# 1.) Yhden otoksen keskiarvotesti, $\sigma^2$ tunnettu

HARJ 13

1) populaatio: erään ruokakaupan asiakkaat


$X$ = asiakkaan viipymisaika (minuuteissa)

$(X_1, \ldots, X_{16})$

Oletetaan, että $X \sim N(\mu, 6^2)$

2) hypoteesit: $\begin{cases} H_0: \mu = 26 \\ H_1: \mu \neq 26 \end{cases}$

3) testisuure: $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$, kun $H_0$ on tosi

4) testisuureen havaittu arvo: $\bar{x} = 25$, $\mu_0 = 26$, $\sigma = 6$ ja $n = 16$

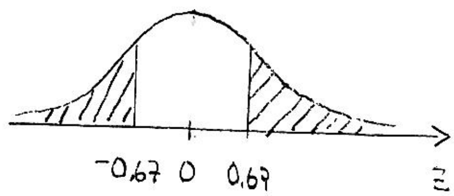
$$z = \frac{25 - 26}{6/\sqrt{16}} \approx -0.67$$

5) P-arvo $= P(Z \leq -0.67 \text{ tai } Z \geq 0.67 \mid H_0)$

$= P(|Z| \geq -0.67 \mid H_0)$

$= 2 \cdot P(Z \geq 0.67)$

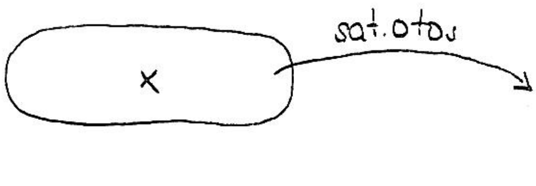
$= 2 \cdot 0.2514 = 0.5028$




6) Johtopäätökset: Aineisto on sopusoinnussa $H_0$:in kanssa, ts. aineiston perusteella näyttäisi sille, että ruokakaupan asiakkaiden keskimääräinen viipymisaika liikkeessä on 26 minuuttia.

# 2.) Yhden otoksen keskiarvotesti, $\sigma^2$ tuntematon

1) populaatio: kuljetukset paikasta A paikkaan B




$X$ = ajoaika (minuuteissa)

$(X_1, \ldots, X_{10})$

Oletetaan, että $X \sim N(\mu, \sigma^2)$

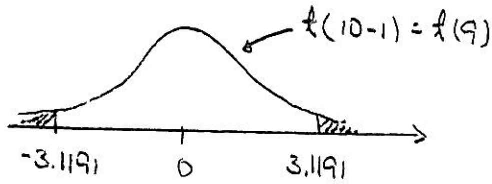
2) hypoteesit: $\begin{cases} H_0: \mu = 48 \\ H_1: \mu \neq 48 \end{cases}$

3) testisuure: $T = \dfrac{\bar{X} - \mu_0}{S_X/\sqrt{n}} \sim t(n-1)$, kun $H_0$ on tosi

4) testisuureen havaittu arvo: harjoitus 12, tehtävä 5: $\bar{x} = 51.1$ ja $s \approx 3.1429$
$\mu_0 = 48$ ja $n = 10$

$$T = \frac{51.1 - 48}{3.1429/\sqrt{10}} \approx 3.1191$$

5) p-arvo $= P(T \leq -3.1191 \text{ tai } T \geq 3.1191 \mid H_0)$
$= 2 \cdot P(T \geq 3.1191 \mid H_0)$




$t(n-1) = t(10-1) = t(9)$-jakauman taulukko:

$2 \cdot 0.005 < \text{p-arvo} < 2 \cdot 0.01$
$\Leftrightarrow 0.01 < \text{p-arvo} < 0.02$

6) johtopäätökset: Aineisto on ristiriidassa $H_0$:n kanssa, $H_1$ uskottavampi.
Kuljetusyrityksen keskimääräinen kuljetusaika paikasta A paikkaan B ei näyttäisi olevan 48 minuuttia.

3. Olkoon $X$ = reikäaiheiden lkm MFP-ryhmässä
$Y$ = –"– –"– vertailuryhmässä

$(X_1, \ldots, X_n)$ on satunnaisotos $N(\mu_x, \sigma_x^2)$-jakaumasta ja
$(Y_1, \ldots, Y_m)$ on –"– $N(\mu_y, \sigma_y^2)$-jakaumasta.
Oletetaan lisäksi, että $\sigma_x^2 = \sigma_y^2$.

Otoksista: $\bar{x} = 19.98$, $s_x = 10.61$, $n = 208$
$\bar{y} = 22.39$, $s_y = 11.96$, $m = 194$

a) kahden riippumattoman otoksen t-testi (keskiarvotesti)

1) Oletetaan, että $X \sim N(\mu_x, \sigma_x^2)$ ja $Y \sim N(\mu_y, \sigma_y^2)$, missä $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ ovat tuntemattomia. Oletetaan lisäksi, että $\sigma_x^2 = \sigma_y^2$.

2) hypoteesit: $\begin{cases} H_0: \mu_x = \mu_y \\ H_1: \mu_x < \mu_y \end{cases}$
$H_0$: eli ryhmien välillä ei ole eroa keskimääräisissä reikäaiheiden lukumäärissä
$H_1$: ← "MFP-fluori ehkäisee..."

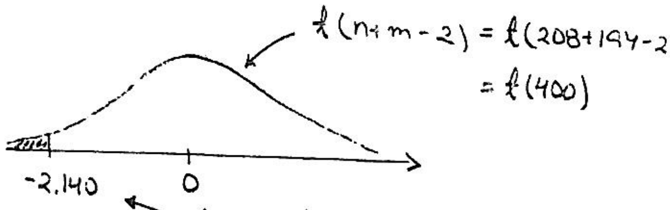
3) testisuure: $T = \dfrac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n + 1/m}} \sim t(n+m-2)$, kun $H_0$ on tosi

4.) testisuureen havaittu arvo:

$$S^2 = \frac{(n-1)\,S_x^2 + (m-1)\,S_y^2}{n+m-2} = \frac{(208-1)\cdot 10.61^2 + (194-1)\cdot 11.96^2}{208+194-2} \approx 127.2736$$

$$T = \frac{19.98 - 22.39}{\sqrt{127.2736}\cdot\sqrt{1/208 + 1/194}} \approx -2.1403$$

5) p-arvo $= P(T \leq -2.140 \mid H_0)$




$t(n+m-2) = t(208+194-2) = t(400)$-jakauman taulukko: $0.01 < \text{p-arvo} < 0.025$

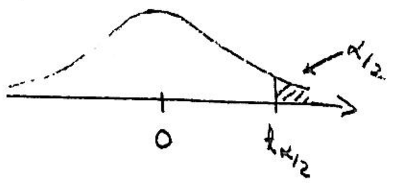
6) johtopäätökset: Aineisto on ristiriidassa $H_0$:in kanssa, $H_1$ uskottavampi. MFP-fluori siis näyttäisi ehkäisevän hampaiden reikiintymistä.

b) $100(1-\alpha)$ %:n luottamusväli odotusarvojen erotukselle $(\mu_x - \mu_y)$ on

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \cdot S\sqrt{1/n + 1/m}$$

, missä $t_{\alpha/2}$ on sellainen vakio, että kun $T \sim t(n+m-2)$

$\Rightarrow P(T \geq t_{\alpha/2}) = \alpha/2$




Nyt $100(1-\alpha) = 95 \Rightarrow \alpha = 0.05$ ja $\alpha/2 = 0.025$

$t(n+m-2) = t(208+194-2) = t(400)$-jakauman taulukko: $P(T \geq 1.966) = 0.025$

$\Rightarrow t_{\alpha/2} = t_{0.025} = 1.966$

$\Rightarrow$ kysytty lv. on: $(19.98 - 22.39) \pm 1.966 \cdot \sqrt{127.2736} \cdot \sqrt{1/208 + 1/194}$

$\Leftrightarrow -2.41 \pm 2.2138$

$\Leftrightarrow \underline{\underline{(-4.624,\ -0.196)}}$

TULKINTA: koska nolla ei kuulu lasketun luottamusvälin sisään, näyttäisi MFP-fluorilla olevan (ehkäisevää) vaikutusta hampaiden reikiintymiseen.

## (4.) Yhden suhteellisen osuuden testi

1) populaatio: uuden tehtaan valmistamat tietyn merkkiset jääkaapit

$X$ = laatuvaatimusten täyttyminen

$$X \xrightarrow{\text{sat.otos}} (X_1, \ldots, X_{500})$$

Merk. $X = \begin{cases} 1, & \text{kun tuote ei täytä laatuvaatimuksia} \\ 0, & \text{-''- -''- täyttää -''-} \end{cases}$

Oletetaan, että $X \sim Bern(\pi)$ ja $P(X=1) = \pi$

2) hypoteesit: $\begin{cases} H_0: \pi = 0.02 \\ H_1: \pi \neq 0.02 \end{cases}$

3) testisuure: $Z = \dfrac{P - \pi_0}{\sqrt{(\pi_0(1-\pi_0))/n}} \sim N(0,1)$ likimain, kun $H_0$ on tosi

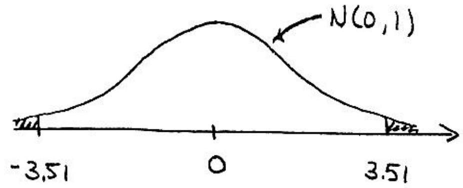
4) testisuureen havaittu arvo: $P = 21/500 = 0.042$, $\pi_0 = 0.02$ ja $n = 500$

$$Z = \frac{0.042 - 0.02}{\sqrt{(0.02 \cdot (1-0.02))/500}} \approx 3.51$$

5) p-arvo $= P(Z \leq -3.51 \text{ tai } Z \geq 3.51 \mid H_0)$

$= 2P(Z \geq 3.51 \mid H_0)$

$= 2 \cdot 0.0002 = 0.0004$




6) johtopäätökset: Aineisto on ristiriidassa $H_0$:n kanssa, $H_1$ uskottavempi. Johdon epäilyt laadun muuttumisesta näyttäisivät pitävän paikkansa.

## (5.) Oletetaan, että $(X_1, \ldots, X_{11000})$ on satunnaisotos $Bern(\pi_1)$-jakaumasta, missä

$X = \begin{cases} 1, & \text{kun aspiriiniryhmään kuuluva henkilö saa sydänkohtauksen 3 v. aikana} \\ 0, & \text{-''- -''- -''- -''- ei saa sydänkohtausta -''- -''-} \end{cases}$

$(Y_1, \ldots, Y_{11000})$ on satunnaisotos $Bern(\pi_2)$-jakaumasta, missä

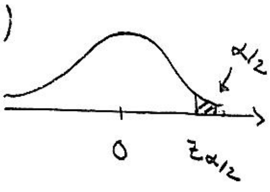
$Y = \begin{cases} 1, & \text{kun verrokkiryhmään kuuluva henkilö saa sydänkohtauksen 3 v. aikana} \\ 0, & \text{-''- -''- -''- -''- ei saa sydänkohtausta -''- -''-} \end{cases}$

a) $100(1-\alpha)$ %:n lv. suhteellisten osuuksien erotukselle $(\pi_1 - \pi_2)$:

$$(P_1 - P_2) \pm z_{\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n} + \frac{P_2(1-P_2)}{m}}$$

, missä $z_{\alpha/2}$ on sellainen vakio, että kun $Z \sim N(0,1)$

$\Rightarrow P(Z > z_{\alpha/2}) = \alpha/2$




Nyt $100(1-\alpha) = 99 \Rightarrow \alpha = 0.01$ ja $\alpha/2 = 0.005$

$N(0,1)$ - jakauman taulukko: $z_{\alpha/2} = z_{0.005} = 2.58$

$\Rightarrow$ kysytty lv. on: $\left(\frac{104}{11000} - \frac{189}{11000}\right) \pm 2.58 \cdot \sqrt{\frac{104/11000\,(1 - 104/11000)}{11\,000} + \frac{189/11000\,(1 - 189/11000)}{11\,000}}$

$\Leftrightarrow -0.00773 \pm 0.00399$

$\Leftrightarrow \underline{\underline{(-0.0117, -0.0037)}}$

TULKINTA: Koska nolla ei kuulu lasketun luottamusvälin sisään, näyttäisi aspiriinilla olevan (ehkäisevää) vaikutusta sydänkohtauksen riskiin

b) <u>Kahden suhteellisen osuuden testi</u>

1) Oletetaan, että $X \sim Bern(\pi_1)$ ja $Y \sim Bern(\pi_2)$ ($X$ ja $Y$ määritelty tehtävän alussa)

2) hypoteesit: $\begin{cases} H_0: \pi_1 = \pi_2 & \text{eli vertailtaville ryhmillä ei eroa} \\ H_1: \pi_1 < \pi_2 & \text{... aspiriinilla ehkäisevä vaikutus} \end{cases}$

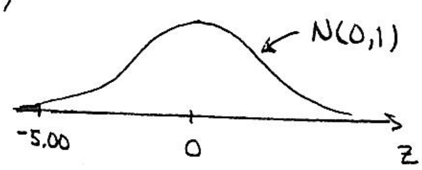
3) testisuure: $Z = \dfrac{P_1 - P_2}{\sqrt{P(1-P)(1/n + 1/m)}} \sim N(0,1)$ likimain, kun $H_0$ on tosi

4) testisuureen havaittu arvo:

$$P = \frac{nP_1 + mP_2}{n+m} = \frac{11000 \cdot \frac{104}{11000} + 11000 \cdot \frac{189}{11000}}{11000 + 11\,000} = \frac{293}{22\,000}$$

$$Z = \frac{104/11000 - 189/11000}{\sqrt{\frac{293}{22000}\cdot\left(1 - \frac{293}{22\,000}\right)\cdot\left(1/11000 + 1/11000\right)}} \approx -5,00$$

5) p-arvo $= P(Z \leq -5,00 \mid H_0) = 0.0000$




6) johtopäätökset: Aineisto on ristiriidassa $H_0$:n kanssa, $H_1$ uskottavampi. Aspiriinille (1 aspiriinitabletti 3 kertaa viikossa) näyttäisi olevan [illegible]

(6.) **$\chi^2$-yhteensopivuustesti**

1) populaatio: lääkäriaseman AB asiakkaat

$$X \xrightarrow{\text{sat. otos}} (X_1, \ldots, X_{400})$$

$X$ = saapumisaika (luokiteltuna)

2) hypoteesit:

$H_0$: $X$:n todennäköisyysjakauma on

| $x_i$ | väh. 10 min. | 0-9 min... | myöh. kork. | myöh. enemm. | $\Sigma$ |
|---|---|---|---|---|---|
| $p_i$ | 0.80 | 0.10 | 0.06 | 0.04 | 1 |

$H_1$: $H_0$ epätosi

3) testisuure: $\chi^2 = \sum_{i=1}^{k} \frac{(F_i - e_i)^2}{e_i} \sim \chi^2(k-m-1)$ likimain, kun $H_0$ on tosi

4) testisuureen havaittu arvo:

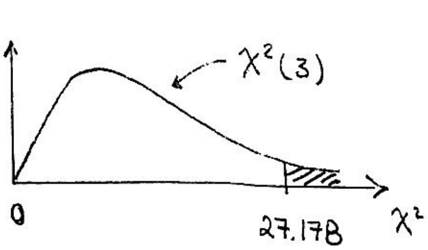
| $x_i$ | $F_i$ | $p_i^\circ$ | $e_i = np_i^\circ$ |
|---|---|---|---|
| väh. 10 min... | 287 | 0.80 | $400 \cdot 0.80 = 320$ |
| 0-9 min... | 49 | 0.10 | $400 \cdot 0.10 = 40$ |
| myöh. kork... | 30 | 0.06 | $400 \cdot 0.06 = 24$ |
| myöh. enemm... | 34 | 0.04 | $400 \cdot 0.04 = 16$ |
| Yhteensä | 400 | 1 | 400 |

(← $\chi^2$-testin ehdot voimassa $e_i$:iden osalta)

$$\chi^2 = \sum_{i=1}^{4} \frac{(F_i - e_i)^2}{e_i} = \frac{(287-320)^2}{320} + \frac{(49-40)^2}{40} + \frac{(30-24)^2}{24} + \frac{(34-16)^2}{16}$$

$$\approx 27.178$$

5) p-arvo = $P(\chi^2 \geq 27.178 \mid H_0)$




$\chi^2(k-m-1) = \chi^2(4-0-1) = \chi^2(3)$-jakauman taulukko:

p-arvo < 0,001

6) johtopäätökset: Aineisto on selvässä ristiriidassa $H_0$:n kanssa, $H_1$ uskottavampi. Lääkäriaseman AB asiakkaiden saapumisaikataulussa on tapahtunut muutos.

## 7. $\chi^2$-riippumattomuustesti

Aineisto ristiintaulukkona:

| | | siemeniä | | |
|---|---|---|---|---|
| | | ei | kyllä | Yhteensä |
| greipin väri | vaaleanpunainen | 20 | 60 | 80 |
| | valkoinen | 40 | 80 | 120 |
| | Yhteensä | 60 | 140 | 200 |

1) populaatio: greipit

$$(x,y) \xrightarrow{\text{sat. otos}} (X_1, Y_1), \ldots, (X_{200}, Y_{200})$$

$X$ = greipin väri
$Y$ = greipin siemenominaisuus

2) hypoteesit: $H_0$: $X$ ja $Y$ riippumattomia toisistaan $(X \perp\!\!\!\perp Y)$
$H_1$: $X$ ja $Y$ riippuvat toisistaan $(X \not\!\perp\!\!\!\perp Y)$

3) testisuure: $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{m} \frac{(F_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2[(r-1)(m-1)]$ likimain, kun $H_0$ on tosi

4) testisuureen havaittu arvo: odotetut frekvenssit $E_{ij} = \frac{F_{i\cdot} F_{\cdot j}}{n}$

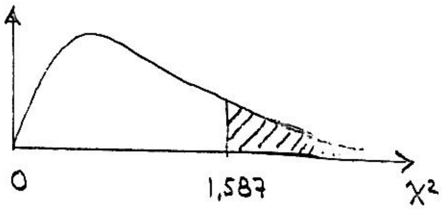
$$E_{11} = \frac{80 \cdot 60}{200} = 24, \quad E_{12} = \frac{80 \cdot 140}{200} = 56, \quad E_{21} = \frac{120 \cdot 60}{200} = 36, \quad E_{22} = \frac{120 \cdot 140}{200} = 84$$

($\chi^2$-testin ehdot voimassa $E_{ij}$:den osalta)

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(F_{ij} - E_{ij})^2}{E_{ij}} = \frac{(20-24)^2}{24} + \frac{(60-56)^2}{56} + \frac{(40-36)^2}{36} + \frac{(80-84)^2}{84}$$

$$\approx 1.587$$

5) p-arvo = $P(\chi^2 \geq 1.587 \mid H_0)$




$\chi^2((r-1)\cdot(m-1)) = \chi^2((2-1)\cdot(2-1)) = \chi^2(1\cdot 1) = \chi^2(1)$-jakauman taulukko:

$$0.20 < \text{p-arvo} < 0.50$$

6) johtopäätökset: Aineisto on sopusoinnussa $H_0$:n kanssa eli greipin väri ja siemenominaisuus ovat toisistaan riippumattomia.