TMA4240 Statistikk H2016 [22]: Hypotesetesting

Hypotesetesting
Tosidig test vs. konfidensintervall [10.4]
P-verdimetoden [10.3]
Ensidig test for ett normalfordelt utvalg [10.4]
To normalfordelte utvalg [10.5]

Mette Langaas

Institutt for matematiske fag, NTNU

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

Hypotese

- En **statistisk hypotese** er et utsagn om egenskaper til en (eller flere) populasjone(r).
- **Nullhypotese:** Hypotesen vi vil undersøke om vi har grunnlag fra data til å forkaste.
- **Alternativ hypotese:** Reflekterer spørsmålet vi stiller (eller påstanden vi kommer med).

To typer feil

- DEF 10.2: Forkasting av nullhypotesen når denne er sann, kalles en type-I-feil.
- DEF 10.3: Å ikke forkaste nullhypotesen når den er gal, kalles en type-II-feil.

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

http://www.pitt.edu/ upjecon/MCG/STAT/DeadlySins.jpg

Hypothesis test and legal trial

Legal trial:

- H_0: The defendent is innocent.
- H_1: The defendent is quilty.

Type I and Type II errors:

- Type I error: Miscarriage of justice, convict an innocent.
- Type II error: not convicting a guilty defendent.

The null- and alternative hypotheses are not commutable - they are handled differently. If we experience doubt we accept the null hypothesis.

Type-I, Type-II and Power

	Not reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 false	Type II error	Correct

- Two sources of error: false positive rate α and false negative rate β.
  - False positives = type I error.
  - False negatives = type II error.
- Power of a test: Ability to detect that the null hypothesis is false (1-β).

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 $|z_0| > z_{\frac{\alpha}{2}}$.

Tosidig hypotesetest vs. konfidensintervall

$$H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu \neq \mu_0$$

- 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.

Skriftlig øving 3, oppgave 1c: Marsvin og C-vitamin

Tilskudd 1: $X_1, X_2, \ldots, X_{n_1} \sim N(\mu, \sigma^2)$ angir odontoblastlengdene til et tilfeldig utvalg av $n_1 = 10$ marsvin.
Tilskudd 2: $Y_1, Y_2, \ldots, Y_{n_2} \sim N(\eta, \tau^2)$ angir odontoblastlengdene til et tilfeldig utvalg av $n_2 = 10$ marsvin.

Forskerne vil gjerne se på forskjellene de to tilskuddene har på differansen i forventet odontoblastlengde. Definer $\delta = \mu - \eta$ som differansen mellom forventet odontoblastlengde for de to tilskuddene.

Fant 95% konfidensintervall til $[1.66, 8.71]$,

Er det grunn til å tro at de to tilskuddene har ulik effekt på forventet odontoblastlengde? (Dette spørsmålet skal vi jobbe mer med videre i kurset, men du skal her basere svaret ditt på konfidensintervallet du har laget.)

$\mu_0 = 0$ er ikke i 95% konfidensintervallet og dermed ville vi forkaste H_0 med signifikansnivå 5% ($\alpha = 0.05$).

P-verdi [10.3]

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 $\mid H_0$ 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.

Vil pedikyrbehandling påvirke målt kroppsfettprosent?

Skriv ned null og alternativ hypoteste, forklar hva utskriften sier.
Velg signifikansnivå 5% og konkluder.

```
length(diff) = 40
mean(diff)  = 0.2975
std(diff) = 1.1002

[h,p,ci,stats]=ttest(diff)
h =     0
p =    0.0952
ci =
   -0.0544
    0.6494
stats =
    tstat: 1.7101
       df: 39
       sd: 1.1002
tcdf(1.73,39,'upper')=0.0458
```

Kroppsfettprosent og pedikyr

$$H_0 : \mu = 0 \text{ mot } H_1 : \mu \neq 0$$

Vi antar at testobservatoren som brukes er t-fordelt

$$T_0 = \frac{\bar{D} - 0}{S\sqrt{\frac{1}{n}}}$$

$$t_0 = \frac{0.3 - 0}{1.10\sqrt{\frac{1}{40}}} = 1.73$$

Hvis signifikansnivå 5%: Kritisk verdi finnes fra t-fordelingen med $n - 1 = 40 - 1 = 39$ frihetsgrader, tabellen: $t(0.025, 35) = 2.03$, siden vi har en tosidig test. Vi forkaster H_0 når $t_0 \geq 2.03$ eller når $t_0 \leq -2.03$. Vi har observert at $t_0 = 1.73$, og dermed vil vi ikke forkaste nullhypotesen.

Kroppsfettprosent og pedikyr

$$H_0 : \mu = 0 \text{ mot } H_1 : \mu \neq 0$$

P-verdi er $2 \cdot P(T > 1.73) = 2 \cdot 0.0458 = 0.0952$.
Med signifikansnivå 5% er p-verdien er større enn valgt signifikansnivå, og vi forkaster ikke nullhypotesen.
Konklusjon: Vi forkaster ikke nullhypotesen og har ikke tilstrekkelig grunnlag til å si at pedikyrbehandlingen har noen effekt på måling av kroppsfettprosent.

P-verdi

- P-verdien gir sannsynligheten for å observere det vi har gjort, eller noe mer ekstremt (i retning H_1) når nullhypotesen er sann.
- P-verdien er *ikke* sannsynligheten for at nullhypotesen er sann (dessverre ikke).
- Vi forkaster nullhypotesen når p-verdien er *mindre* enn det valgte signifikansnivået, f.eks. $\alpha = 0.05$ eller $\alpha = 0.01$.
- Dette vil garantere oss at *sannsynligheten* for å begå en Type I-feil (justismord) er mindre eller lik α.

Reporting results from hypothesis test

- Report p-value rather than e.g. t-statistic, since the p-value a "normalized" test statistic.
- Report confidence interval because that takes into account the precision of the estimated effect size.
- According to the *ISO Guide to the Expression of Uncertainty in Measurement (GUM)* the standard error should also be reported. For the t-test example this is $\frac{s}{\sqrt{n}}$.
- Statistical significance does not imply medical relevance.

Ett utvalg: ensidig test for μ med σ kjent [10.4]

- Ensidig test (større):
  0. $X_1, X_2, ..., X_n$ u.i.f. $\mathrm{N}(\mu, \sigma^2)$ der σ er kjent.
  1. En-sidig test (større): $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu > \mu_0$
  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):
  1. $H_0 : \mu = \mu_0 \qquad H_1 : \mu < \mu_0$
  3. Forkast H_0 hvis $z < z_\alpha$.

Ett utvalg: ensidig test for μ med σ ukjent [10.4]

- Ensidig test (større):

  0. $X_1, X_2, ..., X_n$ u.i.f. $\mathrm{N}(\mu, \sigma^2)$ der σ er ukjent.
     $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
  1. En-sidig test (større), $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu > \mu_0$
  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):

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

Ett utvalg: tosidig test for μ med σ ukjent [10.4]

0. $X_1, X_2, ..., X_n$ u.i.f. $\mathrm{N}(\mu, \sigma^2)$ der σ er ukjent.
   $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
1. To-sidig test: $H_0 : \mu = \mu_0 \qquad vs. \qquad H_1 : \mu \neq \mu_0$
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 $|t_0| > t_{\frac{\alpha}{2},(n-1)}$.

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_1^2 kjente - vi jobber med normalfordelt testobservasjon.
2. σ_1^2 og σ_1^2 ukjente - vi jobber med t-fordelt testobservasjon.

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.

Læringsmål: husk å alltid tegne figur!

- Det er en klar sammenheng mellom en tosidig test (på nivå α) og et $(1-\alpha)100\%$ konfidensintervall.
- Forkastningsmetoden: finner forkastningsgrenser basert på en testobservator før man har observert data. Valgt signifikansnivå avgjør 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.
- En tosidig test bruker $\alpha/2$ for forkastningsgrenser, og p-verdien å ta med nedre og øvre hale.
- En ensidig test bruker α i forkastningsgrenser og p-verdien er basert på en hale. Hvis $H_1 : \mu > 0$ skriver vi $H_0 : \mu = 0$ selv om vi kan mene $H_0 : \mu \leq 0$ fordi 0 er det mest ekstreme tilfellet.
- To-utvalgs t-test er kanskje den mest brukte test-situasjonen.