

TMA4245 – Parallell 1

Geir-Arne Fuglstad, IMF

17. mars 2025

Viktig beskjed – endring angående gule ark

Endring:

IMF kommer ikke lenger til å ha gule og stemplede A5-ark ved instituttkontoret. Dere må selv ordne et A5-ark for å ta med på eksamen, og det trenger ikke å være gult eller stemplet.

Tillatte hjelpemidler på deres eksamen blir dermed:

- Tabeller og formler i statistikk (Akademika).
- Bestemt enkel kalkulator i henhold til NTNUs generelle reglement.
- Ett A5-ark ($148\text{mm} \times 210\text{mm}$) med egne formler og notater. (Du må selv ordne et A5-ark, for eksempel ved å dele et A4-ark i to like store deler som beskrevet i [Store norske leksikon](#)).

Egenskaper til hypotesetester

Anta vi gjør et tilfeldig utvalg fra en populasjon med parameter θ . Basert på det tilfeldige utvalget gjør vi en hypotesetest av $H_0: \theta = \theta_0$ mot en alternativ hypotese H_1 .

Da har vi:

- **Signifikansnivået** til hypotesetesten,
$$\alpha = P(\text{"Type I feil"}) = P(\text{"Forkast } H_0" | "H_0 \text{ sann"}).$$
- **Sannsynlighet for Type II feil** for en θ_1 i H_1 ,
$$\begin{aligned}\beta(\theta_1) &= P(\text{"Type II feil"} | \theta = \theta_1) \\ &= P(\text{"Ikke forkast } H_0" | \theta = \theta_1)\end{aligned}$$
- **Teststyrke** for en θ_1 i H_1 ,
$$\begin{aligned}\text{Teststyrke}(\theta_1) &= P(\text{"Forkast } H_0" | \theta = \theta_1) \\ &= 1 - \beta(\theta_1)\end{aligned}$$

Eksempel 1: Hypotesetest for varians

En maskin produserer sjokoladeplater med forventningsverdi $\mu = 200$ og en ukjent varians σ^2 . Vi tror at $\sigma^2 = \sigma_0^2 = 4$, men mistenker at variansen har økt. For å undersøke dette gjør vi et tilfeldig utvalg av størrelse n . Vi er villige til å anta at sjokoladeplatene utgjør en normalpopulasjon.

- a) Formuler en hypotesetest og bestem en beslutningsregel, ved signifikansnivå $\alpha = 0.05$, som funksjon av n .
- b) Bestem et uttrykk for teststyrke som funksjon av σ^2 og n , og plott som funksjon av σ^2 når $n = 20$.
- c) Vi ønsker at teststyrken skal være minst 95% når $\sigma^2 = 10$. Bestem minste mulige utvalgsstørrelse n .

Ikke-parametrisk statistikk

Parameterisk statistikk:

Metoder basert på modeller/fordelinger beskrevet av parametere.

Ikke-parametrisk statistikk:

Metoder som ikke krever antagelser om bestemte fordelinger.

Hvis det kan rettferdiggjøres ønsker vi å bruke parametriske metoder. Hvis det ikke kan rettferdiggjøres må vi bruke ikke-parametriske metoder.

Eksempel 2: Enkel demonstrasjon

Anta at vi har et tilfeldig utvalg $X_1, \dots, X_{15}$ fra en fordeling med ukjent forventnings verdi μ og ukjent varians σ^2 . Vi ønsker å teste $H_0: \mu = 200$ mot $H_1: \mu < 200$ ved signifikansnivå 0.05.

Under antagelse om normalfordeling er beslutningsregelen:

$$\text{“Forkast } H_0 \text{ hvis } \bar{x} \leq 200 - \frac{t_{0.05,14}}{\sqrt{15}} s \text{”}$$

Hva blir signifikansnivået hvis beslutningsregelen brukes når $X_1, \dots, X_{15} \sim_{\text{iid}} \text{Exp}(200)$?

NB: Her er H_0 sann, men antagelsen om normalfordeling er feil.

Fortegnstest

Antagelser:

- $X_1, \dots, X_n \sim \text{iid } f(x)$
- Observerer $x_1, \dots, x_n$

Mål:

La $\tilde{\mu}$ betegne medianen til $f(x)$. Da ønsker vi å teste

$$H_0: \tilde{\mu} = \tilde{\mu}_0 \quad \text{mot} \quad H_1: \tilde{\mu} > \tilde{\mu}_0$$

eller eventuelt $H_1: \tilde{\mu} < \tilde{\mu}_0$ eller $H_1: \tilde{\mu} \neq \tilde{\mu}_0$.

Prosedyre:

1. Fjern alle verdier eksakt lik $\tilde{\mu}_0$ og la $u = \sum_{i=1}^m \mathbb{I}(x_i > \tilde{\mu}_0)$ der m er antall målinger ulik $\tilde{\mu}_0$.
2. Regn ut p-verdi basert på $U \sim \text{Binom}(m, 0.5)$
3. Sammenlign med ønsket signifikansnivå og konkluder

Eksempel 3: Levetid til lyspærer

Vi tror at median levetid til lyspærene vi produserer er 1000 timer, men mistenker at median levetid har endret seg grunnet endringer i produksjonsprosessen. Vi gjør et tilfeldig utvalg av $n = 15$ lyspærer og måler:

773	13	375	901	2138
2238	49	579	33	1700
1193	672	3078	384	62

- a) Formuler nullhypotese og alternativ hypotese, og utfør en fortegnstest.

To-utvalg permutasjonstest

Antagelser: To grupper av individer: $x_1, \dots, x_n$ (gruppe A), og $y_1, \dots, y_m$ (gruppe B)

Mål: Teste

H_0 : "A og B like egenskaper" mot H_1 : "A og B forskjellige egenskaper"

En-sidig prosedyre (basert på gjennomsnitt):

1. Samle alle målingene $\{x_1, \dots, x_n, y_1, \dots, y_m\}$
2. For $b = 1 \dots B$
 1. Trekk n målinger tilfeldig uten tilbakelegging. Dette er de nye $x_1^{(b)}, \dots, x_n^{(b)}$.
 2. De gjenstående m målingene er $y_1^{(b)}, \dots, y_m^{(b)}$
 3. Regn ut $u^{(b)} = \overline{x^{(b)}} - \overline{y^{(b)}}$
3. Estimer $p = P(U \geq u)$, der u er observert verdi, med
$$\hat{p} = \frac{1}{B+1} \left(1 + \sum_{b=1}^B \mathbb{I}(u^{(b)} \geq u) \right)$$
4. Sammenlign med ønsket signifikansnivå og konkluder

Eksempel 4: Effekt av høydehus

Høydehus	Ikke høydehus
16.3	15.0
15.4	14.3
14.2	14.9
13.7	13.2
14.8	12.2
12.9	13.1
	12.8

Vi ønsker å sjekke om høydehus har en effekt på blodprosent på (tilsynelatende) like idrettsutøvere. Vi har gjort målinger for 13 utøvere som har brukt eller ikke har brukt høydehus.

- a) Formuler hypoteser og beskriv beslutningsregel ved signifikansnivå $\alpha = 0.05$.
- b) Estimer p-verdi.
- c) Hva blir konklusjonen?