

Nemparaméteres próbák

Statisztika II., 3. alkalom
Medián vizsgálata, egy minta

Nemparaméteres eljárások

A következőkben változók középértékét vizsgáló nemparaméteres eljárásokkal foglalkozunk.

Paraméteres eljárások:

- t-próbák (egymintás, kétmintás, páros)

- variancia analízis (egyszempontos, többszempontos)

Feltételeik: normális eloszlás, szóráshomogenitás

Nemparaméteres eljárások:

- Az intervallum skála sem követelmény, de jó, ha a változó folytonos vagy legalább finom beosztású.

- Gyakran nem a tényleges értékekkel, hanem úgynevezett rangokkal dolgoznak.

Nemparaméteres eljárások

	paraméteres eljárások	nemparaméteres eljárások
egy minta	egymintás t-próba	BINOMIÁLIS WALD-WOLFOWITZ
két független minta	független mintás (kétmintás) t-próba	KOLMOGOROV-SMIRNOV WALD-WOLFOWITZ MANN-WHITNEY
két összetartozó minta	páros t-próba	WILCOXON
több független minta egy szempont szerint	egyszempontos varianciaanalízis	KRUSKAL-WALLIS
több összetartozó minta egy szempont szerint	egyszempontos varianciaanalízis, ismételt mérés	FRIEDMAN

Binomiális próba, egy minta

A binomiális próbát használhatjuk nem binomiális eloszlású változó esetén is, egymintás esetben medián vizsgálatára. Ekkor a feltételezett középérték vagy annál kisebb értékeket 0-val, a középértéknél nagyobb értékeket 1-el kódoljuk.

$$B(N, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

H0: A medián megegyezik a feltételezett értékkel

H1: A medián nem egyezik meg a feltételezett értékkel

Ha $n > \frac{9}{np(1-p)}$, akkor jó közelítést ad: $z = \frac{\hat{\pi} - 0.5}{SE(\hat{\pi})} = \frac{\hat{\pi} - 0.5}{\sqrt{\pi(1-\pi)/N}}$

Példa:

IQ ált. isk. osztályban, lehet a medián 100? IQ értékek:

87, 101, 97, 96, 72, 73, 97, 128, 99, 97, 96, 83, 125, 88, 95, 107, 129

Átkódolva

0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1,

$$\hat{\pi} = 0.29 \quad p=0.14$$

Wald-Wolfowitz próba, egy minta

Alapelv: Ha a dichotóm változó értékei véletlenszerűen váltakoznak, akkor az értékek véletlenszerű, azonos értékekből álló sorozatokba szerveződnek.

A próbastatisztika

értéke a szériák száma: R (runs).

n_+ = pozitív szériák elemszáma n_- = negatív szériák elemszáma

$$z = \frac{R - \mu_r}{SE(R)} = \frac{R - \mu_r}{\sigma_r} \quad \mu_r = \frac{2n_+n_-}{n_+ + n_-} + 1 \quad \sigma_r = \sqrt{\frac{(\mu_r - 1)(\mu_r - 2)}{N - 1}}$$

Bár alapvetően dichotóm változók vizsgálatára fejlesztették ki, átlag, medián vizsgálatára is alkalmas. Vagy bármely osztópont esetén megvizsgálható, hogy az értékek alá illetve fölé kerülése véletlenszerű-e. Ha $N < 50$, korrekciót kell végrehajtani.

Példa:

IQ ált. isk. osztályban, lehet a medián 100?

IQ értékek:

87, 101, 97, 96, 72, 73, 97, 128, 99, 97, 96, 83, 125, 88, 95, 107, 129

Átkódolva

-, +, -, -, -, -, -, +, -, -, -, -, +, -, -, +, +,