

5.25 Svangerskapsdiabetes?  $\rightarrow$  GDM (gestational diabetes mellitus)

$X$  = blodsukker 1 time etter sukkerinntak  $X > 140$  mg/dl = GDM

$$X \sim N(125, 10)$$

a) En måling:  $P(X > 140) = 1 - P(X < 140) = 1 - P\left(\frac{X-125}{10} < \frac{140-125}{10}\right)$   
 $= 1 - P(Z < 1.5) = (\text{Table A}) = 1 - 0.9332 \approx 7\%$

b) Gj. sn av 3 ulike dager:  $X_1, X_2, X_3 \rightarrow \bar{X} \sim N(125, \frac{10}{\sqrt{3}})$

Hvorfor er  $E(\bar{X}) = 125$  og  $\sigma_{\bar{X}} = \frac{10}{\sqrt{3}}$ ? **Generelt svar:**

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$\sigma_{\bar{X}} = \sqrt{\text{Var}(\bar{X})}$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{\text{fordi } X_1, \dots, X_n \text{ uavh}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

$$P(\bar{X} > 140) = 1 - P(\bar{X} < 140) = 1 - P\left(\frac{\bar{X}-125}{10/\sqrt{3}} < \frac{140-125}{10/\sqrt{3}}\right)$$
$$= 1 - P(Z < 2.60) = (\text{Table A}) = 1 - 0.9953 \approx 0.005 = 0.5\%$$

(Jfr Tsienobyl)

5.28 Forsikring. Liten risiko, høy kostnad

$X$  = <sup>individuell</sup> kostnad ved brann på et år

$E(X) = \mu = \$250$  per hus

$\sigma = \$10.000$

høyreskjev fordeling



a) Med en slik fordeling, slår ikke CLT, central limit theorem, sentralgrenseteoremet, inn for  $n=12$ . Her må vi ha en  $n$  som er nærmere  $\infty$  enn 12 er.

b)  $n=25000$   
 $\bar{X}_{25000} \xrightarrow{\text{fordi } n \text{ stor tilnærmet}} N(250, \frac{10.000}{\sqrt{25000}})$

$$P(\bar{x} > 270) = 1 - P(\bar{x} < 270) = 1 - P\left(\frac{\bar{x} - 250}{\frac{10.000}{\sqrt{25000}}} < \frac{270 - 250}{\frac{10.000}{\sqrt{25000}}}\right)$$

$$= 1 - P(Z < 0.32) \underset{\substack{\uparrow \\ \text{Table A}}}{=} 1 - 0.6255 \approx \underline{0.37}$$

### 5.50

Binomisk fordeling: La  $X$  være antall suksesser på

- \*  $n$  uavhengige forsøk, der
- \* hvert forsøk har to mulige utfall, suksess (S) og fiasko (F),
- \* og  $P(S) = p$  er lik i hvert forsøk.

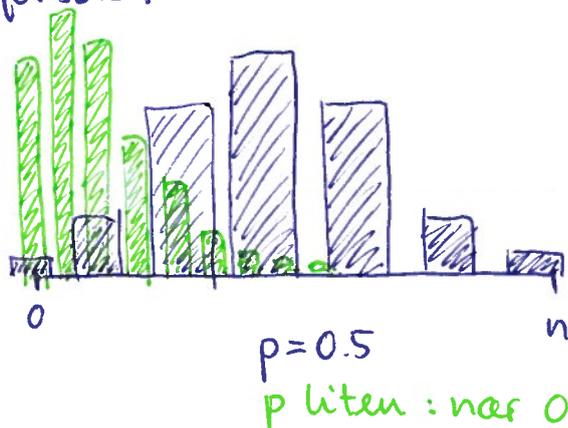
Da er

$$X \sim \text{bin}(n, p)$$

$$E(x) = np$$

$$\text{Var}(x) = np(1-p)$$

$$\sigma_x = \sqrt{\text{Var}(x)} = \sqrt{np(1-p)}$$



a)  $X$  = gjennomsnittlig daglig treningsstid for en gruppe. NEI

b)  $n=20$  tilfeldig valgte sko, enten Feil eller Feilfri  
 $P(\text{Feil på skoen}) = p$  er lik i hvert forsøk, JA

c)  $n$  = tilfeldig valgte studenter, enten nok frukt eller ikke,  
 $P(\text{nok frukt}) = p$  er lik i hvert forsøk, med mindre de  $n$  var valgt som venners venner e. l., hvilket de ikke er. JA,  $X = \#$  som spiser nok frukt, er  $\sim \text{bin}$ .

d)  $X = \#$  dager med skule. NEI, selv om hver dag kan ha enten skule eller ikke, er dagene neppe uavhengige.

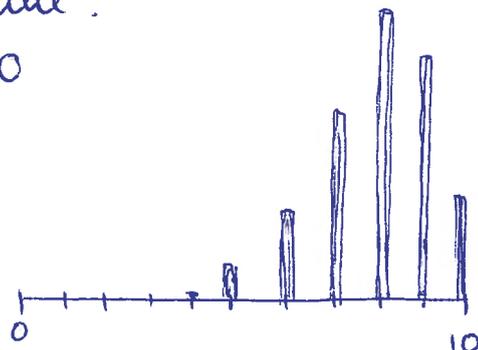
5.51 Spør 20000 studenter om de har stjålet. Andelen som stjeler: 0.2  
 Trekker 10 tilfeldige fra disse, kan anta uavhengighet.  
 Enten stjeler de, eller så stjeler de ikke.

$$P(\text{stjeler}) = 0.20, \quad P(\text{ikke stjele}) = 0.80$$

$X = \#$  som ikke stjeler

a)  $X \sim \text{bin}(10, 0.8)$

$$P(X=x) = \binom{10}{x} 0.8^x \cdot 0.2^{10-x}$$



m/2 desimaler

x	0	1	2	3	4	5	6	7	8	9	10	Sum
$P(X=x)$	0	0	0	0	0.01	0.03	0.09	0.20	0.30	0.27	0.11	1

$\rightarrow R: \text{dbinom}(0:10, 10, 0.8)$

$\text{plot}(0:10, \text{dbinom}(0:10, 10, 0.8))$

b)  $Y \approx \#$  som stjeler

$$Y \sim \text{bin}(10, 0.2)$$

noen ganger får man helt feilteppe og må bare seu alt opp ned...

$$P(\text{minst 4 stjeler}) = P(Y \geq 4) = 1 - P(Y < 4) = 1 - P(Y \leq 3)$$

$$R > 1 - \text{sum}(\text{dbinom}(0:3, 10, 0.2)) \approx \underline{0.12}$$

5.53  $E(X) = np = 10 \cdot 0.8 = 8$ . (Forventer 8 ærlige studenter)

$E(Y) = np = 10 \cdot 0.2 = 2$ . (Forventer 2 tyvaktige studenter)

Sum 10 studenter

$$b) \sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{np(1-p)} = \sqrt{10 \cdot 0.2 \cdot 0.8} = \underline{1.26}$$

$$\text{obs: } \sigma_X = \sqrt{\text{Var}(X)} = \sqrt{10 \cdot 0.8 \cdot 0.2} = \underline{1.26}$$

c)

p	$\sqrt{10 \cdot p(1-p)}$	$\sigma_Y$	$E(Y)$
0.2	$\sqrt{10 \cdot 0.2 \cdot 0.8}$	1.26	2
0.1	$\sqrt{10 \cdot 0.1 \cdot 0.9}$	0.95	1
0.01	$\sqrt{10 \cdot 0.01 \cdot 0.99}$	0.31	0.1



$\sigma$  blir større enn  $\mu$  og fordelingen skjev.

5.60 Det ideelle antallet barn?

La  $p$  være andelen i populasjonen som synes to barn er ideelt  
(Anta at den samme verdien til  $p = 0.53$ , men det vet vi ikke.)

Spørreundersøkelse:  $n = 1020$ , 53% svarer "to barn".

Da er  $\hat{p} = 0.53$  et estimat for  $p$ , og  $\hat{p}$  en estimator for  $p$ .

Gallupfirmaet forteller at feilmarginen her er  $\pm 4$  prosentpoeng

Hva er  $P(0.49 < \hat{p} < 0.57)$ ?

For å besvare dette, må vi vite fordelingen til  $\hat{p}$ .

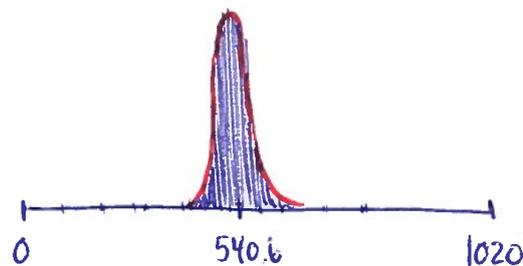
Vi vet: Hvis

$X = \#$  av de 1020 som svarer "2 barn", så er

$X \sim \text{bin}(1020, 0.53)$ , og da er

$$E(X) = 1020 \cdot 0.53 = \underline{540.6}$$

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{1020 \cdot 0.53 \cdot 0.47} = \underline{15.94}$$



Men dette er nesten en normalfordeling:

$$X \sim \text{bin}(1020, 0.53) \approx N(540.6, 15.94) \\ N(n \cdot p, \sqrt{n \cdot p \cdot (1-p)})$$

Generelt om  $\hat{p}$ :

$$\hat{p} = \frac{X}{n}$$

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot np = p$$

$$\sigma_{\hat{p}} = \sqrt{\text{Var}(\hat{p})} = \sqrt{\text{Var}\left(\frac{X}{n}\right)} = \sqrt{\frac{1}{n^2} \text{Var}(X)} = \sqrt{\frac{1}{n^2} \cdot np(1-p)} = \sqrt{\frac{p(1-p)}{n}}$$

nær stør og mindre for nær 0 eller 1

$$\rightarrow \hat{p} = \frac{X}{n} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$N\left(0.53, \sqrt{\frac{0.53 \cdot 0.47}{1020}}\right) = N(0.53, 0.0156)$$

$$\text{Altså: } P(0.49 < \hat{p} < 0.57) =$$

$$= P\left(\frac{0.49-0.53}{0.0156} < \frac{\hat{p}-0.53}{0.0156} < \frac{0.57-0.53}{0.0156}\right)$$

$$= P(-2.56 < \underset{\substack{\uparrow \\ N(0,1), \text{ se table A}}}{Z} < 2.56) = 0.995 - 0.005 = \underline{0.99}$$

5.62  $n = 300, p = 0.53$

$$\text{Da vi} \hat{p} = \frac{X}{n} \approx N\left(0.53, \sqrt{\frac{0.53 \cdot 0.47}{300}}\right) = N(0.53, 0.0288)$$

Og

$$P(0.49 < \hat{p} < 0.57) =$$

$$= P\left(\frac{0.49-0.53}{0.0288} < \frac{\hat{p}-0.53}{0.0288} < \frac{0.57-0.53}{0.0288}\right)$$

$$= P(-1.39 < \underset{\substack{\uparrow \\ N(0,1), \text{ se table A}}}{Z} < 1.39) = 0.9177 - 0.0823 \approx \underline{0.84}$$

$n = 1020, p = 0.53$  (fornise oppgave:

$$P(0.49 < \hat{p} < 0.57) = \dots \approx \underline{0.99}$$

$n = 5000, p = 0.53$

$$\text{Da vi} \hat{p} = \frac{X}{n} \approx N\left(0.53, \sqrt{\frac{0.53 \cdot 0.47}{5000}}\right) = N(0.53, 0.007)$$

$$P(0.49 < \hat{p} < 0.57) = P(5.67 < Z < 5.67) \approx 1$$

Jo større utvalg, jo mindre feil.

6.19 Biomarker: Måler innholdet i blodet av bestemte proteiner, for å få et inntrykk av prosesser i kroppen

Beinmetabolisme: "Osteoblastane byggjar  
Osteoklastane klir."

↳ gir tartrate-resistant acid phosphatase (TRAP) i blodet.

Måler TRAP i  $n=31$  unge kvinner;  $\bar{x}=13.2$  U/l

Antar at  $\sigma$  kjent, og  $\sigma=6.5$  U/l OBS: sier ingenting om fordelingen til TRAP.

Finn feilmarginer & konfidensintervall for

$\mu$  = forventet TRAP-mengde i blodet til unge kvinner.

## FEILMARGINER & KONFIDENSINTERVALL

Feilmarginer har jeg tidligere kalt "slingsingmenn".

Et konfidensintervall vil nesten alltid skrives på formen

estimat  $\pm$  feilmargin,

og størrelsen på feilmarginen avhenger av konfidensgraden.

95% konfidensgrad er et vanlig valg. Velge vi en lavere konfidensgrad, er konklusjonen tulle så sikker, og vi slipper unna med mindre feilmarginer og smalere KI.

Hvis vi derimot vil være sikrere på å ikke bomme, og velge en høyere konfidensgrad, må vi garde oss mer; beregne større feilmarginer, og bredere KI.

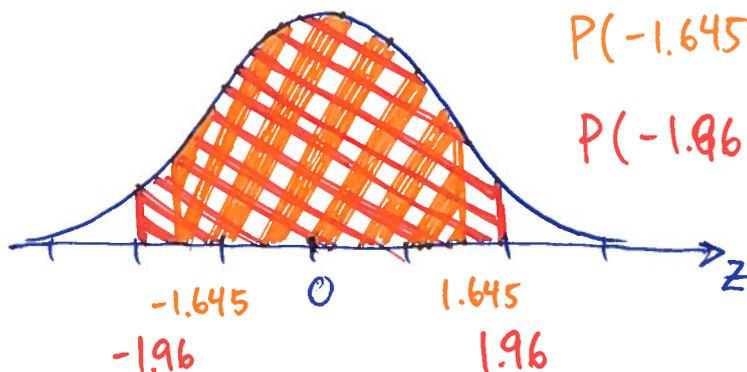
Generelt: Jo større  $n$ , jo smalere KI, og jo sikrere konklusjon

Her: Satser på at  $n=31$  er stort nok til at CLT kan brukes, altså at  $\bar{x} \stackrel{tln}{\sim} N(\mu, \frac{\sigma}{\sqrt{n}}) = N(\mu, \frac{6.5}{\sqrt{31}})$

Deretter: (Viktig)

Fordi  $\bar{X} \stackrel{tln}{\sim} N(\mu, 1.167)$ , så er

$$Z = \frac{\bar{X} - \mu}{1.167} \stackrel{tln}{\sim} N(0, 1)$$



$$P(-1.645 < Z < 1.645) = 0.90$$

$$P(-1.96 < Z < 1.96) = 0.95$$

$$\text{Altså: } P\left(-1.96 < \frac{\bar{X} - \mu}{1.167} < 1.96\right) \approx 0.95$$

↑  
vil isolere  $\mu$   
i midten av ulikhetene

$$P\left(-1.96 \cdot \underbrace{1.167}_{\frac{\sigma}{\sqrt{n}}} < (\bar{X} - \mu) < 1.96 \cdot \underbrace{1.167}_{\frac{\sigma}{\sqrt{n}}}\right) \approx 0.95$$

feilmargin = 2.29                      feilmargin = 2.29

$$P(-\bar{X} - 1.96 \cdot 1.167 < -\mu < -\bar{X} + 1.96 \cdot 1.167) \approx 0.95$$

$$P(\bar{X} + 1.96 \cdot 1.167 > \mu > \bar{X} - 1.96 \cdot 1.167) \approx 0.95$$

$$P(\bar{X} - 1.96 \cdot 1.167 < \mu < \bar{X} + 1.96 \cdot 1.167) \approx 0.95$$

95% konfidensintervall for  $\mu$ :  $\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

her:  $\bar{X} = 13.2$ , og konfidensintervall:  $[10.9, 15.5]$

Tilsvarende: 90% konfidensintervall:

$$P\left(-1.645 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.645\right) = 0.90$$

$$P\left(-1.645 \cdot \frac{\sigma}{\sqrt{n}} < (\bar{x} - \mu) < 1.645 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.90$$

$$P\left(-\bar{x} - 1.645 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + 1.645 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.90$$

$$P\left(\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}}\right) = 0.90$$

90% konfidensintervall for  $\mu$ :  $\bar{x} \pm \underbrace{1.645 \cdot \frac{\sigma}{\sqrt{n}}}_{\text{feilmargin}}$

6.20 Måler osteokalcin,  $n = 31$ ,  $\bar{x} = 33.4$ ,  $\sigma$  kjent,  $\sigma = 19.6$

95% KI for  $\mu$ , forventet osteok. hos unge kvinner:

$$33.4 \pm 1.96 \cdot \frac{19.6}{\sqrt{31}} \rightarrow 33.4 \pm 6.9 \rightarrow \underline{[26.5, 40.3]}$$

6.27

$n = 1200$  intervjuer om radiovaner, og hvor mange timer per uke de lytter til radio.

Gjennomsnittlig lyttetid for de 1200 studentene:  $\bar{x} = 11.5$

Av de 1200, var det bare 83% som hørte på radio.

Mao: Mange, 204, har svart 0 timer.

Anta  $\sigma$  kjent, og  $\sigma = 8.3$ .

Fordelingen:



a) 95% KI for  $\mu =$  forventet lyttetid for studenter.

Antar (?; har ikke andre verktøy i STK1000. Bootstrapping ville vært et godt alternativ her) at selv om fordelingen er skjev, så er  $n = 1200$  nok til at CLT slår inn, og:

$\bar{X} \stackrel{t/n}{\sim} N(\mu, \frac{\sigma}{\sqrt{n}})$  igjen.

95% KI for  $\mu$ :  $\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

$$11.5 \pm 1.96 \cdot \frac{8.3}{\sqrt{1200}} \rightarrow 11.5 \pm 0.45$$

$$[11.05, 11.95]$$

b) Spm: Er det sann at 95 av studentene har en lyttetid mellom 11.05 timer og 11.95 timer? Nei. KI forteller oss noe om hva  $\mu$  er, ikke hva  $X_i$  er.

Mao: KI er ikke det samme som percentiler.

c) Med en så stor  $n$ , vil  $\bar{X}$  ha stabilisert seg, til tross for skjevheten i data.  $\bar{X}$  er imidlertid ikke nødvendigvis en god oppsummering av data, selv om vi kan anta at  $\bar{X}$  er et stabilt tall, og at fordelingen til  $\bar{X}$  er omtrent normalfordelt.

6.28

$\bar{x} = 11.5$  timer, eller 690 min.

a)  $\sigma = 8.3$  timer, eller 498 min.

b) 95% KI for  $\mu$ , forventet lyttetid i minutter:

$$690 \pm 1.96 \cdot \frac{498}{\sqrt{1200}} \rightarrow [662, 718]$$

28.2, feilmargin

c) [11.05, 11.95]  $\xrightarrow[\text{med } 60]{\text{ganger nedre og øvre grænse}}$  [663, 717]

6.30 Bensinforbruk: miles per gallon

$n=20$ , data i fit, antar  $\sigma$  kjent,  $\sigma=3.5$  mpg

RStudio: Import dataset  $\rightarrow$  From local file

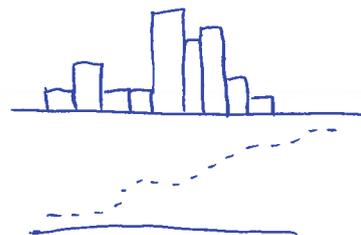
IPSE DataSets > Chapter 6 > ex06-30mpg

kan ikke åpnes.

Åpne fila i Excel, save as CSV

Rstudio på nytt. OK!

b) hist(ex06.30mpg\$MPG)  
qqnorm(ex06.30mpg\$MPG)  
mean(—————)  
>43.17



ikke normalfordelt.

a)  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{20}} = \underline{0.783}$

c) 95% KI for  $\mu$ , forventet bensinforbruk:

Her er  $n$  relativt liten, men nå vet vi at det er rimelig å anta  $N$ -fordeling. Dermed har vi igjen at

$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ , og 95% KI for  $\mu$  er gitt ved

$$43.17 \pm 1.96 \cdot 0.783 \rightarrow [41.6, 44.7]$$

6.33 TRAP igjen, jfr oppg 6.19:

Anta fortsatt at  $\bar{x} = 13,2$ ,  $\sigma = 6.5$ , og finn ut hvor stor  $n$  må være for at vi skal kunne beregne  $\mu$  innenfor feilmargin på 1.5, med 95% konfidensgrad

95% konfidensintervall:

estimat  $\pm$  feilmargin

$$\text{Her: } 13.2 \pm 1.96 \cdot \frac{6.5}{\sqrt{n}}$$

Feilmargin på 1.5, betyr at

$$1.5 = 1.96 \cdot \frac{6.5}{\sqrt{n}}$$

$$\sqrt{n} \cdot 1.5 = 1.96 \cdot 6.5$$

$$\sqrt{n} = \frac{1.96 \cdot 6.5}{1.5}$$

$$n = \left( \frac{1.96 \cdot 6.5}{1.5} \right)^2$$

$$n = 72.1 \longrightarrow \underline{n = 73}$$

Dette kalles en utvalgsberegning.