TMA4240 Statistikk H2016 [23]: Hypotesetesting

Hypotesetesting
To normalfordelte utvalg [10.5]
Utvalgsstørrelse [10.6] - del 1 - mer i F24
Test for en andel [10.8]

Mette Langaas

Institutt for matematiske fag, NTNU

wiki.math.ntnu.no/emner/tma4240/2016h/start/

Hva har vi lært så langt?

- Vi setter opp en nullhypotese H_0 (konservativ, status quo) og en alternativ hypotese H_1 (spørsmålet).
- Vi velger et signifikansnivå α som angir hvor stor vi tåler P(Type I feil), der "Type I feil"=forkaste H_0 når H_0 er sann.
- To metoder:
  - Forkastningsmetoden: finner forkastningsgrenser basert på en testobservator (teoretisk, før man har observert data). Valgt signifikansnivå er med å avgjøre hvor strengt vi setter grensene.
  - P-verdi-metoden: basert på data beregnes en p-verdi som angir sannsynligheten for det man har observer eller noe mer ekstremt (i hht H_1) nå nullhypotesen er sann. Hvis p-verdien er mindre enn signifikansnivået forkaster vi nullhypotesen og sier at vi har funnet en signifikant sammenheng.

Hva mer har vi lært så langt?

- En ensidig test bruker α i forkastningsgrenser og p-verdien er basert på en hale. Hvis $H_1 : \mu > \mu_0$ skriver vi $H_0 : \mu = \mu_0$ selv om vi kan mene $H_0 : \mu \leq \mu_0$ fordi 0 er det mest ekstreme tilfellet.
- En tosidig test bruker $\alpha/2$ for forkastningsgrenser, og for p-verdien må ta med både nedre og øvre hale.
- Det er viktig å tegne en figur når vi skal finne forkastningsgrenser eller p-verdi! (Husk å gjøre det på eksamen.)
- Det er en klar sammenheng mellom en tosidig test (på nivå α) og et $(1-\alpha)100\%$ konfidensintervall: ligger μ_0 i et $(1-\alpha)100\%$ KI for μ vil vi ikke forkaste H_0 på nivå α.
- To-utvalgs t-test er kanskje den mest brukte test-situasjonen - og det skal vi se på mer nå!

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 mellom forveningsverdiene i de to populasjonene
- Sammenligningene kan være parvise eller ikke parvise.
- Vi ser på binomisk situasjon i 10.9.

10.5: To utvalg, normalfordeling

Situasjon: to uavhengige utvalg

- $X_{11}, X_{12}, ..., X_{1n_1}$ er u.i.f., $X_{1i} \sim N(\mu_1, \sigma_1^2)$.
- $X_{21}, X_{22}, ..., X_{2n_2}$ er u.i.f., $X_{2j} \sim N(\mu_2, \sigma_2^2)$.

Problemstilling: Vil teste hypotesen

$$H_0 : \mu_1 - \mu_2 = d_0 \quad \text{mot} \quad H_1 : \mu_1 - \mu_2 \neq d_0$$

(Alternativt: $H_1 : \mu_1 - \mu_2 < d_0$ eller $H_1 : \mu_1 - \mu_2 > d_0$)

Hypotesetest, varianser kjente eller ukjente?

1. σ_1^2 og σ_2^2 kjente - vi jobber med normalfordelt testobservasjon.
2. σ_1^2 og σ_2^2 ukjente - vi jobber med t-fordelt testobservasjon.

To metoder for å lære å lese

Vi vil sammenligne to metoder for å lære barn å lese. Etter metoden er brukt måler vi lesescore, og en høy verdi er bra.

Metode 1 (gammel): $n_1 = 22$ barn lærer med denne metoden og $\bar{x}_1 = 41.05$ er gjennomsnittlig lese score, med empirisk standardavvik $s_1 = 5.64$

Metode 2 (ny): $n_2 = 22$ barn lærer med denne metoden og $\bar{x}_2 = 46.73$ er gjennomsnittlig lese score, med empirisk standardavvik $s_2 = 7.39$.

Spørsmålet: Er det grunn til å tro at de som lærer å lese ved hjelp av metode 2 oppnår høyere lesescore enn de som lærer å lese ved hjelp av metode 1?

To uavhengige utvalg, normalfordeling (forts.)

1. σ_1^2 og σ_2^2 kjente: Bruker at

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \text{under } H_0.$$

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

2. $\sigma_1^2 \neq \sigma_2^2$, σ_1^2 og σ_2^2 ukjente: Tilnærmet t-fordelt med "stygg" ν. Se læreboka.

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t_\nu \quad \text{under } H_0.$$

   Forkast H_0 dersom $|t_0| > t_{\nu,\frac{\alpha}{2}}$, der t_0 er observert verdi for T_0.

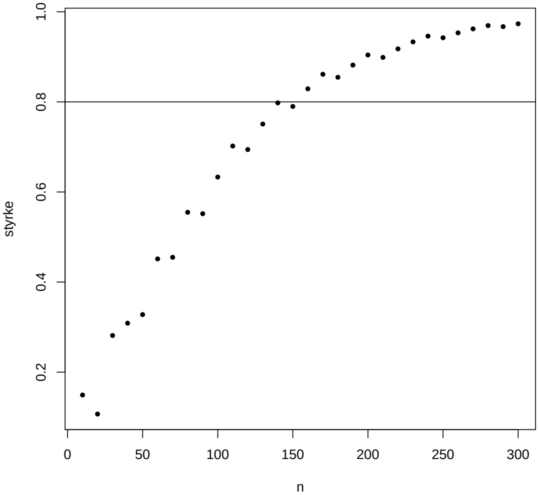
Hypotesetest av en andel [10.1, 10.8]

- 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, p_0 ikke er nær 0 eller 1 (sentralgrenseteoremet).

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

er tilnærmet standard normalfordelt under H_0.

Teststyrke - binomisk andel for ny medisin



Utvalgsstørrelse vs. styrke i p=0.7 for $H_0 : p = 0.6$ vs $H_1 : p > 0.6$ med binomisk fordeling som sorte prikker.

Læringsmål

- To uavhengig normalfordelte utvalg:
  - Kjente varianser: inferens med normalfordeling.
  - Ukjente varianser: inferens med t-fordeling.
- Parrede utvalg: lag differanser og se på det som ett utvalg.
- En andel:
  - Bruk binomisk fordeling og tabeller for å regne ut forkastningsregel og p-verdi.
  - eller ved stor nok n bruk sentralgrenseteoremet og normalfordeling.
- Styrken er P(forkaste H_0 når H_0 er gal)=1-P(type II feil), og den kan vi regne ut for en gitt verdi (vi velger en verdi) under H_1.
- Stryken øker hvis antall observasjoner vi samler inn øker.
- Vi kan finne en formel for hvor mange observasjoner vi må samle inn gitt at vi ønsker en gitt styrke for en verdi i H_1. Vi skal vise formel for ett utvalg normalfordeling med kjent varians i F24.