



**EKSAMENSOPPGAVE I SØK1004**  
**STATISTIKK FOR ØKONOMER**

**Faglig kontakt under eksamen: Päivi Lujala**

**Tlf.: 9 67 60**

**Eksamensdato:** Onsdag 2. juni 2010

**Eksamenssted:** Dragvoll

**Eksamenstid:** 4 timer

**Studiepoeng:** 7,5

**Tillatte hjelpemidler:** Alle trykte og håndskrevne hjelpemidler, samt godkjent kalkulator Citizen SR-270x el. HP 30S.

**Sensur:** 22. juni 2010

Eksamen består av 4 oppgaver med delspørsmål som alle skal besvares. Vekting av oppgavene er gitt i parentes.

Antall sider bokmål: 3

Antall sider nynorsk: 3

## Oppgave 1 (20%)

Lungeavdelingen ved et sykehus regner med at 7 % av den voksne befolkningen lider av KOLS. Lungeavdelingen regner også med at 90 % av de som lider av KOLS, røyker, mens av dem som ikke har KOLS, røyker 25 %.

Vi trekker ut tilfeldig en voksen person.  $K$  er utfallet at personen har en lungesykdom og  $R$  er utfallet at personen røyker.

- Formuler opplysningene i oppgaven som sannsynligheter. Vis at  $Pr(R)=0,296$ .
- Regn ut  $Pr(K/R)$  og forklar kort hva denne sannsynligheten sier oss
- Hva er sannsynligheten for at en person som ikke røyker har KOLS?

## Oppgave 2 (30%)

Farten til en tilfeldig bil  $X$  (km/t) på en bestemt veistrekning kan oppfattes som en normalfordelt stokastisk variabel. Tidligere målinger har vist at gjennomsnittsfarten på denne strekningen var 77 km/t.

En holdningskampanje er gjennomført, og en vil undersøke om fartsnivået på strekningen har avtatt. På en dag ble det målt fart for 12 biler. Målingene  $X_1, X_2, \dots, X_{12}$  antas å være uavhengige og normalfordelte variabler med forventning  $\mu$ .

- Tyder resultatene på at fartsnivået på strekningen nå er lavere enn 77 km/t? Formuler dette spørsmålet som en hypotesetest. Bruk signifikansnivå 5 %. Gjennomfør testingen og angi konklusjonen når målt gjennomsnittsfart for de 12 bilene var 73,33 km/t og populasjonsstandardavvik er  $\sigma=10,0$  km/t.
- Lag et 98 % konfidensintervall for utvalgsgjennomsnittet for de 12 målte bilene basert på tallene gitt i 2a. Kan du bruke dette for å gjennomføre en hypotesetest med et gitt signifikansnivå? Begrunn i så fall hva slags test og vis resultatet av testen.
- Anta nå at  $\sigma$  er ukjent, men at du kjenner standardavviket for utvalget som er 10,0 km/t. Gjennomfør den samme hypotesetesten som i oppgave 2a (samme gjennomsnittsfart og signifikansnivå) og angi konklusjonen.
- Hvis utvalgsstandardavviket hadde endret seg (uten at noe annet gjorde det), måtte det vært større eller mindre enn 10 for å endre konklusjonen i oppgave 2c? Forklar, med ord, hvorfor, og regn så ut hvor mye utvalgsstandardavviket måtte vært for å endre konklusjonen på testen. (Vi antar at gjennomsnittsfart, utvalgsstørrelse og signifikansnivå er som i oppgave 2c.)

### Oppgave 3 (25%)

Følgende tabell angir antall epletrær solgt i begynnelsen av juni i en hagebutikk i Trondheim (D) og utetemperatur (T) (helgene er ikke inkludert).

<b>D</b>	92	70	72	61	75	50	84
<b>T</b>	22	23	19	22	24	18	26

Anta at  $D = \alpha + \beta T + \varepsilon$ , hvor  $\varepsilon$  er et restledd.

- Finn Minste kvadraters metode estimatorene av  $\alpha$  og  $\beta$ . Tolk resultatene.
- I følge Meteorologisk institutt kan utetemperaturen i Trondheim variere mellom 4 og 28 grader i begynnelsen av juni. Prediker antall solgte epletrær for de to ytterpunktene. Hvor troverdige er resultatene?

### Oppgave 4 (25%)

I følge forskning finnes det et bytteforhold (trade-off) mellom lengden på nattsøvn og tiden brukt i inntektsgivende arbeid. En gruppe på 17 kvinner rapporterte totalt antall timer de hadde sovnet (*Søvn*) og jobbet (*Arbeid*) i løpet av en uke. Forskerne brukte deretter data til å estimere en modell hvor lengden på søvn er responsvariabelen. De benyttet Minste kvadraters metode. Følgende er utdrag fra Excelutskrift:

Variansanalyse			
	<i>fg</i>	<i>SK (SS)</i>	
Regresjon	1		
Residualer	15	634,59	
Totalt	16	813,42	

  

	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>
Skjæringspunkt	62,865	4,864	12,923
Arbeid	-0,224	0,109	-2,056

#### AVVIK (UTDATA)

<i>Observasjon</i>	<i>Fremskrevet (estimert)</i>	
	<i>Søvn</i>	<i>Residualer</i>
1	53,3	-4,2
2	51,0	

I tillegg vet vi følgende for datasettet:

	Søvn	Arbeid
Minimumsverdi	38,9	6,2
Maksimumsverdi	66,5	72,1
Gjennomsnittet	53,4	42,2
Median	54,9	45

## SØK1004 Statistikk for økonomer

- a) Skriv den estimerte regresjonslinjen på formatet  $\hat{Y} = a + bX$ .
- b) Hva forteller skjæringspunktet i denne ligningen?
- c) Hvis en person øker den ukentlige arbeidsinnsatsen med 2 timer, hvor mye kan vi forvente at det totale antall timer sovet i løpet av en uke forandrer seg basert på den estimerte modellen?
- d) For den første observasjonen i datasettet, hva er det rapporterte antall timer sovet per uke ( $Y_1$ )?
- e) Regn ut forklart variasjon og restvariasjon for den andre observasjonen som rapporterte 53 timer på jobb og søvn på 55 timer.
- f) Hvor mye av variasjon i *Søvn* er forklart av *Arbeid*? Begrunn svaret ved å bruke tallene gitt i Excelutskrift.

## Oppgåve 1 (20%)

Lungeavdelinga ved et sjukehus reknar med at 7 % av den vaksne befolkninga lider av KOLS. Lungeavdelinga reknar også med at 90 % av dei som lir av KOLS, røyker, mens av dei som ikkje har KOLS, røyker 25 %.

Vi trekker ut tilfeldig ein vaksen person.  $K$  er utfallet at personen har ein lungesjukdom og  $R$  er utfallet at personen røyker.

- Formuler opplysningane i oppgåva som sannsyn. Vis at  $Pr(R)=0,296$ .
- Rekn ut  $Pr(K/R)$  og forklar kort kva dette sannsynet seier oss.
- Kva er sannsynet for at en person som ikkje røyker har KOLS?

## Oppgåve 2 (30%)

Farten til ein tilfeldig bil  $X$  (km/t) på ei bestemd vegstrekning kan oppfattas som ein normalfordelt stokastisk variabel. Tidlegare målingar har vist at gjennomsnittsfarten på denne strekninga var 77 km/t.

Ein haldningskampanje er gjennomført, og ein vil undersøkje om fartsnivået på strekninga har avteke. På ein dag vart det målt fart for 12 bilar. Vi antar at målingane  $X_1, X_2, \dots, X_{12}$  er uavhengige og normalfordelte variablar med forventning  $\mu$ .

- Tydar resultatata på at fartsnivået på strekninga nå er lågare enn 77 km/t? Formuler dette spørsmålet som ein hypotesetest. Bruk signifikansnivå 5 %. Gjennomfør testinga og angje konklusjonen når målt gjennomsnittsfart for dei 12 bilane var 73,33 km/t og populasjonsstandardavvik er  $\sigma=10,0$  km/t.
- Lag eit 98 % konfidensintervall for utvalgsgjennomsnittet for dei 12 målte bilane basert på talla gjett i 2a. Kan du bruke dette for å gjennomføre en hypotesetest med eit gitt signifikansnivå? Grunnge i så fall kva slags test og vis resultatet av testen.
- Anta nå at  $\sigma$  er ukjent, men at du kjenner standardavviket for utvalet som er 10,0 km/t.. Gjennomfør den same hypotesetesten som i oppgåve 2a (same gjennomsnittsfart og signifikansnivå) og angje konklusjonen.
- Om utvalsstandardavviket hadde endra seg, ville det ha måtte vere større eller mindre enn 10 for å endre konklusjonen i oppgåve 2c? Forklar, med ord, kvifor og rekn så ut kor mykje utvalsstandardavviket måtte vere for å endre konklusjonen på testen. (Vi antar at gjennomsnittsfart, utvalsstørrelse og signifikansnivå er som i oppgåve 2c.)

### Oppgave 3 (25%)

Fylgjande tabell gjer talet på epletre selt i byrjinga av juni i ein hagebutikk i Trondheim (D) og utetemperatur (T) (helgene er ikkje inkludert).

<b>D</b>	92	70	72	61	75	50	84
<b>T</b>	22	23	19	22	24	18	26

Anta at  $D = \alpha + \beta T + \varepsilon$ , kor  $\varepsilon$  er eit restledd.

- Finn Minste kvadraters metode estimatorane av  $\alpha$  og  $\beta$ . Tolk resultatata.
- I fylgje Meteorologisk institutt kan utetemperaturen i Trondheim variere mellom 4 og 28 grader i byrjinga av juni. Prediker talet på selde epletre for dei to ytterpunkta. Kor truverdige er resultatata?

### Oppgave 4 (25%)

I fylgje forskning finnes det eit bytteforhold (trade-off) mellom lengda på nattsøvn og tida brukt i inntektsgjevande arbeid. Ei gruppe på 17 kvinner rapporterte totalt tal på timar dei hadde søvt (*Søvn*) og jobba (*Arbeid*) i løpet av ei veke. Forskarane brukte deretter data til å estimere ein modell kor lengda på søvn er responsvariabelen. Dei nytta Minste kvadraters metode. Fylgjande er utdrag frå Excelutskrift:

Variansanalyse

	<i>fg</i>	<i>SK (SS)</i>
Regresjon	1	
Residualer	15	634,59
Totalt	16	813,42

	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>
Skjæringspunkt	62,865	4,864	12,923
Arbeid	-0,224	0,109	2,056

AVVIK (UTDATA)

<i>Observasjon</i>	<i>Fremskrevet (estimert)</i>	
	<i>Søvn</i>	<i>Residualer</i>
1	53,3	-4,2
2	51,0	

I tillegg veit vi fylgjande for datasettet:

	Søvn	Arbeid
Minimumsverdi	38,9	6,2
Maksimumsverdi	66,5	72,1
Gjennomsnittet	53,4	42,2
Median	54,9	45

- a) Skriv den estimerte regresjonslinja på formatet  $\hat{Y} = a + bX$ .
- b) Kva fortel skjeringspunktet i denne likninga?
- c) Om ein person aukar den vekelege arbeidsinnsatsen med 2 timer, kor mykje kan vi forvente at det totale tal på timer søvt i løpet av ei uke endrar seg basert på den estimerte modellen?
- d) For den første observasjonen i datasettet, kva er det rapporterte tall timar søvt per veke (Y1)?
- e) Regn ut forklart variasjon og restvariasjon for den andre observasjonen som rapporterte 53 timer på jobb og søvn på 55 timer.
- f) Kor mykje av variasjon i Søvn er forklart av Arbeid? Grunnge svaret ved å bruke tala gjett i Excelutskrift.