TIIVISTELMÄ PÄÄTTELYTILANTEISTA JA -MENETELMISTÄ ESIMERKKEINEEN

Arvioitavana	Otossuure	Luottamusväli (kaavanumero)	Testi (kaavanumero)	Esimerkki
μ	otoskeskiarvo	(4.1) tai (4.2)	(5.1) tai (5.2)	1, 2
π	prosenttiosuus otoksessa	(4.3)	(5.3)	3
$\mu_1 - \mu_2$	otoskeskiarvojen erotus	(4.4) tai (4.5)	(5.4) tai (5.5)	4, 5

Esim. 1 Rattaan pyörimisajan (s) oletetaan noudattavan normaalijakaumaa odotusarvona 150 s ja hajontana 10 s. Rasvataan laakereita. Halutaan selvittää, onko rasvaus vaikuttanut keskimääräiseen pyörimisaikaan. Mitataan rasvauksen jälkeen 25 kerran pyörimisajat, joiden keskiarvoksi saadaan 155 s.

Tapa 1

Jos pyörimisaika ei ole muuttunut, niin $\bar{X} \sim N(150, \frac{10^2}{25})$. Tällöin todennäköisyys sille, että otoskeskiarvoksi saadaan suurempi kuin 155 on $P(\bar{X} > 155) = 1 - P(\bar{X} \leq 155) = 1 - \Phi\left(\frac{155-150}{2}\right) = 1 - \Phi(2{,}5) = 0{,}0062$. Tämä harvinaista, joten päätellään rasvauksen pidentäneen keskimääräistä pyörimisaikaa.

Jos ei oleteta pyörimisajan noudattavan normaalijakaumaa, niin $\bar{X} \sim N(150, \frac{10^2}{25})$, likimain.

Tapa 2

Kaavan (4.1) mukaisesi 95 %:n luottamusväli odotusarvolle $155 \pm 1{,}96 \cdot 10/\sqrt{25}$. Koska 150 ei kuulu luottamusvälille, päätellään keskimääräisen pyörimisajan muuttuneen.

Tapa 3

H_0: μ = 150
H_1: μ > 150
Jos H_0 tosi, niin $Z = \frac{\bar{X}-150}{10/\sqrt{25}} \sim N(0, 1)$, kaava (5.1). Saadaan $z_{hav.}$ = 2,5 ja p-arvo on 1 - $\Phi(2{,}5)$= 0,0062. Hylätään nollahypoteesi (esim. 1 %:n riskitasolla) ja päätellään rasvauksen pidentäneen keskimääräistä pyörimisaikaa. Jos H_1 ≠150, niin p-arvo on 0,0124. Tällöin H_0 hylätään, jos valitaan tätä suurempi riskitaso.

Esim. 2 Sokerin pussituskoneen pitäisi tuottaa kilon pusseja. Tutkitaan koneen toimivuutta ja valitaan koneen tuottamista pusseista satunnaisesti 20 ja saadaan niiden keskipainoksi 1002 g ja keskihajonnaksi 3,4 g. Toimiiko pussituskone oikein?

Tapa 1

Kaavan (4.2) mukaisesti 95 %:n luottamusväli odotusarvolle $1002 \pm 2{,}093 \cdot 3{,}4/\sqrt{20}$. Koska 1000 ei kuulu luottamusvälille, päätellään koneen toimivan väärin (ei tuota keskimäärin kilon pusseja).

Tapa 2

H_0: μ = 1000
H_1: μ ≠ 1000
Jos H_0 tosi, niin $t = \frac{\bar{X}-1000}{s/\sqrt{n}} \sim t_{n-1}$, kaava (5.2). Saadaan $t_{hav.}$ = 2,631 > 2,093 = $t_{0,025,19.}$ Hylätään nollahypoteesi 5 %:n riskitasolla. Päätellään, että kone ei tuota keskimäärin kilon pusseja.

Esim. 3 Aikaisempien tutkimusten perusteella 10 % kahvin ostajista valitsi kahvin hinnan perusteella. Haluttiin selvittää, onko ostokäyttäytymisessä tapahtunut muutosta. Kysyttiin valinnan perustetta 266 ostajalta, joista 38 teki ostopäätöksen hinnan perusteella. Onko tapahtunut muutosta?

Tapa 1

Jos ei ole tapahtunut muutosta, niin X = otoksessa valintansa hinnan perusteella tekevien lukumäärä ~ Bin(266, 0,10) ja E(X) = 26,6, Var(X) = 23,94. X noudattaa likimain normaalijakaumaa parametrin 26,6, ja 23,94. Tällöin P(X ≥ 38) = 1 − P(X≤ 37) ≈ $1 - \Phi(\frac{37-26{,}2}{\sqrt{23{,}94}})$= 1 - $\Phi(2{,}13) = 1 - 0{,}9834 = 0{,}0166$. Jos tätä todennäköisyyttä pitää pienenä, niin päättelee muutosta tapahtuneen.

Tapa 2

Kuten edellä, mutta tarkastellaan prosenttiosuutta p. Jos muutosta ei tapahtunut, niin $p \sim N\left(10, \frac{10\cdot 90}{266}\right)$, likimain. P(p ≥100·38/266) ≈ $1 - \Phi\left(\frac{14{,}29-10}{\sqrt{10\cdot 90/266}}\right) = 1 - \Phi(2{,}33) = 0{,}0099$. Jos muutosta ei olisi tapahtunut, niin olisi harvinaista saada otos, jossa prosenttiosuus suurempi kuin 100·38/266. Päätellään muutosta tapahtuneen.

Tapa 3

Kaavan (4.3) mukaisesti 95 %:n luottamusväli prosenttiosuudelle $14{,}29 \pm 1{,}96\sqrt{\frac{14{,}29(100-14{,}29)}{266}}$. Koska 10 ei kuulu luottamusvälille, päätellään muutosta tapahtuneen.

Tapa 4

H_0: π = 10
H_1: π ≠ 10
Jos H_0 tosi, niin $Z = \frac{p-10}{\sqrt{10\cdot 90/n}} \sim N(0,1), likimain$. Saadaan $z_{hav.}$ = 2,33 > 1,96 = $z_{0,025.}$ Hylätään nollahypoteesi 5 %:n riskitasolla. Päätellään, että muutosta on tapahtunut. Pienin ristitaso, jolla H_0 voidaan hylätä, on 2·0,0099, yksisuuntaisessa testissä 0,0099.

Esim. 4 Koneiden A ja B pitäisi valmistaa keskimäärin samanmittaisia tankoja. Molempien koneiden tuottamien tankojen pituuksissa X ja Y (cm) on jonkin verran vaihtelua, jota voidaan luonnehtia normaalijakaumalla, jonka varianssi on 0,20 cm^2. Laadunvalvonnassa seurataan koneiden toimintaa ja valitaan satunnaisesti koneen A tuotannosta 20 ja koneen B tuotannosta 10 tankoa. Koneen A tuotannosta valittujen tankojen keskipituus on 40,0 cm ja koneelta B valittujen 39,5 cm. Tuottavatko koneet keskimäärin samanmittaisia tankoja?

Tapa 1

$$\overline{X} - \overline{Y} \sim N\left(0, \frac{0{,}2}{20} + \frac{0{,}2}{10}\right), jos\ tuottavat\ samanmittaisia$$

$$\begin{aligned}
P(|\overline{X} - \overline{Y}| > 0{,}5) &= 1 - P(-0{,}5 \le \overline{X} - \overline{Y} \le 0{,}5) \\
&= 1 - \left(\Phi\left(\frac{0{,}5-0}{\sqrt{0{,}03}}\right) - \Phi\left(\frac{-0{,}5-0}{\sqrt{0{,}03}}\right)\right) \\
&= 1 - (\Phi(2{,}89) - \Phi(-2{,}89)) = 2 - 2\Phi(2{,}89) = 0{,}0038
\end{aligned}$$

Eivät tuota keskimäärin samanmittaisia tankoja. Jos tuottaisivat, niin olisi harvinaista saada otokset, joiden keskiarvojen erotuksen itseisarvo olisi suurempi kuin 0,5 cm.

Tapa 2

Kaavan (4.4) mukaisesi 95 %:n luottamusväli odotusarvojen erotukselle $40 - 39{,}5 \pm 1{,}96\sqrt{\frac{0{,}2}{20} + \frac{0{,}2}{10}}$. Tämä ei sisällä nollaa. Päätellään, että koneet eivät tuota keskimäärin samanmittaisia tankoja.

Tapa 3

H_0: $\mu_A = \mu_B$
H_1: $\mu_A \neq \mu_B$
Kaavan (5.4) perusteella saadaan $z_{hav.} = \frac{40-39{,}5}{\sqrt{\frac{0{,}2}{20}+\frac{0{,}2}{10}}} = 2{,}89 > 1{,}96 = z_{0{,}025}$.

Hylätään nollahypoteesi 5 %:n riskitasolla. Päätellään, että eivät tuota keskimäärin samanmittaisia tankoja. Pienin riskitaso, jolla H_0 voidaan hylätä, on 0,0038.

Esim. 5 Eräs yritys valmistaa 10 metrin teräskaapeleita tehtaissa X ja Y. Tarkasteltiin kaapeleiden murtolujuuksia (kilonewtoneina). Haluttiin selvittää, ovatko murtolujuudet keskimäärin samoja molempien tehtaiden tuotannoissa.

Tehtaan X tuotannosta valittiin satunnaisesti 9 ja tehtaan Y tuotannosta 16 kaapelia, joiden murtolujuudet mitattiin. Saatiin tulokset:

Tehdas X:	$\bar{x} = 30{,}11, \sum(x_i - \bar{x})^2 = 0{,}8013$
Tehdas Y:	$\bar{y} = 29{,}63, \sum(y_i - \bar{y})^2 = 3{,}0206$

Tapa 1

Kaavan (4.5) mukaisesi 95 %:n luottamusväli odotusarvojen erotukselle on $30{,}11 - 29{,}63 \pm 2{,}069 \cdot \underbrace{\sqrt{\frac{0{,}8013+3{,}0206}{9+16-2}}}_{0{,}4076}\sqrt{\frac{1}{9} + \frac{1}{16}}$. Tämä ei sisällä nollaa. Päätellään, että keskimääräisissä murtolujuuksissa on eroja.

Tapa 2

H_0: $\mu_X = \mu_Y$
H_1: $\mu_X \neq \mu_Y$

Kaavan (5.5) perusteella saadaan

$$t_{hav.} = \frac{30{,}11-29{,}63}{0{,}4076\sqrt{\frac{1}{9}+\frac{1}{16}}} = 2{,}83 > t_{0{,}025;9+16-2} = 2{,}069.$$

Hylätään nollahypoteesi 5 %:n riskitasolla. Päätellään, että keskimääräisissä murtolujuuksissa on eroja.

Pienin riskitaso, jolla H_0 voidaan hylätä, on 0,0047·2 = 0,0094, ks. http://stattrek.com/online-calculator/t-distribution.aspx tai http://onlinestatbook.com/2/calculators/t_dist.html