

Regressão para dados de contagem em R

Um experimento foi realizado em uma área dividida em 90 lotes, cada um de 25 x 25 m, com três diferentes intensidades de pH do solo, uma medida de biomassa e riqueza de espécies (número de diferentes espécies no lote). Já é bastante conhecido que a riqueza de espécies diminui com o aumento da biomassa. Um dos objetivos do estudo consiste em avaliar se o coeficiente da relação entre riqueza de espécies e biomassa varia com o pH do solo.

Nota 1. Os dados coletados estão listados ao final deste arquivo. Os dados devem ser lidos para o objeto `dados` de modo que a variável pH seja do tipo fator (*factor*).

```
dados <- . . . # completar

names(dados)

[1] "pH" "biomassa" "especies"

summary(dados)

      pH      biomassa      especies
alto :30   Min.    :0.05017   Min.    : 2.00
baixo:30   1st Qu.:1.44132   1st Qu.:12.25
medio:30   Median  :3.10836   Median  :18.50
          Mean    :3.55701   Mean    :19.46
          3rd Qu.:5.08570   3rd Qu.:25.75
          Max.    :9.98177   Max.    :44.00
```

Os níveis da variável pH, por *default*, são fixados em ordem alfabética. Tendo em vista que a variável é ordinal, por facilidade, os níveis são mudados para ordem decrescente.

```
dados$pH <- factor(dados$pH, levels = c("alto", "medio", "baixo"))
```

Nota 2. Represente graficamente cada uma das variáveis do conjunto de dados.

Diferentes modelos são ajustados a estes dados, conforme abaixo.

```
## Modelos
m1 <- glm(especies ~ pH * biomassa, family = "poisson", data = dados)
m2 <- update(m1, ~ . -pH:biomassa)
m3 <- update(m2, ~ . -pH)
m4 <- update(m3, ~ . -biomassa)
```

Nota 3. Descreva cada um dos quatro modelos ajustados.

Alguns resultados do ajuste do modelo m1 são mostrados em seguida.

```
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 452.346 on 89 degrees of freedom
Residual deviance: 83.201 on 84 degrees of freedom
```

Nota 4. Com base apenas nos resultados acima, pode ser afirmado que o modelo m1 faz um bom ajuste aos dados?

Os quatro modelos são comparados na tabela ANODEV abaixo.

```
## ANODEV
anova(m4, m3, m2, m1, test = "Chi")

Model 1: especies ~ 1
Model 2: especies ~ biomassa
Model 3: especies ~ pH + biomassa
Model 4: especies ~ pH * biomassa
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      89    452.35
2      88    407.67  1    44.673 2.328e-11 ***
3      86     99.24  2   308.431 < 2.2e-16 ***
4      84     83.20  2    16.040 0.0003288 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nota 5. Utilizando a tabela ANODEV justifique a escolha do modelo m1. O que pode ser afirmado em relação aos objetivos do estudo?

Nota 6. Apresente as estimativas dos coeficientes do modelo m1 e interprete-os.

Com as funções `predict` e `lines` representamos graficamente os dados e o ajuste obtido com o modelo m1. Algumas variáveis auxiliares são necessárias.

```
## Ajuste pelo modelo m1
npontos <- 200
lph <- levels(dados$pH)
nlph <- length(lph)
cores <- rainbow(nlph)
bioaux <- seq(min(dados$biomassa), max(dados$biomassa), length = npontos)
bioaux <- rep(bioaux, times = nlph)
phaux <- factor(rep(lph, each = npontos))

with(dados, plot(biomassa, especies, type = "n", xlab = "Biomassa",
  ylab = "Número de espécies", cex.lab = 1.3, cex.axis = 1.3))
for (i in 1:nlph) {
  with(dados, points(biomassa[pH == lph[i]], especies[pH == lph[i]],
    col = cores[i], pch = 20))
  lines(bioaux[phaux == lph[i]], predict(m1, type = "response",
    data.frame(biomassa = bioaux, pH = phaux))[phaux == lph[i]],
    col = cores[i])
}
legend("topright", lph, pch = 20, col = cores, bty = "n")
```

Nota 7. Reescreva o trecho de código acima sem utilizar a função `with`.

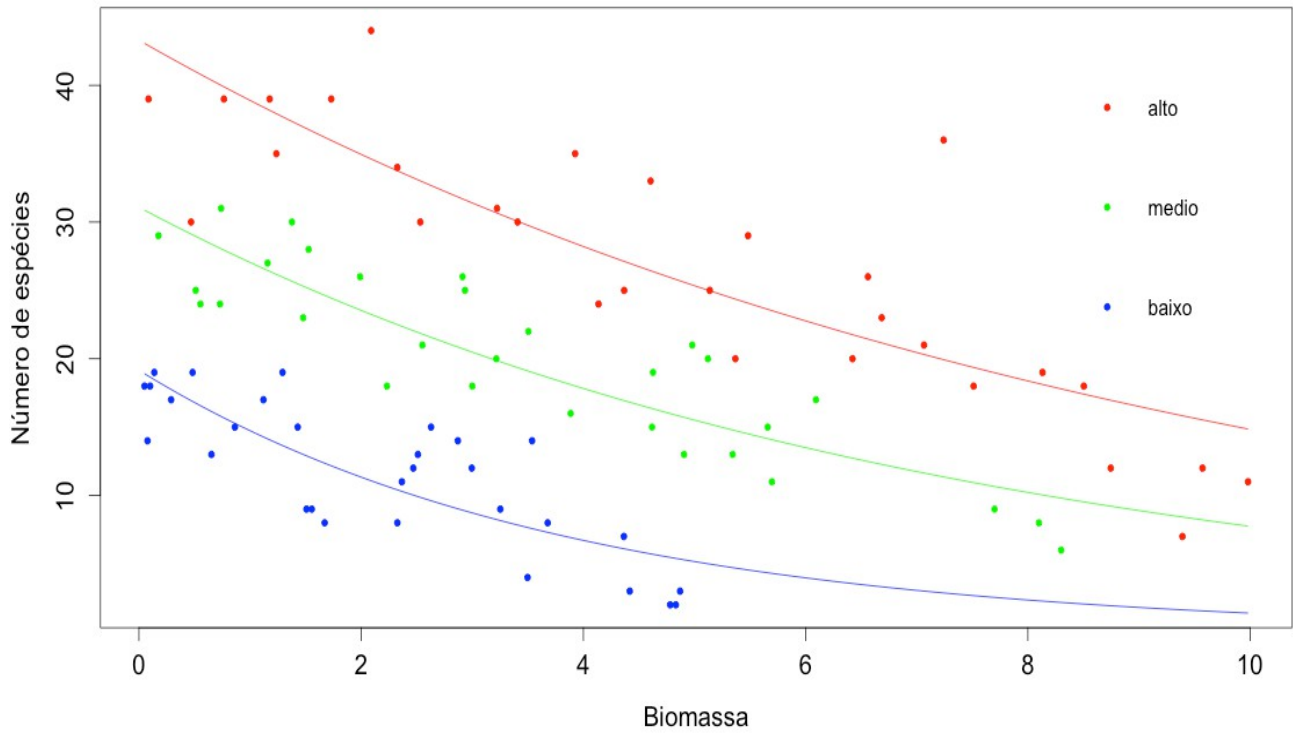


Figura 1. Gráfico de ajuste do modelo m1.

Nota 8. Represente graficamente cada um dos três demais modelos ajustados e comente os resultados, comparando-os.

Nota 9. O que mudaria se os lotes fossem de dimensões diferentes?

Nota 10. Refaça o exemplo utilizando outros pacotes estatísticos.

Dados

pH	biomassa	especies	pH	biomassa	especies	pH	biomassa	especies
alto	0.469297221	30	medio	0.175762697	29	baixo	0.100847904	18
alto	1.730870426	39	medio	1.376778303	30	baixo	0.138596089	19
alto	2.089778481	44	medio	2.551042556	21	baixo	0.863515079	15
alto	3.925787136	35	medio	3.00027434	18	baixo	1.292919034	19
alto	4.366792653	25	medio	4.905623865	13	baixo	2.46916355	12
alto	5.481974683	29	medio	5.343305425	13	baixo	2.366553089	11
alto	6.68468591	23	medio	7.7	9	baixo	2.629217082	15
alto	7.511650634	18	medio	0.553688929	24	baixo	3.252286524	9
alto	8.132202505	19	medio	1.990296437	26	baixo	4.417276192	3
alto	9.572128635	12	medio	2.91263671	26	baixo	4.780810391	2
alto	0.086653665	39	medio	3.216451335	20	baixo	0.050175289	18
alto	1.236973905	35	medio	4.979884675	21	baixo	0.482836915	19
alto	2.532043235	30	medio	5.658722899	15	baixo	0.652667143	13
alto	3.407941525	30	medio	8.1	8	baixo	1.555336555	9
alto	4.605045962	33	medio	0.739569861	31	baixo	1.671638196	8
alto	5.367717087	20	medio	1.526934202	28	baixo	2.870053904	14
alto	6.560842148	26	medio	2.232122387	18	baixo	2.510720525	13
alto	7.242062143	36	medio	3.885288176	16	baixo	3.497603852	4
alto	8.503632993	18	medio	4.626505411	19	baixo	3.678761859	8
alto	9.390953422	7	medio	5.120968439	20	baixo	4.831542449	2
alto	0.76488801	39	medio	8.3	6	baixo	0.289722659	17
alto	1.176470196	39	medio	0.511278577	25	baixo	0.077560095	14
alto	2.325120818	34	medio	1.478232688	23	baixo	1.429020407	15
alto	3.222882069	31	medio	2.934558004	25	baixo	1.120740925	17
alto	4.136129299	24	medio	3.50548891	22	baixo	1.507953837	9
alto	5.137176521	25	medio	4.617909137	15	baixo	2.325963179	8
alto	6.421938112	20	medio	5.696963821	11	baixo	2.995705821	12
alto	7.065526382	21	medio	6.093016876	17	baixo	3.538199085	14
alto	8.745929177	12	medio	0.730062798	24	baixo	4.364541206	7
alto	9.981770129	11	medio	1.158068376	27	baixo	4.870507888	3