

Prática de Computador II: Reprodução

Marco A. R. Mello

Universidade de São Paulo

Instituto de Biociências

Departamento de Ecologia

Tópicos Avançados em Ecologia de Animais (BIE0315)

Profs. José Carlos Motta Jr. & Marco A. R. Mello

Artigo de base: Palaoro et al. (2014 Animal Behaviour)

Agradecimentos: Alexandre Palaoro, primeiro autor do artigo de base, gentilmente nos cedeu os dados usados nesta prática.

Sumário

Preparativos

Teste 1

Teste 2

Teste 3

Teste 4

Teste 5

Para saber mais

Preparativos

Leia o tutorial desta prática, disponível aqui em formato PDF e também no moodle da disciplina.

Defina o diretório de trabalho como sendo a origem deste script.

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

Remova todos os objetos prévios.

```
rm(list= ls())
```

Limpe o console.

```
cat("\014")
```

Carregue os pacotes necessários.

```
library(ggplot2)
library(lme4)
```

Carregando pacotes exigidos: Matrix

Importe os dados e inspecione-os.

```
dados<- read.delim("dados.txt", header=T)
dim(dados)
```

```
## [1] 60 5
```

```
head(dados)
```

```
##      dupla   id    cc   ap   status
## 1 DUPLA1   AG1 18.88 6.88 vencedor
## 2 DUPLA1   AG6 15.31 5.27 perdedor
## 3 DUPLA2   AG5 16.96 5.99 perdedor
## 4 DUPLA2   AG7 20.35 8.35 vencedor
## 5 DUPLA3   AL6 13.78 4.77 vencedor
## 6 DUPLA3  AL21 14.33 4.39 perdedor
```

```
tail(dados)
```

```
##      dupla   id    cc   ap   status
## 55 DUPLA35 AG112 22.20 9.34 perdedor
## 56 DUPLA35 AG113 22.45 9.37 vencedor
## 57 DUPLA36 AG117 22.73 8.78 vencedor
## 58 DUPLA36 AG120 22.52 8.92 perdedor
## 59 DUPLA37 AG89 24.08 11.89 vencedor
## 60 DUPLA37 AG127 23.68 11.54 perdedor
```

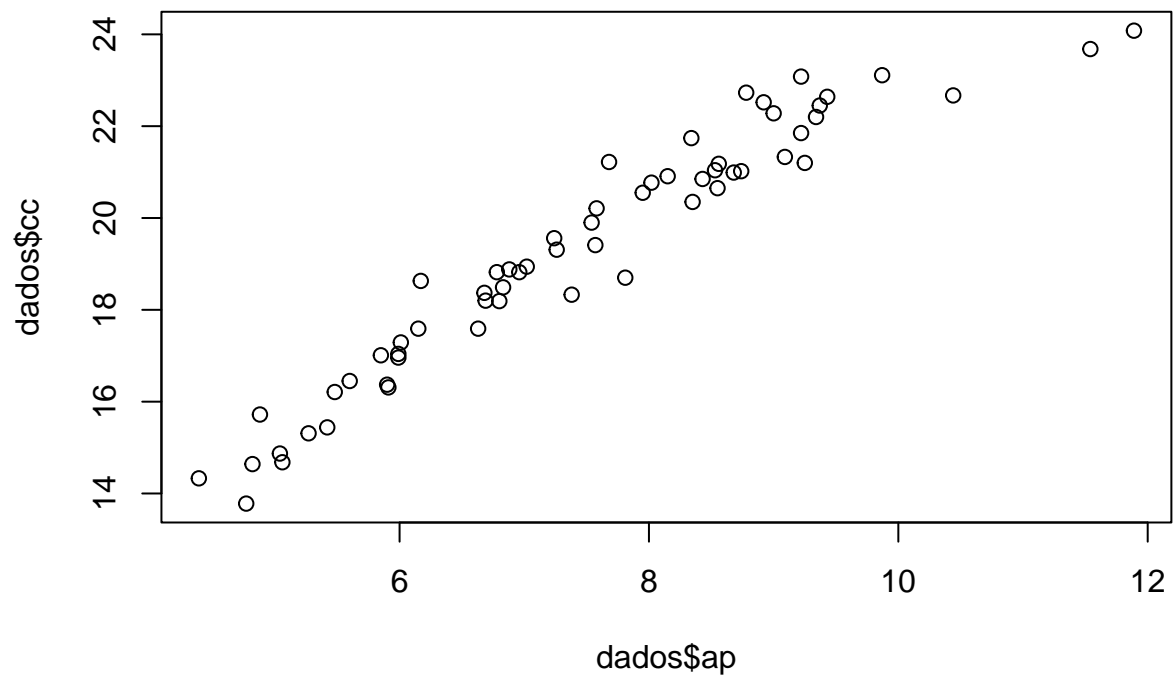
Transforme a variável “status” em numérica e binária, salvando-a como uma nova variável.

```
dados$status2 <- ifelse(dados$status == "perdedor", 0, 1)
head(dados)
```

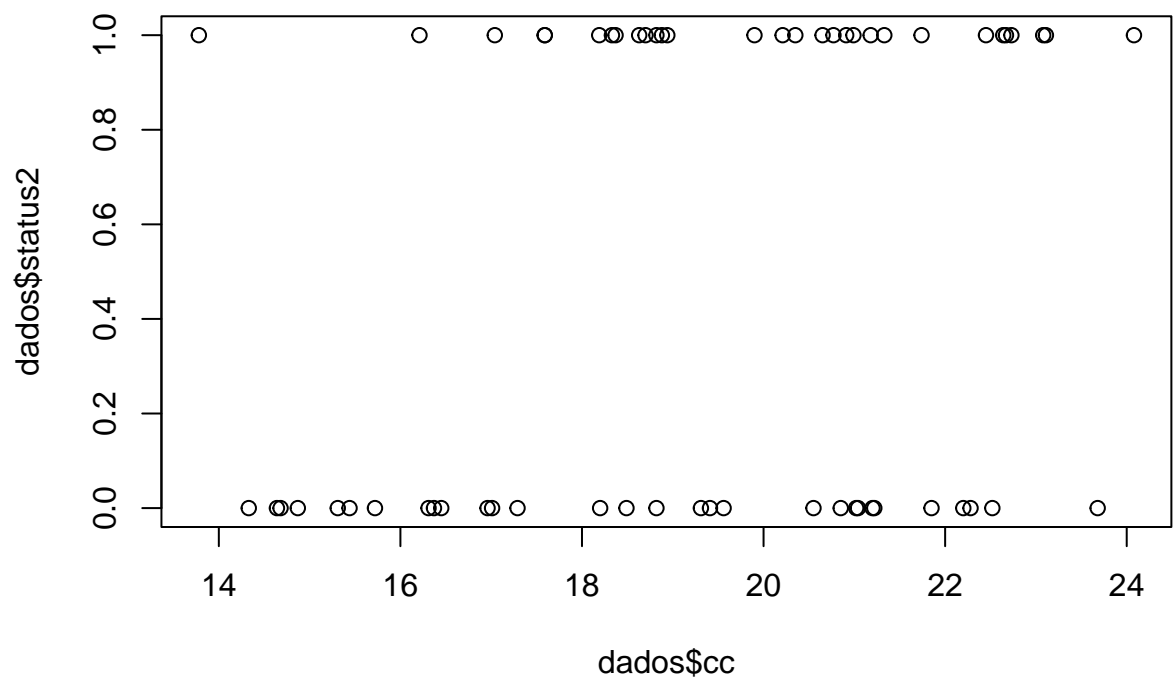
```
##      dupla   id    cc   ap   status status2
## 1 DUPLA1   AG1 18.88 6.88 vencedor         1
## 2 DUPLA1   AG6 15.31 5.27 perdedor         0
## 3 DUPLA2   AG5 16.96 5.99 perdedor         0
## 4 DUPLA2   AG7 20.35 8.35 vencedor         1
## 5 DUPLA3   AL6 13.78 4.77 vencedor         1
## 6 DUPLA3  AL21 14.33 4.39 perdedor         0
```

Examine de forma rápida as relações entre as variáveis.

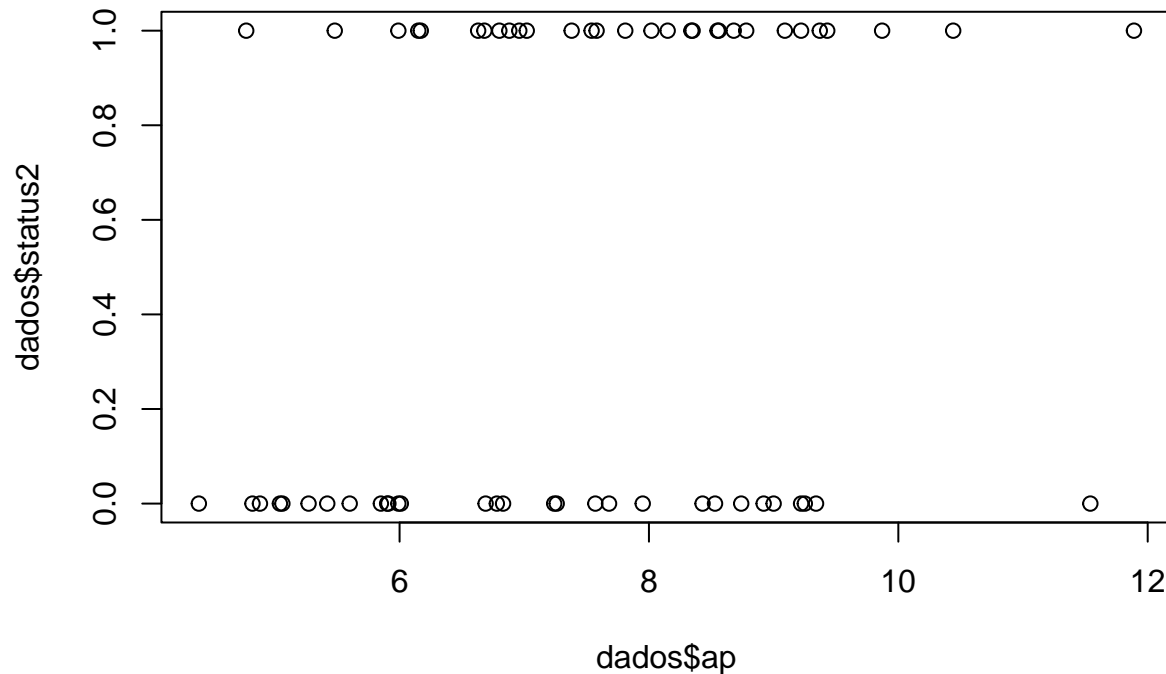
```
plot(dados$cc~dados$ap)
```



```
plot(dados$status2~dados$cc)
```



```
plot(dados$status2~dados$ap)
```



Teste 1

Plote a relação entre o status e o comprimento cefalotorácico, depois exporte o gráfico como um arquivo PNG.

```
png(filename= "figuras/p1.png", res= 300, height= 2000, width= 3000)
p1 = ggplot(dados, aes(x=cc, y=status2)) +
  geom_point(colour = "#1855FA", size=4, alpha = 0.5) +
  stat_smooth(method="glm", method.args=list(family="binomial"), se=FALSE) +
  labs(x="Comprimento cefalotorácico", y = "Status") +
  theme(text = element_text(size=20),
        plot.title = element_text(size=40, hjust=0.5),
        axis.text.x = element_text(size = 20, angle=0, hjust=1),
        axis.text.y = element_text(size = 20, angle=0, vjust=1),
        axis.title.x = element_text(size = 30, angle=0),
        axis.title.y = element_text(size = 30, angle=90))
p1
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
```

```
## 2
```

Teste se a relação é significativa e examine os valores estatísticos, depois exporte os resultados como um arquivo TXT.

```
fit1 = glm(dados$status2~dados$cc, family=binomial)
summary(fit1)
```

```
##
```

```
## Call:
```

```
## glm(formula = dados$status2 ~ dados$cc, family = binomial)
```

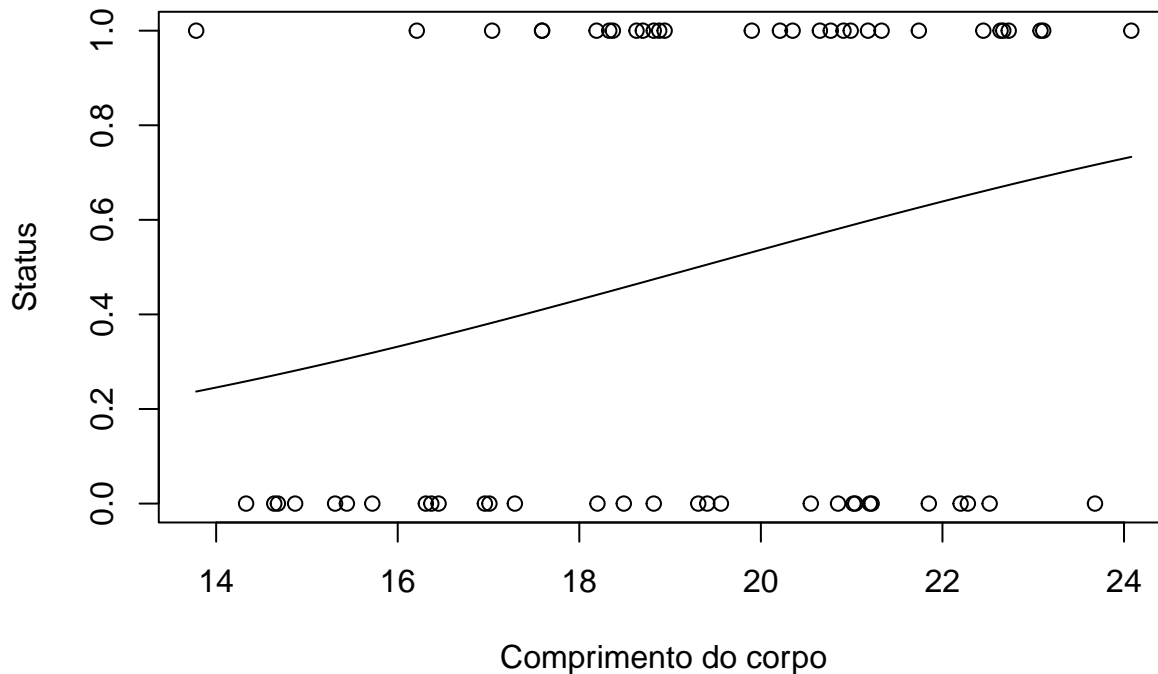
```
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58713  -1.11223   0.00722   1.08821   1.69711
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.0863     2.0550  -1.989   0.0468 *
## dados$cc      0.2117     0.1054   2.009   0.0446 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 83.178  on 59  degrees of freedom
## Residual deviance: 78.819  on 58  degrees of freedom
## AIC: 82.819
##
## Number of Fisher Scoring iterations: 4
res1 = anova(fit1, test="Chisq")
res1

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: dados$status2
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                    59      83.178
## dados$cc  1    4.3587      58      78.819 0.03682 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
capture.output(res1, file = "resultados/resultados-cc.txt")
```

Você também pode plotar o gráfico usando apenas o pacote base, com base nos resultados do teste que acabou de fazer.

```
plot(dados$status2~dados$cc,
     xlab = "Comprimento do corpo",
     ylab = "Status")
curve (exp(fit1$coefficients[[1]]+fit1$coefficients[[2]]*x)/
      (1+exp(fit1$coefficients[[1]]+fit1$coefficients[[2]]*x)),
      add=T)
```



Teste 2

Plote a relação entre o status e a altura do própodo, depois exporte o gráfico como um arquivo PNG.

```
png(filename= "figuras/p2.png", res= 300, height= 2000, width= 3000)
p2 = ggplot(dados, aes(x=ap, y=status2)) +
  geom_point(colour = "#1855FA", size=4, alpha = 0.5) +
  stat_smooth(method="glm", method.args=list(family="binomial"), se=FALSE) +
  labs(x="Altura da garra", y = "Status") +
  theme(text = element_text(size=20),
        plot.title = element_text(size=40, hjust=0.5),
        axis.text.x = element_text(size = 20, angle=0, hjust=1),
        axis.text.y = element_text(size = 20, angle=0, vjust=1),
        axis.title.x = element_text(size = 30, angle=0),
        axis.title.y = element_text(size = 30, angle=90))
p2
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
## 2
```

Teste se a relação é significativa e examine os valores estatísticos, depois exporte os resultados como um arquivo TXT.

```
fit2 = glm(dados$status2~dados$ap, family=binomial)
summary(fit2)
```

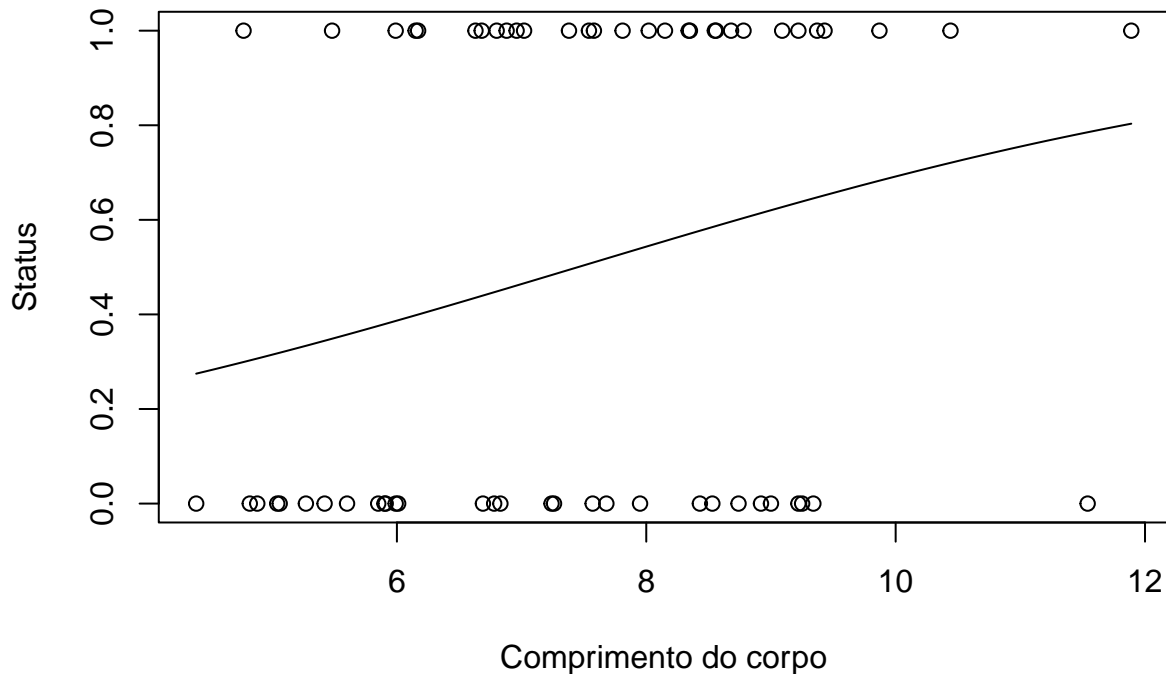
```
##
## Call:
## glm(formula = dados$status2 ~ dados$ap, family = binomial)
##
## Deviance Residuals:
```

```
##      Min      1Q      Median      3Q      Max
## -1.75391 -1.09000 -0.06983  1.10885  1.55317
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.3630      1.2703  -1.860  0.0629 .
## dados$ap      0.3171      0.1670   1.899  0.0576 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 83.178  on 59  degrees of freedom
## Residual deviance: 79.256  on 58  degrees of freedom
## AIC: 83.256
##
## Number of Fisher Scoring iterations: 4
res2 = anova(fit2, test="Chisq")
res2

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: dados$status2
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                59      83.178
## dados$ap  1    3.9217      58    79.256 0.04767 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
capture.output(res2, file = "resultados/resultados-ap.txt")
```

Você também pode plotar o gráfico usando apenas o pacote base, com base nos resultados do teste que acabou de fazer.

```
plot(dados$status2~dados$ap,
     xlab = "Comprimento do corpo",
     ylab = "Status")
curve (exp(fit2$coefficients[[1]]+fit2$coefficients[[2]]*x)/
      (1+exp(fit2$coefficients[[1]]+fit2$coefficients[[2]]*x)),
      add=T)
```



Teste 3

Teste como tanto o comprimento cefalotorácico quanto a altura do própodo determinam o status, depois exporte os resultados como um arquivo TXT.

```
fit3 = glm(dados$status2~dados$ap+dados$cc, family=binomial)
summary(fit3)
```

```
##
## Call:
## glm(formula = dados$status2 ~ dados$ap + dados$cc, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.54118  -1.11800   0.02586   1.08650   1.71999
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.40482    3.33303  -1.322   0.186
## dados$ap    -0.07273    0.59603  -0.122   0.903
## dados$cc     0.25625    0.38102   0.673   0.501
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 83.178  on 59  degrees of freedom
## Residual deviance: 78.804  on 57  degrees of freedom
## AIC: 84.804
##
## Number of Fisher Scoring iterations: 4
```

```
res3 = anova(fit3, test="Chisq")
res3
```

```
## Analysis of Deviance Table
```



```
##
## Model: binomial, link: logit
##
## Response: dados$status2
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                    59      83.178
## dados$ap  1    3.9217      58    79.256 0.04767 *
## dados$cc  1    0.4518      57    78.804 0.50148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

capture.output(res3, file = "resultados/resultados-ap-cc.txt")
```

Teste 4

Teste como tanto o comprimento cefalotorácico quanto a altura do própodo determinam o status, mas considerando a identidade da dupla como um fator aleatório, depois exporte os resultados como um arquivo TXT.

```
fit4 = glmer(status2 ~ ap + cc + (1|dupla), family=binomial, data=dados)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(fit4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula: status2 ~ ap + cc + (1 | dupla)
##   Data: dados
##
##           AIC          BIC    logLik deviance df.resid
##           86.8          95.2     -39.4     78.8       56
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.5097 -0.9319  0.0232  0.8969  1.8410
##
## Random effects:
##   Groups Name          Variance Std.Dev.
##   dupla  (Intercept)  0          0
## Number of obs: 60, groups:  dupla, 30
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.40482    3.33262  -1.322   0.186
## ap          -0.07273    0.59594  -0.122   0.903
## cc           0.25625    0.38097   0.673   0.501
##
## Correlation of Fixed Effects:
##   (Intr) ap
## ap  0.787
```

```
## cc -0.925 -0.961
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

res4 = anova(fit4, test="Chisq")

## Warning in anova.merMod(fit4, test = "Chisq"): additional arguments ignored:
## 'test'

res4

## Analysis of Variance Table
##      npar Sum Sq Mean Sq F value
## ap      1 3.5930  3.5930  3.5930
## cc      1 0.4523  0.4523  0.4523

capture.output(res4, file = "resultados/resultados-ap-cc-dupla.txt")
isSingular(fit4, tol = 1e-05)

## [1] TRUE
```

Teste 5

Teste a relação entre os fatores usados nas análises anteriores, salvando os resíduos dessa relação para fazer uma nova análise.

```
fit5 = lm(cc ~ ap, data=dados)
summary(fit5)

##
## Call:
## lm(formula = cc ~ ap, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90915 -0.42867  0.08534  0.39608  1.59883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.00452    0.42977   18.62  <2e-16 ***
## ap           1.51258    0.05618   26.92  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7331 on 58 degrees of freedom
## Multiple R-squared:  0.9259, Adjusted R-squared:  0.9246
## F-statistic: 724.8 on 1 and 58 DF,  p-value: < 2.2e-16

fit5.res = resid(fit5)
```

Plote a relação entre os fatores, depois exporte o gráfico como um arquivo PNG.

```
png(filename= "figuras/p5.png", res= 300, height= 3000, width= 3000)
p5 = ggplot(dados, aes(x=cc, y=ap), CI = F) +
  geom_smooth(method=lm, colour = "#1855FA") +
  geom_point(colour = "#1855FA", size=4, alpha = 0.5) +
  geom_ribbon(stat='smooth', method = "lm", se=TRUE, alpha=0.1,
            fill = "#1855FA") +
  ggtitle("") +
```

```
labs(x="Comprimento do corpo", y = "Altura da garra") +
theme(text = element_text(size=20),
      plot.title = element_text(size=40, hjust=0.5),
      axis.text.x = element_text(size = 20, angle=0, hjust=1),
      axis.text.y = element_text(size = 20, angle=0, vjust=1),
      axis.title.x = element_text(size = 30, angle=0),
      axis.title.y = element_text(size = 30, angle=90))
```

p5

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
## 2
```

Plote a relação entre status e os resíduos da relação entre comprimento cefalotorácico e altura do própodo, depois exporte o gráfico como um arquivo PNG.

```
p6 = png(filename= "figuras/p6.png", res= 300, height= 2000, width= 3000)
ggplot(dados, aes(x=fit5.res, y=status2)) +
  geom_point(colour = "#1855FA", size=4, alpha = 0.5) +
  stat_smooth(method="glm", method.args=list(family="binomial"), se=FALSE) +
  labs(x="Resíduos corpo-garra", y = "Status") +
  theme(text = element_text(size=20),
        plot.title = element_text(size=40, hjust=0.5),
        axis.text.x = element_text(size = 20, angle=0, hjust=1),
        axis.text.y = element_text(size = 20, angle=0, vjust=1),
        axis.title.x = element_text(size = 30, angle=0),
        axis.title.y = element_text(size = 30, angle=90))
```

```
## `geom_smooth()` using formula 'y ~ x'
```

p6

```
## NULL
```

```
dev.off()
```

```
## pdf
## 2
```

Teste se a relação é significativa e examine os valores estatísticos, depois exporte os resultados como um arquivo TXT.

```
fit6 = glmer(status2 ~ fit5.res + (1|dupla), family=binomial, data=dados)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(fit6)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: status2 ~ fit5.res + (1 | dupla)
## Data: dados
##
##          AIC          BIC    logLik deviance df.resid
```

```
##      88.7      95.0     -41.4      82.7      57
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.21421 -0.99973  0.01677  0.96981  1.26102
##
## Random effects:
##   Groups Name            Variance Std.Dev.
##   dupla  (Intercept)  0          0
## Number of obs: 60, groups:  dupla, 30
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.0001411  0.2591781  -0.001    1.000
## fit5.res     0.2428873  0.3626997   0.670    0.503
##
## Correlation of Fixed Effects:
##              (Intr)
## fit5.res -0.002
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

res6 = anova(fit6, test="Chisq")

## Warning in anova.merMod(fit6, test = "Chisq"): additional arguments ignored:
## 'test'

res6

## Analysis of Variance Table
##              npar  Sum Sq Mean Sq F value
## fit5.res      1 0.44845 0.44845  0.4485

capture.output(res6, file = "resultados/resultados-res-dupla.txt")
isSingular(fit6, tol = 1e-05)

## [1] TRUE
```

Para saber mais

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>

Dobson, A. J., & Barnett, A. J. (2008). *An introduction to generalized linear models* (3rd ed.). CRC Press.

Ellison, A. M., Gotelli, N. J., Inouye, B. D., & Strong, D. R. (2014). P values, hypothesis testing, and model selection: it's déjà vu all over again. *Ecology*, 95(3), 609–610. <https://doi.org/10.1890/13-1911.1>

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* (1st ed.). Springer New York. <https://doi.org/10.1007/978-0-387-87458-6>

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Regressão logística na Wikipedia: https://en.wikipedia.org/wiki/Logistic_regression

GLM na Wikipedia: https://en.wikipedia.org/wiki/Generalized_linear_model