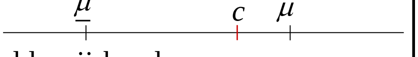
Biomeetria

Populatsiooni keskväärtuse võrdlemine konstandiga.
Kahe populatsiooni keskväärtuste võrdlemine

Populatsiooni keskväärtuse võrdlemine konstandiga

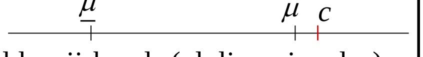
<u>Usalduspiirid</u>

$H_0 : \mu = c$



võetakse vastu siis, kui c kuulub usalduspiirkonda

$H_1 : \mu \neq c$



on tõestatud siis, kui c ei kuulu usalduspiirkonda (olulisusnivool α)

<u>Normaaljaotuse eeldusel *t*-test</u>

Teststatistik: $t = \frac{\overline{x} - c}{s}\sqrt{n} \underset{H_0}{\sim} t_{n-1}$

$$|t| \geq t_{1-\alpha/2,n-1} \quad <=> \quad p \leq \alpha \quad \Rightarrow H_1 : \mu \neq c$$
$$|t| < t_{1-\alpha/2,n-1} \quad <=> \quad p > \alpha \quad \Rightarrow H_0 : \mu = c$$

<u>Suurte valimite (*n*>60) korral *z*-test</u>

Teststatistik:

$$Z = \frac{\overline{x} - c}{s}\sqrt{n} \underset{H_0}{\sim} N(0,1)$$

$$|Z| \geq z_{1-\alpha/2} \quad <=> \quad p \leq \alpha \quad \Rightarrow H_1 : \mu \neq c$$
$$|Z| < z_{1-\alpha/2} \quad <=> \quad p > \alpha \quad \Rightarrow H_0 : \mu = c$$

Populatsiooni keskväärtuse võrdlemine konstandiga

Näide. Kümnel herefordi tõugu veisel Eestis kulus kurjast haigusest paranemiseks vastavalt 3, 5, 4, 6, 2, 6, 5, 6, 5 ja 3 päeva. Teades, et sama tõugu veised Prantsusmaal tervenevad keskmiselt 5,4 päevaga, kontrollida hüpoteesi tervenemisaegade erinevusest Eestis ja Prantsusmaal.

$H_0: \mu = 5{,}4$
$H_1: \mu \neq 5{,}4$

või

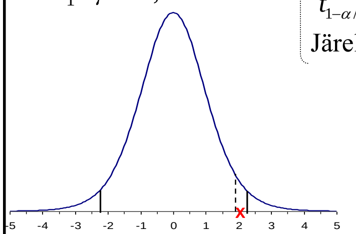
$H_0: \mu \geq 5{,}4$
$H_1: \mu < 5{,}4$

$n = 10;\ \bar{x} = 4{,}5;\ s \approx 1{,}43;\ \alpha = 0{,}05$

Teststatistik: $|t| = \left|\dfrac{\bar{x} - 5{,}4}{s}\sqrt{n}\right| = \left|\dfrac{4{,}5 - 5{,}4}{1{,}43}\sqrt{10}\right| = |-1{,}985| = 1{,}985$

Teststatistiku kriitiline väärtus (kahepoolne hüpotees):
$t_{1-\alpha/2;n-1} = t_{0,975;9} = 2{,}26$
Järeldus: $|t| = 1{,}985 < 2{,}26 = t_{0,975;9} \Rightarrow H_0: \mu = 5{,}4$

Teststatistiku kriitiline väärtus (ühepoolne hüpotees):
$t_{1-\alpha;(n-1)} = t_{0,95;9} = 1{,}83$
Järeldus: $t_{0,95;9} = 1{,}83 < 1{,}985 = |t| \Rightarrow H_1: \mu < 5{,}4$



Näiteks *MS Excelis* funktsioon TDIST(t;n-1;2)

Arvuti abil saab leida ka täpse tõenäosuse teststatistiku väärtuse $|t| = 1{,}985$ saamiseks eeldusel, et kehtib H_0:
$p = 0{,}0784$ (2-poolne hüp.); $p = 0{,}0392$ (1-poolne hüp.)

Kahe grupi keskmiste võrdlus, kolm *t*-testi – mis ja millal?

Kahe populatsiooni keskmiste võrdlus – *t*-test

Eeldus: $X \sim N(\mu;\sigma^2)$
[või suur (>60) n]

Sõltuvad vaatlused (paariviisiline võrdlus)

$H_0: \mu_1 = \mu_2 \quad \Leftrightarrow \quad H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 \neq \mu_2 \quad \quad H_1: \mu_1 - \mu_2 \neq 0$

Teststatistik: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_{1-2}}\sqrt{n} \underset{H_0}{\sim} t_{n-1}$

Sõltumatud vaatlused

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

Kahe populatsiooni dispersioonide võrdlus – *F*-test

$H_0: \sigma_1^2 = \sigma_2^2$
$H_1: \sigma_1^2 \neq \sigma_2^2$

Teststatistik:
$F = \dfrac{s_1^2}{s_2^2} \underset{H_0}{\sim} F_{n_1-1,n_2-1}$

Võrdsed dispersioonid

Teststatistik:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \underset{H_0}{\sim} t_{n_1+n_2-2},$$

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

Mittevõrdsed dispersioonid

Teststatistik: $t = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \underset{H_0}{\sim} t_\nu,$

$$\nu = \left\{(s_1^2/n_1 + s_2^2/n_2)^2 \bigg/ \left[\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}\right]\right\} - 2$$

Sõltuvad vaatlused (paariviisiline võrdlus)

$H_0: \mu_1 = \mu_2$ $\Leftrightarrow$ $H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$ $\quad$ $H_1: \mu_1 - \mu_2 \neq 0$

Teststatistik: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_{1-2}}\sqrt{n} \underset{H_0}{\sim} t_{n-1}$

Näide. Soovitakse uurida, kas lehmade ööpäevane piimatoodang langes pärast seda, kui neile lõpetati juurvilja söötmine (olulisuse nivool $\alpha = 0{,}05$).

Lehm	Piim (kg/ööpäevas) juurviljaga (x_1)	Piim (kg/ööpäevas) juurviljata (x_2)	Piimatoodangu muutused (d)
1	10	11	-1
2	8	7	1
3	11	10	1
4	10	10	0
5	7	6	1
6	8	5	3
7	10	6	4
8	9	4	5
9	8	6	2
10	10	8	2

Kontrollime hüpoteesi piimatoodangu muutuse (μ_d) kohta, s.t. H_0: $\mu_d = 0$

H_1: $\mu_d \neq 0$

$n = 10;\ \bar{d} = 1{,}8;\ s_d = 1{,}81$

Andmetest arvutatud teststatistik: $t = \bar{d}\big/ s_d \sqrt{n} \approx 3{,}14$

Teststatistiku kriitiline väärtus (kahepoolne hüpotees):

$$t_{1-\alpha,n-1} = t_{0{,}975;9} = 2{,}26$$

Järeldus: $t = 3{,}14 > 2{,}26 = t_{1-\alpha,n-1}$

$\Rightarrow H_1: \mu_d \neq 0$ $(p = 0{,}012)$

Sõltumatud vaatlused

Näide. Ettevõttes võrreldi ametiühingusse kuuluvate ja sinna mittekuuluvate töötajate puudumisi aasta jooksul. Viiskümmend vaadeldud ametiühinguliiget puudusid keskmiselt 9,3 päeva, kusjuures standardhälve oli 3,1 päeva. Ametiühingusse mittekuulujad, keda oli 45, puudusid igaüks keskmiselt 8,7 päeva standardhälbega 2,3 päeva. Kontrollida hüpoteesi ettevõtte töötajate keskmiselt puudutud päevade arvu sõltuvusest ametiühingusse kuulumisest olulisuse nivool $\alpha = 0{,}05$.

$n_1 = 50;\ n_2 = 45$

$\bar{x}_1 = 9{,}1;\ \bar{x}_2 = 8{,}7$

$s_1 = 3{,}1;\ s_2 = 2{,}3$

$\alpha = 0{,}05$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

(1) $H_0:\ \sigma_1^2 = \sigma_2^2$

$H_1:\ \sigma_1^2 \neq \sigma_2^2$

Teststatistik: $F = s_1^2 \big/ s_2^2 = 1{,}817$

Teststatistiku kriitiline väärtus:

$$F_{1-\alpha/2;n_1-1;n_2-1} = F_{0{,}975;\,49;\,44} = 1{,}799$$

Järeldus: $F = 1{,}817 > 1{,}799 = F_{0{,}975;\,49;\,44} = F_{1-\alpha/2;n_1-1;n_2-1}$

$\Rightarrow H_1: \sigma_1^2 \neq \sigma_2^2$ $\boxed{(p = 0{,}023)}$

(2) Teststatistik: $t = |\bar{x}_1 - \bar{x}_2| \Big/ \sqrt{s_1^2/n_1 + s_2^2/n_2} = 0{,}969$

Teststatistiku kriitiline väärtus: $t_{1-\alpha/2,\,\nu} = t_{0{,}975;95} = 1{,}985$

Järeldus: $t = 0{,}969 < 1{,}985 = t_{0{,}975;45}$ $\Rightarrow H_0: \mu_1 = \mu_2$ $\boxed{(p = 0{,}335)}$

Märkus: eeldanuks me siiski, et $\sigma_1^2 = \sigma_2^2$, jõudnuks me peale ühise dispersiooni $s^2 = 7{,}566$ ja teststatistiku $t = 0{,}708$ arvutamist samale järeldusele: $\mu_1 = \mu_2$, aga olulisustõenäosus olnuks pisut suurem ($p = 0{,}481$).

Kolm *t*-testi – mis seal vahet on?

Üldine eeldus: $X \sim N(\mu;\sigma^2)$

Andmete olemus		Lisaeeldused	p: $\mu_1 > \mu_2$	p: $\mu_1 \neq \mu_2$
Sõltumatud vaatlused	mittevõrdne varieeruvus	–	0,026	0,053
Sõltumatud vaatlused	võrdne varieeruvus	$\sigma_1^2 = \sigma_2^2$	0,024	0,048
Sõltuvad vaatlused		$H_0: \mu_1 = \mu_2$ $\Updownarrow$ $H_0: \mu_1 - \mu_2 = 0$	0,006	0,012

Piim (kg/ööpäevas)	
juurviljaga (x_1)	juurviljata (x_2)
10	11
8	7
11	10
10	10
7	6
8	5
10	6
9	4
8	6
10	8

Mida enam on lihtsustavaid eelduseid e mida kitsamalt on situatsioon (andmed) enne juhuslikkuse mängu toomist piiritletud, seda suurem on statistilise testi võimsus (seda väiksem erinevus on vajalik populatsioonide erinevuse tõestamiseks e seda väiksem on uurija eksimistõenäosus kummutades nullhüpoteesi)!