TMA4245 Statistikk
Eksamen 10. august 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

En blodgiver ser at hans hemoglobinverdier (X, i g/dl) i perioden 1993 til 2012 tilsynelatende var uavhengige og normalfordelt med forventning 16.14 og standardavvik 0.63.

a) Idrettsutøvere får ikke lov å delta i en konkurranse hvis hemoglobinnivået er over 17.5 (blodet blir for tykt). Hva er sannsynligheten for at blodgiveren ville blitt nektet deltakelse hvis han skulle finne på å melde seg på en konkurranse?

Før større konkurranser, som f.eks. Tour de France på sykkel, blir det tatt hyppige blodprøver av idrettsutøverne for å forhindre doping. Hva er sannsynligheten for at vår blodgiver ville fått minst en av fem målinger over grensa på 17.5? Beskriv kort hvilke forutsetninger som må være oppfylt for at dere skal kunne regne ut denne sannsynligheten.

Utover 2000-tallet fant Blodbanken at den gamle målemetoden hadde for stor usikkerhet, med varians på 0.63^2. En ny målemetode ble derfor introdusert fra og med 2007. Femten verdier til blodgiveren målt med den nye metoden ga $\sum_{i=1}^{15}(x_i - \bar{x})^2 = 2.2$.

b) Formuler en hypotesetest som undersøker om det er grunnlag for å påstå at den nye metoden har lavere varians enn den gamle.

Hvilke antagelser må aksepteres for at testen skal kunne gjennomføres?

Hva blir konklusjonen på testen når dataene er som over, og signifikansnivået er 0.05?

Oppgave 2

Et problem med vindmøller til kraftproduksjon er at fugler kan kollidere med rotorbladene og dø. Som et prøveprosjekt monteres ei mølle (A) på et sted langs norskekysten, og det registreres kontinuerlig hvor mange fugler som blir funnet døde av kollisjonsskader rundt mølla. Erfaring fra Danmark med lignende møller tilsier at forventet antall døde fugler per uke er $\lambda = 1$. Anta at den tilfeldige variabelen Y, antall fugler som kolliderer og dør med mølle A per uke, er poissonfordelt.

a) Hvis vi antar like forhold i Norge som i Danmark og at etterfølgende uker er uavhengige, finn sannsynligheten for at det i løpet av de første 5 uker skal kollidere mer enn 10 fugler med vindmølle A.

Gitt at det kolliderer mindre enn 5 fugler, hva er sannsynligheten for at ingen fugler kolliderte?

b) I løpet av fire år blir det funnet 261 fugler rundt den norske mølla. Estimer parameteren λ med sannsynlighetsmaksimering (maximum likelihood) fra disse data.

Ei anna mølle (B) ble plassert litt lenger inn i landet, på en plass med mindre fugletetthet. Anta at X, antall fugler som kolliderer og dør med mølle B per uke, er poissonfordelt med parameter ν og uavhengig av Y. La så $Z = X + Y$ være summen av døde fugler ved de to møllene

c) Finn den momentgenererende funksjonen $M_Z(t)$ til den tilfeldige variabelen Z.

Bruk den momentgenererende funskjonen til å si hvilken fordeling Z har.

Oppgave 3

Teodor jobber i iskiosken ved campingen denne sommeren. Han undrer seg over fordelingen til inntekten i kiosken, og hvordan den varierer med temperatur. La x_i være temperaturen kl 14 dag i. Vi ser på denne temperaturen og inntekten Y_i for $i = 1, \ldots, n$ ulike dager.

Anta en regresjonsmodell $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, for $i = 1, \ldots, n$, der $\epsilon_1, \ldots, \epsilon_n$ er uavhengige normalfordelte støyledd med forventning 0 og varians τ^2.

Det kan vises at en forventningsrett estimator for β_1 er gitt ved $\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

a) Vis at variansen til $\hat{\beta}_1$ er $\frac{\tau^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

La videre $\hat{\beta}_0$ være en estimator for β_0. Forklar kort hvorfor $s^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$ er en nyttig estimator for τ^2. Hvorfor deler vi her på $n - 2$?

Inntekten kan splittes i salg av fløteiskrem og saftis. Vi definerer disse henholdsvis y_i^f og y_i^s. Teodor tenker at inntekten kan variere ulikt som en funksjon av temperatur og foreslår følgende modell:

$$Y_i^f = \beta_0^f + \beta_1^f x_i + \epsilon_i^f, \quad Y_i^s = \beta_0^s + \beta_1^s x_i + \epsilon_i^s \quad i = 1, \ldots, n.$$

Her er $\epsilon_1^f, \ldots, \epsilon_n^f, \epsilon_1^s, \ldots, \epsilon_n^s$ uavhengige normaldelte støyledd med forventning 0 og varians σ^2.

b) Utled et 90% konfidensintervall for differansen i stigningstall $\beta_1^f - \beta_1^s$.

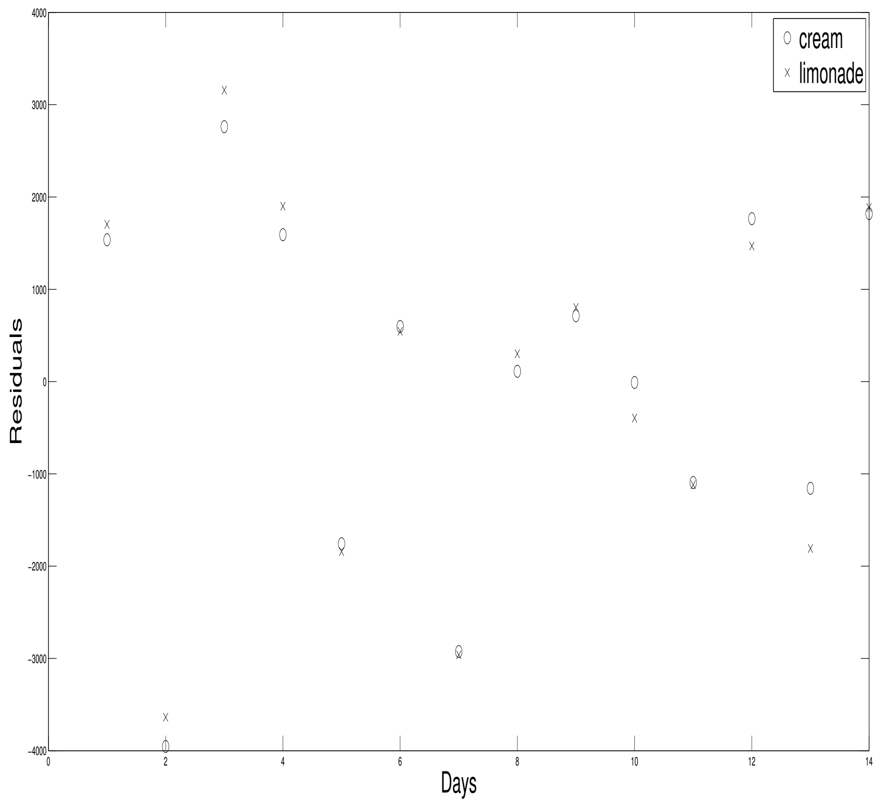
Regn ut intervallet når Teodor har $n = 14$ dager med data. Det oppgis her at $\hat{\beta}_1^f = 1720$ og $\hat{\beta}_1^s = 1442$. Videre er $\sum_{i=1}^n (x_i - \bar{x})^2 = 260.4, \sum_{i=1}^{14} (Y_i^f - \hat{\beta}_0^f - \hat{\beta}_1^f x_i)^2 = 7046^2$ og $\sum_{i=1}^{14} (Y_i^s - \hat{\beta}_0^s - \hat{\beta}_1^s x_i)^2 = 7300^2$.

Merk: Totalt er det 4 parametre som estimeres i regresjonslinjene.

Figur 1 viser residualene etter tilpasning av regressjonslinjer for salg av fløteis og saftis de 14 dagene.

c) Bruk plottet i figur 1 til å drøfte antakelsene gjort om feilleddene i modellen i punkt b).

Oppgave 4



Figur 1: Estimerte residualer etter tilpasning av regresjonslinjer for salg av fløteis og saftis de 14 dagene.

I en enkel versjon av spillet Craps kaster en spiller i hver runde to regulære terninger og registrerer summen av de to, et tall mellom 2 og 12. Det er to mulige måter for en spiller å vinne på: i) *Direkte seier*: Oppnå 7 eller 11 i første runde. ii) *Forsinket seier*: Oppnå 4, 5, 6, 8, 9 eller 10 i første runde og deretter få det samme tallet i en senere runde, men før man får 7.

Regn ut sannsynligheten for å vinne en *direkte seier*.

La A_i = vinne en *forsinket seier* ved å oppnå sum lik i, der $i = 4, 5, 6, 8, 9, 10$. Regn ut $P(A_4)$.

Finn til slutt sannsynligheten for å vinne i Craps. Hint: $\sum_{k=0}^{\infty}(\frac{q}{36})^k = \frac{1}{1-q/36}$, for $q \in \{1, \dots, 35\}$.

Fasit

1. **a**) 0.154,0.075 **b**) Forkaster H_0

2. **a**) 0.014,0.015 **b**) 1.25

3. **b**) $[-32, 588]$

4. 8/36,1/36,0.493