Universitat Oberta de Catalunya

uoc.edu

Modelització de dades longitudinals amb efectes aleatoris de intercepció/pendent i presència de dades perdudes en el context dels estudis d'estabilitat – ANNEX 5

Estudiant: Francesc Bernad Martin
M0.178 TFM-Estadística i Bioinformàtica 2
Màster universitari en Bioinformàtica i bioestadística UOC-UB
Àrea: Bioestadística / Bioinformàtica

Nom Consultor/a: Nuria Perez Alvarez
Nom Professor/a responsable de l'assignatura: Carles Ventura Royo

Barcelona, 04 de Juny de 2019

Resums R dels models simple i d'interaccions de l'anàlisi pràctic LAKE1

Índex

1 RESULTATS DELS MODELS SIMPLES

```
# Fit the full model
Lake1.fm0 <- lm(CargaViral ~ . - LogCargaViral - CargaViral_t - Ident, data = LAKE1)
# Stepwise regression model
Lake1.step.model0 <- stepAIC(Lake1.fm0, direction = "both",
                     trace = F)
(Lake1.step.model0s <- summary(Lake1.step.model0))
##
## Call:
## lm(formula = CargaViral ~ Tiempo + CD4P + Cloro, data = LAKE1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -113727  -55558   -7185   42815  276156
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   910289     478303   1.903   0.0650 .
## Tiempo         -1868        958  -1.950   0.0591 .
## CD4P           -4985       2505  -1.990   0.0542 .
## Cloro          -6948       4748  -1.463   0.1521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 81390 on 36 degrees of freedom
## Multiple R-squared: 0.4226,
## Adjusted R-squared: 0.3745
## F-statistic: 8.783 on 3 and 36 DF,  p-value: 0.0001673
anova(Lake1.step.model0)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Tiempo	1	119689303289	119689303289	18.069892	0.0001440645
CD4P	1	40649071089	40649071089	6.136926	0.0180686480
Cloro	1	14180720145	14180720145	2.140911	0.1520937398
Residuals	36	238452714894	6623686525	*NA*	*NA*

4 rows

```
# Fit the full model
Lake1.fm1 <- lm(LogCargaViral ~ . - CargaViral - CargaViral_t - Ident, data = LAKE1)
# Stepwise regression model
Lake1.step.model1 <- stepAIC(Lake1.fm1, direction = "both",
                     trace = F)
(Lake1.step.model1s <- summary(Lake1.step.model1))
##
## Call:
## lm(formula = LogCargaViral ~ Tiempo, data = LAKE1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4563 -1.6775 -0.0412  1.6012  4.3657
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.7566     0.5966  14.678  < 2e-16 ***
## Tiempo       -0.1343     0.0203  -6.616 8.16e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## s: 2.178 on 38 degrees of freedom
## Multiple R-squared: 0.5353,
## Adjusted R-squared: 0.5231
## F-statistic: 43.78 on 1 and 38 DF,  p-value: 8.162e-08
# Fit the full model
Lake1.fm2 <- lm(CargaViral_t ~ . - CargaViral - LogCargaViral - Ident, data = LAKE1)
# Stepwise regression model
Lake1.step.model2 <- stepAIC(Lake1.fm2, direction = "both",
                      trace = F)
(Lake1.step.model2s <- summary(Lake1.step.model2))
##
## Call:
## lm(formula = CargaViral_t ~ Tiempo, data = LAKE1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -647.6 -238.6   11.4  272.2  352.4
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7834.432     81.191  96.494  < 2e-16 ***
## Tiempo       -21.784      2.762  -7.887 1.61e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 296.5 on 38 degrees of freedom
## Multiple R-squared: 0.6208,
## Adjusted R-squared: 0.6108
## F-statistic:  62.2 on 1 and 38 DF,  p-value: 1.606e-09
```

2 RESULTATS DELS MODELS AMB INTERACCIONS

```
#Original
Inter.0 <- attr(Lake1.step.model0s$terms,"term.labels")
Lake1.step.model0int.s  <-  summary(Lake1.step.model0int  <-  update(Lake1.step.model0,paste(".  ~  .
+",paste(Inter.0,collapse="*"),sep=" ")))
(Lake1.step.model0int2.s2   <-   summary(Lake1.step.model0int2   <-   stepAIC(Lake1.step.model0int,
direction = "both", trace = F)))
##
## Call:
## lm(formula = CargaViral ~ Tiempo + CD4P + Cloro + Tiempo:CD4P +
##     Tiempo:Cloro + CD4P:Cloro + Tiempo:CD4P:Cloro, data = LAKE1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -73263 -18903  -1618  24695 147784
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6447992.94 1022547.10   6.306 4.49e-07 ***
## Tiempo            -176830.68   58543.48  -3.021  0.00493 **
## CD4P              -343452.68   74069.49  -4.637 5.69e-05 ***
## Cloro              -60755.43   10234.76  -5.936 1.31e-06 ***
## Tiempo:CD4P          9148.16    2830.75   3.232  0.00285 **
## Tiempo:Cloro         1649.19     574.98   2.868  0.00725 **
## CD4P:Cloro           3282.16     732.55   4.480 8.93e-05 ***
## Tiempo:CD4P:Cloro     -86.72      27.74  -3.127  0.00375 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 47430 on 32 degrees of freedom
## Multiple R-squared: 0.8257,
## Adjusted R-squared: 0.7875
## F-statistic: 21.65 on 7 and 32 DF,  p-value: 1.856e-10
(Lake1.step.model0int2.an2<-Anova(Lake1.step.model0int2))
```

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
Tiempo	57912478201	1	25.739390229	1.605279e-05
CD4P	18872847563	1	8.388098787	6.757372e-03
Cloro	21648029754	1	9.621537583	3.998059e-03
Tiempo:CD4P	28046007872	1	12.465139869	1.281942e-03
Tiempo:Cloro	4601747	1	0.002045261	9.642094e-01
CD4P:Cloro	23259790306	1	10.337889828	2.975084e-03
Tiempo:CD4P:Cloro	21993544225	1	9.775102619	3.750561e-03
Residuals	71998570517	32	*NA*	*NA*

8 rows

```
(Lake1.step.model0int2.an1_2<-anova(Lake1.step.model0int2,Lake1.step.model0))
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	32	71998570517	*NA*	*NA*	*NA*	*NA*
2	36	238452714894	-4	-166454144377	18.49527	5.807627e-08

2 rows

```
if (Lake1.step.model0int2.an1_2$`Pr(>F)`[2]>0.05){
  intM0 <- Lake1.step.model0
  intM0s <- Lake1.step.model0s
} else {
  intM0 <- Lake1.step.model0int2
  intM0s <- Lake1.step.model0int2.s2
}
  intM0col <- Lake1.step.model0col
  intM0cols <- Lake1.step.model0col.s
```

En el cas dels models transformats no seria necessari en principi ja que només ha quedat un predictor com a model òptim i per tant no té altres predictors per creuar-se. Tot i així es fa la prova on apareixen totes les variables amb interaccions per comprobar la diferència entre models:

```
#Log
#Step1
Lake1.step.model1int.s  <-  summary(Lake1.step.model1int  <-  update(Lake1.step.model1,paste(".  ~  .
+",paste(Inter.0,collapse="*"),sep=" ")))
(Lake1.step.model1int2.s2   <-   summary(Lake1.step.model1int2   <-   stepAIC(Lake1.step.model1int,
direction = "both", trace = F)))
##
## Call:
## lm(formula = LogCargaViral ~ Tiempo + CD4P + Tiempo:CD4P, data = LAKE1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7900 -1.0384 -0.2722  0.7064  4.2465
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.642765   1.468711   9.289 4.29e-11 ***
## Tiempo      -0.350318   0.069405  -5.047 1.30e-05 ***
## CD4P        -0.272305   0.078710  -3.460  0.00141 **
## Tiempo:CD4P  0.010529   0.002956   3.562  0.00106 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 1.89 on 36 degrees of freedom
## Multiple R-squared: 0.6687,
## Adjusted R-squared: 0.6411
## F-statistic: 24.22 on 3 and 36 DF,  p-value: 9.352e-09
(Lake1.step.model1int2.an2<-Anova(Lake1.step.model1int2))
```

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
Tiempo	97.671197	1	27.349144	7.435470e-06
CD4P	6.459032	1	1.808609	1.870872e-01
Tiempo:CD4P	45.305808	1	12.686187	1.058997e-03
Residuals	128.565744	36	*NA*	*NA*

4 rows

```
(Lake1.step.model1int2.an1_2 <- anova(Lake1.step.model1int2,Lake1.step.model1))
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	36	128.5657	*NA*	*NA*	*NA*	*NA*
2	38	180.3306	-2	-51.76484	7.247398	0.002264675

2 rows

```
if (Lake1.step.model1int2.an1_2$`Pr(>F)`[2]>0.05){
```

```
  intM1 <- Lake1.step.model1
  intM1s <- Lake1.step.model1s
} else {
  intM1 <- Lake1.step.model1int2
  intM1s <- Lake1.step.model1int2.s2
}
#Box-Cox
#Step1
Lake1.step.model2int.s  <-  summary(Lake1.step.model2int  <-  update(Lake1.step.model2,paste(".  ~  .
+",paste(Inter.0,collapse="*"),sep=" ")))
(Lake1.step.model2int2.s2   <-   summary(Lake1.step.model2int2   <-   stepAIC(Lake1.step.model2int,
direction = "both", trace = F)))
##
## Call:
## lm(formula = CargaViral_t ~ Tiempo + CD4P + Tiempo:CD4P, data = LAKE1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -631.26 -157.19   31.58  187.95  448.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8267.5540   215.9970  38.276  < 2e-16 ***
## Tiempo       -47.3143    10.2070  -4.635 4.55e-05 ***
## CD4P         -23.5364    11.5756  -2.033   0.0494 *
## Tiempo:CD4P    1.1663     0.4347   2.683   0.0110 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 277.9 on 36 degrees of freedom
## Multiple R-squared: 0.6843,
## Adjusted R-squared: 0.6579
## F-statistic: 26.01 on 3 and 36 DF,  p-value: 3.979e-09
(Lake1.step.model2int2.an2<-Anova(Lake1.step.model2int2))
```

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
Tiempo	3308923.010	1	42.83920038	1.306654e-07
CD4P	3352.261	1	0.04340027	8.361476e-01
Tiempo:CD4P	555922.233	1	7.19728559	1.095743e-02
Residuals	2780659.473	36	*NA*	*NA*

4 rows

```
(Lake1.step.model2int2.an1_2 <- anova(Lake1.step.model2int2,Lake1.step.model2))
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	36	2780659	*NA*	*NA*	*NA*	*NA*
2	38	3339934	-2	-559274.5	3.620343	0.03692995

2 rows

```
if (Lake1.step.model2int2.an1_2$`Pr(>F)`[2]>0.05){
  intM2 <- Lake1.step.model2
  intM2s <- Lake1.step.model2s
} else {
  intM2 <- Lake1.step.model2int2
  intM2s <- Lake1.step.model2int2.s2
}
```