Plan:

1. Repetisjon:
   - (A) Sannsynlighetsteori og -regning
   - (B) Statistikk
   - (C) Lineær regresjon

2. Gjennomgang av Prøve-eksamen II

Minimum forberedelse:

- Prøve-eksamen I – II
- Oppsummering A – B – C
- Siste eksamensoppgaver

Prøve-eksamen II

1. a) X = antall feilvare (Type A)

<u>binomisk</u> $(n,p) = \underline{(50, 0.02)}$

$P(X \geq 1) = 1 - P(X \leq 0) = 1 - \binom{n}{0} p^0 (1-p)^{50}$

$= 1 - 0.98^{50} \simeq \underline{\underline{0.636}} = \underline{\underline{63.6\%}}$

b) Y = antall feilvare (Type B)

<u>binomisk</u>, $(n,p) = (100, 0.06)$

$P(Y \geq 10) = 1 - P(Y \leq 9)$

Y tilnærmet normalfordelt,

$E(Y) = np = 100 \cdot 0.06 = \underline{6} \leftarrow \mu$

$Var(Y) = np(1-p) = 6 \cdot 0.94 = 5.64 \leftarrow \sigma^2$

god tilnærming siden $Var(Y) > 5$

$\simeq 1 - P\left(Z \leq \frac{9-6+0.5}{\sqrt{5.64}}\right)$

$\simeq 1 - P\left(Z \leq \frac{3.5}{2.375}\right) = 1 - G(3.5/2.375)$

$\simeq \underline{\underline{0.070}} = \underline{\underline{7.0\%}}$

$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$

$\binom{n}{i}$: n $\boxed{nCr}$ i

$P(Y \leq a) \simeq P\left(Z \leq \frac{a-\mu+0.5}{\sigma}\right)$

heltallskorreksjon

<u>Kalk:</u>

6 : $\boxed{Z \rightleftarrows P}$

c) $X+Y=$ antall feilvare samlet

X, Y uavhengige

$$p(X+Y\geq 2) = 1 - p(X+Y\leq 1)$$

$$= 1 - p(X=0, Y=0) - p(X=1, Y=0) - p(X=0, Y=1)$$

$$= 1 - \underbrace{0.98^{50}}_{X=0} \cdot \underbrace{0.94^{100}}_{Y=0} - \underbrace{50 \cdot 0.02 \cdot 0.98^{49}}_{X=1} \cdot \underbrace{0.94^{100}}_{Y=0}$$

$$- \underbrace{0.98^{50}}_{X=0} \cdot \underbrace{100 \cdot 0.06 \cdot 0.94^{99}}_{Y=1} \simeq$$

$$\simeq 1 - 0.00075 - 0.00076 - 0.00478$$

$$\simeq 0.994 = \underline{\underline{99.4\,\%}}$$

Kalk:

$\boxed{\text{C STAT}}$

211.7 $\boxed{\Sigma+}$

2.2 $\boxed{\Sigma+}$

⋮

-0.8 $\boxed{\Sigma+}$

$\boxed{\bar{x}}$

$\boxed{S_x}$

2. a) $x_1, x_2, \ldots, x_n$: datapunkt $(n=10)$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \simeq \underline{36.54\,\%}$$

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \simeq \underline{85.05\,\%}$$

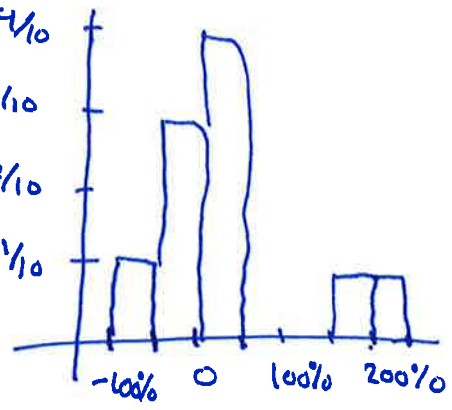
Intervall: $[\bar{x}-S_x, \bar{x}+S_x] = [\underline{-48.5\%, 121.6\%}]$

Andel: $7/10 = 0.7 = \underline{\underline{70\%}}$ av observasjoner er innenfor intervallet $\bar{x} \pm S_x$

Frekvenstabell:

	Antall	Frekvens
[-100%, -50%]	1	1/10
[-50%, 0%]	3	3/10
[0%, 50%]	4	4/10
(50%, 100%]	0	0
[100%, 150%]	0	0
[150%, 200%]	1	1/10
[200%, 280%]	1	1/10
	10	

Histogram:



b) **Konfidensintervall:** μ (σ ukjent)

Antar: $x_1, \ldots, x_{10}$ normalfordelte med $E(x_i) = \mu$, $Var(x_i) = \sigma^2$, uavhengige

Intervall: $\bar{x} \pm t^{n-1}_{\alpha/2} \cdot s/\sqrt{n}$

$\bar{x} = 36.54\%$
$s_x = s = 85.05\%$ } fra a)

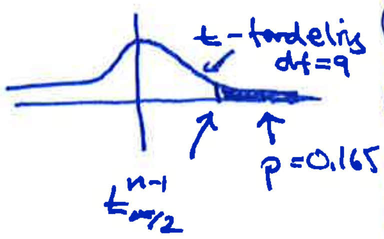
$n-1 = 9$

$\alpha/2 = 0.33/2 = 0.165$

(siden $1-\alpha = 0.67$)

$t^{n-1}_{\alpha/2} = t^{9}_{0.165}$

$\simeq 1.0298$



Kalk:

9 $\boxed{\text{INV}}$ $\boxed{\text{df}, t \Rightarrow p}$

0.835 $\boxed{=}$

siden

$p(T < t^{n-1}_{\alpha/2})$

$= 1 - 0.165 = 0.835$

Konfidensintervall:

$36.5\% \pm 1.0298 \cdot 85.05\%$

$= 36.5\% \pm 27.7\%$

$[8.8\%, 64.2\%]$

c) Hypotese-test (T-test) for μ

$$\left.\begin{array}{l} H_0: \mu \le 0 = \mu_0 \\ H_1: \mu > 0 = \mu_0 \end{array}\right\} \text{hypoteser}$$

(høyresidig)

Testobservator: $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

Forkastningsområde: $T > t_{\alpha}^{n-1} \simeq 1.383$

$\alpha = 0.01$, $n-1 = 9$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{36.54\% - 0\%}{85.05\% / \sqrt{10}} \simeq 1.359$$

Konklusjon:

Siden t ikke er i forkastnings-området, er vår beslutning å behalde H_0 ($\mu \le 0$) på $\alpha = 0.01$

Anta:
Samme antagelser som i b).

Kalk:
9 $\boxed{\text{INV}}$ $\boxed{df, t \rightleftharpoons P}$
0.99 $\boxed{=}$
↑
areal til venstre for $t_{\alpha}^{n-1} = 1 - \alpha = 0.99$

3. a) Vi skriver $(x_1, y_1), \dots, (x_n, y_n)$ for observasjonene, med $n = 11$.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}}$$

$$\simeq 0.8164 = 0.8164$$

$$\hat{\beta} = r_{xy} \cdot \frac{S_y}{S_x} \simeq 50.01$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \simeq 300.0$$

Anta:
Standard antagelser for lineær regresjon

Kalk:
$\boxed{\text{C STAT}}$
10 $\boxed{\text{INPUT}}$ 804 $\boxed{\Sigma+}$
8 " 695 "
⋮
5 " 568 $\boxed{\Sigma+}$

$\boxed{r} \to r$
$\boxed{m} \to \hat{\beta}$
$\boxed{b} \to \hat{\alpha}$

Estimat for regresjonslinjen:

$$y = 300 + 50x$$

b) Konfidensintervall for β

$$\hat{\beta} \pm t^{n-2}_{\alpha/2} \cdot SE(\hat{\beta})$$

$\hat{\beta} = 50.01$ (fra a))

$t^{n-2}_{\alpha/2} = t^{9}_{0.02} \simeq 2.398$

siden $\alpha = 0.04$ $(1-\alpha = 0.96)$

$$SE(\hat{\beta})^2 = \frac{1}{n-2} \frac{s_y^2}{s_x^2}(1-r^2)$$

$$= \frac{1}{9} \cdot \frac{41273}{4}(1-0.8164^2)$$

$$\simeq 139.0$$

Konfidensintervall:

$$50.01 \pm 2.398 \cdot \sqrt{139.0}$$

$$= 50.01 \pm 28.3$$

$$[21.7, 78.3]$$

$SE(\hat{\beta})$: standard feil for $\hat{\beta}$ = std. avvik til $\hat{\beta}$

$$= \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

Kalk:

9 $\boxed{\text{INV}}$ $\boxed{\text{df}, t \geq P}$

0.98 $\boxed{=}$

↑

1−0.02

c) $r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$ er den samme om vi bytter om x og y, dvs $r_{xy} = r_{yx}$.

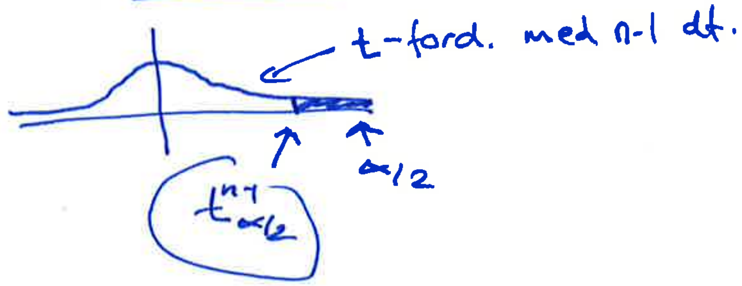
Siden $\hat{\beta} = r_{xy} \cdot \frac{S_y}{S_x}$ har samme fortegn som r_{xy}, er fortegnet til $\hat{\beta}$ uavhengig av om vi bytter om rekkefølgen på x og y; fortegnet blir fortsatt positivt

4. a) Konfidensintervall $\bar{x} \pm t^{n-1}_{\alpha/2} \cdot s/\sqrt{n}$

$\bar{x}$: utvalgsgjennomsnitt

s: —"— std avvik

$t^{n-1}_{\alpha/2}$: kritisk t-verdi med n-1 df slik at $P(T > t^{n-1}_{\alpha/2}) = \alpha/2$



$\frac{s}{\sqrt{n}} = \frac{s}{\sqrt{100}}$ er estimat for $\frac{\sigma}{\sqrt{n}} = SE(\bar{x})$ siden

$$Var(\bar{x}) = Var\left(\frac{1}{n}(x_1 + \dots + x_n)\right) = \frac{1}{n^2}(n \cdot \sigma^2) = \sigma^2/n$$

og $\sqrt{\sigma^2/n} = \sigma/\sqrt{n}$.

b) Bredden til konfidensintervallet:

$$B - A = \left(\bar{x} + t^{n-1}_{\alpha/2}\cdot s/\sqrt{n}\right) - \left(\bar{x} - t^{n-1}_{\alpha/2}\, s/\sqrt{n}\right)$$

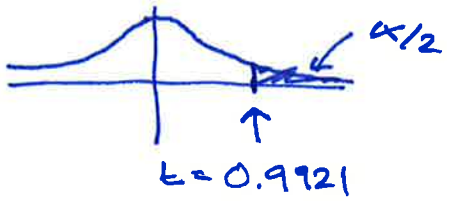
$$= 2\, t^{n-1}_{\alpha/2}\cdot s/\sqrt{n}$$

For å få halve bredden, må vi halvere $t^{n-1}_{\alpha/2}$.

$\alpha = 0.05$: $t^{n-1}_{\alpha/2} = t^{99}_{0.025} = 1.9842$

(95%)

Ønsker: α slik at $t^{n-1}_{\alpha/2} = \frac{1.9842}{2} = 0.9921$



Markert areal:

$$P(T > 0.9921) = 1 - P(T \leq 0.9921)$$

$$\simeq 1 - 0.8382 = 0.162$$

$$\Downarrow$$

$$\alpha/2 = 0.162$$

$$\alpha = 0.324 = \underline{\underline{32.4\%}}$$

Kalk:

99 $\boxed{\text{dt,tzp}}$

0.9921 $\boxed{=}$

5. $R_1 \sim N(0.12, 0.20)$
$R_2 \sim N(0.20, 0.30)$ ← $N(\mu, \sigma)$ skrivemåte

a) $p(R_1 > 0.20) = 1 - p(R_1 \leq 0.20)$
$= 1 - p\left(Z \leq \frac{0.20 - 0.12}{0.20}\right)$
$= 1 - G(8/20)$
$= 1 - 0.655 = 0.345 = \underline{\underline{34.5\%}}$

$p(R_2 < 0) = p\left(Z < \frac{0 - 0.20}{0.30}\right) = G\left(-\frac{0.20}{0.30}\right)$
$= 0.252 = \underline{\underline{25.2\%}}$

b) $p(R \geq 0.20)$
$= 1 - p(R < 0.20)$
$= 1 - p\left(Z < \frac{0.20 - 0.16}{0.18}\right)$
$= 1 - G\left(\frac{0.04}{0.18}\right) \simeq \underline{\underline{41.2\%}}$

$R = 0.5 R_1 + 0.5 R_2$
R_1, R_2 uavh

i) R normalfordelt $N(\mu, \sigma)$
ii) $\mu = E(0.5R_1 + 0.5R_2)$
$= 0.5 \cdot 0.12 + 0.5 \cdot 0.20$
$= 0.06 + 0.10 = \underline{0.16}$
iii) $\sigma^2 = Var(R) = Var(0.5R_1 + 0.5R_2)$
$= 0.5^2 \cdot 0.20^2 + 0.5^2 \cdot 0.30^2$
$+ 2 \cdot 0.5 \cdot 0.5 \cdot Cov(R_1, R_2)$, $Cov(R_1,R_2) = 0$
$= 0.0325$
$\sigma = \sqrt{0.0325} \simeq 0.180$

c) $\mathrm{Var}(R) = \mathrm{Var}(0.5R_1 + 0.5R_2)$

$= 0.5^2 \cdot 0.20^2 + 0.5^2 \cdot 0.30^2 + 2 \cdot 0.5 \cdot 0.5 \cdot \gamma$

$= \underline{0.0325 + 0.5\gamma}$

Hva er mulige verdier av $\underline{\gamma = \mathrm{Cov}(R_1, R_2)}$?

$$\rho = \frac{\mathrm{Cov}(R_1,R_2)}{\sigma_{R_1} \cdot \sigma_{R_2}} \longleftarrow$$

korrelasjonskoeff. til R_1, R_2 må være i $\underline{[-1,1]}$

$$= \frac{\gamma}{0.20 \cdot 0.30} = \frac{\gamma}{0.06}$$

$$-1 \leq \frac{\gamma}{0.06} \leq 1 \qquad | \cdot 0.06$$

$$-0.06 \leq \gamma \leq 0.06$$

$\Big\}\ \gamma \in [-0.06, 0.06]$

<u>Minste verdi for $\mathrm{Var}(R)$:</u> $0.0325 + 0.5 \cdot (-0.06)$

$= \underline{\underline{0.0025}}$

når $\underline{\underline{\gamma = -0.06}}$

$\Downarrow$

$\sigma_R = \sqrt{0.0025} = \underline{\underline{0.05}}$

minste std. avvik