TMA4245 Statistikk
Eksamen mai 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Kvalitetsavdelinga i ein fabrikk som produserer klokker ynskjer å sjå nøyare på dei defekte klokkene som av og til kjem frå produksjonen. Dei bestemmer seg for å nytta $k = 3$ defekte klokker i inspeksjonen. Frå ei produksjonsline kjem det ein kontinuerleg straum av klokker og kvar klokke som vert produsert har sannsyn p for å vera defekt, uavhengig av kvarandre.

La X vere det minste talet på klokker ein må inspisera frå produksjonsstraumen frå produksjonslina for å identifisera eksakt $k = 3$ defekte klokker. Me veit då at den tilfeldige variablen X er negativ-binomisk fordelt med sannsynsfordeling

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \qquad ; \qquad x = k, k+1, \ldots$$

a) Anta i dette punktet at defektsannsynet er $p = 0.1$ og rekn ut sannsyna

$$P(X > 3),$$
$$P(X < 6),$$
$$P(X \geq 6 | X > 3).$$

Defektsannsynet p er no anteke ukjend og skal estimerast. Kvalitetsavdelinga gjentar forsøket med å identifisera $k = 3$ defekte klokker n gonger, og får eit tilfeldig utval: $X_1, \ldots, X_n$. Basert på dette tilfeldige utvalet ynskjer ein å estimera p.

b) Utlei sannsynsmaksimeringsestimatoren $\hat{p}$ for p, basert på det tilfeldige utvalet.

Anta no at klokkefabrikken faktisk har to separate produksjonsliner, namngjeve høvesvis A og B og med ulike defektsannsyn p_A og p_B

Statistikaren i avdelinga får ein observasjon $X = x$ på talet på klokker som må inspiserast før $k = 3$ defekte er identifisert frå ei av de to produksjonslinene. Han veit ikkje om observajonen er henta frå produksjonsline A eller B, så han antar derfor i utgangspunktet sannsyn 0.5 for kvart av høvene.

c) Nytt Bayes sin regel til å utleia eit uttrykk for sannsynet for at observasjonen kjem frå produksjonsline A gjeve at $X = x$.

La så $p_A = 0.1$ og $p_B = 0.2$, samt $x = 5$, og rekn ut talsvaret for sannsynet for at observasjonen er frå produksjonsline A.

Oppgave 2

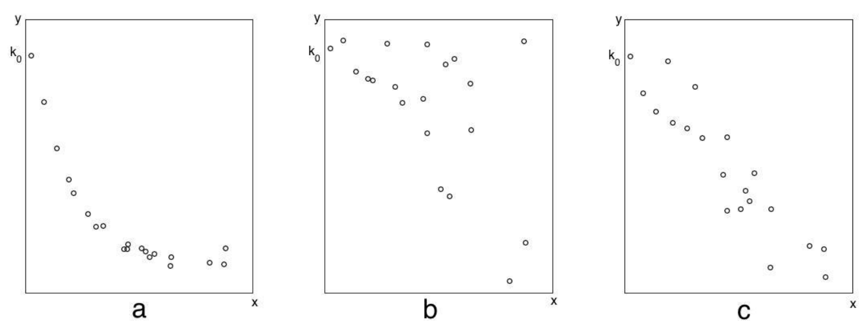
Ein bilprodusent vil evaluere slitasjen på bremseklossane på bilane som vert produsert. Ein definerer ein enkel lineær regresjonsmodell,

$$Y = k_0 - \beta x + \epsilon,$$

der responsvariablen Y er tjukkleiken på bremseklossane, forklaringsvariablen x er talet på kilometer køyrd, k_0 er klosstjukkleiken for ein ny bil og β er slitasjeraten. Feilleddet ϵ antas å vera normalfordelt, $n(\epsilon; 0, \sigma)$. Me antar at k_0 er kjend, medan raten β og variansen σ^2 er ukjende modellparametrar som skal estimerast.

Bilprodusenten designar eit forsøk for å estimera β og σ^2. Ei gruppe av n testsjåførar køyrer ulike bilar over eit varierande tal på kilometer og deretter målast klosstjukkleiken. Dette definerer eit tilfeldig utval frå modellen, $(x_1, Y_1), \ldots, (x_n, Y_n)$.

I Figur 1 presenterast tre plott av moglege utfall $(x_1, y_1), \ldots, (x_n, y_n)$ for $n = 20$.



Figur 1: Tre moglege utfall av forsøket i Oppgåve 2.

a) For kva for eit av desse tre plotta i Figur 1 synast den enkle lineære regresjonsmodellen definert over å vera ein god modell? Grunngje svaret.

Kvifor er ikkje modellen god for dei to andre plotta?

Den enkle lineære regresjonsmodellen definert over må naudsynleg vere approksimativ og gyldig berre for eit intervall av forklaringsvariablen x. Forklar kort kvifor.

b) Bruk anten minste kvadraters metode eller sannsynsmaksimeringsprinsippet til å utleia ein estimator $\hat{\beta}$ for β basert på det tilfeldige utvalet. Vis at estimatoren blir

$$\hat{\beta} = \frac{k_0 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i Y_i}{\sum_{j=1}^{n} x_j^2}.$$

Utlei uttrykk for forventningsverdien og variansen til $\hat{\beta}$.

Som estimator for σ^2 basert på det tilfeldige utvalet er det rimeleg å nytta

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[Y_i - (k_0 - \hat{\beta} x_i) \right]^2 .$$

Vidare er det oppgjeve at $\hat{\beta}$ er normalfordelt, at

$$V = \frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

er kji-kvadratfordelt med $(n-1)$ fridomsgrader, og at $\hat{\beta}$ og V er uavhengige tilfeldige variabler.

c) Utlei eit $100(1-\alpha)\%$-konfidensintervall for β.

Grei kort ut korleis konfidensintervallet kan nyttast til å testa om slitasjeraten er eksakt lik β_0.

Oppgave 3

Ein bonde frå Sogn dyrkar eple. Han pakkar og sel epla i det som er nemnd '3-kilo-poser'. Talet på eple i kvar pose er sjølvsagt eit heiltal, så posane varierer naudsynleg i vekt. Ein tilfeldig pose veg X kilogram, der X er normalfordelt med forventning μ og varians σ^2. Gå utifrå at μ er ukjend og la $\sigma^2 = 0.4^2$. Det er sjølvsagt ynskjeleg at forventninga μ er 3 kilogram.

Lageret til Rema 1000 på Sandmoen får eit stort billass med '3-kilo-poser' med eple frå bonden. Innkjøpsavdelingen på Rema 1000 ynskjer å kontrollera at posane er tunge nok. Dei tar eit tilfeldig utval på $n = 3$ posar frå billasset, veg disse posane, og registrerer følgjande vekter: X_1, X_2, X_3. Ein rimelig estimator for forventa vekt μ er

$$\hat{\mu} = \frac{1}{3}\sum_{i=1}^{3} X_i = \bar{X}.$$

a) Estimatoren $\hat{\mu}$ er normalfordelt, forklar kort med ord kvifor.

Utlei uttrykk for forventninga og variansen til denne normalfordelingen.

Er estimatoren $\hat{\mu}$ forventningsrett? Grunngje svaret.

Forklar kort med ord kva det inneber at ein estimator er forventningsrett.

Innkjøpsavdelinga ynskjer å sikra seg at forventa vekt av posane, μ, er minst 3 kilogram. Statistikaren i avdelinga formulerer vektkontrollen som eit hypotesetestingsproblem,

$$H_0 : \mu = 3 \text{ mot } H_1 : \mu < 3$$

og nyttar signifikansnivå $\alpha = 0.05$ i ein test med estimatoren $\hat{\mu}$ som testobservator.

b) Utlei forkastningsområdet for $\hat{\mu}$ med omsyn til hypotesane definert over.

c) Utlei eit uttrykk for styrkefunksjonen for testen som blei definert i punkt **b)**.

Skisser grafisk korleis styrkefunksjonen ser ut.

Dersom forventa vekt μ er på kun 2.9 kilogram, så ynskjer statistikaren å avsløra at posane veg for lite med sannsyn minst 0.9. Rekn ut kor mange posar n det då må vere i det tilfeldige utvalet som hentast frå billasset.

Statistikaren fortset å leika seg litt med problemet etter arbeidstid. Han ser på den ordna versjonen av det tilfeldige utvalet, $X_{(1)}, X_{(2)}, X_{(3)}$ i stigande orden. Deretter definerer han ein alternativ estimator for μ,

$$\tilde{\mu} = X_{(2)}$$

d) Utlei eit uttrykk for sannsynsfordelinga til estimatoren $\tilde{\mu}$.

Vis at $\tilde{\mu}$ er ein forventningsrett estimator for μ.

Fasit

1. a) 0.999, 0.00856, 0.9924 **b)** $\widehat{p} = \frac{nk}{\sum_{i=1}^{n}} X_i$ **c)** 0.1366

2. b) $\mathrm{E}\left[\widehat{p}\right] = \beta$, $\mathrm{Var}\left[\widehat{\beta}\right] = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

3. a) $\mathrm{E}\left[\widehat{\mu}\right] = \mu$, $\mathrm{Var}\left[\widehat{\mu}\right] = \frac{\sigma^2}{n}$, $\widehat{\mu}$ er forventningsrett **b)** Forkast H_0 dersom $\widehat{\mu} < 2.62$ **c)** 138 **d)** $f_{X_{(2)}}(x) = 6F_X(x)f_X(x)\left[1 - F_X(x)\right]$