

## Ei-parametriset ja robustit menetelmät, syksy 2012

### Harjoitus 5

1. Tarkastellaan kahden riippumattoman otoksen tapausta, jossa ensimmäisen otoksen havainnot  $x_1, \dots, x_m$  tulevat  $N(\mu, \sigma^2)$ -jakaumasta ja toisen otoksen havainnot  $y_1, \dots, y_n$  tulevat  $N(\mu + \Delta, \sigma^2)$ -jakaumasta. Käsittelyjen eroa kuvaavan parametrin  $\Delta$  piste-estimaattina käytetään  $t$ -testin tapauksessa otoskeskiarvojen erotusta  $\bar{y} - \bar{x}$ . Tutki millä parametrin  $\lambda$  arvolla piste-estimaatin varianssi  $\text{var}(\hat{\Delta}) = \text{var}(\bar{y} - \bar{x})$  saa pienimmän arvon.
2. Vertaile kahden otoksen  $t$ -testin, Moodin testin ja kahden otoksen Wilcoxonin testin voimakkuuksia graafisesti  $p$ -arvon kertymäfunktionestimaatin avulla, kun  $m = n = 10$  ja havainnot tulevat a) normaalijakaumasta b) Laplace-jakaumasta c)  $t$ -jakaumasta vapausastein  $\nu = 3$  ja d) tasaisesta jakaumasta. Voit käyttää hyväksi luentomonisteen kappaleiden 2.2.2, 2.2.3 ja 2.2.4 R-funktioita.
3. Tyttöjen ja poikien syntymäpainoista on saatu seuraavanlaiset riippumattomat satunnaisotokset ( $m = n = 10$ ): Aineisto löytyy myös tiedostosta `syntpain.data`.

Tytöt	Pojat
3500	3440
3900	3600
3870	4530
3600	3200
3730	2830
3240	3370
3530	3750
2600	3560
3370	3370
3560	4070

Tutki tyttöjen ja poikien syntymäpainoja kahden otoksen  $t$ -testin, Moodin testin ja kahden otoksen Wilcoxonin testin avulla. Laske myös parametrin  $\Delta$  piste-estimaatit ja likimääräiset 95% luottamusvälit. Vertaile eri menetelmillä saatuja tuloksia. Laske Moodin testiä ja kahden otoksen Wilcoxonin testiä vastaaville luottamusväleille tarkat peitetodennäköisyydet ja normaaliapproksimaatioon perustuvat likimääräiset peitetodennäköisyydet. Voit käyttää hyväksi luentomonisteen kappaleiden 2.2.2, 2.2.3 ja 2.2.4 R-funktioita.

4. Tarkastellaan edelleen kahden riippumattoman otoksen tapausta. Tee R-funktio, joka laskee  $t$ -testisuurelle
  - a)  $t$ -jakaumaan perustuvan  $P$ -arvon ja
  - b) permutaatiotestin  $P$ -arvon likiarvon simuloimalla.

Kokeile funktiota simuloimalla aineistoja normaalijakaumasta, Laplace-jakaumasta ja tasaisesta jakaumasta.

Avuksi: Permutaatiotestissä käydään läpi kaikki mahdolliset yhdistetyn otoksen havaintojen jaot kahteen ryhmään (ryhmien koot  $m$  ja  $n$ ) ja lasketaan testisuureen arvo jokaiselle jaolle erikseen. Nollahypoteesin ( $H_0 : \Delta = 0$ ) vallitessa kaikki jaot ovat yhtä todennäköisiä. Kaksisuuntaisen testin P-arvo on silloin

$$\frac{1}{\binom{N}{m}} \#\{|t_i| \geq |t_0|\} = \frac{1}{\binom{N}{m}} \sum_{i=1}^{\binom{N}{m}} I(|t_i| \geq |t_0|)$$

jossa  $t_i$  on jaon  $i$  antama testisuureen arvo ja  $t_0$  on alkuperäisen jaon antama testisuureen arvo. Permutaatiotestin P-arvoa voi approksimoida simuloimalla. Tällöin likimääräiseksi P-arvoksi saadaan

$$\frac{1}{M} \#\{|t_i| \geq |t_0|\} = \frac{1}{M} \sum_{i=1}^M I(|t_i| \geq |t_0|),$$

jossa  $t_i$  on  $i$ :nnettä simuloitua (arvottua) jakoa vastaava  $t$ -testisuureen arvo ja  $M$  on simuloitavien jakojen lukumäärä (esim.  $M = 1000$ ).

Avuksi:

```
> m<-5
> n<-5
> N<-m+n
> z<-rnorm(N) #Yhdistetty otos z=(z1,...,zN)=(x1,..xm,y1,...,yn)
> z
[1] 1.7673729 0.2983867 -1.5357762 0.4138721 -0.4233122 -1.8539434
[7] 1.0616170 -0.4996802 -0.1155978 -1.5055158
> indx<-sample(N,m) #Ens.otoksen indeksit arvotaan yhdistetyn
#otoksen indekseista 1,...,N
> indx
[1] 6 7 1 9 10
> z[indx] #Ensimm. otos
[1] -1.8539434 1.0616170 1.7673729 -0.1155978 -1.5055158
> z[-indx] #Toinen otos
[1] 0.2983867 -1.5357762 0.4138721 -0.4233122 -0.4996802
```

Esimerkki for-silmukan käytöstä:

```
koe<-function()  
{  
  M<-10  
  a<-NULL  
  b<-NULL  
  for(i in 1:M){  
    a[i]<-2*i-1  
    b[i]<-i^2  
  }  
  list(a=a,b=b)  
}
```

```
> koe()
```

```
$a
```

```
[1] 1 3 5 7 9 11 13 15 17 19
```

```
$b
```

```
[1] 1 4 9 16 25 36 49 64 81 100
```