7 Základní pojmy matematické statistiky

7.1 Bodové odhady výběrových charakteristik

Příklad 7.1. Ve 12-ti náhodně vybraných internetových obchodech byly zjištěny následující ceny deskriptoru artefaktů (v Kč): $102, 99, 106, 103, 96, 98, 100, 105, 103, 98, 104, 107$. Těchto 12 hodnot považujeme za realizace náhodného výběru $X_1, \ldots, X_{12}$ z rozdělení, které má střední hodnotu μ a rozptyl σ^2.

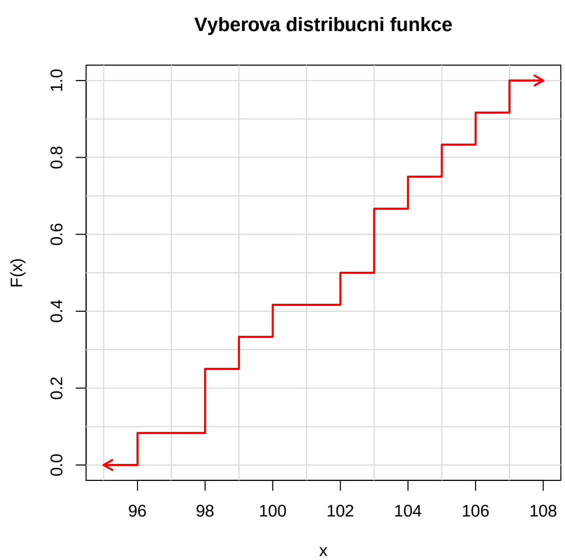
a) Určete nestranné bodové odhady neznámé střední hodnoty μ a neznámého rozptylu σ^2.

b) Najděte výběrovou distribuční funkci $F_{12}(x)$ a nakreslete její graf.

```
## [1] 101.75
## [1] 12.38636
```

Pro usnadnění výpočtu hodnot výběrové distribuční funkce $F_{12}(x)$ uspořádáme ceny podle velikosti: 96, 98, 98, 99, 100, 102, 103, 103, 104, 105, 106, 107. Číselnou osu rozdělíme na 11 intervalů a v každém intervalu stanovíme hodnotu výběrové distribuční funkce:

```
##        [,1] [,2]   [,3]   [,4] [,5]   [,6] [,7]   [,8]   [,9] [,10]
## [1,] 0.0833 0.25 0.3333 0.4167  0.5 0.6667 0.75 0.8333 0.9167     1
```



Příklad 7.2. Přírůstky cen akcií v % na burze v New Yorku u 10 náhodně vybraných společností dosáhly těchto hodnot: $10, 16, 5, 10, 12, 8, 4, 6, 5, 4$.

a) Odhadněte střední hodnotu, rozptyl a směrodatnou odchylku růstu cen akcií.

b) Odhadněte pravděpodobnost růstu cen akcií aspoň o 8.5 %.

ad a)
```
##            m    s2    s
## akcie 101.75 15.78 3.97
```

ad b)
```
## [1] 0.4
```

Průměrný růst cen akcií odhadujeme na 8 % se směrodatnou odchylkou 3.97 %. Dále, u 40 % akcií vzrostla cena alespoň o 8.5 %.

Příklad 7.3. 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 $(X_1, Y_1), \ldots, (X_9, Y_9)$ z dvourozměrného rozdělení s kovariancí σ_{12} a koeficientem korelace ρ. Najděte bodové odhady kovariance σ_{12} a koeficientu korelace ρ.

```
## [1] 130
## [1] 0.8049892
```

Výběrová kovariance veličin X, Y se realizuje hodnotou 130. Výběrový koeficient korelace veličin X, Y nabyl hodnoty 0.805, tedy mezi veličinami X, Y existuje silná přímá lineární závislost.

7.2 Intervalové odhady výběrových charakteristik

Nechť $X_1, \ldots, X_n$ je náhodný výber z rozdělení $L(\theta)$, θ je sledovaný parametr a $\alpha \in (0, 1)$.

Interval (D, H) se nazývá $100(1-\alpha)$ % *oboustranný interval spolehlivosti* parametru θ, pokud pro každé $\theta \in \Theta$ platí

$$\Pr(D < \theta < H) = 1 - \alpha.$$

Interval (D, ∞) se nazývá $100(1-\alpha)$ % *levostranný interval spolehlivosti* parametru θ, pokud pro každé $\theta \in \Theta$ platí

$$\Pr(D < \theta) = 1 - \alpha.$$

Interval $(-\infty, H)$ se nazývá $100(1-\alpha)$ % *pravostranný interval spolehlivosti* parametru θ, pokud pro každé $\theta \in \Theta$ platí

$$\Pr(\theta < H) = 1 - \alpha.$$

Číslo α se nazývá *riziko*, číslo $1-\alpha$ se nazývá *spolehlivost*.

Příklad 7.4. 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 rozdělením se směrodatnou odchylkou $\sigma = 20\,h$. 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.

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

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.

ad a)
```
## [1] 2987.121
## [1] 3012.879
```

2987 h a 6 min $< \mu <$ 3012 h a 54 min s pravděpodobností 0.99.

ad b)
```
## [1] 2993.592
```

2993 h a 36 min $< \mu$ s pravděpodobností 0.9.

ad c)
```
## [1] 3008.224
```

3009 h a 48 min $> \mu$ s pravděpodobností 0.95.

7.3 Úvod do testování hypotéz

Příklad 7.5. Víme, že výška hochů ve věku 9.5 až 10 let má normální rozdělení 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?

Testujeme $H_0 : \mu \geq 142$ proti $H_1 : \mu < 142$ na hladině významnosti $\alpha = 0.05$.

a) Test provedeme pomocí kritického oboru.
Pro úlohy o střední hodnotě normálního rozdělení při známém rozptylu používáme pivotovou statistiku

$$U = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

```
## [1] "t0 =  -1.777348"
## [1] "Wh = -1.644854"
```

$$W = (-\infty, -1.645\rangle$$

b) Test provedeme pomocí intervalu spolehlivosti.
Meze $100(1-\alpha)$ % empirického pravostranného intervalu spolehlivosti pro střední hodnotu μ při známém rozptylu σ^2 jsou

$$(-\infty, h) = (-\infty, m - \frac{\sigma}{\sqrt{n}} u_\alpha).$$

```
## [1] "hh =  141.786052"
```

$$(-\infty, 141.79).$$

c) Test provedeme pomocí p-hodnoty.

$$p = \Pr(T_0 \leq t_0)$$

```
## [1] "p-hodnota =  0.037755"
```

Závěr testování: Tvrzení lékaře lze tedy akceptovat s rizikem omylu 5 %.