Opgaver til Statistik-dag på Aalborg Universitet

1 Den centrale grænseværdisætning

1.1 Uniformfordelte variable

Lad $Z_1, \ldots Z_n$ være uafhængige of uniformt fordelte på intervallet $[-1/2; 1/2]$.

1. Summen af 2 uafhængige bidrag? Lad $U = \sum_{j=1}^{2} Z_j$. Simuler mange gange og tegn histogram. Ligner det en normalfordeling?
2. Summen af 4 uafhængige bidrag? Lad $U = \sum_{j=1}^{4} Z_j$. Simuler mange gange og tegn histogram. Ligner det en normalfordeling?
3. For en stokastisk variabel Z, der er uniformt fordelt $[-1/2; 1/2]$, vis at $\mathbb{E}(Z) = 0$ og $\mathbb{V}ar(Z) = 1/12$.

1.2 Binomialfordelt variabel

Lad X være en binomialfordelt stokastisk variabel, $X \sim \text{bin}(N, \theta)$.

1. Find $\mathbb{E}(X)$ og $\mathbb{V}ar(X)$
2. Argumenter for, at X approximativt er normalfordelt for store N
3. Hvad er den approximative fordeling af $\frac{X}{N}$?
4. Hvad er den approximative fordeling af $\dfrac{\frac{X}{N} - \theta}{\sqrt{\theta(1-\theta)/N}}$?

2 Regression

2.1 Opgave

Vis at LS estimaterne fra slide 18 i "Statistisk modellering og regressionsanalyse" giver residualer med middelværdi 0.

2.2 Opgave

Eksperimenter med app'en "regression", hvor man kan simulere data og inspicere residualerne. Data kan genereres under antagelserne for regression samt situationer hvor data afviger fra modellen.

2.3 Opgave

NB! Denne og den nedenstående opgave omtales også i oplægget "Konfidensintervaller og vurdering af usikkerhed på estimerede størrelser" (samt er hhv. opgaverne 3.4 og 3.5)

Mindste kvadraters estimaterne af α og β kan udregnes som

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (X^\top X)^{-1} X^\top Y,$$

hvor X er $n \times 2$-matricen med søjler $(1, \ldots, 1)^\top$ og $(x_1, \ldots, x_n)^\top$ og $Y = (Y_1, \ldots, Y_n)^\top$. Brug dette til at udregne kovariansmatricen for $(\hat{\alpha}, \hat{\beta})^\top$.

Hint: Kovarians-matricen Σ for en stokastisk vektor Z har indgange $\Sigma_{ij} = \mathbb{C}ov(Z_i, Z_j)$. Hvis A er en matrix af passende dimension, så har AZ kovariansmatrix $A\Sigma A^\top$.

2.4 Opgave

Vis følgende: Middelværdien for x_i, μ_i, estimeres ved $\hat{\mu}_i = \hat{\alpha} + \hat{\beta} x_i$. Den tilhørende varians er givet ved

$$\mathbb{V}ar(\hat{\mu}_i) = \mathbb{V}ar(\hat{\alpha}) + x_i^2 \mathbb{V}ar(\hat{\beta}) + 2x_i \mathbb{C}ov(\hat{\alpha}, \hat{\beta}),$$

hvor

$$\mathbb{C}ov(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \frac{\bar{x}_.}{\sum_i (x_i - \bar{x}_.)^2}$$

3 Usikkerhed og konfidensinterval

Slide numre henviser til oplægget "Konfidensintervaller og vurdering af usikkerhed på estimerede størrelser"

3.1 Opgave

Check resultaterne på slide 8 når $\approx$ er erstattet med $=$ og X_i'erne antages at være normalfordelte.

3.2 Opgave

Vedr. slide 9: Hvad gør landmåleren hvis sandsynligheden ikke er under 99.9%?

3.3 Opgave

Check (jf slide 20) at

3.3.1 Opgave

$$\mathbb{V}ar(l_i) = \theta(1-\theta).$$

Hint: brug slide 39.

3.3.2 Opgave

$$\hat{\theta}(1-\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(l_i - \bar{l}_.)^2.$$

Dvs variansestimatet er $(n-1)/n$ gange det sædvanlige variansestimat

3.4 Opgave

Mindste kvadraters estimaterne af α og β kan udregnes som

$$\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} = (X^\top X)^{-1}X^\top Y,$$

hvor X er $n \times 2$-matricen med søjler $(1,\ldots,1)^\top$ og $(x_1,\ldots,x_n)^\top$ og $Y = (Y_1,\ldots,Y_n)^\top$. Brug dette til at udregne kovariansmatricen for $(\hat{\alpha},\hat{\beta})^\top$.

Hint: Kovarians-matricen Σ for en stokastisk vektor Z har indgange $\Sigma_{ij} = \mathbb{C}ov(Z_i, Z_j)$. Hvis A er en matrix af passende dimension, så har AZ kovariansmatrix $A\Sigma A^\top$.

3.5 Opgave

Vis (1) på slide 25 ved at bruge resultatet og hintet fra forrige opgave.

3.6 Opgave

(slide 33) Vis at $[g(L); g(U)]$ er et α-konfidensinterval for η når g er voksende.

3.7 Opgave

Check resultaterne på slide 35-37

4 Hypotese-test og p-værdier

4.1 Normalitet

Vis at estimatet for en andel i et binomialforsøg er approksimativt normalfordelt samt redegør for hvad i normalfordelingen er.

4.2 Approksimativt konfidensinterval

I et studie om kræft i spiserøret (Frankrig, 1980), var der i en bestemt gruppe (baseret på alder samt rygning- og alkoholdindtag) konstateret kræft for 1 person og 14 personer uden konstateret kræft.

Konstruer et konfidensinterval for populationsandelen af kræft baseret på disse oplyste tal samt normalfordelingsapproksimationen (se eventuelt formelsamlingen). Hvordan skal dette tolkes?

4.3 Eksperimenter med `binom_test.{mw, xlxs, tns}`

Benyt en eller flere af filerne `binom_test.{mw, xlxs, tns}` til at lave simuilationsbaserede hypotesetest for binomial forsøg.