TMA4245
Statistikk

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer b6

Oppgave 1 Kalibrering ved regresjon — Eksamen juni 2005, oppgave 3 av 3

Et apparat for registrering av stråling er tatt inn for kalibrering og kontroll. Vi skal i denne oppgaven anta at følgende relasjon gjelder mellom apparatets registrerte måleverdi Y og strålingsintensiteten x til en strålingskilde som plasseres i henhold til et gitt forsøksoppsett:

$$Y = \alpha + \beta\, x + \varepsilon$$

Her er α og β konstanter, og ε er en stokastisk (tilfeldig) variabel som, sammen med α, representerer effekten av bakgrunnsstrålingen. Det antas at ε er normalfordelt med forventningsverdi $\mathrm{E}(\varepsilon) = 0$ og varians $\mathrm{Var}(\varepsilon) = \sigma^2$.

a) Kalibreringen innledes ved å registrere m uavhengige måleverdier $y_1, \ldots, y_m$ for Y uten noen strålingskilde, dvs. med bare bakgrunnsstråling. Disse måleverdiene kan da betraktes som et tilfeldig utvalg fra en normalfordelt populasjon.

La $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ være middelet (gjennomsnittet) basert på dette tilfeldig utvalget. Hvilken fordeling har $\overline{Y}$?

Forklar at en rimelig estimator for α i dette tilfellet er $\overline{Y}$.

Anta i resten av oppgaven at α og σ er *kjente* parametere.

Andre fase i kalibreringen foregår ved å foreta målinger med et utvalg strålingsintensiteter $x_1, \ldots, x_n$, som gir måleverdiene $y_1, \ldots, y_n$. Vi kan da betrakte $y_i - \alpha - \beta\, x_i$, $i = 1, \ldots, n$, som et tilfeldig utvalg fra en normalfordelt populasjon med forventning 0 og varians σ^2.

b) Bruk prinsippet for sannsynlighetsmaksimering (maximum likelihood) til å finne en estimator for koeffisienten β. Alle steg i utledningen av uttrykket for estimatoren skal vises.

Utled også minste kvadratsums-estimatoren for β.

Sammenlign de to estimatorene.

Oppgave 2 Mosjonisten — Eksamen mai 2000, oppgave 4 av 4

En 45-åring startet med løpetrening for 9 år siden, og har hvert år siden deltatt i samme mosjonsløp. Anvendt tid, i minutter, er gitt i tabellen nedenfor.

år i	1	2	3	4	5	6	7	8	9
alder x_i	37	38	39	40	41	42	43	44	45
tid y_i	45.54	41.38	42.50	38.80	41.26	37.20	38.19	38.05	37.45

Det oppgis at $\sum_{i=1}^{9} x_i = 369$, $\sum_{i=1}^{9} y_i = 360.37$, $\sum_{i=1}^{9}(x_i - \bar{x})^2 = 60$, $\sum_{i=1}^{9}(y_i - \bar{y})^2 = 63.28$ og $\sum_{i=1}^{9}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{9}(x_i - \bar{x})y_i = -52.57$.

Vi skal anta at observasjonene kan ses på som realisasjoner av uavhengige normalfordelte variable $Y_1, \ldots, Y_9$, hvor $\mathrm{E}(Y_i) = \alpha + \beta x_i$ og $\mathrm{Var}(Y_i) = \sigma^2$.

a) Skriv opp de vanlige forventningsrette estimatorene $\hat{\alpha}$, $\hat{\beta}$ og $\hat{\sigma}^2$ for α, β og σ^2. Regn ut estimatene for α og β for de gitte dataene. Plott datasettet og den estimerte regresjonslinjen.

Det oppgis at estimatet for σ^2 er 1.568^2.

b) Regn ut et uttrykk for variansen til estimatoren $\hat{\beta}$.

Gjennomfør en test av $H_0:\ \beta = 0$ mot $H_1:\ \beta \neq 0$, på signifikansnivå 1%.

Hva blir den praktiske fortolkningen av testen over?

Løperen ønsker å predikere anvendt tid på mosjonsløpet neste gang (alder $x_0 = 46$ år).

c) Regn ut predikert tid.

Det oppgis at $\mathrm{Var}(\hat{\alpha} + \hat{\beta}x_0) = \sigma^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2})$. Utled et 95% prediksjonsintervall for Y ved $x_0 = 46$ år. Hva blir intervallet med de oppgitte data?

Hvis løperen ber deg predikere anvendt tid om 15 år (alder 60 år), hva vil du svare da?

Fasit

2. **a**) $\widehat{\alpha} = 75.96, \widehat{\beta} = -0.876$ **b**) $\mathrm{Var}(\widehat{\beta}) = \sigma^2 / \sum_{i=1}^{n}(x_i - \overline{x})^2$, Forkast H_0 **c**) $35.66, [31.09, 40.23]$