

Exemplos de código do **R** usados na análise

Wagner Hugo Bonat
Henrique Silva Dallazuanna

Vamos começar lendo o arquivo de dados:

```
> dados <- read.table("correto.txt", header = T)
> bairros <- split(dados, dados$SITIO)
> BT <- bairros[[1]]
> names(BT)

[1] "DATA_OBS"          "TIPO_IMOVEL"      "QUINTAL"
[4] "SOMBRA"            "res.pq"           "AGUA_REDE"
[7] "FREQ_ABASTECIMENTO" "CANALIZADA_COMODO" "res.grd.com"
[10] "res.grd.sem"       "res.peq.com"      "res.peq.sem"
[13] "NRO_OVOS"          "GRUPO"            "SITIO"
[16] "PRECIPM.1"         "PRECIPM.2"        "PRECIPM.3"
[19] "PRECIPM.4"         "PRECIPM.5"        "PRECIPM.6"
[22] "PRECIPM.7"         "PRECIPM.8"        "PRECIPM.9"
[25] "PRECIPM.10"        "PRECIPM.11"       "PRECIPM.12"
[28] "UMIDM.1"           "UMIDM.2"          "UMIDM.3"
[31] "UMIDM.4"           "UMIDM.5"          "UMIDM.6"
[34] "UMIDM.7"           "UMIDM.8"          "UMIDM.9"
[37] "UMIDM.10"          "UMIDM.11"         "UMIDM.12"
[40] "MINM.1"            "MINM.2"           "MINM.3"
[43] "MINM.4"            "MINM.5"           "MINM.6"
[46] "MINM.7"            "MINM.8"           "MINM.9"
[49] "MINM.10"           "MINM.11"          "MINM.12"
[52] "MAXM.1"            "MAXM.2"           "MAXM.3"
[55] "MAXM.4"            "MAXM.5"           "MAXM.6"
[58] "MAXM.7"            "MAXM.8"           "MAXM.9"
[61] "MAXM.10"           "MAXM.11"          "MAXM.12"
[64] "coords.x1"         "coords.x2"        "COD_ARMADILHA"
```

Temos que transformar as covariáveis categóricas em fatores.

```
> BT$TIPO_IMOVEL <- as.factor(BT$TIPO_IMOVEL)
> BT$QUINTAL <- as.factor(BT$QUINTAL)
> BT$res.pq <- as.factor(BT$res.pq)
> BT$AGUA_REDE <- as.factor(BT$AGUA_REDE)
> BT$FREQ_ABASTECIMENTO <- as.factor(BT$FREQ_ABASTECIMENTO)
> BT$CANALIZADA_COMODO <- as.factor(BT$CANALIZADA_COMODO)
> BT$res.grd.com <- as.factor(BT$TIPO_IMOVEL)
> BT$res.grd.sem <- as.factor(BT$TIPO_IMOVEL)
```

```
> BT$res.peq.com <- as.factor(BT$TIPO_IMOVEL)
> BT$res.peq.sem <- as.factor(BT$TIPO_IMOVEL)
```

Para a construção dos fatores climáticos através da análise fatorial, é necessário a determinação do número de fatores e o tipo de rotação que vamos usar, no caso do fator unidade o número de fatores foi definido como 4 e no caso da temperatura foi definido 3, a rotação utilizada para compor os fatores foi a promax.

```
> fac.umi <- factanal(as.formula(paste("~", paste(grep("PRECIP|UMID",
+   names(dados), value = T), collapse = "+"))), 4, data = BT,
+   rotation = "promax")
> fac.list.umi <- split(apply(fac.umi$loadings, 1, max), apply(fac.umi$loadings,
+   1, which.max))
> fac.list.umi <- fac.list.umi[order(unlist(lapply(fac.list.umi,
+   function(x) as.numeric(gsub(patt = "[a-Z]|\\". ", rep = "",
+   names(x)[1])))))]
> names(fac.list.umi) <- 1:4
> fac.list.umi
```

```
$`1`
PRECIPM.1 PRECIPM.2 PRECIPM.3  UMIDM.1  UMIDM.2  UMIDM.3
0.8406929 0.7907277 0.5497182 0.8512350 0.7575812 0.5167023
```

```
$`2`
PRECIPM.4 PRECIPM.5 PRECIPM.6  UMIDM.4  UMIDM.5
0.6503820 0.7002525 0.5706700 0.6303387 0.5662656
```

```
$`3`
PRECIPM.7 PRECIPM.8 PRECIPM.9  UMIDM.6  UMIDM.7  UMIDM.8
0.7189305 0.7535654 0.6244347 0.5463465 0.6522713 0.5695280
```

```
$`4`
PRECIPM.10 PRECIPM.11 PRECIPM.12  UMIDM.9  UMIDM.10  UMIDM.11  UMIDM.12
0.5042061 0.7352809 0.8492636 0.5587580 0.7836866 0.9336651 0.9179453
```

Até aqui obteve-se as cargas fatoriais, porém para utiliza-las a função abaixo as padroniza para estarem no intervalo 0 a 1. Além disso para terminar a construção dos fatores cada carga tem que ser multiplicada pelo sua devida covariável, a função abaixo é usada para fazer isto:

```
> f <- function(x) {
+   x <- x/sum(x)
+   rowSums(sweep(BT[names(x)], 2, x, FUN = "*"))
+ }
```

Usando a função para padronizar as cargas e compor os fatores:

```
> fac.umid <- sapply(fac.list.umi, f)
> dim(fac.umid)
```

```
[1] 2580 4
```

O mesmo procedimento é feito para a construção dos fatores referentes a temperatura.

```
> fac.temp <- factanal(as.formula(paste("~", paste(grep("MAX|MIN",
+   names(dados), value = T), collapse = "+"))), 3, data = BT,
+   rotation = "promax")
> fac.list.temp <- split(apply(fac.temp$loadings, 1, max), apply(fac.temp$loadings,
+   1, which.max))
> fac.list.temp <- fac.list.temp[order(unlist(lapply(fac.list.temp,
+   function(x) as.numeric(gsub(patt = "[a-Z]|\\.", rep = "",
+   names(x)[1])))))]
> names(fac.list.temp) <- 1:3
> fac.list.temp

$`1`
  MINM.1  MINM.2  MINM.3  MINM.4  MAXM.1  MAXM.2  MAXM.3  MAXM.4
0.7697783 0.8451731 0.7576782 0.5878682 0.9563984 0.9662133 0.8193940 0.6642899

$`2`
  MINM.5  MINM.6  MINM.7  MINM.8  MAXM.5  MAXM.6  MAXM.7
0.3933101 0.5514749 0.8087383 0.5295014 0.4859633 0.6312209 0.5440980

$`3`
  MINM.9  MINM.10  MINM.11  MINM.12  MAXM.8  MAXM.9  MAXM.10  MAXM.11
0.5337888 0.7623467 0.8629555 0.9697453 0.4165318 0.5961280 0.7655081 0.8820747
  MAXM.12
0.8490514
```

Construindo o fator temperatura.

```
> fac.temp <- sapply(fac.list.temp, f)
> head(fac.temp)

      1      2      3
1 27.59949 26.27974 26.87626
2 27.70973 26.71035 27.43842
3 28.15205 27.09293 26.07249
4 28.18552 26.84908 27.61092
5 27.71967 26.33368 27.31212
6 27.39602 26.35193 26.65828
```

Criando um arquivo .txt com os fatores umidade e temperatura.

```
> nomes <- c("DATA_OBS", "TIPO_IMOVEL", "QUINTAL", "SOMBRA", "res.pq",
+   "AGUA_REDE", "FREQ_ABASTECIMENTO", "CANALIZADA_COMODO", "res.grd.com",
+   "res.grd.sem", "res.peq.com", "res.peq.sem", "NRO_OVOS",
+   "GRUPO", "SITIO", "coords.x1", "coords.x2", "COD_ARMADILHA",
+   "UMIDADE.1", "UMIDADE.2", "UMIDADE.3", "UMIDADE.4", "TEMP.1",
+   "TEMP.2", "TEMP.3")
> BT <- data.frame(BT, fac.umid, fac.temp)
> BT <- BT[, -c(16:63)]
> names(BT) <- nomes
> write.table(BT, "BT.txt")
```

Com o arquivo pronto podemos proceder o ajuste dos modelos.

```
> require(MASS)
> require(mgcv)
```

This is mgcv 1.3-31

```
> BT$SEQ <- rep(1:length(table(BT$DATA_OBS)), table(BT$DATA_OBS))
```

Modelo completo covariáveis.

```
> mod.tudo <- glm(NRO_OVOS ~ AGUA_REDE + res.pq + CANALIZADA_COMODO +
+ TIPO_IMOVEL + QUINTAL + FREQ_ABASTECIMENTO + res.grd.com +
+ res.grd.sem + res.peq.com + res.peq.sem, family = negative.binomial(1),
+ data = BT)
> summary(mod.tudo)
```

Call:

```
glm(formula = NRO_OVOS ~ AGUA_REDE + res.pq + CANALIZADA_COMODO +
TIPO_IMOVEL + QUINTAL + FREQ_ABASTECIMENTO + res.grd.com +
res.grd.sem + res.peq.com + res.peq.sem, family = negative.binomial(1),
data = BT)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8867	-0.9077	-0.3503	0.3227	4.8460

Coefficients: (4 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.52082	0.16841	44.658	< 2e-16 ***
AGUA_REDE1	-0.31768	0.09657	-3.290	0.001019 **
res.pq1	-0.17084	0.05059	-3.377	0.000746 ***
CANALIZADA_COMODO1	-0.12299	0.06527	-1.884	0.059654 .
TIPO_IMOVEL1	0.01480	0.13347	0.111	0.911697
QUINTAL1	-0.03791	0.05763	-0.658	0.510749
FREQ_ABASTECIMENTO1	0.03196	0.16059	0.199	0.842258
res.grd.com1	NA	NA	NA	NA
res.grd.sem1	NA	NA	NA	NA
res.peq.com1	NA	NA	NA	NA
res.peq.sem1	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1) family taken to be 1.078863)

Null deviance: 2349.4 on 2201 degrees of freedom
Residual deviance: 2323.7 on 2195 degrees of freedom
(378 observations deleted due to missingness)
AIC: 37099

Number of Fisher Scoring iterations: 8

Modelo selecionado covariáveis.

```
> mod.final <- gam(NRO_OVOS ~ AGUA_REDE + res.pq + CANALIZADA_COMODO,  
+ family = negative.binomial(1), data = BT)  
> summary(mod.final)
```

Family: Negative Binomial(0.9343)
Link function: log

Formula:

NRO_OVOS ~ AGUA_REDE + res.pq + CANALIZADA_COMODO

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.53999	0.03551	212.350	< 2e-16 ***
AGUA_REDE1	-0.32154	0.09577	-3.358	0.000800 ***
res.pq1	-0.16110	0.04638	-3.474	0.000524 ***
CANALIZADA_COMODO1	-0.12833	0.06417	-2.000	0.045651 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.00717 Deviance explained = 1.07%
GCV score = 1.0018 Scale est. = 1 n = 2202

Modelo completo climáticas

```
> mod.cli <- gam(NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + TEMP.1 + TEMP.2 +  
+ UMIDADE.1 + UMIDADE.4 + TEMP.3, family = negative.binomial(1),  
+ data = BT)  
> summary(mod.cli)
```

Family: Negative Binomial(0.9357)
Link function: log

Formula:

NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + TEMP.1 + TEMP.2 + UMIDADE.1 +
UMIDADE.4 + TEMP.3

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.181644	1.785724	1.782	0.0749 .
UMIDADE.2	0.031902	0.006906	4.619	4.07e-06 ***
UMIDADE.3	0.013667	0.007288	1.875	0.0609 .
TEMP.1	0.081821	0.059865	1.367	0.1718
TEMP.2	-0.048021	0.050960	-0.942	0.3461
UMIDADE.1	-0.010830	0.006652	-1.628	0.1036
UMIDADE.4	0.021290	0.008667	2.456	0.0141 *
TEMP.3	0.042781	0.049057	0.872	0.3833

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0482 Deviance explained = 5.16%
GCV score = 1.0036 Scale est. = 1 n = 2202

Modelo selecionado climaticas

```
> mod.final <- gam(NRO_OVOS ~ UMIDADE.2 + UMIDADE.3, family = negative.binomial(1),  
+ data = BT)  
> summary(mod.final)
```

Family: Negative Binomial(0.9126)
Link function: log

Formula:
NRO_OVOS ~ UMIDADE.2 + UMIDADE.3

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.852136	0.158734	36.867	< 2e-16 ***
UMIDADE.2	0.022371	0.005639	3.967	7.50e-05 ***
UMIDADE.3	0.022456	0.005667	3.963	7.64e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0438 Deviance explained = 4.7%
GCV score = 1.0014 Scale est. = 1 n = 2202

Modelo combinado

```
> mod.comb <- glm(NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + AGUA_REDE +  
+ res.pq + CANALIZADA_COMODO, family = negative.binomial(1),  
+ data = BT)  
> summary(mod.comb)
```

Call:

```
glm(formula = NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + AGUA_REDE +  
res.pq + CANALIZADA_COMODO, family = negative.binomial(1),  
data = BT)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8631	-0.8468	-0.3078	0.3073	4.9097

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.947023	0.156534	37.992	< 2e-16 ***
UMIDADE.2	0.022289	0.005478	4.068	4.90e-05 ***
UMIDADE.3	0.023176	0.005504	4.210	2.65e-05 ***
AGUA_REDE1	-0.345741	0.094119	-3.673	0.000245 ***

```
res.pq1          -0.170111    0.045584   -3.732 0.000195 ***
CANALIZADA_COMODO1 -0.128888    0.063077   -2.043 0.041139 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for Negative Binomial(1) family taken to be 1.033750)

```
Null deviance: 2349.4 on 2201 degrees of freedom
Residual deviance: 2210.8 on 2196 degrees of freedom
(378 observations deleted due to missingness)
AIC: 36984
```

Number of Fisher Scoring iterations: 6

Ajuste do modelo aditivo generalizado.

```
> mod.temp <- gam(NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + AGUA_REDE +
+   res.pq + CANALIZADA_COMODO + s(coords.x1, coords.x2) + s(SEQ,
+   bs = "cr"), family = negative.binomial(1), data = BT)
> summary(mod.temp)
```

```
Family: Negative Binomial(1.1171)
Link function: log
```

Formula:

```
NRO_OVOS ~ UMIDADE.2 + UMIDADE.3 + AGUA_REDE + res.pq + CANALIZADA_COMODO +
s(coords.x1, coords.x2) + s(SEQ, bs = "cr")
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.809273	0.146965	39.528	< 2e-16 ***
UMIDADE.2	0.019860	0.005125	3.875	0.000110 ***
UMIDADE.3	0.027932	0.005143	5.431	6.23e-08 ***
AGUA_REDE1	-0.254891	0.107392	-2.373	0.017709 *
res.pq1	-0.153729	0.052282	-2.940	0.003313 **
CANALIZADA_COMODO1	-0.286158	0.074266	-3.853	0.000120 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Approximate significance of smooth terms:

	edf	Est.rank	F	p-value
s(coords.x1,coords.x2)	26.567	29	9.924	<2e-16 ***
s(SEQ)	5.682	9	2.190	0.0202 *

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.105   Deviance explained = 15.6%
GCV score = 1.0177   Scale est. = 1           n = 2202
```

A análise de pressuposto é feita usando a função gam.check.

```
> gam.check(mod.temp)
```

fit method: performance iteration - magic
Smoothing parameter selection converged after 7 iterations.
The RMS GCV score gradient at convergence was $2.0867e-06$.
The Hessian was positive definite.
The estimated model rank was 44 (maximum possible: 44)

