

Anvendt Statistik

Lektion 3

Punkt- og intervalestimer

Konfidensintervaller

Valg af stikprøvestørrelse

Punkt- og intervalestimer: Motivation

■ Motiverende eksempel:

- I en undersøgelse er andelen af rygere 0.27. Det anslås at populationsandelen er mellem 0.25 og 0.29.

■ Begreber:

- De 0.27 er et **punkt-estimat**
- Intervallet 0.25 til 0.29 er et **interval-estimat**.
- Dvs. populationsandelen falder (anslået) indenfor **punkt-estimat +/- fejl-margin**.
- **Fejl-margin** er her 0.02

Punkt-estimat og -estimator

- En **estimator** er en "generel formel", der bruges til at estimere en parameter med, fx.

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

- Et **estimat** er en konkret udregning af en estimator, ved at indsætte data.
- Der kan være **mange estimatore for den samme parameter**. Hvis populationsfordelingen er symmetrisk er både stikprøve-median og -gennemsnit estimatore for populations-middelværdien.

En god estimator

- En god estimator har typisk følgende egenskaber:
 - Den er **Unbiased** – dvs. at estimatoren i gennemsnit er lig parameteren.
 - Den er **Efficient** – dvs. fejl-margin bliver mindre jo mere data der er til rådighed.
- Stikprøve-medianen er en **biased** estimator for populationsmiddelværdien, hvis fordelingen ikke er symmetrisk.

Notation – en på hatten

- Stikprøve-gennemsnittet $\bar{y}$ er en estimator for populationsmiddelværdien μ .
- Generelt vil vi betegne en estimator med en "hat" $\wedge$.
- Fx. betegner $\hat{\mu}$ en estimator for μ .

Konfidensinterval

■ Motivation:

- Ifølge undersøgelse: 54% er vilde med pålægschokolade!
- **Spørgsmål:** Hvor sikkert er dette estimat?

■ Konfidensinterval

- Et konfidensinterval angiver et interval, hvor vi tror parameteren ligger. Sandsynligheden for at konfidensintervallet indeholder parameteren betegnes **konfidensniveauet**. Konfidensniveauet er typisk 0.95 eller 0.99.

Konfidensinterval: Typisk opskrift

- I mange tilfælde er stikprøvefordelingen for estimatoren (tilnærmelsesvis) normalfordelt. Fx stikprøvegennemsnittet.
 - I disse tilfælde er konfidensintervallet givet ved
punkt-estimat fejl-margin
 - **Spørgsmål:** Hvordan finder vi fejl-margin?
-

Konfidensinterval for andele

■ Notation

□ π : populations-andel

□ $\hat{\pi}$: stikprøve-andel

■ Bemærk: $\hat{\pi}$ er en estimator for π .

■ Antag

□ $y = 1$: succes / vild med pålægschokolade

□ $y = 0$: fiasko / ikke vild med pålægschokolade

■ Vi har

$$P(1) = \pi \quad \text{og} \quad P(0) = 1 - \pi.$$

■ Middelværdi og standard-afvigelse for y (populationen) er hhv.
 $\mu = \pi$ og $\sigma = \sqrt{\pi(1-\pi)}$

Andelen er et gennemsnit

- Bemærk:

$$\hat{\pi} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{\sum_i y_i}{n}$$

Dvs. **stikprøve-andelen er et gennemsnit.**

- For gennemsnit ved vi at standard-fejlen er $\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$
så for stikprøve-andelen er den

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Konfidensinterval for π for stort n

- Da $\hat{\pi}$ er et gennemsnit siger CLT, at $\hat{\pi}$ ca. følger en normalfordeling hvis n bare er stor nok.

- Med 95% sandsynlighed vil $\hat{\pi}$ falde i intervallet

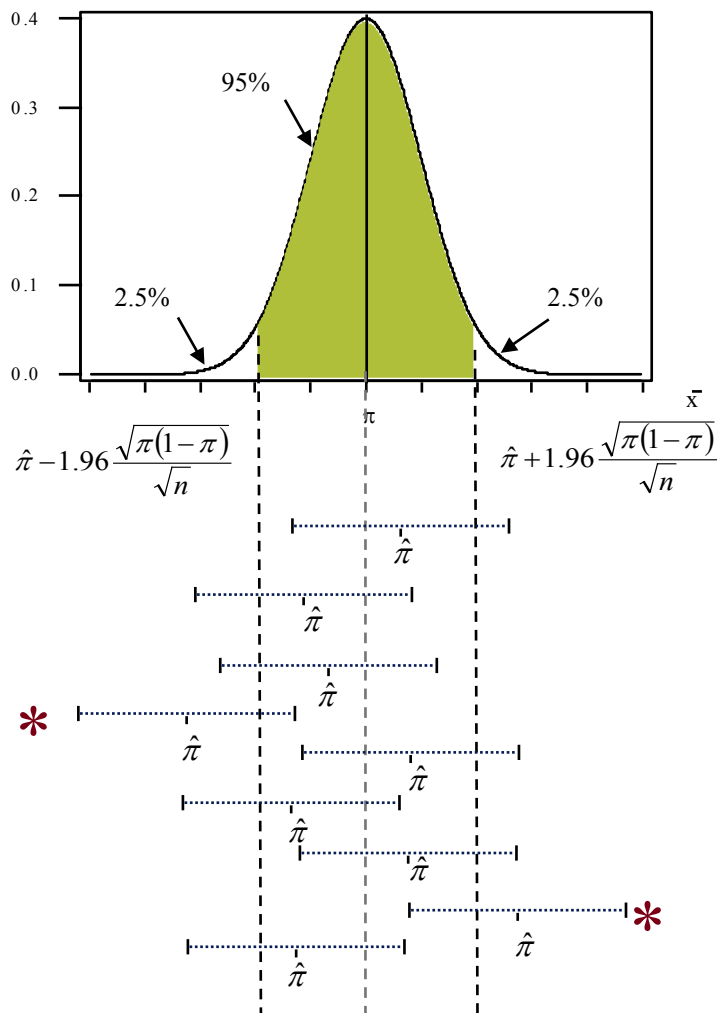
$$\pi \pm 1.96 \cdot \sigma_{\hat{\pi}}$$

- Omvendt: Med 95% sandsynlighed vil π ligge i intervallet

$$\hat{\pi} \pm 1.96 \cdot \sigma_{\hat{\pi}}$$

Konfidens-interval: En figur

Stikprøvefordelingen for $\hat{\pi}$



Konfidensinterval

- I praksis kender vi ikke π , dvs. vi kender ikke standard fejlen:

$$\sigma_{\hat{\pi}} = \sqrt{\pi(1-\pi)/n}$$

- I stedet for π bruger vi estimatet $\hat{\pi}$:

- Et 95% konfidens-interval for π er nu givet ved

$$\hat{\pi} \pm 1.96 \cdot se$$

hvor

$$se = \sqrt{\hat{\pi}(1-\hat{\pi})/n}$$

- se = "standard error"

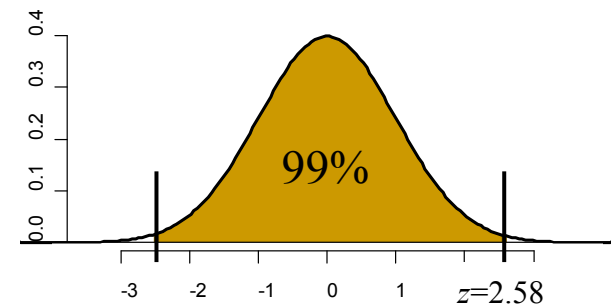
Eksempel

- Af 1200 adspurgte i Florida svarer 396 ja til reduktion af abortrettigheder. Find et 95% konfidensinterval for populationsandelen af ja-sigere.
- $\hat{\pi} =$
- $se =$
- 95% konfidensinterval:

Hvad med et 99% konf. interval?

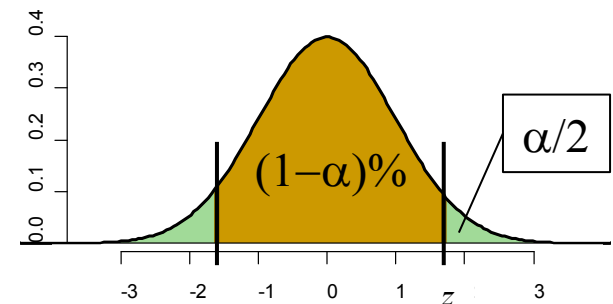
- Et 99% konfidensinterval:

$$\hat{\pi} \pm 2.58 \cdot se$$



- Et $(1-\alpha)100\%$ konfidensinterval

$$\hat{\pi} \pm z \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$



- **Egenskaber ved konfidensintervaller:**

- Jo højere konfidensniveau, jo større z og jo længere konfidens-interval
- Jo større stikprøve (n) jo kortere konfidensinterval
- Firdobling af n sikre halvering af konfidensinterval.

Kofidensinterval for middelværdi

- Ingredienser:
 - μ : Populations-middelværdi
 - $\bar{y}$: Stikprøve-gennemsnit
- $\bar{y}$ er et punktestimat for μ .
- For store stikprøver er $\bar{y}$ normalfordelt. Standardfejlen er altid

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

hvor σ er standard-afvigelsen for populationen.

- Den estimerede standard-fejl er

$$se = \frac{s}{\sqrt{n}}$$

Eksempel

- Konfidensinterval for middelværdi er

$$\bar{y} \pm z \cdot se \quad , \text{ hvor } se = \frac{s}{\sqrt{n}}$$

- **Eksempel:**
- På et spørgsmål om antal seksuelle partnere blandt $n = 231$ kvinder, var gennemsnittet $\bar{y} = 4.96$ og standard-afvigelsen 6.81.
- Find et 95% konfidensinterval for populations-middelværdien μ .

Konfidensinterval for middelværdi – små stikprøver

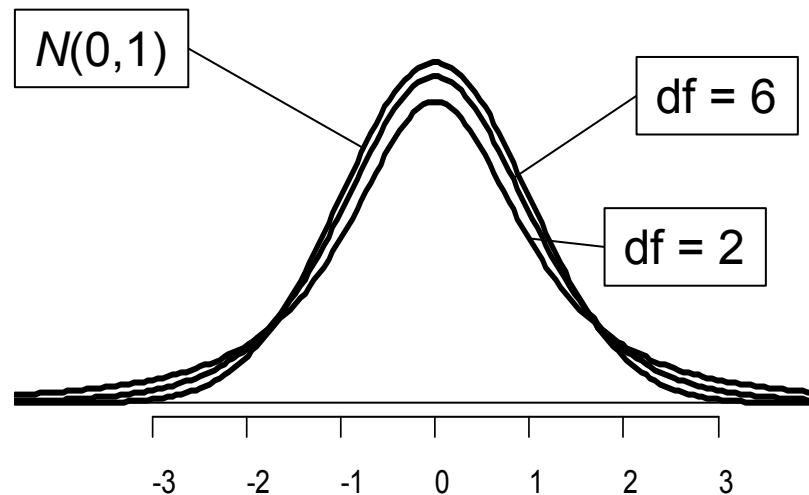
- **Antag:** populationen er normal-fordelt.
- Da er y normalfordelt – uanset stikprøve-størrelse n .
- Hvis vi **kender pop. standard-afvigelsen** σ er et (eksakt) konfidensinterval givet ved

$$\bar{y} \pm z \frac{\sigma}{\sqrt{n}}$$

- Hvis **σ er ukendt**, erstatter vi med stikprøve standard-afvigelsen s .
- **Problem:** For små stikprøver medfører dette en ekstra usikkerhed.
- **Løsning:** Erstat z med t ...!

t -fordelingen

- t -fordelingen er
 - Klokkeformet og symmetrisk omkring 0
 - Standard-afvigelsen er lidt større end 1
 - Faconen afhænger af antal *frihedsgrader* (df).
 - Har lidt "tykkere haler" end standard normalfordelingen.
 - Ligner en normalfordeling for $df \geq 30$.

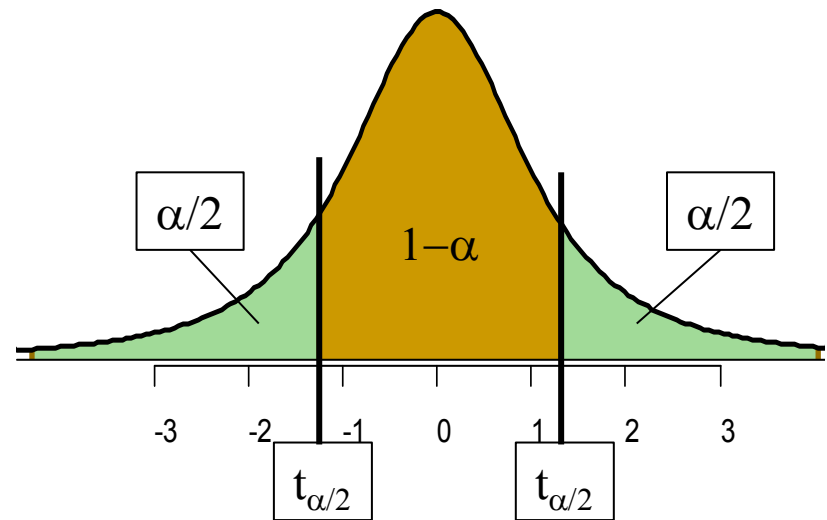


Konfidensinterval for små stikprøver

- For en normalfordelt population er et $(1-\alpha)100\%$ konfidensinterval for μ

$$\bar{y} \pm t_{\alpha/2} \cdot se, \text{ hvor } se = \frac{s}{\sqrt{n}}$$

hvor $df = n-1$.



- **Eksempel:**

Vi har observeret 29 vægtændringer, hvor $\bar{y} = 3.01$ og $s = 7.31$.
Find et 95% konf. interval for μ :

Løsning:

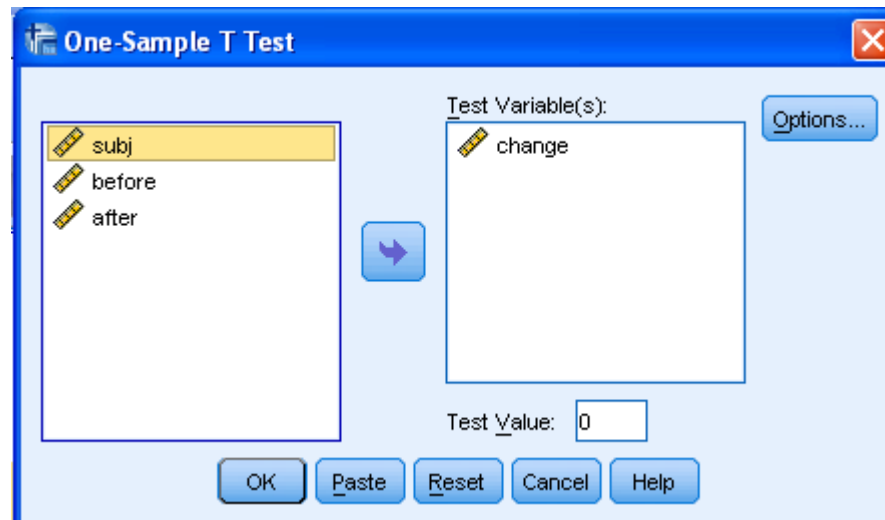
$df = n-1 = 28$, $\alpha = 0.025$, så $t_{0.025} = 2.048$

t -tabellen

df	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
	Right-Tail Probability					
	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396

I SPSS

- SPSS: Analyze→Compare Means→One-Sample T-Test...



One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
change	2,216	28	,035	3,00690	,2269	5,7869

Valg af stikprøvestørrelse

- Hvordan **vælger man stikprøvestørrelse n** så vi opnår en
 - **Given fejl-margin** ved et
 - **Givet konfidensniveau**

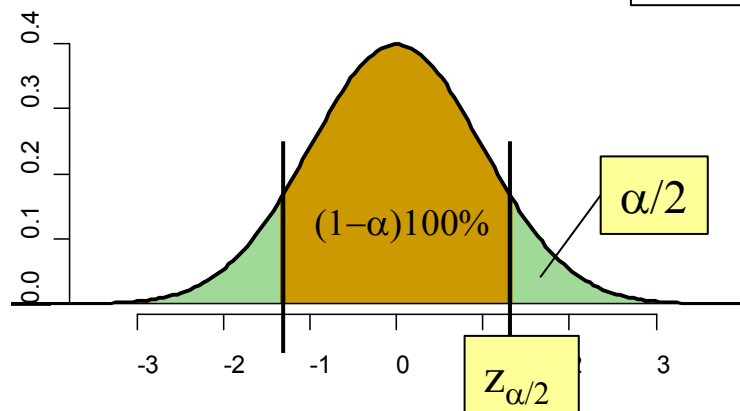
- **Eksempel:**
 - Vi ønsker at bestemme et konfidensinterval for π , så
 - Fejl-margin : Max 0.04
 - Konfidensniveau : 95%
 - Løsning:

Generel løsning for andele

- For at populationsandel π vælges
 - Fejl-margin: M
 - Signifikansniveau: $(1-\alpha)100\%$
- Stikprøvestørrelsen skal da være:

$$n = \pi(1 - \pi) \left(\frac{z}{M} \right)^2$$

Hvis populations-andelen π er helt og aldeles ukendt bruges $\pi = 0.5$ i formlen.

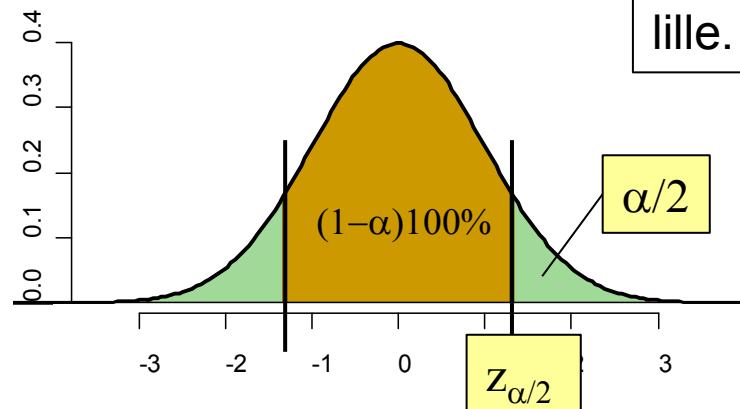


Generel løsning for middelværdi

- For middelværdien μ vælg
 - Fejl-margin: M
 - Signifikansniveau: $(1-\alpha)100\%$
- Stikprøvestørrelsen skal da være:

$$n = \sigma^2 \left(\frac{z}{M} \right)^2$$

Hvis populationsstandardafvigelsen σ er man nød til at gætte sig frem til. Hellere lidt for stor end for lille.



Eksempel

- Middel antal års uddannelse blandt indianere
- Ønsker:
 - Fejl-margin: $M = 1\text{ år}$
 - Konfidensniveau: 99%
- Først skal vi gætte σ !
- Vi tror (næsten) alle har mellem 5 og 20 års uddannelse...
- Derfor er vores gæt $\sigma = 2.5\text{ år...!}$