



Universitetet  
i Stavanger

EKSAMEN I: STA100 SANNSYNLIGHETSREGNING OG STATISTIKK 1

VARIGHET: 4 TIMER

DATO: 12. MAI 2017

BOKMÅL

TILLATTE HJELPEMIDLER: Godkjent enkel kalkulator (HP30S, Casio FX82, TI-30,  
Citizen SR-270X , Texas BA II Plus eller HP17bII+ .

OPPGAVESETTET BESTÅR AV 5 OPPGAVER PÅ 9 SIDER INKL. VEDLEGG

---

formelark,  $N(0,1)$ ,  $t$

Oppgave 1

Tolv personer er søkere på en jobb. Ni av søkerne er kompetente for jobben; tre er ikke kompetente. To skal ansettes.

- a) Hvor mange ulike par av personer er mulig? Hvor mange par er det der én eller begge er kompetente? Dersom to velges tilfeldig blant de tolv, hva er sannsynligheten for ingen er kompetente?

Oppgave 2

To oljereservoar betraktes, 1 og 2. Sannsynligheten for å finne olje i reservoar 1 er 0.6, og sannsynligheten for å finne olje i reservoar 2 er 0.7. Sannsynligheten for å finne olje i begge reservoarene er 0.45.

- a) Finn sannsynligheten for at man finner olje i minst ett av reservoarene, og finn sannsynligheten for at man ikke finner olje i noen av reservoarene.
- b) Hva er sannsynligheten for at man finner olje kun i reservoar 1? Hva er sannsynligheten for å finne olje i reservoar 2 gitt at man finner i reservoar 1?
- c) Finn sannsynligheten for at man finner olje i kun ett av reservoarene og ikke i begge.

Oppgave 3

I forvaltningen av et fond brukes det mange ulike typer investeringer. En medarbeider i fondet har en stund hatt to industriselskaper i kikkerten, SOLENRG og SOLIN.

Gjennomsnittlig verdistigning pr. uke er registrert for disse to selskapene i fem uker.  
Verdistigninger:

uke:	1	2	3	4	5
SOLENRG ( $x_i$ ):	0.35	-0.13	0.20	0.70	0.30
SOLIN ( $y_i$ ):	0.28	0.01	0.41	0.8	0.45

- a) Hva er gjennomsnitt, median og datastandardavvik i de to datasettene?  
Lag et prikkdiagram for datasettene (begge i samme figur).

Det oppgis at:  $\sum_{i=1}^5 (x_i - \bar{x})^2 = 0.356$  og  $\sum_{i=1}^5 (y_i - \bar{y})^2 = 0.329$ .

Videre ser vi først på det ene selskapet, SOLENRG.

- b) Lag et konfidensintervall for forventet verdistigning til SOLENRG. La  $\mu_X = E(X_i)$  og betrakt dataene  $x_1, \dots, x_5$  som utfall av de tilfeldige variablene  $X_1, \dots, X_5$ . Hvilke antakelser må du gjøre?
- c) Er det grunnlag for å hevde at forventet verdistigning er positiv? Besvar spørsmålet ved å gjennomføre en test av  $H_0 : \mu_X = 0$  mot  $H_0 : \mu_X > 0$ . Bruk 5% signifikansnivå. Gjør klart hva som er teststørrelse, nullfordeling og forkastningsområde.

Vi vil så sammenligne forventet prisstigning for de to selskapene.

Det oppgis at "pooled" varians-estimatoren er:  $S_P^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$ .

- d) Lag et 95% konfidensintervall for forskjell i forventet verdistigning.
- e) Er det grunn til å hevde at det i virkeligheten er forskjellig forventet verdistigning i de to selskapene? Svar på spørsmålet ved å gjennomføre en to-utvalgs hypotesetest. Still opp null- og alternativhypotese.

De fem målingene for de to selskapene er fra de samme fem ukene, og det vil rimeligvis kunne være korrelasjon mellom slike målinger fra samme uke.

- f) Lag et 95% konfidensintervall for forskjellen i forventet verdistigning når du bruker parplanen, og forklar hvordan du kan bruke dette til å besvare spørsmålet i punkt e). Hva blir konklusjonen nå?  
Det oppgis at med  $d_i = x_i - y_i$ , er  $\sum_{i=1}^5 (d_i - \bar{d})^2 = 0.045$ .

#### Oppgave 4

Ved en medikamentell behandling av en bestemt sykdom gis pasienten en dose medisin der mengden er forsøkt tilpasset pasienten og sykdomstilstanden. Mengden medisin i pasientens kropp måles etter en viss tid. For ti ulike pasienter fikk man følgende resultat mellom dose ( $x_i$ ) og målt medisinkonsentrasjon ( $y_i$ ); begge i passende enhet:

$x_i$	6.5	6	4.5	4	5.5	7	8	8.5	9	5
$y_i$	7.2	6.4	5.2	4.2	6.3	7.7	8	9.5	9.3	6

Vi vil bruke regresjonsanalyse for å studere sammenhengen mellom gitt dose og målt konsentrasjon. En R-utskrift av enkel regresjonsanalyse av dataene er vist under.

```
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-0.53636 -0.20455  0.09545  0.18068  0.47727

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.75455     0.45806   1.647   0.138
x            0.97273     0.06937  14.022 6.49e-07

Residual standard error: 0.3564 on 8 degrees of freedom
Multiple R-squared:  0.9609,    Adjusted R-squared:  0.956
F-statistic: 196.6 on 1 and 8 DF,  p-value: 6.491e-07
```

- a) Lag et spredningsdiagram over dataene. Modellen for enkel regresjon er:  
 $Y_i = \alpha + \beta x_i + e_i, i = 1, \dots, 10$ . Hvilke antakelser gjøres?  
Hva er predikert konsentrasjon i kroppen ved en dose på 5, og hva med en dose på 9?
- b) Forklar hvilken parameter som sier noe om eventuell sammenheng mellom gitt dose og målt konsentrasjon. Er det virkelig sammenheng? Besvar spørsmålet ved å gjennomføre en hypotesetest. Still opp null- og alternativhypotese.

Stoffet i medisinen som blir gitt er ikke naturlig i menneskekroppen. Derfor bør det i denne situasjonen være slik at forventet målt konsentrasjon i kroppen er null når dosen som er gitt er null.

- c) Hvilken parameter i modellen for enkel regresjon sier noe om dette, og hva sier resultatene av regresjonsanalysen om dette?
- d) Dersom vi har følgende regresjonsmodell  $Y_i = \beta x_i + e_i, i = 1, \dots, 10$ , vis at minstekvadratersestimatoren for  $\beta$  er

$$\hat{\beta} = \frac{\sum_{i=1}^{10} x_i Y_i}{\sum_{i=1}^{10} x_i^2}.$$

### Oppgave 5

Vi tar nå utgangspunkt i situasjonen som beskrevet i 4d). Vi har da med standard antakelser at

$$\frac{\hat{\beta} - \beta}{\sqrt{S_e^2 / \sum_{i=1}^{10} x_i^2}} \sim t(9)$$

For dataene i oppgave 4 oppgis det at  $\sum_{i=1}^{10} x_i y_i = 472.4$  og  $\sum_{i=1}^{10} x_i^2 = 436$ , og estimert  $\text{Var}(e_i)$  med denne modellen er  $s_e^2 = 0.3889^2$ .

- a) Finn et 95% konfidensintervall for  $\beta$  på bakgrunn av dette.