TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Skriftlig innlevering 6 (blokk 2)

Dette er den siste skriftlige innleveringen. Den er basert på det som er diskutert i forelesningene frem til og med uke 46. Spesielt er det i denne innleveringen fokus på hypotesetesting og enkel lineær regresjon. Alle deloppgaver teller like mye.

Oppgave 1

I forbindelse med valg og folkeavstemninger er det i dag vanlig at meningsmålingsinstitutter foretar såkalte "exit polls". Dette utføres ved at et tilfeldig utvalg av de som har avgitt stemme, i det de kommer ut fra stemmelokalet, blir spurt om hva de stemte. I denne oppgaven skal vi regne litt på denne situasjonen for et valg mellom kun to kandidater, som vi skal benevne henholdsvis G og B. Vi skal altså se bort fra at noen kan stemme blankt og at stemmer kan bli forkastet. Vi skal også se bort fra muligheten av at noen ikke vil svare hva de har stemt, eller at noen svarer usant.

La N betegne antall personer som avgir stemme og la p betegne andelen av disse som stemte på G. La videre n betegne antall personer som ble spurt av meningsmålingsinstituttet om sin stemmegivning og la X betegne antall av disse n som stemte på G.

a) Hvilke(n) betingelse(r) må være oppfylt for at X skal være tilnærmet binomisk fordelt ?

Under forutsetning om at X er (tilnærmet) binomisk fordelt, utled sannsynlighetsmaksimeringsestimatoren (SME) for p.

Bestem estimatorens forventning og varians.

I resten av oppgaven skal vi også forutsette at X er (tilnærmet) binomisk fordelt.

Som kjent kan en binomisk fordeling tilnærmes med en normalfordeling dersom n er stor nok. I resten av oppgaven skal vi forutsette at dette er tilfelle slik at vi har at

$$\frac{X - np}{\sqrt{np(1-p)}}$$

er (tilnærmet) standard-normalfordelt.

Anta at meningsmålingen utføres på oppdrag fra et fjernsynsselskap. Dersom et tilstrekkelig stort flertall av de spurte stemte på en av kandidatene, vil fjernesynsselskapet gå ut og erklære denne kandidaten som vinner av valget.

b) Formuler dette som en hypotesetest. Spesifiser nullhypotese og alternativ hypotese og bestem forkastningskriteriet når signifikansnivået er α.

Hva blir konklusjonen på hypotesetesten dersom 5 000 ble spurt, 2562 av disse hadde stemt på G, og en benytter signifikansnivå $\alpha = 0.10$?

Oppgave 2

En viktig vitenskapelig oppdagelse fant sted i 1929 da Edwin Hubble oppdaget at universet er ekspanderende. Hubble's tallmateriale bestod blant annet av; x_i = avstanden til galakse i (målt i millioner lysår), og y_i = hastigheten til galakse i (målt i 1000 km/s). Verdiene Hubble benyttet i en av sine analyser er som følger:

Navn	Avstand, x_i	Hastighet, y_i
Virgo	22	1.2
Pegasus	68	3.8
Perseus	108	5.1
Coma Berenices	137	7.5
Ursa Major 1	255	14.9
Leo	315	19.2
Corona Borealis	390	21.4
Gemini	405	23.0
Bootes	685	39.2
Ursa Major 2	700	41.6
Hydra	1100	60.8

Det oppgis her at $\sum_{i=1}^{11} x_i = 4185$, $\sum_{i=1}^{11} y_i = 237.7$, $\sum_{i=1}^{11} x_i^2 = 2685141$ og $\sum_{i=1}^{11} x_i y_i = 152224$.

Hubble foreslo en modell for hastighet som funksjon av avstand på formen $y = \beta x$, der β senere har blitt kalt Hubble's konstant. En statistisk versjon av ligningen kan gis ved:

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, 11, \tag{2.1}$$

der ε_i, $i = 1, \ldots, 11$, er uavhengige og normalfordelte stokastiske variabler med forventning 0 og varians σ^2.

a) Vi vil i første omgang finne en estimator for β.

Bruk minste kvadraters metode (method of least squares) til å estimere β med utgangspunkt i ligning (2.1), og vis at estimatoren for β da blir gitt ved $\hat{\beta} = \frac{\sum_{i=1}^{11} x_i Y_i}{\sum_{i=1}^{11} x_i^2}$. Regn ut estimatet for β basert på dataene over.

Finn også forventning og varians til $\hat{\beta}$.

b) Anta at en annen galakse befinner seg en avstand $x_0 = 900$ millioner lysår borte.

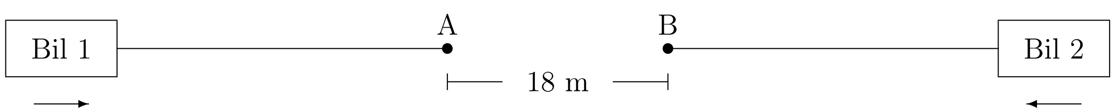
Finn predikert hastighet, $\hat{y}_0$, til denne galaksen.

Utled et 95% prediksjonsintervall for en måling av hastigheten til denne galaksen. Det oppgis at $\sum_{i=1}^{11}(y_i - \hat{y}_i)^2 = 9.87$, der $\hat{y}_i = \hat{\beta} x_i$.

Oppgave 3

Bremselengden Y (målt i meter) for en bil som kjører i en hastighet $10x$ km/time antas å være normalfordelt med forventning θx^2 og standardavvik σx^2. (For en bil som kjører i 40 km/t vil dermed bremselengden være normalfordelt med forventning 16θ og standardavvik 16σ.) Parametrene θ og σ vil avhenge av forsøksbetingelsene, f.eks. bremsenes egenskaper, dekktype, veidekke, vær- og føreforhold, målenøyaktighet. Anta i punkt a) at $\theta = 0.5$ og $\sigma = 0.1$.

a) Anta at to biler, bil I og bil II, kjører mot hverandre i 40 km/t langs den samme rette linje. Ved punktene A og B, henholdsvis, begynner de å bremse. Avstanden mellom punktene A og B antas å være 18 meter (se figur).



Hva er sannsynligheten for at de to bilene treffer hverandre? (Anta at de også under bremsing følger den samme rette linjen.)

I resten av oppgaven antas θ å være en ukjent parameter, mens σ antas å være kjent, $\sigma = 0.1$. Anta at det på en forsøksbane gjøres 8 målinger av bremselengde med den samme bilen, hver gang med hastigheten 40 km/t. Anta videre at målingene er stokastisk uavhengige.

De observerte verdiene er

$$8.4 \quad 7.8 \quad 10.3 \quad 8.7 \quad 10.7 \quad 10.0 \quad 8.8 \quad 12.1$$

b) Tyder målingene på at θ er større enn 0.5? Bruk en test med 5% signifikansnivå.

c) Utled og tegn sannsynligheten for type-II feil for testen i punkt b). Hvor mange observasjoner må en ha for at sannsynligheten for å forkaste nullhypotesen når $\theta = 0.6$ er minst 0.99?

Fasit

1. **a)** $\frac{X}{n}$, θ, $\frac{\theta(1-\theta)}{n}$ **b)** Forkaster H_0

2. **a)** 0.0567 **b)** 51.03, (48.5,53.5)

3. **a)** 0.1894 **b)** Forkast H_0 **c)** Antall observasjoner ≥ 16