Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

# TMA4240 Statistikk
Høst 2018

**Anbefalt øving 11**

Oppgavene i denne øvingen dreier seg om hypotesetesting og sentrale begreper som nullhypotese og alternativ hypotese, testobservator, forkastning og akseptanse, og kritisk verdi.

### Oppgave 1

La $X_1, \ldots, X_5$ være uavhengige og normalfordelt med ukjent forventningsverdi $\mu$ og ukjent varians $\sigma^2$. Observerte verdier for $X_1, \ldots, X_5$ er:

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x_i$ | 5.00 | 4.48 | 5.60 | 4.25 | 5.44 |

**a)** Angi rimelige estimatorer for forventningsverdien $\mu$ og for variansen $\sigma^2$.

Hva blir estimatene med dataene gitt over ?

**b)** Utled et 95% konfidensintervall for $\mu$ basert på observasjonene over.

Forklar kort hvordan en fra dette konfidensintervallet kan avgjøre konklusjonen i en hypotesetest (signifikansnivå 5%) med alternativ hypotese $H_1 : \mu \neq \mu_0$ for ulike verdier for $\mu_0$.

### Oppgave 2

Vi skal i denne oppgaven se for oss at vi kaster en mynt flere ganger. Mynten har 0.5 sannsynlighet for utfall ≪mynt≫ og 0.5 sannsynlighet for ≪kron≫. Vi antar at utfallene av ulike myntkast er uavhengige.

**a)** Anta at vi kaster mynten 5 ganger.

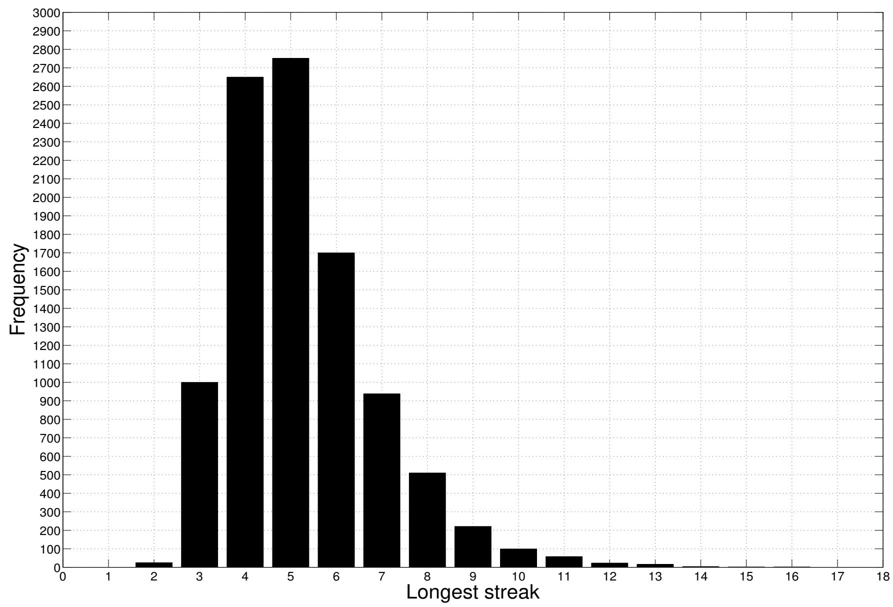
Hva er sannsynligheten for å få 5 ≪kron≫?

Hva er sannsynligheten for å få 3 ≪kron≫?

Hva er sannsynligheten for å få minst 4 ≪kron≫ på rad, det vil si en sekvens (*streak*) av bare ≪kron≫ utfall som minst er av lengde 4?

**b)** Vi kaster mynten 30 ganger. Fordelingen til lengste sekvens med ≪kron≫ er vanskelig å regne ut. I stedet kan vi få en datamaskin til å generere 30 stokastiske og uavhengige myntkast. Vi registrerer lengste sekvens av ≪kron≫. Prosedyren gjentas $B$ ganger, og resultatet er representativt for fordelingen for lengste sekvens av ≪kron≫, se Figur 1.

Anslå sannsynligheten for å få en lengste sekvens på 5 eller 6.




Figur 1: Figuren viser et histogram av sekvenser. Disse er resultat av $B = 10000$ gjentak av 30 myntkast. Høyden på stolpene angir i hvor mange av de 10000 forsøkene lengste sekvens av ≪kron≫ var en bestemt lengde

Miriam har fått hjemmelekse å kaste en mynt 30 ganger. Resultatet er som følger, der 0 betyr ≪mynt≫ og 1 betyr ≪kron≫:
$(0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1)$.

Læreren mistenker at Miriam har jukset og bare funnet på tallene istedenfor å faktisk kaste en mynt, og læreren vil undersøke dette. Formuler dette som en hypotesetest om lengste sekvens av ≪kron≫. Bruk histogrammet i Figur 1 til å svare.

## Oppgave 3

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

?

Gitt at denne betingelsen er oppfylt, bestem følgende sannsynligheter for $n = 20$ og $p = 0.50$:

$$\mathrm{P}(X = 9) \quad , \quad \mathrm{P}(X > 9) \quad \text{og} \quad \mathrm{P}(X > 9 | X \leq 12)$$

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

**b)** Utled sannsynlighetsmaksimeringsestimatoren (SME) for $p$.

Bestem estimatorens forventning og varians.

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.

**c)** Formuler dette som en hypotesetest. Spesifiser nullhypotese og alternativ hypotese og bestem forkastningskriteriet når signifikansnivået er $\alpha$.

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 4

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 |

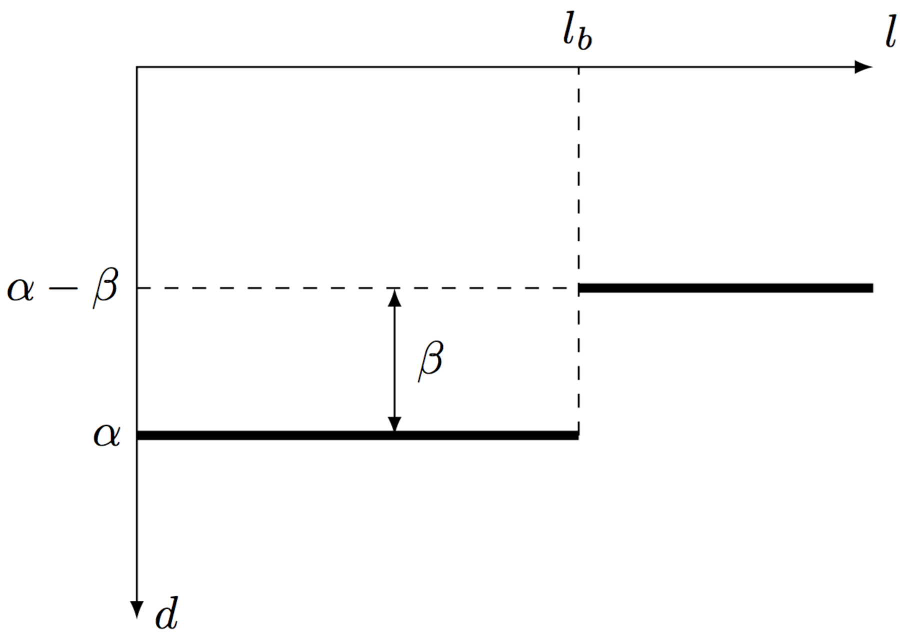
hvor $l$ er lengden langs vannledningen fra en gitt kum og $d$ er dypet til ledningen. Det er altså åtte observasjoner $(d_i, l_i)$; $i = 1, \ldots, 8$. Ved utregning får en at $\sum_{i=1}^{8} d_i = 22.44$ og $\sum_{i=1}^{8} d_i^2 = 63.0162$. Anta først følgende modell for observasjonene:

$$D_i = \alpha + \epsilon_i$$

der $\alpha$ er det konstante dyp for vannledningen og $\epsilon_i$; $i = 1, \ldots, 8$ er uavhengige normalfordelte variable med kjent forventning 0 og ukjent varians $\sigma_1^2$.

En rimelig estimator for dypet til ledningen er

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




Figur 2: Vannledning med opprinnelig dybde $\alpha$, men hevet for $l > l_b$ grunnet et vertikalt brudd i posisjon $l_b$.

**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 $\beta$ (se figur 2).

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 $\alpha$ og $\beta$ som er

$$\hat{\alpha} = \frac{1}{5} \sum_{i=1}^{5} D_i$$

$$\hat{\beta} = \frac{1}{5} \sum_{i=1}^{5} D_i - \frac{1}{3} \sum_{i=6}^{8} D_i$$

Vis videre at $\hat{\alpha}$ og $\hat{\beta}$ er forventningsrette.
Finn estimatene for $\alpha$ og $\beta$ 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.

# Fasit

**1**. **b**) $[4.23, 5.68]$

**2**. **a**) 0.031,0.313,0.09 **b**) 0.44, forkast $H_0$

**3**. **a**) 0.1602, 0.588, 0.525 **b**) $\frac{X}{n}$, $\theta$, $\frac{\theta(1-\theta)}{n}$ **c**) Forkaster $H_0$

**4**. **a**) $\mathrm{E}(\widehat{\alpha}) = \alpha$, $\mathrm{Var}(\widehat{\alpha}) = \frac{\sigma_1^2}{8}$ **b**) $[2.72, 2.89]$ **c**) $\widehat{\alpha} = 2.852$, $\widehat{\beta} = 0.125$ **d**) Forkast $H_0$