MTTTP5, luento 3.12.2018

6.1.2 Yhdessä populaatiossa tietyn tyyppisten alkioiden prosentuaalista osuutta koskeva päättely

$$\boxed{H_0 : \pi = \pi_0}$$

Oletetaan, että populaatiossa viallisia π %. Olkoon X_1, X_2, . . . , X_n satunnaisotos tästä populaatiosta.

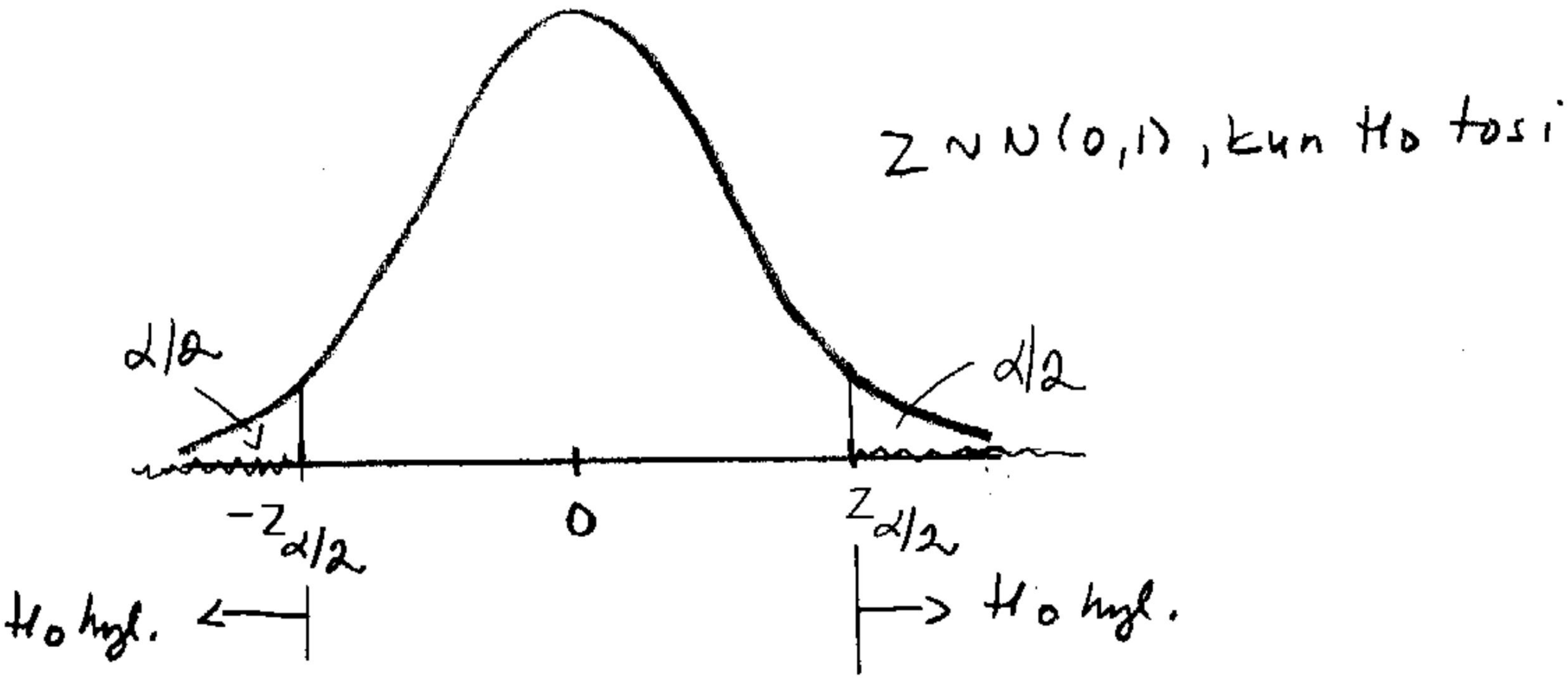
Jos H_0:n on tosi, niin viallisten prosenttiosuus otoksessa

$$p \sim N\left(\pi_0, \frac{\pi_0(100-\pi_0)}{n}\right), likimain,$$

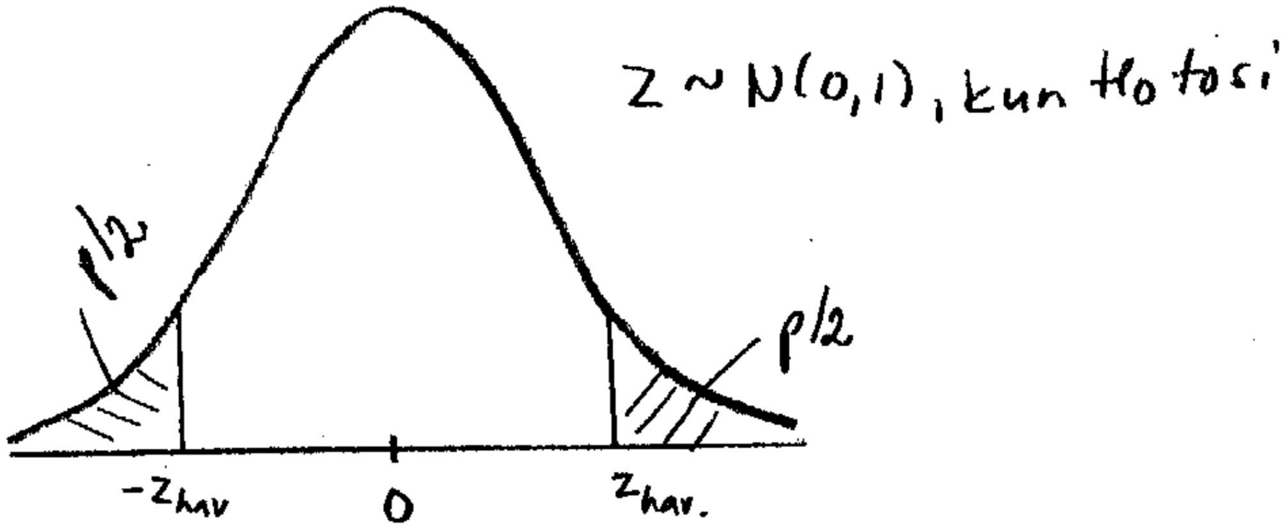
joten

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(100 - \pi_0)/n}} \sim N(0,1), likimain.$$

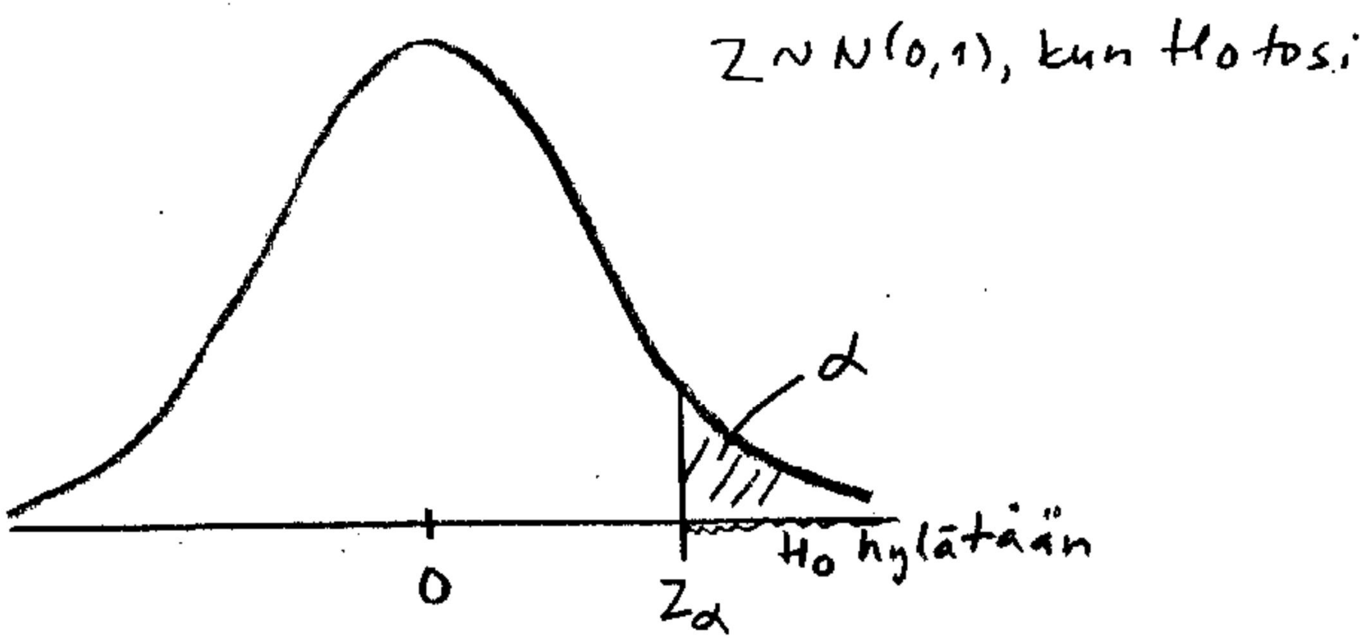
- Jos $H_1 : \pi \neq \pi_0$, niin H_0 hylätään riskitasolla α, jos otoksesta laskettu $|z_{hav}| > z_{\alpha/2}$.



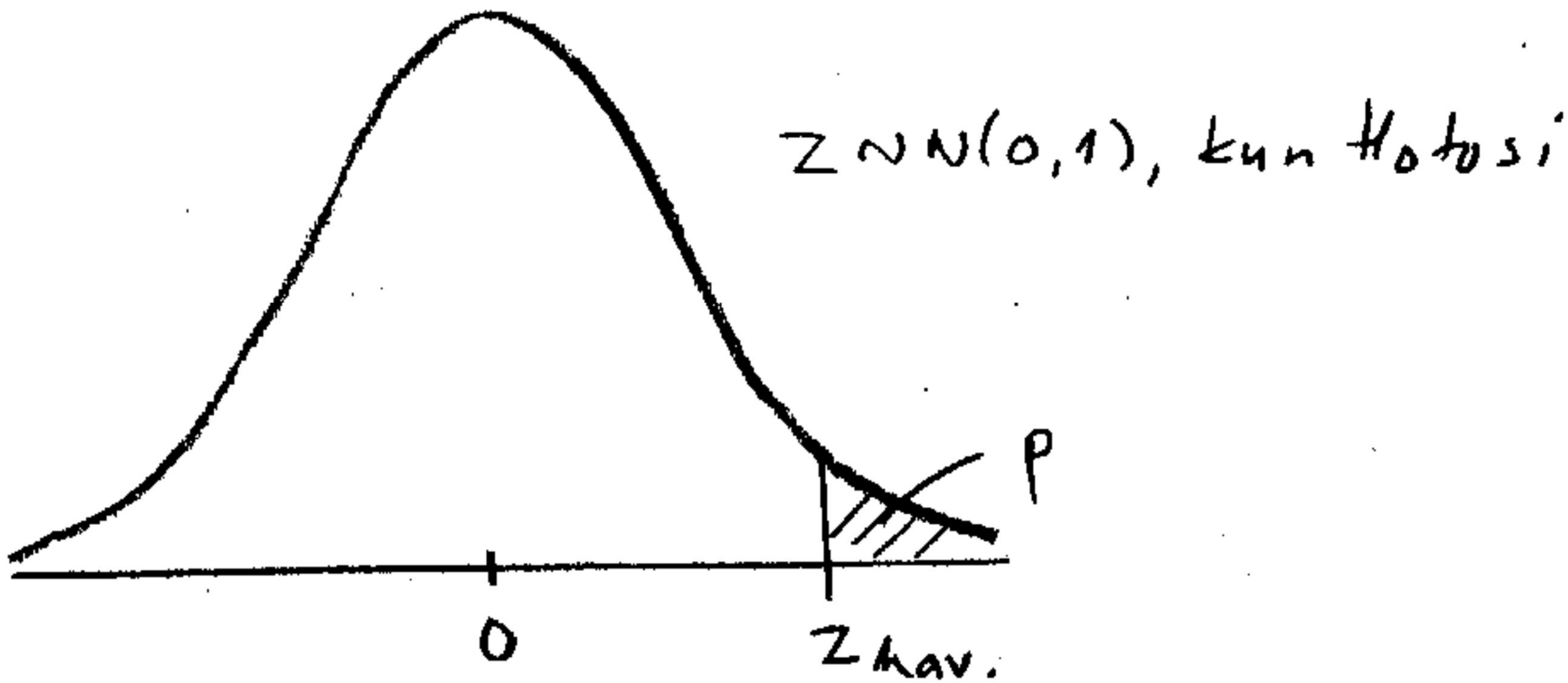
Pienin riskitaso p, jolla H_0 voidaan hylätä, on $2P(Z > |z_{hav}|)$.



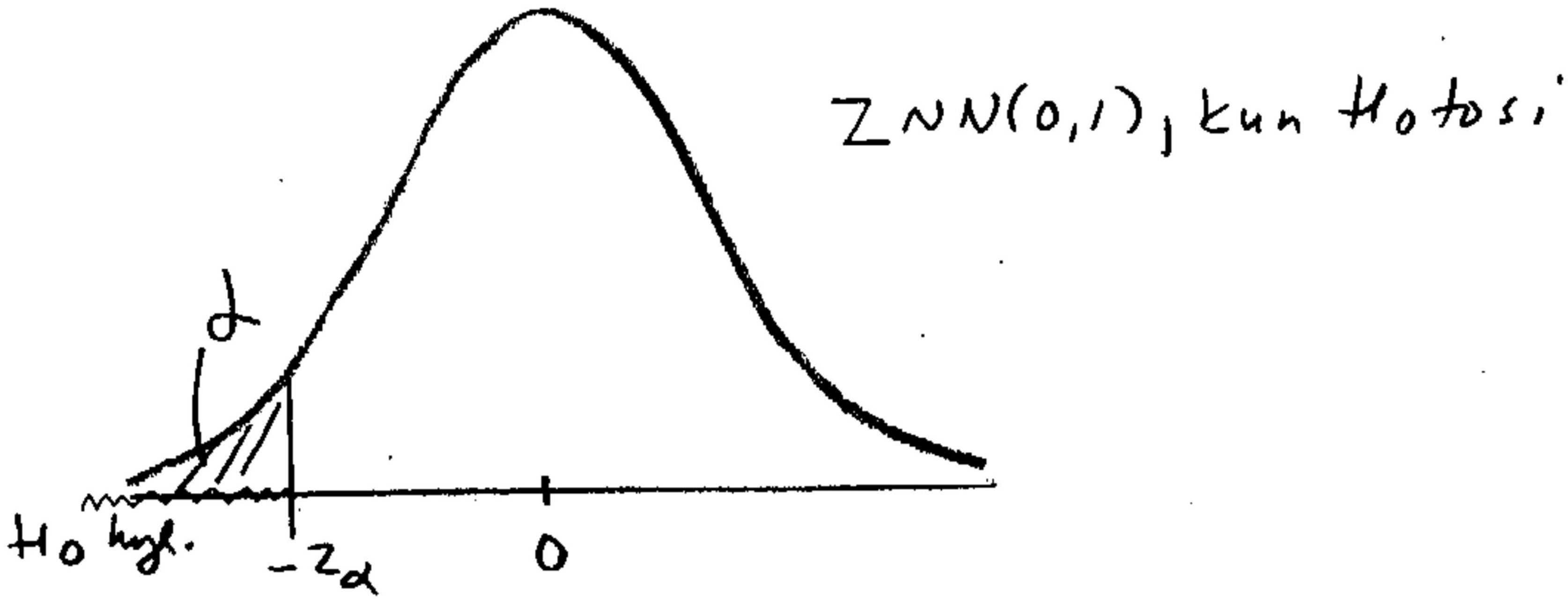
- Jos $H_1 : \pi > \pi_0$, niin H_0 hylätään riskitasolla α, jos otoksesta laskettu $z_{hav} > z_\alpha$.



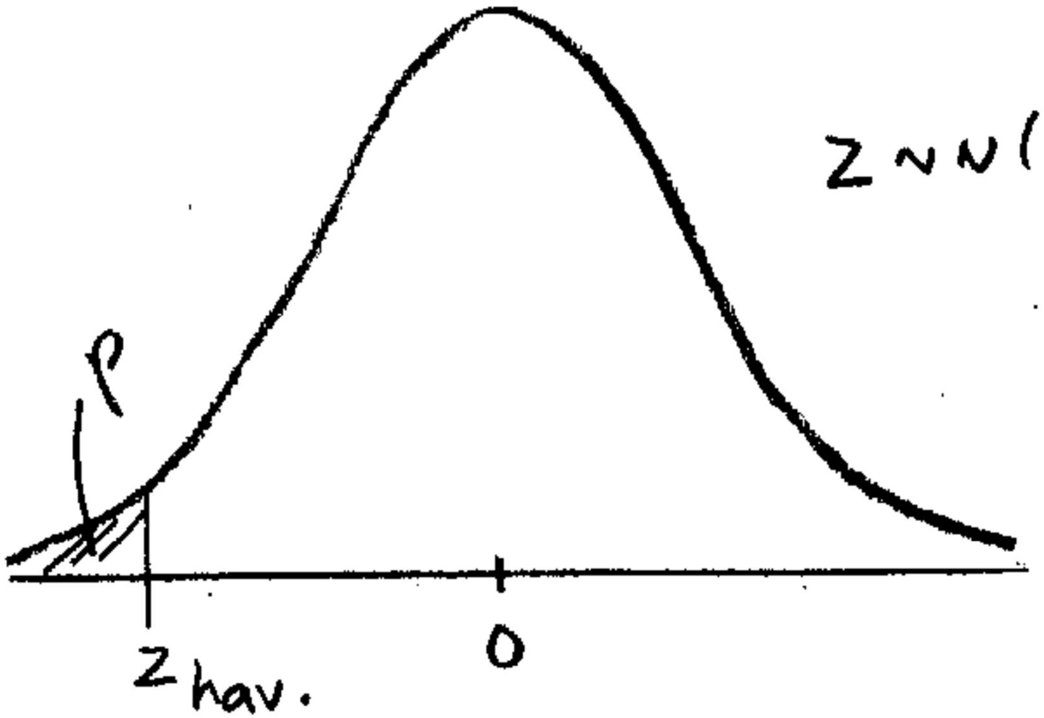
Pienin riskitaso p, jolla H_0 voidaan hylätä, on $P(Z > z_{hav})$.



- Jos H_1 : $\pi < \pi_0$, niin H_0 hylätään riskitasolla α, jos otoksesta laskettu $z_{hav} < -z_\alpha$.



Pienin riskitaso p, jolla H_0 voidaan hylätä, on $P(Z < z_{hav})$.



$Z \sim N(0,1)$, kun H_0 tosi

Esim. 6.1.6. Ystäväsi väittää, että suomalaisista on 10% vasenkätisiä. Tutkit asiaa ja valitset satunnaisesti 400 suomalaista, joista vasenkätisiä on 47. Uskotko ystäväsi väitteen?

$$H_0 : \pi = 10$$

$$H_1 : \pi > 10$$

Jos Ho tosi,

$$Z = \frac{p - 10}{\sqrt{10(100 - 10)/n}} \sim N(0,1), likimain.$$

Otoksesta laskettu z:n arvo on

$$z = \frac{11{,}75 - 10}{\sqrt{10(100-10)/400}} = 1{,}17$$

Pienin riskitaso, jolla H_0 voidaan hylätä yksisuuntaisessa testissä, on $P(Z > 1{,}17) = 1 - \Phi(1{,}17) = 1 - 0{,}8790 = 0{,}121$. Uskotaan siis ystävän väite.

Jos valitaan 5 %:n riskitaso, niin yksisuuntaisessa testissä kriittinen arvo on $z_{0,05} = 1{,}64$ (koska $\Phi(1{,}64) = 0{,}9495$) ja kaksisuuntaisessa testissä $z_{0,05/2} = 1{,}96$ (koska $\Phi(1{,}96) = 0{,}975$).

Esim. Ruletissa on 37 numeroa, joista pyöritettäessä jokaisella pitäisi olla sama todennäköisyys tulla tulokseksi. Pelipaikka voittaa numerolla nolla. Rulettia pyöritetään 3700 kertaa. Saadaan nollia 140 eli 3,78 %. Toimiiko ruletti oikein?

$H_0 : \pi = 100 \cdot 1/37$

$H_1 : \pi > 100 \cdot 1/37$

Jos H_0 tosi, niin

$$Z = \frac{p - \frac{100}{37}}{\sqrt{\frac{100}{37}\left(100 - \frac{100}{37}\right)/n}} \sim N(0,1), likimain.$$

Otoksesta laskettu z:n arvo on

$$z = \frac{\frac{140}{37} - \frac{100}{37}}{\sqrt{\frac{100}{37}\left(100 - \frac{100}{37}\right)/3700}} \approx 4{,}06$$

Pienin riskitaso, jolla H_0 voidaan hylätä yksisuuntaisessa testissä, on $P(Z > 4{,}06) = 1 - \Phi(4{,}06) \approx 0$. Ruletti ei toimi oikein.

Esim. 6.1.7. Öljy-yhtiö väittää, että erään kaupungin asunnoista 20 % lämmitetään öljyllä. Onko kuitenkin syytä olettaa, että vähemmän kuin viidesosa asunnoista lämmitetään öljyllä, jos 1000 satunnaisesti valitusta kaupungin asunnosta vain 160 lämmitettiin öljyllä?

Nyt

$$H_0: \pi = 20$$
$$H_1: \pi < 20.$$

Jos H_0 tosi, niin

$$Z = \frac{p - 20}{\sqrt{20(100 - 20)/n}} \sim N(0,1), likimain.$$

Aineiston perusteella testisuureen arvoksi saadaan

$$z = \frac{16-20}{\sqrt{20(100-20)/1000}} = -3{,}16 < -z_{0,001} = -3{,}08$$

H_0 hylätään 0,1 %:n riskitasolla. Päätellään, että alle viidesosa lämmitetään öljyllä.

Pienin riskitaso, jolla H_0 voidaan hylätä, on P(Z < -3,16) = 1 - Φ(3,16) = 1 – 0,9992 = 0,0008.

Esim. Aiempien tutkimusten perusteella 20 % kahvin juojista valitsi kahvin hinnan perusteella. Haluttiin selvittää, onko ostokäyttäytymisessä tapahtunut muutosta. Kysyttiin valinnan perusteita 100 kahvin juojalta. Päätellään muutosta tapahtuneen, jos p > 28 % (p = otoksessa valintansa hinnan perusteella tekevien %-osuus). Mikä on tällöin testissä käytetty α?

H_0: π = 20

Jos H_0 tosi, niin $p \sim N\left(20, \frac{20(100-20)}{100}\right), likimain.$

Käytetty riskitaso

α = P(H_0 hylätään, kun se on tosi)

= P(p > 28, kun π = 20)

= 1 – P(p ≤ 28, kun π = 20)

$$= 1 - \Phi\left(\frac{28-20}{\sqrt{20(100-20)/100}}\right) = 1 - \Phi(2) = 0{,}0228$$

Esim. Verkkokauppa pyrkii toimimaan siten, että tuotteet lähetetään kolmen työpäivän kuluttua tilauksesta. Tämä ei kuitenkaan aina ole mahdollista. Verkkokauppa haluaa toimia siten, että satunnaisia viivästymisiä voi olla 10 %. Viimeisen kuukauden ajalta satunnaisesti valituista 150 tilauksesta 21 lähetettiin myöhässä. Verkkokauppa toteaa, että lähetyksissä ei ole 10 % suurempaa viivästymistä. Millä riskitasolla päättely on voitu tehdä?

H_0: $\pi = 10$
H_1: $\pi > 10$

Jos Ho tosi,

$$Z = \frac{p - 10}{\sqrt{10(100 - 10)/n}} \sim N(0,1), likimain.$$

Aineiston perusteella testisuureen arvoksi saadaan

$$z = \frac{14 - 10}{\sqrt{10(100-10)/150}} = 1{,}63$$

Pienin riskitaso, jolla H_0 voidaan hylätä, on P(Z > 1,63) = 1 - Φ(1, 63) = 1 – 0,9484 = 0,0516.

Päättely on voitu tehdä esim. 5 %:n riskitasolla.

6.1.3 Kahden jakauman sijainnin vertailu

$\boxed{H_0 : \mu_1 = \mu_2}$ tai $\boxed{H_0\ \mu_1 - \mu_2 = 0}$

$X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu_1, \sigma_1^2)$:sta,

$Y_1, Y_2, \ldots, Y_m$ on satunnaisotos $N(\mu_2, \sigma_2^2)$:sta.

Oletetaan, että varianssit tunnettuja ja satunnaisotokset riippumattomia.

Kun H_0 on tosi, niin

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1).$$

Tätä käytetään testisuureena, päättely kuten aiemminkin normaalijakaumaa noudattavien testisuureiden tapauksissa.

Esim. Hedelmien viljelijä tietään pitkäaikaisen seurannan perusteella, että samanlaisissa olosuhteissa viljeltyinä lajike A ja B tuottavat keskimäärin yhtä suuren sadon. Lisäksi molempien lajikkeiden sadon vaihtelu on normaalisti jakautunut varianssina 66 g^2. Viljelijä ottaa käyttöön uudet viljelyalueet. Hän epäilee, että uudet kasvumaat eivät enää tuota keskimäärin samansuuruisia satoja. Hän valitsee satunnaisesti molempien lajikkeista sadoista 25 saaden lajikkeen A sadon keskiarvoiksi 116 g ja lajikkeen B sadon keskiarvoksi 111 g. Onko viljelijän epäily aiheellista?

Aseta tilanteeseen sopiva H_0 ja H_1. Suorita testaus 1 %:n riskitasolla. Laske pienin riskitaso, jolla H_0 voidaan hylätä.

$$H_0 : \mu_1 = \mu_2$$

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

Kun H_0 tosi, niin $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$,

$$z_{hav.} = \frac{116 - 111}{\sqrt{\frac{66}{25} + \frac{66}{25}}} = 2{,}18.$$

Koska $z_{0,01/2} = 2{,}58$, niin H_0 hyväksytään 1 %:n riskitasolla. Viljelijän epäily ei aiheellinen. Pienin riskitaso, jolla H_0 voidaan hylätä, on

$$\begin{aligned} 2(1 - P(Z < 2{,}18)) &= 2(1 - \Phi(2{,}18)) \\ &= 2(1 - 0{,}9854) \\ &= 0{,}0292. \end{aligned}$$