

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1000 — Innføring i anvendt statistikk
Eksamensdag: Onsdag 2. desember 2015
Tid for eksamen: 09.00 – 13.00
Oppgavesettet er på 3 sider.
Vedlegg: Ingen
Tillatte hjelpemidler: Godkjent kalkulator, ordliste for STK1000 og lærebok (alle utgaver og det er lov til å notere i læreboken).

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

Oppgave 1

I perioden 2005–2014 var ca. 2% av fødsleene i Norge tvillingfødsler. La X være antall tvillingfødsler av 1000 tilfeldig valgte fødsler fra denne perioden.

1a

Hva er forventning og standardavvik til X ?

1b

Bruk tilnærming til normalfordelingen og beregn sannsynligheten for at ikke mer enn 11 av 1000 tilfeldig valgte fødsler fra denne perioden var tvillingfødsler.

Er normalfordelingen en god tilnærming i dette tilfellet? Begrunn svaret.

1c

Vi tenker oss at vi for hvert av disse 10 årene registrerer utfallet av 1000 tilfeldig valgte fødsler. Vi antar at det er uavhengighet mellom årene. Hva er sannsynligheten for at det i alle av disse ti årene er fler enn 11 tvillingfødsler blant de 1000 fødsleene?

Forklar hvordan du kom fram til svaret.

Oppgave 2

Værmeldinger er usikre og temperaturen som meldes dagen i forveien er ikke alltid den som faktisk observeres når dagen kommer. Vi skal i denne oppgaven se på differensen D mellom målt temperatur og den temperaturen

(Fortsettes på side 2.)

værmeldingen predikerer dagen før. Nedenfor ser du et datasett med slike differenser registrert 10 tilfeldige dager, etterfulgt av en utskrift fra en analyse av dataene gjort i Minitab. Du skal anta at observasjonene er uavhengige og kommer fra en normalfordelt populasjon med forventning μ og standardavvik σ .

Differenser: -1, 1, 2, 0, 0, 2, 4, -3, 1, -1

One-Sample T: Differense

Test of $\mu = 0$ vs $\neq 0$

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Differense	10	0,500	1,958	0,619	(-0,901; 1,901)	?	0,440

2a

Foreslå en forventningsrett estimator for μ .

Angi estimatorens standardavvik.

Hvilken fordeling har estimatoren?

2b

Sett opp nullhypotese og alternativ hypotese for å teste om det er grunnlag i dataene for å påstå at du ikke kan stole på værmeldingen.

Sett opp formelen for tilhørende teststatistikk, som er utelatt fra utskriften, og beregn denne.

Benytt utskriften fra Minitab over til å formulere en konklusjon på testen.

2c

Minitabutskriften inneholder også et 95% konfidensintervall. Vis hvordan dette intervallet er regnet ut.

Gi en fortolkning av intervallet og sammenhold denne med konklusjonen av testen i forrige punkt.

Oppgave 3

Lungefunksjon kan måles ved FEV_1 (Forced Expiratory Flow in one second), som er det volumet luft målt i liter som en person kan puste ut i løpet av et sekund. I denne oppgaven skal vi se på data om FEV_1 for 109 barn i alderen 3–5 år. Spesielt skal vi se på hvordan denne størrelsen avhenger av barnets vekt i kg og barnets høyde målt i meter (med cm som desimal).

3a

Under ser du utskrift fra en lineær regresjon med FEV_1 som responsvariabel og vekt som forklaringsvariabel. Finn på bakgrunn av utskriften korrelasjonen mellom FEV_1 og vekt.

(Fortsettes på side 3.)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,237162	31,75%	31,12%	25,97%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	0,177	0,119	1,49	0,140
vekt	0,04899	0,00694	7,06	0,000

3b

Sett opp modellen som regresjonsanalysen er basert på.

Formuler en nullhypotese om at FEV_1 ikke avhenger av vekt samt en alternativ hypotese om at det er slik en sammenheng ved hjelp av denne modellspesifikasjonen.

Formuler en konklusjon på denne hypotesetesten basert på utskriften.

3c

Hva er estimert forskjell i FEV_1 mellom et barn som veier 19 kg og et barn som veier 18 kg?

Beregn et 95% konfidensintervall for den forventede forskjellen i FEV_1 .

3d

I dette punktet har vi utvidet analysen til en multippel lineær regresjon der også høyde er med som forklaringsvariabel. En utskrift fra denne analysen er gitt under. Korrelasjonskoeffisienten mellom vekt og høyde er også satt opp. Test på ny om det er sammenheng mellom vekt og FEV_1 .

Gi en forklaring på forskjellene mellom konklusjonen i dette og i det forrige punktet.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-2,022	0,298	-6,78	0,000
høyde	0,02630	0,00338	7,79	0,000
vekt	0,01318	0,00722	1,83	0,071

Correlation: vekt; høyde

Pearson correlation of vekt and høyde = 0,637

SLUTT