TMA4245 Statistikk
Eksamen 10. august 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Ein blodgivar ser at hans hemoglobinverdier (X, i g/dl) i perioden 1993 til 2012 tilsynelatande var uavhengige og normalfordelte med forventning 16.14 og standardavvik 0.63.

a) Idrettsutøvarar får ikkje lov å delta i ein konkurranse om hemoglobinnivået er over 17.5 (blodet blir for tjukt). Kva er sannsynet for at blodgivaren ville blitt nekta å delta dersom han skulle finne på å melde seg på ein konkurranse?

Før større konkurranser, som f.eks. Tour de France på sykkel, vert det teke hyppige blodprøver av idrettsutøvarane for å hindre doping. Kva er sannsynet for at vår blodgivar ville fått minst ei av fem målinger over grensa på 17.5? Beskriv kort kva forutsetninger som må vere oppfylt for at du skal kunne rekne ut dette sannsynet.

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

b) Formuler ein hypotesetest som undersøkjer om det er grunnlag for å påstå at den nye metoden har lågere varians enn den gamle.

Kva antagelser må til for at testen skal kunne gjennomførast?

Kva vert konklusjonen på testen når data er som over, og signifikansnivået er 0.05?

Oppgave 2

Eit problem med vindmøller til kraftproduksjon er at fuglar kan kollidere med rotorblada og døy. Som eit prøveprosjekt monterast ei mølle (A) på ein stad langs noregskysten, og det vert tella kor mange fuglar som vart funne døde av kollisjonsskader rundt mølla. Erfaringer frå Danmark med liknande møller tilsei at forventa tal på døde fuglar per veke er $\lambda = 1$. Anta at den tilfeldige variabelen Y, talet på fuglar som kolliderer of døyr med mølle A per veke, er poissonfordelt med denne parameteren.

a) Dersom vi antar like forhold i Noreg som i Danmark, og at etterfølgjande veker er uavhengige, finn sannsynet for at det i løpet av dei 5 første vekene skal kollidere meir enn 10 fuglar med vindmølle A.

Gitt at det kolliderer mindre enn 5 fuglar, kva er sannsynet for at ingen fuglar kolliderte?

b) I løpet av fire år vert det funne 261 fuglar rundt den norske mølla. Estimer parameteren λ med sannsynlighetsmaksimering (maximum likelihood) frå desse data.

Ei anna vindmølle (B) vart sett lenger inn i landet, på ein stad med mindre fugletettleik. Anta at X, talet på fuglar som kolliderer og døyr med mølle B per veke, er poissonfordelt med parameter ν og uavhengig av Y. La $Z = X + Y$ være summen av døde fuglar ved dei to møllene.

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

Bruk den momentgenererande funksjonen til å seie kva fordeling Z har.

Oppgave 3

Teodor arbeider i iskiosken ved campingen. Han undrar seg over fordelinga til inntekta i kiosken, og korleis den samvarierer med temperaturen. La x_i vere temperaturen kl 14 dag i. Vi ser på denne temperaturen og inntekta Y_i for $i = 1, \ldots, n$ ulike dagar.

Anta ein 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.

Du kan bruke at ein forventningsrett estimator for β_1 er $\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

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

La vidare $\hat{\beta}_0$ vere ein estimator for β_0. Forklar kort kvifor $s^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$ er ein nyttig estimator for τ^2. Kvifor deler vi på $n - 2$?

Inntekta kan splittast i sal av fløyteiskrem og saftis. Vi definerer desse henholdsvis y_i^f og y_i^s. Teodor tenker at inntekta kan variere ulikt som en funksjon av temperatur og foreslår følgjande 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) Utlei eit 90% konfidensintervall for differansen i stigningstall $\beta_1^f - \beta_1^s$.

Rekn ut intervallet når Teodor har $n = 14$ dagar med data. Her er $\hat{\beta}_1^f = 1720$ og $\hat{\beta}_1^s = 1442$. Vidare 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 $\sum_{i=1}^{14} (Y_i^s - \hat{\beta}_0^s - \hat{\beta}_1^s x_i)^2 = 7300^2$.

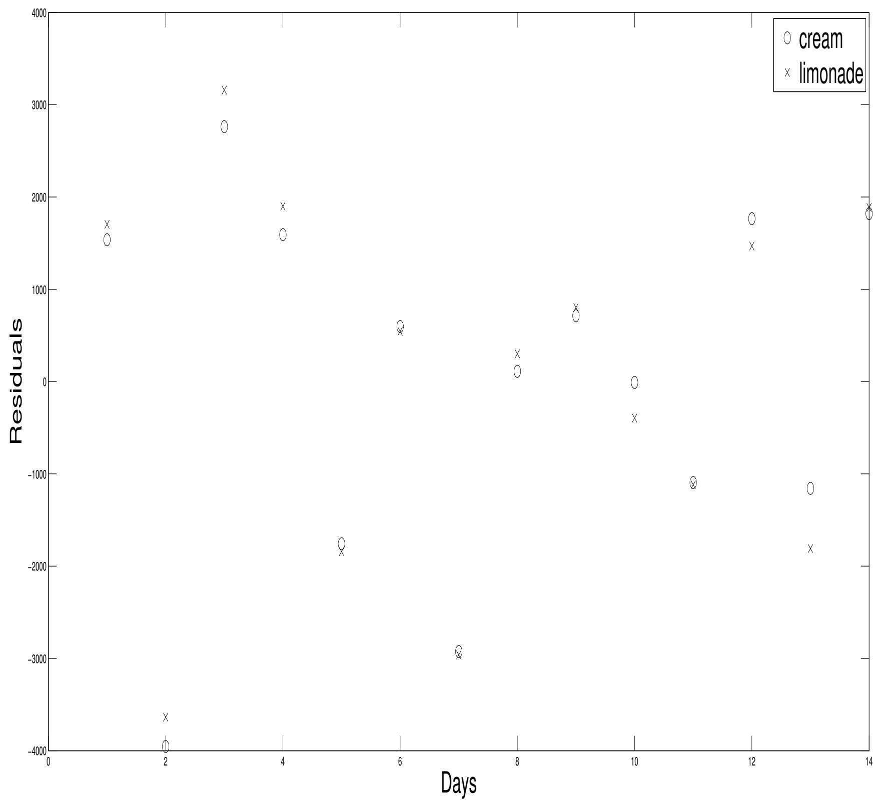
Merk: Totalt er det 4 parametre som estimerast i regresjonslinjene.

Figur 1 viser residuala etter tilpassing av regressjonslinjene for sal av fløyteis og saftis dei 14 dagane.

c) Bruk plottet i figur 1 til å drøfte vurderingane om feilledd i modellen i punkt b).

Oppgave 4

I en enkel versjon av spelet Craps kaster ein spelar kvar runde to regulære terningar og registrerer summen av dei to, eit tal mellom 2 og 12. Det er to moglege måtar for ein spelar å



Figur 1: Estimerte residual etter tilpassing av regresjonslinjer for sal av fløyteis og saftis dei 14 dagane.

vinne på: i) *Direkte siger*: Oppnå 7 eller 11 i første runde. ii) *Forsinka siger*: Oppnå 4, 5, 6, 8, 9 eller 10 i første runde og deretter få same tal i ei seinare runde, men før ein får 7.

Rekn ut sannsynet for å vinne ein *direkte siger*.

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

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

Fasit

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

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

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

4. 8/36,1/36,0.493