Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4245 Statistikk
Vår 2011

Tavleøving nummer 12

Oppgave 1 Eksamen mai 2001, oppgave 1 av 4

Vi ser på konsentrasjonen av et giftstoff i havbunnen like utenfor en fabrikk. Miljøforskriftene sier at konsentrasjonen ikke skal overstige 12 [g/cm^3]. For å kontrollere dette tas prøver av havbunnen. Anta at en prøveverdi Y er normalfordelt med forventning μ og standardavvik σ. Sett $\mu = 13$ og $\sigma = 1{,}5$ i punkt a), og la de være ukjent i resten av oppgaven.

a)

b) De observerte måleverdiene er

$$11{,}7 \quad 12{,}4 \quad 12{,}8 \quad 12{,}9 \quad 13{,}3.$$

Kan vi på grunnlag av dette konkludere med at giftkonsentrasjonen på havbunnen like ved fabrikken er over 12? Formuler problemstillingen som en hypotesetest og utfør testen på signifikansnivå 0,05.

c) Det blir tatt 10 nye målinger, men denne gang i ulike avstander x fra fabrikken. Målingene er

x	10	20	30	40	50	60	70	80	90	100
y	9,9	11,1	9,3	10,6	9,2	9,3	10,0	9,2	10,3	8,4

I tillegg kommer de fem målingene i b). Her er $x = 0$. Det oppgis at $\sum x_i = 550$, $\sum (x_i - \bar{x})^2 = 18\,333{,}33$, $\sum y_i = 160{,}4$ og $\sum x_i y_i = 5245$.

Vi velger å utføre en lineær regresjonsanalyse med Y som avhengig variabel og x som uavhengig variabel. Modellen er

$$E(Y \mid x) = \alpha + \beta x.$$

Beregn estimatene for α og β. Forklar hva estimatet for α beskriver i dette eksemplet.

Regresjonsanalysen gir oss ikke grunnlag for å konkludere med at $\alpha > 12$. Hvorfor ikke? Sammenlign resultatet fra denne analysen med resultatet i b) og kommenter. Hvorfor kan det skje at to slike analyser gir forskjellig konklusjon? Bruk gjerne figur i forklaringen.

Oppgave 2 Aksjekurser – Eksamen juni 2004, oppgave 2 av 3

Selskapet Agderfrukt er notert på børsen. Vi antar at endringen X i verdien på en Agderfrukt-aksje i løpet av en dag er normalfordelt med forventningsverdi $\mu_X = 0.15$ kroner og standardavvik $\sigma_X = 0.60$ kroner. Har du en aksje i Agderfrukt vil $X > 0$ bety fortjeneste, mens $X < 0$ er tap.

a)

Selskapet Trønderfrukt er også notert på børsen. Vi kaller endringen på en Trønderfrukt-aksje i løpet av en dag for Y, der Y er normalfordelt med forventningsverdi $\mu_Y = 0.15$ kroner og standardavvik $\sigma_Y = 0.80$ kroner.

b) Vi ser på aksjekursendringen på den samme dagen for Agderfrukt og Trønderfrukt og antar i dette punktet at aksjekursendringene X og Y er uavhengige.

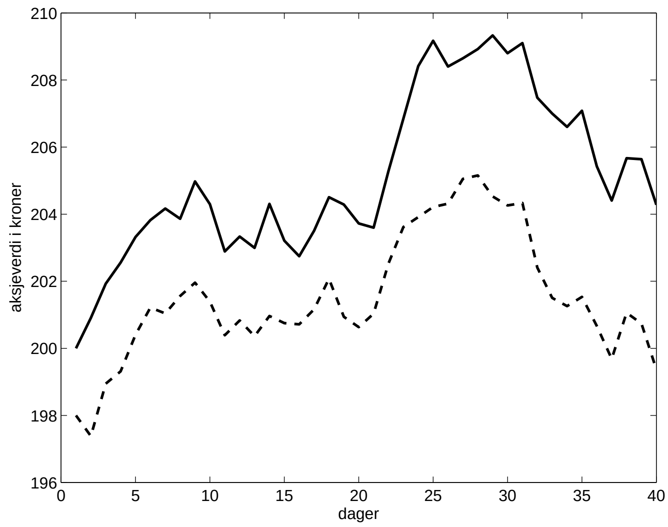
Idag er verdien til en Agderfrukt-aksje den samme som verdien til en Trønderfrukt-aksje. Vi ønsker å undersøke tre mulige strategier for aksjekjøp, der vi kjøper aksjer idag og selger imorgen.

i) Kjøp to aksjer i Agderfrukt.

ii) Kjøp en aksje i Agderfrukt og en aksje i Trønderfrukt.

iii) Kjøp to aksjer i Trønderfrukt.

Dersom du vil ha minst mulig risiko for investeringen din, hvilken av de tre investeringsstrategiene over vil du velge? Begrunn svaret.



Figuren viser utviklingen av aksjekursen til Agderfrukt (stiplet) sammen med aksjekursen til Trønderfrukt (heltrukket).

Kursendringen dag i for Agderfrukt kaller vi X_i, og vi antar at X_i er normalfordelt med forventning $\mu_X = 0.15$ kroner og standardavvik $\sigma_X = 0.60$ kroner.

Kursendringen dag i for Trønderfrukt kaller vi Y_i, og vi antar at Y_i er normalfordelt med forventning $\mu_Y = 0.15$ kroner og standardavvik $\sigma_Y = 0.80$ kroner.

Kursendringer for ulike dager antas å være uavhengige.

Vi sammenlikner de to selskapene ved å måle differansen mellom de daglige kursendringene, $D_i = X_i - Y_i$, og ta gjennomsnitt. Vi ser på 10 dager og får $\overline{D} = \frac{1}{10}\sum_{i=1}^{10} D_i = \frac{1}{10}\sum_{i=1}^{10}(X_i - Y_i)$.

c) Gir figuren grunn til å tro at endringene i de to aksjekursene samme dag, X_i og Y_i, er uavhengige?

Korrelasjonen mellom X_i og Y_i for disse to selskapene, $\rho(X_i, Y_i)$, er enten -0.5, 0.0 eller 0.5. Hvilken av disse verdiene virker mest rimelig fra figuren? Begrunn kort.

Hva blir forventningsverdi og varians for $\overline{D}$? Benytt verdien for korrelasjonen, $\rho(X_i, Y_i)$, som du valgte over.

Oppgave 3 Hubble — Eksamen mai 2006, oppgave 4 av 4

En viktig vitenskapelig oppdagelse fant sted i 1929 da Edwin Hubble oppdaget at universet er ekspanderende. Hubble's tallmateriale bestod blant annet av; x_i = avstanden til galakse i (målt i millioner lysår), og y_i = hastigheten til galakse i (målt i 1000 km/s). Verdiene Hubble benyttet i en av sine analyser er som følger:

Navn	Avstand, x_i	Hastighet, y_i
Virgo	22	1.2
Pegasus	68	3.8
Perseus	108	5.1
Coma Berenices	137	7.5
Ursa Major 1	255	14.9
Leo	315	19.2
Corona Borealis	390	21.4
Gemini	405	23.0
Bootes	685	39.2
Ursa Major 2	700	41.6
Hydra	1100	60.8

Det oppgis her at $\sum_{i=1}^{11} x_i = 4185$, $\sum_{i=1}^{11} y_i = 237.7$, $\sum_{i=1}^{11} x_i^2 = 2685141$ og $\sum_{i=1}^{11} x_i y_i = 152224$.

Hubble foreslo en modell for hastighet som funksjon av avstand på formen $y = \beta x$, der β senere har blitt kalt Hubble's konstant. En statistisk versjon av ligningen kan gis ved:

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \dots, 11, \tag{3.1}$$

der ε_i, $i = 1, \dots, 11$, er uavhengige og normalfordelte stokastiske variabler med forventning 0 og varians σ^2.

a) Vi vil i første omgang finne en estimator for β.

Bruk minste kvadraters metode (method of least squares) til å estimere β med utgangspunkt i ligning (3.1), og vis at estimatoren for β da blir gitt ved $\hat{\beta} = \frac{\sum_{i=1}^{11} x_i Y_i}{\sum_{i=1}^{11} x_i^2}$. Regn ut estimatet for β basert på dataene over.

Finn også forventning og varians til $\hat{\beta}$.

b) Anta at en annen galakse befinner seg en avstand $x_0 = 900$ millioner lysår borte.

Finn predikert hastighet, $\hat{y}_0$, til denne galaksen.

Utled et 95% prediksjonsintervall for en måling av hastigheten til denne galaksen. Det oppgis at $\sum_{i=1}^{11}(y_i - \hat{y}_i)^2 = 9.87$, der $\hat{y}_i = \hat{\beta} x_i$.

Oppgave 4 Automatisert laboratorium — Eksamen november 2002, oppgave 3 av 3

I eit laboratorium ynskjer ein å evaluere samanhengen mellom to variablar Y og x. Apparaturen er sett opp slik at ein kan fastsetje x for deretter å måle Y. Ein vel å nytte følgjande modell for samanhengen mellom variablane

$$Y = \alpha + \beta x + \varepsilon$$

der α og β er to ukjende koeffisientar og ε er ein tilfeldig variabel som er normalfordelt med forventing 0 og ukjend varians σ^2. La $x_1, x_2, \ldots, x_n$ vere n verdiar av variabelen x og $y_1, y_2, \ldots, y_n$ dei tilhøyrande verdiane som blir målt for Y. Desse skal sjåast på som realiseringar av n uavhengige variablar $Y_1, Y_2, \ldots, Y_n$. Minste kvadratsums (least squares) estimatorane, A og B, for koeffisientane α og β er då gitt ved

$$A = \frac{1}{n}\sum_{i=1}^{n} Y_i - B\bar{x} \qquad \text{og} \qquad B = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{der} \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

a) Vis at estimatorane A og B er forventingsrette estimatorar for α og β.

Bruk at kovariansen mellom $\bar{Y} = (1/n)\sum_{i=1}^{n} Y_i$ og B er 0 og utlei variansen til estimatorane A og B.

Kva sannsynsfordelingar har estimatorane A og B? Grunngjev svaret.

Laboratorieforsøka er svært arbeidskrevjande, men apparaturen er automatisert slik at forsøka kan utførast automatisk for ekvidistante verdiar av x. Sjå på to måleseriar med $n = 10$

$$\begin{array}{ll} \text{Serie 1:} & x_1 = 1, x_2 = 2, \ldots, x_{10} = 10 \\ \text{Serie 2:} & x_1 = 2, x_2 = 4, \ldots, x_{20} = 20 \end{array}$$

Målet med forsøket er å predikere Y_0 for $x_0 = 5.5$. Følgjande prediktor blir brukt: $\widehat{Y}_0 = A + Bx_0$.

b) Utlei variansen til $Y_0 - \widehat{Y}_0$.

Kva måleserie bør nyttast for å predikere Y_0 best mogeleg for $x_0 = 5.5$? Grunngjev og kommenter svaret du har funne.

Fasit

1. b) Forkaster H_0

3. a) 0.0567 **b)** 51.03, (48.5,53.5)

4. b) Bør benytte måleserie 2