

## 9 Statistika, odhady parametrů

### Teorie: Základní statistické zpracování a bodové odhady

**Populace** je soubor objektů (statistických jednotek), který je vymezen jejich výčtem nebo charakterizací jejich vlastností, může být proto konečný (rozsah označujeme  $N$ ) i nekonečný. Typ rozdělení ani jeho parametry neznáme.

**Statistická šetření** umožní na základě opakovaných pozorování  $X_1, X_2, \dots, X_n$  usuzovat na vlastnosti populace.

**Náhodný výběr** je pokud pozorování  $X_1, X_2, \dots, X_n$  jsou nezávislá a stejně rozdělená.

**Data** je souhrn konkrétních číselných údajů  $x_1, x_2, x_3, \dots, x_n$  (výsledek opakovaných pokusů).

Populace	←Výběr	←Excel
stř. hodnota $EX$	průměr $\bar{x} = 1/n \sum_{i=1}^n x_i$ (včetně textových buněk)	PRŮMĚR(data) AVERAGEA(data)
rozptyl $varX$	výběrový rozptyl $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	VAR.S(data) VAR.VÝBĚR(data)
směr.odchylka $\sqrt{varX}$	výběrová odchylka $s = \sqrt{s^2}$	SMODCH.VÝBĚR.S(data) STDEV.S(data)
medián $x_{50\%}$	výběrový medián	MEDIAN(data)
kvartily $x_{0\%}, x_{25\%}, x_{50\%}, x_{75\%}, x_{100\%}$		QUARTIL.INC(data;0.4) QUARTIL.EXC(data;0.4)
kvantil $x_\alpha$		PERCENTIL.INC(data;\alpha) PERCENTIL.EXC(data;\alpha)
pořadí (0-sestupně,1-vzestupně)		RANK.EQ(hodnota;data;0,1) RANK.AVG(hodnota;data;0,1)
funkce hustoty	histogram	ČETNOSTI(data; hodnoty)
distribuční funkce	kumulované četnosti	

### Teorie: Intervalové odhady parametrů $\Theta$ ve tvaru $(D; H)$ se spolehlivostí $1 - \alpha$

Pro parametr  $\mu$  normálního rozdělení (pro známý rozptyl  $\sigma^2$ )  $\left( \bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$

Pro parametr  $\mu$  normálního rozdělení (pro neznámý rozptyl)  $\bar{x} \pm t_{1-\frac{\alpha}{2}}(\nu = n-1) \cdot \frac{s}{\sqrt{n}}$

Pro parametr  $\sigma^2$  normálního rozdělení  $\left( \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(\nu = n-1)}; \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(\nu = n-1)} \right)$

Pro parametr  $p$  alternativního rozdělení  $\left( \hat{p} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$

(9.1) Z náhodného výběru

9.99 10.42 10.34 10.49 10.02 11.23

10.22 9.95 10.32 10.76 9.70 10.86

odhadněte střední hodnotu, rozptyl a směrodatnou odchylku.

$$[\bar{x} = 10.358; s^2 = 0.188; s = 0.433]$$

(9.2) Z náhodného výběru rozsahu  $n = 25$  byl určen aritmetický průměr  $\bar{x} = 151.25$  a výběrový rozptyl  $s^2 = 32.5$ .

(a) Sestrojte 95% interval spolehlivosti pro střední hodnotu  $\mu$ .

$$[\text{pro známý rozptyl } D = \bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 151.25 - u_{0.975} \sqrt{\frac{32.5}{25}} = 149.02, ]$$

[kde  $u_{0.975} = 1.96$  podle tabulek nebo v Excelu NORM.S.INV(0.975) ]

[CONFIDENCE.NORM( $\alpha$ ;s;n) ]

$$[ \text{pro neznámý rozptyl } D = \bar{x} - t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 151.25 - t_{0.975} \sqrt{\frac{32.5}{25}} = 148.90, ]$$

[kde  $t_{0.975} = 2.06$  podle tabulek nebo v Excelu T.INV.2T(0.05;25-1) ]

(b) Sestrojte 90% interval spolehlivosti pro střední hodnotu  $\mu$ .

[ pro známý rozptyl: (149.37; 153.13) ]

[CONFIDENCE.T( $\alpha$ ;s;n) ]

(c) Určete rozsah výběru  $n$  tak, aby šířka 95% intervalu spolehlivosti pro střední hodnotu  $\varepsilon = H - D$  byla 2.

[ přibližně  $n = 127$  ]

(d) Sestrojte 90% a 95% intervaly spolehlivosti pro rozptyl  $\sigma^2$ .

[ 90% interval : (21.42; 56.32) ;95% interval : (19.82; 62.90) ]

[ kvantily  $\chi^2$  rozdělení v tabulkách nebo v Excelu CHISQ.INV( $\alpha$ ;volnost) ]

(9.3) Proved'te  $n$  krát pokus hození mincí a na základě získaných dat odhadněte bodově a intervalově pravděpodobnost, že padne líc.

(9.4) Proved'te základní statistické zpracování výsledků prvních testů.

(a) spočtete základní charakteristiky souboru - průměr, rozptyl, směrodatná odchylka

(b) vykreslete histogram dat

(c) sestrojte intervalové odhady počtu získaných bodů

## Výpočty kvantilů v Excelu

normální rozdělení	$u_\alpha$	NORM.S.INV( $\alpha$ )
	$u_{1-\alpha/2}$	NORM.S.INV( $1 - \alpha/2$ ) nebo NORMSINV( $1 - \alpha/2$ )
t- rozdělení	$t_\alpha$	T.INV( $\alpha$ ;volnost)
	$t_{1-\alpha/2}$	T.INV( $1 - \alpha/2$ ;volnost) nebo T.INV.2T( $\alpha$ ;volnost)
		nebo TINV( $\alpha$ ;volnost)
$\chi^2$ - rozdělení	$\chi_\alpha^2$	CHISQ.INV( $\alpha$ ;volnost) nebo CHIINV( $1-\alpha$ ;volnost)

**Odhady ve WolframAlpha** confidence interval for mean, sample mean, variance , standard deviation