Noregs teknisk–
naturvitskaplege universitet
Institutt for matematiske fag

Faglege kontakter under eksamen:
Mette Langaas 98847649
Turid Follestad 98066880

EKSAMEN I FAG TMA4245 STATISTIKK

Laurdag 11.juni 2005

Tid: 09:00–13:00

Tillatne hjelpemiddel:
Gult A5-ark med eigne handskrivne notatar.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Kalkulator: HP30S.

NYNORSK

Sensur: 4.juli 2005.

Oppgåve 1 Meiningsmålingar

I forkant av eit stortingsvalg blir det gjennomført ei meiningsmåling der eit representativt utvalg av velgarane blir spurte om dei ønskjer eit regjeringsskifte eller ikkje. Gå ut ifrå at andelen av velgarane som ønskjer eit skifte er p, og la X vere talet på personar blant n spurte som svarer JA på spørsmålet "Ønskjer du eit regjeringsskifte ved valget til hausten?"

a) Under kva for føresetnader vil X her vere binomisk fordelt? Du må relatere føresetnadene til situasjonen som er skildra i oppgåveteksten.

Føreset i resten av dette punktet at andelen av velgarane som ønskjer eit regjeringsskifte er $p = 0.7$, og at $n = 20$ personar blir spurte. Bruk at X er binomisk fordelt.

Kva er sannsynet for at 18 eller fleire av dei 20 spurte svarer JA på spørsmålet om regjeringsskifte?

Kva er sannsynet for at fleire enn 10, men færre enn 15, av dei 20 seier JA?

Føreset at to aviser på ein bestemt dag presenterer resultat frå to meiningsmålingar, gjennomførte av kvart sitt meiningsmålingsinstitutt Byrå A og Byrå B. La n_1 vere talet på spurte og X_1 talet som svarer JA i målinga frå Byrå A, og n_2 og X_2 tilsvarande storleikar for Byrå B. Vi føreset at X_1 er binomisk fordelt med parametrar n_1 og p og X_2 er binomisk fordelt med parametrar n_2 og p, og at X_1 og X_2 er uavhengige.

Vi ønskjer å estimere p ved å kombinere resultata frå dei to målingane. To aktuelle estimatorar 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 kvar av dei to estimatorane $\hat{P}$ og P^*.

Dersom $n_1 = 500$ og $n_2 = 1000$, kva for ein estimator vil du då velge? Grunngi svaret.

Vi føreset no at $n_1 = n_2 = n$, slik at X_1 og X_2 er uavhengige og binomisk fordelte, med *same* parametrar p og n. Dette medfører at

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

Utlei eit tilnærma 95% konfidensintervall for p ved å bruke at fordelinga til

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

er tilnærma standard normalfordelt.

Eit tredje meiningsmålingsinstitutt, Byrå C, har annonsert at dei snart kjem med resultat frå ei tilsvarande måling med n_3 spurte. La X_3 vere talet på personar som svarer JA på spørsmålet om regjeringsskifte i målinga fra Byrå C, og anta at X_3 er uavhengig av X_1 og X_2. Vi vil no bruke resultata frå Byrå A og Byrå B til å predikere kor mange som svarer JA i den nye målinga. Vi føreset i resten av oppgåva at $n_1 = n_2 = n_3 = n = 1000$, og at observerte verdiar for X_1 og X_2 er $x_1 = 645$ og $x_2 = 692$.

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

Grunngi at det i vår situasjon er rimeleg å føresette at Y er tilnærma normalfordelt, og vis at variansen til Y er $\frac{3}{2}np(1 - p)$.

Bruk dette til å utleie eit tilnærma 95% prediksjonsintervall for talet på spurte som i målinga frå Byrå C svarer JA på spørsmålet om regjeringsskifte.

Bestem også intervallet numerisk basert på dei observerte verdiane.

Oppgåve 2 Vegprosjektet

Vi ser på kostnaden av eit vegprosjekt. La X vere ein kontinuerlig stokastisk (tilfeldig) variabel som representerer den faktiske kostanden pr. meter for vegen som skal byggast. Kostnaden pr. meter er avhengig av grunnforhold og pris på materialar. Vi føreset at X er normalfordelt med forventningsverdi $\mathrm{E}(X) = \mu$ og standardavvik $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sigma$.

a) La oss føresette (kun i dette punktet) at det er kjent at $\mu = 10000$ kr/meter og $\sigma = 2500$ kr/meter. Før vegen blir bygt, ønskjer ein å rekne ut sannsynet for ulike utfall.

Kva er sannsynet for at kostnaden pr. meter for vegen vil overskride 13000 kr/meter?

Finn eit tal, k, som er slik at sannsynet er 0.05 for at den faktiske kostanden pr. meter for vegen blir mindre enn k.

Gitt at vi veit at kostnaden pr. meter blir minst 10000 kr/meter, kva er då sannsynet for at kostnaden pr. meter blir høgare enn 13000 kr/meter?

Vi føreset i resten av oppgåva at både μ og σ er ukjente storleikar.

Ei ekspertgruppe meiner at forventa kostnad pr. meter for vegen som skal byggast blir 10000 kr/meter. I tillegg er det samla inn data frå $n = 9$ vegprosjekt med tilsvarande grunnforhold og materialkostnader, og dataene fins i tabell 1. Her er x_i kostnaden i kr/meter for vegprosjekt nummer i, og det blir oppgitt 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 $n = 9$ vegprosjekt.*

Prosjektleiaren er skeptisk til kostnadsanslaget på $\mu = 10000$ kr/meter, og meiner at grunnforholda tilseier at forventa kostnad i kr/meter er høgare.

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

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

Rekn 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: *Kumulativt sannsyn i t-fordelinga. For ein stokastisk variabel T som er t-fordelt med ν fridomsgrader, så viser tabellen $P(T \leq t)$ for ulike verdiar av t.*

Den totale kostnaden for vegen er avhengig av lengda på vegen (i meter) og kostnad pr. meter (i kr/meter).

La Y vere ein kontinuerlig stokastisk variabel som representerer den faktiske lengda av vegen som skal byggast. Vi føreset at lengda av vegen er normalfordelt, med forventing $E(Y) = \eta$ og standardavvik $SD(Y) = \sqrt{\text{Var}(Y)} = \tau$. Vi har frå før at kostanden pr. meter, X, er normalfordelt, med forventning $E(X) = \mu$ og standardavvik $\text{SD}(X) = \sqrt{\text{Var}(X)} = \sigma$.

Vi føreset at X og Y er uavhengige stokastiske variablar.

Kostnaden for vegen kan uttrykkast som $W = X \cdot Y$.

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

Oppgåve 3 Kalibrering ved regresjon

Eit apparat for registrering av stråling er tatt inn for kalibrering og kontroll. Vi skal i denne oppgåva føresette at følgjande relasjon gjeld mellom den registrerte måleverdien Y for apparatet og strålingsintensiteten x til ei strålingskjelde som blir plassert i følgje eit gitt forsøksoppsett:

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

Her er α og β konstantar, og ε er ein stokastisk (tilfeldig) variabel som, saman med α, representerer effekten av bakgrunnsstrålinga. Det blir føresett at ε er normalfordelt med forventningsverdi $\mathrm{E}(\varepsilon) = 0$ og varians $\mathrm{Var}(\varepsilon) = \sigma^2$.

a) Kalibreringa blir innleia ved å registrere m uavhengige måleverdiar $y_1, \ldots, y_m$ for Y utan noka strålingskjelde, dvs. med berre bakgrunnsstråling. Desse måleverdiane kan vi då sjå på som eit tilfeldig utvalg fra ein normalfordelt populasjon.

La $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ vere middelet (gjennomsnittet) basert på dette tilfeldige utvalget. Kva for ei fordeling har $\overline{Y}$?

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

Føreset i resten av oppgåva at α og σ er *kjente* parametrar.

Andre fase i kalibreringa går føre seg ved å gjere målingar med eit utvalg strålingsintensitetar $x_1, \ldots, x_n$, som gir måleverdiane $y_1, \ldots, y_n$. Vi kan då sjå på $y_i - \alpha - \beta\, x_i$, $i = 1, \ldots, n$, som eit tilfeldig utvalg frå ein normalfordelt populasjon med forventning 0 og varians σ^2.

b) Bruk prinsippet for sannsynsmaksimering (maximum likelihood) til å finne ein estimator for koeffisienten β. Alle steg i utleiinga av uttrykket for estimatoren skal visast.

Utlei også minste kvadratsums estimatoren for β.

Samanlikn dei to estimatorane.