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

Bokmål

Faglig 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, stemplet A5-ark med egne håndskrevne notater.

Sensuren faller: 24. juni 2011

## Oppgave 1 Bolig på Hawaii

Fra 1832 til 1950 hadde vulkanen Mauna Loa på Hawaii 37 utbrudd, altså med gjennomsnittlig rate (intensitet) på 0.026 utbrudd pr. måned. Over korte geologiske perioder kan en vulkans utbrudd antas å være en Poisson-prosess i tid. I forbindelse med studier ved University of Hawaii har du fått et svært gunstig tilbud på en fem års leiekontrakt for et hus ved foten av Mauna Loa.

**a)** Beskriv kort egenskapene som må være oppfylt for at en vulkans utbrudd skal kunne regnes som en Poisson-prosess i tid.

Hva blir sannsynligheten for at det i løpet av leieperioden på fem år skal komme minst ett vulkanutbrudd? Bruk raten (intensiteten) som fra 1832 til 1950.

Du får vite at Mauna Loa har hatt et utbrudd seks måneder før overtagelsesdato. Hva er sannsynligheten for at det skal gå mer enn tre år fra overtagelsesdato til neste utbrudd?

**Oppgave 2** Boligmarkedet i Trondheim

Petter har nettopp fått seg jobb i Trondheim og ser etter bolig. Ettersom han jobber i Midtbyen, ser han etter bolig der. For et utfallsrom som består av alle boliger som er til salgs i Trondheim på `finn.no` trekker Petter en bolig tilfeldig (alle boligene har samme sannsynlighet for å bli trukket). Vi definerer disse hendelsene:

$$M: \text{Boligen ligger i Midtbyen}$$
$$T: \text{Boligen har prisantydning på under 2 mill. kr}$$

**a)** Tegn hendelsene i et Venn-diagram.

Det er 381 boliger til salgs i Trondheim. Av disse er 94 i Midtbyen, og 190 har prisantydning på under 2 mill. Av boligene i Midtbyen har 50 prisantydning på under 2.000.000.

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

Begrunn og kommenter svarene.

Vi antar at hver bolig selges etter en budrunde til høyest bydende og at salgsprisen per kvadradmeter for boliger i Midtbyen kan modelleres med regresjonsmodellen

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

der $Y$ er salgspris (kr per m$^2$), $x$ er prisantydningen (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 en kjent parameter.

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

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

En bolig på 60 m$^2$ legges ut med prisantydning 1.8 mill. kr. Definer $W$ som prisen (i millioner kroner) som betales for denne boligen.

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

Finn sannsynligheten for at salgsprisen for boligen er større enn 2 mill. kr.

Petter ønsker å se hvordan forholdet mellom prisantydning og salgspris er, og observerer prisantydninger og salgspriser (i 1000 kr/m$^2$) for $N = 30$ boliger i Midtbyen. Dataene er plottet i figur 1a, og $\sum_{i=1}^{N} \frac{y_i}{x_i} = 32.98$.

**c)** Dersom vi antar uavhengige observasjoner, er sannsynlighetsmaksimeringsestimatoren til $\beta_1$ (skal ikke vises)

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

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

Utled et 95 %-konfidensintervall for $\beta_1$, og finn intervallet numerisk basert på dataene fra Midtbyen.

For Petter er også Tyholt et aktuelt sted å bo. Vi antar regresjonsmodellen

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

for salgspris per kvadratmeter for boliger 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 ønsker å teste om Midtbyen og Tyholt har samme forhold mellom prisantydning og forventet salgspris.

**d)** Sett dette opp som en hypotesetest.

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

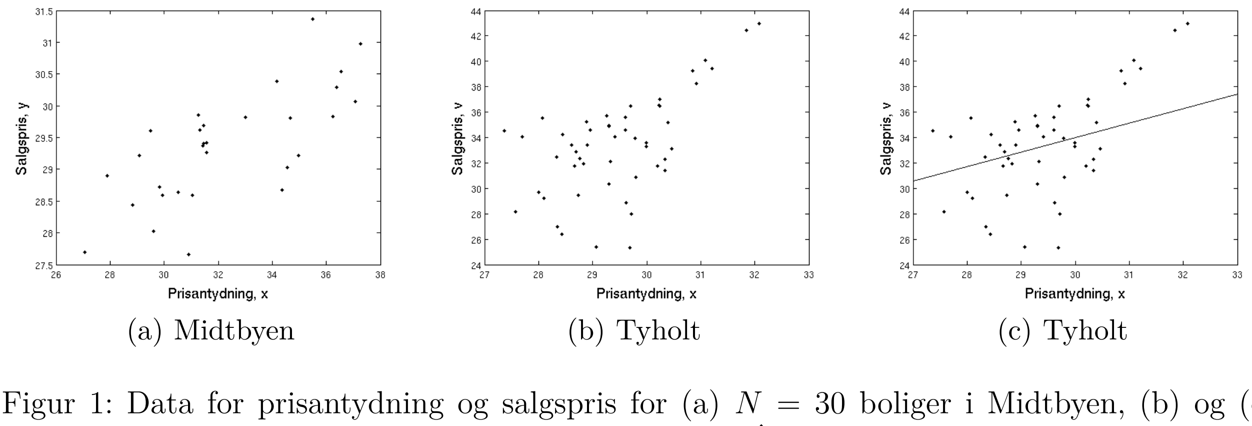
Utfør testen basert på dataene fra Midtbyen og Tyholt. Hva blir konklusjonen av testen med signifikansnivå 0.05?

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

List opp modellantagelsene som er gjort.

Ser antagelsene i modellen ut til å stemme ut fra figuren? Forklar.

Grei også ut om hvordan du grafisk kan gjøre videre undersøkelser av antagelsene.




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

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

**Oppgave 3** Ferdighusfabrikken

Fabrikken Hus-i-tide produserer ferdighusmoduler. Produksjonstida, dvs. antall uker fra et hus blir påbegynt til det er klart til utkjøring, er eksponentielt fordelt med forventningsverdi $\mu$, altså med sannsynlighetstetthet gitt ved $\frac{1}{\mu}e^{-x/\mu}$ for $x > 0$.

**a)** La $\mu = 2$ (uker) i dette punktet. Hva er sannsynligheten for at produksjonstida er mindre enn 1 uke?

Hva er sannsynligheten for at det av 5 hus ikke er noen produksjonstid som er mindre enn 1 uke? Anta at produksjonstidene er uavhengige av hverandre.

Fabrikken er ikke fornøyd med gjennomsnittlig produksjonstid og engasjerer konsulentfirmaet GodeDyreRåd. GodeDyreRåd foreslår en alternativ produksjonsprosess som ifølge konsulentene skal ha en gjennomsnittlig produksjonstid som er $c$ ganger kortere enn med den gamle produksjonsprosessen, dvs. at produksjonstida for den nye produksjonsprosessen har sannsynlighetstetthet gitt ved $\frac{c}{\mu}e^{-cy/\mu}$ for $y > 0$.

Anta videre at $\mu$ er ukjent. Hus-i-tide ønsker å estimere $\mu$ basert på $n_1$ produksjonstider $X_1$, $X_2$, ..., $X_{n_1}$ med den gamle produksjonsprosessen og $n_2$ produksjonstider $Y_1$, $Y_2$, ..., $Y_{n_2}$ med den nye. De 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 $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. Anta at alle produksjonstidene er uavhengige av hverandre, og at $c$ er kjent.

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

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

Utled et 95 %-konfidensintervall for $\mu$ basert på $\tilde{\mu}$, og finn tallsvar når $n_1 = 30$, $n_2 = 20$, $\bar{x} = 2.07$ og $\bar{y} = 0.59$. Gjør greie for eventuelle tilnærminger eller antakelser du gjør.

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

Hva blir variansen?

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

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

Bestem også sannsynlighetsmaksimeringsestimatoren $\mu^*$ for $\mu$ når $c$ antas kjent, og sammenlikn med estimatoren $\tilde{\mu}$ fra (b).

(Du trenger ikke argumentere for at kritiske punkter er maksima.)