**Norges teknisk-naturvitenskapelige universitet**
**Institutt for matematiske fag**



Nynorsk

Fagleg kontakt under eksamen:

| | |
|---|---|
| Ingelin Steinsland | 926 63 096 |
| Øyvind Bakke | 990 41 673 |
| Ola Diserud | 932 18 823 |


# EKSAMEN I TMA4245 STATISTIKK

3. juni 2011
Tid: 09:00–13:00

Hjelpemiddel: *Tabeller og formler i statistikk*, Tapir Forlag
K. Rottmann: *Matematisk formelsamling*
Kalkulator HP30S / CITIZEN SR-270X
Gult, stempla A5-ark med eigne handskrivne notat.

Sensuren fell: 24. juni 2011

## Oppgave 1 Bustad på Hawaii

Frå 1832 til 1950 hadde vulkanen Mauna Loa på Hawaii 37 utbrot, altså med gjennomsnittleg rate (intensitet) på 0.026 utbrot pr. månad. Over korte geologiske periodar kan ein gå ut frå at utbrota til ein vulkan er ein Poisson-prosess i tid. I samband med studium ved University of Hawaii har du fått eit svært gunstig tilbod på ein fem års leigekontrakt for eit hus ved foten av Mauna Loa.

**a)** Beskriv kort eigenskapane som må vere oppfylte for at utbrota til ein vulkan skal kunne reknast som ein Poisson-prosess i tid.

Kva blir sannsynet for at det i løpet av leigeperioden på fem år skal kome minst eitt vulkanutbrot? Bruk raten (intensiteten) som frå 1832 til 1950.

Du får vete at Mauna Loa har hatt eit utbrot seks månader før overtakingsdato. Kva er sannsynet for at det skal gå meir enn tre år frå overtakingsdato til neste utbrot?

**Oppgave 2** Bustadmarknaden i Trondheim

Petter har nettopp fått seg jobb i Trondheim og ser etter bustad. Ettersom han jobbar i Midtbyen, ser han etter bustad der. For eit utfallsrom som består av alle bustader som er til sals i Trondheim på `finn.no` trekkjer Petter ein bustad tilfeldig (alle bustadene har same sannsyn for å bli trekte). Vi definerer desse hendingane:

$$M: \text{Bustaden ligg i Midtbyen}$$
$$T: \text{Bustaden har rettleiande pris på under 2 mill. kr}$$

**a)** Teikn hendingane i eit Venn-diagram.

Det er 381 bustader til sals i Trondheim. Av desse er 94 i Midtbyen, og 190 har rettleiande pris på under 2 mill. Av bustadene i Midtbyen har 50 rettleiande pris under 2 mill.

Er hendingane $M$ og $T$ disjunkte?
Er hendingane $M$ og $T$ uavhengige?

Grunngje og kommenter svara.

Vi antar at kvar bustad blir seld etter ein bodrunde til høgast bydande og at salsprisen per kvadradmeter for bustader i Midtbyen kan modellerast med regresjonsmodellen

$$Y = \beta_1 x + \epsilon(x),$$

der $Y$ er salspris (kr per m$^2$), $x$ er den rettleiande prisen (kr per m$^2$) og $\epsilon(x)$ er normalfordelt (gaussisk fordelt) med forventningsverdi $E(\epsilon) = 0$ og varians $\text{Var}(\epsilon) = \tau^2 x^2$, der $\tau = 0.1$ er ein kjend parameter.

**b)** Anta i dette punktet at $\beta_1 = 1.1$.

Forklar betydninga av at $\beta_1$ er større enn 1.

Ein bustad på 60 m$^2$ blir lagt ut med rettleiande pris 1.8 mill. kr. Definer $W$ som prisen (i millioner kroner) som blir betalt for denne bustaden.

Vis at $W$ er normalfordelt med forventningsverdi 1.98 og standardavvik 0.18.

Finn sannsynet for at salsprisen for bustaden er større enn 2 mill. kr.

Petter ønskjer å sjå korleis forholdet mellom rettleiande pris og salspris er, og observerer rettleiande prisar og salsprisar (i 1000 kr/m$^2$) for $N = 30$ bustader i Midtbyen. Dataa er plotta i figur 1a, og $\sum_{i=1}^{N} \frac{y_i}{x_i} = 32.98$.

**c)** Dersom vi antek uavhengige observasjonar, er sannsynsmaksimeringsestimatoren til $\beta_1$ (skal ikkje visast)

$$\hat{\beta}_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{x_i}.$$

Vis at $\hat{\beta}_1$ er normalfordelt med forventningsverdi $\beta_1$ og varians $\frac{\tau^2}{N}$.

Utlei eit 95 %-konfidensintervall for $\beta_1$, og finn intervallet numerisk basert på dataa frå Midtbyen.

**d)** For Petter er òg Tyholt ein aktuell stad å bu. Vi antek regresjonsmodellen

$$V = \beta_2 x + \epsilon(x)$$

for salspris per kvadratmeter for bustader på Tyholt, der $\epsilon$ er som i punkt (b), dvs. at $\epsilon(x)$ er normalfordelt med forventningsverdi 0 og varians $\tau^2 x^2$, der $\tau = 0.1$. Han ønskjer å teste om Midtbyen og Tyholt har same forhold mellom rettleiande pris og forventa salspris.

Sett dette opp som ein hypotesetest.

Data for $M = 50$ bustader på Tyholt er gjeve i figur 1b, og $\sum_{i=1}^{M} \frac{v_i}{x_i} = 56.66$.

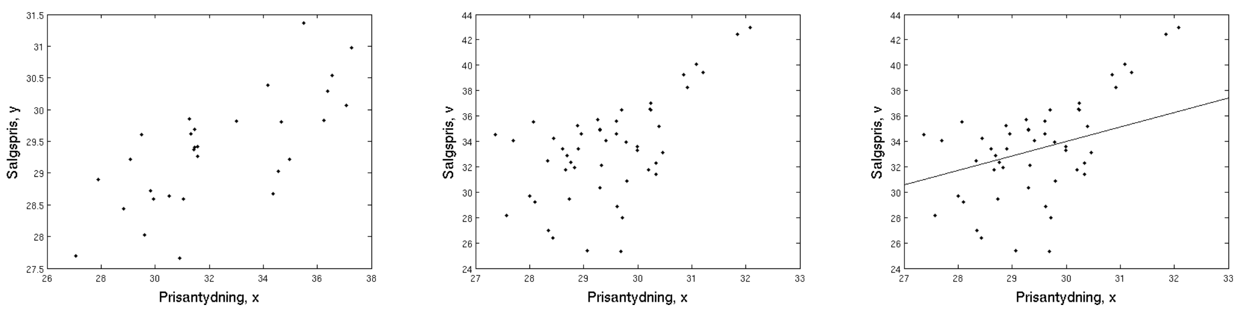
Utfør testen basert på dataa frå Midtbyen og Tyholt. Kva blir konklusjonen av testen med signifikansnivå 0.05?

**e)** For Tyholt er $\beta_2$ estimert frå data til $\hat{\beta}_2 = 1.13$, og vi får regresjonslinja som vist i figur 1c.

List opp modellantakingane som er gjort.

Ser antakingane i modellen ut til å stemme ut frå figuren? Forklar.

Grei òg ut om korleis du grafisk kan gjere vidare undersøkingar av antakingane.




(a) Midtbyen (b) Tyholt (c) Tyholt

Figur 1: Data for rettleiande pris (prisantydning) og salspris for (a) $N = 30$ bustader i Midtbyen, (b) og (c) $M = 50$ bustader på Tyholt. I (c) er regresjonslinja $v = \hat{\beta}_2 x$ teikna inn.

**Oppgave 3** Ferdighusfabrikken

Fabrikken Hus-i-tide produserer ferdighusmodular. Produksjonstida, dvs. talet på veker frå eit hus blir byrja med til det er klart til utkøyring, er eksponentielt fordelt med forventningsverdi $\mu$, altså med sannsynstettleik gitt ved $\frac{1}{\mu}e^{-x/\mu}$ for $x > 0$.

**a)** La $\mu = 2$ (veker) i dette punktet. Kva er sannsynet for at produksjonstida er mindre enn 1 veke?

Kva er sannsynet for at det av 5 hus ikkje er noka produksjonstid som er mindre enn 1 veke? Gå ut frå at produksjonstidene er uavhengige av kvarandre.

Fabrikken er ikkje nøgd med gjennomsnittleg produksjonstid og engasjerer konsulentfirmaet GodeDyreRåd. GodeDyreRåd foreslår ein alternativ produksjonsprosess som ifølgje konsulentane skal ha ei gjennomsnittleg produksjonstid som er $c$ gonger kortare enn med den gamle produksjonsprosessen, dvs. at produksjonstida for den nye produksjonsprosessen har sannsynstettleik gitt ved $\frac{c}{\mu}e^{-cy/\mu}$ for $y > 0$.

Gå vidare ut frå at $\mu$ er ukjend. Hus-i-tide ønskjer å estimere $\mu$ basert på $n_1$ produksjonstider $X_1, X_2, \ldots, X_{n_1}$ med den gamle produksjonsprosessen og $n_2$ produksjonstider $Y_1, Y_2, \ldots, Y_{n_2}$ med den nye. Dei bruker estimatoren $\tilde{\mu} = \alpha\bar{X} + \beta\bar{Y}$, der $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ og $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. Gå ut frå at alle produksjonstidene er uavhengige av kvarandre, og at $c$ er kjend.

**b)** Gå i dette punktet ut frå at $c = 2$, $\alpha = \frac{1}{2}$ og $\beta = 1$.

Rekn ut forventningsverdien og variansen til $\tilde{\mu}$. Er $\tilde{\mu}$ forventningsrett?

Utlei eit 95 %-konfidensintervall for $\mu$ basert på $\tilde{\mu}$, og finn talsvar når $n_1 = 30$, $n_2 = 20$, $\bar{x} = 2.07$ og $\bar{y} = 0.59$. Gjer greie for eventuelle tilnærmingar eller føresetnader du gjer.

**c)** Bestem $\alpha$ og $\beta$ slik at $\tilde{\mu}$ blir forventningsrett og får minst mogleg varians (blant forventningsrette estimatorar av denne forma).

Kva blir variansen?

Som nytilsett sivilingeniør ved Hus-i-tide med ansvar for kvalitet og leveringspålitelegskap får du mistanke om at verdien GodeDyreRåd har oppgitt for $c$ ikkje stemmer. Basert på $X_1, X_2, \ldots, X_{n_1}$ og $Y_1, Y_2, \ldots, Y_{n_2}$ ønskjer du å estimere også $c$ samtidig med $\mu$.

**d)** Finn sannsynsmaksimeringsestimatoren $(\hat{\mu}, \hat{c})$ for $(\mu, c)$

Bestem også sannsynsmaksimeringsestimatoren $\mu^*$ for $\mu$ når vi ser på $c$ som kjend, og samanlikn med estimatoren $\tilde{\mu}$ frå (b).

(Du treng ikkje argumentere for at kritiske punkt er maksima.)