

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Tirsdag 4. desember 2007.

Tid for eksamen: 09.00 – 12.00.

Oppgavesettet er på 4 sider.

Vedlegg: Tabell over normalfordeling, 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.

Samferdselsetaten i Oslo kommune ønsker å vurdere ordningen med gebyr for å kjøre med piggdekk innenfor byens grenser. Spesielt er de interessert i om det er forskjell i sannsynligheten for at en bil er involvert i minst en ulykke i løpet av en vintersesong ved bruk av piggdekk sammenlignet med andre vinterdekk.

- a) Samferdselsetaten velger å trekke ut 1000 biler fra registret over biler med betalt piggdekkoblat og 1000 biler fra øvrige biler som er utstyrt med andre vinterdekk. Er dette et fornuftig forsøksopplegg? Drøft spesielt betydningen av "skjulte" variable (lurking variables).

(Fortsettes side 2.)

La tilfeldige variabler  $X$  og  $Y$  betegne henholdsvis antall av de 1000 bilene med piggdekk og de 1000 bilene med andre vinterdekk som er involvert i minst en ulykke i løpet av en vintersesong. Samferdselsetaten antar at  $X$  og  $Y$  er binomisk fordelt henholdsvis  $(1000, p)$  og  $(1000, q)$ . Her er  $p$  og  $q$  sannsynlighetene for at en bil er involvert i minst en ulykke i løpet av en vintersesong med henholdsvis piggdekk og andre vinterdekk.

- b) Hva syns du om disse antagelsene? Begrunn svaret.
- c) Anta  $p = 0.01$ . Samferdselsetaten ønsker å beregne en tilnærmet sannsynlighet for at  $X$  er mindre enn eller lik 5. Utfør denne beregningen og begrunn den tilnærmelsen som blir brukt.
- d) Foreslå en forventningsrett estimator for  $p - q$  og beregn standardavviket til denne estimatoren. Du kan anta at  $X$  og  $Y$  er uavhengige. Hva blir estimatet for  $p - q$  hvis en observerer  $X = x = 12$  og  $Y = y = 20$ ? Kommenter resultatet.

## Oppgave 2.

I en innsjø i Finland er det fanget 12 fisk av slaget abbor. For hver fisk har en målt vekt i gram samt lengde og bredde i cm. Anta at vektene er uavhengige og normalfordelte med ukjent forventning  $\mu$  og ukjent standardavvik  $\sigma$ . Data er analysert i **Minitab** og nødvendige utskrifter til bruk i oppgaven er lagt ved sist i oppgaven.

- a) I den første **Minitab**-utskriften er det gitt et 95% konfidensintervall (intervallestimat) for  $\mu$ . Vis hvordan en har kommet frem til det.
- b) Vi ønsker å teste

$$H_0 : \mu = 200 \quad \text{mot} \quad H_a : \mu > 200$$

for å finne ut om forventet vekt av abboren er større enn 200 gram. Utfør en hypotesetest med signifikansnivå 5%. Hva blir konklusjonen?

- c) Vi ønsker å gjennomføre en regresjonsanalyse med vekt som responsvariabel og lengde 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 vekten til en nyfisket abbor med lengde 40 cm. Hvor mange frihetsgrader har  $t$ -fordelingen som brukes i beregningene, og hvilken verdi  $t^*$  brukes i beregningen av intervallet?

(Fortsettes side 3.)

- e) Kommenter med utgangspunkt i Minitab-utskriften hva som skjer hvis en bruker bredde som forklaringsvariabel i stedet for lengde.
- f) Kommenter med utgangspunkt i Minitab-utskriften hva som skjer hvis en bruker både lengde og bredde som forklaringsvariabler.

## One-Sample T: Weight

Variable	N	Mean	StDev	SE Mean	95% CI
Weight	12	365,5	333,2	96,2	(153,8; 577,2)

## One-Sample T: Weight

Test of mu = 200 vs &gt; 200

Variable	N	Mean	StDev	SE Mean	T	P
Weight	12	365,5	333,2	96,2	1,72	0,057

## Regression Analysis: Weight versus Length

The regression equation is  
 Weight = - 469 + 28,5 Length

Predictor	Coef	SE Coef	T	P
Constant	-468,91	92,55	-5,07	0,000
Length	28,462	2,967	9,59	0,000

S = 109,416    R-Sq = 90,2%    R-Sq(adj) = 89,2%

## Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	669,6	44,8	(569,8; 769,3)	(406,2; 933,0)

## Values of Predictors for New Observations

New Obs	Length
1	40,0

(Fortsettes side 4.)

Regression Analysis: Weight versus Width

The regression equation is  
Weight = - 449 + 175 Width

Predictor	Coef	SE Coef	T	P
Constant	-449,44	89,27	-5,03	0,001
Width	174,63	17,93	9,74	0,000

S = 107,904    R-Sq = 90,5%    R-Sq(adj) = 89,5%

Regression Analysis: Weight versus Length; Width

The regression equation is  
Weight = - 467 + 13,3 Length + 94,6 Width

Predictor	Coef	SE Coef	T	P
Constant	-466,78	92,18	-5,06	0,001
Length	13,33	14,82	0,90	0,392
Width	94,61	90,82	1,04	0,325

S = 108,953    R-Sq = 91,3%    R-Sq(adj) = 89,3%

SLUTT