

MTTTP1, luento 4.10.2018

## KERTAUSTA

Luottamisvälin avulla voidaan arvioida populaation tuntematonta parametria.

Muodostetaan väli, joka peittää parametrin etukäteen valitulla todennäköisyydellä, nk. luottamustasolla.

Olkoon populaatiossa  $\pi$  % "viallisia".

Nyt  $\pi$ :n  $100(1 - \alpha)$  %:n luottamusväli on

$$p \pm z_{\alpha/2} \sqrt{p(100 - p)/n}$$

kaava (8)

Esim. Rahapelin pitäisi antaa voitto 20 %:lle pelatuista peleistä. Pelaat peliä 200 kertaa ja voitat 32 kertaa. Voitko uskoa, että 20 % peleistä voittaa?

Muodostetaan 95 %:n luottamusväli prosenttiosuudelle. Nyt  $\alpha = 0,05$ ,  $z_{0,05/2} = z_{0,025} = 1,96$ ,  $n = 200$ ,  $p = 16$ , luottamusväli

$$16 \pm 1,96\sqrt{16(100 - 16)/200}$$

$$16 \pm 5,1$$

Koska 20 kuuluu luottamusvälille, päätellään pelin toimivan luvatusalla tavalla.

Esim. 7.6.5. Yritys valvoo tuotantoaan. Virheellisten komponenttien osuus ei saisi olla suurempi kuin 4 %. Laaduntarkkailussa tehtiin 500 komponentin otos, jossa 28 komponenttia osoittautui virheellisiksi. Onko tuotanto keskeytettävä?

95 %:n luottamusväli virheellisten komponenttien prosenttiosuudelle

$$5,6 \pm 1,96 \sqrt{5,6(100-5,6)/500}$$

Virheellisten osuuden arvellaan olevan välillä 3,6 % - 7,6 %, joten vaihtelu on sallituissa rajoissa, koska 4 % kuuluu luottamusvälille.

## 7.6.2 Populaation odotusarvon luottamusväli

Esim. 7.6.6. Arvioidaan poikien keskimääräistä syntymäpituutta, siis poikapopulaation keskiarvoa. Otoksessa 65 pojan syntymäpituuden keskiarvo 50,95 cm ja keskihajonta 1,97 cm. Arvio populaation odotusarvon luottamusvälin avulla, määrittämisessä käytetään otoskeskiarvoa ja otoshajontaa. Poikien keskipituuden arvellaan olevan välillä 50,5 cm – 51,4 cm.

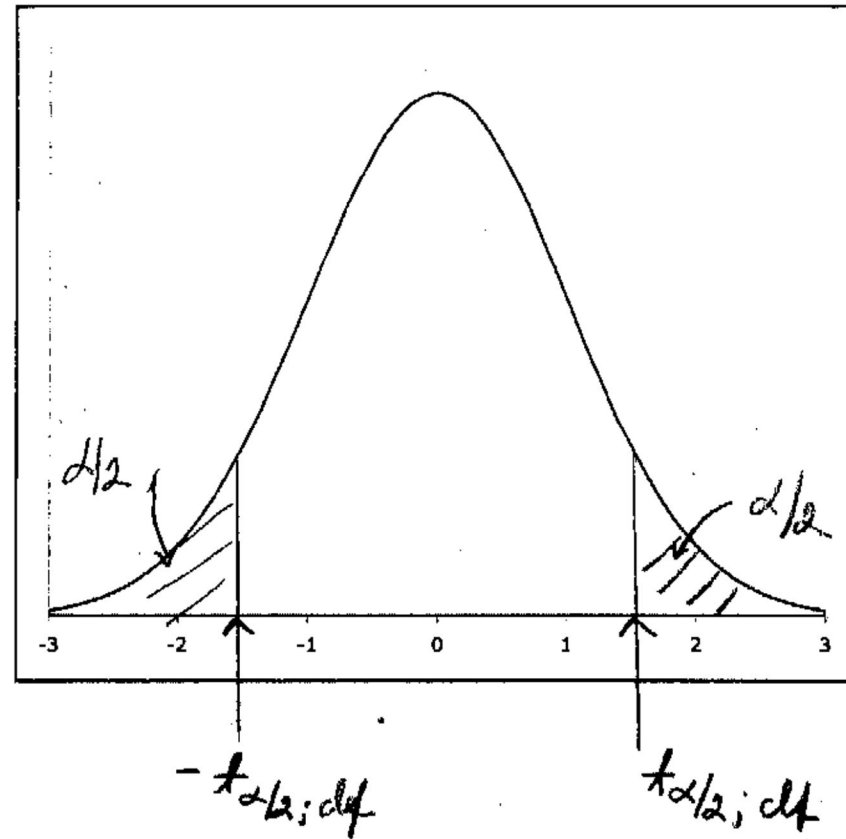
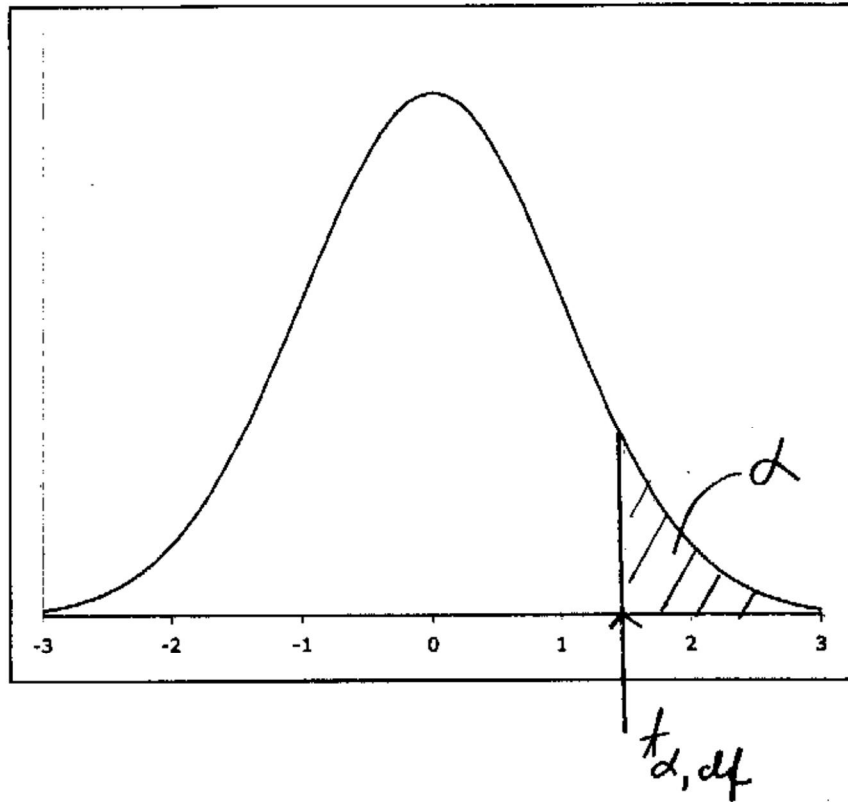
SPSS-tulos:

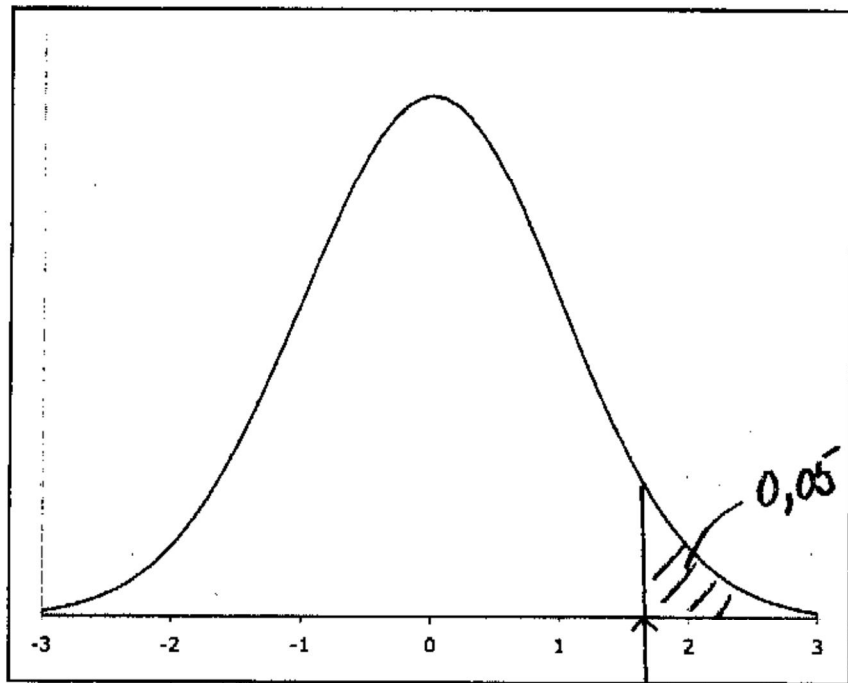
**Descriptives**

			Statistic	Std. Error
pituus	Mean		50,95	,245
	95% Confidence Interval for Mean	Lower Bound	50,47	
		Upper Bound	51,44	
	Std. Deviation		1,972	

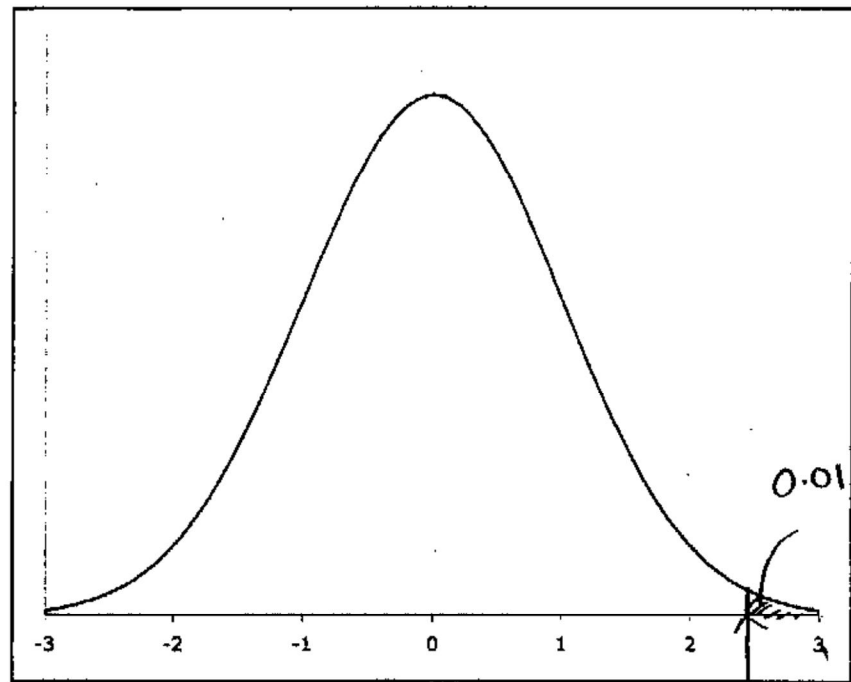
Kaava (9),  $100(1 - \alpha)$  %:n luottamusväli odotusarvolle

$$\bar{X} \pm t_{\alpha/2; n-1} s / \sqrt{n}$$

Studentin t-jakauman taulukkoarvot  $t_{\alpha,df}$  ja  $t_{\alpha/2,df}$ 



$$t_{0.05,10} = 1,812$$



$$t_{0.01,30}$$



Esim. 7.6.9. Tiedetään, että eräs kirjailija käyttää tuotannossaan virkkeitä, joiden keskipituus on 32 sanaa. Tutkija lukee erään tekstin, jossa on 30 virkettä. Näiden 30 virkkeen keskipituus on 35,0 sanaa ja keskihajonta 6,8 sanaa. Voisiko teksti olla peräisin kyseisen kirjailijan tuotannosta?

Muodostetaan odotusarvon 95 %:n luottamusväli. Nyt  $t_{0,05/2;30-1} = 2,045$  ja luottamusväli  $35,0 \pm 2,045 \cdot 6,8 / \sqrt{30}$ . Saadaan väliksi 32,5 – 37,5, jolle 32 ei kuulu. Päätellään, että teksti ei ole kyseisen kirjailijan tuotantoa.

Esim. Eräs tehdas valmistaa tiiliä, joiden keskipaino pitkän aikavälin seurannassa on ollut 2,000 kg. Erään päivän tuotannosta valittiin satunnaisesti 16 tiiltä. Näiden keskipainoksi saatiin 1,972 kg ja keskihajonnaksi 0,054 kg. Onko keskipainossa tapahtunut muutosta?

Muodostetaan odotusarvon 95 %:n luottamusväli.

Nyt  $t_{0,05/2;16-1} = t_{0,025;16-1} = 2,131$  ja luottamusväli

$$1,972 \pm 2,131 \cdot 0,054 / \sqrt{16}.$$

Saadaan väliksi  $1,972 \pm 0,029$ .

Koska 2,000 kuuluu välille, päätellään ettei muutosta.

Esim. 7.6.7. Esimerkissä 7.6.6, syntymäpituuden tarkastelu, otoksessa 65 poikaa

**Descriptives**

		Statistic	Std. Error
pituus	Mean	50,95	,245
	95% Confidence Interval for Mean	Lower Bound	50,47
		Upper Bound	51,44
	Std. Deviation	1,972	

Luottamusväli on laskettu

$$50,95 \pm 2 \cdot 1,972 / \sqrt{65}, (t_{0,05/2;65-1} \approx 2)$$

$$50,95 \pm 2 \cdot 0,245$$

Esim. Lepakoiden tunnistusmatkat, ks.

[http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/esimerkit\\_kaavoihin.pdf](http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/esimerkit_kaavoihin.pdf)

Esim. 7.6.8. Esimerkin 5.1.30 kovuusindeksien erotukset -5, 1, -2, -5, 2, -7, -1, -7, 1, 0, joista keskiarvo -2,3 ja keskihajonta 3,4.

Odotusarvon 95 %:n luottamusväli

$$-2,3 \pm 2,262 \cdot 3,4 / \sqrt{10}$$

$$-2,3 \pm 2,4$$

Lisäaineilla ei eroja, koska nolla kuuluu luottamusvälille.

#### Descriptives

		Statistic	Std. Error
Erotus	Mean	-2,3000	1,08577
	95% Confidence Interval for Mean	Lower Bound Upper Bound	-4,7562 ,1562
	Std. Deviation	3,43350	

### 7.6.3 Kahden populaation odotusarvon erotuksen luottamusväli

Jos halutaan arvioida kahden populaation odotusarvojen yhtäsuuruutta, niin voidaan arvioida odotusarvojen erotusta  $\mu_1 - \mu_2$ .

Esim. 7.6.10. Arvio lähiö- ja keskusta-asuntojen keskineliöhintojen erotukselle on  $(-989,844, -798,862)$ , muodostettu luottamusväli odotusarvojen erotukselle. Koska nolla ei kuulu luottamusvälille, niin päätellään: odotusarvojen olevan eri suuret (keskihinnat eivät samoja).

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=78>

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Neliöhinta	Equal variances assumed	1,235	,268	-18,455	227	,000	-894,35342	48,46101	-989,844	-798,862

Aineisto Asunnot\_2006

## 7.7 Hypoteesien testausta

Tutkimusongelmia:

- Puolueen kannatus?

Väite:  $\pi = 18 \%$

- Virheellisten komponenttien osuus tuotannossa?

Väite:  $\pi = 4 \%$

- Kynttilöiden keskimääräinen palamisaika?

Väite:  $\mu = 20 \text{ h}$



- Asuntojen keskimääräiset neliöhinnat keskustassa ja lähiössä?

Väite:  $\mu_1 = \mu_2$

- Painon ja pituuden välinen lineaarinen riippuvuus?

Väite:  $\rho = 0$

- Opintosuunnan vaikutus kurssiarviointiin?

Väite: ei riippuvuutta

Tilastollinen hypoteesi on väite populaatiosta, usein populaation jakauman parametrissa.

Esim.  $H_0: \pi = \pi_0$

$H_0: \mu = \mu_0$

Hypoteesin testaus on väitteen tutkimista otoksen perusteella.

Käytetään sopivaa otossuuretta (testisuuretta), jonka jakauma tunnetaan, kun  $H_0$  tosi.

Esim.

$$H_0: \pi = \pi_0$$

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

$$H_0: \mu = \mu_0$$

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

Otossuureen (testisuureen) arvon perusteella  $H_0$  hyväksytään tai hylätään.

Jos testisuureen arvoa pidetään tavanomaisten arvojen joukkoon kuuluvaksi, niin  $H_0$  hyväksytään. Jos arvoa pidetään harvinaisten arvojen joukkoon kuuluvaksi, niin  $H_0$  hylätään.

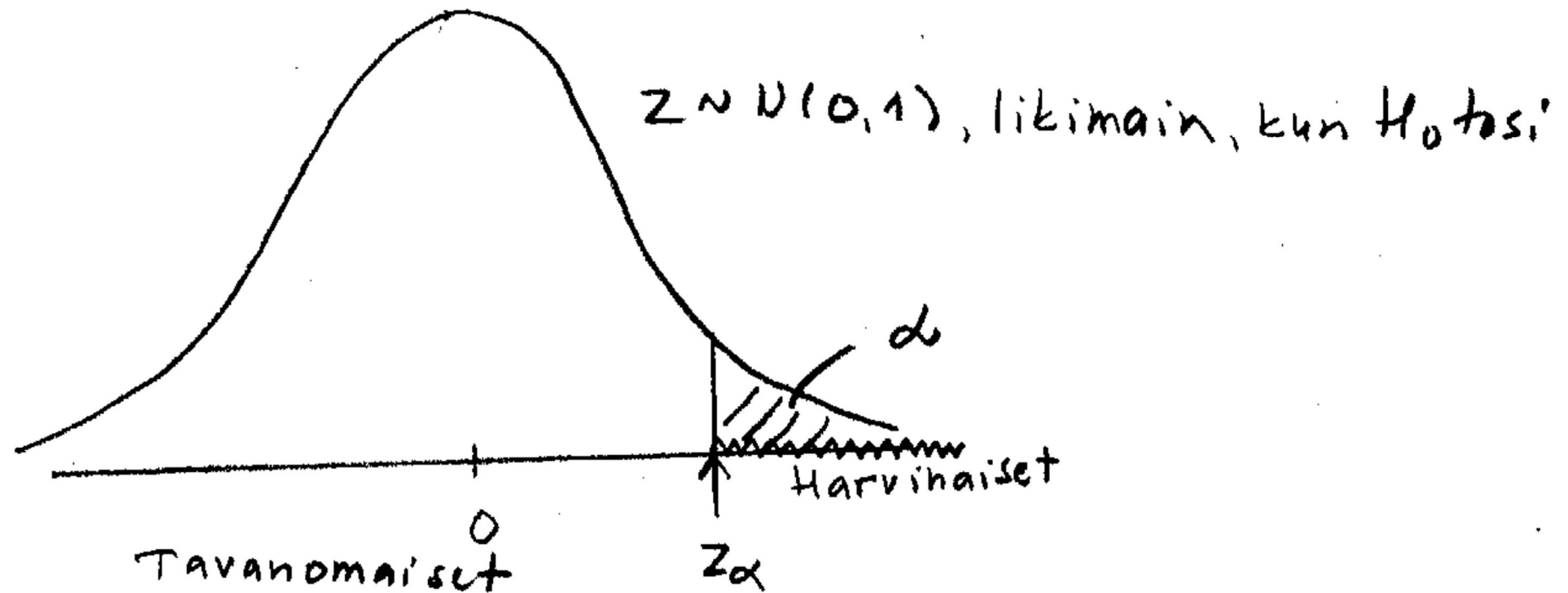
Mikä on harvinaista? Mihin  $H_0$ :n hylkääminen johtaa?

Esim.  $H_0: \pi = \pi_0$

$H_1: \pi > \pi_0$  vaihtoehtoinen hypoteesi

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

Nyt harvinaisiksi arvoiksi katsotaan "suuret arvot", suuremmat kuin  $z_\alpha$ . Tällöin merkitsevyys- eli riskitaso on  $\alpha$ . Usein  $\alpha = 0,05, 0,025, 0,01$  tai  $0,001$ .



Jos  $H_0$  hylätään, niin  $H_1$  hyväksytään.

Miten  $H_1$  asetetaan?

Esim.  $H_0: \pi = \pi_0$

$H_1: \pi > \pi_0$  (yksisuuntainen testi) tai

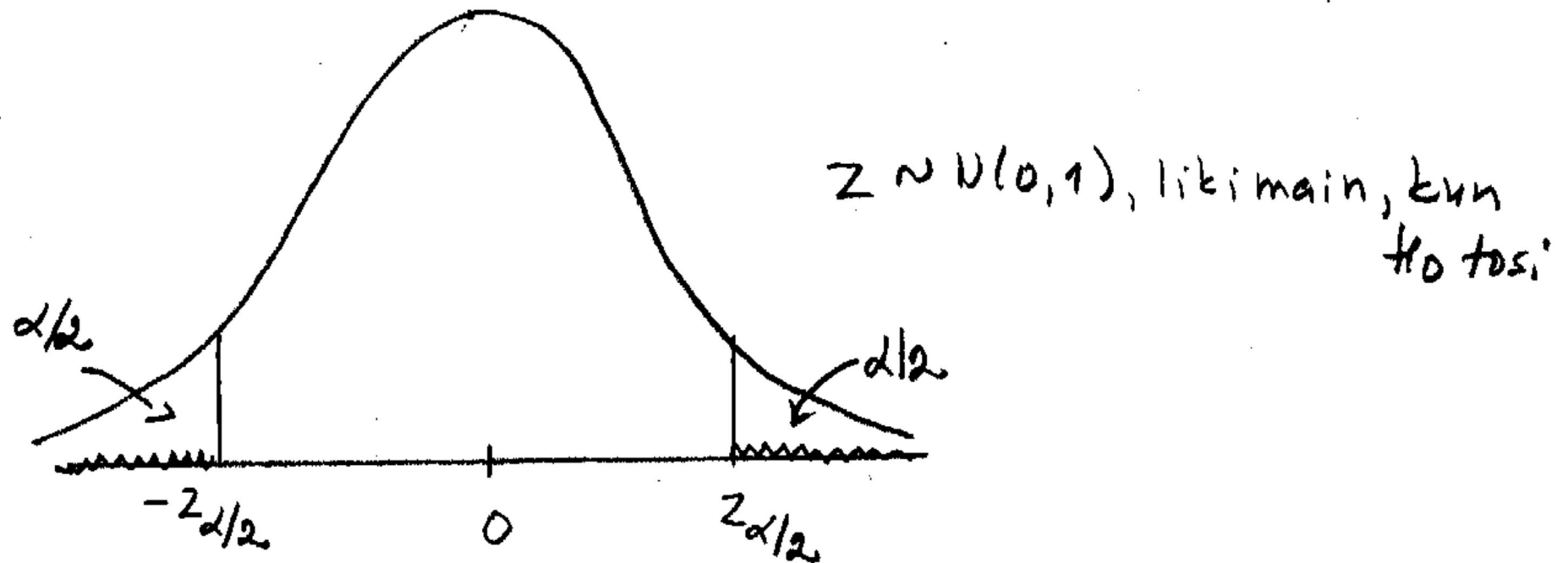
$H_1: \pi < \pi_0$  (yksisuuntainen testi) tai

$H_1: \pi \neq \pi_0$  (kaksisuuntainen testi)

Kaksisuuntaisessa testissä harvinaisten arvojen joukko on jakauman "hännillä".

Esim.

$$H_0: \pi = \pi_0$$

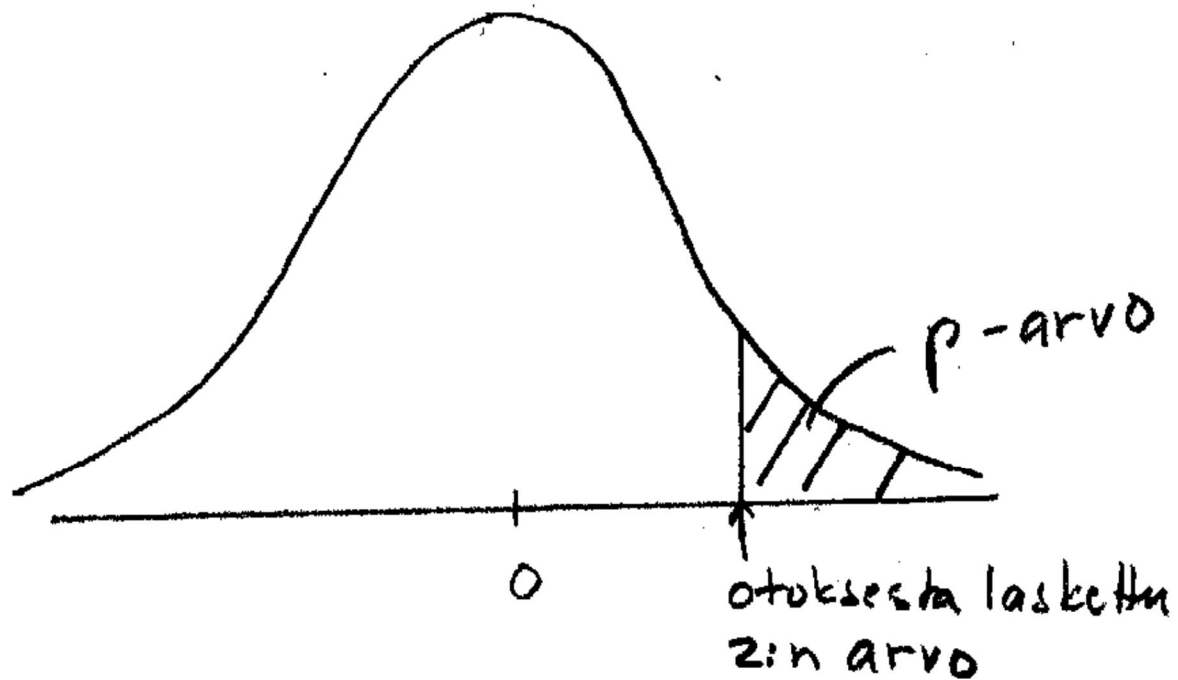
$$H_1: \pi \neq \pi_0 \text{ (kaksisuuntainen testi)}$$


Päätätely p-arvon perusteella, p-arvo on pienin riskitaso, jolla  $H_0$  voidaan hylätä.



Esim.  $H_0: \pi = \pi_0$   
 $H_1: \pi > \pi_0$

Jos p-arvo on pienempi kuin valittu riskitaso, niin  $H_0$  hylätään.



## Testisuureita (seuraavalla luennolla)

$$H_0: \pi = \pi_0$$

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

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

Nyt

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

$$H_1: \pi > 50 \%.$$

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>