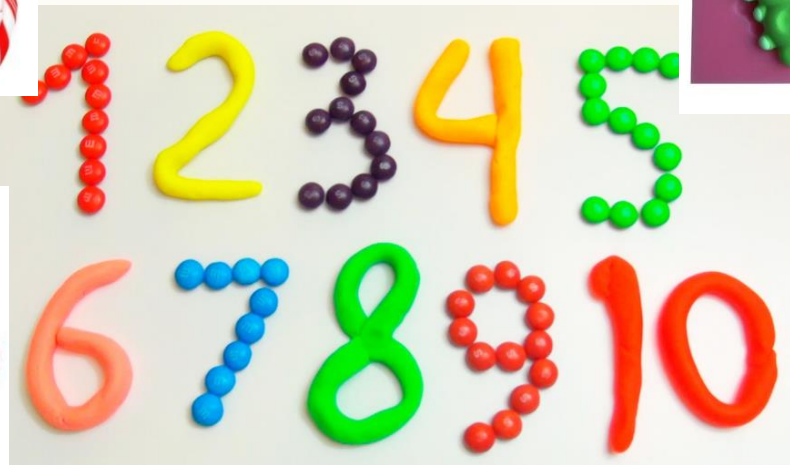


Aula 20 Goodies*



* Goodies related to animals, plants and numbers...



Whatsapp ?


Eco numérica está a correr bem?

Aqui entre nós, acho que estou com mais dificuldade em mante-los interessados... c'est la vie...!

09:07 ✓✓

:/ chato. Mas sim, cest la vie. Eu disse que nos éramos mais fixes 😏
É pena porque qnd eles chegarem onde eu estou vão querer ter tirado o máximo disso xd dados e dados e dados
N se fax nada aqui sem dados

09:12

MAAAS não desanime 🙏 força nisso 😊

09:28



10:10 ✓✓

[ncse-members] [ncse-all] New MSc in Statistical Ecology



ncse-members-request@st-andrews.ac.uk on behalf of Stephen E
To ncse-all@st-andrews.ac.uk



Mon 2:39 PM

St Andrews University now offers an MSc Degree and Postgraduate Diploma in Statistical Ecology. These programmes (one year fulltime or two years part-time) are taught by staff at the world-leading [Centre for Research into Ecological and Environmental Modelling \(CREEM\)](#). They are designed to give students a solid understanding of the statistical foundations of modern statistical ecology methods and to equip them with the skills to use these methods effectively, or to pursue a research career in statistical ecology.

[CREEM](#) collaborates with researchers across the globe and the MSc incorporates a summer research dissertation that includes an optional placement, with collaborators either in the UK or (at additional cost) abroad. Placements are designed to connect theoretical training with real field studies and professionals, and to give hands-on experience using these methods under the supervision of experts, including some of the leading researchers in this field. We also have an innovative and flexible dissertation format, that can include producing a podcast, web page, poster, field report, training materials, or a short film.

Further information, including details of entry requirements and how to apply can be found here:

<https://www.st-andrews.ac.uk/subjects/statistics/statistical-ecology-msc/#65966>

Applicants may also apply for a fee waiver. If successful, their course fee will be covered by the University. Details will be posted at the above website shortly.

Steve Buckland

Stephen T. Buckland
CREEM, The Observatory, Buchanan Gdns, St Andrews KY16 9LZ, Scotland
Tel. +44 (0)1334 461841 Secretary +44 (0)1334 461842
e-mail steve@st-andrews.ac.uk

Professor of Statistics <http://www.st-andrews.ac.uk/>
NCSE <http://www.ncse.org.uk/>
CREEM <http://creem2.st-andrews.ac.uk/>
Personal website <http://steve7397.wix.com/personal>

The University of St Andrews is a charity registered in Scotland:No SC013532



Se alguém achar que isto é um caminho que gostava de seguir, venha falar comigo!



A cool package for EDA

Required package

And hence sharing with you

```
texugo <- read.csv2("DataTP8texugo.csv")
library(dlookr)
eda_report(texugo, target = Densidade, output_format = "html",
           output_file = "EDA.html", output_dir = getwd())
```

Dados da ficha 8,
Aula TP 11

data.frame with the data

response variable

Puts doc in the working directory

Exploratory Data Analysis Report

Report by dlookr package

2018-11-15

- 1 Introduction
 - 1.1 Information of Dataset
 - 1.2 Information of Variables
 - 1.3 About EDA Report
- 2 Univariate Analysis
 - 2.1 Descriptive Statistics
 - 2.2 Normality Test of Numerical Variables
 - 2.2.1 Statistics and Visualization of (Sample) Data
- 3 Relationship Between Variables
 - 3.1 Correlation Coefficient
 - 3.1.1 Correlation Coefficient by Variable Combination
 - 3.1.2 Correlation Plot of Numerical Variables
- 4 Target based Analysis
 - 4.1 Grouped Descriptive Statistics
 - 4.1.1 Grouped Numerical Variables
 - 4.1.2 Grouped Categorical Variables
 - 4.2 Grouped Relationship Between Variables
 - 4.2.1 Grouped Correlation Coefficient
 - 4.2.2 Grouped Correlation Plot of Numerical Variables

1 Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and

2.2.1 Statistics and Visualization of (Sample) Data

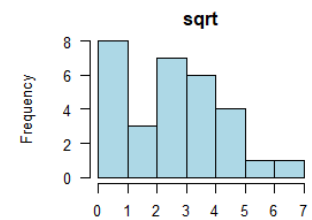
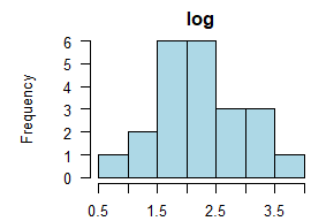
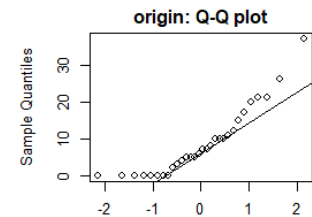
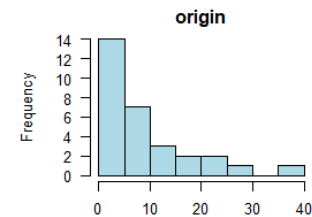
[b]

normality test : Shapiro-Wilk normality test

```
statistic : 0.86037, p-value : 0.0010304
```

skewness and kurtosis

| type | skewness | kurtosis |
|---------------------|-----------|----------|
| original | 1.2700514 | 4.333604 |
| log transformation | NaN | NaN |
| sqrt transformation | 0.0226179 | 2.090171 |



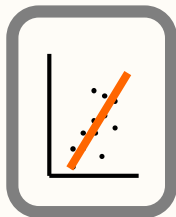


With too little data, you won't be able to make any conclusions that you trust. With loads of data you will find relationships that aren't real... Big data isn't about bits, it's about talent.

— *Douglas Merrill* —

AZ QUOTES

<https://www.azquotes.com/quote/603350>



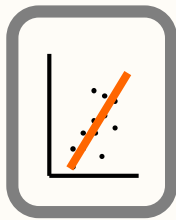
correlação

Correlação múltipla

- O R^2 numa análise da regressão múltipla expressa a correlação entre a variável resposta e as predições feitas pelo modelo com base em todas as variáveis.

$$R^2 = \frac{SS_{REGRESSÃO}}{SS_{TOTAL}}$$

$$0 \leq R^2 \leq 1$$



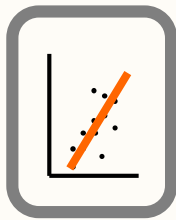
correlação

O coeficiente de correlação múltipla

Variabilidade explicada pelo modelo

- O coeficiente de determinação múltiplo R^2 depende de p e n e, por isso, uma estimativa mais adequada da correlação múltipla é o R^2 ajustado, dado por:

$$R_a^2 = 1 - \frac{MS_{RESIDUAL}}{MS_{TOTAL}} = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$



correlação

Correlação múltipla: testes de hipóteses

- O teste F, que avalia a igualdade a zero de todos os coeficientes da regressão, testa também a correlação múltipla entre todas as variáveis.

$$F = \frac{\frac{SS_{REGRESSÃO}}{v_{REGRESSÃO}}}{\frac{SS_{ERRO}}{v_{ERRO}}} = \frac{MS_{REGRESSÃO}}{MS_{ERRO}}$$

```
> xs=runif(100,0,10)
> zs=runif(100,30,40)
> ms=runif(100,50,90)
> ys=3+0.3*xs+0.2*zs+rnorm(100)
>
> lm1=lm(ys~xs+zs+ms)
> summary(lm1)
```

Call:

```
lm(formula = ys ~ xs + zs + ms)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|---------|---------|--------|--------|--------|
| | -2.0733 | -0.7459 | 0.0795 | 0.7042 | 3.2557 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|-----------|------------|---------|----------|-----|
| (Intercept) | 3.846878 | 1.372926 | 2.802 | 0.00614 | ** |
| xs | 0.297329 | 0.036258 | 8.200 | 1.07e-12 | *** |
| zs | 0.204510 | 0.035826 | 5.708 | 1.27e-07 | *** |
| ms | -0.014804 | 0.009647 | -1.534 | 0.12821 | |

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

Residual standard error: 1.016 on 96 degrees of freedom

Multiple R-squared: 0.5149, Adjusted R-squared: 0.4997

F-statistic: 33.96 on 3 and 96 DF, p-value: 4.761e-15

```
> cor(predict(lm1),ys)^2
[1] 0.5148676
```



correlação

Correlação múltipla não-paramétrica

O coeficiente de concordância de Kendall

$$W = \frac{\sum R_i^2 - \frac{(\sum R_i)^2}{n}}{\frac{M^2(n^3 - n)}{12}}$$

$$0 \leq W \leq 1$$

- onde M é o número de variáveis, n o número de observações por variável e R_i a soma das ordens associada com a linha i .
- Existe uma outra expressão para situações em que existem empates nas ordens das observações.

Table 1. Illustrative example. Upper panel: Hellinger-transformed abundances of four mite species at 10 sites selected along the long axis of Figure 5. The Hellinger transformation was computed for the full dataset (70 sites). Lower panel: the same data transformed into ranks (with ties); last column: sum of the ranks for each site.

| | <i>Hellinger-transformed abundances</i> | | | | |
|---------|---|-------------------|-------------------|-------------------|---------------------------------------|
| | <i>Species 13</i> | <i>Species 14</i> | <i>Species 15</i> | <i>Species 23</i> | |
| Site 4 | .25087 | .40538 | .24380 | .08362 | |
| Site 9 | .40324 | .25503 | .39303 | .00000 | |
| Site 14 | .26577 | .47620 | .27267 | .06097 | |
| Site 22 | .32350 | .63337 | .47003 | .00000 | |
| Site 31 | .26312 | .29089 | .39223 | .08771 | |
| Site 34 | .33675 | .44836 | .53727 | .10153 | |
| Site 45 | .07956 | .19487 | .19487 | .11251 | |
| Site 53 | .00000 | .18570 | .26261 | .11744 | |
| Site 61 | .00000 | .15430 | .15430 | .00000 | |
| Site 69 | .12769 | .62987 | .27584 | .34578 | |
| | <i>Ranks (species-wise)</i> | | | | <i>Sum of ranks R_i</i> |
| | <i>Species 13</i> | <i>Species 14</i> | <i>Species 15</i> | <i>Species 23</i> | |
| Site 4 | 5 | 6 | 3 | 5 | 19.0 |
| Site 9 | 10 | 4 | 8 | 2 | 24.0 |
| Site 14 | 7 | 8 | 5 | 4 | 24.0 |
| Site 22 | 8 | 10 | 9 | 2 | 29.0 |
| Site 31 | 6 | 5 | 7 | 6 | 24.0 |
| Site 34 | 9 | 7 | 10 | 7 | 33.0 |
| Site 45 | 3 | 3 | 2 | 8 | 16.0 |
| Site 53 | 1.5 | 2 | 4 | 9 | 16.5 |
| Site 61 | 1.5 | 1 | 1 | 2 | 5.5 |
| Site 69 | 4 | 9 | 6 | 10 | 29.0 |



correlação

Testes de hipóteses

Hipótese:

H0: não há associação entre as variáveis (ou temos M juízes independentes)

H1: há associação entre as variáveis

Estatística de teste:

$$\chi_r^2 = M(n-1)W$$

Teste de Friedman

Valor crítico:

$$\chi^2_{\alpha, n-1}$$

Critério de decisão:

Rejeitar H0 se:

$$\chi^2 > \chi^2_{\alpha, n-1}$$

Não rejeitar H0 caso contrário

Coeficiente de concordância de Kendall

```
> library(vegan)
> kendall.global(iris[,1:4])
$`Concordance_analysis`
      Group.1
W      4.864873e-01
F      2.842115e+00
Prob.F  2.778329e-17
Chi2    2.899464e+02
Prob.perm 1.000000e-03
```

Rejeitamos a H0 de que a correlação múltipla é nula

Teste de Friedman

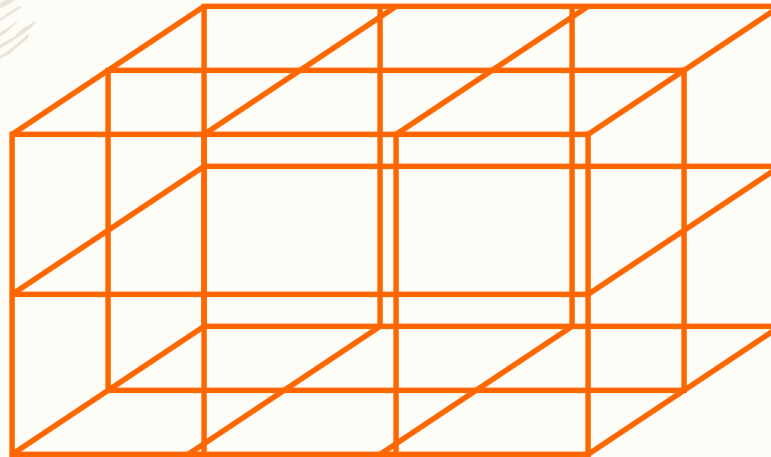
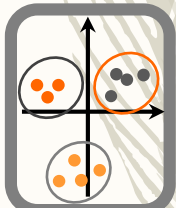
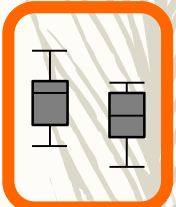
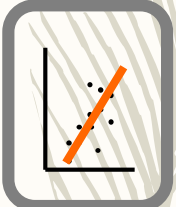
```
> friedman.test(as.matrix(iris[,1:4]))

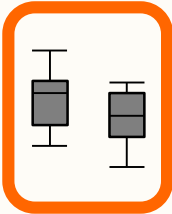
Friedman rank sum test

data:  as.matrix(iris[, 1:4])
Friedman chi-squared = 410, df = 3, p-value < 2.2e-16
```

Ecologia Numérica

tabelas de contingência





tabelas de contingência

- Quais os testes adequados para analisar dados de contagens de variáveis categóricas?
- Quais as condições para a sua aplicação?
- Como interpretar os seus resultados?

Como obter uma tabela de contingência no R? Se tivermos observações em factores, e quisermos contar quantos temos em cada classe, basta usar a função `table`

```
with(npk, table(N, P, K))  
, , K = 0
```

| | P | |
|---|---|---|
| N | 0 | 1 |
| 0 | 3 | 3 |
| 1 | 3 | 3 |

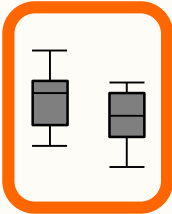
```
, , K = 1
```

| | P | |
|---|---|---|
| N | 0 | 1 |
| 0 | 3 | 3 |
| 1 | 3 | 3 |

`npk` – in package `datasets` - A classical N, P, K (nitrogen, phosphate, potassium) factorial experiment on the growth of peas conducted on 6 blocks.

```
with(npk, table(N, P))
```

| | P | |
|---|---|---|
| N | 0 | 1 |
| 0 | 6 | 6 |
| 1 | 6 | 6 |



tabelas de contingência

Tabelas de Contingência

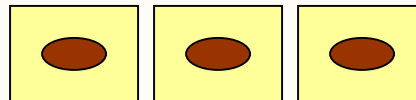
Dados: contagens de variáveis nominais (categóricas) -
(mas por vezes podem ser variáveis contínuas discretizadas em classes)

Variável B

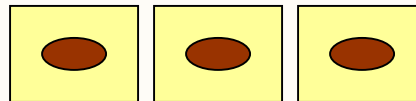
Variável A

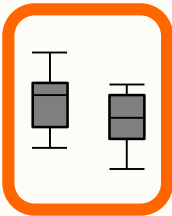
A1 A2 A3

B1



B2





tabelas de contingência

Tabelas de Contingência

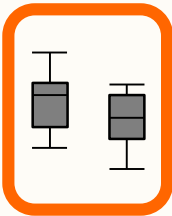
Hipóteses:

H0: As frequências referentes às linhas **são independentes** das registadas nas colunas

H1: ... **não são independentes** ...

daí estes testes serem chamados ...

Testes de Independência



tabelas de contingência

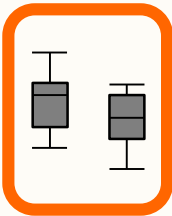
Teste do qui-quadrado

Estatística de teste:

$$\chi^2 = \sum \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

Onde f_{ij} é a frequência observada para a célula ij e

\hat{f}_{ij} é a frequência esperada para a célula ij



tabelas de contingência

Teste do qui-quadrado

Estatística de teste:

$$\chi^2 = \sum \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

Valor crítico:

$$\chi^2_{\alpha, \nu}$$

Critério de decisão:

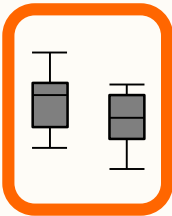
Rejeitar H0 se:

$$\chi^2 < \chi^2_{\alpha, \nu}$$

Não rejeitar H0, caso contrário

Graus de liberdade:

$$\nu = (n^{\circ}linhas - 1)(n^{\circ}colunas - 1) = (nl - 1)(nc - 1)$$



tabelas de contingência

Tabelas de Contingência

| | Castanho | Preto | Branco | |
|-----------|-------------------------|------------------------|-------------------------|-----|
| sexo\cor | | | | |
| Masculino | 10 (85*41/160=21.78) | 23 (85*48/160=25.5) | 52 (85*71/160=37.72) | 85 |
| Feminino | 31 (75*41/160=19.22) | 25 (75*48/160=22.5) | 19 (75*71/160=33.28) | 75 |
| | 41 | 48 | 71 | 160 |

$$\begin{aligned} X^2 = & (10-21.78)^2/21.78 + (23-25.5)^2/25.5 + (52-37.72)^2/37.72 + \\ & (31-19.22)^2/19.22 + (25-22.5)^2/22.5 + (19-33.28)^2/33.28 \\ & 6.371368 + 0.245098 + 5.406108 + 7.22 + 0.2777778 + 6.127356 = 25.64771 \end{aligned}$$

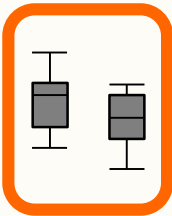
```
dados=matrix(c(10,23,52,31,25,19),ncol=3,nrow=2,byrow=TRUE)
```

```
> chisq.test(dados)
```

Pearson's Chi-squared test

```
data: dados  
X-squared = 25.653, df = 2, p-value = 2.689e-06
```

```
> 1-pchisq(25.653,2)  
[1] 2.688573e-06
```



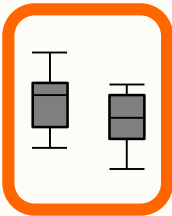
tabelas de contingência

Teste G

Hipóteses:

H0: As frequências referentes às linhas **são independentes** das registadas nas colunas

H1: ... **não são independentes** ...



tabelas de contingência

Teste G

Estatística de teste:

$$G = 2 \left[\sum_i \sum_j f_{ij} \ln f_{ij} \left(\frac{f_{ij}}{\hat{f}_{ij}} \right) \right]$$

Graus de liberdade:

$$\nu = (nl - 1)(nc - 1)$$

Valor crítico:

$$\chi^2_{\alpha, \nu}$$

Critério de decisão:

Rejeitar H0 se:

$$\chi^2 < \chi^2_{\alpha, \nu}$$

Não rejeitar H0 caso contrário

From library DescTools

```
> dados=matrix(c(10,23,52,31,25,19),ncol=3,nrow=2,byrow=TRUE)
> chisq.test(dados)
```

Pearson's Chi-squared test

```
data: dados
X-squared = 25.653, df = 2, p-value = 2.689e-06
```

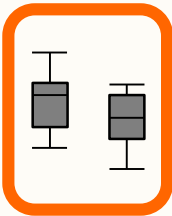
```
> GTest(dados,correct="none")
```

Log likelihood ratio (G-test) test of independence without correction

```
data: dados
G = 26.686, X-squared df = 2, p-value = 1.604e-06
```

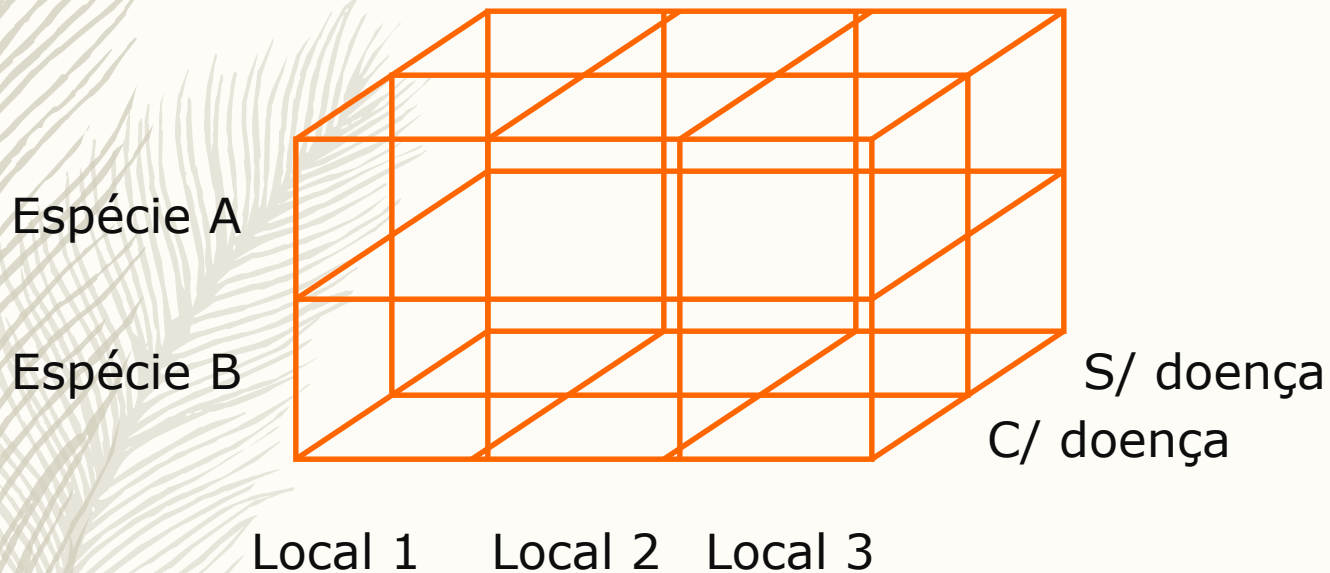
The G and Chi-Squared tests are asymptotically the same...!

<https://stats.stackexchange.com/questions/169142/g-test-vs-pearsons-chi-squared-test>

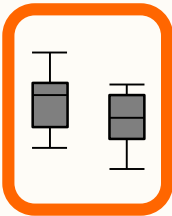


tabelas de contingência

Tabelas de contingência tridimensionais



O cálculo da estatística de teste é idêntico aos casos mais simples.



tabelas de contingência

Algumas notas importantes sobre o teste do qui-quadrado:

- No cálculo da estatística de teste são utilizadas sempre as frequências absolutas;
- Células com zeros - "No more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater" (Yates, Moore & McCabe, 1999, p. 734)
- Podemos proceder ao agrupamento de linhas ou colunas.

```
> dados=matrix(c(10,23,52,31,25,19),ncol=3,nrow=2,byrow=TRUE)
> chisq.test(dados)
```

Pearson's Chi-squared test

```
data: dados
X-squared = 25.653, df = 2, p-value = 2.689e-06
```

```
> chisq.test(dados,simulate.p.value = TRUE, B = 10000)
```

Pearson's Chi-squared test with simulated p-value (based on 10000 replicates)

```
data: dados
X-squared = 25.653, df = NA, p-value = 9.999e-05
```

```
> GTest(dados,correct="none")
```

Log likelihood ratio (G-test) test of independence without correction

```
data: dados
G = 26.686, X-squared df = 2, p-value = 1.604e-06
```



Existe também um teste de qui-quadrado para testar o ajustamento de modelos (*goodness-of-fit test*)

(sendo Z uma distribuição qualquer, e.g. Gaussiana, Binomial, Poisson, etc.)

H0 : A amostra **foi** retirada de uma população com distribuição Z

H1 : A amostra **não foi** retirada de uma população com distribuição Z

Sendo mais adequado no caso das variáveis discretas, também pode ser usado para variáveis contínuas – depois de agrupadas em classes

A ideia é comparar os valores esperados (E_i) de acordo com uma determinada distribuição, que assumimos sob H0, com os valores observados (O_i)

Os valores esperados, E_i , são obtidos de acordo com a função de distribuição da variável cuja distribuição queremos testar, sob H_0

A estatística de teste é a mesma que anteriormente (como se fosse calculada numa tabela de contingência com k colunas e uma linha):

| $-\infty, B_1$ | B_1, B_2 | B_2, B_3 | B_3, B_4 | ... | B_{k-3}, B_{k-2} | B_{k-2}, B_{k-1} | $B_k, +\infty$ |
|----------------|------------|------------|------------|-----|--------------------|--------------------|----------------|
| O_1 | O_2 | O_3 | O_4 | ... | O_{k-2} | O_{k-1} | O_k |
| E_1 | E_2 | E_3 | E_4 | ... | E_{k-2} | E_{k-1} | E_k |

(a soma dos quadrados das diferenças entre valores esperados e observados a dividir pelos valores esperados)

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Sob H_0 , a estatística de teste tem uma distribuição com $k-1$ graus de liberdade (se houver p parâmetros a estimar, $k-p-1$)

7 classes = 6+1

```
> set.seed(123)
> N=200;mean=3
> data=rpois(N,mean)
> obscounts=table(data)
> obscounts=c(obscounts[1:6],sum(obscounts[7:length(obscounts)]))
> #get the Poisson probabilities
> ps=c(dpois(0:5,mean),1-ppois(5,mean))
> chisq.test(obscounts,p=ps)
```

Chi-squared test for given probabilities

```
data: obscounts
X-squared = 5.3336, df = 6, p-value = 0.5018
```

```
> sum(ps)
[1] 1
>
> ET=sum((obscounts-N*ps)^2/(N*ps))
> ET
[1] 5.333647
```

```
> 1-pchisq(ET,df=6)
[1] 0.5017862
```

Uso (e abuso) mais comum: Distribuição uniforme

Lancei um dado 100 vezes, seria um dado equilibrado?

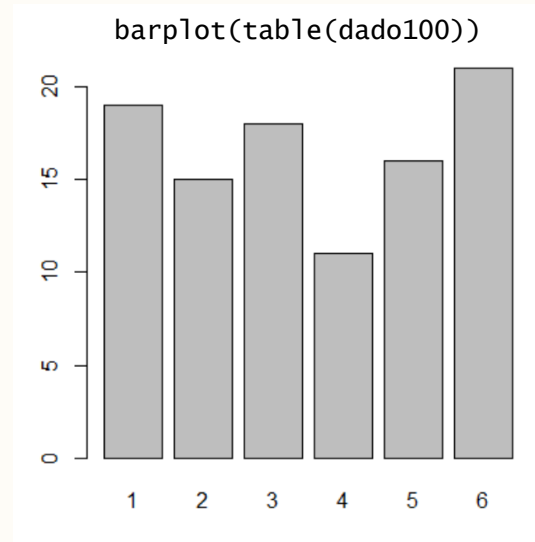
```
library(purrr)
set.seed(123); dado100=rdunif(n=100, b=6, a = 1)
table(dado100)
```

```
dado100
 1  2  3  4  5  6
19 15 18 11 16 21
```

```
chisq.test(table(dado100), p=rep(1/6, 6))
```

Chi-squared test for given probabilities

```
data: table(dado100)
X-squared = 3.68, df = 5, p-value = 0.5963
```



Porquê o abuso? Porque na realidade muitas vezes já sabemos que as coisas não são uniformes, mas ficamos mais felizes por conseguir demonstrá-lo com uma significância estatística, mas não ficamos nem um passo mais perto de conseguir interpretar a significância biológica do padrão encontrado.

“Entre” os testes de independência e os testes de ajustamento temos ainda os testes de homogeneidade, de uso menos frequente que os seus “irmãos”.

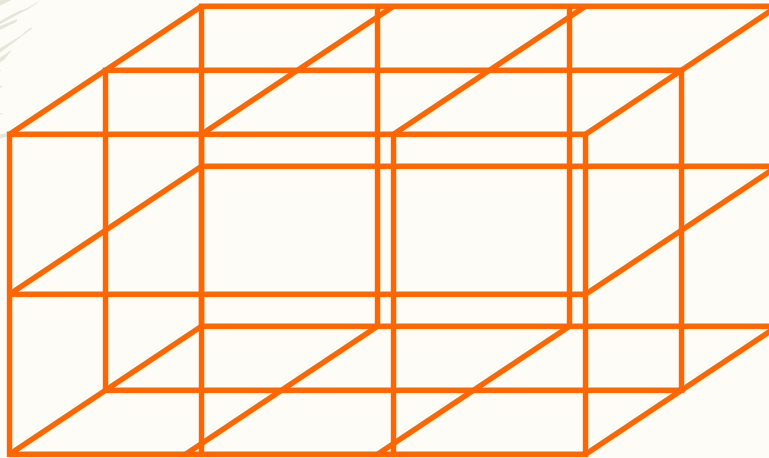
Neste caso a hipótese nula é que temos P amostras e queremos ver se poderiam ter vindo todas da mesma distribuição parente. As observações são recolhidas em K classes.

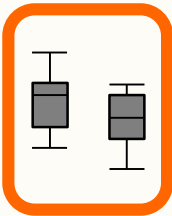
Curiosamente, a estatística de teste calcula-se ainda da mesma forma que para o teste da independência

$$\chi^2 = \sum \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

e tem a mesma distribuição sob H0 (um qui-quadrado com P-1 vezes K-1 graus de liberdade), apenas a H0 é que é diferente

Um caso especial de
GLMs
aplicados a
tabelas de contingencia

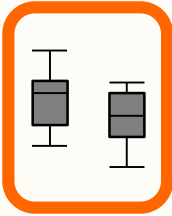




tabelas de contingência

Modelos log-lineares

- Metodologia mais usual para casos multivariados (geralmente para tabelas de dimensão superior a 2)
- Expressa cada célula como um somatório de diversos contributos (dos vários factores e interacções)
- Pretende-se concluir que um modelo mais simples que o saturado descreve bem a situação observada



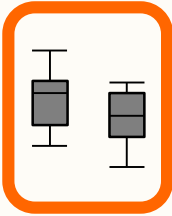
tabelas de contingência

Modelos log-lineares

Exemplo com 2 variáveis:

$$\ln E = [\theta] + [A] + [B] + [AB]$$

onde $[\theta]$ é a média dos logaritmos das frequências esperadas e A, B e AB aquilo que é devido aos factores A, B e interacção.



tabelas de contingência

Modelos log-lineares

Exemplo com 2 variáveis:

$$\ln E = [\theta] + [A] + [B] + [AB]$$

Se A e B forem independentes, $AB=0$ e então, segundo H_0 as frequências seriam dadas por:

$$\ln E = [\theta] + [A] + [B]$$

Mais um caso particular de um **GLM**!

1. distribuição Poisson (“contagens em caixas”)
2. função de ligação log

To learn more about loglinear models, we'll explore the following data from Agresti (1996, Table 6.3). It summarizes responses from a survey that asked high school seniors in a particular city whether they had ever used alcohol, cigarettes, or marijuana.

| Alcohol Use | Cigarette Use | Marijuana Use | |
|-------------|---------------|---------------|-----|
| | | Yes | No |
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |



| | cigarette | marijuana | alcohol | Freq |
|---|-----------|-----------|---------|------|
| 1 | yes | yes | yes | 911 |
| 2 | no | yes | yes | 44 |
| 3 | yes | no | yes | 538 |
| 4 | no | no | yes | 456 |
| 5 | yes | yes | no | 3 |
| 6 | no | yes | no | 2 |
| 7 | yes | no | no | 43 |
| 8 | no | no | no | 279 |



Link function by default for Poisson is the "log"

```
> mod0 <- glm(Freq ~ cigarette + marijuana + alcohol,
              data = seniors.df, family = poisson)
> summary(mod0)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|--------------|----------|------------|---------|--------------|
| (Intercept) | 4.17254 | 0.06496 | 64.234 | < 2e-16 *** |
| cigaretteyes | 0.64931 | 0.04415 | 14.707 | < 2e-16 *** |
| marijuanayes | -0.31542 | 0.04244 | -7.431 | 1.08e-13 *** |
| alcoholyes | 1.78511 | 0.05976 | 29.872 | < 2e-16 *** |

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2851.5 on 7 degrees of freedom
 Residual deviance: 1286.0 on 4 degrees of freedom

And finally introducing interactions between covariates

```
> mod1 <- glm(Freq ~ (cigarette + marijuana + alcohol)^2,
               data = seniors.df, family = poisson)
> summary(mod1)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|---------------------------|----------|------------|---------|--------------|
| (Intercept) | 5.63342 | 0.05970 | 94.361 | < 2e-16 *** |
| cigaretteyes | -1.88667 | 0.16270 | -11.596 | < 2e-16 *** |
| marijuanayes | -5.30904 | 0.47520 | -11.172 | < 2e-16 *** |
| alcoholyes | 0.48772 | 0.07577 | 6.437 | 1.22e-10 *** |
| cigaretteyes:marijuanayes | 2.84789 | 0.16384 | 17.382 | < 2e-16 *** |
| cigaretteyes:alcoholyes | 2.05453 | 0.17406 | 11.803 | < 2e-16 *** |
| marijuanayes:alcoholyes | 2.98601 | 0.46468 | 6.426 | 1.31e-10 *** |

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2851.46098 on 7 degrees of freedom
Residual deviance: 0.37399 on 1 degrees of freedom