

FORELESNING

15

HET 3431

EIVIND ERIKSEN

MAR 13 2012

STATISTIKK

PLAN:

- ① Repetisjon: Kap 6 (Normalfordeling)
Kap 7 (Konfidensintervall) < $\begin{matrix} \text{populasjons andel} \\ \text{gjennomsnitt} \end{matrix}$
- ② Oppgaver fra Arbeidskrav 5
- ③ Konfidensintervall i SAS JMP
- ④ Hypotesetesting [T] kap. 8

Husk:

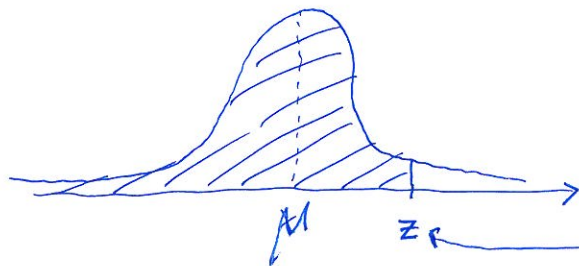
(a) Tidligere eksamensoppgaver er lagt ut i H's Learning.

(b) Studentveiledning:

20/03	kl. 17.15 - 20.15
10/04	— " —
24/04	— " —
08/05	— " —

① Repetisjon:

Normalfordeling:



z standard normal-fordelt

Hvis X er normal fordelt med parametre μ, σ ,
Så regner vi ut z-verdier:

$$Z = \frac{X - \mu}{\sigma}$$

Vi finner skravert areal

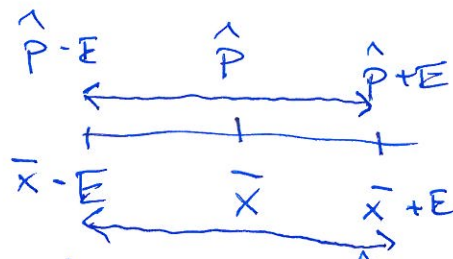
= Sannsynligheten for at variabelen Z har verdi mindre enn eller lik gitt

z-verdi

$$= G(z)$$

i Tabell A-2

Konfidensintervall:



a) Populasjonsandel p :

$$\hat{p} \pm E = p \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

b) Gjennomsnitt $\bar{x}$:

$$\bar{x} \pm E = \bar{x} \pm t_{\alpha/2}^{(n-1)} \cdot \frac{s}{\sqrt{n}}$$

I tillegg a) bruker vi $Z_{\alpha/2}$ fra std. normalfordeling Tabell A-2

-||- b) -||- $t_{\alpha/2}^{(n-1)}$ fra t-fordeling Tabell A-3
(med $n-1$ frihetsgrader)

Felles for begge:

Intervall bredde = $2 \cdot E$

Midtpunkt = punktestimat ($\hat{p}$ eller $\bar{x}$)



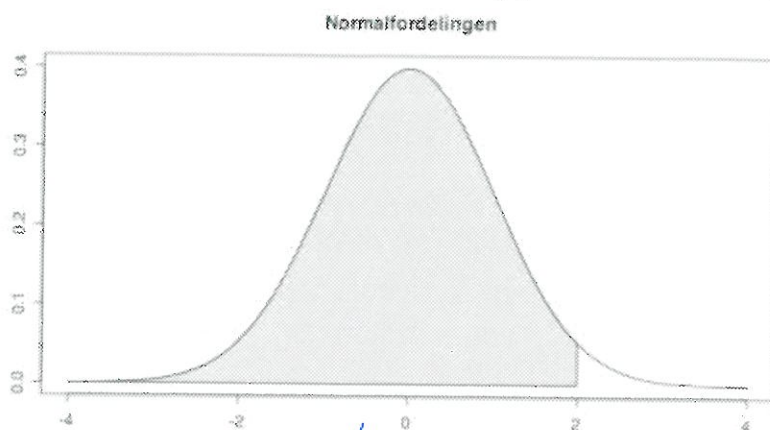
Arbeidskrav 5

[Back to Arbeidskrav 5 - result report for Eriksen, Eivind \(12 March 2012\)](#)

Question 4

Grafen viser standard normalfordelingen.

Finn arealet til det markerte området. Bruk tabell i boka og gi svaret med 4 desimaler. Bruk "." og ikke "," for



desimaltegn.

Your answer:

$$\Phi(2) = P(Z \leq 2) = \underline{0.9772}$$

slå opp i
tabell A-Z



Arbeidskrav 5

Back to Arbeidskrav 5 - result report for Eriksen, Eivind (12 March 2012)

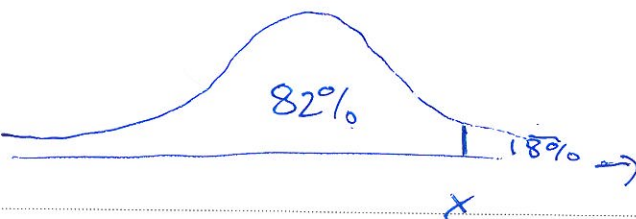
Question 5

Vaskemaskiner av typen XX har en levetid som er normalfordelt med gjennomsnitt 10.8 år og et standardavvik som er 1.2 år. Hvilken levetid er slik at bare 18% av maskinene lever lengre, mens 82% lever kortere? Med andre ord, hva er 82-persentilen?

Your answer

- ☐ 11.9 år
- ☐ 11.0 år
- ☐ 11.1 år
- ☐ 9.7 år

$X = \text{levetid i år}$
normalfordelt : $\mu = 10.8$
 $\sigma = 1.2$



$$z = \frac{x - \mu}{\sigma}$$

$$\sigma z = x - \mu$$

$$\sigma z + \mu = x$$

$$\begin{aligned} x &= \mu + \sigma \cdot z \\ &= 10.8 + 0.915 \cdot 1.2 \\ &\approx \underline{\underline{11.9}} \end{aligned}$$

z
 $\uparrow$
Finn z slik at
 $\Phi(z) = 0.82$
Fra tabell A-2 :
 $\underline{\underline{z = 0.915}}$



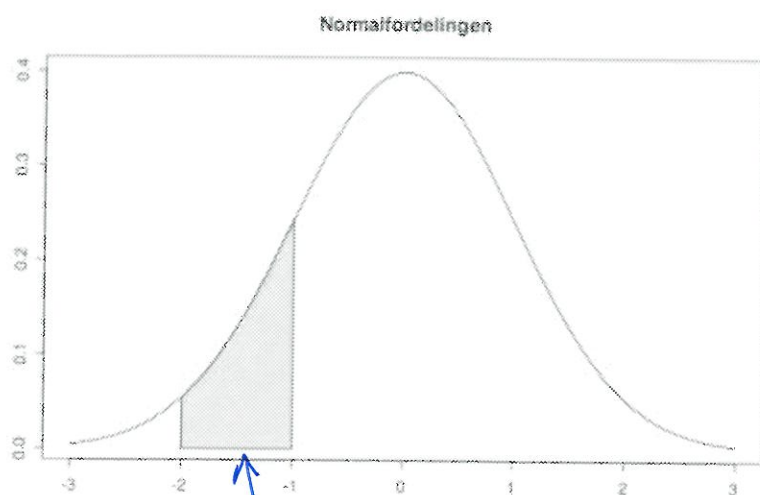
Arbeidskrav 5

[Back to Arbeidskrav 5 - result report for Eriksen, Eivind \(12 March 2012\)](#)

Question 6

Grafen viser standard normalfordelingen.

Bruk tabell til å finne arealet av det markerte området, gi svaret med 4 desimaler. Bruk "." og ikke "," i svaret ditt.



$$P(a \leq Z \leq b) \\ = \Phi(b) - \Phi(a)$$

Your answer:

$$\begin{aligned} & \Phi(-1) - \Phi(-2) \\ & = 0.1587 - 0.0228 = \underline{\underline{0.1359}} \end{aligned}$$



Arbeidskrav 5

[Back to Arbeidskrav 5 - result report for Eriksen, Eivind \(12 March 2012\)](#)

Question 8

Anta at 13 av 54 kunder i et kjøpesenter kjente til en spesiell type majones.

Konstruer et 95% konfidensintervall for andelen i populasjonen som kjenner til majonesen.

$$\alpha = 0.05$$

$$\hat{p} \pm E = \hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Your answer

☐ $0.126 < p < 0.356$

☒ $0.127 < p < 0.355$

☐ $0.144 < p < 0.338$

☐ $0.145 < p < 0.337$

$$\hat{p} = \frac{13}{54} \approx 0.2407$$

$$E = 1.96 \cdot \sqrt{\frac{0.2407 \cdot 0.7593}{54}} \approx 0.1140$$

$$\hat{p} + E = 0.3547 \approx 0.355$$

$$\hat{p} - E = 0.1267 \approx 0.127$$



Arbeidskrav 5

Back to Arbeidskrav 5 - result report for Eriksen, Eivind (12 March 2012)

Question 9

Anta at en populasjon har en normalfordeling og at vi har en stikkprøve med størrelse $n=10$, der gjennomsnittet er $\bar{x} = 14.4$ og standardavviket er $s=3.8$.

Lag et 95% konfidensintervall for populasjonsgjennomsnittet μ .

$\alpha = 0.05, n-1 = 9$

$$\bar{x} \pm E = \bar{x} \pm t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}$$

Your answer

- ☐ $12.20 < \mu < 16.60$
- ☒ $11.68 < \mu < 17.12$
- ☐ $11.72 < \mu < 17.08$
- ☐ $11.69 < \mu < 17.11$

$$\bar{x} = 14.4$$

$$s = 3.8$$

$$E = 2.262 \cdot \frac{3.8}{\sqrt{10}} \approx 2.72$$

$$\bar{x} + E \approx 14.4 + 2.72 = 17.12$$

$$\bar{x} - E \approx 14.4 - 2.72 = 11.68$$

Konfidensintervall i JMP

Konfidensintervall i JMP

- Legg inn 23, 25, 12, 30, 20 i en kolonne i JMP
- *Analyze > Distribution*
- JMP oppgir 95% konfidensintervall for μ

Moments

Mean	22
Std Dev	6.670832
Std Err Mean	2.9832868
Upper 95% Mean	30.282932
Lower 95% Mean	13.717068
N	5

Lengde på telefonsamtaler 2

Example

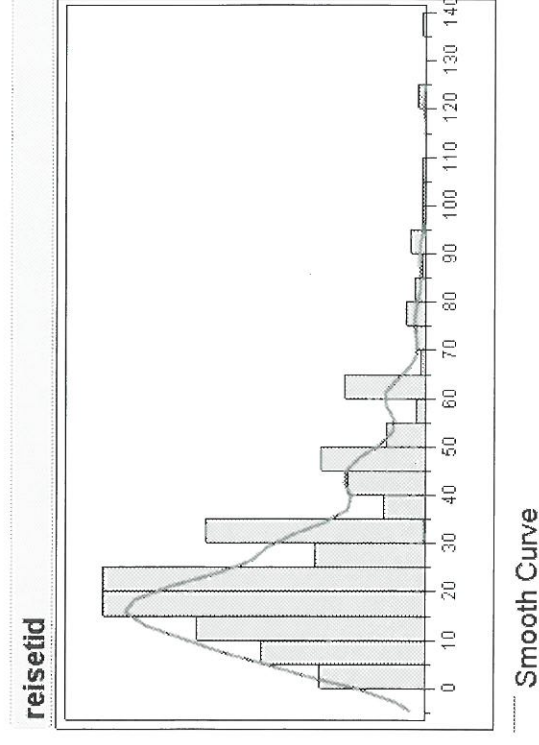
- I en annen bedrift har vi 4, 12, 10, 3, 5 og 8 minutter.
- Lag et 90% konfidensintervall for gjennomsnittlig samtaleid
- Løsning
 - 1 Regn ut $\bar{x} = 7.00$ og $s = 3.578$
 - 2 Finn $t_{\alpha/2} = 2.015$ for 5 frihetsgrader. Tabell A3
 - 3 Feilmarginen:
$$E = 2.015 \cdot \frac{3.578}{\sqrt{6}} = 2.94$$
 - 4 90% konfidensintervall for μ går ifra 7.00 – 2.94 til 7.00 + 2.94:
$$4.06 < \mu < 9.94$$
 - 5 Vi er 90% sikre på at gjennomsnittlig samtaleid er mellom 4.06 og 9.94 minutter

Er lengden på telefonsamtaler normalfordelt?

Forutsetninger for å bruke t-fordeling

- Stikkprøven må være tilfeldig utvalgt
- Originaldataene x er normalfordelt, eller $n > 30$

- I eksemplene over har vi færre enn $n = 30$ objekter i stikkprøven
- Forutsatte derfor at lengden på telefonsamtaler var normalfordelt
- Det er mer sannsynlig at lengden er høyreskjev, siden det finnes en naturlig nedre skranke: 0 minutter for lengden
- Vi skulle nok hatt stikkprøve på minst 30 telefonlengder...



Figur: Eksempel på høyreskjev fordeling. Reisetid til skolen for 1937 BI studenter

Norsk kundebarometer

Example

For variabelen *Anbefale* hos DnB NOR kunder:

- 54 menn: Gjennomsnitt: 5.426, Standardavvik: 2.618
- 37 kvinner: Gjennomsnitt 4.865, Standardavvik: 2.699

Finn 99% konfidensintervall for anbefalingsgjennomsnittet for mannlige og for kvinnelige DnB NOR kunder. Løsning:

❶ Menn: $t_{\alpha/2} = 2.678$ (runder ned til 50 df) og kvinner
 $t_{\alpha/2} = 2.719$

❷ Menn: $E = 2.678 \cdot 2.618 / \sqrt{54} = 0.954$. Kvinner:
 $E = 2.719 \cdot 2.699 / \sqrt{37} = 1.206$

❸ Menn 99% konfidensintervall:

$$4.47 < \mu_M < 6.38$$

❹ Kvinner 99% konfidensintervall:

Eksamensoppgaver Met 8006 vaaren 05

Spørsmål 24 Konfidensintervall for en andel

Av 100 nystartede bedrifter viste 20 seg å være oppegående fem år etter. Et 95 % konfidensintervall for overlevelsesandelen for nystartede bedrifter går fra:

- A) 0,12 til 0,28
- B) 0,14 til 0,26
- C) 0,16 til 0,24
- D) 0,18 til 0,22
- E) Jeg velger å ikke besvare spørsmålet

Spørsmål 30 Konfidensintervall

Du har gjort 400 målinger av en tilfeldig variabel. Målingenes standardavvik var 10. Et 90 % konfidensintervall for populasjonsgjennomsnittet har vidde:

- A) 0,82
- B) 1,65
- C) 16,45
- D) 32,90
- E) Jeg velger å ikke besvare spørsmålet

$$(\bar{x} - E, \bar{x} + E)$$

$$\text{bredden} = 2E = 2 \cdot \underbrace{t_{\alpha/2}^{(n-1)} \cdot \frac{s}{\sqrt{n}}}_{E} = 2 \cdot 1.649 \cdot \frac{10}{\sqrt{400}} = \underline{\underline{1.65}}$$

$$\alpha = 10\%$$

$$n-1 = 399$$

Spørsmål 26

Median

Du har intervjuet 100 personer og spurt om hvordan de liker det nye produktet i forhold til det gamle. 40 svarte "mye bedre", 15 "noe bedre", 20 "like bra", 10 "noe dårligere", og 15 "mye dårligere". Hva er mediansvaret?

- A) noe bedre
- B) like bra
- C) noe dårligere
- D) mye dårligere
- E) Jeg velger å ikke besvare spørsmålet

mye bedre	noe bedre	like bra	noe dårligere	mye dårligere
40	15	10	10	15

median = noe bedre

Spørsmål 27

Konfidensintervall

I en kvalitetskontroll har du gjort 64 observasjoner av x , og funnet et gjennomsnitt på 99,0 og et standardavvik på 5,0. Et 95 % konfidensintervall for μ har vidde:

- A) 20,0
- B) 10,0
- C) 5,0
- D) 2,5

$$\text{Bredde} = 2E = 2 \cdot t_{\alpha/2}^{(n-1)} \cdot \frac{s}{\sqrt{n}} = 2 \cdot 2.5 = 5.0$$

$\uparrow$
 $\alpha = 0.05$
 $n-1 = 63$

$\nwarrow$
 $\frac{5.0}{\sqrt{64}}$

1 8-1: Oversikt

2 8-2: Grunnleggende hypotesetesting

3 Section 8-3: Å teste påstander om andeler

4 Section 8-5: Teste en påstand om gjennomsnittet

Definisjoner

Hypotese

- En *hypotese* er en påstand om noe i en populasjon
- En *hypotesetest* er en standard prosedyre for å teste om en hypotese er sann eller ikke.

Example

- Hypotese: Andelen iPhone brukere på BI er $p = 0.23$
- Hypotesetest for en andel:
 - ➊ Hent inn en representativ stikkprøve
 - ➋ Beregn andelen iPhone brukere i stikkprøven
 - ➌ Bruk denne andelen til å avgjøre om hypotesen $p = 0.23$ er sann
 - ➍ Detaljer om dette kommer i seksjon 8-3...

Hypotesetesting

Rare event rule

- Vi har en hypotese om noe. Vi antar den er sann.
- Vi observerer noe i stikkprøven som ser uvanlig ut
- Konklusjon: Da er det noe galt med hypotesen, og vi avgjør at den er usann.

Hypotesetest = Beslutningsregel

- Du må velge en av to konkurrerende hypoteser.
 - Nullhypotesen H_0 inneholder alltid =
 - Alternativhypotesen H_1 inneholder $<$, $>$ eller $\neq$
- Du må velge H_0 eller H_1 . Du antar H_0 er sann
- Beslutningsgrunnlaget er stikkprøven.
- Dersom stikkprøven ser "uvanlig" ut, så velger vi H_1
- Dersom stikkprøven ikke er uvanlig, så velger vi H_0

Fiktivt eksempel: FemiPower

Example

- Selskapet FemiPower selger piller til kvinner som prøver å bli gravide.
- Selskapet påstår at pillene øker sjansen for å få en jente

For å teste om FemiPower gir økt sjanse til jente så gir vi pillene til flere kvinner, og 100 av disse blir gravide. Som utgangspunkt antar vi at FemiPower *ikke* virker. (Nullhypotesen)

- Bruk sunn fornuft. Hva vil du konkludere med i disse to scenariene:
 - ❶ 52 av 100 kvinner fikk jente?
 - ❷ 91 av 100 kvinner fikk jente?

Scenarie 1; 52 av 100 er jenter

52 av 100 er jenter

Nullhypotesen er at FemiPower ikke virker.

- Da vil vi normalt forvente omkring 50 jenter
- Vi fikk 52, noe som er nær 50, og det kan lett skje ved en tilfeldighet.
- Antagelsen om ingen effekt ser ut til å være korrekt
- *Det er ikke tilstrekkelig grunnlag til å hevde at FemiPower virker*

Scenarie 2; 91 av 100 er jenter

91 av 100 er jenter

Nullhypotesen er at FemiPower ikke virker.

- Da skal det utrolig mye til at 91 av 100 babyer er jenter! Vi kan forklare dette på to måter
 - ① En ekstremt uvanlig ting har skjedd, eller
 - ② FemiPower virker
- Det ekstremt usannsynlige i 91 av 100 jenter tyder på at FemiPower virker
- *Det er tilstrekkelig grunnlag til å hevde at FemiPower virker*