

Příklad 5.: Při kontrolních zkouškách životnosti 16 žárovek byl stanoven odhad m = 3000 h střední hodnoty jejich životnosti. Z dřívějších zkoušek je známo, že životnost žárovky se řídí normálním rozložením se směrodatnou odchylkou σ = 20 h. Vypočtěte

a) 99% empirický interval spolehlivosti pro střední hodnotu životnosti
b) 90% levostranný empirický interval spolehlivosti pro střední hodnotu životnosti
c) 95% pravostranný empirický interval spolehlivosti pro střední hodnotu životnosti.

Upozornění: Výsledek zaokrouhlete na jedno desetinné místo a vyjádřete v hodinách a minutách.

Řešení:

ad a)

$$d = m - \frac{\sigma}{\sqrt{n}} u_{0,995} = 3000 - \frac{20}{\sqrt{16}} 2,57583 = 2987,1,$$

$$h = m + \frac{\sigma}{\sqrt{n}} u_{0,995} = 3000 + \frac{20}{\sqrt{16}} 2,57583 = 3012,9$$

2987 h a 6 min < μ < 3012 h a 54 min s pravděpodobností 0,99

Výpočet pomocí systému STATISTICA

Otevřeme nový datový soubor o dvou proměnných d, h a jednom případu.
Do Dlouhého jména proměnné d napíšeme vzorec =3000-20/sqrt(16)*VNormal(0,995;0;1)
Do Dlouhého jména proměnné h napíšeme vzorec =3000+20/sqrt(16)*VNormal(0,995;0;1)

ad b)

$$d = m - \frac{\sigma}{\sqrt{n}} u_{0,9} = 3000 - \frac{20}{\sqrt{16}} 1,28155 = 2993,6$$

2993 h a 36 min < μ s pravděpodobností 0,9

Výpočet pomocí systému STATISTICA

Otevřeme nový datový soubor o jedné proměnné d a jednom případu.
Do Dlouhého jména proměnné d napíšeme vzorec =3000-20/sqrt(16)*VNormal(0,9;0;1)

ad c)

$$h = m + \frac{\sigma}{\sqrt{n}} u_{0,975} = 3000 + \frac{20}{\sqrt{16}} 1,95996 = 3009,8$$

3009 h a 48 min > μ s pravděpodobností 0,95

Výpočet pomocí systému STATISTICA

Otevřeme nový datový soubor o jedné proměnné h a jednom případu.
Do Dlouhého jména proměnné h napíšeme vzorec =3000+20/sqrt(16)*VNormal(0,975;0;1)

Užitečný odkaz: na adrese http://www.prevody-jednotek.cz je program, s jehož pomocí lze převádět různé fyzikální jednotky, v našem případě hodiny na minuty.

Základní poznatky o testování hypotéz

Předpokládáme, že testujeme nulovou hypotézu H_0: $h(\vartheta) = c$, kde $c \in \mathbb{R}$ buď proti oboustranné alternativě H_1: $h(\vartheta) \neq c$ nebo proti levostranné alternativě H_1: $h(\vartheta) < c$ nebo proti pravostranné alternativě H_1: $h(\vartheta) > c$.

Testování pomocí kritického oboru

Najdeme testovou statistiku $T_0 = T_0(X_1, ..., X_n)$. Množina všech hodnot, jichž může testová statistika nabýt, se rozpadá na obor nezamítnutí nulové hypotézy (značí se V) a obor zamítnutí nulové hypotézy (značí se W a nazývá se též kritický obor). W aV jsou odděleny kritickými hodnotami (pro danou hladinu významnosti α je lze najít ve statistických tabulkách).
Jestliže číselná realizace t_0 testové statistiky T_0 padne do kritického oboru W, pak nulovou hypotézu zamítáme na hladině významnosti α a znamená to skutečné vyvrácení testované hypotézy. Jestliže t_0 padne do oboru nezamítnutí V, pak jde o pouhé mlčení, které platnost nulové hypotézy jenom připouští.
Stanovení kritického oboru pro danou hladinu významnosti α:
Označme t_{min} (resp. t_{max}) nejmenší (resp. největší) hodnotu testového kritéria.
Kritický obor v případě oboustranné alternativy má tvar
$W = (t_{min}, K_{\alpha/2}(T)\rangle \cup \langle K_{1-\alpha/2}(T), t_{max})$, kde $K_{\alpha/2}(T)$ a $K_{1-\alpha/2}(T)$ jsou kvantily rozložení, jímž se řídí testové kritérium T_0, je-li nulová hypotéza pravdivá.
Kritický obor v případě levostranné alternativy má tvar:
$W = (t_{min}, K_{\alpha}(T)\rangle$.
Kritický obor v případě pravostranné alternativy má tvar:
$W = \langle K_{1-\alpha}(T), t_{max})$.

Testování pomocí intervalu spolehlivosti

Sestrojíme 100(1-α)% empirický interval spolehlivosti pro parametrickou funkci $h(\vartheta)$.
Pokryje-li tento interval hodnotu c, pak H_0 nezamítáme na hladině významnosti α, v opačném případě H_0 zamítáme na hladině významnosti α.
Pro test H_0 proti oboustranné alternativě sestrojíme oboustranný interval spolehlivosti.
Pro test H_0 proti levostranné alternativě sestrojíme pravostranný interval spolehlivosti.
Pro test H_0 proti pravostranné alternativě sestrojíme levostranný interval spolehlivosti.

Testování pomocí p-hodnoty

p-hodnota udává nejnižší možnou hladinu významnosti pro zamítnutí nulové hypotézy:
je-li $p \leq \alpha$, pak H_0 zamítáme na hladině významnosti α, je-li $p > \alpha$, pak H_0 nezamítáme na hladině významnosti α.
Způsob výpočtu p-hodnoty:
Pro oboustrannou alternativu $p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\}$.
Pro levostrannou alternativu $p = P(T_0 \leq t_0)$.
Pro pravostrannou alternativu $p = P(T_0 \geq t_0)$.

Příklad 6.: Víme, že výška hochů ve věku 9,5 až 10 let má normální rozložení s neznámou střední hodnotou μ a známým rozptylem $\sigma^2 = 39{,}112\ cm^2$. Dětský lékař náhodně vybral 15 hochů uvedeného věku, změřil je a vypočítal realizaci výběrového průměru m = 139,13 cm. Podle jeho názoru by výška hochů v tomto věku neměla přesáhnout 142 cm s pravděpodobností 0,95. Lze tvrzení lékaře akceptovat?

Řešení: Testujeme H_0: $\mu = 142$ proti H_1: $\mu < 142$ (to je tvrzení lékaře) na hladině významnosti 0,05.

a) Test provedeme pomocí kritického oboru.

Pro úlohy o střední hodnotě normálního rozložení při známém rozptylu používáme pivotovou statistiku $U = \dfrac{M-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$. Testová statistika tedy bude $T_0 = \dfrac{M-c}{\frac{\sigma}{\sqrt{n}}}$ a bude mít rozložení N(0, 1), pokud je nulová hypotéza pravdivá. Vypočítáme realizaci testové statistiky:

$$t_0 = \frac{139{,}13 - 142}{\frac{\sqrt{39{,}112}}{\sqrt{15}}} = -1{,}7773 .$$

Stanovíme kritický obor: $W = (-\infty, u_\alpha \rangle = (-\infty, u_{0,05} \rangle = (-\infty, -u_{0,95} \rangle = (-\infty, -1{,}6449 \rangle$.

Protože $-1{,}7773 \in W$, H_0 zamítáme na hladině významnosti 0,05. Tvrzení lékaře lze tedy akceptovat s rizikem omylu 5 %.

b) Test provedeme pomocí intervalu spolehlivosti.

Meze 100(1-α)% empirického pravostranného intervalu spolehlivosti pro střední hodnotu μ při známém rozptylu σ^2 jsou: $(-\infty, h) = (-\infty, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha})$.

V našem případě dostáváme: $h = 139{,}13 + \frac{\sqrt{39{,}112}}{\sqrt{15}} u_{0,95} = 139{,}13 + \frac{\sqrt{39{,}112}}{\sqrt{15}} 1{,}645 = 141{,}79$.

Protože $142 \notin (-\infty; 141{,}79)$, H_0 zamítáme na hladině významnosti 0,05.

c) Test provedeme pomocí p-hodnoty

$p = P(T_0 \le t_0) = \Phi(-1{,}7773) = 0{,}0378$

Jelikož $0{,}0378 \le 0{,}05$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Při řešení tohoto příkladu použijeme systém STATISTICA pouze jako inteligentní kalkulátor.