

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i STK1000 — Innføring i anvendt statistikk.

Eksamensdag: 2. desember 2008.

Tid for eksamen: 9.00–12.00.

Oppgavesettet er på 4 sider.

Vedlegg: Ingen.

Tillatte hjelpemidler: Læreboken Moore & McCabe: Introduction to the practice of statistics, ordliste for STK1000, godkjent kalkulator

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

### Oppgave 1

Siden eneggede tvillinger har samme genetiske materiale, er tvillingstudier spesielt egnet for å kartlegge hvordan miljøet virker inn på f.eks. personligheten. I en bok av den amerikanske forskeren Susan Faber finner vi data for  $n = 31$  tvillingpar der den ene tvillingen vokste opp hos biologiske foreldre (Twin A) og den andre vokste opp hos andre familiemedlemmer, foster- eller adoptiv-foreldre (Twin B). Nedenfor finner du en oppsummering av målt IQ for disse personene, samt resultatet av en statistisk analyse. Spørsmålet vi ønsker å belyse er om det er forskjell i IQ hos eneggede tvillinger der den ene tvillingen har vokst opp hos biologiske foreldre, og den andre ikke.

	N	Mean	StDev	SE Mean
Twin A	31	93.32	15.41	2.77
Twin B	31	96.58	13.84	2.49
Difference	31	-3.26	8.81	1.58

95% CI for mean difference: (-6.49, -0.03)

T-Test of mean difference = 0 (vs not = 0): T-Value = -2.06 P-Value = 0.048

a) Hvilken type statistisk prosedyre er benyttet? Begrunn hvorfor denne typen prosedyre er best egnet for problemstillingen.

b) Kall forventet forskjell mellom Twin A og Twin B for  $\mu$ . Sett opp nullhypotese og alternativ hypotese for testen som er utført. Skriv opp formelen for testobservatoren, og finn dennes numeriske verdi basert på utskriften.

(Fortsettes på side 2.)

c) Vis hvordan P-verdien er beregnet. Spesifiser antall frihetsgrader. Bruk signifikansnivå  $\alpha = 0.05$  og formuler din konklusjon på testen.

d) I utskriften finner du et 95% konfidensintervall for  $\mu$ . Hva betyr det at dette intervallet dekker kun negative verdier? Beregn i stedet et 99% konfidensintervall for  $\mu$ . Forklar hvorfor det ene intervallet blir lenger enn det andre.

## Oppgave 2

Et bestemt legemiddel er kjent for å irritere slimhinnen i magesekken, og man ønsker derfor å dekke middelet med et belegg som forsinker utskilling av virkestoffet til etter at tablettene har passert magesekken. Samtidig er det viktig at belegget ikke hindrer at virkemiddelet blir tatt opp i tarmen etterpå.

47 fastende frivillige ble randomisert i to grupper, der den ene gruppen ( $n_1 = 23$ ) fikk tablettene uten belegg og den andre ( $n_2 = 24$ ) fikk tablettene med belegg. Konsentrasjonen av virkemiddelet (i ng/ml) i blodet ble målt 12 timer etter administrasjon; en oppsummering av resultatene er gjengitt nedenfor (coating=belegg). Målingene kan antas som tilfeldige utvalg fra normalfordelte populasjoner, og de to utvalgene er uavhengige av hverandre. Vi ønsker å finne ut om det er signifikant lavere konsentrasjon av virkestoffet etter 12 timer for legemiddelet med belegg, sammenlignet med uten belegg.

Population	sample size	sample mean	sample st.dev.
Without coating	23	131.62	6.41
With coating	24	127.00	6.14

a) Svar på spørsmålet om belegget gir lavere konsentrasjon etter 12 timer (hindrer opptak i tarmen) ved hjelp av en hypotesetest: Velg passende notasjon, sett opp nullhypotese og alternativ og angi type test du vil bruke med begrunnelse.

b) Skriv ned formelen for egnet testobservator og beregn numerisk verdi fra data. Beregn (evt. tilnærmet) P-verdi. Beskriv nøye hvordan du er kommet frem til svaret. Velg et signifikansnivå og konkluder.

c) Hadde konklusjonen i b) vært troverdig selv om normalantakelsene skulle vise seg å ikke stemme? Begrunn svaret.

## Oppgave 3

Leonardo da Vinci (1452-1519) har tegnet en meget berømt tegning av en mann som står med armene rett ut fra kroppen inni et kvadrat (og en sirkel).

(Fortsettes på side 3.)

Tegningen viser at mannens armspenn er like langt som høyden på kroppen. For å sjekke om dette er riktige proposjoner, har vi målt 8 personers høyde og armspenn (i tommer). Resultatet av en enkel lineær regresjonsanalyse med høyde som forklaringsvariabel og armspenn som respons, finnes nedenfor.

a) Sett opp en enkel lineær regresjonsmodell for dette datasettet. Finn estimater for de tre parametrene i modellen fra utskriften. Finn et estimat for forventet armspenn for en person som er 66 tommer høy.

b) Dersom Leonardo da Vinci hadde rett, skulle vi forvente at stigningstallet  $\beta_1$  er omtrent lik 1. Beregn et 95% konfidensintervall for  $\beta_1$  og bruk det til å avgjøre om Leonardo tok feil.

c) Diskuter kort hvor godt modellen passer til data på bakgrunn av vedlagte materiale. Dersom vi i stedet hadde brukt armspenn som forklaringsvariabel og høyde som responsvariabel, hva hadde  $r^2$  blitt?

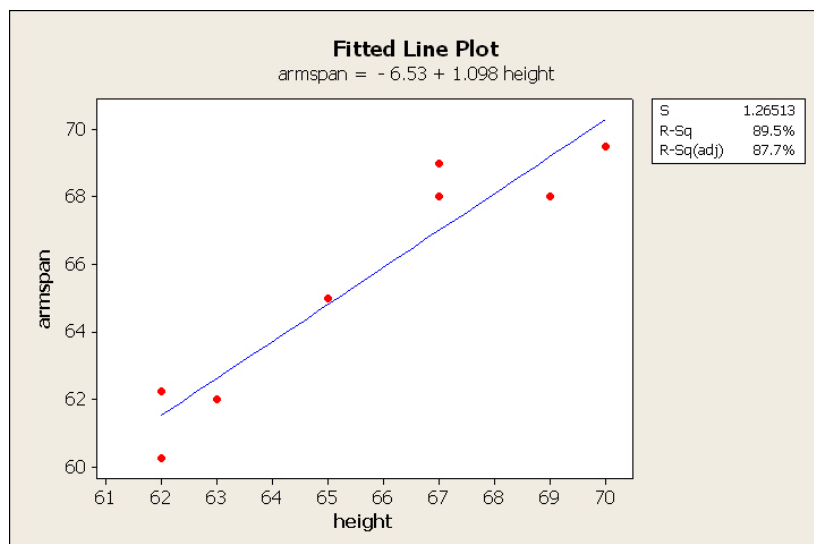
Regression Analysis: armspan versus height

The regression equation is

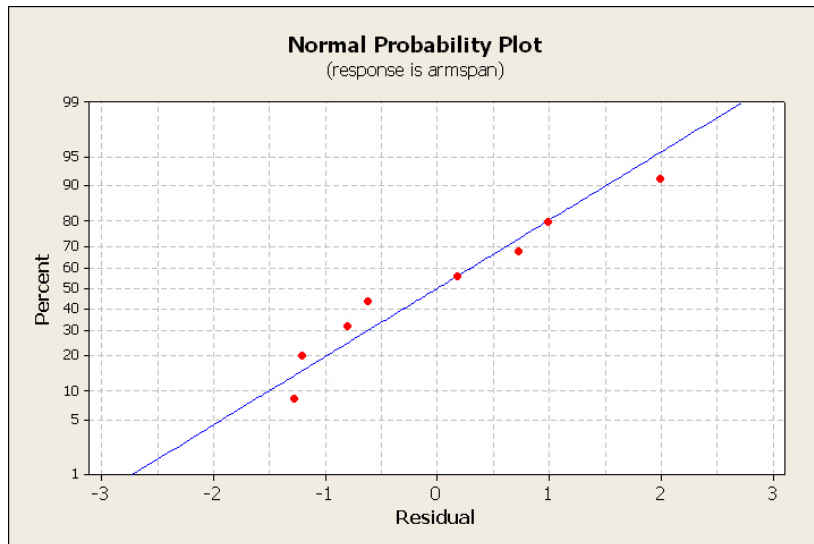
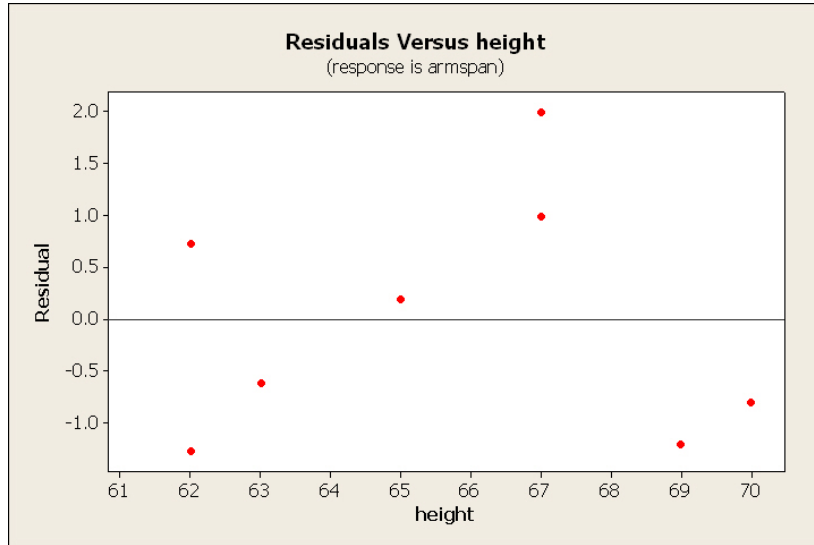
armspan = - 6.5 + 1.10 height

Predictor	Coef	SE Coef	T	P
Constant	-6.53	10.09	-0.65	0.541
height	1.0976	0.1536	7.15	0.000

S = 1.26513    R-Sq = 89.5%    R-Sq(adj) = 87.7%



(Fortsettes på side 4.)



SLUTT