

MALI VZORCI

problem ocen iz malih zorcev
rešitev: t-porazdelitev

as. dr. Nino RODE
prof. dr. Blaž MESEC

1

VELIKI IN MALI VZORCI

- Vzorčne statistike se porazdeljujejo normalno okoli parametra populacije

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- Problem: **ne poznamo** standardnega odklona populacije

- Pri malih vzorcih standardni odklon vzorca **NI** dovolj zanesljiva ocena standardnega odklona populacije

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \neq \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

2

MALI VZORCI t-porazdelitev

- Pri malih vzorcih je porazdelitev statistike

$$t_n = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

drugačna: odvisna je samo od velikosti vzorca, natančneje od stopinj prostosti

3

MALI VZORCI stopinje prostosti

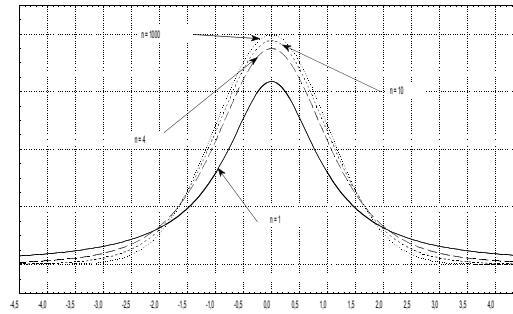
- Vsako določeno vrednost statistike lahko dobimo iz različnih vzorcev

$$\bar{x}[1; 2; 3; 4; 5] = \bar{x}[3; 2; 3; 2; 5] = \bar{x}[100; 25; 10; 15; -135] = 3$$

- Ti vzorci imajo skupno lastnost: enako statistiko
- Možnost izbire v teh vzorcih je zato manjša, kot če ne bi imeli skupne vrednosti statistike, to pa večja napako vzorčenja

4

MALI VZORCI t - porazdelitev



MALI VZORCI standardne napake različnih statistik

- Aritmetična sredina

$$s_{\bar{x}} = \frac{s}{\sqrt{n-1}}$$

- Mediana

$$s_{Me} = \frac{1,2535 \cdot s}{\sqrt{n-1}}$$

- Odstotek

$$s_p = \sqrt{\frac{P \cdot (100-P)}{n-1}}$$

- Koeficient korelacije

$$s_r = \frac{(1-r)}{\sqrt{n-2}}$$

- Razlika aritmetičnih sredin (neodvisni vzorci)

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 + n_2) \cdot ((n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2)}{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}}$$

6

Ocenjevanje parametrov mali vzorci

- so jih 120 (n) anketirali glede poučenosti o aidsu
- Rezultati:
 - Povprečno število točk na testu = 43 točk
 - s = 5,1 točke
- Določi meje intervala zaupanja v aritmetično sredino pri stopnji tveganja 0,05

7

Ocenjevanje parametrov

$$s_x = \frac{s}{\sqrt{n-1}} = \frac{5,1}{\sqrt{119}} = 0,4675$$

$$\bar{x} - t_{0,05;119} \cdot s_x = 43 - 1,9801 \cdot 0,4675 = 43 - 0,93 = 42,07$$

$$\bar{x} + t_{0,05;119} \cdot s_x = 43 + 1,9801 \cdot 0,4675 = 43 + 0,93 = 43,93$$

S 95% zaupanjem lahko trdimo, da je aritmetična sredina populacije med 42,02 in 43,93

8

Ocenjevanje parametrov

- Pri istih rezultatih in vzorcu 20 študentk/ov

$$s_x = \frac{s}{\sqrt{n-1}} = \frac{5,1}{\sqrt{19}} = 1,1700$$

$$\bar{x} - t_{0,05;19} \cdot s_x = 43 - 2,0930 \cdot 1,1700 = 43 - 2,449 = 40,55$$

$$\bar{x} + t_{0,05;19} \cdot s_x = 43 + 2,0930 \cdot 1,1700 = 43 + 2,449 = 45,45$$

S 95% zaupanjem lahko trdimo, da je aritmetična sredina populacije med 40,55 in 45,45

9

Sklepanje na podlagi malih vzorcev:

Primer aritmetičnih sredin

- Otroci, stari 2 do 4 leta
- 9 deklic in 6 dečkov
- Pri dečkih so opazili povprečno 6 agresivnih dejanj (varianca 4), pri deklicah pa 1,9 (varianca 5,4)
- Pri stopnji tveganja 0,05 preizkusimo hipotezo, da med dečki in deklicami ni razlik v povprečnem številu agresivnih dejanj

10

Sklepanje na podlagi malih vzorcev:

Primer aritmetičnih sredin

- Poglejmo izračune:
- Razlika med povprečnima ocenama je:

$$\bar{x}_1 - \bar{x}_2 = 6 - 1,9 = 4,1$$

- Standardna napaka za razliko med aritmetičnimi sredinami je:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 + n_2) \cdot ((n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2)}{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}}$$

11

Sklepanje na podlagi malih vzorcev:

Primer aritmetičnih sredin

- Izračun Standardne napake:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 + n_2) \cdot ((n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2)}{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(6 + 9) \cdot ((6 - 1) \cdot s_1^2 + (9 - 1) \cdot s_2^2)}{6 \cdot 9 \cdot (6 + 9 - 2)}} = \sqrt{\frac{15 \cdot (5 \cdot 4 + 8 \cdot 5,4)}{6 \cdot 9 \cdot 13}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(15) \cdot (20 + 43,2)}{702}} = \sqrt{\frac{(15) \cdot (63,2)}{702}} = \sqrt{\frac{948}{702}} = \sqrt{1,35} = 1,16$$

12

Sklepanje na podlagi malih vzorcev:

Primer aritmetičnih sredin

- Izračun t-vrednosti za razliko:

$$t_{sp;\alpha} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{6 - 1,9}{1,16} = \frac{4,1}{1,16} = 3,53$$

- Primerjava izračuna s tabelirano vrednostjo:

$$t_{iz\ podatkov} = 3,53 > 2,160 = t_{13;0,05}$$

- Odgovor: **z manj kot 5%** (tudi manj kot 1%) tveganjem **lahko zavrnemo** ničelno hipotezo, da med dečki in deklicami ni razlik v povprečnem številu agresivnih dejanj

13

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Redni študentje ($n = 131$) in specializanti ($n = 34$) socialnega dela
- Med rednimi jih je 14,5 odstotka izjavilo, da so dobro poučeni o AIDS-u, med specializanti pa 20,6
- Preizkusi hipotezo, da med rednimi študenti in specializanti ni razlik v odstotku tistih, ki menijo, da so dobro poučeni o aidsu

14

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Razlika med odstotkoma je:

$$P_1 - P_2 = 20,6 - 14,5 = 6,1$$

- Odstotki se v vzorcih **NE porazdeljujejo po normalni porazdelitvi**, zato pri majhnih vzorcih ne moremo preprosto uporabiti enake logike za izračun standardne napake za razliko, kot pri aritmetičnih sredinah

15

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Uporabimo metodo skupnega odstotka
 - Če v populaciji med odstotkoma ni razlik (H_0), sta oba odstotka iz vzorca oceni istega odstotka
 - Skupni odstotek dobimo po formuli:

$$P_s = \frac{P_1 \cdot n_1 + P_2 \cdot n_2}{n_1 + n_2} = \frac{20,6 \cdot 34 + 14,5 \cdot 131}{34 + 131} = \frac{2599,9}{165} = 15,76$$

16

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Standardna napaka za razliko med vzorčnima ocenama na podlagi skupnega odstotka je:

$$s_{P_1 - P_2} = \sqrt{P_s \cdot (100 - P_s) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{1327,41 \cdot \left(\frac{1}{34} + \frac{1}{131}\right)}$$

$$s_{P_1 - P_2} = \sqrt{1327,41 \cdot (0,029 + 0,008)} = \sqrt{1327,41 \cdot 0,037} = \sqrt{49,17} = 7,01$$

17

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Razliko med odstotkoma v vzorcu primerjamo s to standardno napako te razlike

$$t_{sp; \alpha} = \frac{P_1 - P_2}{s_{P_1 - P_2}} = \frac{6,1}{7,01} = 0,87$$

- Ugotovimo, da je izračunani t **premajhen**, da bi lahko zavrnilo ničelno hipotezo

$$t_{iz\ podatkov} = 0,87 < 1,975 = t_{163; 0,05}$$

- Odgovor: Na podlagi danih podatkov ne moremo trditi, da je razlika med odstotki statistično značilna na nivoju 0,05 (tudi pri 0,35 ne)

18

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Izračun Standardne napake:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 + n_2) \cdot ((n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2)}{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(6 + 9) \cdot ((6 - 1) \cdot s_1^2 + (9 - 1) \cdot s_2^2)}{6 \cdot 9 \cdot (6 + 9 - 2)}} = \sqrt{\frac{15 \cdot (5 \cdot 4 + 8 \cdot 5,4)}{6 \cdot 9 \cdot 13}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(15) \cdot (20 + 43,2)}{702}} = \sqrt{\frac{(15) \cdot (63,2)}{702}} = \sqrt{\frac{948}{702}} = \sqrt{1,35} = 1,16$$

19

Sklepanje na podlagi malih vzorcev:

Primer odstotkov

- Izračun t-vrednosti za razliko:

$$t_{sp; \alpha} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{6 - 1,9}{1,16} = \frac{4,1}{1,16} = 3,53$$

- Primerjava izračuna s tabelirano vrednostjo:

$$t_{iz\ podatkov} = 3,53 > 2,160 = t_{13; 0,05}$$

Odgovor: **z manj kot 5%** (tudi manj kot 1%) tveganjem **lahko zavrne** ničelno hipotezo, da med dečki in deklicami ni razlik v povprečnem številu agresivnih dejanj

20