Statistik

To gennemsnit – store

Test og Konfidensinterval for forskellen på to middelværdier:

Population 1:
$\mu_1 = Middelværdi$

Stikprøve 2:
$n_1 = antal\ observationer$
$\overline{x_1} = stikprøve\ gennemsnit$
$s_1 = stikprøve\ standardafvigelse$

Population 2:
$\mu_2 = Middelværdi$

Stikprøve 2:
$n_1 = antal\ observationer$
$\overline{x_1} = stikprøve\ gennemsnit$
$s_1 = stikprøve\ standardafvigelse$

Forudsætninger:
- Stikprøven er tilfældig udvalgte

Hypoteser:
Hypoteserne / opskrivningerne af hypoteserne er <u>helt</u> identiske med kapitel 21.

Beregninger:

$$Ts = \frac{\overline{x_1} - \bar{x}_2 - \mu_0}{SD(\overline{x_1} - \overline{x_2})}$$

$$Ts = \frac{\bar{x}_1 - \overline{x_2} - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Testen følger en z fordeling:
Vagt af korrekt hypotese følger reglerne fra kap 21.

<u>Eller brug p-værdi metoden:</u>
- Er p-værdien under testniveau - accepteres H1.

Hypotese:

$$H_0: \sigma_1 - \sigma_2 \leq 0 \rightarrow \ \sigma_1 \leq \sigma_2$$
$$H_1: \sigma_1 - \sigma_2 > 0 \rightarrow \ \sigma_1 > \sigma_2$$

$$Ts = \frac{1{,}8^2}{1{,}4^2} = 1{,}653061$$

Kritiske værdi:
$C = 2{,}11$ - Fundet i exceldokumentet "model for forskelle i to standardafvigelser - F test".

Da Ts < C kan H0 ikke afvises.

P-værdi = 0,133

Konklusion:
VI kan ikke vise at den nye procedure har givet lavere standardafgivelse.

Opgaver:

Opgave 1:

Et forsikringsselskab har kunderne opdeling i to grupper (1 og 2) og vil undersøge om den gennemsnitlige forsikringspris er ens i de to grupper. Der er foretaget to stikprøveudtagninger med følgende resultat.

Obs.	Gr. 1	Gr. 2
1	591	617
2	783	765
3	793	840
4	749	738
5	787	820
6	570	612
7	608	553
8	714	702
9	696	695
10	649	756
11	668	720

Test på 5 % niveau om populations-standardafvigelserne er ens.

Forudsætningstjek:
- Kan/bør laves!
    - Normalfordelingsplot

Tester om P1 = P2 (eller om P1 er forskellig for P2) - Dobbelt sidet test.

Beregning:

Når $P_0 = 0$:
Kræver sammenvægtning af $\widehat{p_1}$ og $\widehat{p_2}$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$SD(\widehat{p_1} - \widehat{p_2}) = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n_1} + \frac{\hat{p} \cdot (1 - \hat{p})}{n_2}}$$

$$Ts = \frac{\widehat{p_1} - \widehat{p_2} - 0}{SD(\widehat{p_1} - \widehat{p_2})}$$

Ellers!!:

$$SD(\widehat{p_1} - \widehat{p_2}) = \sqrt{\frac{\widehat{p_1} \cdot (1 - \widehat{p_1})}{n_1} + \frac{\widehat{p_2} \cdot (1 - \widehat{p_2}}{n_2}}$$

$$Ts = \frac{\widehat{p_1} - \widehat{p_2} - p_0}{SD(\widehat{p_1} - \widehat{p_2})}$$

Konfidensinterval:

$(\widehat{p_1} - \widehat{p_2}) \pm Z \cdot SD(\widehat{p_1} - \widehat{p_2})$ - Uden sammenvægtet $\hat{p}$

Test og Konfidensinterval følger en Z-fordeling.

Typeopgave:

Århus:

$$n_1 = 400$$
$$x_1 = 128$$

$$\widehat{p_1} = \frac{128}{400} = 0{,}32$$

København:

$$n_2 = 600$$
$$x_2 = 174$$

$$SD(\widehat{p_1}-\widehat{p_2})=\sqrt{\frac{0{,}332\cdot(1-0{,}332)}{1000}+\frac{0{,}396\cdot(1-0{,}396)}{900}}=0{,}0152$$

Grænseværdier:

$$(0{,}332-0{,}396)\pm(1{,}96\cdot 0{,}0152)$$

Nedre grænse:

$$(0{,}332-0{,}396)-(1{,}96\cdot 0{,}0152)=-0{,}093792$$

Øvre grænse:

$$(0{,}332-0{,}396)+(1{,}96\cdot 0{,}0152)=-0{,}034208$$

Lineær regression:

Mål:

At få fastlagt en matematisk sammenhæng (lineær) mellem en x-variable og en y-variable.

På baggrund af et datasæt (x og y), tegnes en ret linje og forskriften for den fastlægges.

Ret linje: (I matematik)

$$f(x)=a\cdot x+b$$

a: hvor meget stiger/falder med når x stiger med 1 $\frac{\Delta y}{\Delta x}$ - Hældningskoefficient

b: skæring på 2- aksen - begyndelsesværdien - skæring på y-aksen

Ret linje: (I statistik)