

MTTTP1, luento 9.10.2018

KERTAUSTA TESTAUKSESTA, p-ARVO

- Asetetaan

H_0

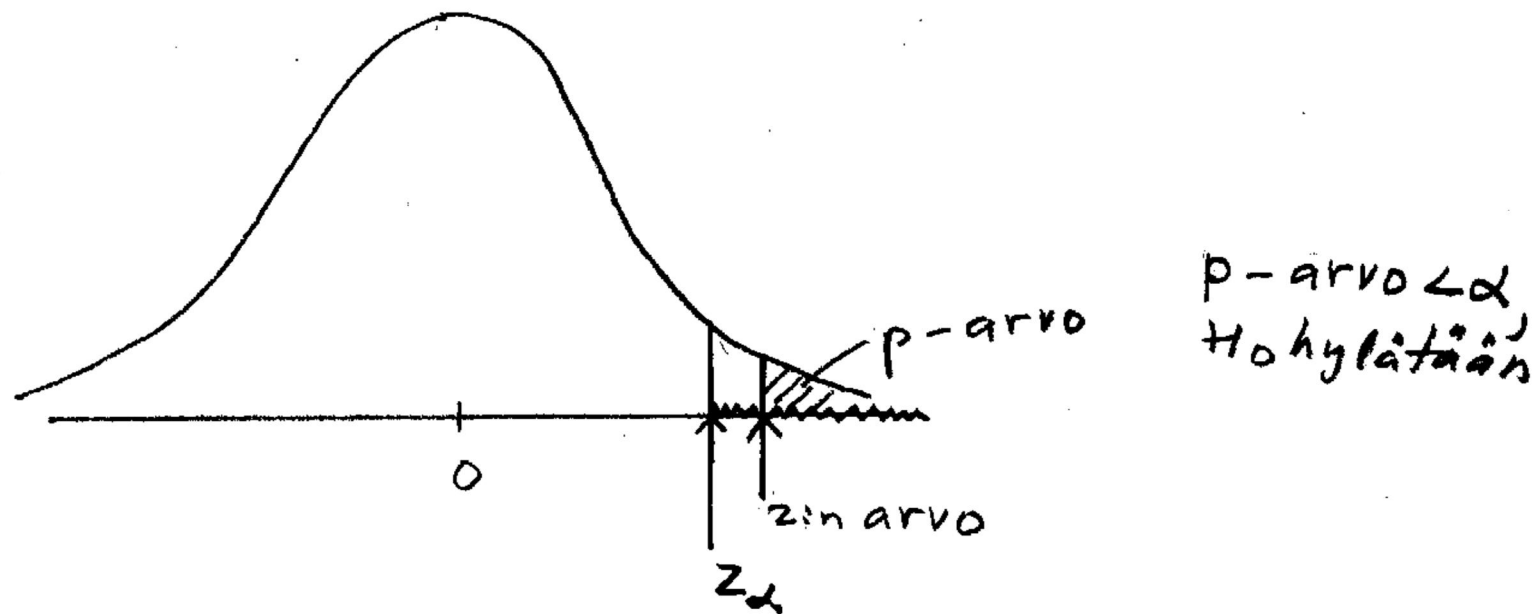
H_1

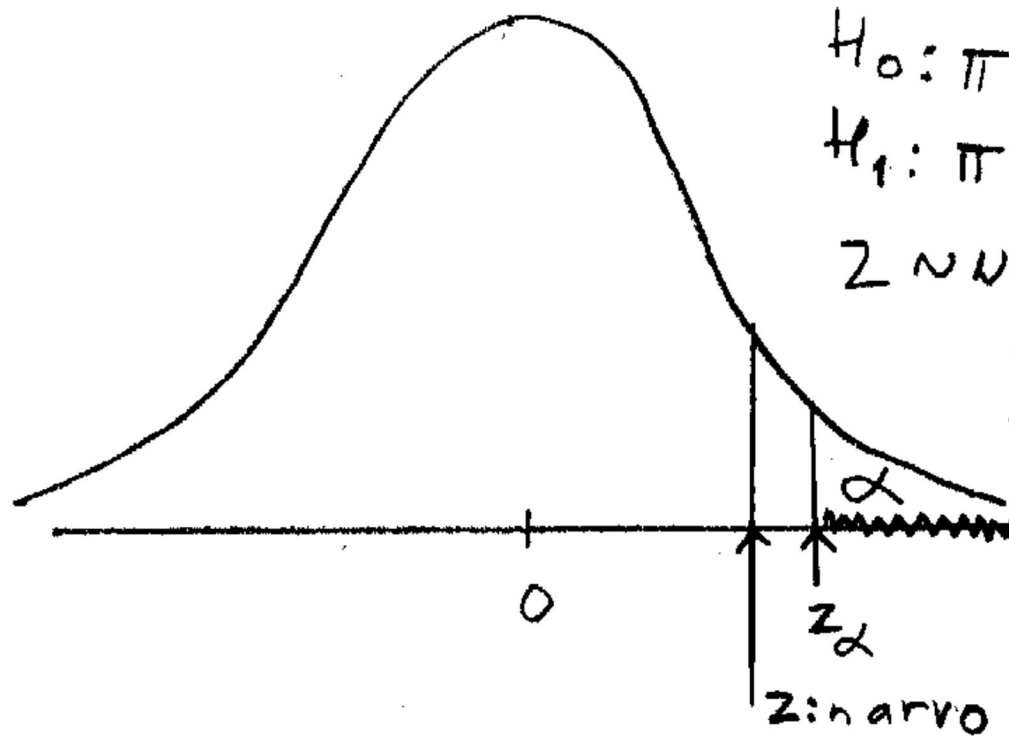
- Valitaan testisuure, jonka jakauma tunnetaan H_0 :n ollessa tosi.
- Lasketaan otoksesta testisuureelle arvo.

- Kiinnitetään riskitaso α (eli todennäköisyys, että tosi H_0 hylätään). Määritetään testisuureen harvinaisten arvojen joukko, joka riippuu riskitasosta ja vaihtoehtoisesta hypoteesista.
- Hylätään H_0 , jos saatu arvo kuuluu harvinaisten arvojen joukkoon, muulloin hyväksytään.
- H_0 :n hylkääminen johtaa H_1 :n hyväksymiseen.

- Päätelmä voidaan myös tehdä nk. p-arvon perusteella, p-arvo on pienin riskitaso, jolla H_0 voidaan hylätä. Jos p-arvo on valittua riskitasoa α pienempi, niin H_0 hylätään, muulloin hyväksytään.

$H_0: \pi = \pi_0, H_1: \pi > \pi_0, Z \sim N(0, 1)$ likimain, kun H_0 tosi.





<u>p-arvo</u>	<u>tulos tilastollisesti</u>
$< 0,05$	melkein merkitsevä
$< 0,01$	merkitsevä
$< 0,001$	erittäin merkitsevä

7.7.1 Prosenttiosuuden testaus (kertausta)

$$H_0: \pi = \pi_0$$

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(100 - \pi_0)/n}} \stackrel{\text{likimain}}{\sim} N(0,1), \text{ kun } H_0 \text{ tosi}$$

Esim. 7.7.1. Tarkastellaan erään liikkeen asiakkaita. Halutaan tutkia, ovatko asiakkaista yli puolet naisia. Tehdään 200 asiakkaan satunnaisotos jossa naisia on 113.

Nyt

$$H_0: \pi = 50 \% \text{ ja}$$

$$H_1: \pi > 50 \%. \text{}$$

Otoksessa naisia 56,5 %. Aineiston perusteella testisuureen arvoksi saadaan

$$z = \frac{56,5 - 50}{\sqrt{50(100 - 50) / 200}} = 1,838 > z_{0,05} = 1,6449$$

H_0 hylätään 5 %:n riskitasolla, mutta ei 2,5 %:n riskillä, joten $0,025 < p\text{-arvo} < 0,05$.

Ks.

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=81>

Esim. 7.7.2. Eräs puolue väittää, että suomalaisista 40 % kannattaa sen tekemää ehdotusta. Väitteen tutkimiseksi teit kyselyn 1000 henkilölle, joista 360 ilmoitti kannattavansa kyseistä ehdotusta. Onko puolue arvioinut ehdotuksen kannattajat oikein?

Nyt

$$H_0: \pi = 40 \% \text{ ja } H_1: \pi < 40 \%$$

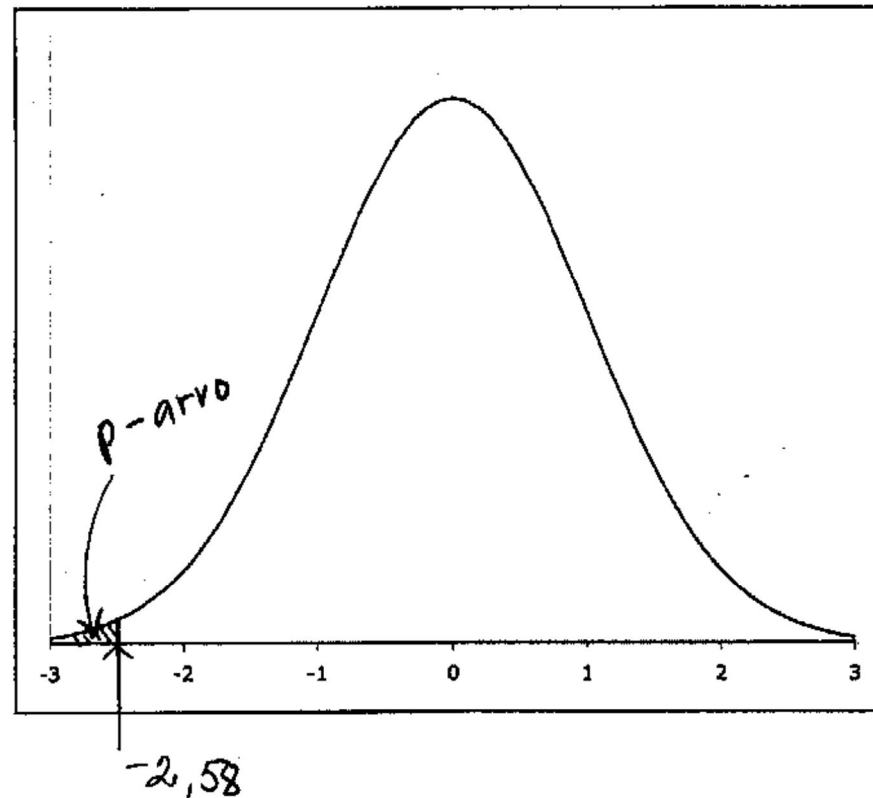
Otoksessa kannattajia 36 %. Testisuureen arvoksi saadaan

$$z = \frac{36 - 40}{\sqrt{40(100 - 40) / 1000}} = -2,58$$

Onko tämä harvinaisten arvojen joukkoon kuuluva?

Jos tehdään testaus 1 %:n riskitasolla, niin harvinaisiksi arvoiksi katsotaan lukua $-2,3264 = -z_{0,01}$ pienemmät ja täten hylätään nollahypoteesi ($-2,58 < -2,3264$).

Pienin riskitaso, jolla H_0 voidaan hylätä, on $P(Z \leq -2,58) \approx 0,005$, koska $P(Z \leq -2,5758) = 0,005$.



Vaihtoehtoinen hypoteesi voidaan asettaa myös kaksisuuntaisena, jolloin $H_1: \pi \neq 40\%$.

Tällöin p -arvo on $P(Z \leq -2,58) + P(Z \geq 2,58) \approx 0,01$.

Jos valitaan kaksisuuntaisessa testissä tätä suurempi riskitaso (esimerkiksi 0,025), niin nollahypoteesi hylätään ja päätellään, että puolue on arvioinut ehdotuksen kannattajien osuuden väärin.

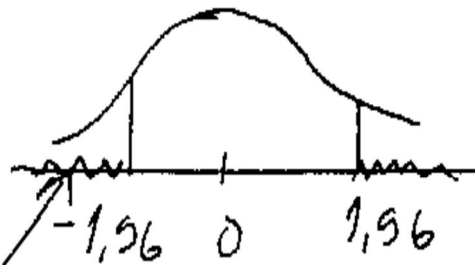
Esim. Kahvin myyjä väittää, että 15 % kahvin juojista valitsee kahvimerkin hinnan perusteella. Tehdään tutkimus, jossa 250 kahvin juojalta kysytään kahvimerkin valintaa vaikuttavia tekijöitä. Vastanneista 25 valitsi kahvin hinnan perusteella. Uskotko myyjän väitteen?

$$H_0: \pi = 15 \quad H_1: \pi \neq 15 \quad Z = \frac{p - \pi_0}{\sqrt{\pi_0(100 - \pi_0)/n}} \stackrel{\text{likimain}}{\sim} N(0,1), \text{ kun } H_0 \text{ totsi.}$$

$$Z_{\text{havaittu}} = \frac{10 - 15}{\sqrt{15(100 - 15)/258}} = -2,21$$

$$\alpha = 0,05 \quad \alpha/2 = 0,025$$

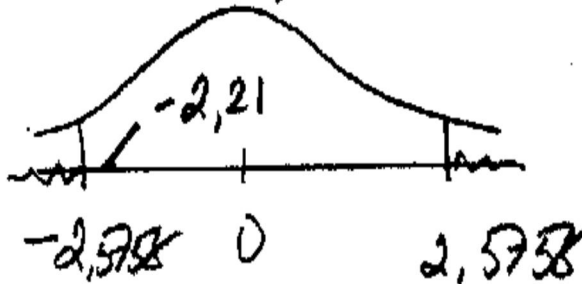
$$Z_{0,025} = 1,96$$



-2,21

$$\alpha = 0,01, \alpha/2 = 0,005$$

$$Z_{0,005} = 2,5758$$



H_0 hylätään 5%:n riskitasolla, mutta ei 1%:n riskitasolla.

$$0,01 < p\text{-arvo} < 0,05$$

7.7.2 Odotusarvon testaus

$$H_0 : \mu = \mu_0$$

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t_{n-1}, \text{ kun } H_0 \text{ tosi}$$

Esim. 7.7.3. Kauppias väitti, että kananmunien keskipaino on 50 g. Tehdään 36 alkion satunnaisotos ja saadaan $\bar{x} = 47$, $s = 6$. Onko kauppiaan väittämään uskominen? Nyt

$$H_0: \mu = 50 \text{ g}$$

$$H_1: \mu < 50 \text{ g}$$

Saadaan

$$t = \frac{47-50}{6/\sqrt{36}} = -3,$$

joka on pienempi kuin $-t_{0,005;35} \approx -2,75$, joten hylätään nollahypoteesi 0,5 %:n riskitasolla (ei uskota kauppiaan väitettä).

Voidaan arvioida, että p -arvo eli $P(t_{35} < -3) < 0,005$.

Vaihtoehtoinen hypoteesi voidaan myös asettaa kaksisuuntaisena eli

$$H_1: \mu \neq 50 \text{ g}$$

Jos päättely tehdään 1 %:n riskitasolla, niin harvinaisten arvojen rajoina ovat $\pm t_{0,01/2;35} \approx \pm 2,75$, joiden ulkopuolelle -3 jää. Päätellään, että $\mu \neq 50 \text{ g}$.

Nyt p -arvo = $P(t_{35} < -3) + P(t_{35} > 3) < 0,01$.

Esim. 7.7.4. Perunalastujen valmistaja ilmoittaa perunalastupussiensa keskipainoksi 340 g. Tutkitaan väitettä ja tehdään 16 pussin satunnaisotos ja saadaan keskipainoksi 336 g ja keskihajonnaksi 11 g. Voitko uskoa valmistajan väitteen?

$$H_0: \mu = 340 \text{ g}$$

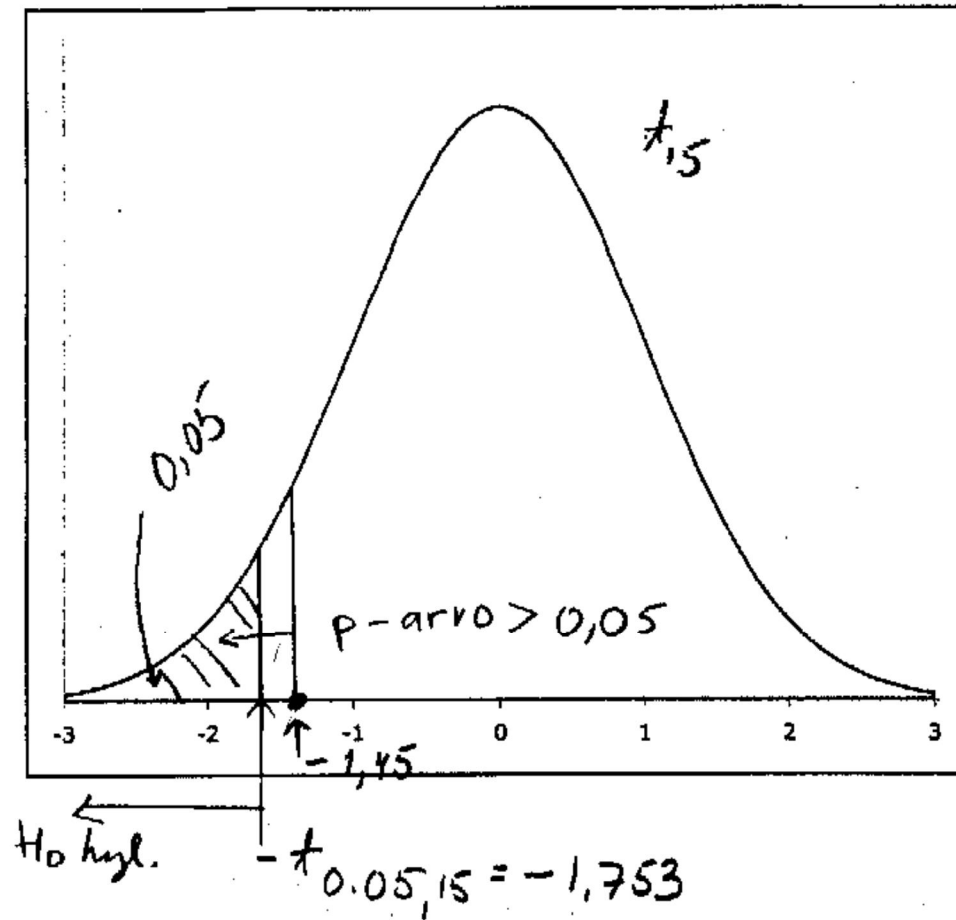
$$H_1: \mu < 340 \text{ g}$$

$$t = \frac{336 - 340}{11 / \sqrt{16}} = -1,45$$

5 %:n riskitasolla harvinaisten arvojen raja on

$-t_{0,05;15} = -1,753$, joten H_0 hyväksytään.

Nyt p -arvo on $P(t_{15} < -1,45) > 0,05$, graafisesti:



Sama johtopäätelmä tehdään, jos vaihtoehtoinen hypoteesi asetetaan kaksisuuntaiseksi eli $H_1: \mu \neq 340$ g.

Koska $t_{0,05/2;15} = 2,131$, niin H_0 hyväksytään 5 %:n riskitasolla.

Nyt p-arvo = $P(t_{15} < -1,45) + P(t_{15} > 1,45) > 0,10$.

Esim. Lepakoiden tunnistusmatkat, ks.

http://www.sis.uta.fi/tilasto/mttp1/syksy2018/esimerkit_kaavoihin.pdf

7.7.3 χ^2 -riippumattomuustesti

Kahden muuttujan välinen riippuvuustarkastelu ristiintaulukon perusteella, asetetaan

H_0 : ei riippuvuutta

H_1 : on riippuvuutta

H_0 hylätään, jos p-arvo pieni (esim. $< 0,05$).

7.7.4 Odotusarvojen yhtäsuuruuden testaaminen t-testillä

Tutkitaan kahden populaation odotusarvojen yhtäsuuruutta riippumattomien otosten t-testillä, asetetaan

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

H_0 hylätään, jos p-arvo pieni (esim. $< 0,05$).