Kapittel 10: Hypotesetesting

TMA4245 Statistikk

10.1, 10.2, 10.3: Introduksjon, 10.5, 10.6, 10.7: Test for μ i normalfordeling, 10.4: p-verdi

Estimering og hypotesetesting

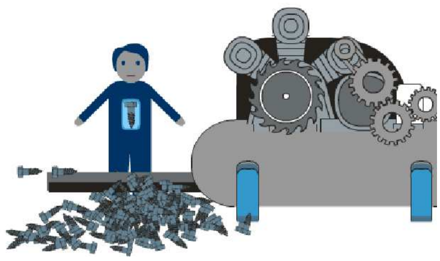
Situasjon:

	Eks 1: Studentar si oppfatning av køyreeigenskapar	**Eks 2: Høgda til studentar**
Problemstilling	Synest studentar dei er flinkare enn gj.snittet til å køyre bil?	Kor høge er studentane?
Populasjon	Alle studentar, (evt. menn og kvinner som to pop.)	Alle studentar, (evt. menn og kvinner som to pop.)
Vil observere	Om dei synest dei er "bedre enn gj.snittet" eller ikkje	Høgda
Fordeling	Binomisk	Normal
Parameter θ	Andelen p som synest dei er flinkare enn gj.snittet	Forventa høgde, μ.
Utvalg, storleik n	Alle studentar som svarte på spørreundersøking	Alle studentar som svarte på spørreundersøking
Data: observert verdi av $X_1, \ldots, X_n$ u.i.f. SV	"Bedre enn gj.snittet" eller ikkje. $X = X_1 + \cdots + X_n$ = antall "bedre"	Høgda, $X_1, \ldots, X_n$

Estimering og hypotesetesting (forts)

	Eks 1: Studentar si oppfatning av køyreeigenskapar	**Eks 2: Høgda til studentar**
Spørsmål, estimering:	Kor stor andel av studentane synest dei er flinkare enn gj.snittet til å køyre bil?	Kva er forventa høgde for ein student?
Estimator	$\hat{p} = \frac{X}{n}$, andel "betre" blant dei n spurte	$\hat{\mu} = \bar{X}$, gjennomsnittleg høgde i utvalget.
Storleik med kjent fordeling	$Z = \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx \mathcal{N}(0,1)$ for stor n, og p ikkje nær 0 eller 1.	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ er t-fordelt med $n-1$ fridomsgrader.
Konfidensintervall:	$[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$	$[\bar{x} - t_{\alpha/2,(n-1))}\frac{s}{\sqrt{n}}$, $\bar{x} + t_{\alpha/2,(n-1))}\frac{s}{\sqrt{n}}]$
Spørsmål, hypotesetesting:	Synest fleire enn 50% av studentene at dei er betre enn gj.snittet til å køyre bil? Trur fleire menn enn kvinner at dei er betre enn gj.snittet?	Er forventa høgde på kvinnelige studentar mindre enn 170 cm?

Eks: Kvalitetskontroll av skruar



- Ser på produksjon av skruar, der lengda på ein skrue skal vere 15 mm.
- I praksis er det små variasjonar, og lengda på skruane har standardavvik $\sigma = 0.1$ mm. Skruane vert antatt å følgje spesifikasjonen dersom forventa lengde er 15 mm.
- Tar jamnlig stikkprøver frå prosessen, for å sjekke om skruane svarer til spesifikasjonen.
- Dersom stikkprøva tyder på at dei produserte skruane ikkje er 15 mm , må maskina som lagar skruane kalibrerast på nytt.
- Vil bruke hypotesetesting basert på stikkprøva til å avgjere om maskina må kalibrerast.

Hypotese

DEF 10.1: Ein statistisk hypotese er ein antakelse eller påstand om eigenskapar ved ein eller fleire populasjonar.

Nullhypotese (H_0): Hypotesen vi vil undersøke om vi har grunnlag frå data for å forkaste. Ein bestemt verdi for ein parameter.

Alternativ hypotese (H_1): Hypotesen vi aksepterer dersom vi forkastar nullhypotesen. Ofte meir enn ein verdi for ein parameter.

Eks:	Bilkøyring:	$H_0 : p_1 - p_2 = 0,$	$H_1 : p_1 - p_2 > 0$
	Høgde:	$H_0 : \mu = 170,$	$H_1 : \mu < 170$
	Skruar:	$H_0 : \mu = 15,$	$H_1 : \mu \neq 15$

Statistisk hypotesetesting

- Undersøke om dataene gir tilstrekkelig "bevis" for at den alternative hypotesen er sann.
- Sml. rettssak: Tiltalte er uskuldig inntil det motsette er bevist.

To typer testar:

To-sidig test: $H_0 : \theta = \theta_0$ mot $H_1 : \theta \neq \theta_0$

Ein-sidig test:

- $H_0 : \theta = \theta_0$ mot $H_1 : \theta < \theta_0$, eller
- $H_0 : \theta = \theta_0$ mot $H_1 : \theta > \theta_0$

To typer feil

Type I-feil: Forkasting av nullhypotesen når denne er sann (DEF 10.2)

- Eks, skruar: Type I-feil kan føre til unødvendig produksjonsstans. Vil ikkje stoppe produksjonen for å kalibrere utan at vi er sikre på at skruane ikkje er 15 mm.

Type II-feil: Ikkje forkaste nullhypotesen når denne er usann (DEF 10.3).

- Eks, skruar: Type-II-feil kan føre til mangel på nødvendig kalibrering. noko som er ugunstig overfor kundar. For lange og for korte skruar påfører kjøpar problem.

Signifikansnivået for ein test = $P(\text{Type I-feil}) = \alpha$

Type-I og type-II-feil

	H_0 sann	H_0 falsk
Aksepter H_0	Korrekt	Type-II feil
Forkast H_0	Type-I feil	Korrekt

Analogi, rettssak:

	H_0 sann	H_0 falsk
Aksepter H_0	Uskyldig, frikjent	Skyldig, går fri
Forkast H_0	Uskyldig, dømt	Skyldig, dømt

Eitt utvalg: to-sidig test for μ med σ kjent

La $X_1, X_2, ..., X_n$ vere u.i.f., $X_i \sim \mathcal{N}(\mu, \sigma^2)$, der σ er kjent.

1. Sett opp **hypoteser:** $H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu \neq \mu_0$
2. Bestem **signifikansnivå** α.
3. - Velg **testobservator**: $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

     Under H_0 er
     $$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$
   - Finn **forkastningsområde**: Forkast H_0 dersom $|z_0| > z_{\frac{\alpha}{2}}$, der z_0 er observert verdi for Z_0.
4. Berekn $\bar{x}$ frå utvalget, og vidare observert verdi for Z_0: $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

   Samanlikn z_0 med $z_{\frac{\alpha}{2}}$, og forkast H_0 dersom $|z_0| > z_{\frac{\alpha}{2}}$.

Eks, test for μ, σ kjent: Skruar

Vil undersøke om det er grunn til å tru at skruane som blir produserte ikkje er 15 mm lange i forventning.

	Generell framgangsmåte	**Kvalitetskontroll av skruar**
Utvalg:	$X_1, X_2, ..., X_n$ uavhengige, med $X_i \sim \mathcal{N}(\mu, \sigma^2)$, der σ er kjent.	Stikkprøve med $n = 10$ skruar, antar normalfordeling, kjent $\sigma = 0.1$mm.
1.	Hypoteser: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$	$H_0 : \mu = 15$ vs. $H_1 : \mu \neq 15$
2.	Bestem signifikansnivå α	Velger $\alpha = 0.05$
3.	Testobservator $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ under H_0 Forkastingsområde: Forkast H_0 dersom $z_0 > z_{\frac{\alpha}{2}}$ eller $z_0 < -z_{\frac{\alpha}{2}}$.	
4.	Observerer $\bar{x}$ frå utvalget Bereknar $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ Finn $z_{\frac{\alpha}{2}}$ Samanliknar $-z_{\frac{\alpha}{2}}$, z_0 og $z_{\frac{\alpha}{2}}$ Forkast H_0 og konkluder med H_1, eller hald på H_0.	$\bar{x} = 15.05$ mm. $z_0 = \frac{15.05-15}{0.1/\sqrt{10}} = 1.58$ $z_{\frac{0.05}{2}} = z_{0.025} = 1.96$ -1.96<1.58<1.96 Forkastar ikkje H_0. Har ikkje sterke nok bevis for at $\mu \neq 15$ mm.

Hypotesetesting, generell prosedyre

1. Spesifi sér nullhypotese og alternativ **hypotese**.
2. Bestem **signifikansnivå** α.
3. - Velg passande **testobservator**, med kjent fordeling, og
   - etablér **forkastningsområde** (kritisk område) basert på denne og α.
4. Berekn observert verdi av testobservatoren. Forkast H_0 dersom denne verdien ligg i forkastningsområdet, og behald H_0 dersom ikkje.
5. Trekk konklusjonar.

10.4: p-verdi

DEF 10.5: Ein p-verdi er det lavaste (signifi kans-)nivået som gjer at den observerte verdien for testobservatoren er signifi kant.

Alternativ def: Ein p-verdi er sannsynet for å få det vi har observert eller noko meir ekstremt, gitt at H_0 er sann.

Generell prosedyre, hypotesetest ved bruk av p-verdi:

- Bestem null- og alternativ hypotese.
- Velg passande testobservator.
- Berekn p-verdien basert på observert verdi av testobservatoren.
- Bestem om vi vil forkaste eller behalde nullhypotesen på grunnlag av p-verdien og kunnskap om systemet.

10.6: Hypotesetest og konfidensintervall

Anta $X_1, X_2, ..., X_n$ u.i.f., $X_i \sim \mathcal{N}(\mu, \sigma^2)$ der σ er kjent.

To-sidig test:

- $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$
- Signifikansnivå α.
- Testobservator $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ under H_0
- Forkastingsområde: Forkast H_0 dersom $z_0 > z_{\frac{\alpha}{2}}$ eller $z_0 < -z_{\frac{\alpha}{2}}$

Konfidensintervall for μ: Under H_0 har vi

$$P(-z_{\frac{\alpha}{2}} < Z_0 < z_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- Dersom eit $(1-\alpha)100\%$ konfidensintervall inneheld μ_0 vil vi ved hypotesetesting ikkje forkaste H_0 på nivå α.
- Dersom eit $(1-\alpha)100\%$ konfidensintervall ikkje inneheld μ_0 vil vi ved hypotesetesting forkaste H_0 på nivå α.

10.7: Eitt utvalg, to-sidig test for μ med σ ukjent

La $X_1, X_2, ..., X_n$ vere u.i.f., $X_i \sim \mathcal{N}(\mu, \sigma^2)$, der σ er ukjent.

Estimator for σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

1. Sett opp **hypoteser:** $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$
2. Bestem **signifikansnivå** α.
3. - **Testobservator**: $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}}$

     Under H_0 er $T_0 = \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$ t-fordelt med $n-1$ fridomsgrader.
   - **Forkastingsområde**: Forkast H_0 dersom $|t_0| > t_{\frac{\alpha}{2},n-1}$, der t_0 er observert verdi for T_0.
4. Berekn $\bar{x}$ og s frå utvalget, og vidare observert verdi for T_0:
   $t_0 = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$.
   Samanlikn t_0 med $t_{\frac{\alpha}{2},n-1}$, og forkast H_0 dersom $|t_0| > t_{\frac{\alpha}{2},n-1}$.

Rep: Viktige begrep, hypotesetesting

Hypoteser:

- **Nullhypotese (H_0):** Hypotesen vi vil undersøke om vi har grunnlag frå data for å forkaste. Ein bestemt verdi for ein parameter.
- **Alternativ hypotese (H_1):** Hypotesen vi aksepterer dersom vi forkastar nullhypotesen. Ofte meir enn ein verdi for ein parameter.

Statistisk hypotesetesting: Undersøke om dataene gir tilstrekkelig "bevis" for at den alternative hypotesen er sann.

To typer testar:

- **To-sidig test:** $H_0 : \theta = \theta_0 \quad \text{mot} \quad H_1 : \theta \neq \theta_0$
- **Ein-sidig test:**
  - $H_0 : \theta = \theta_0 \quad \text{mot} \quad H_1 : \theta < \theta_0$, eller
  - $H_0 : \theta = \theta_0 \quad \text{mot} \quad H_1 : \theta > \theta_0$

To typer feil: Type-I og type-II-feil.

	H_0 sann	H_0 falsk
Aksepter H_0	Korrekt	Type-II feil
Forkast H_0	Type-I feil	Korrekt

Signifi kansnivå og teststyrke

- Defi nerer

$$\begin{aligned} \alpha &= P(\text{Type I-feil}) \\ \beta &= P(\text{Type II-feil}) \end{aligned}$$

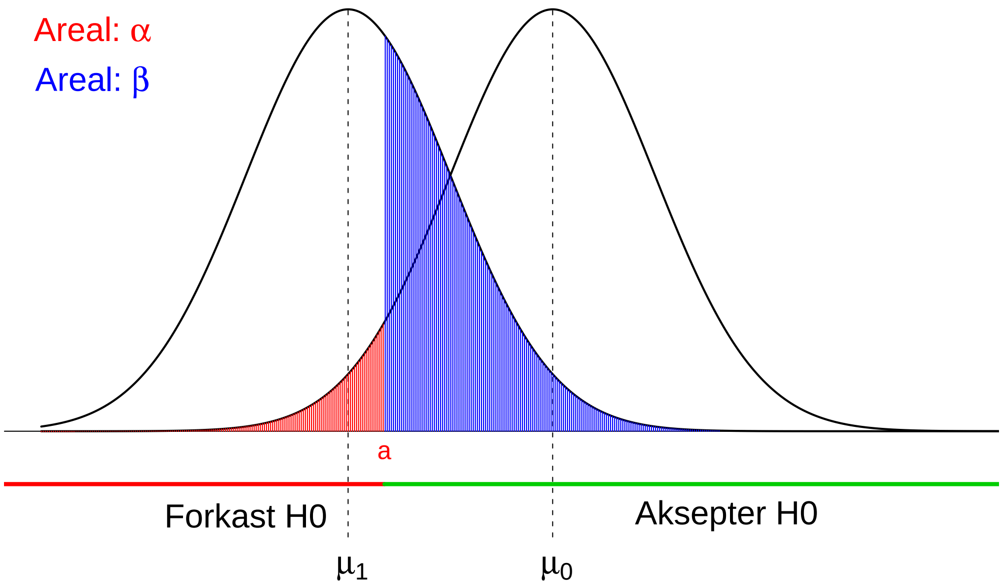
- **Signifi kansnivået** for ein test $= P(\text{Type I-feil}) = \alpha$.
- **Styrken** for ein test er sannsynet for å forkaste H_0 når *eit bestemt alternativ* er sant (DEF 10.4), dvs.
  Styrken $= 1 - P(\text{Type II-feil, bestemt alternativ}) = 1 - \beta$.

Har at

- Reduserer $\alpha \Rightarrow \beta$ aukar og $1 - \beta$ (styrken) minkar (og omvendt).
- Aukar $n \Rightarrow \alpha$ minkar og $1 - \beta$ (styrken) aukar.

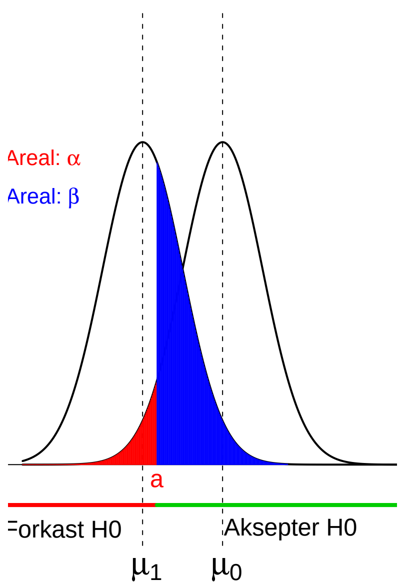
Teststyrke, illustrasjon

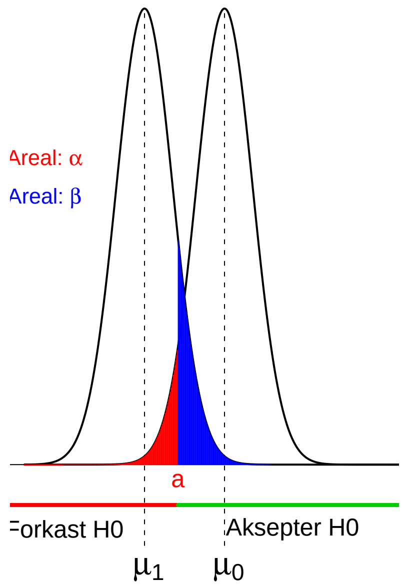
- Testar hypotesen $H_0 : \mu = \mu_0 \quad$ mot $\quad H_1 : \mu < \mu_0$.
- Forkastar H_0 dersom $z_0 < -z_\alpha$, eller ekvivalent $\bar{x} < a$, der $a = \mu_0 - z_\alpha \sigma/\sqrt{n}$.
- Anta sann verdi $\mu = \mu_1$, kva er $\beta = P(\text{Type II-feil}, \mu = \mu_1)$ og teststyrken $1 - \beta$?



Teststyrke, illustrasjon (forts)

- Testar hypotesen $H_0 : \mu = \mu_0 \quad$ mot $\quad H_1 : \mu < \mu_0$.
- Forkastar H_0 dersom $z_0 < -z_\alpha$, eller ekvivalent $\bar{x} < a$, der $a = \mu_0 - z_\alpha \sigma/\sqrt{n}$.
- La $\alpha = 0.05$. Teststyrke $1 - \beta$ for $\mu = \mu_1$ ($n_{\text{høgre}} = 2n_{\text{venstre}}$):





Valg av utvalgsstorleik

- Einsidig test for μ, σ kjent.
  - $H_0 : \mu = \mu_0$ mot $H_1 : \mu > \mu_0$ (eller $H_1 : \mu < \mu_0$)
  - Dersom vi vil ha sannsyn $(1-\beta)$ for å oppdage $\mu = \mu_0 + \delta$ (for gitt δ), og ønskjer signifikansnivå α, må storleiken n på utvalget vere minst

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{\delta^2}$$

- Tosidig test for μ, σ kjent.
  - $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$
  - Med tilsvarande situasjon som over, blir minste utvalgsstorleik (tilnærma)

$$n = \frac{(z_{\frac{\alpha}{2}} + z_\beta)^2\sigma^2}{\delta^2}$$