

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Mandag 2. juni 2008.

Tid for eksamen: 09.00 – 12.00.

Oppgavesettet er på 4 sider.

Vedlegg: Tabell over t -fordeling.

Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics", 3., 4. el. 5. utgave. Ordliste for bruk i STK 1000, kalkulator.

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

Oppgave 1.

Aksjekjøp på Internett har økt dramatisk de siste årene. Vi har data fra 10 Internett aksjemeglerfirmaer. For hvert aksjemeglerfirma har en registrert kapitalbeholdningen (assets) i milliarder dollar, antall tusen Internett kontoer (accounts) samt markedsandel (market share) i prosent. Anta at kapitalbeholdningene er uavhengige og normalfordelte med ukjent forventning μ og ukjent standardavvik σ . Data er analysert i Minitab og nødvendige utskrifter til bruk i oppgaven er lagt ved sist i oppgaven.

- I den første Minitab-utskriften er det gitt et 95% konfidensintervall (intervallestimat) for μ . Vis hvordan en har kommet frem til det.
- Vi ønsker å teste:

$$H_0 : \mu = 100 \quad \text{mot} \quad H_a : \mu \neq 100$$

(Fortsettes side 2.)

for å finne ut om forventet kapitalbeholdning er forskjellig fra 100 milliarder dollar. Utfør en hypotesetest med signifikansnivå 5%. Hva blir konklusjonen?

- c) Vi ønsker å gjennomføre en regresjonsanalyse med kapitalbeholdning som responsvariabel og antall tusen Internett kontoer som forklaringsvariabel. Sett opp en enkel lineær regresjonsmodell for disse dataene. Spesifiser antakelsene som ligger til grunn. Angi estimater for alle de 3 parametrene i modellen.
- d) Finn et 95% prediksjonsintervall for kapitalbeholdningen til et Internett aksjemeglerfirma med 1 million kontoer. Hvor mange frihetsgrader har t -fordelingen som brukes i beregningene, og hvilken verdi t^* brukes i beregningen av intervallet?
- e) Kommenter med utgangspunkt i Minitab-utskriften hva som skjer hvis en bruker både markedsandel og antall tusen Internett kontoer som forklaringsvariabler.

One-Sample T: Assets

Variable	N	Mean	StDev	SE Mean	95% CI
Assets	10	48,9	76,2	24,1	(-5,6; 103,4)

Regression Analysis: Assets versus Accounts

The regression equation is
 Assets = - 17,1 + 0,0832 Accounts

Predictor	Coef	SE Coef	T	P
Constant	-17,121	8,778	-1,95	0,087
Accounts	0,083205	0,007592	10,96	0,000

S = 20,1877 R-Sq = 93,8% R-Sq(adj) = 93,0%

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	66,08	6,57	(50,93; 81,24)	(17,12; 115,04)

Values of Predictors for New Observations

New Obs	Accounts
1	1000

(Fortsettes side 3.)

Regression Analysis: Assets versus Marketshare; Accounts

The regression equation is

$$\text{Assets} = -21,5 + 1,16 \text{ Marketshare} + 0,0756 \text{ Accounts}$$

Predictor	Coef	SE Coef	T	P
Constant	-21,45	10,24	-2,09	0,074
Marketshare	1,158	1,344	0,86	0,418
Accounts	0,07559	0,01173	6,44	0,000

$$S = 20,5216 \quad R\text{-Sq} = 94,4\% \quad R\text{-Sq}(\text{adj}) = 92,7\%$$

Oppgave 2.

I en studie av jernmangel blant spebarn sammenlignet en data fra to grupper. Den første gruppen av spebarn fikk morsmelk, mens den andre gruppen av spebarn fikk standard babymat uten jerntilsetning. En målte så blodets hemoglobin nivå når barna fylte et år. En oppsummering av dataene er gitt i følgende tabell:

Morsmelk gruppen	Babymat gruppen
$n_1=23$	$n_2= 19$
$\bar{x}_1= 13.3$	$\bar{x}_2= 12.4$
$s_1= 1.7$	$s_2= 1.8$

Målingene av hemoglobin nivået for spebarn som fikk morsmelk antas å komme fra en normalfordeling med forventning μ_1 og standardavvik σ . Tilsvarende kan målingene av hemoglobin nivået for spebarn som fikk standard babymat antas å komme fra en normalfordeling med forventning μ_2 og standardavvik σ . Målingene fra de to gruppene antas uavhengige av hverandre.

- Sett opp en estimator for standardavviket σ . Beregn et estimat for σ fra data i tabellen.
- Forklar hvorfor vi velger å teste hypotesene:

$$H_0 : \mu_1 = \mu_2 \quad \text{mot} \quad H_a : \mu_1 > \mu_2$$

(Fortsettes side 4.)

for å finne ut om morsmelk har en positiv effekt på hemoglobin nivået i blodet hos spebarn. Sett opp en passende testobservator for hypotesene og angi hvilken fordeling testobservatoren har.

- c) Beregn verdien av testobservatoren fra dataene i tabellen og anslå tilhørende P -verdi. Hvilken konklusjon trekker du?
- d) Hadde konklusjonen ovenfor vært troverdig selv om normalantagelsene skulle vise seg å ikke stemme? Begrunn svaret.
- e) Finn et 90% konfidensintervall (intervallestimat) for forskjellen mellom morsmelk gruppen og babymat gruppen i forventet hemoglobin nivå i blodet.

SLUTT