Uvod u linearne modele

Višestruka linearna regresija

Analiza kovarijanse

Analiza varijanse

maj 2012. godine

Sadržaj

Definisanje modela

- Proširenje osnovne formule:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_m X_{im} + \epsilon_i,$$

odnosno, u matričnom formatu:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{1m} \\ 1 & X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Sadržaj

Plauzibilni model

- Vratićemo se našim podacima iz tabele *Prestige*:

```
> require(car)
> names(Prestige)
```

```
[1] "education" "income"    "women"     "prestige"  "census"    "type"
```

- Regresiraćemo prestiž zanimanja, y, sa nizom potencijalno efikasnih prediktora:
  - godine obrazovanja, x_1
  - prosečni prihodi, x_2
  - procenat žena u datom zanimanju, x_3

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

Osnovni rezultati funkcije `lm()`

```
> prestige.lm <- lm(prestige ~ education + income + women, data=Prestige)
> summary(prestige.lm)

Call:
lm(formula = prestige ~ education + income + women, data = Prestige)
Residuals:
     Min       1Q   Median       3Q      Max
-19.8246  -5.3332  -0.1364   5.1587  17.5045

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.7943342  3.2390886  -2.098   0.0385
education    4.1866373  0.3887013  10.771  < 2e-16
income       0.0013136  0.0002778   4.729 7.58e-06
women       -0.0089052  0.0304071  -0.293   0.7702

Residual standard error: 7.846 on 98 degrees of freedom
Multiple R-squared: 0.7982, Adjusted R-squared: 0.792
F-statistic: 129.2 on 3 and 98 DF,  p-value: < 2.2e-16
```

Standardizovani regresioni koeficijenti

- Sirovi koeficijenti nam ne dopuštaju da poredimo doprinose prediktora
- Ovakvu mogućnost daju nam standardizovani regresioni koeficijenti
- Treba znati da mnogi autori imaju rezerve prema ovom tipu koeficijenata, pre svega, zbog standardne greške standardizovanih regresionih koeficijenata (vidi npr.: Fox, 2002)

TEŠKA PITANJA:

Zbog čega ne možemo porediti sirove regresione koeficijente?

Koji problem/opasnost postoji u vezi sa pomenutom standardnom greškom standardizovanih regresionih koeficijenata?

Standardizovani regresioni koeficijenti

- U R-u, dobijanje standardizovanih regresionih koeficijenata je relativno jednostavno
- Jedna mogućnost je da originalne podatke standardizujemo: pretvaranje u $z - vrednosti$, sa $\bar{X} = 0$ i $s = 1$, pomoću funkcije `scale()`

```
> Prestige.z <- data.frame(scale(
+  Prestige[,c('prestige', 'income', 'education', 'women')]))
> Prestige.z[1:5,]
```

```
                    prestige    income education      women
gov.administrators  1.2767988 1.3078662 0.8693455 -0.5616725
general.managers    1.2942361 4.4939820 0.5578127 -0.7867320
accountants         0.9629272 0.5824643 0.7447324 -0.4185673
purchasing.officers 0.5793063 0.4868431 0.2499449 -0.6262904
chemists            1.5499834 0.3780328 1.4227745 -0.5452816
```

STANDARDIZOVANI REGRESIONI KOEFICIJENTI

```
> prestige.lm.z <- lm(prestige ~ education + income + women, data=Prestige.z)
> summary(prestige.lm.z)

Call:
lm(formula = prestige ~ education + income + women, data = Prestige.z)
Residuals:
     Min       1Q   Median       3Q      Max
-1.15229 -0.30999 -0.00793  0.29984  1.01744

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.396e-17  4.516e-02   0.000     1.00
education    6.640e-01  6.164e-02  10.771  < 2e-16
income       3.242e-01  6.855e-02   4.729 7.58e-06
women       -1.642e-02  5.607e-02  -0.293     0.77

Residual standard error: 0.4561 on 98 degrees of freedom
Multiple R-squared: 0.7982, Adjusted R-squared: 0.792
F-statistic: 129.2 on 3 and 98 DF,  p-value: < 2.2e-16
```


STANDARDIZOVANI REGRESIONI KOEFICIJENTI

```
> summary(prestige.lm.z2 <- lm(prestige ~ -1 + education + income +
+  women, data=Prestige.z))

Call:
lm(formula = prestige ~ -1 + education + income + women, data = Prestige.z)
Residuals:
     Min       1Q   Median       3Q      Max
-1.15229 -0.30999 -0.00793  0.29984  1.01744

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
education  0.66396    0.06133  10.826  < 2e-16
income     0.32418    0.06821   4.753 6.81e-06
women     -0.01642    0.05579  -0.294    0.769

Residual standard error: 0.4538 on 99 degrees of freedom
Multiple R-squared: 0.7982, Adjusted R-squared: 0.7921
F-statistic: 130.5 on 3 and 99 DF,  p-value: < 2.2e-16
```

Univerzitetski centar za primenjenu statistiku

Pitanja

1. U čemu je razlika između dva prethodna modela: `prestige.lm.z` i `prestige.lm.z2`?
2. Šta smo uklonili?
3. Zbog čega smo to smeli da uradim, odnosno, zbog čega nam ova vrednost nije potrebna?

Sadržaj

Univerzitetski centar za primenjenu statistiku

ANCOVA MODELI

- Model u kojem želimo da testiramo efekte i numeričkih (kvantitativnih) i kategoričkih prediktora, istovremeno, zove se opštim imenom kovarijansni model
- Neki autori ovu vrstu analize zovu neformalno i regresiona analiza sa prividnom promenljivom (*dummy-variable regression*)
- Kako inače zovemo kategorijske prediktore?
- Koje su promenljive u tabeli 'Prestige' kategoričke – faktori?

DEFINISANJE ANCOVA MODELA

- Formalna specifikacija ovakvog modela nije baš najočigledniji zadatak
- Uzimajući u obzir da su svi ostali elementi nepromenjeni, osnovno pitanje je kako se to promenila matrica nacrta
- Štaviše, numeričke promenljive (vektori kolona) ostaju nepromenjeni, pa je pitanje kako specifikovati kategorije

Definisanje ANCOVA modela

- Kada prikažemo samo deo koji se odnosi na kategoričke promenljive, onda matrica **X** kojom bismo prikazali faktor sa tri nivoa (ili tri grupe), sa po dva slučaja za svaki nivo izgleda ovako:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- Međutim ovo nije matrica čiji je rang jednak broju kolona; tj. postoji linearna međuzavisnost kolona ove matrice

Definisanje ANCOVA modela

- Svaka kolona matrice može se iskazati kao linearna kombinacija preostalih kolona
- Prema tome, mi ne ocenjujemo model sa 4 već sa najviše 3 parametra
- Geometrijski, prostor modela ima, dakle, jednu dimenziju manje
- Numerički, "defektnost" ranga matrice će se manifestovati u nepostojanju inverza za $\mathbf{X}'\mathbf{X}$
- Rešenje ovog problema je u izostavljanju jedne kolone matrice
- Ovim, matrica dobija puni rang kolona, a prostor modela je nepromenjen i svi potrebni parametri mogu biti ocenjeni

Definisanje ANCOVA modela

- R automatski izostavlja prvu kolonu, pa matrica nacrta izgleda ovako:

$$\mathbf{X}^d = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Definisanje ANCOVA modela

- Moguće je dati opšti obrazac za izostavljanje kolone, ako matricu $\mathbf{X}$ izrazimo u izdeljenom (particionisanom) obliku:

$$\mathbf{X} = [\mathbf{1} : \mathbf{X}_1]$$

gde je $\mathbf{1}$ kolona jedinica

- Na osnovu prethodnog, možemo definisati:

$$\mathbf{X}^d = [\mathbf{1} : \mathbf{X}_1\mathbf{C}_1]$$

pri čemu je $\mathbf{C}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, matrica kontrasta

Sadržaj

Kategorijska promenljiva

- Uzmimo ponovo tabelu ‘Prestige’ i promenljivu ’type’:

```
> attach(Prestige)

The following object(s) are masked from 'package:datasets':
    women

> type

  [1] prof prof prof prof prof prof prof prof prof prof prof prof prof prof prof
 [16] prof prof prof prof prof prof prof prof prof prof prof prof bc   prof prof
 [31] wc   prof wc   <NA> wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc
 [46] wc   wc   wc   wc   wc   wc   wc   <NA> bc   wc   wc   wc   bc   bc   bc
 [61] bc   bc   <NA> bc   bc   bc   <NA> bc   bc   bc   bc   bc   bc   bc   bc
 [76] bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc
 [91] bc   bc   bc   bc   bc   prof bc   bc   bc   bc   bc   bc
Levels: bc prof wc

> levels(type)

[1] "bc"   "prof" "wc"
```

KATEGORIJSKA PROMENLJIVA

- Promenljiva ima nedostajućih podataka i njih možemo ukloniti

```
> detach(Prestige)
> Prestige2 <- na.omit(Prestige)
> attach(Prestige2)

The following object(s) are masked from 'package:datasets':
    women

> type

 [1] prof prof prof prof prof prof prof prof prof prof prof prof prof prof prof
[16] prof prof prof prof prof prof prof prof prof prof prof prof bc   prof prof
[31] wc   prof wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc
[46] wc   wc   wc   wc   wc   wc   bc   wc   wc   wc   bc   bc   bc   bc   bc
[61] bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc
[76] bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc
[91] bc   prof bc   bc   bc   bc   bc   bc
Levels: bc prof wc
```


DEFINISANJE KONTRASTA

- R automatski obavlja kodiranje prividne promenljive (*dummy coding*)

```
> contrasts(type)

     prof wc
bc      0  0
prof    1  0
wc      0  1
```

- Prvi nivo kategorijalne promenljive se uzima za referentni nivo (*reference or baseline*)
- Međutim, sam izbor kategorije je arbitraran i mi ga u R-u možemo lako izmeniti

```
> contrasts(type) <- contr.treatment(levels(type), base=2)
> contrasts(type)

     bc wc
bc    1  0
prof  0  0
wc    0  1
```

Definisanje kontrasta

- R dopušta i druge postupke za kodiranje matrice kontrasta
- Jedan interesantan slučaj jeste Helmertov postupak

```
> contrasts(type) <- NULL
> contrasts(type)

     prof wc
bc      0  0
prof    1  0
wc      0  1

> contrasts(type) <- contr.helmert(levels(type))
> contrasts(type)

     [,1] [,2]
bc     -1   -1
prof    1   -1
wc      0    2
```


Definisanje kontrasta

- Helmertov postupak kontrastira dati nivo sa prosekom “prethodnih”
- Helmertov postupak daje nekorelirane kontraste (ortogonalne) kada je broj opservacija na svim nivoima faktora jednak
- Putem R-ovog sistema za pomoć informišite se koji vam još postupci za kodiranje kontrasta stoje na raspolaganju

Uređeni faktori

- Prema potrebi i kada je to logički prihvatljivo, možemo definisati uređeni faktor u R-u:

```
> type.ord <- ordered(type, levels=c('bc', 'wc', 'prof'))
> type.ord
```

```
 [1] prof prof prof prof prof prof prof prof prof prof prof prof prof prof prof
[16] prof prof prof prof prof prof prof prof prof prof prof prof bc   prof prof
[31] wc   prof wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc   wc
[46] wc   wc   wc   wc   wc   wc   bc   wc   wc   wc   bc   bc   bc   bc   bc
[61] bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc
[76] bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc   bc
[91] bc   prof bc   bc   bc   bc   bc   bc
Levels: bc < wc < prof
```

- Za uređene faktore kontrasti se određuju kao nezavisni (ortogonalni) polinomi – `contr.poly`, kada su nivoi faktora sa jednakim razmacima
- Red polinoma je za jedan manji od broja nivoa

```
> round(contrasts(type.ord), 3)
```

```
          .L     .Q
[1,] -0.707  0.408
[2,]  0.000 -0.816
[3,]  0.707  0.408
```


Plauzibilni kovarijansni model

- Uključimo sada varijablu `type` u naš prethodni višesmerni model sa prestižom zanimanja, y i prediktorima:
  - godine obrazovanja, x_1
  - prosečni prihodi, x_2
  - tip zanimanja, definisan preko prividnih promenljivih, d_1 i d_2
- Kovarijansni model je definisan sledećim izrazom:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 d_{i1} + \gamma_2 d_{i2} + \epsilon_i$$

Osnovni rezultati funkcije lm()

```
> prestige.ancova1 <- lm(prestige ~ education + income + type,
+  data=Prestige2)

> summary(prestige.ancova1)

Call:
lm(formula = prestige ~ education + income + type, data = Prestige2)
Residuals:
     Min       1Q   Median       3Q      Max
-14.9529  -4.4486   0.1678   5.0566  18.6320

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.6229292  5.2275255  -0.119    0.905
education    3.6731661  0.6405016   5.735 1.21e-07
income       0.0010132  0.0002209   4.586 1.40e-05
typeprof     6.0389707  3.8668551   1.562    0.122
typewc      -2.7372307  2.5139324  -1.089    0.279

Residual standard error: 7.095 on 93 degrees of freedom
Multiple R-squared: 0.8349, Adjusted R-squared: 0.8278
F-statistic: 117.5 on 4 and 93 DF,  p-value: < 2.2e-16
```


Testiranje efekta za faktore

- t vrednosti u rezultatima regresione analize su sasvim adekvatni za prediktore (efekte) sa jednim stepenom slobode ($df = 1$)
- Međutim, kako je referentna kategorija i način prividnog kodiranja potpuno arbitraran, istraživači obično traže ukupni efekat za takvu varijablu

```
> anova(prestige.ancova1)

Analysis of Variance Table
Response: prestige
          Df  Sum Sq Mean Sq  F value    Pr(>F)
education  1 21282.5 21282.5 422.8056 < 2.2e-16
income     1  1792.0  1792.0  35.5999 4.355e-08
type       2   591.2   295.6   5.8721  0.003966
Residuals 93  4681.3    50.3
```

Testiranje efekta za faktore

- Funkcija `anova()` zapravo daje rezultate tzv. sekvencijalnog testiranja
- Za korelirane (neortogonalne) podatke, kada su koeficijenti različitih članova u modelu korelirani, sekvencijalni test ne odgovara pretpostavkama o parametrima
- U suštini, postoje različiti postupci izračunavanja tipova sume kvadrata
- U paketu CAR postoji prigodna funkcija `Anova` koja daje simultani test (Tip II ili Tip III)

```
> Anova(prestige.ancova1)

Anova Table (Type II tests)
Response: prestige
          Sum Sq Df F value    Pr(>F)
education 1655.5  1 32.8882 1.205e-07
income    1058.8  1 21.0339 1.405e-05
type       591.2  2  5.8721  0.003966
Residuals 4681.3 93
```


Još nešto o funkciji `anova()`

- Funkcija `anova()` se može koristiti i za poređenje alternativnih modela, kada se F-testom kontrastiraju ugnežđeni linearni modeli

```
> prestige.ancova0 <- lm(prestige ~ education + income,
+   data=Prestige2)
> anova(prestige.ancova0, prestige.ancova1)

Analysis of Variance Table
Model 1: prestige ~ education + income
Model 2: prestige ~ education + income + type
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     95 5272.4
2     93 4681.3  2    591.16 5.8721 0.003966
```

I JOŠ NEŠTO O ANALIZI KOVARIJANSE

- Izostavljanje regresione konstante – odsečka, ne daje smislene rezultate
- Na ovaj način, mi bismo forsirali R da udešava intercept za svaku grupu posebno
- Hipoteza da ne postoji efekat varijable (npr., 'type') postaje hipoteza da su koeficijenti za svaki nivo varijable jednaki nuli
- Konačno, R^2 i F-test nemaju svoje uobičajeno značenje, odnosno, interpretaciju

SADRŽAJ

ANOVA MODELI

- Analiza varijanse, zapravo, nema posebni status u porodici linearnih modela
- Važno je samo zapamtiti da ona pruža osnovu za lakše (i bolje) razumevanje interakcije dva ili više faktora, kao i načina na koji je se vrši ispitivanje interakcije u R-u

- Na ovo pitanje vratićemo se nešto kasnije . . .

PODACI ZA JEDAN ANOVA MODEL

- Uzmimo podatke u kojima su prikazani rezultati o ponašanju pacova u lavirintu:

```
> rats = read.table('data/rats.txt', header=TRUE, sep='\t')
> head(rats)

  envirnm strain errors
1    FREE BRIGHT     26
2    FREE BRIGHT     14
3    FREE BRIGHT     41
4    FREE BRIGHT     16
5    FREE  MIXED     41
6    FREE  MIXED     82

> str(rats)

'data.frame': 24 obs. of  3 variables:
 $ envirnm: Factor w/ 2 levels "FREE","RESTRCTD": 1 1 1 1 1 1 1 1 1 1 ...
 $ strain : Factor w/ 3 levels "BRIGHT","DULL",..: 1 1 1 1 3 3 3 3 2 2 ...
 $ errors : num  26 14 41 16 41 82 26 86 36 87 ...
```

KAKO "ŽIVE" PACOVI U LAVIRINTU?

- Ono što ovde želimo da saznamo jeste da li uslovi (*environment*) u kojima pacovi žive, kao i njihova priroda (*strain*), utiču na njihovu uspešnost u "rešavanju" lavirinta
- Možemo, najpre, uporediti prosečne uspehe za svaku grupu pacova:

```
> attach(rats)
> envirnm <- factor(envirnm, levels=c('RESTRCTD', 'FREE'))
> strain <- factor(strain, levels=c('DULL', 'MIXED', 'BRIGHT'))

> means <- tapply(errors, list(strain, envirnm), mean)
> means

        RESTRCTD  FREE
DULL       97.75 65.25
MIXED      87.25 58.75
BRIGHT     54.50 24.25
```


KAKO "ŽIVE" PACOVI U LAVIRINTU?

```
> tapply(errors, list(strain, envirnm), function(x) sqrt(var(x)))

        RESTRCTD     FREE
DULL   41.12076 32.43840
MIXED  33.40035 29.83706
BRIGHT 28.61818 12.33896

> tapply(errors, list(strain, envirnm), length)

        RESTRCTD FREE
DULL           4    4
MIXED          4    4
BRIGHT         4    4
```

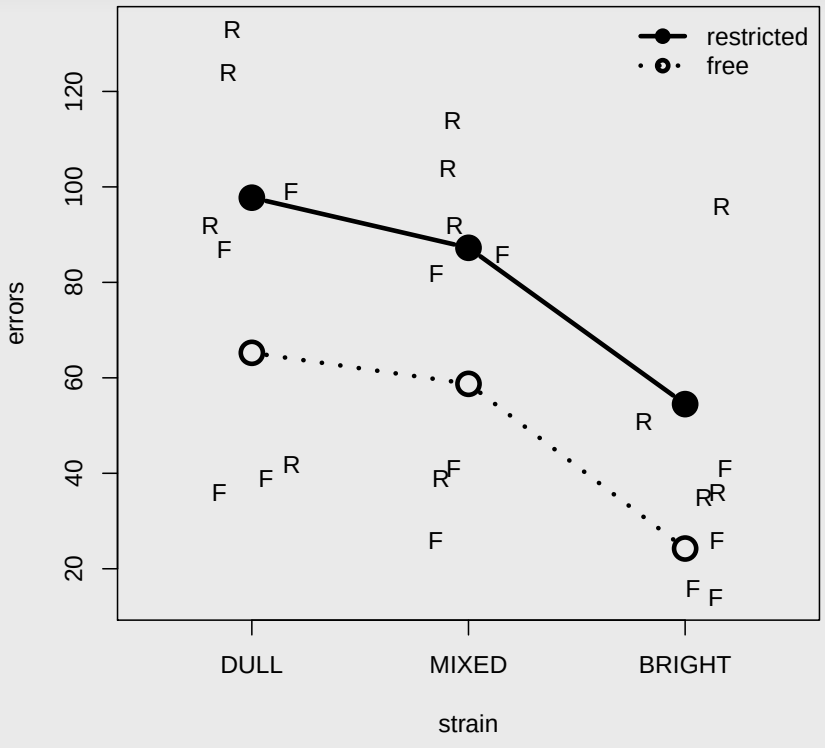
- Prvi poziv `taply()` je pomalo neobičan
- To je samo demonstracija kako da pozovete funkciju koju sami možete definisati
- `function(x) sqrt(var(x))` Specifikuje korisničku (i privremenu) funkciju, koja uzima parametar `x`, a to je u konkretnom slučaju promenljiva "errors"
- `tapply(errors, list(strain, envirnm), sd)` daje identičan rezultat

KAKO "ŽIVE" PACOVI U LAVIRINTU?

- Konačno, mnogo bolji pregled imamo ako rezultate prikažemo grafički:

```
> Strain <- as.numeric(strain)
> plot(c(0.5, 3.5), range(errors), xlab='strain',
+     ylab='errors', type='n', axes=F)
> axis(1, at=1:3, labels=c('DULL', 'MIXED', 'BRIGHT'))
> axis(2)
> points(jitter(Strain[envirnm=='RESTRCTD']),
+     errors[envirnm=='RESTRCTD'], pch='R')
> points(jitter(Strain[envirnm=='FREE']),
+     errors[envirnm=='FREE'],, pch='F')
> lines(1:3, means[,1], lty=1, lwd=3, type='b', pch=19, cex=2)
> lines(1:3, means[,2], lty=3, lwd=3, type='b', pch=1, cex=2)
> legend('topright', c('restricted','free'), bty='n', lty=c(1,3),
+  lwd=c(3,3), pch=c(19,1))
> box()
```


PROSEČNI BROJ GREŠAKA U LAVIRINTU



Analiza grafikona

- Kako ćemo obaviti analizu dobijenog grafikona?
- Postoje li potencijalne “opasnosti” u podacima?

```
> identify(Strain, errors)
```


Anova za tabelu ‘rats’

```
> rats.aov <- lm(errors ~ envirnm + strain, data=rats)
> summary(rats.aov)

Call:
lm(formula = errors ~ envirnm + strain, data = rats)
Residuals:
    Min      1Q  Median      3Q     Max
-54.708 -18.833  -0.875  24.604  41.417

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)        24.17      11.96   2.020  0.05699
envirnmRESTRCTD    30.42      11.96   2.542  0.01939
strainDULL         42.12      14.65   2.875  0.00936
strainMIXED        33.62      14.65   2.295  0.03270

Residual standard error: 29.31 on 20 degrees of freedom
Multiple R-squared: 0.4399, Adjusted R-squared: 0.3559
F-statistic: 5.236 on 3 and 20 DF,  p-value: 0.007868

> anova(rats.aov)

Analysis of Variance Table
Response: errors
          Df  Sum Sq Mean Sq F value  Pr(>F)
envirnm    1  5551.0  5551.0  6.4634 0.01939
strain     2  7939.7  3969.9  4.6224 0.02238
Residuals 20 17176.8   858.8
```

Sažeto i najvažnije o analizi varijanse

- Cilj ANOVE jeste testiranje hipoteze o razlikama između proseka uslova (*cell means*) i ukupnog proseka
- Kodiranje kontrasta i izračunavanje sume kvadrata mora biti u skladu sa postavljenim hipotezama
- Određeni problemi mogu nastati kada podaci (odnosno, nacrt) nisu balansirani
- Ako se pridržavamo principa marginaliteta, šanse za grešku su minimalne
- Takav pristup dovodi do sume kvadrata Tip II: ignoriši efekte viěg reda dok testiraš efekte nižeg reda
- Postupno (inkrementirano) testiranje sume kvadrata (Tip III), testira efekat nakon što su ostali uzeti u obzir (npr., glavni efekat faktora A, iz proseka nivoa drugog faktora B)

Sažeto i najvažnije o analizi varijanse

- Razmislite o razlikama F-testova s obzirom na Tip sume kvadrata:
  - Anova(rats.aov, type='II')
  - Anova(rats.aov, type='III')
- Ima li razlike između ova dva tipa sume kvadrata?
- A u odnosu na onaj koji daje funkcija `anova()`

- Funkcija `aov()` nam štedi vreme, sprovodi analizu varijanse pozivajući funkciju `lm()`

Sadržaj

- funkcije za demonstraciju:

```
> source('data/demoUvodLM.R')
```

  - `regConst()`
    - min → minimum za varijablu X
    - max → maksimum za varijablu X
    - a → odsečak na Y-osi
    - b → nagib regresione funkcije
    - it → broj pokušaja (iteracija)
    - m → aritmetička sredina za grešku (epsilon)
    - s → standardna devijacija za grešku (epsilon)

- funkcije za demonstraciju:

```
> source('data/demoUvodLM.R')
```

  - `regVar()`
    - min → minimum za varijablu X
    - max → maksimum za varijablu X
    - a → odsečak na Y-osi
    - b → nagib regresione funkcije
    - it → broj pokušaja (iteracija)
    - m → aritmetička sredina za grešku (epsilon)
    - s → standardna devijacija za grešku (epsilon)
    - g → koeficijent priraštaja greške (epsilon) u funkciji X

- funkcije za demonstraciju:

```
> source('data/demoUvodLM.R')
```

  - `regCorr()`
    - n → veličina uzorka
    - m → aritmetičke sredine za dve varijable
    - r → koeficijent korelacije
    - it → broj pokušaja (iteracija)

K R A J