Kvantitativne metode za analize v upravi

VZORČENJE

Prof.dr. Srečko Devjak
Dr. Jože Benčina

VZORČENJE

Spoznavanje:

- Osnovnih pojmov in pomena
- Načinov oblikovanja vzorcev,
- Zakonitosti množice velikih vzorcev,
- Metod ocenjevanja vrednosti parametrov množic
- Prizkušanje hipotez

OSNOVNI POJMI

- Opazovanje pojavov je lahko:
- ***popolno**, kadar opazujemo vse enote statistične množice,*
- ***delno**, kadar opazujemo le nekatere enote statistične množice*

- Ugotovimo lahko, da:
  - *je vzorec statistična množica,*
  - *so spremenljivke na vzorcu iste kot na osnovni populaciji,*
  - *na osnovni populaciji izračunane vrednosti parametrov so prave vrednosti parametrov,*
  - *na vzorcu izračunane vrednosti parametrov so ocene za vrednosti parametrov*

Označevanje

PARAMETER	OSNOVNA POPULACIJA	VZOREC
SPLOŠNO	**G**	g
ŠTEVILO ENOT	N	n
Vrednost spremenljivke	y	y
Aritmetična sredina	$\bar{Y}$ =M_y=μ	$\bar{y}$
Varianca	$\sigma_y^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N}$ $\sigma_y^2 = \frac{\sum_{j=1}^{k} f_j (y_j - \bar{Y})^2}{N}$	$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$ $s_y^2 = \frac{\sum_{j=1}^{k} f_j (y_j - \bar{y})^2}{n-1}$
Standardni odklon	σ_y	s_y

Postopki vzorčenja

- ***enostavno vzorčenje*** *- naključen način izbire enot v vzorec, poznavanje celotne množice ni potrebno,*

- ***stratificirano vzorčenje*** *- razdelitev množice v homogene dele - stratume, variabilnost v stratumih je majhna, v njih izvedemo slučajno vzorčenje,*

- ***vzorčenje v skupinicah*** *- izbiramo skupinice slučajno in te opazujemo v celoti,*

- ***vzorčenje v več stopnjah*** *- nadaljevanje vzorčenja v skupinicah, ko je možna delitev populacije v hierarhične skupine,*

- ***sistematično vzorčenje*** *- naključno je izbrana prva enota, nato se izbirajo enote po neki zakonitosti.*

***Primer** za prikaz značilnosti odnosov vrednosti aritmetične sredine : populacije, vzorcev in množice vzorcev*

- *Statistična množica ima 4 enote, spremenljivka y ima vrednosti:2, 3, 7, 8.*

i	**y_i**	**$y_i - \bar{Y}$**	**$(y_i - \bar{Y})^2$**
y_1	2	-3	9
y_2	3	-2	4
y_3	7	2	4
y_4	8	3	9
SKUPAJ	20	0	26

Parametri populacije so:
N = 4

$$\bar{Y} = \frac{20}{4} = 5$$

$$\sigma_y^2 = \frac{26}{4} = 6{,}50$$

$$\sigma_y = \sqrt{6{,}5} = 2{,}55$$

Parametri vzorca

- **Izračunajmo parametre enega od vzorcev** *z dvema enotama:*
- *n = 2 naj bo:* $y_1=2$ *in* $y_2=7$

i	yi	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	2	-2,5	6,25
2	7	2,5	6,25
SKUPAJ	**9**	**0**	**12,5**

$$\bar{y} = \frac{9}{2} = 4{,}5$$

$$s_y^2 = \frac{12{,}5}{2-1} = 12{,}5$$

$$s_y = \sqrt{12.5} = 3{,}53$$

Število vseh vzorcev:

$$\binom{N}{n} = \binom{4}{2} = \frac{1*2*3*4}{(1*2)*(1*2)} = 6$$

Vrednosti aritmetičnih sredin za vzorce:

	2	**3**	**7**	**8**
2	xxxx	2,5	4,5	5
3		xxxx	5	5,5
7			xxxx	7,5
8				xxxx

Parametri množice vseh vzorcev:

$\overline{y}_j$	f_j	$f_j * \overline{y}_j$	$\overline{y}_j - \overline{Y}_{\bar{y}}$	$f_j * (\overline{y}_j - \overline{Y}_{\bar{y}})^2$
2,5	1	2,5	-2,5	6,25
4,5	1	4,5	-0,5	0,25
5	2	10	0	0
5,5	1	5,5	0,5	0,25
7,5	1	7,5	2,5	6,25
	6	30		13

Aritmetična sredina:

$$\overline{Y}_{\bar{y}} = \frac{30}{6} = 5$$

$$\sigma^2_{\bar{y}} = \frac{\sum f_j (\overline{y}_j - \overline{Y}_{\bar{y}})^2}{\sum f_j} = \frac{13}{6} = 2{,}17$$

$$\sigma_{\bar{y}} = \sqrt{\sigma^2_{\bar{y}}} = 1{,}47$$

<u>Ugotovitve:</u>

•pri vzorčenju nastopajo tri statistične množice:

..osnovna množica,

.. vzorec,

.. množica vseh vzorcev

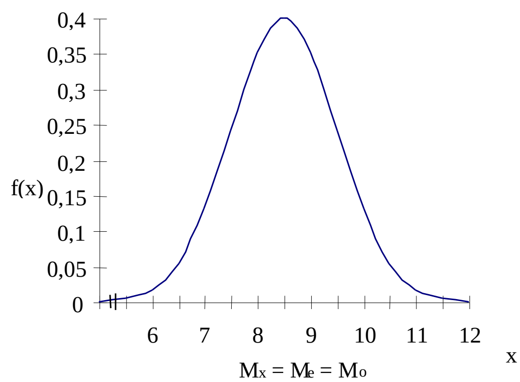
•aritmetična sredina osnovne populacije in aritmetična sredina množice vseh vzorcev sta enaki,

•pri velikih vzorcih (n> 50) je množica vseh vzorcev porazdeljena normalno.

Množica vseh vzorcev – normalna porazdelitev

Zakonitost normalne porazdelitve

Razmik	Delež vseh enot populacije %
$M_x \pm \sigma$	68,3
$M_x \pm 2\sigma$	95,4
$M_x \pm 3\sigma$	99,7



Zakonitost porazdelitve množice vseh vzorcev - veliki vzorci

Delež vseh vzorcev	Interval
68,28 %	$\overline{Y}_{\bar{y}} - SE_{\bar{y}} < \overline{Y} < \overline{Y}_{\bar{y}} + SE_{\bar{y}}$
95,45%	$\overline{Y}_{\bar{y}} - 2SE_{\bar{y}} < \overline{Y} < \overline{Y}_{\bar{y}} + 2SE_{\bar{y}}$
99,73%	$\overline{Y}_{\bar{y}} - 3SE_{\bar{y}} < \overline{Y} < \overline{Y}_{\bar{y}} + 3SE_{\bar{y}}$

$SE_{\bar{y}}$ *standardni odklon vzorčnih vrednosti aritmetične sredine v množici vzorcev.*

2. 3 Točkovna in intervalna ocena parametrov statistične množice

- *Točkovna ocena parametra G* - vrednost parametra izračunana na vzorcu, označena z "*g*"

- ***Intervalna ocena***,
- *Odvisna je od stopnje tveganja:*
  - **Tveganje** pomeni verjetnost, da se dogodek, ki ga pričakujemo, ne zgodi.

- Vrste intervalnih ocen (trditev):
  - **dvostranske** (kadar opredlejujemo interval možnih vrednosti) in
  - **enostranske** (kadar določamo vrednost od katere je parameter večji ali manjši).

- Do **intervalne ocene parametra *G* statistične množice** pridemo na osnovi tveganja z določanjem intervala, na katerem se lahko vrednost parametra *G* nahaja.

Intervalna ocena- splošni obrazec

Intervalno oceno parametra *G* statistične množice določimo:

- na osnovi tveganja in
- ocene intervala, na katerem se vrednost G lahko nahaja.

Splošni obrazec ima obliko:

$$g - d_g < G < g + d_g$$

pri tem pomeni:

G – pravo vrednost parametra G
***g**-točkovno oceno izračunano na vzorcu,*

***d_g** - odklon zaupanja za parameter G na množici vseh vzorcev, določen glede na porazdelitev množice vzorcev in stopnjo tveganja*

Odklon zaupanja d_g

Odklon zaupanja d_g izračunamo po obrazcu:

$$z_\alpha . SE(g) = d_g$$

*G - **parameter**, za katerega določamo intervalno oceno,*

*g - **ocena (točkovna) parametra** G na vzorcu ,*

*$z\alpha$ - **standardizirani odklon** pri stopnji tveganja α,*

*SE(g) **- standardna napaka** ocene parametra G, (standardni odklon na množici vseh vzorcev).*

Vrednosti $z\alpha$ in ocene za odklon zaupanja d_g

Stopnja tveganja α	Vrednost z_α pri dvostranski trditvi	Vrednost z_α pri enostranski trditvi
0,05	1,96	1,64
0,01	2,58	2,32
0,005	3,29	3,09

Oceno za dklon zaupanja d_g:

$$z_\alpha . se(g) = d_g$$

se(g) – je približek za SE(g), za vsak parameter G obstajajo obrazci, s katerimi se izračunavajo **približki se(g)**

se(g) so pogosto množeni s popravki za končnost:

$$f = \sqrt{\frac{N-n}{N-1}}$$

Intervalna ocena in velikost vzorca:

Intervalna ocena parametra:

$$g - z_\alpha se(g) < G < g + z_\alpha se(g)$$

Velikost vzorca, *n*:

Velikost vzorca izhaja iz želene natančnosti ocene:

-naj bo d_g poznan, potem: $z_\alpha se(g) = d_g$

- *naj bo delež* $d_g/g = p$ *poznan, potem bo:*

$$\frac{z_\alpha . se(g)}{g} = p$$

Ocenjevanje aritmetične sredine

Točkovna ocena:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Intervalna ocena:

$$\bar{y} - z_\alpha \frac{s_y}{\sqrt{n}} < \bar{Y} < \bar{y} + z_\alpha \frac{s_y}{\sqrt{n}}$$

$$se(\bar{y}) = \frac{s_y}{\sqrt{n}}$$

$$s_y^2 = \frac{\sum f_j * (y_j - \bar{y})^2}{n-1}$$

$$s_{y,pop.}^2 = s_y^2 - \frac{d^2}{12}$$

Velikost vzorca:

$$d_{\bar{y}} = z_\alpha \frac{s_y}{\sqrt{n}}$$

$$n = \left(\frac{z_\alpha s}{d_{\bar{y}}} \right)^2$$

Ocenjevanje vsote (totala, agregata)

Točkovna ocena:

$$\boxed{\overleftrightarrow{Y} = N\bar{y}}$$

Intervalna ocena:

$$\boxed{\overleftrightarrow{Y} - z_{\alpha}.N\frac{s_y}{\sqrt{n}} < Y < \overleftrightarrow{Y} + z_{\alpha}.N\frac{s_y}{\sqrt{n}}}$$

$$se(\overleftrightarrow{Y}) = N\frac{s_y}{\sqrt{n}}$$

Velikost vzorca:

$$d_{\overleftrightarrow{Y}} = z_{\alpha}N\frac{s_y}{\sqrt{n}}$$

$$\boxed{n = \left(\frac{z_{\alpha}Ns_y}{d_{\overleftrightarrow{Y}}}\right)^2}$$

Ocenjevanje deleža

Delež *P:*

$$P = \frac{N_a}{N}$$

Točkovna ocena deleža *p*:

$$\boxed{p = \frac{n_a}{n}}$$

Intervalna ocena deleža:

$$se(p) = \sqrt{\frac{p(1-p)}{n}}$$

$$\boxed{p - z_\alpha \sqrt{\frac{p(1-p)}{n}} < P < p + z_\alpha \sqrt{\frac{p(1-p)}{n}}}$$

Velikost vzorca:

$$d_p = z_\alpha se(p)$$

$$d_p = z_\alpha \sqrt{\frac{p(1-p)}{n}}$$

$$\boxed{n = \frac{z_\alpha^2 p(1-p)}{d_p^2}}$$

Ocenjevanje razlike med aritmetičnima sredinama: $\overline{Y}_1 - \overline{Y}_2$

Točkovna ocena razlike: $\boxed{\overline{y}_1 - \overline{y}_2}$

Intervalna ocena razlike aritmetičnih sredin:

$$se(\overline{y}_1 - \overline{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\boxed{(\overline{y}_1 - \overline{y}_2) - z_\alpha . se(\overline{y}_1 - \overline{y}_2) < (\overline{Y}_1 - \overline{Y}_2) < (\overline{y}_1 - \overline{y}_2)' + z_\alpha . se(\overline{y}_1 - \overline{y}_2)}$$

Uporaba: ocenjevanje različnosti

2.4 Statistično preizkušanje domnev

- Izhodišče
- Kakšen je odnos med hipotetično parametra G_0 in vrednostjo parametra g na vzorcu?
- Velja zveza:

$$g - z_{\alpha} se(g) < G < g + z_{\alpha} se(g)$$

Kdaj je $G = G_0$?

$$g - z_{\alpha} se(g) = G_o$$

Kadar bo *z* enak:

$$z = \frac{g - G_0}{se(g)}$$