UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1110 — Statistiske metoder og dataanalyse 1.
Eksamensdag:	Onsdag 5. desember 2007.
Tid for eksamen:	9.00 – 12.00.

Oppgavesettet er på 4 sider.

Vedlegg:	Tabell for t-fordelingen.
Tillatte hjelpemidler:	Godkjent lommeregner og Formelsamling for STK1100 og STK1110.

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

Oppgave 1.

I USA er Thanksgiving Day, som feires den fjerde torsdagen i november, en nasjonal høytidsdag. På denne dagen spises det tradisjonelt kalkun som festmat. Tabell 1 gir vekten (i pund) for tilfeldig valgte Thanksgiving-kalkuner fra to delstater i USA.

Tabell 1: Slaktevekt for kalkuner fra to delstater i USA.

Delstat	Virginia	Wisconsin
Vekt	13.1	11.5
(i pund)	12.4	14.2
	13.2	15.4
	11.8	13.1
		13.8
Gjennomsnitt	12.625	13.600
Empirisk standardavvik	0.655	1.440

Vi lar μ_1 være forventet vekt for en tilfeldig valgt kalkun fra Virginia, mens μ_2 er forventet vekt for en tilfeldig valgt kalkun fra Wisconsin.

(Fortsettes side 2.)

a) Bestem et 90% konfidensintervall for $\mu_2 - \mu_1$ og diskuter hva intervallet sier deg. Beskriv hvilke forutsetninger konfidensintervallet bygger på.

b) Angi en test med signifikansnivå 5% for nullhypotesen $H_0 : \mu_1 = \mu_2$ mot den alternative hypotesen $H_1 : \mu_1 \neq \mu_2$. Hva blir konklusjonen av testen?

c) Forklar hva vi mener med P-verdien til en test. Bestem (så godt du kan) P-verdien for testen i punkt b.

Oppgave 2.

Thanksgiving-kalkunene i oppgave 1 var ikke like gamle da de ble slaktet. Tabell 2 viser kalkunenes alder (i uker) ved slakting. For oversiktens skyld gir vi også slaktevektene i tabellen.

Tabell 2: Slaktevekt og slaktealder for kalkuner.

Virginia		Wisconsin	
Alder	Vekt	Alder	Vekt
29	13.1	21	11.5
27	12.4	27	14.2
28	13.2	29	15.4
26	11.8	23	13.1
		25	13.8

a) Plott vekten av kalkunene mot alderen ved slakting. Utfør plottingen i ett diagram, men benytt forskjellige symboler for de to delstatene. Kommenter plottet.

For å studere hvilken betydning delstat og alder har for slaktevekten til kalkunene, setter vi opp regresjonsmodellen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i. \tag{1}$$

I modellen (1) er:

- Y_i slaktevekten til kalkun nummer i (fra begge delstatene sett under ett),
- x_{i1} alderen ved slakting til kalkun nummer i,
- $x_{i2} = 1$ hvis kalkun nummer i er fra Wisconsin, $x_{i2} = 0$ hvis kalkun nummer i er fra Virginia.

(Fortsettes side 3.)

b) Gi en fortolkning av parametrene β_1 og β_2 i regresjonsmodellen (1), og beskriv hvilke forutsetninger regresjonsmodellen bygger på. Syns du modellformuleringen (1) virker rimelig i lys av plottet i punkt a?

Ut fra tallene i tabell 2 finner vi at minste kvadraters estimater er $\hat{\beta}_0 = 0.311$, $\hat{\beta}_1 = 0.448$ og $\hat{\beta}_2 = 2.094$, og at estimert standardfeil til $\hat{\beta}_2$ er $S_{\hat{\beta}_2} = 0.235$. (Du skal *ikke* vise dette.)

c) Angi en test med signifikansnivå 5% for nullhypotesen $H_0 : \beta_2 = 0$ mot den alternative hypotesen $H_1 : \beta_2 \neq 0$. Hva blir konklusjonen av testen? Kommenter resultatet i lys av testen i punkt b i oppgave 1.

Ut fra tallene i tabell 2 har vi at residualkvadratsummen er

$$RSS = \sum_{i=1}^{9} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right)^2 = 0.5648.$$

Vi innfører 9×3 matrisen

$$\mathbf{X} = \begin{pmatrix} 1 & 29 & 0 \\ 1 & 27 & 0 \\ 1 & 28 & 0 \\ 1 & 26 & 0 \\ 1 & 21 & 1 \\ 1 & 27 & 1 \\ 1 & 29 & 1 \\ 1 & 23 & 1 \\ 1 & 25 & 1 \end{pmatrix}$$

Da er

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 17.06 & -0.611 & -1.778 \\ -0.611 & 0.0222 & 0.0556 \\ -1.778 & 0.0556 & 0.589 \end{pmatrix}$$

d) Finn den estimerte standardfeilen $S_{\hat{\beta}_1}$ til $\hat{\beta}_1$ og bestem et 95 % konfidensintervall for β_1. Forklar hva intervallet sier deg.

Oppgave 3.

En bedrift produserer lysstoffrør. Levetiden (i timer) av et tilfeldig valgt lysstoffrør antas å være en stokastisk variabel med sannsynlighetstetthet

$$f(y \mid \theta) = \begin{cases} \frac{y}{\theta}\, e^{-y^2/2\theta} & \text{for } y > 0 \\ 0 & \text{ellers} \end{cases} \tag{2}$$

der $\theta > 0$.

(Fortsettes side 4.)

a) La $F(y \,|\, \theta)$ være den kumulative fordelingsfunksjonen som svarer til sannsynlighetstettheten (2). Medianen $y_{0.50}$ er gitt ved at $F(y_{0.50} \,|\, \theta) = 0.50$. Vis at $y_{0.50} = \sqrt{2\theta \log 2}$. (Her står "log" for den naturlige logaritmen.)

Bedriften ønsker å studere levetiden til lysstoffrørene. For å gjøre det, observerer de levetidene $Y_1, Y_2, \ldots, Y_{20}$ for 20 tilfeldig valgte lysstoffrør. Vi antar at $Y_1, Y_2, \ldots, Y_{20}$ er uavhengige og identisk fordelte med sannsynlighetstetthet gitt ved (2).

b) Vis at maksimum likelihood estimatoren $\hat{\theta}$ er gitt ved $\hat{\theta} = \frac{1}{40} \sum_{i=1}^{20} Y_i^2$.

c) Bedriften ønsker å teste nullhypotesen at median levetid er høyst 1000 timer mot alternativet at den er mer enn 1000 timer. Vis at det er det samme som å teste

$$H_0 : \theta \leq \theta_0 \quad \text{mot} \quad H_1 : \theta > \theta_0$$

der $\theta_0 = 10^6/(2 \log 2)$.

d) En rimelig test forkaster H_0 hvis $\hat{\theta} > k$. Bestem k slik at testen får signifikansnivå 5%. (*Vink:* En kan vise at $\frac{1}{\theta} \sum_{i=1}^{20} Y_i^2$ er kji-kvadratfordelt med 40 frihetsgrader. Du kan bruke dette resultatet uten å vise det. Utvalgte fraktiler i kji-kvadratfordelingen med 40 frihetsgrader er gitt i tabell 3 nedenfor.)

e) Bestem (så godt du kan) styrken for testen i punkt d hvis median levetid er 1250 timer. Hva blir da sannsynligheten for feil av type II?

Tabell 3: Fraktiler for kji-kvadratfordelingen med 40 frihetsgrader.

p	0.025	0.05	0.10	0.20	0.30	0.40	0.50
χ^2_p	24.43	26.51	29.05	32.34	34.87	37.13	39.33
p	0.60	0.70	0.80	0.90	0.95	0.975	
χ^2_p	41.62	44.16	47.27	51.81	55.76	59.34	

SLUTT