

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

- Eksamen i: STK 1000 — Innføring i anvendt statistikk.
- Eksamensdag: Mandag 7. juni 2004.
- Tid for eksamen: 09.00 – 12.00
- Oppgavesettet er på 3 sider.
- Vedlegg: Ingen.
- Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics", 3. el. 4. utgave. Ordliste for bruk i STK 1000, kalkulator.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1.

Etter utvidelsen av EU 1. mai 2004 er en professor i statsvitenskap interessert i om studentene ved Universitetet i Oslo ønsker norsk medlemskap i EU eller ikke.

- Hun velger å bruke studentregistret over de studentene som har betalt semesteravgiften og trekker tilfeldig 400 kvinnelige og 400 mannlige studenter til å delta i undersøkelsen. Er dette et fornuftig forsøksopplegg? Begrunn svaret kort.
- La de tilfeldige variable Y og X betegne henholdsvis antall av de 400 kvinnelige og 400 mannlige studentene som er for norsk medlemskap i EU. Professoren antar at Y og X er binomisk fordelt henholdsvis $(400, p)$ og $(400, q)$ der p og q er henholdsvis sannsynligheten for at en tilfeldig valgt kvinnelig og mannlig student ved Universitetet i Oslo er for norsk medlemskap i EU. Er dette fornuftige antagelser? Begrunn svaret.

(Fortsettes side 2.)

- c) Foreslå en estimator for p og vis at den er forventningsrett.
- d) Anta $p = 0.5$. Professoren ønsker å beregne en tilnærmet sannsynlighet for at Y er større eller lik 220. Utfør denne beregningen og begrunn den tilnærmelsen som blir brukt.
- e) Foreslå en forventningsrett estimator for $p - q$ og beregn standardavviket til denne estimatoren. Du kan anta at Y og X er uavhengige. Hva blir estimatet for $p - q$ hvis en observerer $Y = y = 180$ og $X = x = 240$? Kommenter resultatet.

Oppgave 2.

34 studenter i et statistikkurs ved et amerikansk universitet gjennomførte følgende forsøk. For hver enkelt student registrerte en først kjønn (en setter mann = 1, kvinne = 2), høyde i tommer (1 tomme = 2.54 cm) og vekt i pund (1 pund = 0.454 kg). For hver student målte en så hvilepuls i antall hjerteslag per minutt. Deretter ble alle studentene satt til å løpe på stedet i et minutt, hvoretter en umiddelbart etterpå registrerte det vi kaller belastningspuls i antall hjerteslag per minutt. Som responsvariabel i denne oppgaven vil vi for hver enkelt student benytte belastningspuls minus hvilepuls. Denne differensen kalles pulsendring og gir for små verdier indikasjon på god fysisk form.

- a) Betrakt først vekten i pund som eneste forklaringsvariabel. Sett opp en modell for enkel lineær regresjon og gjør rede for antagelsene en baserer seg på.

I det følgende er det gitt en Minitab-utskrift av den enkle lineære regresjonsanalysen.

Regression Analysis: pulsendring versus vekt

The regression equation is
pulsendingring = 70.8 - 0.342 vekt

| Predictor | Coef | SE Coef | T | P |
|-----------|----------|---------|-------|-------|
| Constant | 70.79 | 15.23 | 4.65 | 0.000 |
| vekt | -0.34194 | 0.09930 | -3.44 | 0.002 |

S = 13.10 R-Sq = 26.4%

(Fortsettes side 3.)

- b) Finn et 95% konfidensintervall (intervallestimat) for stigningskoeffisienten for populasjons-regresjonslinjen. Hvordan kan du bruke intervallet til å avgjøre om stigningskoeffisienten er signifikant forskjellig fra null?
- c) I Minitab-utskriften står det at $R - Sq = 26.4\%$. Gjør rede for hva det betyr.
- d) Betrakt nå også høyden i tommer og kjønn som forklaringsvariable i tillegg til vekten i pund. Sett opp en multippel regresjonsmodell og gjør rede for de antagelsene en baserer seg på. Ved analysen av denne baserer en seg på fordelingen til en bestemt tilfeldig variabel. Hvordan avviker denne fordelingen fra fordelingen brukt i den enkle lineære regresjonsanalysen i b)?
- e) I det følgende er det gitt en Minitab-utskrift for regresjonsmodellen i d).

Regression Analysis: pulsendring versus vekt; høyde; kjønn

The regression equation is

pulsendring = 21.0 - 0.136 vekt - 0.007 høyde + 14.5 kjønn

| Predictor | Coef | SE Coef | T |
|-----------|---------|---------|-------|
| Constant | 21.03 | 62.84 | 0.33 |
| vekt | -0.1361 | 0.1505 | -0.90 |
| høyde | -0.0069 | 0.9535 | -0.01 |
| kjønn | 14.466 | 6.476 | 2.23 |

S = 12.47

Kommenter de viktigste resultatene fra denne utskriften. Sammenlign med resultatet fra den enkle lineære regresjonsanalysen.

SLUTT