NTNU
Det skapende universitet

Kapittel 10

Ett- og toutvalgs hypotesetesting

TMA4240 H2006: Eirik Mo

Estimering og hypotesetesting

Fenomen	Bilkjøring	Høyden til studenter
Spørsmål	Hvor stor andel av studentene synes de er flinkere enn gjennomsnittet til å kjøre bil?	Hvor høye er studentene?
Populasjon	Alle studenter, eller evt. menn og kvinner som to populasjoner.	Alle studenter, eller evt. menn og kvinner som to populasjoner.
Parameter	Andelen p som synes de er flinkere enn gjennomsnittet.	Forventet høyde, μ.
Utvalg	Alle studenter som svarte på spørreundersøkelse.	Alle studenter som svarte på spørreundersøkelse.
Data, u.i.f og representative?	Flinkere eller ikke enn gjennomsnittet.	Høyden.

Estimering og hypotesetesting

Fenomen	**Bilkjøring**	**Høyden til studenter**
Estimator	$\hat{p} = \frac{X}{n}$, X antall som synes de er flinkere enn gjennomsnittet av n spurte.	$\hat{\mu} = \bar{X}$ gjennomsnittlig høyde.
Størrelse med kjent fordeling	For store n, og p ikke for nært 0 eller 1, så er $Z = \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}}$ *tilnærmet* normalfordelt	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ t-fordelt med $n-1$ frihetsgrader.

Estimering og hypotesetesting

Fenomen	**Bilkjøring**	**Høyden til studenter**
Kvantiler i fordeling	$-z_{\alpha/2}$ og $z_{\alpha/2}$	$-t_{\alpha/2,(n-1)}$ og $t_{\alpha/2,(n-1)}$
Intervall	$[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$ $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$	$[\bar{X} - t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}},$ $\bar{X} + t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}}]$
Hypotesetesting:	Er andelen av studenter som synes de er flinkere enn gjennomsnittet til å kjøre bil større enn 0.5? Tror flere menn enn kvinner at de er gode sjåfører?	Er dette årets mannlige studenter høyerer enn gjennomsnittet for værnepliktige, 179.8cm? Er byggstudenter høyere enn studenter fra marin?

Hypotese

DEF 10.1: En statistisk hypotese er en antakelse eller påstand om egenskaper ved en eller fl ere populasjoner.

Nullhypotese: Hypotesen vi vil undersøke om vi har grunnlag fra data for å forkaste. En bestemt verdi for en parameter.

Alternativ hypotese: Hvis vi forkaster nullhypotesen så aksepterer vi den alternative hypotesen. Ofte mer enn en verdi for en parameter (større enn, mindre enn og ulik).

Kvalitetskontroll av skruer

— Produksjon av skruer.

— Lengden på produsert skrue skal være 15 mm.

— Tar jevnlig stikkprøve fra prosessen, for å sjekke om skruene som produseres er 15 mm lange.

— Hvis stikkprøven tyder på at de produserte skruene ikke er 15 mm, må maskinen som lager skruene kalibreres på nytt.

— Hvilken nullhypotese og alternativ hypotese vil vi undersøke?

Hypotesetesting og rettsak

— Spørsmål: Er grunn til å tro at skruene som produseres ikke er 15 mm lange?

— Statistisk hypotesetesting: Undersøke om det er nok bevis som underbygger at skruene ikke er 15 mm lange. Som i rettssak: tiltalte er antatt uskyldig til han er bevist skyldig.

— Null hypotesen: skruene som produseres er 15 mm.

— Alternativ hypotese: skuene som produseres er ikke 15 mm.

$$H_0 : \mu = 15\text{mm} \quad \text{vs.} \quad H_1 : \mu \neq 15\text{mm}$$

To typer feil

— DEF 10.2: Forkasting av nullhypotesen når denne er sann, kalles en type-I-feil.

- Vi vil være sikre på at skruene ikke er 15 mm før vi bestemmer oss for å stoppe produksjonen for å kalibrere. Produksjonsstopp for kalibrering av maskin gjør at produsenten taper penger pga. forsinket produksjon.

— DEF 10.3: Å ikke forkaste nullhypotesen når den er gal, kalles en type-II-feil.

- Vi vil gjerne kalibrere maskinen på nytt hvis skruene som produseres ikke er 15 mm. For lange og for korte skruer påfører kjøper problemer.

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

Ett utvalg: test for μ med σ kjent

	Generell fremgangsmåte	Kvalitetskontroll av skruer
0	$X_1, X_2, ..., X_n$ u.i.f. normal(μ, σ) der σ er kjent.	Stikkprøve (utvalg) av $n = 10$ skruer, antar normalfordeling og kjenner $\sigma =$0.1mm.
1	To-sidig test $H_0 : \mu = \mu_0 \quad vs. \quad H_1 : \mu \neq \mu_0$	Er grunn til å tro at skruene som produseres ikke er 15 mm lange? $H_0 : \mu = 15 \quad vs. \quad H_1 : \mu \neq 15$
2	Signifikansnivå α bestemmes.	Velger $\alpha = 0.05$
3	Testobservator $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ er under H_0 standard normalfordelt Forkast H_0 hvis $z_0 > z_{\frac{\alpha}{2}}$ eller $z_0 < -z_{\frac{\alpha}{2}}$.	

Ett utvalg: test for μ med σ kjent

	Generell fremgangsmåte	Kvalitetskontroll av skruer
4	$z_{\frac{\alpha}{2}}$	$z_{\frac{0.05}{2}} = 1.96$
	Observerer $\bar{x}$ fra utvalget (stikkprøven)	$\bar{x} = 15.05$ mm.
	Beregner $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$z_0 = \frac{15.05-15}{0.1/\sqrt{10}} = 1.58$
	Sammenligner $-z_{\frac{\alpha}{2}}$, z_0 og $z_{\frac{\alpha}{2}}$	-1.96<1.58<1.96
	Forkast H_0 og konkluder med H_1, eller behold H_0.	Beholder H_0. Har ikke sterke nok bevis for at $\mu \neq 15$mm.

Ett utvalg: tosidig test for μ med σ kjent

— $X_1, X_2, ..., X_n$ u.i.f. normal(μ, σ) der σ er kjent.

— To-sidig test:

1. $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu \neq \mu_0$
2. Signifikansnivå α bestemmes.
3. Testobservator under H_0 er Z_0. $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ er under H_0 standard normalfordelt.
   Forkast H_0 hvis $|Z_0| > z_{\frac{\alpha}{2}}$.
4. Beregn $\bar{x}$ fra utvalget, og videre $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
   Sammenlign z_0 og $z_{\frac{\alpha}{2}}$, og forkast H_0 hvis $|z_0| > z_{\frac{\alpha}{2}}$.

Kvalitetskontroll av skruer

- Produksjon av skruer.
- Lengden på produsert skrue skal være 15 mm.
- Tar jevnlig stikkprøve fra prosessen, for å sjekke om skruene som produseres er 15 mm lange.
- Hvis stikkprøven tyder på at de produserte skruene ikke er 15 mm, må maskinen som lager skruene kalibreres på nytt.
- Hvordan skal vi bestemme om maskinen skal rekalibreres?

Hypoteser og tester

Hypoteser:

- **Nullhypotese (H_0):** Hypotesen vi vil undersøke om vi har grunnlag fra data for å forkaste. Inneholder en bestemt verdi for en parameter.
- **Alternativ hypotese (H_1):** Hypotesen vi aksepterer dersom vi forkastar nullhypotesen. Ofte mer enn en verdi for en parameter.

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

To typer tester:

- **To-sidig test:** $H_0 : \theta = \theta_0$ mot $H_1 : \theta \neq \theta_0$
- **En-sidig test:**
  - $H_0 : \theta \geq \theta_0$(evt. $\theta = \theta_0$) mot $H_1 : \theta < \theta_0$, eller
  - $H_0 : \theta \leq \theta_0$(evt. $\theta = \theta_0$) mot $H_1 : \theta > \theta_0$

To typer feil

- Type-I: forkaste H_0 gitt at H_0 er sann. Justismord.
- Type-II-feil: ikke forkaste H_0 gitt at H_0 er falsk. La skyldig tiltalt gå fri.

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

Ett utvalg: tosidig test for μ med σ kjent [10.5]

- $X_1, X_2, ..., X_n$ u.i.f. N(μ, σ^2) der σ er kjent.
- To-sidig test:
  1. $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu \neq \mu_0$
  2. Signifikansnivå α bestemmes.
  3. Testobservator under H_0 er Z_0. $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ er under H_0 standard normalfordelt.
     Regel: Forkast H_0 hvis $|Z_0| > z_{\frac{\alpha}{2}}$.
  4. Beregn $\bar{x}$ fra utvalget, og videre $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
     Sammenlign z_0 og $z_{\frac{\alpha}{2}}$, og forkast H_0 hvis $|z_0| > z_{\frac{\alpha}{2}}$.

www.ntnu.no mo@math.ntnu.no (utarbeidet av Mette Langaas), TMA4240 H2006

P-verdi [10.4]

DEF 10.5: En P-verdi er det laveste nivået hvor den observerte verdien til testobservatoren er signifikant.

Utregning: P-verdi = P(for det vi har observert eller noe verre | H_0 er sann)

Steg:
- Bestem null- og alternativ hypotese.
- Velg testobservator.
- Beregn P-verdien basert på testobservatoren.
- Bestem om vi vil forkaste eller beholde nullhypotesen basert på P-verdien og kunnskap om systemet.

Tilleggsinformasjon: Kan også gjøre hypotesetesting basert på signifikansnivå og forkastningsregion og oppgi P-verdi som tilleggsinformasjon.

Kvalitetskontroll: lengde av skruer

- $X_1, X_2, \ldots, X_n$ er lengden på n skruer.
- Anta at $X_1, X_2, \ldots, X_n$ er u.i.f N$(\mu, \sigma^2 = 0.1^2)$.

Estimering	**Hypotesetest**
Gi et anslag (punktestimat) og intervall (konfidensintervall) der vi har 95% tillit til at sann lengde for produserte skruer ligger.	Undersøk om det er grunn til å tro at de produserte skruene ikke er 15 mm lange (test hypotese). Bruk signifikansnivå 5%. $H_0 : \mu = 15 \quad vs. \quad H_1 : \mu \neq 15$
$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ er standard normalfordelt, $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	

Kvalitetskontroll: lengde av skruer

Estimering	Hypotesetest
95% konfidensintervall for μ. $\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	Forkast H_0 hvis $z_0 > z_{\frac{\alpha}{2}}$ eller $z_0 < -z_{\frac{\alpha}{2}}$. Behold H_0 hvis $-z_{\frac{\alpha}{2}} < z_0 < z_{\frac{\alpha}{2}}$ dvs. behold hvis $\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
95 % konfidensintervall: [14.99, 15.11]	$\lvert z_0 \rvert = 1.58$, $z_{0.025} = 1.96$, dermed ikke forkast H_0. p-verdi 0.11.

— Hvis et $(1-\alpha)100\%$ konfidensintervall inneholder μ_0 vil vi med en tosidig hypotesetest med signifikansnivå α ikke forkaste H_0 på nivå α.

— Hvis et $(1-\alpha)100\%$ konfidensintervall ikke inneholder μ_0 vil vi med en tosidig hypotesetest med signifikansnivå α forkaste H_0 på nivå α.

Ett utvalg: ensidig test for μ med σ kjent [10.5]

— $X_1, X_2, ..., X_n$ u.i.f. N(μ, σ^2) der σ er kjent.

— En-sidig test (større):

1. $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$
2. Signifikansnivå α bestemmes.
3. Testobservator $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ er under H_0 standard normalfordelt. Forkast H_0 hvis $Z_0 > z_\alpha$.
4. Observerer $\bar{x}$ fra utvalget, beregn $z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$. Sammenlign z_0 og z_α, og forkast H_0 hvis $z_0 > z_\alpha$.

— En-sidig test (mindre):

- $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$
- ...
- Forkast H_0 hvis $z < z_\alpha$.

Ett utvalg: ensidig test for μ med σ ukjent [10.7]

— $X_1, X_2, ..., X_n$ u.i.f. N(μ, σ^2) der σ er ukjent. $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

— En-sidig test (større):

1. $H_0 : \mu = \mu_0 \quad vs. \quad H_1 : \mu > \mu_0$
2. Signifikansnivå α bestemmes.
3. Testobservator $T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ er under H_0 t-fordelt med $n-1$ frihetsgrader.
   Forkast H_0 hvis $T_0 > t_{\alpha,(n-1)}$.
4. Beregn $\bar{x}$ og s fra utvalget, og videre $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$.
   Sammenlign t_0 og $t_{\alpha,(n-1)}$, og forkast H_0 hvis $t > t_{\alpha,(n-1)}$.

— En-sidig test (mindre):

- $H_0 : \mu = \mu_0 \quad H_1 : \mu < \mu_0$
- ...
- Forkast H_0 hvis $t_0 < t_{\alpha,(n-1)}$.

Ett utvalg: tosidig test for μ med σ ukjent [10.7]

— $X_1, X_2, ..., X_n$ u.i.f. N(μ, σ^2) der σ er ukjent. $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

— To-sidig test:

1. $H_0 : \mu = \mu_0 \quad vs. \quad H_1 : \mu \neq \mu_0$
2. Signifikansnivå α bestemmes.
3. Testobservator $T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ er under H_0 t-fordelt med $n-1$ frihetsgrader.
   Forkast H_0 hvis $|T_0| > t_{\frac{\alpha}{2},(n-1)}$.
4. Beregn $\bar{x}$ og s fra utvalget, og videre $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$.
   Sammenlign t_0 og $t_{\frac{\alpha}{2},(n-1)}$, og forkast H_0 hvis $|t_0| > t_{\frac{\alpha}{2},(n-1)}$.

Signifikansnivå og teststyrke

— Definerer

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

— **Signifikansnivået** for en test $= P(\text{Type I-feil}) = \alpha$.

— **Styrken** for en test er sannsynligheten for å forkaste H_0 når *et bestemt alternativ* er sant (DEF 10.4), dvs.
Styrken $= 1 - P(\text{Type II-feil, bestemt alternativ}) = 1 - \beta$.

Har at

— Reduserer $\alpha \Rightarrow \beta$ øker og $1 - \beta$ (styrken) minker.

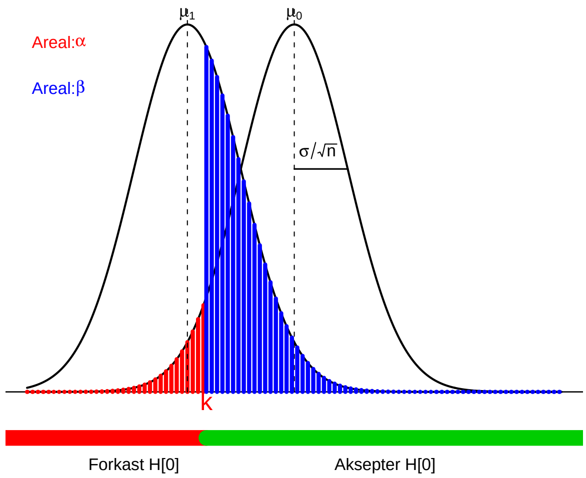
— Øker $n \Rightarrow \alpha$ minker, β minker og $1 - \beta$ (styrken) øker.

Teststyrke, illustrasjon

— Tester hypotesen $H_1 : \mu = \mu_0$ mot $H_1 : \mu < \mu_0$.

— Forkastar H_0 dersom $z_0 < -z_\alpha$, eller ekvivalent $\bar{x} < k$, der $k = \mu_0 - z_\alpha \sigma/\sqrt{n}$.

— Anta sann verdi $\mu = \mu_1$, hva er teststyrken $1 - \beta$?



Fartskontroll med laser

- Ved fartskontroll benytter ofte politiet laser til å måle farten til bilene.
- Hvis Y er målt fart (km/t) til en tilfeldig valgt bil, antar vi at Y er normalfordelt med forventning μ og standardavvik $\sigma = 1.5$ km/t.
- Politiet gjennomfører en fartskontroll i en 50-sone der farten til hver bil måles med en lasermåling.
- Politiet vil fastsette en verdi k slik at sannsynligheten for at en bilist feilaktig beskyldes for fartsovertredelse blir høyst 0.01.
- a) Formuler hypotesetest og finn minste verdi k kan være.
- b) Hva er sannsynligheten for at en bilist som kjører i 55 km/t ikke blir beskyldt for fartsovertredelse?
- c) Hvor mange målinger må vi har for å oppdager at bilisten kjører for fort med styrke 0.95 når bilisten kjører i 55 km/t?

Fasit: k=53.5, ikke beskyldt=0.16, minst 2 observasjoner.

Fartskontroll med laser

- Ved fartskontroll benytter ofte politiet laser til å måle farten til bilene.
- Hvis Y er målt fart (km/t) til en tilfeldig valgt bil, antar vi at Y er normalfordelt med forventning μ og standardavvik $\sigma = 1.5$ km/t.
- Politiet gjennomfører en fartskontroll i en 50-sone der farten til hver bil måles med en lasermåling.
- Politiet vil fastsette en forkastningsregel slik at sannsynligheten for at en bilist feilaktig beskyldes for fartsovertredelse blir høyst 0.01.
- a) Formuler hypotesetest og finn forkastningsregel.
- b) Hva er sannsynligheten for at en bilist som kjører i 55 km/t ikke blir beskyldt for fartsovertredelse?
- c) Hvor mange målinger må vi har for å oppdager at bilisten kjører for fort med styrke 0.95 når bilisten kjører i 55 km/t?

Fasit: k=53.5, ikke beskyldt=0.16, minst 2 observasjoner.

Hypotesetest: generell fremgangsmåte

	Generell fremgangsmåte	**Kvalitetskontroll av skruer**
0	Observasjoner $X_1, X_2, ..., X_n$ u.i.f. fra fordeling med kjente og ukjente parametere.	Stikkprøve (utvalg) av $n = 10$ skruer, antar normalfordeling og kjenner σ =0.1mm.
1	Ensidig eller to-sidig test H_0vs. H_1	Er grunn til å tro at skruene som produseres ikke er 15 mm lange? $H_0 : \mu = 15$ vs. $H_1 : \mu \neq 15$
2	Signifikansnivå α bestemmes.	Velger $\alpha = 0.05$
3	Testobservator: størrelse med kjent fordeling under nullhypotesen. Forkasningsområde fra P(forkaste $H_0 \| H_0$ sann)$\leq \alpha$.	$Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ er under H_0 standard normalfordelt Forkast H_0 hvis $z_0 > z_{\frac{\alpha}{2}}$ eller $z_0 < -z_{\frac{\alpha}{2}}$.

Hypotesetest: generell fremgangsmåte

	Generell fremgangsmåte	**Kvalitetskontroll av skruer**
4	Konklusjon basert på observasjoner Forkast H_0 og konkluder med H_1, eller behold H_0.	$z_{\frac{0.05}{2}} = 1.96, \bar{x} = 15.05$ mm. $z_0 = \frac{15.05-15}{0.1/\sqrt{10}} = 1.58$ -1.96<1.58<1.96 Beholder H_0. Har ikke sterke nok bevis for at $\mu \neq 15$mm.
5	Tilleggsinformasjon: p-verdi=P(det vi har observert eller noe verre| H_0 er sann), teststyrke ved bestemt alternativ	p-verdi 0.11.

Hypoteser og tester

Hypoteser:

- **Nullhypotese (H_0):** Hypotesen vi vil undersøke om vi har grunnlag fra data for å forkaste. Inneholder en bestemt verdi for en parameter.
- **Alternativ hypotese (H_1):** Hypotesen vi aksepterer dersom vi forkastar nullhypotesen. Ofte mer enn en verdi for en parameter.

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

To typer tester:

- **To-sidig test:** $H_0 : \theta = \theta_0$ mot $H_1 : \theta \neq \theta_0$
- **En-sidig test:**
  - $H_0 : \theta \geq \theta_0($ evt. $\theta = \theta_0)$ mot $H_1 : \theta < \theta_0$, eller
  - $H_0 : \theta \leq \theta_0($ evt. $\theta = \theta_0)$ mot $H_1 : \theta > \theta_0$

To typer feil

- Type-I: forkaste H_0 gitt at H_0 er sann. Justismord.
- Type-II-feil: ikke forkaste H_0 gitt at H_0 er falsk. La skyldig tiltalt gå fri.

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

Signifikansnivå og teststyrke

— Definerer

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

— **Signifikansnivået** for en test $= P(\text{Type I-feil}) = \alpha$.

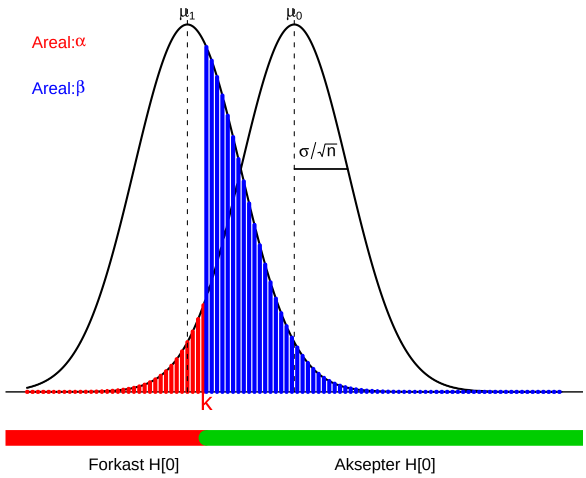
— **Styrken** for en test er sannsynligheten for å forkaste H_0 når *et bestemt alternativ* er sant (DEF 10.4), dvs.
Styrken $= 1 - P(\text{Type II-feil, bestemt alternativ}) = 1 - \beta$.

Teststyrke, illustrasjon

— Tester hypotesen $H_1 : \mu = \mu_0$ mot $H_1 : \mu < \mu_0$.

— Forkastar H_0 dersom $z_0 < -z_\alpha$, eller ekvivalent $\bar{x} < k$, der $k = \mu_0 - z_\alpha \sigma/\sqrt{n}$.

— Anta sann verdi $\mu = \mu_1$, hva er teststyrken $1 - \beta$?



Utvalgsstørrelse [10.9]

— Ensidig test, σ kjent.

- $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu > \mu_0$
- Hvis vi ønsker å ha sannsynlighet $(1-\beta)$ for å oppdage $\mu = \mu_0 + \delta$ (for gitt δ) og ønsker signifikansnivå α, må vi minst ha utvalgsstørrelse

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

— Tosidig test, σ kjent.

- $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu \neq \mu_0$
- Som over, da blir minst utvalgsstørrelsen (tilnærmet)

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

To utvalg: eksempler

— Betong: to ulike oppskrifter, A og B, skal sammenlignes. Er det forskjell i styrken (“crushing strength”) for betong fra oppskrift A og fra oppskrift B?

— Sykdom: tester ut ny blodtrykksmedisin. Er den nye medisinen bedre enn den nåværende markedsledende blodtrykksmedisin?

— Kosthold: får jeg en vektreduksjon på mer enn 10 kg hvis jeg følger Dr Fedon Lindbergs kostråd i et halvt år? (balanse i blodsukker, lav glykemisk indeks)

— Bildekk: to typer dekk, A og B, skal sammenlignes mhp slitasje. Slites A og B dekk forskjellig?

To utvalg: statistisk situasjon

- Ønsker å sammenligne to populasjoner basert på et u.i.f. utvalg fra hver populasjon.
- Nå: Studerer en egenskap som kan sies å være normalfordelt i hver populasjon,
- og ønsker å utføre en hypotesetest om forholdet medllom forveningsverdiene i de to populasjonene
- Sammenligningene kan være parvise eller ikke parvise.
- I 10.12 ser vi på egenskaper som er binomisk fordelt.

10.8: To utvalg, normalfordeling

Situasjon:

- $X_1^A, X_2^A, ..., X_{n_1}^A$ er u.i.f., $X_i^A \sim N(\mu_A, \sigma_A^2)$.
- $X_1^B, X_2^B, ..., X_{n_2}^B$ er u.i.f., $X_j^B \sim N(\mu_B, \sigma_B^2)$.

Problemstilling: Vil teste hypotesen

$$H_0 : \mu_A - \mu_B = d_0 \quad \text{mot} \quad H_1 : \mu_A - \mu_B \neq d_0$$

(Alternativt: $H_1 : \mu_A - \mu_B < d_0$ eller $H_1 : \mu_A - \mu_B > d_0$)

Hypotesetest, tre tilfelle:

1. σ_A^2 og σ_B^2 kjente.
2. $\sigma_A^2 = \sigma_B^2 = \sigma^2$, der σ^2 er ukjent
3. $\sigma_A^2 \neq \sigma_B^2$, σ_A^2 og σ_B^2 ukjente.

To utvalg, normalfordeling (forts.)

1. σ_A^2 og σ_B^2 kjente: Bruker at

$$Z_0 = \frac{(\bar{X}_A - \bar{X}_B) - d_0}{\sqrt{\frac{\sigma_A^2}{n_1} + \frac{\sigma_B^2}{n_2}}} \sim N(0,1) \quad \text{under } H_0.$$

   Forkast H_0 dersom $|z_0| > z_{\frac{\alpha}{2}}$, der z_0 er observert verdi for Z_0.

2. $\sigma_A^2 = \sigma_B^2 = \sigma^2$, der σ^2 er ukjent: T-fordeling med $n_A + n_B - 2$ frihetsgrader.
3. $\sigma_A^2 \neq \sigma_B^2$, σ_A^2 og σ_B^2 ukjente: Se læreboka.

To utvalg, normalfordeling (forts.)

1. σ_A^2 og σ_B^2 kjente: Normalfordeling.
2. $\sigma_A^2 = \sigma_B^2 = \sigma^2$, der σ^2 er ukjent:
   - Estimator for σ^2:

$$S_p^2 = \frac{1}{n_A + n_B - 2}\left[\sum_{i=1}^{n_A}(X_i^A - \overline{X}_A)^2 + \sum_{j=1}^{n_B}(X_j^B - \overline{X}_B)^2\right]$$

   - Bruker at

$$T_0 = \frac{(\overline{X}_A - \overline{X}_B) - d_0}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

     er t-fordelt med $n_A + n_B - 2$ frihetsgrader under H_0. Forkast H_0 dersom $|t_0| > t_{\frac{\alpha}{2},(n_A+n_B-2)}$, der t_0 er observert verdi for T_0.

3. $\sigma_A^2 \neq \sigma_B^2$, σ_A^2 og σ_B^2 ukjente: Se læreboka.

Parvist eksempel: Dekkslitasje

Spørsmål: Er slitasjen for A-dekka større enn for B-dekka?
Forsøk: Utstyr n tilfeldig valgte biler med to dekk av type A og to av type B.

— La $X_i,\ i=1,\ldots,n$ være slitasje til type A-dekka på de n bilene (gj.snitt over to dekk).

— La $Y_i,\ i=1,\ldots,n$ være slitasje til de tilsv. parene av type B-dekk (gj.snitt over to dekk).

— Da er $D_i = X_i - Y_i,\ i=1,\ldots,n$ uavhengige, og $D_i \sim N(\mu_D, \sigma_D^2)$.

Observasjoner:

— $n = 15$ forsøk med observerte verdier $\bar{d} = 0.72$ og $s_d^2 = 0.97$.

Parvist eksempel: Dekkslitasje

Hypotesetest:

— $H_0 : \mu_D = \mu_0$ mot $H_1 : \mu_D > \mu_0$, der $\mu_0 = 0$.

—

$$T_0 = \frac{\bar{D} - \mu_0}{S_D/\sqrt{n}} = \frac{\bar{D} - 0}{S_D/\sqrt{n}}$$ er t-fordelt med $n-1$ frihetsgrader under H_0.

— Gjennomfør testen som for ett utvalg.

Hypotesetest av varians (10.13)

— Inspirert av eksamen, august 2003, oppgave 1.

- En laborant skal undersøke måleusikkerheten til et instrument som benyttes til å bestemme konsentrasjonen av et stoff i en oppløsning.
- Det gjennomføres n målinger med instrumentet på en oppløsning.
- Observasjonene $X_1, X_2, \ldots, X_n$ kan antas å være uavhengige og normalfordelte med forventning μ og varians σ^2.

— I oppgaven arbeider man med kjent konsentrasjon av stoffet, men vi skal her anta at konsentrasjonen er ukjent.

— Vi kan tenke oss at produsenten av måleinstrumentet reklamerer med at måleusikkerheten i instrumentet ikke er høyere enn $\sigma_0^2 = 0.04$. Vi ønsker å teste om dette er tilfellet.

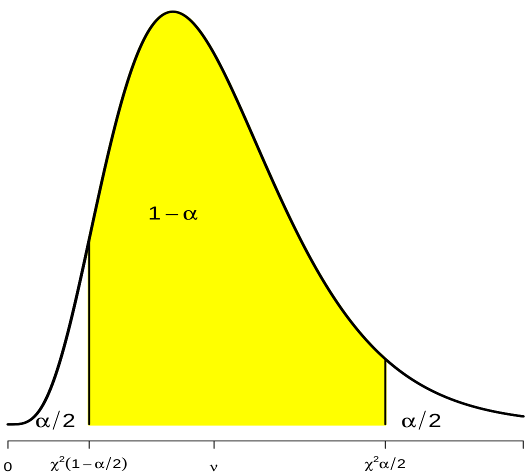
— Data fra oppgaven: $n = 10$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 0.43$ og $\alpha = 0.05$.

[10.13] Hypotesetest av varians

— La $X_1, X_2, ..., X_n$ være et tilfeldig utvalg fra en populasjon som beskrives av en normalfordeling med forventning μ og varians σ^2.

— $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ er en estimator for σ^2 (forventningsrett, men ikke SME).

— Størrelsen $V = \frac{(n-1)S^2}{\sigma^2}$ er kjikvadrat-fordelt med $n - 1$ frihetsgrader.



Laban strakk seg ikke lenger, men smaker den bedre?

— Vi ønsker å finne ut om studenter synes at Nidar Laban smaker bedre enn COOP Seigmenn. Formuler spørsmålet som en hypotesetest.

— Etter seigmann-strekkingen på forelesningen, svarte de studentene som hadde strukket (og spist) både Laban og Seigmenn på hvilket av merkene som smakte best.

— Data: $n = 51$ studenter svarte, av disse likte $x = 30$ studenter Laban bedre enn COOP Seigmann.

— Gjennomfør testen. Hva blir konklusjonen?

— Hva ville konklusjonen blitt hvis vi hadde observert samme andel, $\hat{p} = \frac{30}{51} = 0.59$, men

- $n = 10$ og $x = 6$,
- $n = 100$ og $x = 59$.

[10.11] Hypotesetest av en andel

— X er antall suksesser i et binomisk forsøk med parametere antallet n og andelen p.

— Vi vil teste en hypotese om p, dvs. relatere p til bestemte verdier (ensidig eller tosidig test).

— Estimator $\hat{p} = \frac{X}{n}$, der $\mathrm{E}(\hat{p}) = p$ og $\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$.

— Forkastningsområdet kan enten finnes fra

- binomisk fordeling: relatert til verdien av X, trenger å finne forkastningsområde fra tabell over kumulativ binomisk fordeling,
- fra normaltilnærming av $\hat{p}$ når n er stor, og $np > 5$, $n(1-p) > 5$ og p ikke er nær 0 eller 1.

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{1}{n} p_0(1-p_0)}}$$

er tilnærmet standard normalfordelt under H_0.

Studenter og bilkjøring

— Her angir n antall studenter i utvalget som hadde sertifikat, og x antall studenter som svarte at de er “bedre enn gjennomsnittet av Norges befolkning” til å kjøre bil.

	n	x	$\frac{x}{n}$
Menn	102	50	0.49
Kvinner	37	9	0.24
Alle	139	59	0.42

— a) Finn punktestimat og 99% konfidensintervall for andelen av studenter som synes sine kjøreegenskaper er “bedre enn gjennomsnittet”.

Fasit: $0.42 \pm 2.576\sqrt{\frac{0.42\cdot(1-0.42)}{139}}=[0.32, 0.53]$.

Studenter og bilkjøring

	n	x	$\frac{x}{n}$
Menn	102	50	0.49
Kvinner	37	9	0.24
Alle	139	59	0.42

— b) La p være sannsynligheten for at en tilfeldig valgt student synes han/hun er bedre enn gjennomsnittet til å kjøre bil. Ville hypotesen

$$H_0 : p = 0.5 \text{ vs. } H_0 : p \neq 0.5$$

blitt forkastet? Hvilke signifikansnivå ville ført til forkastning? Baser resonnementet på fasit fra a), dvs. uten regning. Hvilke hypoteser (valg av p_0) ville ikke blitt forkastet på nivå 0.01?

Studenter og bilkjøring, forts.

	n	x	$\frac{x}{n}$
Menn	102	50	0.49
Kvinner	37	9	0.24
Alle	139	59	0.42

— c) Finn punktestimat og 99% konfidensintervall for differensen mellom andelen av mannlige studenter og kvinnlige studenter som synes sine kjøreegenskaper er "bedre enn gjennomsnittet". Fasit: $[0.03, 0.47]$.

— d) La p_1 være sannsynligheten for at en tilfeldig valgt mannlig student synes han er bedre enn gjennomsnittet til å kjøre bil, og tilsvarende p_2 for kvinner. Ville hypotesen

$$H_0 : p_1 - p_2 = 0 \text{ vs. } H_0 : p_1 - p_2 \neq 0$$

blitt forkastet? Hvilke signifikansnivå ville ført til forkastning? Baser resonnementet på fasit fra c), dvs. uten regning. Hvilke hypoteser ville ikke blitt forkastet på nivå 0.01?

Lovlydige bilførere?

— BOT: Kjøring med motorvogn på eller over sperrelinje og/eller i sperreområde begrenset av heltrukken linje, på fortau, gangveg/gangbane, sykkelveg/sykkelbane og gang- og sykkelveg/gang- og sykkelbane. Kr. 2500,- sub. 5 dgs fengsel.

— Målinger av kryssing av hvit heltrukket sperrelinje ved fartsdempere ved bussholdeplass Gløshaugen Nord (29.03.2004, fra 08:20 til 08:35).

- n=antall observasjoner
- X=antall bilister som kjører rundt fartsdemperen (over hvit heltrukket sperrelinje).

	n	x	$\frac{x}{n}$
Menn	74	29	0.39
Kvinner	37	10	0.27
Alle	111	39	0.35

Lovlydige bilførere?

	n	x	$\frac{x}{n}$
Menn	74	29	0.39
Kvinner	37	10	0.27
Alle	111	39	0.35

— a) Er det grunn til å tro at det er flere enn 25% av bilistene som svinger rundt fartsdemperen?
(Fasit: forkast ensidig test på nivå 0.01, p-verdi= 0.004)

— b) Er det grunn til å tro at kvinner og menn er like lovlydige i denne situasjonen?
(Fasit: ikke forkast tosidig hypotese, p-verdi 0.21)