



Høgskolen i Telemark

**EKSAMEN**

**6019 Statistikk II**

**09.12.2011**

Tid: *9-13 (4 timer)*

Målform: *Bokmål*

Sidetal: *6 (inkludert denne forsiden)*

Hjelpemiddel: *Alle trykte og skrevne samt kalkulator.*

Merknader:

Vedlegg:

**Eksamensresultata blir offentliggjort på Studentweb.**





## Oppgave 1

Vi skal teste om oppslutningen om to mulige regjeringsalternativer har endret seg siden forrige stortingsvalg i 2009. Ved valget i 2009 fikk de "rød-grønne" (Arbeiderpartiet, Sosialistisk Venstreparti og Senterpartiet) 47.8 % av stemmene, mens et mulig regjeringsalternativ av Høyre og Fremskrittspartiet fikk 40.1 %. 12.1 % av stemmene gikk til partier som ikke var med i noen av de to regjeringsalternativene.

I en meningsmåling gjengitt i Dagens Næringsliv 18. november 2011 svarte 793 velgere på hvilket parti de ville ha stemt på om det var stortingsvalg i dag. Av disse ville 319 stemt på et av de "rød-grønne" partiene, 375 på Høyre og Fremskrittspartiet, mens 99 velgere ville stemt på partier som ikke er med i de to mulige regjeringsalternativene.

- a) Bruk Karl Pearsons test til å teste om oppslutningen om de to mulige regjeringsalternativene er endret siden forrige stortingsvalg. Angi nullhypotese og alternativ hypotese, gjennomfør testingen og angi konklusjonen når signifikansnivået velges lik 1 %.

## Oppgave 2

Vi betrakter måling av radonstråling inne i husene i et boligområde (område I).

Det er foretatt målinger,  $X_1, X_2, \dots, X_{14}$  (Becquerel pr  $m^3$ ), i 14 forskjellige hus. Resultatene er gitt på neste side.

- a) I dette punktet antar vi at  $X_1, X_2, \dots, X_{14}$  er uavhenige med en symmetrisk kontinuerlig sannsynlighetstetthet som har median  $m$ .

Vi skal teste om  $m$  er over tiltaksgrensen på 200 Becquerel pr  $m^3$ . Formuler dette problemet som en hypotesetest. Gjennomfør testingen og angi konklusjonen. Bruk signifikansnivå 5%.

I et annet boligområde (område II) er det målt radonstråling i 10 forskjellige hus. De målte verdiene  $Y_1, Y_2, \dots, Y_{10}$  (Becquerel pr  $m^3$ ) er gitt på neste side.

- b) Vi antar i dette punktet at de til sammen 24 målte radonverdiene i de to boligområdene er uavhengige og *normalfordelte* med forventninger henholdsvis  $\mu_1$  ( $X$ -ene) og  $\mu_2$  ( $Y$ -ene) og felles standardavvik  $\sigma$ .

Estimer  $\sigma$  ut fra de 24 radonmålingene i de to boligområdene.

Tyder resultatene på at gjennomsnittlig radonstråling er lavere i område II enn i område I. Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen og angi konklusjonen på testen når signifikansnivået velges lik 5 %.



### Resultater område I

$X$  (Becquerel pr  $m^3$ ): 230 195 176 172 217 171 258 217 227 244 263 231 192 256

$$\bar{X} = 217.79 \quad \sum (X_i - \bar{X})^2 = 13434.4$$

### Resultater område II

$Y$  (Becquerel pr  $m^3$ ): 197 222 148 250 226 157 161 188 150 179

$$\bar{Y} = 187.80 \quad \sum (Y_i - \bar{Y})^2 = 11339.6$$

- c) Gjør kort rede for valg av modell/analyse som kan brukes på følgende problemstillinger knyttet til måling av radonstråling i bolighus:
- 1) Radonstråling inne i hus i fem boligområder skal sammenlignes. Siktemålet er å finne ut om det er forskjeller mellom områdene når det gjelder radonstråling i husene.  
  
Hvilken test kan en bruke om en antar at målingene er uavhengige og normalfordelte som i b), og hvilken test kan en bruke dersom en antar at målingene er uavhengige, men ikke normalfordelte?
  - 2) Flere metoder er aktuelle for måling av radonstråling i bolighus. I undersøkelsen av de fem boligområdene vil en benytte to forskjellige målemetoder: *kullboks* (metode I) og *sporfilm* (metode II). Siktemålet er nå både å finne ut om det er forskjeller mellom de fem områdene, og om de to forskjellige målemetodene gir forskjellig resultat. Hvilken modell og analyse kan en bruke om en antar at målingene er uavhengige og normalfordelte?  
  
Kan en i dette tilfellet bruke to-utvalgs t-test for å teste om de to målemetodene gir forskjellig resultat? Begrunn svaret.

### Oppgave 3

Vi skal studere tidsrekken  $Y_t$  ( $t = 1, 2, 3, \dots, 60$ ) der  $Y_t$  er produksjonen av elektrisk kraft i Telemark i måned nr  $t$ . Måleenheten er GWh (= gigawatt-time), og dataene er hentet fra hjemmesidene til Statistisk sentralbyrå (ssb.no). I denne oppgaven skal vi analysere data fra de 5 siste årene fram til og med juli 2011. Måned nr 1 er august 2006, nr 2 er september 2006 osv., mens måned nr 60 er altså juli 2011.

Dataene er gitt i utskriften fra Minitab til slutt i oppgaven. Studer denne før du svarer på spørsmålene i oppgaven.



Vi skal først bruke en enkel regresjonsmodell (Modell I):

$$Y_t = \beta_0 + \beta_1 t + U$$

- a) Skriv opp den estimerte regresjonsmodellen og forklar kort hva de estimerte parametrene uttrykker.

Gir bruk av denne modellen grunnlag for å påstå at produksjonen av elektrisk kraft i Telemark har økt eller avtatt i denne fem-årsperioden? Formuler dette spørsmålet som en hypotesetest og angi konklusjonen ut fra resultatene i Minitab-utskriften. Bruk signifikansnivå 5 %.

I denne modellen er  $R^2$  (R-Sq i Minitab) 5.9 %. Forklar kort hva dette innebærer.

Produksjonen og forbruket av elektrisk kraft er større om vinteren enn om sommeren. Vi skal bruke en modell som tar hensyn til dette (Modell II):

$$Y_t = \beta_0 + \beta_1 t + \beta_2 Q + U$$

der  $Q$  er 1 i vintermånedene november, desember, januar, februar, mars og april og 0 ellers.

- b) Skriv også her opp den estimerte modellen og forklar kort hva de estimerte parametrene uttrykker. Kan en i denne modellen påstå at produksjonen av elektrisk kraft i Telemark har økt eller avtatt i denne fem-årsperioden, og er sesongeffekten signifikant? Formuler selv relevante hypoteser og angi konklusjonene med 5 % signifikansnivå.

Hva er autokorrelasjon? Hvilken størrelse i Minitab-utskriften gir oss informasjon om eventuell autokorrelasjon. Hva kan du si om autokorrelasjonen i modell II?

Bruk modell II til å beregne prediksjoner for produksjonen av elektrisk kraft i Telemark for resten av 2011 (august-desember).

### Minitab – Oppgave 3

| Row | Måned      | Y    | t  | Q |
|-----|------------|------|----|---|
| 1   | Aug. 2006  | 782  | 1  | 0 |
| 2   | Sept. 2006 | 942  | 2  | 0 |
| 3   | Okt. 2006  | 1051 | 3  | 0 |
| 4   | Nov. 2006  | 1083 | 4  | 1 |
| 5   | Des. 2006  | 1131 | 5  | 1 |
| 6   | Jan. 2007  | 1321 | 6  | 1 |
| 7   | Feb. 2007  | 1409 | 7  | 1 |
| 8   | Mars 2007  | 1235 | 8  | 1 |
| 9   | April 2007 | 1047 | 9  | 1 |
| 10  | Mai 2007   | 1180 | 10 | 0 |
| 11  | Juni 2007  | 824  | 11 | 0 |
| 12  | Juli 2007  | 1089 | 12 | 0 |
| 13  | Aug. 2007  | 989  | 13 | 0 |
| 14  | Sept. 2007 | 697  | 14 | 0 |
| 15  | Okt. 2007  | 867  | 15 | 0 |
| 16  | Nov. 2007  | 1100 | 16 | 1 |
| 17  | Des. 2007  | 1283 | 17 | 1 |
| 18  | Jan. 2008  | 1398 | 18 | 1 |
| 19  | Feb. 2008  | 1410 | 19 | 1 |



|    |            |      |    |   |
|----|------------|------|----|---|
| 20 | Mars 2008  | 1527 | 20 | 1 |
| 21 | April 2008 | 1323 | 21 | 1 |
| 22 | Mai 2008   | 1439 | 22 | 0 |
| 23 | Juni 2008  | 1195 | 23 | 0 |
| 24 | Juli 2008  | 949  | 24 | 0 |
| 25 | Aug. 2008  | 864  | 25 | 0 |
| 26 | Sept. 2008 | 993  | 26 | 0 |
| 27 | Okt. 2008  | 1140 | 27 | 0 |
| 28 | Nov. 2008  | 1355 | 28 | 1 |
| 29 | Des. 2008  | 1458 | 29 | 1 |
| 30 | Jan. 2009  | 1353 | 30 | 1 |
| 31 | Feb. 2009  | 1368 | 31 | 1 |
| 32 | Mars 2009  | 1428 | 32 | 1 |
| 33 | April 2009 | 1115 | 33 | 1 |
| 34 | Mai 2009   | 862  | 34 | 0 |
| 35 | Juni 2009  | 587  | 35 | 0 |
| 36 | Juli 2009  | 506  | 36 | 0 |
| 37 | Aug. 2009  | 601  | 37 | 0 |
| 38 | Sept. 2009 | 590  | 38 | 0 |
| 39 | Okt. 2009  | 887  | 39 | 0 |
| 40 | Nov. 2009  | 1188 | 40 | 1 |
| 41 | Des. 2009  | 1486 | 41 | 1 |
| 42 | Jan. 2010  | 1611 | 42 | 1 |
| 43 | Feb. 2010  | 1418 | 43 | 1 |
| 44 | Mars 2010  | 1413 | 44 | 1 |
| 45 | April 2010 | 1033 | 45 | 1 |
| 46 | Mai 2010   | 829  | 46 | 0 |
| 47 | Juni 2010  | 412  | 47 | 0 |
| 48 | Juli 2010  | 400  | 48 | 0 |
| 49 | Aug. 2010  | 694  | 49 | 0 |
| 50 | Sept. 2010 | 807  | 50 | 0 |
| 51 | Okt. 2010  | 991  | 51 | 0 |
| 52 | Nov. 2010  | 1138 | 52 | 1 |
| 53 | Des. 2010  | 1309 | 53 | 1 |
| 54 | Jan. 2011  | 1298 | 54 | 1 |
| 55 | Feb. 2011  | 1180 | 55 | 1 |
| 56 | Mars 2011  | 1190 | 56 | 1 |
| 57 | April 2011 | 717  | 57 | 1 |
| 58 | Mai 2011   | 542  | 58 | 0 |
| 59 | Juni 2011  | 391  | 59 | 0 |
| 60 | Juli 2011  | 914  | 60 | 0 |

## Modell I

### Regression Analysis: Y versus t

The regression equation is

$$Y = 1189 - 4,38 t$$

| Predictor | Coef    | SE Coef | T     | P     |
|-----------|---------|---------|-------|-------|
| Constant  | 1189,25 | 80,43   | 14,79 | 0,000 |
| t         | -4,380  | 2,293   | -1,91 | 0,061 |

S = 307,616    R-Sq = 5,9%    R-Sq(adj) = 4,3%



## Modell II

### Regression Analysis: Y versus t; Q

The regression equation is

$$Y = 967 - 4,38 t + 444 Q$$

| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | 967,40 | 61,50   | 15,73 | 0,000 |
| t         | -4,380 | 1,572   | -2,79 | 0,007 |
| Q         | 443,70 | 54,45   | 8,15  | 0,000 |

S = 210,902    R-Sq = 56,5%    R-Sq(adj) = 55,0%

#### Unusual Observations

| Obs | t    | Y      | Fit    | SE Fit | Residual | St Resid |
|-----|------|--------|--------|--------|----------|----------|
| 22  | 22,0 | 1439,0 | 871,0  | 40,8   | 568,0    | 2,74R    |
| 57  | 57,0 | 717,0  | 1161,4 | 56,7   | -444,4   | -2,19R   |

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 1,33569