

Ühe tunnuse analüüs

Eesmärk - üksiktunnuse kirjeldamine ja tema statistilise olemuse selgitamine

Andmetel kaks tendentsi:

- 1) koonduda mingi keskme ümber (keskmised),
- 2) samal ajal hälbida sellest keskmest (variatsioonikarakteristikud)

Tunnuse karakteristikud

1) keskmised

- mahukeskmised
 - aritmeetiline-
 - harmooniline-
 - astme-
 - geomeetriline-
 - jt
- asendikeskmised
 - mood
 - mediaan
 - kvantiilid:
 - * kvartiilid
 - * detsiilid
 - * tsentiilid
 - * jt

2) variatsioonikarakteristikud

KESKMISED

Mahukeskmised

1. Aritmeetiline keskmine (AK)

- lihtne (grupeerimata andmed)

$$M = (\sum X_i) / N, \quad i=1, \dots, N, \quad \text{kus}$$

X_i - statistilise rea liige (variant),

N - rea liikmete (objektide) arv.

- kaalutud (grupeeritud andmed)

$$M = (\sum X_j * f_j) / (\sum f_j), \quad j=1, \dots, K, \quad \text{kus}$$

f_j - variandi X_j esinemissagedus,

K - variantide (tunnuse skaalaväärtuste) arv.

2. Harmooniline keskmine

- $M_H = N / (\sum 1/X_i)$

- $M_H = (\sum f_j) / (\sum f_j / X_j)$

3. Astmekeskmine

$$M_A = \sqrt[T]{(\sum X_i^T) / N}, \quad \text{kus } T - \text{astendaja}$$

4. Geomeetriline keskmine

$$M_G = \sqrt[N]{(\prod X_i)}, \quad i=1, \dots, N, \quad \text{kus } N - \text{objektide arv}$$

Asendikeskmised

1. Mood - kõige sagedamini esinev väärtus

Intervallitud variatsioonirea korral:

$M_O = X_{M_O} + P(f_{M_O} - f_{M_O-1}) / [(f_{M_O} - f_{M_O-1}) + (f_{M_O} - f_{M_O+1})]$, kus

X_{M_O} - moodintervalli (MoI) täpne alumine piir,

P - intervalli pikkus,

f_{M_O} - MoI-le vastav sagedus,

f_{M_O-1} - MoI-le eelneva intervalli sagedus,

f_{M_O+1} - MoI-le järgneva intervalli sagedus.

2. Mediaan - korrastatud statistilise rea keskliige

Intervallitud variatsioonirea korral:

$M_E = X_{M_E} + (0,5 \cdot N - S) \cdot P / f_{M_E}$, kus

X_{M_E} - mediaanintervalli (MeI) täpne alumine piir,

N - objektide arv,

S - MeI-le eelnevate intervallide sageduste summa,

P - intervalli pikkus,

f_{M_E} - MeI sagedus.

3. Kvantiilid

- kvartiilid - korrastatud statistilise rea liikmed, mis jagavad rea 4-ks võrdse liikmete arvuga osaks

| | | | |
|-------|-------|-------|-------------|
| Q_1 | Q_2 | Q_3 | $Q_2 = M_E$ |
| 25% | 50% | 75% | |

$Q_1 = X_{Q_1} + P(0,25N - S_{Q_1}) / f_{Q_1}$,

$Q_3 = X_{Q_3} + P(0,75N - S_{Q_3}) / f_{Q_3}$, kus

X_{Q1} - esimese kvartiilintervalli (KI_1) alumine piir,
 X_{Q3} - kolmanda kvartiilintervalli (KI_3) alumine piir,
 N - objektide arv,
 P - intervalli pikkus,
 S_{Q1} - KI_1 -le eelnevate intervallide sageduste summa,
 S_{Q3} - KI_3 -le eelnevate intervallide sageduste summa,
 f_{Q1} - KI_1 sagedus,
 f_{Q3} - KI_3 sagedus;

- detsiilid - ..., mis jagavad rea 10-ks võrdse liikmete arvuga osaks
- tsentiilid - ..., mis jagavad rea 100-ks võrdse liikmete arvuga osaks

Kasuta

- *aritmeetilist keskmist*, kui
 - tulemused on jaotunud enam-vähem sümmeetriliselt keskmise ümber
 - ülesanne nõuab sellise keskmise määramist, mis on aluseks mingile teisele meetodile
 - nõutakse eri gruppide sama tunnuse mõõtmisel saadud tulemuste võrdlemist
 - on tarvis analüüsida populatsiooni, millest uuritav valim pärineb;
- *mediaani*, kui
 - ülesanne nõuab rea täpse keskpunkti leidmist
 - erandlikud vaatlusandmed moonutavad aritmeetilist keskmist
 - tegu on “veidrakujulise” jaotusega;
- *moodi*, kui
 - ülesandeks kiiresti määrata kesktendents
 - keskmise all mõeldakse “tüüpilist”.

Variatsioonikarakteristikud

- variatsiooniulatus

$$U = X_{\text{MAX}} - X_{\text{MIN}}$$

- keskmise lineaarhälve

$$d = (\sum |X_i - M|) / N, \quad i=1, \dots, N, \quad \text{kus}$$

X_i

- statistilise rea liige (variant),

N

- rea liikmete (objektide) arv;

$$d = (\sum |X_j - M| * f_j) / (\sum f_j), \quad j=1, \dots, K, \quad \text{kus}$$

f_j

- variandi X_j esinemissagedus,

K

- variantide (tunnuse skaalaväärtuste) arv.

- dispersioon

$$\sigma^2 = [\sum (X_i - M)^2] / N, \quad i=1, \dots, N;$$

$$\sigma^2 = [\sum (X_j - M)^2 * f_j] / (\sum f_j), \quad j=1, \dots, K;$$

$$\sigma^2 = p * q = p(1-p), \quad (0 \dots 0,25)$$

- standardhälve

$$\sigma = \sqrt{\sigma^2}$$

- kvartiilhälve

$$Q = (Q_3 - Q_1) / 2$$

vahemikku $M_E \pm Q$ langeb 50% objektidest

- variatsioonikoefitsiendid:

- variatsiooniulatuse alusel

$$V_R = U / M * (100\%)$$

- keskmise lineaarhälbe alusel

$$V_d = d / M * (100\%)$$

- standardhälbe alusel

$$V_\sigma = \sigma / M * (100\%)$$

Jaotuskõvera kuju hindamine

1. Asümmeetriakoeffitsient

$$r_3 = \mu_3 / \sigma^3, \quad \text{kus}$$
$$\mu_3 = [\sum (X_j - M)^3 \cdot f_j] / (\sum f_j),$$

$r_3 > 0$, positiivse (parempoolse) kaldega jaotus,

$r_3 < 0$, negatiivse (vasakpoolse) kaldega jaotus.

2. Ekstsess

$$E = (\mu_4 / \sigma^4) - 3, \quad \text{kus}$$
$$\mu_4 = [\sum (X_j - M)^4 \cdot f_j] / (\sum f_j),$$