Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Henning Omre 73 59 35 31
Bjørn Kåre Hegstad 73 59 35 39

EKSAMEN I FAG 75515 STATISTIKK 1

Lørdag 10. januar 1998
Tid: 09:00–14:00

Hjelpemidler:
Godkjent lommekalkulator, utdelt ordliste, Statistiske tabeller og formler (Tapir forlag).

Oppgave 1 Helseundersøkelsen

Det har kommet en effektiv behandling for en dødelig sykdom der en tidligere ikke hadde noen behandling. Det planlegges en masseundersøkelse av alle voksne nordmenn, dvs. 3 mill. individer. En vet at sannsynligheten for at en voksen nordmann har denne sykdommen er 0.003. En ønsker altså å identifisere de individer som har sykdommen med en laboratoriumsprøve.

En vet at for en person med denne sykdommen så vil prøven slå ut med sannsynlighet 0.99. På den andre side så vil prøven slå ut feilaktig på personer uten sykdommen med sannsynlighet 0.005.

Hendelsen at en person har sykdommen benevnes D og hendelsen at prøven slår ut benevnes A.

a) Hvor mange voksne nordmenn ville en forvente har sykdommen?

Hvor mange av disse vil en forvente ikke blir avslørt av prøven?

Benytt Bayes regel og utled svarene i den notasjon som er gitt over før tallsvar regnes ut:

b) Hva er sannsynligheten for at en person som prøven har slått ut for, ikke har sykdommen?

Hva er sannsynligheten for at en person som prøven ikke har slått ut for, har sykdommen?

Gi en kort kommentar til svarene.

For å forbedre undersøkelsen, utføres en tilleggsprøve på de personene hvor den første prøven slo ut. For disse personene som den første prøven har slått ut for, vet en at tilleggsprøven har effektivitet 10, dvs. at tilleggsprøven slår ut med ti ganger større sannsynlighet for de som har sykdommen, enn for de som ikke har sykdommen.

Hendelsen at tilleggsprøven slår ut benevnes B.

Benytt Bayes regel og utled svaret i den notasjon som er gitt over før tallsvar regnes ut:

c) Gitt en person som begge prøvene har slått ut for, hva er sannsynligheten for at personen ikke har sykdommen?

Oppgave 2 Jagland

Et meningsmålingsinstitutt ønsket i høst før stortingsvalget å anslå oppslutningen Arbeiderpartiet ville få ved valget. Instituttet spurte $n = 1200$ tilfeldig utvalgte personer av totalt $N = 3$ mill. stemmeberettigede om de ville stemme Arbeiderpartiet ved valget. Av de spurte svarte 413 "JA". De resterende svarte "NEI". La X være antallet som svarer "JA" ved en slik spørreundersøkelse.

a) Hvilken type fordeling har X? Begrunn svaret.

For N mye større enn n tilnærmes gjerne denne fordelingen med en binomisk fordeling. Sett opp forutsetningene for at antallet som svarer "JA" skal være binomisk fordelt og begrunn kort hvorfor disse forutsetningene er tilnærmet oppfylt i dette tilfellet.

Anta at X er binomisk fordelt med forventning $n\theta$ der n er antall spurte personer.

b) Vis at sannsynlighetsmaksimeringsestimatoren for θ er $\hat{\theta} = \frac{X}{n}$.
Finn estimatet for θ.

Vis videre at $E(\hat{\theta}) = \theta$ og at $\text{Var}(\hat{\theta}) = \theta(1-\theta)/n$.
Hvilken tolkning har θ i denne sammenhengen med hensyn på valgresultatet?

Siden det er rimelig at både $n\theta$ og $n(1-\theta)$ er større enn 5, kan du heretter anta at

$$\frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{\text{Var}(\hat{\theta})}}$$

er tilnærmet standard normalfordelt.

Jagland sa før valget at han ville trekke seg som statsminister dersom oppslutningen ble under 36.9%. Løssalgsavisen som bestilte undersøkelsen, ønsker en salgsfremmende forside. De vil vite om de med bare 5% sannsynlighet for å ta feil, kan påstå at Jagland kommer til å trekke seg som statsminister.

c) Formuler dette som et hypotesetestproblem og utfør testen.
Hva ville du råde løssalgsavisen til å gjøre?

Et annet meningsmålingsinstitutt fikk samme estimat for θ, men motsatt konklusjon på hypotesetesten. Hvor mange hadde de spurt?
Her trenger du bare å finne en øvre eller nedre grense.

Oppgave 3 Vannlekkasje

Det er lekkasje på en nedgravd vannledning og kommunens peilepersonell er sendt ut for å undersøke dette. Det gjøres målinger av dypet ned til vannledningen i noen utvalgte posisjoner. Posisjonene kan identifiseres eksakt, men dybdemålingene er beheftet med målefeil.

Følgende observasjoner gjøres:

i	1	2	3	4	5	6	7	8
l_i	1.0	3.0	5.0	7.0	9.0	11.0	13.0	15.0
d_i	2.88	2.92	2.82	2.73	2.91	2.76	2.62	2.80

hvor l er lengden langs vannledningen fra en gitt kum og d er dypet til ledningen. Det er altså åtte observasjoner (d_i, l_i); $i = 1, \ldots, 8$. Ved utregning får en at $\sum_{i=1}^{8} d_i = 22.44$ og $\sum_{i=1}^{8} d_i^2 = 63.0162$. Anta først følgende modell for observasjonene:

$$D_i = \alpha + \epsilon_i$$

der α er det konstante dyp for vannledningen og ϵ_i; $i = 1, \ldots, 8$ er uavhengige normalfordelte variable med kjent forventning 0 og ukjent varians σ_1^2.

En rimelig estimator for dypet til ledningen er

$$\hat{\alpha} = \frac{1}{8} \sum_{i=1}^{8} D_i \quad .$$

a) Finn forventning og varians til $\hat{\alpha}$.

Hvilken type sannsynlighetsfordeling har $\hat{\alpha}$? Begrunn svaret.

b) Utled et 95% konfidensintervall for dypet til vannledningen.

Vannledningen er som kjent lekk og en blir enige om at det muligens er et vertikalt brudd i ledningen i posisjon $l_b = 10.0$. En antar at ledningen etter denne posisjonen er hevet med β.

En antar altså følgende modell:

$$D_i = \alpha - \beta \cdot I_{\geq l_b}(l_i) + \epsilon_i$$

hvor $\epsilon_i;\ i = 1, \ldots, 8$ er uavhengige normalfordelte variable med kjent forventning 0 og kjent varians $\sigma_2^2 = 0.1^2$, og hvor

$$I_{\geq l_b}(l) = \begin{cases} 1 & \text{hvis} \quad l \geq l_b \\ 0 & \text{ellers} \end{cases}$$

c) Utled sannsynlighetsmaksimeringsestimatorene for α og β som er

$$\begin{aligned} \hat{\alpha} &= \frac{1}{5}\sum_{i=1}^{5} D_i \\ \hat{\beta} &= \frac{1}{5}\sum_{i=1}^{5} D_i - \frac{1}{3}\sum_{i=6}^{8} D_i \end{aligned}$$

Vis videere at $\hat{\alpha}$ og $\hat{\beta}$ er forventningsrette.
Finn estimatene for α og β basert på observasjonene i tabellen over.

Da peilepersonellet kom tilbake til kontoret, stilte kommuneingeniøren seg svært tvilende til om det var noe vertikalt brudd i posisjon l_b. Han påstod ganske enkelt at $\beta = 0$.

d) Bruk statistisk hypotesetesting til å undersøke om det er noen grunn til å forkaste kommuneingeniørens påstand. Bruk signifikansnivå 0.05.