# Pearsoni e. $\chi^2$-kriteeriumi kasutamine empiirilise jaotusseaduse võrdlemisel teoreetilisega

Eelnevates arutlustes eeldasime, et käsitletava juhusliku suuruse jaotus on teada. Kuidas aga tegelikkuses kindlaks teha, millise jaotusega on vaadeldav juhuslik suurus? Kuna praktikas on meie poolt käsitletav juhuslik suurus antud tema teatud väärtuste hulgaga, siis võib sellele küsimusele vastamiseks võrrelda juhusliku suuruse praktikast saadud väärtusi nende väärtustega, mis tal oleksid, kui ta oleks jaotunud oletatava teoreetilise jaotuse järgi. Eelpool öeldu võib kirja panna järgmise hüpoteeside paarina:

**$H_0$:** uuritav üldkogum on etteantud teoreetilise jaotusega (s.t. jaotuse tihedusfunktsioon on $p(x)$);
**$H_1$:** uuritav üldkogum ei ole etteantud teoreetilise jaotusega.

<u>Hüpoteesipaari kontrollitakse järgneva algoritmi kohaselt:</u>

1) Punktidega $a_0, a_1, \ldots, a_m$ jaotatakse valim klassideks ning leitakse igasse klassi kuuluvate objektide arv. Tulemuseks on sagedustabel:

| | $[a_0 ; a_1)$ | $[a_1 ; a_2)$ | ... | $[a_{m-1} ; a_m]$ |
|---|---|---|---|---|
| $p_i^* = \frac{n_i}{n}$ | $\frac{n_1}{n}$ | $\frac{n_2}{n}$ | ... | $\frac{n_m}{n}$ |

Klassid tuleb moodustada nii, et ühtegi klassi ei satuks alla viie elemendi (kui see nõue pole täidetud, siis tuleb osa klasse ühendada nii, et igas uues klassis oleks vähemalt 5 elementi).

2) Leitakse, mitu elementi $n_i'$ peaks kuuluma igasse klassi antud teoreetilise jaotuse korral:

$$n_i' = n \cdot \int_{a_{i-1}}^{a_i} p(x)dx = n \cdot (F(a_i) - F(a_{i-1})), \quad i = 1, 2, \ldots, m.$$

Näiteks normaaljaotuse korral oleks see valem kujul $n_i' = n\left[\Phi\left(\frac{a_i - \bar{x}}{s}\right) - \Phi\left(\frac{a_{i-1} - \bar{x}}{s}\right)\right]$.

3) Arvutatakse välja teststatistiku empiiriline väärtus vastavalt valemile

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'}.$$

On tõestatud, et kui $n \to \infty$, siis selle juhusliku suuruse jaotus läheneb $\chi^2$-jaotusele vabadusastmete arvuga $k = m - 1 - r$, kus $r$ on uuritava jaotuse parameetrite arv ja $m$ on klasside arv. Normaaljaotuse korral $r = 2$, seega $k = m - 3$.

4) Kasutatakse parempoolset kriitilist piirkonda $P(\chi^2 > \chi^2_{kr}(\alpha, k)) = \alpha$. Ette antakse olulisuse nivoo $\alpha$ ning leitakse kriitiline punkt $\chi^2_{kr}(\alpha, k)$ $\chi^2$-jaotuse täiendkvantiilide tabelist (või *MS Exceli* keskkonnas kasutades funktsiooni *CHIINV*($\alpha$; $k$)).

5) Kui $\chi^2_{emp} < \chi^2_{kr}(\alpha, k)$, siis pole alust nullhüpoteesi tagasi lükata. Vastupidisel juhul lükatakse nullhüpotees tagasi ja võetakse vastu sisukas hüpotees (s.t. jaotused tunnistatakse erinevaks).

<u>Näide.</u> Järgnevas sagedustabelis on antud detailide (võllide) diameetrite hälbed (mikromeetrites) etteantud normatiivist. Kontrollida andmete vastavust normaaljaotusele olulisuse nivool $\alpha = 0{,}05$.

| | [-20 ; -15) | [-15 ; -10) | [-10 ; -5) | [-5 ; 0) | [0 ; 5) | [5 ; 10) | [10 ; 15) | [15 ; 20) | [20 ; 25) | [25 ; 30) |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 7 | 11 | 15 | 24 | 49 | 41 | 26 | 17 | 7 | 3 |

Ülesande lahendus on toodud lisana esitatud Exceli failis.