Cvičení 12: Limitní věty, normální rozdělení, náhodný výběr

Teorie:

Náhodným výběrem rozsahu n rozumíme n-tici **stochasticky nezávislých** a náhodných veličin $X_1, \ldots, X_n$, které mají totéž rozdělení. S náhodným výběrem se obvykle setkáváme při opakovaném provádění téhož pokusu.

Statistika je náhodná veličina vzniklá transformací náhodného výběru.

- Výběrový průměr $M = \frac{1}{n}\sum_{i=1}^{n} X_i$, a jsou-li navíc $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, pak $M \sim N(\mu, \sigma^2/n)$.
- Výběrový rozptyl $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - nM\right), S = \sqrt{S^2}$.

Intervalovým odhadem parametru θ rozumíme interval (T_L, T_U), kde $T_L(X_1, \ldots, X_n)$ a $T_U(X_1, \ldots, X_n)$ jsou statistiky výběru $(X_1, \ldots, X_n)$. Platí-li

$$P(T_L \leq \theta \leq T_U) = 1 - \alpha,$$

říkáme, že (T_L, T_U) je $100 \cdot (1-\alpha)\%$ interval spolehlivosti pro parametr θ. **Horním odhadem** parametru θ na hladině významnosti $1-\alpha$ je statistika U, pro níž

$$P(\theta < U) \geq 1 - \alpha,$$

dolním odhadem θ na hladině významnosti $1-\alpha$ je pak statistika L, pro níž

$$P(L < \theta) \geq 1 - \alpha.$$

Případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$:

- M a S^2 jsou nezávislé náhodné veličiny.
- $M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.
- $K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- $\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.
- $T = (M - \mu)/(S/\sqrt{n}) \sim t(n-1)$.

Příklad 162. Pravděpodobnost, že zasazený strom se ujme, je 0,8. Jaká je pravděpodobnost, že z 500 zasazených stromů se jich ujme aspoň 360?

Výsledek. 0,987.

Příklad 163. Pravděpodobnost, že semeno vyklíčí, je 0,9. Kolik semen je třeba zasadit, aby s pravděpodobností aspoň 0,995 vyklíčilo cca 90% semen (což přesněji formulujeme se zpřesňujícím požadavkem, aby odchylka podílu vyklíčených semen od 0,9 nepřevýšila 0,034).

Výsledek. $n \approx 600$.

Příklad 164. Životnost (v hodinách) určité elektrické součástky má exponenciální rozdělení s parametrem $\lambda = \frac{1}{10}$. Pomocí centrální limitní věty odhadněte pravděpodobnost, že celková životnost 100 takových součastek bude mezi 900 a 1050 hodinami.

Výsledek. $\mu = 10, \sigma^2 = 100, P(900 \leq \sum X_i \leq 1050) = P\left(\frac{900-n\mu}{\sigma\sqrt{n}} \leq \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \leq \frac{1050-n\mu}{\sigma\sqrt{n}}\right) = \Phi(0{,}5) - \Phi(-1) \approx 0{,}533$.

Příklad 165. Při 600 hodech kostkou padla jednička pouze 45 krát. Rozhodněte, jestli je možné tvrdit, že jde o ideální kostku na hladině $\alpha = 0{,}01$. Vše zdůvodněte a svůj závěr explicitně formulujte.

Příklad 166. Do bedny ukládáme výrobky se střední hodnotou 3 kg a směrodatnou odchylkou 0,8 kg. Jaký maximální počet výrobků můžeme do bedny uložit, aby celková hmotnost s pravděpodobností 0,9738 nepřekročila jednu tunu?

Výsledek. $n \approx 324$.

Příklad 167. Předpokládejme, že výška desetiletých chlapců má normální rozdělení $N(\mu, \sigma^2)$. S neznámou střední hodnotou μ a rozptylem $\sigma^2 = 39{,}112$. Změřením výšky 15 chlapců jsme určili výběrový průměr $M = 139{,}13$. Určete

a) 99% oboustranný interval spolehlivosti pro parametr μ,

b) dolní odhad μ na hladině významnosti 95%.

Výsledek. a) $(136{,}12; 142{,}14)$; b) $136{,}474$.

Příklad 168. Odběratel provádí kontrolu jakosti námi dodaných výrobků namátkovou kontrolou testovaného rozměru u 21 náhodně vybraných výrobků. Dodávka bude přijata, pokud nebude výběrová směrodatná odchylka překračovat hodnotu 0,2 mm. Víme přitom, že naše stroje produkují výrobky, u nichž má sledovaný rozměr normální rozdělení tvaru $N(10\,\mathrm{mm}; 0,0737\,\mathrm{mm}^2)$. S využitím statistických tabulek určete pravděpodobnost, s níž bude dodávka přijata. Jak se změní odpověď, pokud odběratel kvůli nákladům na testy začne testovat pouze 4 výrobky?
(V případě chybějících údajů v tabulce hodnoty, které máte k dispozici, lineárně interpolujte).