Kursus 02403
Introduktion til Statistik

Klaus Kaae Andersen

Informatics and Mathematical Modelling
Building 321 - room 011
Technical University of Denmark
2800 Lyngby – Denmark
e-mail: kka@imm.dtu.dk

Ikke parametriske metoder (kapitel 10)

- Sign Test (afsnit 10.2)
- Rank-Sum Test (afsnit 10.3)
- Test for tilfældighed (afsnit 10.4)

Ikke parametriske metoder

- De metoder vi har anvendt indtil nu har været baseret på antagelse om at data følger en givet fordeling (især normal fordelingen)
- Ikke parametriske metoder kan anvendes som alternativ, såfremt normalfordelingsantagelsen ikke holder

Vi betragter

- ◇ Sign Test (afsnit 10.2)
- ◇ Rank-Sum Test (afsnit 10.3)
- ◇ Test for tilfældighed (afsnit 10.4)

Sign Test

Sign test kan bruges som alternativ til

- ◇ hypotesetest for en middelværdi
- ◇ parret t-test
- ◇ hypotesetest for store stikprøver

når ovenstående test ikke kan bruges pga. antagelse om normalfordelingen

Sign Test

Sign test kan bruges til at teste hypotese om <u>medianen</u>

$$H_0: \quad \tilde{\mu} = \tilde{\mu}_D$$
$$H_1: \quad \tilde{\mu} \neq \tilde{\mu}_D$$

hvor $\tilde{\mu}_D$ er den værdi vi ønsker at teste

Sign Test

Beregning af teststørrelse/p-værdi:

- Antal af observationer større end medianen optælles, X_+.
- Testets p-værdi kan nu findes ved at beregne sandsynligheden for (ensidet test)

$P(X_+ \leq X)$ hvor $X \sim b(x; n, \frac{1}{2})$

- Såfremt p-værdi er mindre end signifikansniveau, forkastes H_0

(evt. kan man basere testet på antal observationer mindre end medianen, X_-)

Rank-Sum Test

- Rank-Sum Test (også kaldet U-Test eller Wilcoxon test eller Mann-Whitney test) kan bruges som alternativ til almindelig t-test for 2 uafhængige stikprøver, i tilfælde af at normalfordelingsantagelsen ikke holder
- Rank-Sum Test kan altså bruges til at sammenligne medianen for 2 uafhængige stikprøver:

$$x_1, x_2, \ldots x_{n_1}$$

$$y_1, y_2, \ldots y_{n_2}$$

Rank-Sum Test

<u>Beregning af teststørrelse:</u> Data sorteres og rangeres (eng: ranks) i stigende rækkefølge. For hver af de to stikprøver summeres de tilhørende ranks, her benævnt W_1 og W_2

Vi beregner nu

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2}$$

eller

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

Rank-Sum Test

Det gælder nu, at såfremt de to stikprøver kommer fra den samme fordeling, så haves

$$\mu_{U_1} = \frac{n_1 n_2}{2}$$

$$\sigma^2_{U_1} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

når n_1 og n_2 er tilpas store (> 8) kan vi nu anvende

$$Z = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}} \sim N(0, 1^2)$$

Test for tilfældighed

I mange undersøgelser er det vigtigt at afgøre om en stikprøve er fremkommet tilfældigt

Hvis vi har en sekvens med n_1 af den ene type og n_2 af en anden type (og hverken n_1 eller n_2 er mindre end 10), f.eks.

$$K\ K\ K\ P\ K\ K\ P\ P\ K\ P\ P\ K\ P\ K\ P...$$

Test for tilfældighed

Det totale antal skift, u, approximeres med en normalfordeling med

$$\mu_u = \frac{2 n_1 n_2}{n_1 + n_2} + 1$$

og

$$\sigma_u = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

Test for tilfældighed

Vi kan nu beregne p-værdien ved

$$Z = \frac{u - \mu_u}{\sigma_u}$$

Idet

$$Z \sim N(0, 1^2)$$

Ikke parametriske metoder (kapitel 10)

- Sign Test (afsnit 10.2)
- Rank-Sum Test (afsnit 10.3)
- Test for tilfældighed (afsnit 10.4)