Modeli s ponovljenim mjerenjima.
MANOVA

STATISTIČKI PRAKTIKUM 2

12

Do sada

1. $Y = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_p x_p + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$, $x_1, \ldots, x_p$ kvantitativne varijable prediktori
2. prediktori mogu biti i faktorske varijable (paralelni pravci), npr. $Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 d_1 + \alpha_3 d_2 + \varepsilon$
3. uvodimo interakciju između prediktora $Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 d_1 + \alpha_3 d_2 + \alpha_4 x_1 d_1 + \alpha_5 x_1 d_2 + \varepsilon$
4. generalizirani linearni modeli (dozvoljavamo ne-Gaussovsku i diskretnu strukturu varijable Y)

5. Y višedimenzionalna, prediktori $x_1, \ldots, x_k$ utječu na cijeli niz varijabli

Analiza varijance ponovljenih mjerenja

Dizajn analize varijance u kojima se provode ponovljena mjerenja (različiti tretmani) na istim subjektima čime se postiže ekonomičnost (na jednom subjektu se primjenjuje više tretmana - trebamo manje subjekata) te se grešku na svakom subjektu može bolje objasniti (imamo više informacija).

Kao i ranije, imamo pretpostavku nezavisnosti unutar jednog ponavljanja te pretpostavku nezavisnosti između subjekata (ljudi), dok je očito da će postojati korelacija unutar bloka odnosno da će različita mjerenja nad istim subjektom biti povezana - tu varijabilnost ćemo moći bolje objasniti.

Model je *balansiran* ako za svakog subjekta imamo mjerenja na svim razinama.

Jednofaktorski model s ponavljanim mjerenjima

Model je sljedeći:

$$X_{i,j} = \mu + \xi_i + \delta_j + \varepsilon_{i,j},$$

gdje je

- μ .. ukupna aritmetička sredina
- ξ_i .. efekt subjekta (definira osobu koju promatramo), donosi varijabilnost između subjekata (zato što su promatrani subjekti različiti)
- δ_j .. efekt j-tog tretmana, donosi varijabilnost unutar svakog subjekta
- $\varepsilon_{i,j}$ .. greška, normalno distribuirana

Pretpostavke:

1. *normalnost*: za svaki nivo tretmana, greške su normalno distribuirane, a sami nivoi imaju multivarijatnu normalnu razdiobu (tj. ponovljena mjerenja za svaki subjekt možemo promatrati kao model s multivarijatnom normalnom razdiobom)
2. *homogenost*: greške su jednako distribuirane (s očekivanjem 0 i konstantnom varijancom), a za svaki subjekt, varijanca razlika između svaka dva podatka je konstantna (sferičnost kovarijacijske matrice)
3. *nezavisnost*: subjekti su međusobno nezavisni (i odabrani su na slučajan način) te su razlike između sredina po nivoima tretmana međusobno nezavisne

Primjer

Provedeno je istraživanje o utjecaju 4 vrste lijeka na čovjeka. U istraživanju je sudjelovalo 5 ljudi. Nakon što im je dan lijek, mjereno je koliko im je vremena potrebno da obave niz standardiziranih fizičkih radnji koje su im prije objašnjene. Podaci se nalaze u datoteci `lijekovi.txt`. Imamo 5 osoba i 4 lijeka, ukupno 20 podataka. Dizajn je balansiran.

Koristeći naredbu `aov` testirajte efekt lijeka na sposobnosti čovjeka.

Rješenje

```
> lijekovi = read.table("lijekovi.txt", header=T)
> lijekovi$lijek = as.factor(lijekovi$lijek)
> lijekovi$osoba = as.factor(lijekovi$osoba)

> model1 = aov(vrijeme ~ lijek + Error(osoba/lijek),
      data=lijekovi)
```

```
> summary(model1)

Error: osoba
Df  Sum Sq  Mean Sq  F value Pr(>F)
Residuals  4   680.8   170.2

Error: osoba:lijek
Df Sum Sq  Mean Sq  F value   Pr(>F)
lijek      3  698.2   232.7    24.76   1.99e-05 ***
Residuals 12  112.8     9.4
```

Dakle, postoji značajna razlika medu tretmanima (tj. utjecaj lijeka je značajan).

Također, moguće je provesti i test značajnosti modela (kao i ranije) čija je p-vrijednost jako mala pa možemo zaključiti da je model značajan.

`Error(subject/value)` ... razdvaja podatke za svaku osobu, tj. definira podatke koji pripadaju istoj osobi kroz varijablu `subject` (tipa `factor`)

Ako želimo izračunati očekivanja za svaku razinu djelujućeg faktora (odnosno ako nas zanima kako pojedina razina faktora utječe na zavisnu varijablu), možemo koristiti naredbu

```
> model.tables(model1, "means")
Tables of means
Grand mean

24.9

  lijek
lijek
   1    2    3    4
26.4 25.6 15.6 32.0
```

Dvofaktorski model s ponavljanim mjerenjima

Model je sljedeći:

$$X_{i,j} = \mu + \xi_i + \delta_j + \xi\delta_{i,j} + \varepsilon_{i,j},$$

gdje je $\xi\delta_{i,j}$ interakcija između dva faktora i mjeri je li utjecaj faktora δ konstantan za svaku osobu.

U ovom slučaju bismo pokrenuli naredbu `aov` slično kao ranije: nezavisne varijable povezujemo veznikom *, a kroz faktor `Error` reguliramo rezultate koji su dobiveni za istu osobu.

MANOVA

MANOVA služi za provođenje statističkih testova s više od jedne zavisne varijable. Radi se o tzv. *analizi nezavisnosti*, odnosno, testira se pretpostavka o uzročno–posljedičnoj vezi između 1 ili više prediktora i 2 ili više varijabli odaziva.

Test MANOVA je u principu generalizacija tzv. jednofaktorske ili dvofaktorske ANOVE.

Neka imamo p promatranih zavisnih varijabli (obilježja) $Y_1, \ldots, Y_p$ na koje djeluje nezavisna faktorska varijabla sa s razina. Za svaku razinu faktora i promatramo n_i subjekata i za svakog od njih mjerimo svih p obilježja, tj. imamo uzorak oblika $Y_{i,j,k}$, gdje je $i = 1, \ldots, s, j = 1, \ldots, n_i$, a $k = 1, \ldots, p$.

Promatrani slučajni vektori $Y_{i,j}$ moraju zadovoljavati uobičajene pretpostavke:

1. $Y_{i,1}, \ldots, Y_{i,n_i}$ imaju multivarijatnu normalnu razdiobu, za $i = 1, \ldots, s$ (jer su $\varepsilon_{i,j} \sim N(0, \Sigma)$)
2. sva mjerenja su međusobno nezavisna.

Testiramo hipotezu:

$$H_0 : \mu_1 = \cdots = \mu_s$$
$$H_a : \text{barem dvije aritmetičke sredine su različite}$$

Primjer

Provedena je kemijska analiza 26 posuda pronađenih u iskapanjima u Velikoj Britaniji na 4 različita mjesta. Podaci su dani u datoteci `pottery.txt`. Analizirao se postotak nekoliko vrsta metala u svakoj posudi. Možemo li tvrditi da mjesto pronalaska posude utječe na kemijski sastav posude?

Rješenje

```
> podaci = read.table("pottery.txt", header=T)

> mod2 = manova(cbind(Al,Fe,Mg,Ca,Na) ~ Site,
             data=podaci)
> summary(mod2)

          Df Pillai approx F num Df den Df    Pr(>F)
Site       3 1.5539   4.2984     15     60 2.413e-05 ***
Residuals 22
```

Za testiranje pretpostavki modela možemo koristiti ranije naučene metode (npr. Bartlettov test za homogenost varijanci).