**Stat100 vår 2018**

**Oppgaver til uke 11**

**Oppgave 1 (fin oppgave, må gjøres)**

Anta at du har som masteroppgave å komponere et nytt svinefôr. Du plukker ut 3 griser tilfeldig og gir disse av det nye fôret, deretter slakter du grisene når de er 180 dager gamle.

Slaktevektene ble: 70 kg, 80 kg og 90 kg. Vi antar at slaktevekt ved det nye fôret er normalfordelt.

En statistisk modell for dette forsøket kan vi skrive som:

**Yi =  + ei, der ei ~ N(0,), og uavhengige*.***

***i =1,2 3. Alle observasjoner er uavhengige***

**a)** Vis at modellen medfører at

E(Yi) = 

SD(Yi) = 

Yi-ene er normalfordelte,

Modellen er dermed identisk med modellen: *Yi* ~ N(****)

**b)** Modellen har to ukjente parametre, gi disse en tolkning.

**c)** Vis at ei = Yi  ***.***

Hvordan vil du tolkeei?

Hva vil du si om gris nummer i hvis den har en negativ verdi av ei?

Hvorfor kan vi ikke beregne ei?

**d)** Estimer begge parametre.

**e)** Finn standardfeilen til estimatoren for .

**Oppgave 2:**

Reisevanene til 36 tilfeldige personer kartlegges i en undersøkelse, og en fant at 29 av disse foretrekker privatbilen framfor kollektivtrafikken.

1. Estimere andelen som ønsker å reise kollektivt.
2. Hva er standardfeilen til estimatet?
3. Dersom du vil halvere usikkerheten til estimatet, hvor mange personer måtte du ha kartlagt?

**Oppgave 3:**

I en spørreundersøkelse svarte 360 av 1050 at de støtter arbeiderpartiet.

1. Estimer partiets oppslutning i befolkningen.
2. Hvorfor tror du at det står i avisene at usikkerheten til denne målingen er ca. 1,5 %?

**Oppgave 4**

Før valget spurte et meningsmålingsfirma 1000 personer og et annet meningsmålingsfirma 2000 (andre) personer om hva de stemte. En avis foreslo å ta Arbeiderpartiets gjennomsnitts-oppslutning på disse to målingene som estimat på den sanne oppslutningen til partiet.

1. Vis at dette er en forventningsrett estimator.
2. Hvorfor er denne ikke optimal? Foreslå en bedre estimator.
3. Dersom 360 svarte Arbeiderpartiet i måling 1, og 680 i måling 2, hva blir estimert oppslutning?
4. Hva blir standardfeilen til estimatet?

**Oppgåve 5**

I ei mære er det samla mange tusen laks. Ein undrast på kva gjennomsnittsvekta (forventninga) av desse er. Odd fiskar ein laks som veg 800 gram, medan Kåre fiskar to laks med ein gjennomsnittvekt på 1100 gram.

1. **(frå midtsemestereksamen vår 2009)**

Kva for eit av desse forventningsrette estimata for forventa laksevekt er det beste?

1. 1100 gram b) 800 gram c) 1000 gram d) 900 gram e) 950 gram

Kan du vise at alle estimatene er forventningsrette?

Kan du argumentere for hvilken som skal velges?

**b**

1. Lag eit 90 % konfidensintervall for gjennomsnittsvekt for all fisk i mæra, dersom du antek at (populasjons-)standardavviket til vekt er 400 gram.
2. Sett opp modellen for observasjonane med antagelser.
3. Forklar fiskeoppdrettaren kva intervallet tyder.
4. Dersom han berre vil slakte fisk viss gjennomsnittet i mæra ligg mellom 0,9 og 1,1 kilo, kva vil du råde han til?

**Oppgave 6** (Oppgave 6.4 i læreboka)

I en by i Sør-Amerika er innbyggerne enten fattige eller rike. En tilfeldig fattig person har inntekt *X* med forventningsverdi  og varians \sigma^2. En tilfeldig rik person har inntekt *Y* med forventningsverdi 15 og varians 4\sigma^2. En norsk forsker ønsker å anslå verdien av . Han samler inn fem *X*-verdier og fem *Y-*verdier. Hvilken av de to følgende estimatorer for  er best:

\hat{\mu}_1 = \frac{\bar{X}+\bar{Y}}{16}    \hat{\mu}_2 = \frac{\bar{X} + 4\bar{Y}}{61}

**Oppgave 7**

På en landsomfattende prøve i matematikk ble det gitt poeng på en skala fra 0 til 100.

Anta at poeng for gutter er normalfordelt med forventning g og standardavvik .

Anta at poeng for jenter er normalfordelt med forventning j og standardavvik .

Anta at du samler inn resultater fra 10 jenter og 12 gutter. Dette ga følgende resultat:

Gjennomsnitt standardavvik (S)

Jenter 75 8

Gutter 65 10

1. Hvor mange ukjente parametere er det her?
2. Hvordan vil du tolke j - g?
3. Hvordan vil du estimere j - g?
4. Vis at denne estimatoren er forventningsrett og finn dens standardavvik.

Det beste estimatoren for standardavviket () er 

Der Sj er utvalgsstandardavviket for jenter og Sg tilsvarende for gutter.

1. Beregn standardfeilen til estimatoren.

**Oppgave 8**

Måling av kroppstemperatur i øret er forbundet med ganske stor variabilitet, men vi antar at apparaturen er slik at det ikke medfører systematiske feil.

En sykepleier velger å ta mange (n) målinger, X1,X2,…Xn av samme pasient for å få et sikrere resultat i form av et gjennomsnitt.

La oss anta at standardavviket (σ) er kjent fra *mange* tidligere målinger.

Vi har målt mange ganger på en pasient

Og vi har gjort dette på menge pasienter.

PÅ denne måten har vi sett at variasjon i temperaturmåling på samme pasient er konstant (den uavhengig av feber).

Vi antar at σ = 0.5.

På en pasient får sjukepleieren disse 8 målingene tatt rett etter hverandre på en pasient:

39.1, 38.8, 39.9, 38.9, 38.8, 38.7, 39.0, 38.4

1. Sett opp en modell for observasjonene.
2. Gi en tolkning av begge parametre i modellen

For oppgavene nedenfor 3 og 5 nedenfor, bruk helst R-commander.

1. Estimer den sanne kroppstemperaturen til denne pasienten.
2. Hva er standardavviket til dette estimatet?
3. Lag et 95% Konfidensintervall for sann kroppstemperatur (for denne pasienten i nåværende helsetilstand).
4. Gi en tolkning av intervallet.
5. Hvor mange målinger måtte hun ha utført for at intervallbredden skal bli mindre enn 0.2 grader?
6. Hvor mange målinger må du ta hvis teknologien forbedres slik at σ reduseres til 0.3? (Bredden skal fortsatt være mindre enn 0,2 grader).
7. Hvor mange målinger med dagens teknologi må du ta for at et 99% konfidensnivå skal ha bredde mindre enn 0.2? Lar dette seg gjøre i praksis.
8. Hva er intervallbredden med bare én observasjon (95% konfidensnivå)?

Lar det seg gjøre med en måling?

**Oppgave 9** **(Tidkrevende, men lærerik)**

To laboratorier har hver sin metode for å måle en viss størrelse, la oss kalle denne  (for eksempel c-vitamininnhold i et stor tank appelsinjuice). Begge metoder er svært kompliserte, og derfor gir de en viss måleusikkerhet. Det betyr at **samme** prøve kan gi forskjellige resultat dersom du måler den flere ganger. Laboratorium A har nøyaktigere (og dyrere?) instrument enn laboratorium B, og grunnet langvarig bruk av instrumentet vet vi at standardavviket (****** er dobbelt så stort i B som i A.

La *Xi* være måling nr. *i* for A og *Yi* være måling nr. *i* for B. Anta at målingene er uavhengig med

*Xi* ~ N(****) og *Yi* ~ (****).

Anta at du har en prøve som blir målt 10 ganger i laboratorium A og 20 ganger i laboratorium B. Følgende 8 estimatorer for  ble foreslått:

















1. Vis at ikke alle estimatorene er forventingsrette. Disse må ekskluderes.
2. Finn variansen til alle forventningsrette estimatorer.
3. Hvilken ville du ha valgt? Begrunn svaret.

I Lab A fikk vi et gjennomsnitt på 20,1 og i Lab B et gjennomsnitt på 21,2 (mg/100 g saft):

1. Finn verdien på de forskjellige estimatene som ikke er ekskludert.

***Tilleggsspørsmål for matematisk interesserte studenter (ta kontakt med foreleser for løsningsforslag):***

*Vis at alle estimatorer for på formen a + (1-a) er forventningsrette.*

*Vis at det valget av a som gir minst varians er a = 2/3.*

*Vis at  oppfyller dette.*

*Kan du tenke deg en måte å estimre  på?*

**Matematikkoppgave,** (for matematisk orienterte studenter, spør foreleser om hjelp)

Vi skal finne E(S2) der 

Vi har at 2 = var(X) = E(X –)2, der  er forventning til X.

Vis at 2 = E(X2) –2. (Hint 2. kvadratsenting) Vis at E(X2) = 2 +2

Vis at E(2) =  +2

Vis at = 

Vis at: E() = n2 +2) – n+2)

Begrunn hvorfor du må dividere på n - 1 dersom du skal ha en forventningsrett estimator for 2.