UNIVERZITA PALACKÉHO V OLOMOUCI
KATEDRA MATEMATICKÉ INFORMATIKY

STOCHASTICKÉ METODY - STATISTIKA

Kolokviální příklad

Květen 2001

Jan Outrata
Informatika, III. ročník

Zadání

Byla změřena velikost 30-ti CD z časopisu. Hodnoty v MB:

175, 581, 647, 650, 390, 641, 487, 590, 648, 647,
556, 647, 621, 601, 642, 633, 657, 648, 651, 668,
664, 670, 649, 669, 660, 663, 646, 630, 624, 585.

Vypočítejte charakteristiky tohoto stat. souboru a zjistěte, zda má normální rozložení. Lze považovat odchylky od jmenovité hodnoty CD 650 MB za normální?

Řešení

Mezi charakteristiky stat. souboru patří:

a) charakteristiky úrovně

charakteristika	vzorec	hodnota
aritmetický průměr	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	608.
harmonický průměr	$\overline{x}_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$	567.754
geometrický průměr	$\overline{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$	593.977
modus	hodnota znaku s největší četností	647.
medián	uspořádání souboru a $\tilde{x} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n+2}{2}})$	646.5

b) charakteristiky variability

charakteristika	vzorec	hodnota
variační rozpětí	$R = x_{max} - x_{min}$	495.
rozptyl	$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$	$9823.\overline{3}$
směrodatná odchylka	$s = \sqrt{s^2}$	99.113
variační koeficient	$v = \frac{s}{\overline{x}}$	0.163
průměrná odchylka	$\overline{d_{\overline{x}}} = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \overline{x}\rvert$	$59.9\overline{3}$

c) charakteristiky šikmosti a špičatosti

charakteristika	vzorec	hodnota
koeficient šikmosti	$a_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2)^{\frac{3}{2}}}$	-17.162
koeficient špičatosti	$a_4 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2)^2} - 3$	10.034

Má soubor normální rozložení?

Hypotéza H: Soubor má normální rozložení
Alternativní hypotéza A: Soubor nemá normální rozložení.

Použijeme tzv. A-test, který je založen na porovnávání empirických a teoretických koeficientů šikmosti a špičatosti.
Hypotézu H zamítneme, když

$$|a_3| \geq u_{1-\frac{\alpha}{2}}\sqrt{D^2(a_3)} \text{ a } |a_4 + \tfrac{6}{n+1}| \geq u_{1-\frac{\alpha}{2}}\sqrt{D^2(a_4)},$$

$D^2(a_3) = \frac{6(n-1)}{(n+1)(n+3)}$, $D^2(a_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$, $u_{1-\frac{\alpha}{2}}$ je kvantil normálního rozložení.

$$17.162 \geq 0.808 \text{ a } 10.228 \geq 1.373$$

$$\Downarrow$$

Hypotézu H zamítneme a soubor tedy normální rozložení nemá.

Pokud by soubor měl normální rozložení, řešili bychom poslední úkol ze zadání testováním hypotézy o střední hodnotě.

Jelikož soubor nemá normální rozložení, nemůžeme testovat hypotézy podle vět z přednášek, protože ty předpokládají soubor s normálním rozložením.