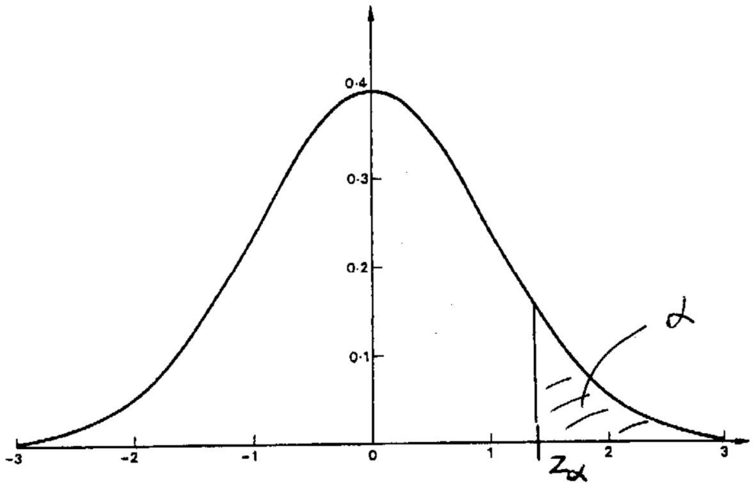
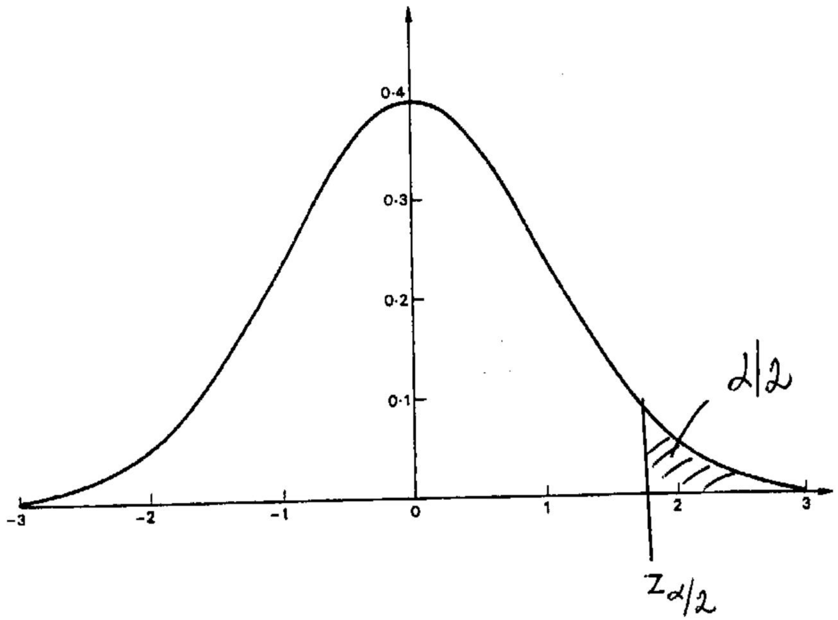
MTTTP1, luento 2.10.2018

7.4 Normaalijakauma (kertausta)

Olkoon $Z \sim N(0, 1)$. Määritellään z_α siten, että $P(Z \geq z_\alpha) = \alpha$, graafisesti:



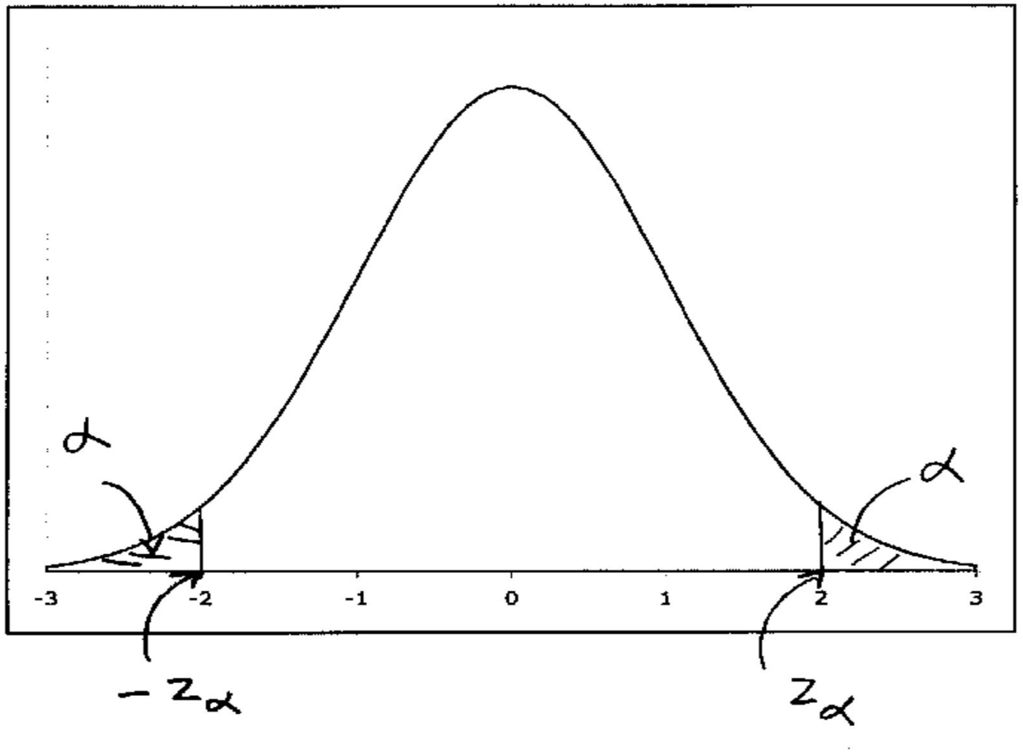
Samoin $z_{\alpha/2}$ siten, että $P(Z \geq z_{\alpha/2}) = \alpha/2$, graafisesti:



Esim.

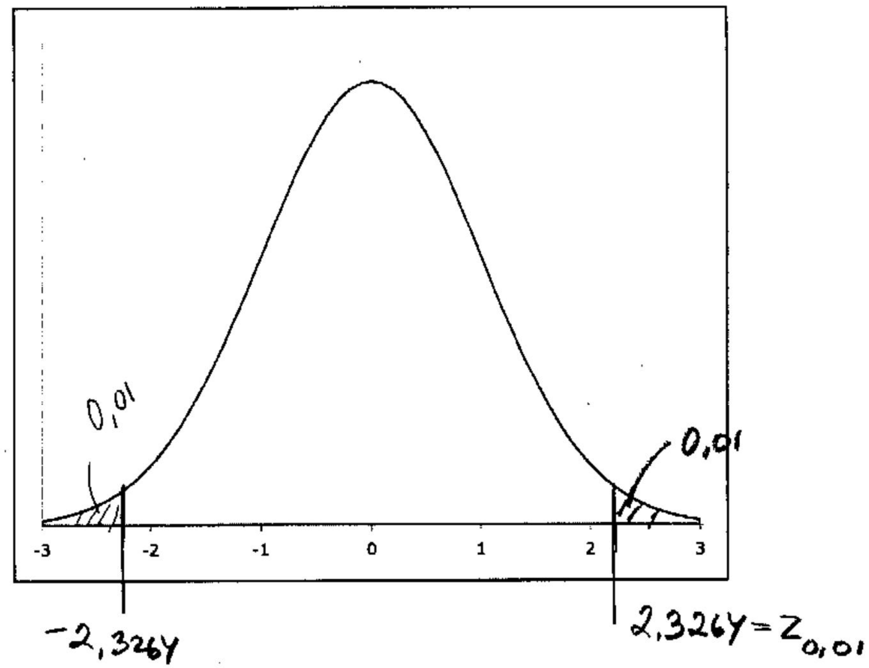
$$z_{0,05} = 1,6449$$
$$z_{0,01} = 2,3264$$
$$z_{0,05/2} = z_{0,025} = 1,96$$

Standardoitu normaalijakauman symmetrinen nollan suhteen



Esim. 7.4.4. Olkoon $Z \sim N(0, 1)$.

$P(Z \geq 2{,}3264) = 0{,}01$, $P(Z \leq -2{,}3264) = 0{,}01$



$P(Z \geq 1{,}96) = 0{,}025$, $P(Z \leq -1{,}96) = 0{,}025$

$P(Z \geq 1{,}6449) = 0{,}05$, $P(Z \leq -1{,}6449) = 0{,}05$

7.5 Satunnaisotos, otossuure ja otantajakauma

Päätelmät populaatiosta otoksen perusteella

- puolueen kannatus
- kynttilöiden keskimääräinen palamisaika
- asuntojen keskimääräiset neliöhinnat keskustassa ja lähiössä

Miten päättely tehdään? Miten tulosten luotettavuutta voidaan arvioida?

Päättely tehdään satunnaisotoksen perusteella.

Satunnaismuuttujajono $X_1, X_2, ..., X_n$ on satunnaisotos, jos X_i:t ovat riippumattomia ja noudattavat samaa jakaumaa.

Esim. Satunnaisotos $X_1, X_2, ..., X_n$ normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin jokainen X_i noudattaa normaalijakaumaa parametrien μ, σ^2 ja X_i:t ovat toisistaan riippumattomia.

Otossuure on satunnaisotoksen perusteella määritelty funktio.

Olkoon satunnaisotos X_1, X_2, ..., X_n normaalijakaumasta $N(\mu, \sigma^2)$, tällöin

$$\bar{X} \sim N(\mu, \sigma^2/n), \text{ kaava (6).}$$

Otoskeskiarvo on otossuure, jonka todennäköisyysjakauma tiedetään. Se on normaalijakauma, havainnollistaminen simuloiden

http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Olkoon populaatiossa π % tietyn tyyppisiä alkioita ja p = tietyn tyyppisten alkioiden % -osuus otoksessa.

Tällöin

$$p \sim N(\pi, \pi(100-\pi)/n), \text{ likimain, kaava (7).}$$

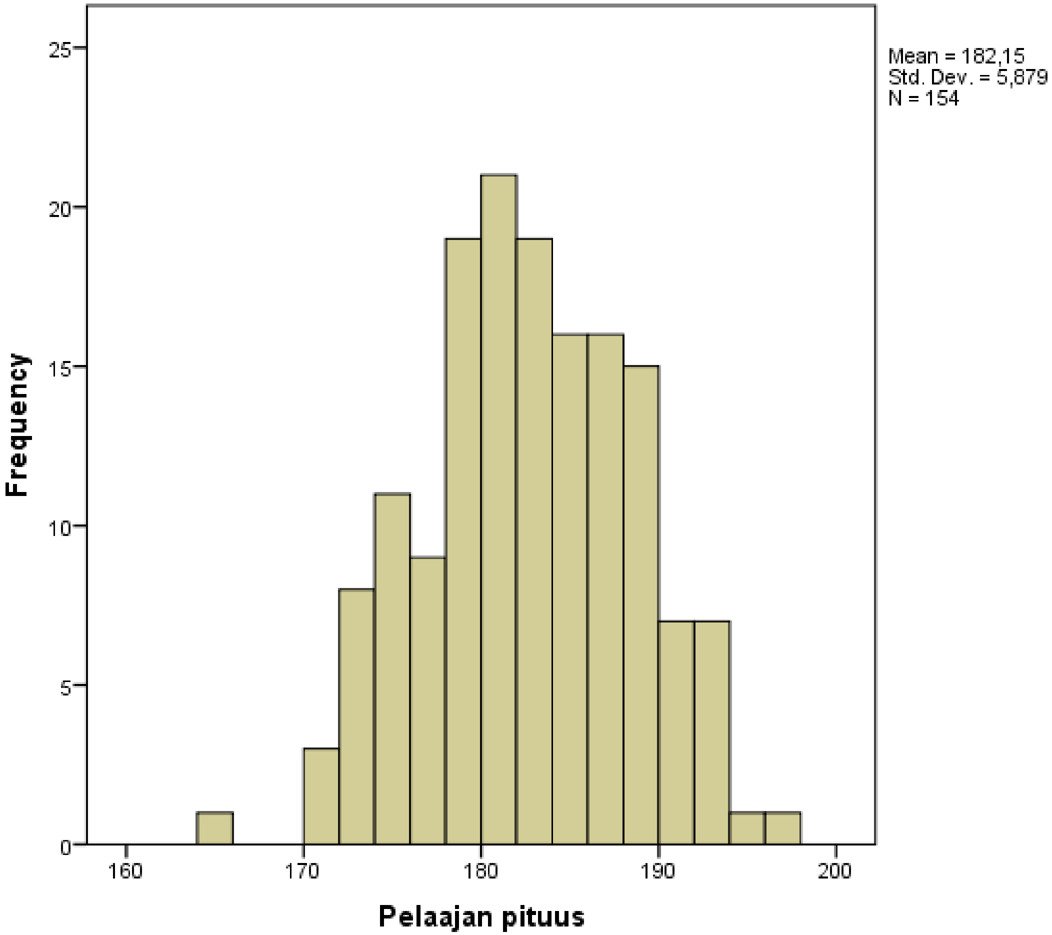
Viallisten prosenttiosuus otoksessa (p) on otossuure, jonka jakauma on likimain normaalijakauma.

Otossuureiden jakaumia käytetään päättelyyn liittyvien tulosten luotettavuuden arvioinnissa.

7.6 Piste-estimointi ja luottamusvälejä

Esim. Vuonna 2007 suomalaisen miesten keskipituuden arvioitiin olevan 179,6 cm, naisten 165,9, http://fi.wikipedia.org/wiki/Ihmisen_pituus#Ihmisten_keskipituus_eri_maissa

Esim. Jalkapalloilijat 2006, jalkapalloilijoiden keskipituuden arviointi. Arvioidaan keskipituuden olevan 182,15 cm.



Esim. Puolueen kannatusarviot,
https://yle.fi/uutiset/3-10387592 (6.9.2018)

Arvioidaan SDP:n kannatuksen olevan 20,3 %.

Estimointi

populaation tuntemattoman parametrin arviointia otossuureen avulla (piste-estimointi)

Estimaattori

otossuure, jolla estimoidaan tuntematonta parametria

Estimaatti

estimaattorin arvo (tehdyn otoksen perusteella laskettu)

Estimaattorin keskivirhe

estimaattorin hajonta

Estimoitava parametri	Esti-maattori	Estimaattorin keskivirhe	Estimoitu keskivirhe
μ	$\bar{X}$	$\sigma/\sqrt{n}$	$s/\sqrt{n}$
π	p	$\sqrt{\pi(100-\pi)/n}$	$\sqrt{p(100-p)/n}$
σ	s		

Esim. Puolueen kannatuksen arviointi

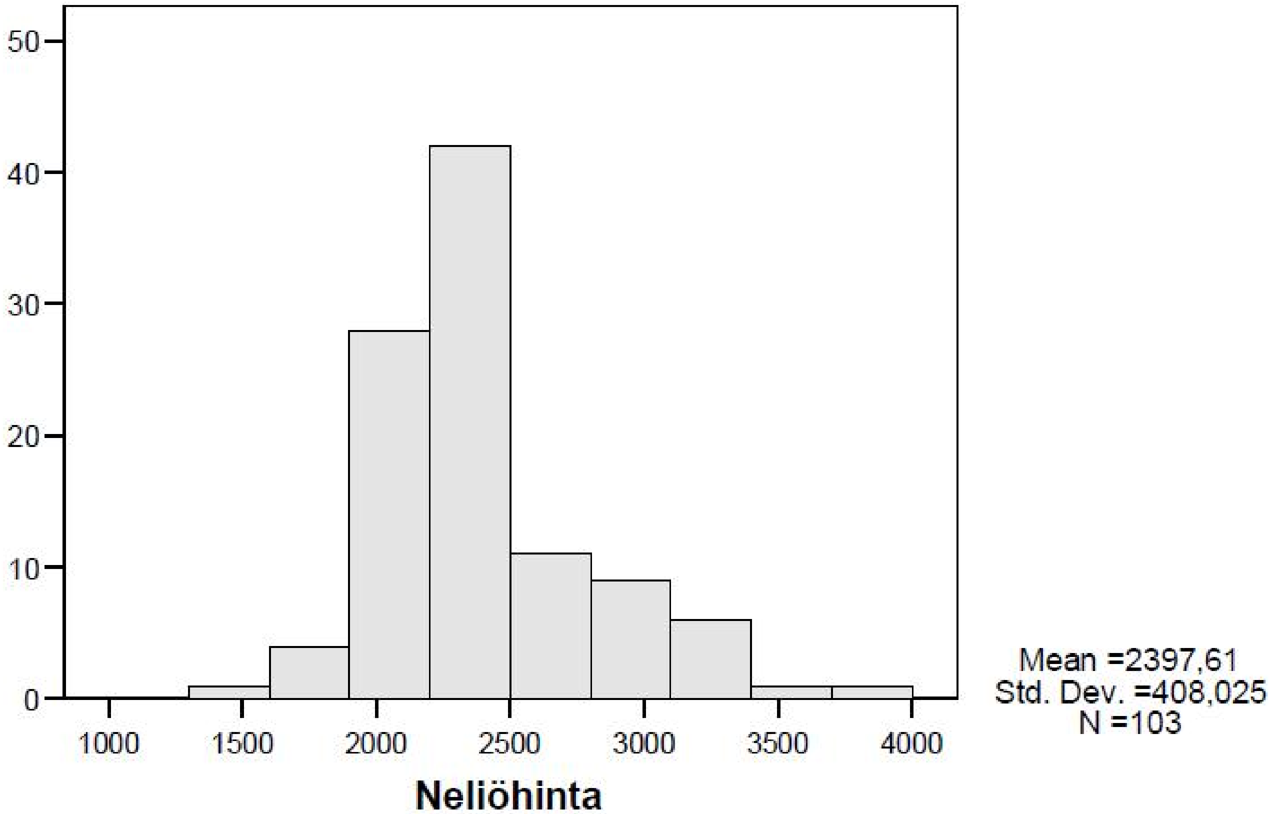
p = 20,3 %,

n = 1460 (kantansa ilmoittaneet).

Kannatuksen estimoitu keskivirhe

$$\sqrt{20{,}3(100-20{,}3)/1460} = 1{,}05.$$

Esim. 7.6.1. Kerrostalohuoneistojen keskimääräisen neliöhinnan estimointi, $\bar{x}$ = 2398, s = 408, joten estimoitu keskivirhe on $408/\sqrt{103}$ = 40,2.



Esim. Jalkapalloilijoiden keskipituuden estimoitu keskivirhe $5{,}879/\sqrt{154} = 0{,}474$.

Myös nk. luottamisvälin avulla voidaan arvioida populaation tuntematonta parametria, tällöin kyse väliestimoinnista.

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

Luottamusväli on satunnaisväli, joka sisältää estimoitavan parametrin todennäköisyydellä $1 - \alpha$. Valitaan α esim. 0,05 tai 0,01. Tällöin kyse 95 %:n tai 99 %:n luottamusvälistä.

7.6.1 Prosenttiosuuden luottamusväli

Kaava (8), 100(1 - α) %:n luottamusväli prosenttiosuudelle

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

95 %:n luottamusväli, α = 0,05, $z_{0,05/2} = z_{0,025} = 1,96$

99 %:n luottamusväli, α = 0,01, $z_{0,01/2} = z_{0,005} = 2,5758$

Esim. 7.6.4. Satunnaisesti valituista 100 henkilöstä puoluetta kannatti 18 %. Puolueen kannatuksen 95 %:n luottamusväli

$$18 \pm 1{,}96\sqrt{18(100-18)/100}$$

$$18 \pm 7{,}5$$

Arvioidaan kannatuksen olevan välillä 10,5 – 25,5. Virhemarginaali ±7,5 %-yksikköä.

Esim. Puolueen kannatusarviot ja virhemarginaali, Puolueen kannatusarviot, Puolueen kannatusarviot, https://yle.fi/uutiset/3-10387592 (6.9.2018)

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

Nyt n = 250, p = 100·25/250 = 10

95 %:n luottamusväli hinnan perusteella valintansa tekevien prosenttiosuudelle

$$10 \pm 1{,}96\sqrt{10(100-10)/250}$$

$$10 \pm 3{,}7$$

Koska 15 ei kuulu luottamusvälille, ei uskota väitettä.

99 %:n luottamusväli

$$10 \pm 2{,}5758\sqrt{10(100-10)/250}$$

$$10 \pm 4{,}9,$$

sama päättely.

Sivut 20-24 seuraavalle luennolle

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:

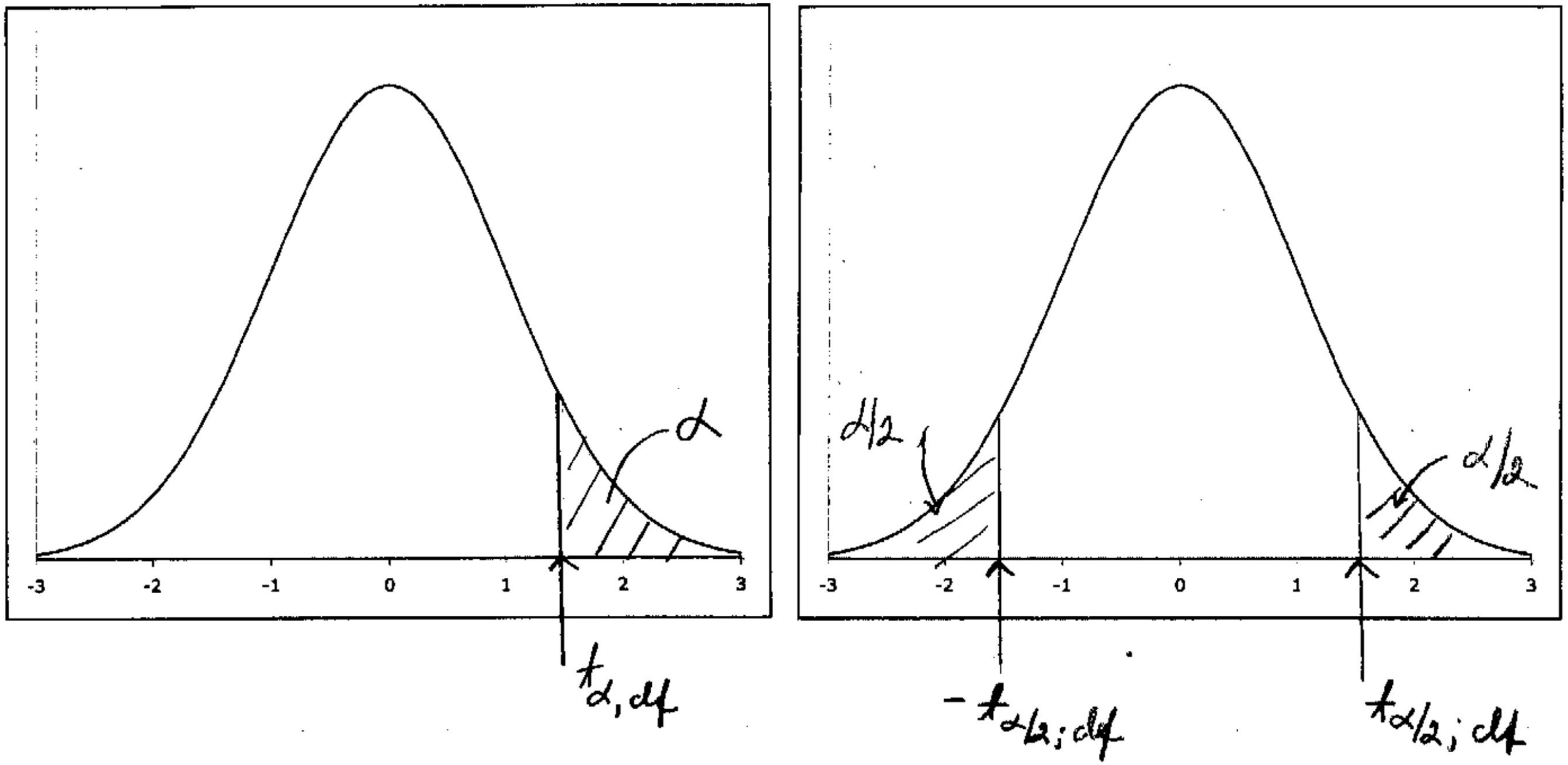
Descriptives

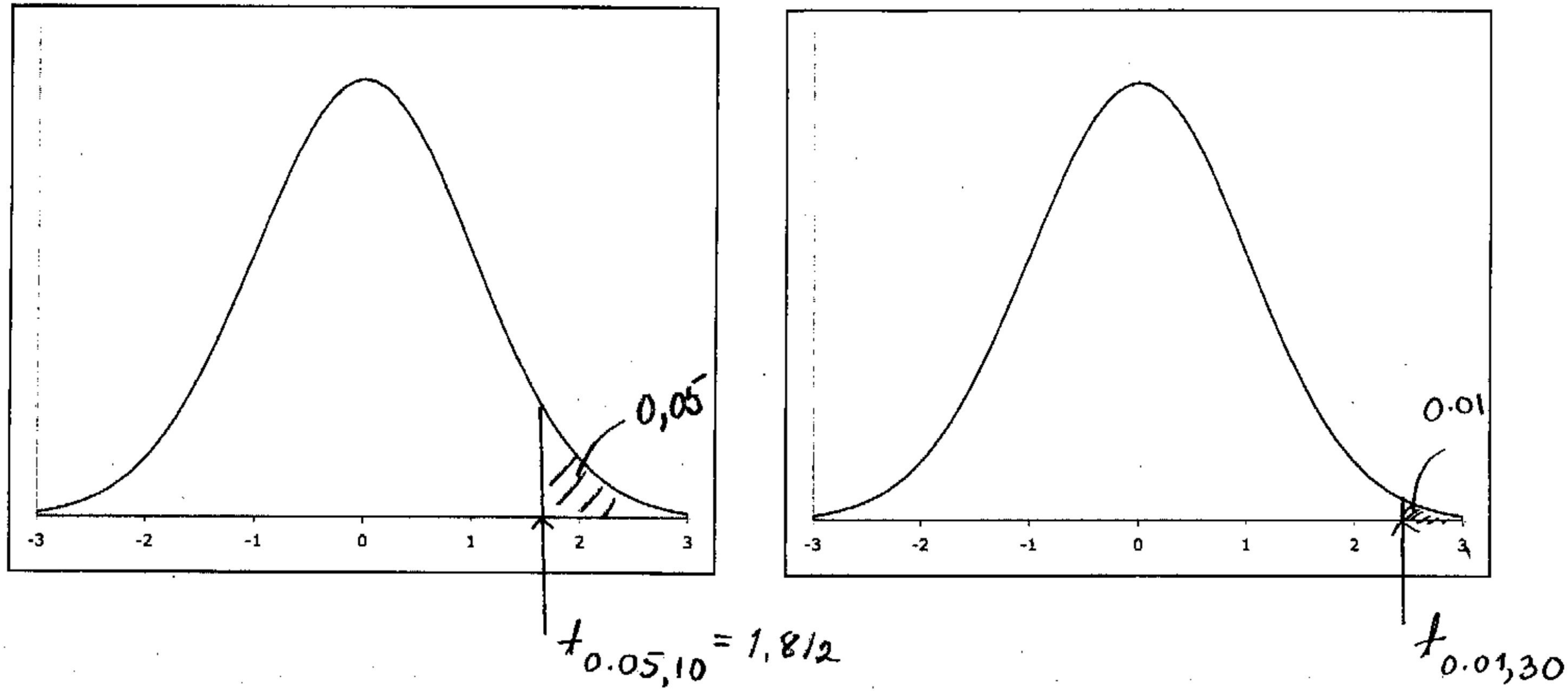
			Statistic	Std. Error
pituus	Mean		50,95	,245
	95% Confidence Interval for Mean	Lower Bound	50,47	
		Upper Bound	51,44	
	Std. Deviation		1,972	

Kaava (9), 100(1 - α) %:n luottamusväli odotusarvolle

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

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




0,05
-3
-2
-1
0
1
2
3
t0.05,10 = 1,812
0.01
-3
-2
-1
0
1
2
3
t0.01,30

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 35,5 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.