Tirsdag 18/10

OPPSUMMERING AV FØRSTE HALVDEL AV STK1000

Bakgrunn
↓
Problemstilling
Konkret spm som vi vil ha svar på

Ch 3

Design
Plan for å besvare problemstilling

Observasjonelt | Eksperimentelt manipulering

Ch 1

Tallfeste det du vil ha svar på
Hovedeffektmål
Hovedvariabel
Responsvariabel

kategorisk | numerisk: diskret, kontinuerlig

skjev

Analyse:

Deskriptiv statistikk: Beskrive én og én variabel

- antall, andel, tabeller (kake, søyle)
- trenger flere tall for å oppsummere fordelingen: Q_1, median, Q_3; min, Q_1, median, Q_3, max
- oppsummeres utmerket med to tall: $\bar{X}$, SD

Ch 2

Beskrive/oppsummere sammenhengen mellom to variabler

kategorisk × kategorisk: KRYSSTABELL

kont × kategorisk:

kont × kontinuerlig: Scatter

> R table(blodtype, yrke) boxplot(lønn ~ yrke) plot(glukose, insulin)

Formalisering: Sannsynlighet

Sannsynlighetsmodell

stokastisk / random / tilfeldig variabel X

X har et sett med verdier den kan være, og en sh fordeling

x		
$P(X=x)$		1

$E(X)$, $Var(X)$, $\sigma_X = \sqrt{Var(X)}$

Ch 4

Populasjon & parameter vs Utvalg & statistic observator

∞ $\rightarrow$ μ, p

n $\rightarrow$ $\left.\begin{matrix} \bar{x} \\ \hat{p} \end{matrix}\right\}$ fordelingen til $\bar{x}$, og til $\hat{p}$

Ch 5

RESTEN AV STK1000

Uke 42 — Estimering, inklusive konfidensintervall for μ, for p } Ch 5 + Ch 6.1

Uke 43 — Hypotesetesting

Sammenligne grupper og finne sammenhengen mellom to variabler: Bivariate tester

Tabellanalyse Ch 9 ikke pensum

t-test Ch 7.2

Mann-Whitney test Ch 15.1

Korrelasjon Ch 2.3

> R

Uke 44 — Mer estimering & hypotesetesting

Ettutvalgstester, parede tester } Ch 7.1

Uke 45 — Regresjon: Ch 2.4 + 2.5 + Ch 10.1

> R

Uke 46 — Mer regresjon! Ch 10.1 + Ch 11 Forklaring vs prediksjon

> R

Uke 47 — Repetisjon

Estimering og konfidensintervall

← trenger å vite noe om fordelingen til estimatoren

∞

populasjon

ukjent fordeling

Er interessert i "populasjonens gjennomsnitt"

Kalles forventningsverdien μ

trekker tilfeldig →

n

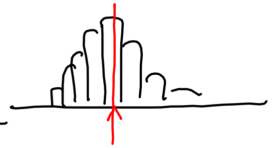
utvalg

kjent

Ch1 →

data

$X_1, \dots, X_n$



gjennomsnitt

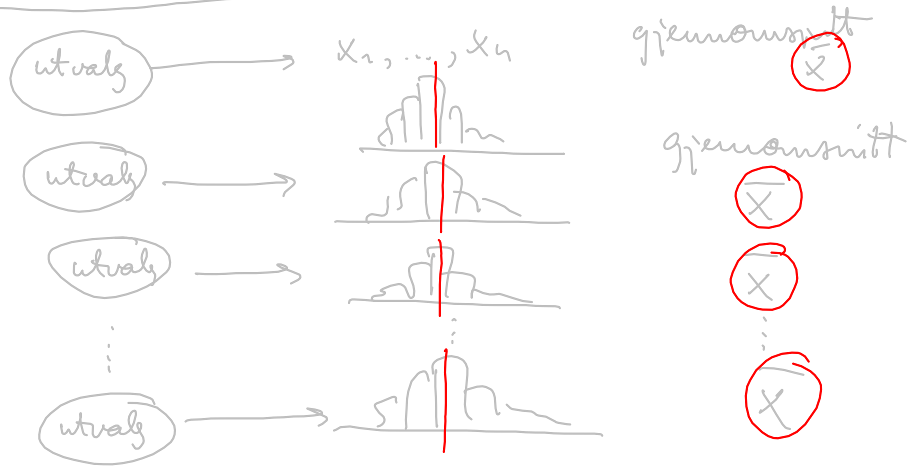
$\bar{X}$

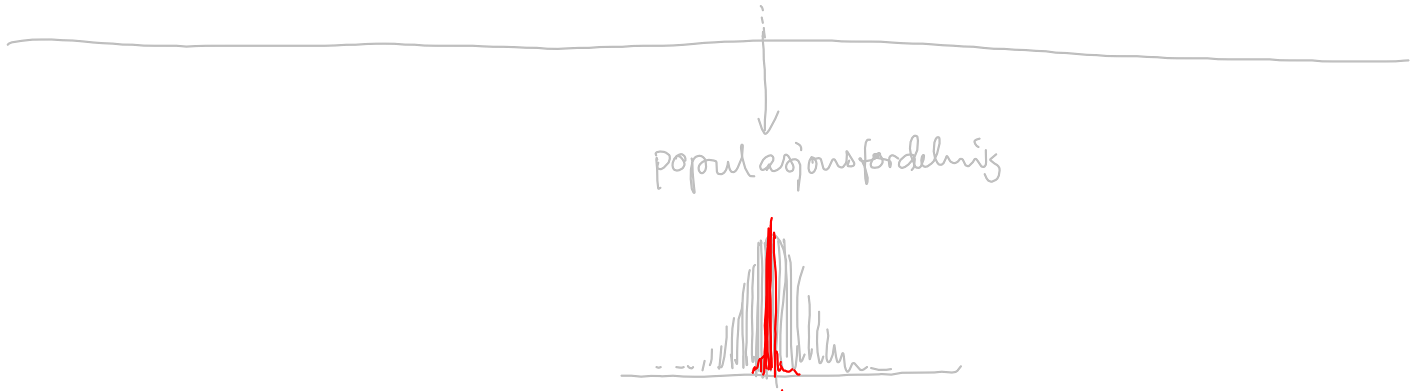
inferens:

gjennomsnittet i utvalget ligner på gjennomsnittet i populasjonen: $\bar{X} \approx \mu$

$\bar{x}$ er et estimat for μ. Men det er ikke nødvendig lik μ: $\bar{x} \pm ?$

Hvor stort slingringsmonn? Dette uttrykker vi presist med et konfidensintervall. For at vi skal kunne lage et konf. int, må vi tenke oss:



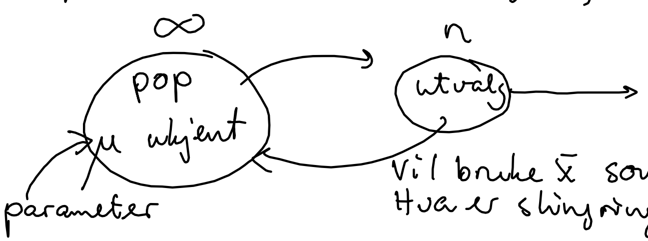


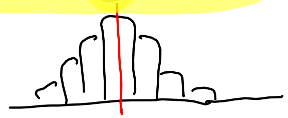
fordelingen til $\bar{x}$

$\boxed{\text{Fordelingen til } \bar{x}}$ er den vi bruker når vi skal regne ut et konfidensintervall.

Hvordan denne ser ut avhenger av hvilken fordeling dataene har (er de normalfordelte eller ikke, og er σ for dataene kjent eller ikke), og av hvor stor n er.

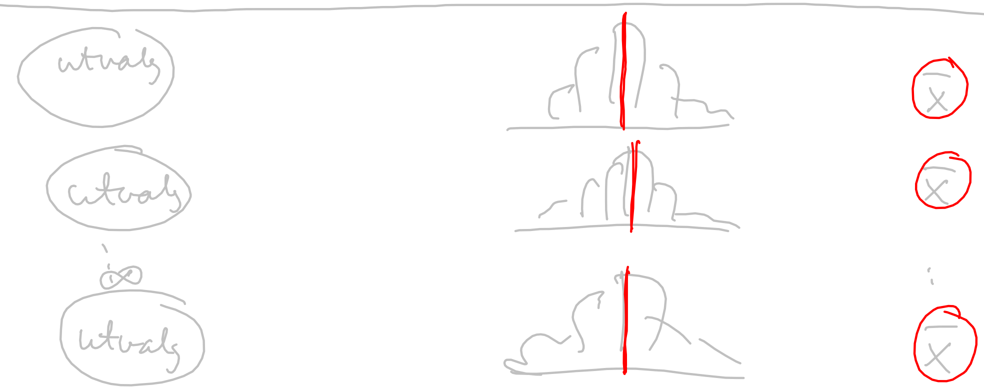
Situasjon 2: Data: $X_1, \dots, X_n$: $X_i \sim N(\mu, \sigma_x)$, σ_x kjent. σ_x ukjent





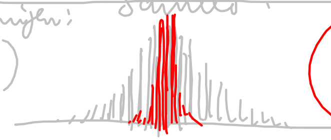
Gjennomsnitt $\bar{x}$

Vil bruke $\bar{x}$ som estimator for μ.
Hva er slingringsmonnet $\bar{x} \pm ?$



populasjon fordelingen: $N(\mu, \sigma_x)$

Samler:



fordelingen til $\bar{X}$: $N\left(\mu, \frac{\sigma_x}{\sqrt{n}}\right)$

Nei, må estimere σ_x, bruker sd_x i stedet.

senere.

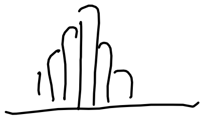
$$\bar{X} \sim N\left(\mu, \frac{\sigma_x}{\sqrt{n}}\right) \longrightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma_x}{\sqrt{n}}} \sim N(0,1)$$

Når σ_x ukjent: ikke samme, det er mer usikkerhet inn i bildet

$$t = \frac{\bar{X} - \mu}{\frac{sd_x}{\sqrt{n}}} \sim T_{n-1}$$

Eksempel:

$X_1, \ldots, X_{25}$ IQ målinger : STK1000 $\sigma = 15$



$\approx$ N-fordelt, σ kjent. $n = 25$, $\bar{X} = 110$

Estimat for μ, forventet IQ for statistikkstudenter : $\bar{X}$

$\bar{X}$ er estimator for μ

Estimat for μ: 110

N

95% Konfidensintervall for μ: $\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$: $110 \pm 1.96 \cdot \frac{15}{5}$

90% 1.645

Men hva hvis σ = ukjent? Data $\sim N$, $n = 25$: Situasjon 2

sd = 14,3

$$\frac{\bar{X} - \mu}{sd/\sqrt{n}} \sim T_{n-1}$$

25-1=24

Table D df = 24 $\rightarrow$ upper-tail 0.025 $\hookrightarrow$ 2.064

95% konfidensintervall

$\longrightarrow$ $110 \pm 2.064 \cdot \frac{14.3}{5}$