

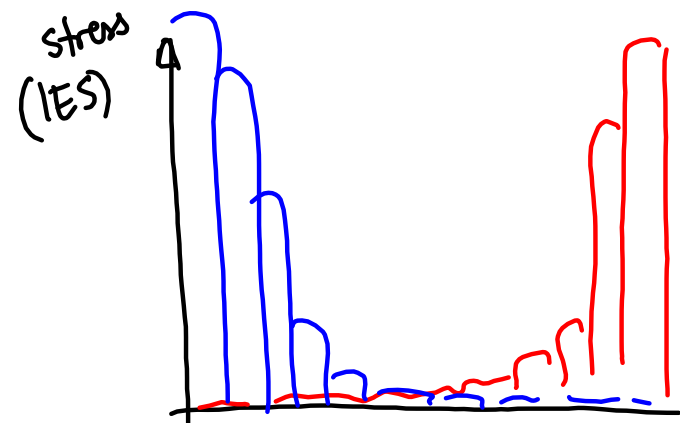
13/11 To oppklaringsark først.

Deretter: konfidensintervall og
prediksjonsintervall: REGRESJON

via oppgave 3 a og 4 fra H-2016-eksamen

Oppklaring 1: Hvis vi skal sammenligne to grupper med kontinuerlige data:
Har tidligere gått gjennom 4 situasjoner, avhengig av om
(σ er kjent eller ikke), n er stor eller ikke, data er
normalfordelt eller ikke, i tråd med boka ca ch 5-7

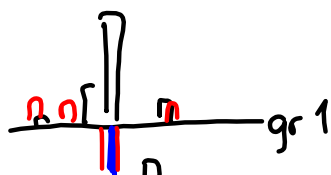
	Symmetrisk	Skjev
liten n	(t-outvalgs) t-test	WRS test
stor n	t-test (\approx z-test)	z-test \rightarrow må håndregnes (\approx t-testen) \rightarrow R, SPSS, ...



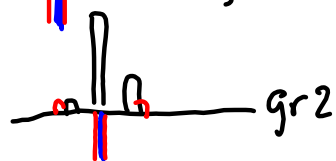
Oppklaring 2: Hvorfor kan vi tillate oss å bruke t-test når n er stor, hvis data er skjeve? Blir det ikke feil å basere seg på gjennomsnitt når data er skjeve, selv om n er stor?

Tenk situasjonen:

$$n = 6 + 6 \text{ (to grupper)}$$



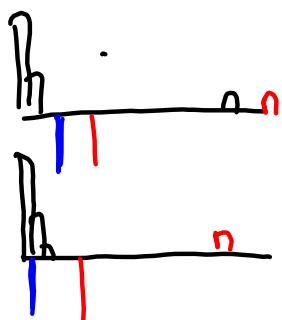
gjennomsnittene i de to gruppene er ikke oppsummeringer av det typiske.



Hva skjer hvis jeg observerer én ekstra observasjon i hver gruppe?

Gjennomsnittet er ganske stabilt.

$$n = 6 + 6$$



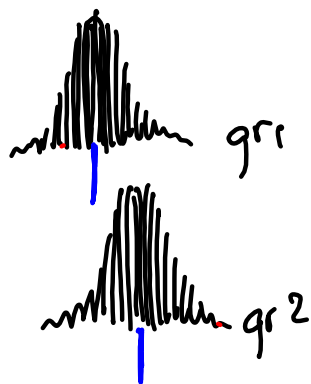
Gjennomsnittet er ikke et oppsummeringsstall av det typiske

Hva med en ny observasjon?

Gjennomsnittet er slikkkelig ustabil også.

} Det er ikke lenger gjennomsnittet selv som oppsummering eller til å sammenligne grupper.

$n = 600 + 600$



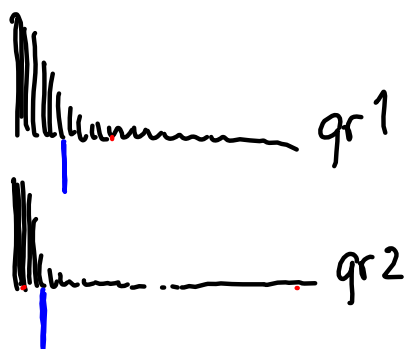
Gjennomsnittene er gode oppsummerings tall for det typiske

Hva hvis vi fikk en ekstra observasjon?

Gjennomsnittet påvirkes veldig lite
 n er stabilt

Brake gj. sn både til å oppsummere & til å samle grupper.

$n = 600 + 600$



gjennomsnittene er ikke gode oppsummeringer av det typiske i data, men ett nytt datapunkt påvirker gjennomsnittet lite, gjennomsnittet er stabilt nok til å brukes for å samle to grupper. \rightarrow CLT

Brake gj. sn til å undersøke om to grupper er like, men vil kanskje ikke bruke gj. sn. som oppsummerings tall.

REGRESJON: Konfidensintervall for regresjonskoeffisienter. Boka Ch10, s574

$$\hat{\beta}_1 \pm t \cdot SE(\hat{\beta}_1)$$

Vi bruker eksamensoppgave 3a ^{H2016} som illustrasjon. (3b: neste uke)

a) **Effektmaß**: Et tall som oppsummerer effekten av (variasjon i) blodsukker på fødselsvekt. I en regresjonsanalyse er det regresjonskoeffisienten som viser stigningstallet til regresjonslinja, altså

$$\beta_1 \text{ i regresjonslikningen } y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

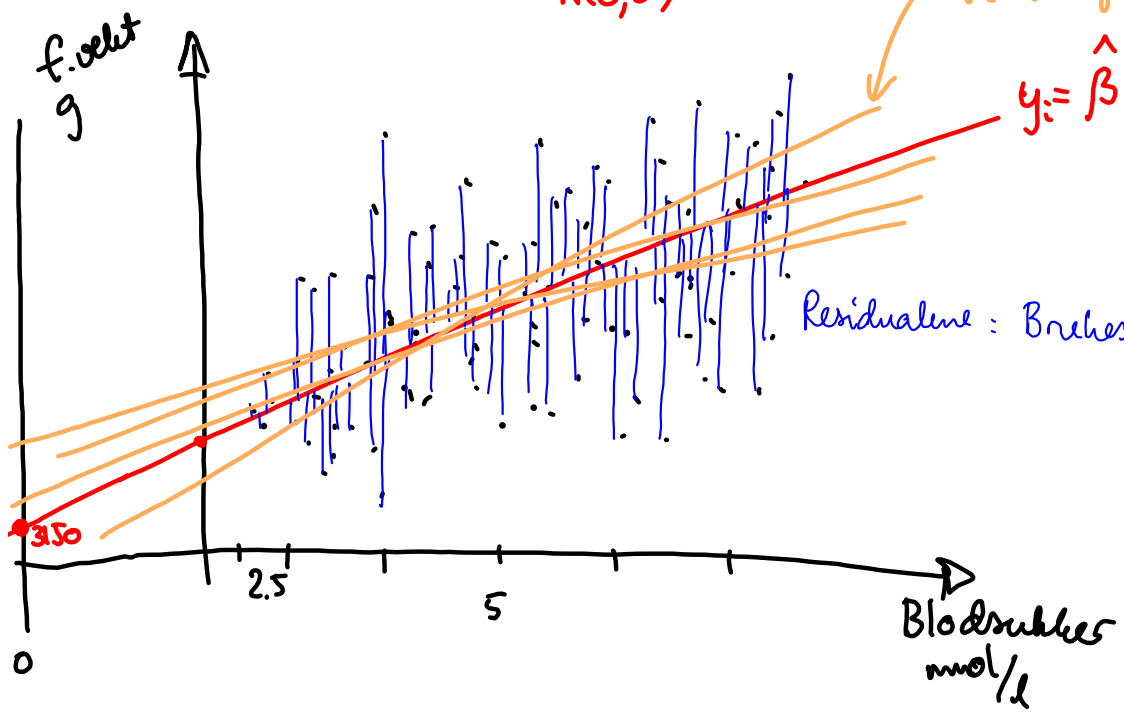
β_1 er stigningstallet til regresjonslikningen: Det viser hvor mange enheters forskjell i fødselsvekt (y) som forventes når blodsukkeret øker med en enhet.

Estimat for β_1 : $\hat{\beta}_1$, fra utskriften: $\hat{\beta}_1 = 172$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad n = 200$$

$\varepsilon_i \sim N(0, \sigma)$

Kan linja like gjens ha gått her?
Vi trenger konfidensintervall for β_1 .



$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

\downarrow
172 g/mol/l

Residualene: Behovs til å finne et estimat for σ .

95% Konfidensintervall for β_1 :

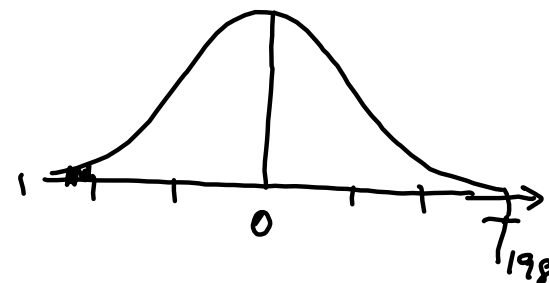
Minner om: $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{tln}{\sim} N(0,1)$

$$\frac{\bar{X} - \mu}{SD/\sqrt{n}} \sim T_{n-1} \rightarrow \frac{\bar{X} - \mu}{S.E(\bar{X})} \sim T_{n-1} \quad H_0: \mu=0 \quad \frac{\bar{X}}{SE(\bar{X})} \sim T_{n-1}$$

Nå: $\frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} \sim T_{n-2} \quad H_0: \beta_1 = 0: \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \sim T_{n-2}$

Her: $n=200 \quad T_{n-2} = t_{198}$

$$P\left(-t_{198,0.025} < \frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} < t_{198,0.025}\right) = 0.95$$



$$P\left(-t_{198,0.025} \cdot S.E(\hat{\beta}_1) < (\hat{\beta}_1 - \beta_1) < t_{198,0.025} \cdot S.E(\hat{\beta}_1)\right) = 0.95$$

95% KI for β_1 :

$$\hat{\beta}_1 \pm t_{198, 0.025} \cdot S.E(\hat{\beta}_1) \rightarrow 172 \pm 1.984 \cdot 82$$

Table D: $t_{100} = \underline{1.984}$

$t_{1000} = 1.962$

[9, 335]

Det betyr at med 95% sikkerhet er effekten av blodsukker på fødselsvekt mellom 9 gram og 335 g per enhets blunng i blodsukker.

Eller: en gruppe med gravide med blodsukker på 4

en gruppe med gravide med blodsukker på 7

Forskjell i fødselsvekt

$30g < 3 \cdot 9g$

$3 \cdot 355g > 1kg$

$3 \cdot 172g$

$\approx 500g$