Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglige kontakter under eksamen:
Mette Langaas 98847649
Turid Follestad 98066880

EKSAMEN I FAG TMA4245 STATISTIKK

Lørdag 11.juni 2005

Tid: 09:00–13:00

Tillatte hjelpemidler:
Gult A5-ark med egne håndskrevne notater.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Kalkulator: HP30S.

BOKMÅL

Sensur: 4. juli 2005.

Oppgave 1 Meningsmålinger

I forkant av et stortingsvalg blir det gjennomført en meningsmåling der et representativt utvalg av velgerne blir spurt om de ønsker et regjeringsskifte eller ikke. Anta at andelen av velgerne som ønsker et skifte er p, og la X være antall personer blant n spurte som svarer JA på spørsmålet "Ønsker du et regjeringsskifte ved høstens valg?".

a) Under hvilke antagelser vil X her være binomisk fordelt? Du må relatere antagelsene til situasjonen som er beskrevet i oppgaveteksten.

Anta i resten av dette punktet at andelen av velgerne som ønsker et regjeringsskifte, er $p = 0.7$, og at $n = 20$ personer blir spurt. Bruk at X er binomisk fordelt.

Hva er sannsynligheten for at 18 eller flere av de 20 spurte svarer JA på spørsmålet om regjeringsskifte?

Hva er sannsynligheten for at flere enn 10, men færre enn 15, av de 20 sier JA?

Anta at to aviser på en bestemt dag presenterer resultater fra to meningsmålinger, gjennomført av hvert sitt meningsmålingsinstitutt, Byrå A og Byrå B. La n_1 være antall spurte og X_1 antall som svarer JA i målingen fra Byrå A, og n_2 og X_2 tilsvarende størrelser for Byrå B. Vi antar at X_1 er binomisk fordelt med parametre n_1 og p, og X_2 er binomisk fordelt med parametre n_2 og p, og at X_1 og X_2 er uavhengige.

Vi ønsker å estimere p ved å kombinere resultatene fra de to målingene. To aktuelle estimatorer er

$$\begin{aligned} \hat{P} &= \frac{1}{2}\left(\frac{X_1}{n_1} + \frac{X_2}{n_2}\right) \quad \text{og} \\ P^* &= \frac{X_1 + X_2}{n_1 + n_2}. \end{aligned}$$

b) Finn forventning og varians til hver av de to estimatorene $\hat{P}$ og P^*.

Dersom $n_1 = 500$ og $n_2 = 1000$, hvilken estimator vil du da velge? Begrunn svaret.

Anta nå at $n_1 = n_2 = n$, slik at X_1 og X_2 er uavhengige og binomisk fordelte, med *samme* parametre p og n. Dette medfører at

$$\hat{P} = P^* = \frac{X_1 + X_2}{2n}.$$

Utled et tilnærmet 95% konfidensintervall for p ved å bruke at fordelingen til

$$\frac{\hat{P} - p}{\sqrt{\frac{1}{2n}\hat{P}(1-\hat{P})}}$$

er tilnærmet standard normalfordelt.

Et tredje meningsmålingsinstitutt, Byrå C, har annonsert at de snart kommer med resultater fra en tilsvarende måling med n_3 spurte. La X_3 være antall som svarer JA på spørsmålet om regjeringsskifte i målingen fra Byrå C, og anta at X_3 er uavhengig av X_1 og X_2. Vi vil nå bruke resultatene fra Byrå A og Byrå B til å predikere hvor mange som svarer JA i den nye målingen. Vi antar i resten av oppgaven at $n_1 = n_2 = n_3 = n = 1000$, og at observerte verdier for X_1 og X_2 er $x_1 = 645$ og $x_2 = 692$.

c) La $Y = X_3 - n\hat{P}$, der $\hat{P} = \dfrac{X_1 + X_2}{2n}$.

Begrunn at det i vår situasjon er rimelig å anta at Y er tilnærmet normalfordelt, og vis at variansen til Y er $\frac{3}{2}np(1-p)$.

Bruk dette til å utlede et tilnærmet 95% prediksjonsintervall for antallet spurte som i målingen fra Byrå C svarer JA på spørsmålet om regjeringsskifte.

Bestem også intervallet numerisk basert på de observerte verdiene.

Oppgave 2 Veiprosjektet

Vi ser på kostnaden av et veiprosjekt. La X være en kontinuerlig stokastisk (tilfeldig) variabel som angir den faktiske kostnaden pr. meter for veien som skal bygges. Kostnaden pr. meter er avhengig av grunnforhold og pris på materialer. Vi antar at X er normalfordelt med forventningsverdi $\mathrm{E}(X) = \mu$ og standardavvik $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sigma$.

a) La oss anta (kun i dette punktet) at det er kjent at $\mu = 10000$ kr/meter og $\sigma = 2500$ kr/meter. Før veien bygges, ønsker man å regne ut sannsynligheten for ulike utfall.

Hva er sannsynligheten for at kostnaden pr. meter for veien vil overskride 13000 kr/meter?

Finn et tall, k, som er slik at sannsynligheten er 0.05 for at den faktiske kostnaden pr. meter for veien blir mindre enn k.

Gitt at vi vet at kostnaden pr. meter blir minst 10000 kr/meter, hva er da sannsynligheten for at kostnaden pr. meter blir høyere enn 13000 kr/meter?

Vi antar i resten av oppgaven at både μ og σ er ukjente størrelser.

En ekspertgruppe mener at forventet kostnad pr. meter for veien som skal bygges blir 10000 kr/meter. I tillegg er det samlet inn data fra $n = 9$ veiprosjekter med tilsvarende grunnforhold og materialkostnader, og dataene finnes i tabell 1. Her er x_i kostnaden i kr/meter for veiprosjekt nummer i, og det oppgis at $\sum_{i=1}^{9} x_i = 106480$ og $\sum_{i=1}^{9}(x_i - \bar{x})^2 = 49295335$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
10099	10925	15397	11676	11823	15788	12652	8337	9783

Tabell 1: *Data over kostnad i kr/meter for $n = 9$ veiprosjekter.*

Prosjektlederen er skeptisk til kostnadsanslaget på $\mu = 10000$ kr/meter, og mener at grunnforholdene tilsier at forventet kostnad i kr/meter er høyere.

b) Formulér dette som en hypotesetest ved å definere nullhypotese og alternativ hypotese.

Sett opp en testobservator og finn forkastingsområdet. Hva blir konklusjonen på testen, med data gitt i tabell 1, når signifikansnivået er $\alpha = 0.01$?

Regn ut p-verdien ved å bruke tabell 2 (øverst på neste side).

t	1.8	1.9	2.0	2.1	2.2	2.3	2.4
$\nu = 7$	0.943	0.950	0.957	0.963	0.968	0.973	0.976
$\nu = 8$	0.945	0.953	0.960	0.966	0.971	0.975	0.978
$\nu = 9$	0.947	0.955	0.962	0.967	0.972	0.977	0.980

Tabell 2: *Kumulativ sannsynlighet i t-fordelingen. For en stokastisk variabel T som er t-fordelt med ν frihetsgrader, så viser tabellen $P(T \leq t)$ for ulike verdier av t.*

Den totale kostnaden for veien er avhengig av veiens lengde (i meter) og kostnad pr. meter (i kr/meter).

La Y være en kontinuerlig stokastisk variabel som angir den faktiske lengden av veien som skal bygges. Vi antar at lengden av veien er normalfordelt, med forventingsverdi $E(Y) = \eta$ og standardavvik $SD(Y) = \sqrt{\mathrm{Var}(Y)} = \tau$. Vi har fra før at kostanden pr. meter, X, er normalfordelt, med forventningsverdi $E(X) = \mu$ og standardavvik $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sigma$.

Vi antar at X og Y er uavhengige stokastiske variabler.

Kostnaden for veien kan uttrykkes som $W = X \cdot Y$.

c) Finn forventningsverdien og variansen til W uttrykt ved μ, η, σ og τ.

Oppgave 3 Kalibrering ved regresjon

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.