

FORELESNING

18

Eivind ERIKSEN

MAR 27 2012

MET 3431

STATISTIKK

PLAN:

- ① Repetisjon: Sammenlikning av uavhengige stikkprøver (9.3)
 ——————
 ② —————— relaterete .. (9.4)
 ③ Formel for rette linjer.

Husk: Nytt kontor: 84i - 032
 Kontortid: Onsdag 10-12 (som før)

- ① Repetisjon: Sammenlikning av uavhengige stikkprøver

Stikkprøve 1	Stikkprøve 2
;	;
data	data
;	;
;	;
;	;
gjennomsnitt: \bar{x}_1	gjennomsnitt: \bar{x}_2
Std. avvik: s_1	Std. avvik: s_2
antall: n_1	antall: n_2

t-test for gjennomsnitt:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

For å finne kritisk t-verdi / forskningsområdet:

- i) Sett opp H_0 / H_1 — ensidig eller tosidig test?
 ii) Sla opp kritisk t-verdi
 i tabell A-3: } Finn fram til riktig α
 der n er minste
 verdi av n_1 og n_2 .

Oppgave 5

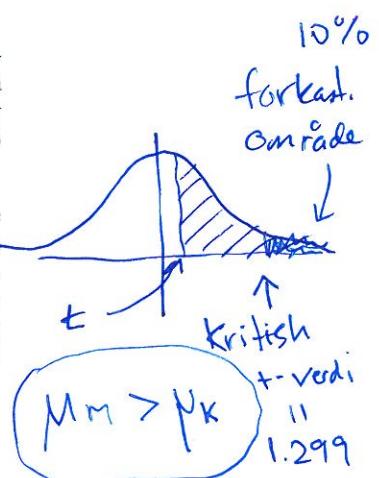
- (a) I hypotesetesting, hva vil det si å begå en type I feil? Anta at du tester på signifikansnivå $\alpha = 0.05$, og at H_0 er sann. Hva er sannsynligheten for å begå en type I feil?
- (b) I en stikkprøve på 157 studenter ved BI Bergen våren 2011 hadde 13 studenter mobiltelefon av merket Samsung. Test påstanden om at mindre enn ti prosent av studentene ved BI Bergen har Samsung mobiltelefon. Skriv opp nullhypotesen og alternativhypotesen. Foreta testen på et $\alpha = 0.05$ signifikansnivå og formuler konklusjonen i et lettfattelig språk.

- (c) Du har tilgang til stikkprøver av ukentlige timer på jobb for 60 mannlige studenter og for 99 kvinnelige studenter. Er de to stikkprøvene med menn og kvinner uavhengige eller relaterte? Test hypotesen om at mannlige studenter gjennomsnittlig jobber mer enn kvinnelige studenter basert på de to stikkprøvene vist i Figur 2. Skriv opp nullhypotesen og alternativhypotesen og bruk signifikansnivå $\alpha = 0.1$. Skisser p-verdien som arealet under en graf.

$\mu_m = \text{gi snitt}$
for menn

$\mu_k = \text{gi snitt}$
for kvinner

- (a) Figur 3 viser et scatterplott av timer på jobb vs studietimer per uke for en tilfeldig valgt stikkprøve av åtte BI studenter våren 2011. Hvor mange studenter i stikkprøven oppga at de studerte 30 timer i uka? Hvor mye jobbet hver av disse studentene i uka?
- (b) Korrelasjonskoeffisienten for dataene i Figur 3 er $r = 0.73$. Test på $\alpha = 0.05$ nivået om det er en korrelasjon mellom timer på jobb og timer brukt på studier. Skriv opp H_0 og H_1 .
- (c) Gi et eksempel på to kontinuerlige variable som du tror er signifikant negativt korrelerte.

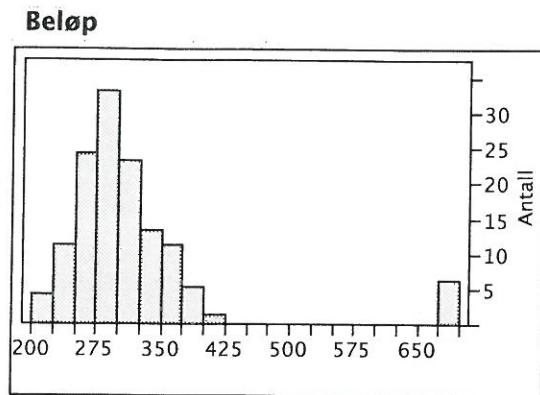


$$t = \frac{\bar{x}_M - \bar{x}_K}{\sqrt{\frac{s_M^2}{n_M} + \frac{s_K^2}{n_K}}}$$

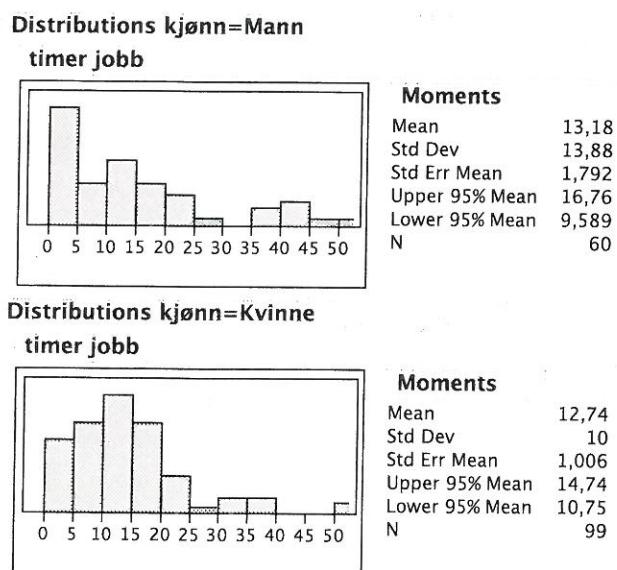
Oppgave 7

Figur 4 gir krysstabellen for sammenhengen mellom kjønn og studium i en tilfeldig stikkprøve av 291 studenter fra BI Trondheim. I hver celle oppgis det faktiske og det forventede antallet.

- (a) Hvilkens hypotesetest kan du bruke for å avgjøre om det er en sammenheng mellom kjønn og studium i populasjonen av studenter ved BI Trondheim?
- (b) Foreta en test på 1% nivået og konkluder i et lettfattelig språk.



Figur 1: Oppgave 1



Figur 2: Oppgave 5

$$\begin{aligned}
 t &= \frac{13,18 - 12,74}{\sqrt{\frac{13,88^2}{60} + \frac{10^2}{99}}} = \frac{\overline{x}_M - \overline{x}_K}{\sqrt{\frac{s_M^2}{n_M} + \frac{s_K^2}{n_K}}} \\
 &\approx 0,214
 \end{aligned}$$

(2)

Sesjon (9.4).

Test for to relaterte stikkprøver

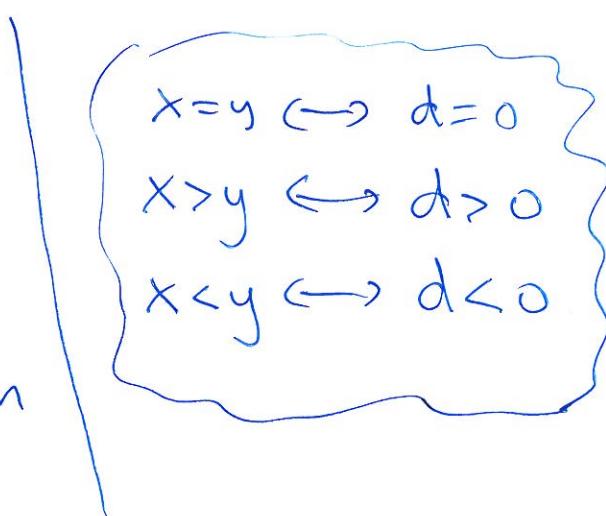
Regn ut differansen i hvert par

- Er stikkprøven relatert, så regner vi ut differansene parvis
- Da har vi egentlig bare en stikkprøve av differanser d
- Da vil $\mu_1 = \mu_2$ være det samme som at gjennomsnittet av differansene er null: $\mu_d = 0$

Forutsetninger

- Tilfeldig stikkprøve
- Enten større enn $n = 30$
- Eller par av differanser kommer fra en normalfordelt populasjon

Stikkprøv.	Stikkprøv.	Differansen
1	2.	$d = x - y$
x_1	y_1	$d_1 = x_1 - y_1$
x_2	y_2	$d_2 = x_2 - y_2$
:	:	:
x_n	y_n	$d_n = x_n - y_n$



Notasjon

- d er differansen mellom to verdier i et par $d = x - y$
- μ_d er populasjonsgjennomsnittet for alle differansene
- \bar{d} er gjennomsnittet av differansene i stikkprøven
- s_d er standardavviket for differansene i stikkprøven
- n er antall par i stikkprøvene (begge stikkprøvene har denne størrelsen)

Test for to relaterte stikkprøver

Testobservator

For to relaterte stikkprøver:

- Nullhypotesen er $H_0 : \mu_1 = \mu_2 \Leftrightarrow H_0 : M_d = 0$
- Testobservatoren blir da

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

- Har $n - 1$ frihetsgrader

Test for to relaterte stikkprøver

To relaterte stikkprøver og t -testen

- To relaterte stikkprøver x og y : Regn ut de parvise differansene
- Regn ut differansenes gjennomsnitt \bar{d} og standardavvik s_d
- Hypotesene er

$$H_0 : \mu_x = \mu_y \quad \text{vs.} \quad H_1 : \mu_x \neq \mu_y$$

- Dersom H_0 er sann, er \bar{d} gjennomsnittlig lik null. Da blir testobservatoren

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

- Forkast H_0 dersom t går utenfor kritisk verdi t_α fra tabell A-3

Example

Selger man flere pakker Cornflakes hvis pakkene står i øyehøyde? Her er salgstall for 13 par av butikker, der hvert par er likt iforhold til beliggenhet, nabolag, størrelse. x er salgstall for øyehøyde, y er for plassering nede ved gulvet.

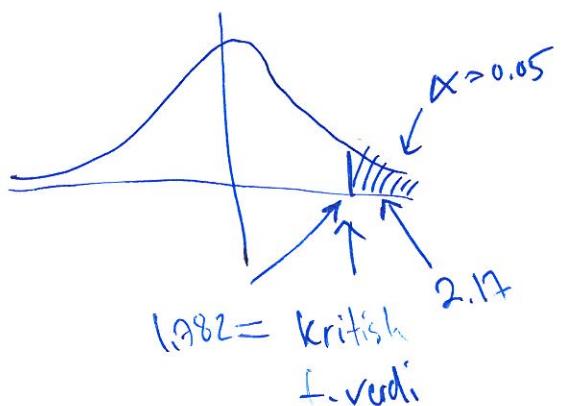
x	111	150	130	154	67	112	84	123	71	62	38	51	77
y	71	121	133	126	93	49	109	96	27	58	36	37	69
d	40	29	-3	28	-26	63	-25	27	44	4	2	14	8

$$d = x - y$$

- Ensidig test: Er øyehøyde best? $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x > \mu_y$
- $\bar{d} = 15.8$ og $s_d = 26.2$ gir t -verdi $\frac{\bar{d}}{s_d} = \frac{15.8}{26.2} = 0.59$

$$t = \frac{15.8}{\frac{26.2}{\sqrt{13}}} = 2.17$$

- Kritisk verdi fra tabell A-3 $t_\alpha = 1.78$ (for $\alpha = 0.05$). t -verdien er større enn kritisk verdi: Forkast H_0
- Det er tilstrekkelig grunnlag til å hevde at øyehøyde selger bedre



Example

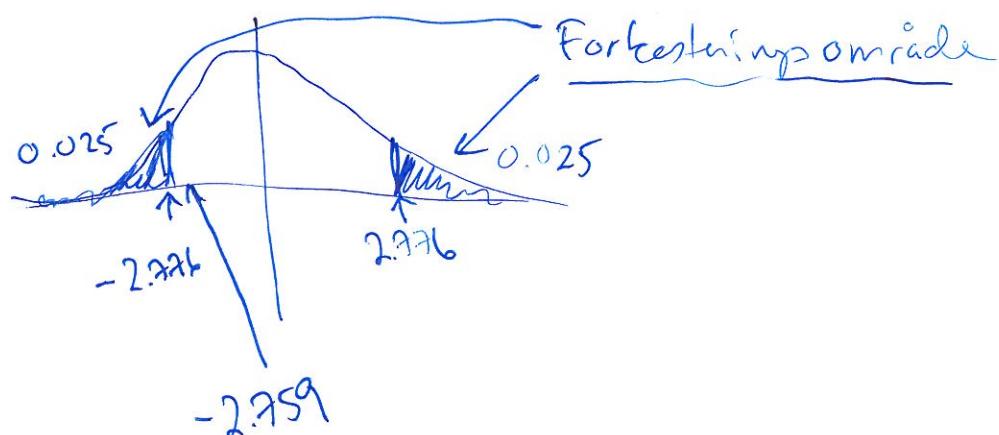
Værmeldingen angir minimumstemperaturen fem dager frem i tid. For fem datoer noterte vi den meldte temperaturen og den faktiske temperaturen (i $^{\circ}\text{F}$) 4 dager senere.

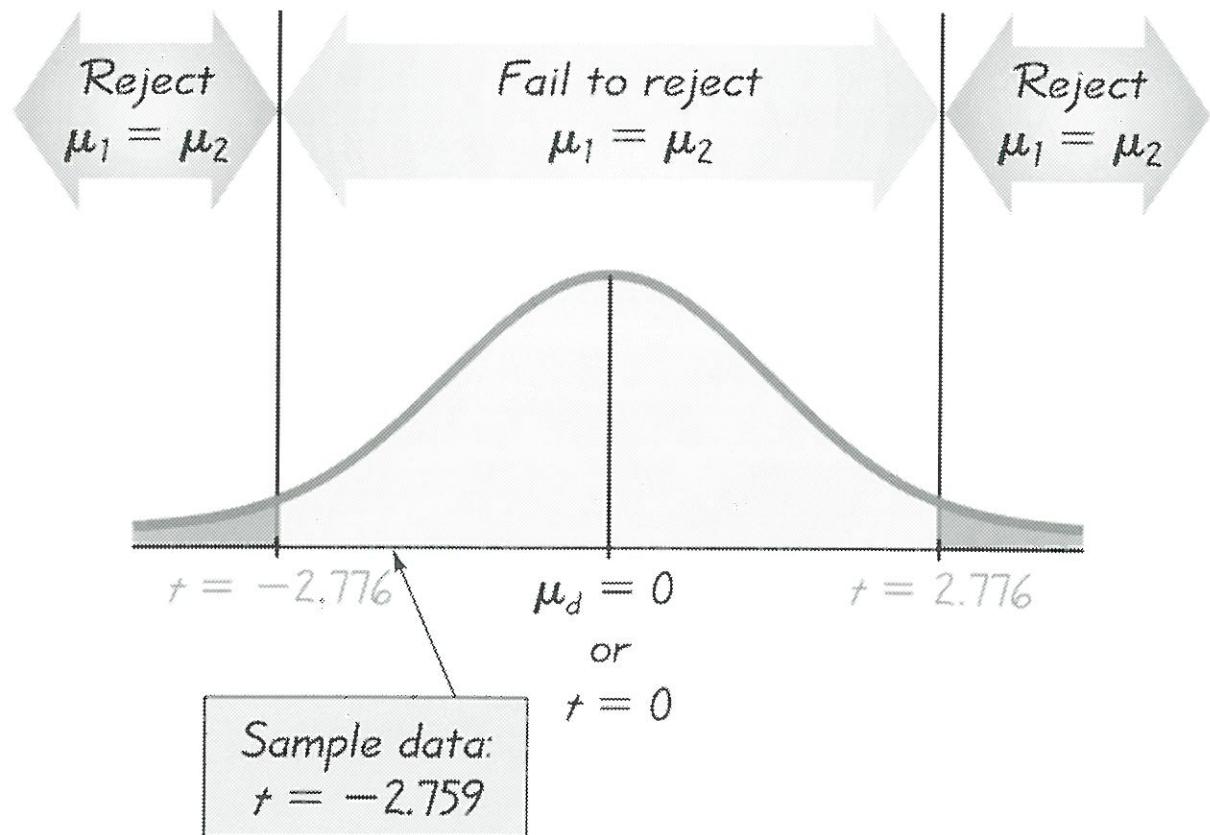
Actual and Forecast Temperature					
	1	-5	-5	23	9
Actual low	1	-5	-5	23	9
Low forecast five days earlier	16	16	20	22	15
Difference $d = \text{actual} - \text{predicted}$	-15	-21	-25	1	-6

$$\begin{aligned} x \\ y \\ d = x - y \end{aligned}$$

Vi ønsker å teste om værmeldingen i gjennomsnitt har rett (bruker $\alpha = 0.05$): $(M_d = 0)$ $(M_d \neq 0)$

- $H_0 : \mu_{\text{meldt}} = \mu_{\text{faktisk}}$ vs $H_1 : \mu_{\text{meldt}} \neq \mu_{\text{faktisk}}$
- $\bar{d} = -13.2$, $s = 10.7$, $n = 5$ og $t_{\alpha/2} = 2.776$ med $df = 4$.
- Testobservatoren $t = \frac{-13.2}{10.7/\sqrt{5}} = -2.759$
- Vi kan ikke forkaste H_0 , det er ikke tilstrekkelig grunnlag til å hevde at værmeldingen i gjennomsnitt bommer





JMP

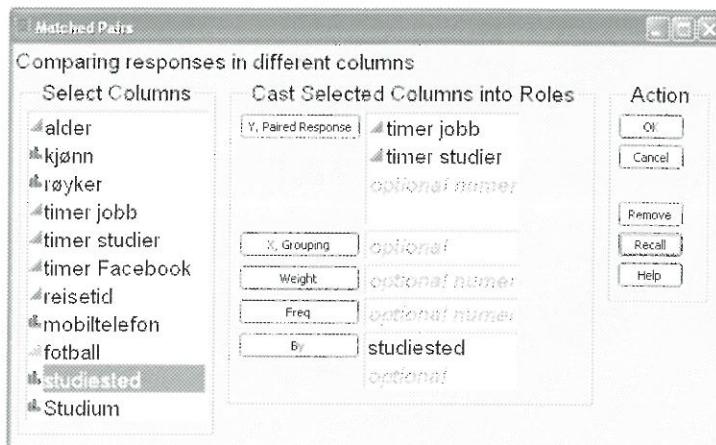
Example

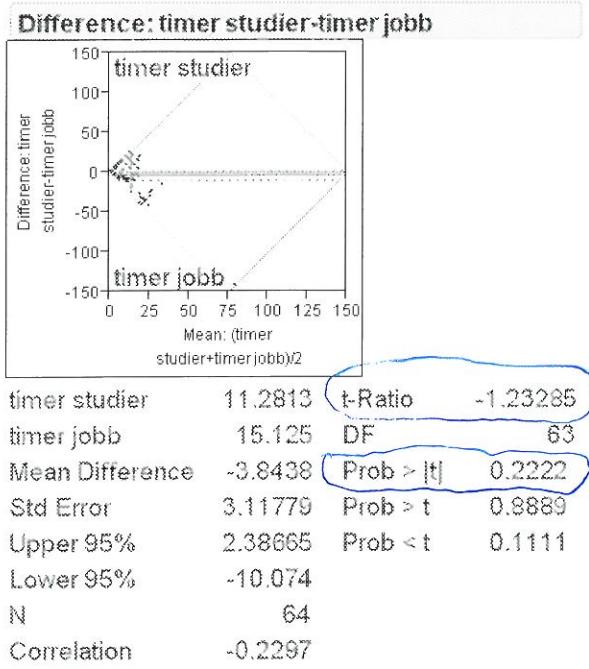
BI Drammen:

- Bruker studentene i gjennomsnitt like mye tid på jobb som på studier ?

JMP for relaterte stikkprøver

Analyze > Matched Pairs





Figur: klassens data alle ny

Example

- Vi setter $\alpha = 0.05$

$$H_0 : \mu_{\text{jobb}} = \mu_{\text{studier}} \quad \text{vs.} \quad H_1 : \mu_{\text{jobb}} \neq \mu_{\text{studier}}$$

- Testobservatoren

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-3.84}{3.12} = -1.23$$

- Vi kan bruke tradisjonell metode med kritisk verdi fra tabell A3. Men JMP gir p -verdien: 0.22
- Siden $0.22 > \alpha = 0.05$, forkastes ikke H_0 . Det er ikke grunnlag til å hevde at antall timer brukt på jobb og på studier i gjennomsnitt er forskjellige ved BI Drammen

Alle rette linjer har sin egen formel

Rett linje

Formelen for en rett linje i $x - y$ koordinatsystemet er

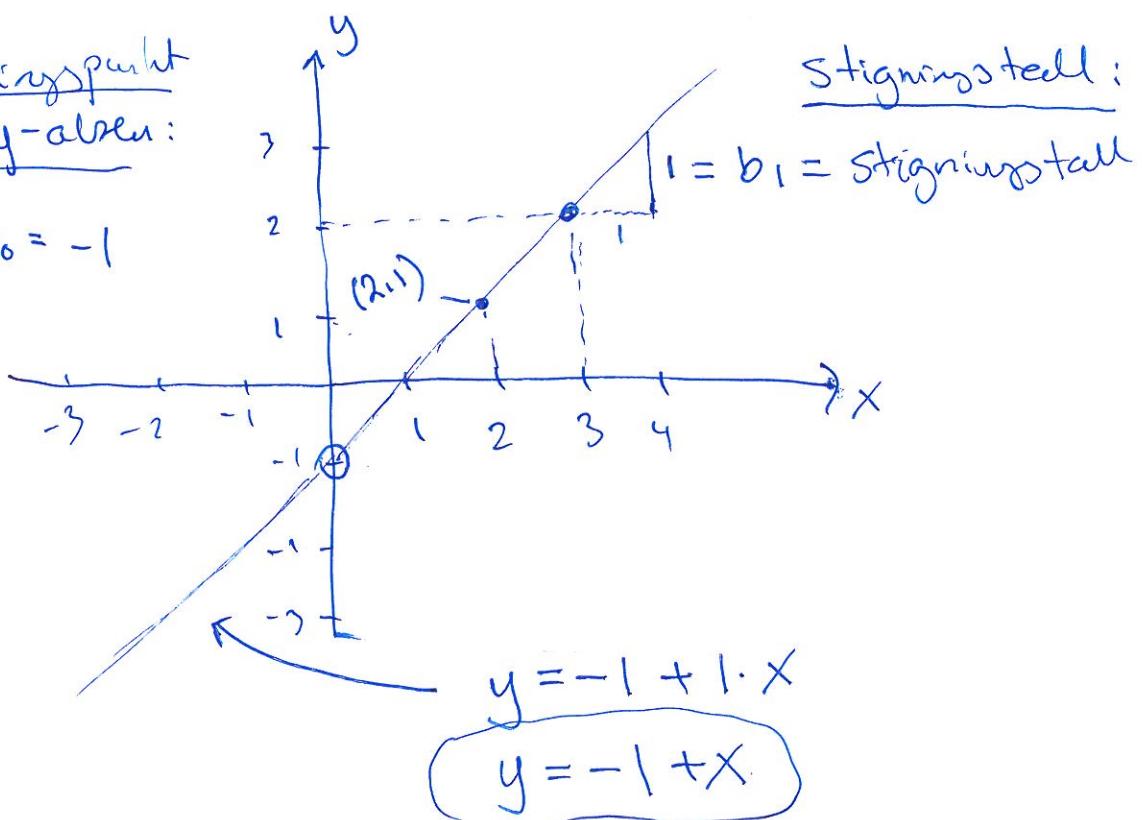
$$y = b_0 + b_1 x$$

Skj. punkt med y-aksen *stigningstall*

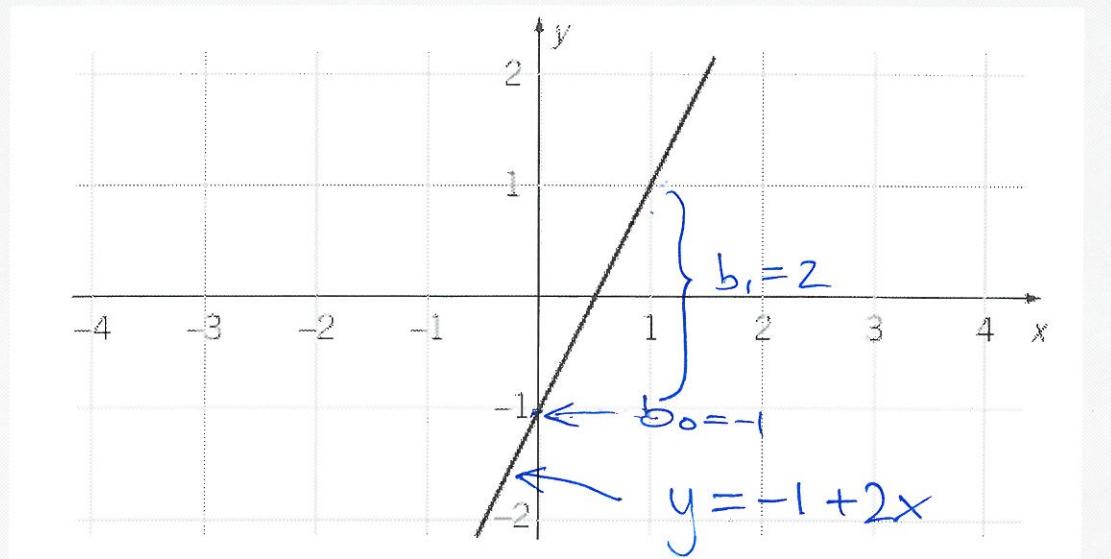
- b_0 er et tall som angir hvor linja krysser y -aksen
- b_1 er *stigningstallet* som angir hva stigningen til linja er

Skjæringspunkt med y-aksen:

Eks: $b_0 = -1$



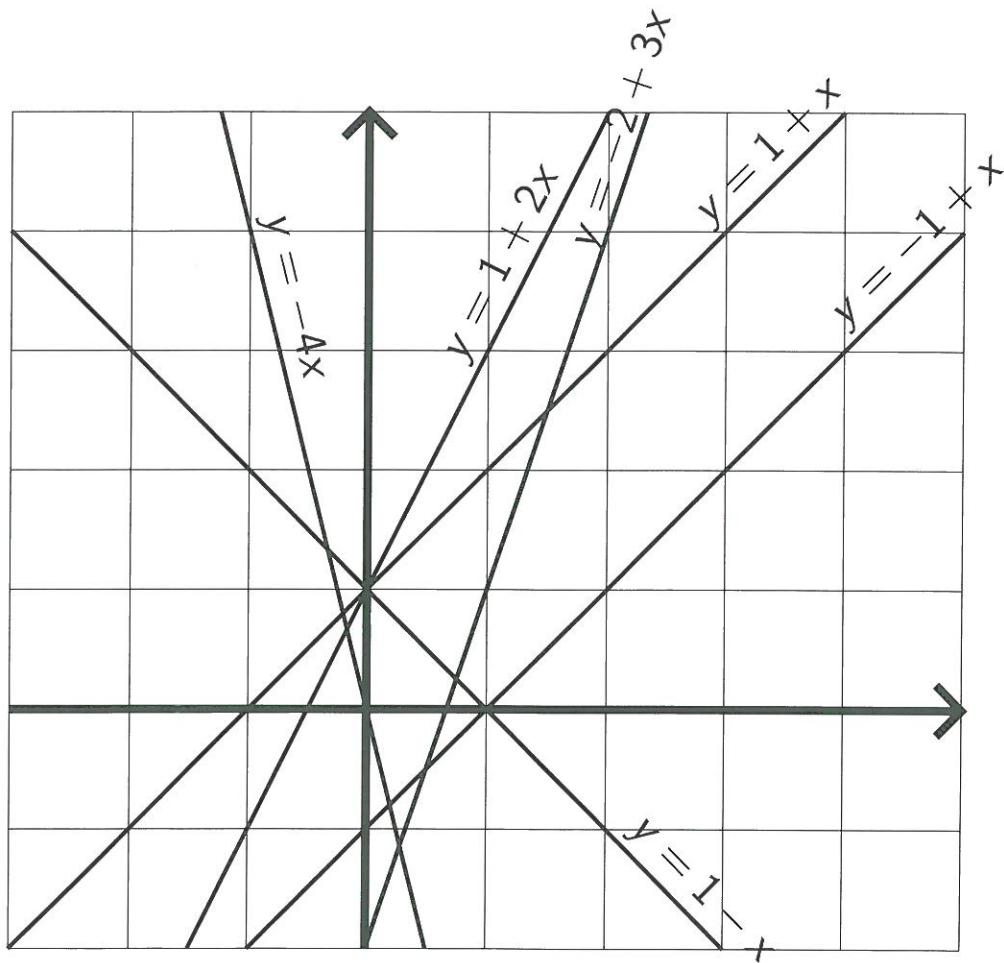
Example



- Krysser y -aksen i -1 : $b_0 = -1$
- Stiger med 2 hver gang x øker med 1: $b_1 = 2$
- Formelen til linja er derfor

$$y = -1 + 2x$$

Øvingsoppgaver



Skriv opp likningen for disse linjene:

