Universitetet i Stavanger

EKSAMEN I: BMF100 SANNSYNLIGHETSREGNING OG STATISTIKK 1

VARIGHET: 4 TIMER DATO: 11. MAI 2012 BOKMÅL

TILLATTE HJELPEMIDLER:

KALKULATOR: **HP30S**, **Citizen SR-270X**, **Casio FX82** eller **TI-30**

OPPGAVESETTET BESTÅR AV 3 OPPGAVER PÅ 9 SIDER INKL. VEDLEGG

MERK: For å få bestått må også minst to av de *-merkede punktene 2c) 3a), og 3c) være bra besvart.

Oppgave 1

For tre hendelser A_1, A_2 og A_3 har vi at $P(A_1) = P(A_2) = P(A_3) = \frac{5}{30}$.

a) Dersom A_1, A_2 og A_3 er statistisk uavhengige begivenheter, finn sannsynligheten for at A_1 eller A_2 inntreffer, $P(A_1 \cup A_2)$, og finn sannsynligheten for at A_1 og A_2 inntreffer, $P(A_1 \cap A_2)$.
Hva er sannsynligheten for at A_1 inntreffer og ikke A_3?

b) Hvordan er statistisk uavhengighet mellom begivenheter definert?
Vi ser så på situasjonen der A_1 og A_2 er avhengige og antar at $P(A_1 \cap A_2) = \frac{2}{30}$, finn den betingede sannsynlighten $P(A_2|A_1)$.
Hva blir den betingede sannsynligheten for A_2 gitt at A_1 ikke inntreffer, $P(A_2|\overline{A_1})$?

Oppgave 2

Antall regndøgn

Vi skal se på hvor ofte det er dager med regn. Tabellen til høyre viser data fra høst- og vintermånedene fom. 1989-90, tom. 2008-09. Nærmere bestemt er det resultatet som vises for år 1990, samlet antall dager med nedbør fra og med 1. oktober 1989 til og med 31. mars 1990. Dvs.: det regnet 142 av døgnene i de seks høst-vintermånedene i 1989-1990. Tilsvarende for 1991, osv.

Skuddårsdager er tatt vekk slik at alle høst-vinterperiodene er like lange med 182 døgn.

Venstre halvdel av tabellen viser resultatene sortert etter år. Høyre halvdel viser de samme resultatene men nå sortert i stigende rekkefølge for antallene. Dataene er fra Bergen.

$n = 20$

Gjennomsnitt: $\overline{x} = 113.45$

Empirisk standardavvik:

$$s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2} = 18.94.$$

Tidsortert			Antallsortert	
år	i	$x_i =$ regndager	år	regndager
1990	1	142	2003	80
1991	2	94	1996	81
1992	3	123	2006	89
1993	4	116	1994	92
1994	5	92	1991	94
1995	6	133	2001	96
1996	7	81	2004	111
1997	8	115	2009	114
1998	9	116	1997	115
1999	10	122	1993	116
2000	11	135	1998	116
2001	12	96	2005	119
2002	13	133	2008	120
2003	14	80	1999	122
2004	15	111	1992	123
2005	16	119	1995	133
2006	17	89	2002	133
2007	18	138	2000	135
2008	19	120	2007	138
2009	20	114	1990	142

a) Lag et histogram eller et prikkdiagram over dataene. Finn første- og tredje kvartil, kvartilavvik og variasjonsbredde.
Hva er medianen for dataene? I hvilke situasjoner kan det være stor forskjell på median og gjennomsnitt?

Antall regndager er en diskret fordelt tilfeldig variabel. Men som en tilnærming, kan vi likevel bruke at antall regndager er normalfordelt (dvs. kontinuerlig fordelt).

b) Dersom vi bruker at antall regndager, X, er normalfordelt med forventning $\mu = 100$ og standardavvik $\sigma = 20$, finn $P(X < 80)$ og $P(X > 140)$.
Hva er sannsynligheten for at antall regndager er mellom 90 og 120?
Dersom X_1 og X_2 er uavhengige med samme fordeling som X, finn $P\left(\frac{X_1 + X_2}{2} > 140\right)$.

c*) Vi ser på dataene for de 20 periodene, $x_1, \ldots, x_{20}$, som utfall av 20 u.i.f. tilfeldige variable $X_1, \ldots, X_{20}$. Forventet antall regnedager, $\mu_X = \mathrm{E}(X_i)$ og $\mathrm{Var}(X_i) = \sigma_X^2$ er ukjente parametre. Vi antar normalfordeling: $X_i \sim N(\mu_X, \sigma_X)$

Lag et 95% konfidensintervall for forventet antall regnedager, μ_X.

d) I meteorologisk forstand er det "normale" antall regndager omtrent 100. Dette er basert på lange dataserier fra tidligere perioder.
Er det forventede antall regndager høyere enn normalt nå? Besvar spørsmålet ved å gjennomføre en hypotesetest av $H_0 : \mu_X = 100$ mot $H_1 : \mu_X > 100$. Bruk 5% signifikansnivå.
Gjør klart hva som er teststørrelse, nullfordeling og forkastningsområde.

e) Hvordan blir testen i d) dersom du kan anta at $\sigma_X = 20$ er kjent (og at X_i'ene er normalfordelte)?
For testen med kjent $\sigma_X = 20$ og normalantakelse, finn styrken i alternativet $\mu_X = 110$.

Oppgave 3

Tilsvarende data som i oppgave 2 for de 30 foregående periodene er også innhentet. Dvs.: vi har antall regndager i de seks høst-vintermånedene i 1959-1960 og antall regndager fram tom. høst-vintermånedene i 1988-1989. For disse $n_y = 30$ dataene er gjennomsnittet $\overline{y} = 102.27$ og empirisk standardavvik $s_y = \sqrt{\sum_{i=1}^{30}(y_i - \overline{y})^2} = 17.07$. Vi ser på dataene $y_1, \ldots, y_{30}$, som utfall av 30 u.i.f. tilfeldige variable $Y_1, \ldots, Y_{30}$ med forventning μ_Y og varians σ_Y^2. Tilsvarende for x_i'ene (som i oppg. 2).

a*) Vi vil sammenligne det gjennomsnittllige antall regndager i de to periodene. Vi har altså:

Tidsperiode	Antall data	gjennomsnitt	empirisk standardavvik
Fom. 1959-60 tom. 1988-89	$n_y = 30$	$\overline{y} = 102.27$	$s_y = \sqrt{\frac{1}{n_y-1}\sum_{i=1}^{30}(y_i - \overline{y})^2} = 17.07$
Fom. 1989-90 tom. 2008-09	$n_x = 20$	$\overline{x} = 113.45$	$s_x = \sqrt{\frac{1}{n_x-1}\sum_{i=1}^{20}(x_i - \overline{x})^2} = 18.94$

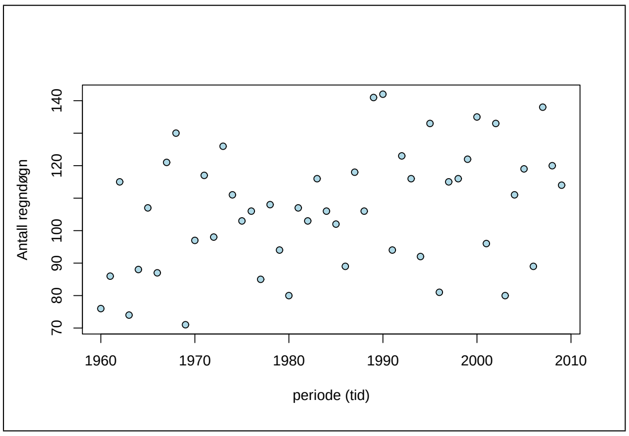
Er det grunnlag for å hevde at forventet antall regndager er høyere i den seineste perioden enn i den tidligste? Svar på spørsmålet ved å gjennomføre en hypotesetest av $H_0 : \mu_X = \mu_Y$ mot $H_1 : \mu_X > \mu_Y$. Bruk 5% signifikansnivå.
Gjør klart hva som er teststørrelse, nullfordeling og forkastningsområde.

b) Lag et konfidensintervall for forskjell i forventet antall regndager, $\mu_X - \mu_Y$.

Vi skal så på alle 50 dataene samlet. Et spredningsdiagram av antall regndager vs. periode (tid) er vist i figuren under.

Vi vil bruke regresjonsanalyse til å undersøke om det er noen trend i antall regndager. Vi lar nå x_i være periode nr i og y_i antall regndager i perioden. Vi setter første periode til 1960, andre til 1961 osv.

i	x_i	y_i
1	1960	76
2	1961	86
$\vdots$	$\vdots$	$\vdots$
50	2009	114



For disse dataene er:

$\overline{x} = 1984.5$, $s_{xx} = \sum_{i=1}^{50}(x_i - \overline{x})^2 = 10412.5$,

$\overline{y} = 106.74$, $s_{yy} = \sum_{i=1}^{50}(y_i - \overline{y})^2 = 16767.62$ og

$s_{xy} = \sum_{i=1}^{50}(x_i - \overline{x})(y_i - \overline{y}) = 4774.5$.

c*) Still opp modellen for enkel lineær regresjonsanalyse.
Estimer konstantledd og stigningstall.
Gi en fortolkning av stigningstallet.
Hva er predikert forventet antall regndager i 1970 og hva er det i 2005?

d) Er det grunnlag for å hevde at det er trend i forventet antall regndager? Besvar spørsmålet ved å gjennomføre en hypotesetest i regresjonsmodellen.
Det oppgis at estimatet av støyvariansen er $\dfrac{SSE}{50-2} = \dfrac{14578.3}{48} = 303.7$.

e) Hvor stor del av totalvariasjonen blir forklart av regresjonen?