

Eksamen i: Stat100 - Statistikk.

Tid: Mandag 29. april 2002, kl. 9.00 - 14.00 (fem timer).

Alle hjelpemidler er tillatt, også lommekalkulator (hjelpemiddelkode C1).

Oppgaveteksten er på tre sider.

Ved bedømmelsen teller alle delpunkter likt.

Kontakt under eksamen er Torfinn Torp, telefon 64948862.

Oppgave 1 Kosthold og kreft

Mange har stilt seg spørsmålet om det er noen sammenheng mellom kosthold og kreft. En sammenheng som er undersøkt er om, og eventuelt hvordan, kostholdsfaktoren fettkonsum innvirker på dødeligheten av prostatakreft. Tabellen nedenfor inneholder fettkonsum (gram per dag) og dødsrate fra prostatakreft (per 100.000 personer) for 20 ulike land.

| Land | Fettkonsum | Dødsrate | Land | Fettkonsum | Dødsrate |
|-------------|------------|----------|--------------|------------|----------|
| El Salvador | 38 | 0,9 | Spania | 97 | 10,1 |
| Filippinene | 29 | 1,3 | Portugal | 73 | 11,4 |
| Japan | 42 | 1,6 | Finland | 112 | 11,1 |
| Mexico | 57 | 4,5 | Storbritania | 143 | 12,4 |
| Hellas | 96 | 4,8 | Canada | 142 | 13,4 |
| Polen | 93 | 6,4 | Frankrike | 137 | 14,4 |
| Israel | 95 | 8,4 | Nederland | 152 | 14,4 |
| Romania | 67 | 8,8 | Østerrike | 119 | 13,9 |
| Italia | 86 | 9,4 | USA | 147 | 16,3 |
| Ungarn | 100 | 13,1 | Norge | 133 | 16,8 |

For disse 20 landene gjelder: Gjennomsnittlig fettkonsum = 97,9 og gjennomsnittlig dødsrate = 9,67.

- a) Framstill disse dataene i en hensiktsmessig figur. Synes trenden i dødsraten å være lineær som funksjon av fettkonsum?**

Tilpasning av en lineær regresjonsmodell, med alle de vanlige forutsetningene, til datamaterialet ovenfor gav blant annet følgende resultater fra statistikkpakken Minitab:

| Regression Analysis | | | | | Analysis of Variance | | |
|---------------------|---------|---------|-------|-------|----------------------|----|--------|
| Predictor | Coef | StDev | T | P | Source | DF | SS |
| Constant | -1,680 | 1,488 | -1,13 | 0,274 | Regression | 1 | 371,10 |
| Fettkonsum | 0,11593 | 0,01421 | 8,16 | 0,000 | Residual Error | 18 | 100,36 |
| | | | | | Total | 19 | 471,46 |

- b) Tegn den tilpassede linjen inn i figuren fra punkt a). Hvor stor del av variasjonen i dødsratene er forklart med den tilpassede modellen?**
- c) Sverige har et fettkonsum på 132 gram per dag. Finn et 95% prediksjonsintervall for dødsraten i Sverige.**
- d) Danmark har et fettkonsum på 156 gram per dag. Finn et 95% konfidensintervall for forventet differanse mellom Danmarks og Sveriges dødsrater.**

Oppgave 2 Dekkslitasje

Slitestykken for to ulike dekktyper, A og B, for bakhjul på traktorer skal sammenlignes. Det er planlagt og utført et forsøk. Begge dekktypene brukes på hver av 10 traktorer. For hver traktor er det trukket lodd om A eller B skal monteres og brukes på høyre bakhjul. Den andre typen brukes på venstre bakhjul. Etter et år i vanlig bruk hos traktorenes respektive eiere måles slitasjen (i millimeter). Forsøksplan og målt slitasje er gitt nedenfor.

| Traktor | Høyre bakhjul | Venstre bakhjul |
|---------|---------------|-----------------|
| 1 | A; 3,9 mm | B; 3,5 mm |
| 2 | B; 2,6 mm | A; 2,8 mm |
| 3 | B; 3,0 mm | A; 3,3 mm |
| 4 | B; 3,9 mm | A; 4,1 mm |
| 5 | A; 3,4 mm | B; 3,3 mm |
| 6 | B; 3,1 mm | A; 3,2 mm |
| 7 | A; 3,0 mm | B; 2,5 mm |
| 8 | A; 3,7 mm | B; 3,3 mm |
| 9 | B; 1,9 mm | A; 2,1 mm |
| 10 | A; 3,1 mm | B; 3,0 mm |

- a) Gi en kort begrunnelse for hvorfor vi i denne situasjonen kan og bør bruke en paret test for å sammenligne slitasjen på de to dekktypene. Spesifiser de forutsetningene som ligger til grunn for en paret t-test, utfør testen og konkluder.
- b) Tenk deg at du blir bedt om å planlegge et tilsvarende forsøk for å sammenligne slitestykken på fire ulike dekktyper for personbiler. Til disposisjon får du 10 ulike biler som skal brukes av sine respektive eiere i forsøksperioden. Beskriv og skisser kort den forsøksplanen som bør brukes. Hva kalles forsøksplanen og hvilken analysemetode vil du bruke for å analysere resultatene fra forsøket?

Oppgave 3 Snorking

I en studie for å undersøke om det synes å være noen sammenheng mellom hyppighet av snorking og forekomst av hjertesykdom er det samlet opplysninger fra 2484 personer. Dette datamaterialet (observerte antall) er gitt ved:

| <u>Hjertesykdom</u> | <u>H y p p i g h e t a v s n o r k i n g</u> | | | | <u>Totalt</u> |
|---------------------|--|------------------|-------------------------|------------------|---------------|
| | <u>Aldri</u> | <u>Av og til</u> | <u>Nesten hver natt</u> | <u>Hver natt</u> | |
| Ja | 24 | 35 | 21 | 30 | 110 |
| Nei | 1355 | 603 | 192 | 224 | 2374 |
| Totalt | 1379 | 638 | 213 | 254 | 2484 |

- a) Formuler problemstillingen ovenfor som en hypotese med alternativ.
- b) Bruk datamaterialet ovenfor til å teste hypotesen fra punkt a).

Oppgave 4 Radiosamtaler

I kontrolltårnet på en bestemt flyplass skal vi la X betegne antall påbegynte radiosamtaler i et tilfeldig valgt ti minutters intervall.

- a) Hva må være oppfylt for at X skal være Poissonfordelt?
- b) Anta nå at X er Poissonfordelt med forventning lik åtte. Finn $P(X = 14)$, $P(X \geq 14)$ og $P(X \geq 14 \mid X \geq 10)$.

For å undersøke fordelingsegenskapene til X nærmere har en registrert antall påbegynte radiosamtaler i hvert av $n = 1000$ disjunkte ti minutters intervaller. Det gav datamaterialet:

| Antall samtaler påbegynt | Registrert i antall intervaller | Antall samtaler påbegynt | Registrert i antall intervaller | Antall samtaler påbegynt | Registrert i antall intervaller |
|--------------------------|---------------------------------|--------------------------|---------------------------------|--------------------------|---------------------------------|
| 0 | 0 | 7 | 135 | 14 | 20 |
| 1 | 3 | 8 | 141 | 15 | 14 |
| 2 | 10 | 9 | 127 | 16 | 6 |
| 3 | 30 | 10 | 93 | 17 | 2 |
| 4 | 56 | 11 | 75 | 18 | 0 |
| 5 | 94 | 12 | 49 | 19 | 1 |
| 6 | 120 | 13 | 23 | 20 | 1 |

- c) La μ være forventningen til X , og la $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ der X_i er antall påbegynte samtaler i intervall i . Gi en kort begrunnelse for hvorfor det er naturlig å bruke \bar{X} som estimator for μ . Bruk datamaterialet ovenfor til å beregne dette estimatet.
- d) Skisser kort hvordan datamaterialet ovenfor kan brukes for å teste om X faktisk er Poissonfordelt med forventning lik åtte (du skal ikke utføre selve testen).
- e) Anta $\mu = 8$ og la T være tiden mellom starten på to tilfeldig valgte og på hverandre følgende samtaler. Bestem sannsynlighetsfordelingen til T ved å finne uttrykket for $F(t) = P(T \leq t) = 1 - P(T > t)$, der t er et generelt reellt tall. Beregn $F(2)$.

Håvard Rue

John Tyssedal

Torfinn Torp