UNIVERSIDADE ABERTA
www.uab.pt

Planeamento Experimental usando ANOVA de 1 e 2 fatores com R

– uma breve abordagem prática –

N. Sousa

ÍNDICE

Introdução

Este texto destina-se a expor o procedimento para realizar testes ANOVA de 1 e 2 fatores recorrendo ao software estatístico R. Em estatística, a ANOVA (*analysis of variance*) é um teste de hipóteses muito poderoso, que permite dar corpo a várias formas de planeamento experimental. Planeamentos experimentais, por vezes também designados por delineamentos experimentais (*design of experiments*), são formas de planear tarefas de recolha e tratamento de dados, por forma a identificar eventuais diferenças estatísticas entre os contextos em que os dados são recolhidos.

Em suma, o planeamento experimental pretende *enunciar* uma, ou várias, pergunta(s) sobre uma realidade a analisar, e a ANOVA é uma técnica que permite dar *resposta* a essa(s) pergunta(s).

Existem muitos tipos de planeamentos e técnicas de análise. Neste texto exploraremos apenas os planeamentos mais simples que se podem analisar pela técnica da ANOVA de 1 e 2 fatores, ressalvando que esta técnica permite também analisar alguns planeamentos mais complexos, como p.ex. planeamentos por blocos casualisados e outros. Falaremos também brevemente de uma alternativa não-paramétrica à ANOVA de 1 fator, o teste de Kruskal-Wallis.

O software R será a ferramenta de cálculo para todas estas técnicas. Recomenda-se por isso a leitura do texto de apoio em R "Visualização de dados e testes de hipóteses com R: uma breve abordagem prática" **[1]**, já disponibilizado no módulo anterior do DW *Tópicos de Estatística Experimental e Análise de Dados na WEB*.

Referências

[1] Sousa N (2016). "Visualização de dados e testes de hipóteses com R: uma breve abordagem prática." Repositório Aberto, Universidade Aberta (Ed.). Lisboa, Portugal.
Link: http://hdl.handle.net/10400.2/5952

[2] Oliveira A, Oliveira T (2011). "Elementos de estatística descritiva." Repositório Aberto, Universidade Aberta (Ed.). Lisboa, Portugal.
Link: http://hdl.handle.net/10400.2/1986

[3] Conover W, Iman R (1981). "Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics" *The American Statistician* 35(**3**):124-129.
Link:
https://www.researchgate.net/publication/247331493_Rank_Transformations_as_a_Bridge_Between_Parametric_and_Nonparametric_Statistics_Rejoinder

[4] Sawilowsky S (1990). "Nonparametric tests of interaction in experimental design." *Review of Educational Research* 60:91–126.
Link: dx.doi.org/10.3102/00346543060001091

ANOVA de 1 fator

Como vimos em **[1]**, a ANOVA de 1 fator permite comparar médias da mesma variável estatística entre amostras provenientes de vários grupos de observações (ou tratamentos). A melhor forma de apresentar este tema é recorrendo a um exemplo.

Um produtor de vinho no Douro tem três vinhas. O produtor pretende responder à questão

"Terão as três vinhas igual produtividade, ou haverá alguma (ou algumas) com produtividade diferente das outras"?

Para poder obter alguma forma de resposta, o produtor recorreu ao seguinte planeamento experimental: em cada das vinhas, anotou a produtividade por hectare de terreno nos últimos 5 anos, tendo obtido

Vinha	**São Gabriel**	**São Miguel**	**São Rafael**
Produtividade (ton/ha)	8.7, 8.3, 7.6, 6.0, 7.9	5.7, 8.0, 7.0, 8.7, 8.7	4.7, 4.0, 6.7, 6.3, 5.0
Média amostral ($\bar{x}_i$)	7.7	7.62	5.64

Em linguagem formal a pergunta pode ser expressa pela confrontação das hipóteses: [1]

$$H_0 : \forall_{ij}\ \mu_i = \mu_j \quad vs. \quad H_1 : \exists_{ij}\ \mu_i \neq \mu_j$$

em que os índices *i* e *j* indicam os grupos (1 = São Gabriel, 2 = São Miguel, 3 = São Rafael) e μ_i são as médias reais de produção das vinhas. A técnica estatística que aqui se pode usar para obter uma resposta, a um dado nível de significância, é a ANOVA de 1 fator.

Convém aqui frisar que há uma diferença entre as médias dos últimos anos (*médias amostrais*, i.e. média de cada amostra, designadas por $\bar{x}_i, i = 1,2,3$) e médias reais (*médias das populações*, i.e. média de *todas* as colheitas que se podem fazer na vinha, passadas e futuras, designadas por $\mu_i, i = 1,2,3$). C.f. a referência **[2]** para um breve esclarecimento sobre este ponto. O que o teste ANOVA faz é, através das amostras (cujas médias são $\bar{x}_i$), dizer algo sobre as médias de populações (μ_i).

Passemos então à execução em R do teste ANOVA.

Formas de entrada de dados

Previamente a correr a ANOVA, há que entrar os dados no R. No texto **[1]** foi apresentada uma forma de entrada de dados para ANOVA em R baseada em construção e justaposição de vetores pelo comando `data.frame`. Abaixo vamos recordar esse procedimento e apresentar uma alternativa.

Linha de comandos R

Para combinar as 15 observações usamos a função `c()`:

```
> prod = c(8.7, 8.3, 7.6, 6.0, 7.9, 5.7, 8.0, 7.0, 8.7, 8.7, 4.7,
4.0, 6.7, 6.3, 5.0)
```

Atenção que no R a vírgula é um separador de valores. O ponto é que é o separador decimal. Para gerar os grupos a que correspondem as observações temos de replicar cópias de strings de texto com `rep()`:

```
> vinha = c(rep("SGab",5),rep("SMig",5),rep("SRaf",5))
```

E finalmente construimos a tabela de dados com `data.frame`:

```
> dados = data.frame(prod,vinha)
> dados
   prod vinha
1   8.7  SGab
2   8.3  SGab
3   7.6  SGab
4     6  SGab
5   7.9  SGab
6   5.7  SMig
7     8  SMig
8     7  SMig
9   8.7  SMig
10  8.7  SMig
11  4.7  SRaf
12    4  SRaf
13  6.7  SRaf
14  6.3  SRaf
15    5  SRaf
```

Janela de edição do R

O R tem um ferramentas interativa para construir esta tabela, que é a função `edit`. Antes da edição há que dizer que a variável dados2 é um data frame (caso contrário o R devolve um erro), e só depois editar, i.e.:

```
> dados2 = data.frame(prod=numeric(),vinha=factor())
> dados2 = edit(dados2)
```

O parâmetro `prod=numeric()` indica que a produtividade vai ser um número e `vinha=factor()` que vinha será um fator, i.e. uma string.

Nota 1: tem mesmo de se escrever `dados2 = edit(dados2)`. Caso contrário quando fecharmos o editor nada será gravado na variável `dados2`. É verdade que isto não faz sentido, mas é assim que funciona ☺

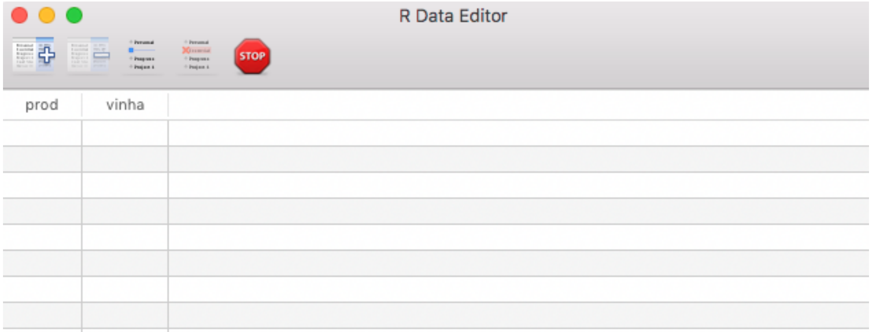
Nota 2: nas versões unix de R (Linux, MacOS) pode acontecer ter de se instalar um pacote de ferramentas gráficas. P.ex. se no R para MacOS ao editar aparecer a mensagem

```
> dados2 = edit(dados2)
```

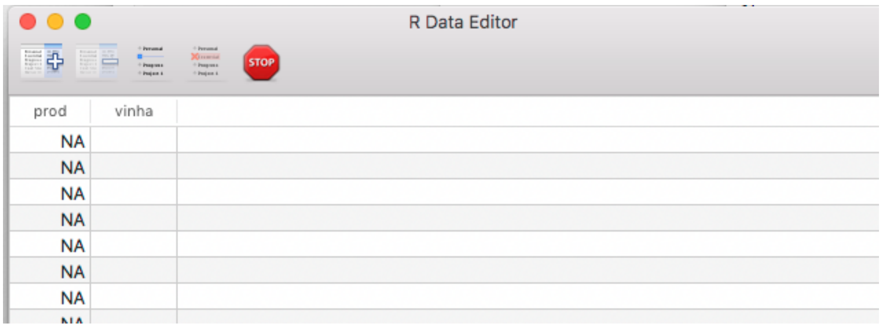
```
Error in check_for_XQuartz() :
  X11 library is missing: install XQuartz from xquartz.macosforge.org
```

então é ir ao endereço indicado e instalar o XQuartz.

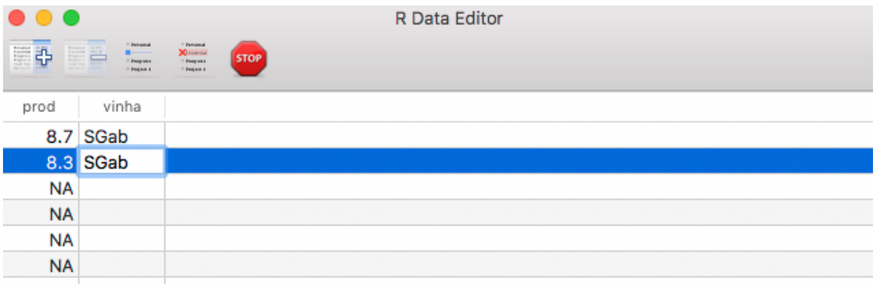
Mas se tudo correr bem, o editor abrirá e aparecerá uma janela mais ou menos com este aspeto:



Os botões ao lado do stop permitem adicionar/retirar linhas e colunas. Vamos precisar de adicionar agora 15 linhas. By the way.. o stop cancela as edições. A forma de sair é clicando no botão de fecho (canto superior esquerdo, a vermelho). Criando as linhas temos



E bastará agora fazer duplo clique ou usar “tab” nas colunas `prod` e `vinha` para entrar os dados:



Este duplo clique/tab é algo maçador e pouco prático. Mas sempre evita entrar os dados à mão. No final, como já foi dito, é clicar no botão vermelho ao canto. NÃO no stop!!

Nota sobre data frames

Como o leitor notará, praticamente todos os comandos R explorados abaixo contém o fragmento `prod ~ vinha, data = dados`. Este fragmento serve para indicar a relação entre os dados e a sua localização, mas só funciona se se tiver conseguido construir o data frame corretamente.

Ora acontece que alguns comandos do R para construir data frames (os procedimentos acima não são únicos, há mais) por vezes forçam todos valores a ser do mesmo tipo e pode acontecer que valores numéricos sejam inadvertidamente transformados em carateres (*strings*). Caso os comandos abaixo vos devolvam mensagens como p.ex.

```
Warning messages:
1: In model.response(mf, "numeric") :
  using type = "numeric" with a factor response will be ignored
2: In Ops.factor(y, z$residuals) : '-' not meaningful for factors
3: In Ops.factor(object$residuals, 2) : '^' not meaningful for
factors
```

isso quer dizer que houve algum percalço na criação do data frame, o que pode acontecer quando se criou o data frame com o comando `cbind`, comando esse que agrega colunas mas força-as a ser do mesmo tipo. E em caso string/número, passa tudo a string e o R deixa de ver p.ex. o valor `8.7` como um número, e passa a vê-lo apenas como um conjunto de carateres. O comando `is.numeric` permite ver se uma coluna de um data frame é numérica ou string:

```
> is.numeric(dados$prod)
[1] TRUE
```

O valor `TRUE` significa que a coluna `prod` do data frame `dados` é de facto numérica. Se o resultado tivesse sido `FALSE`, haveria que reconstruir o data frame. À partida o simples comando indicado acima, `dados = data.frame(prod,vinha)`, bastará para que a coluna `prod` saia numérica, mas se o leitor tiver problemas a este respeito, provavelmente será devido à questão acima.

Análise descritiva preliminar

Retirar resultados de um planeamento começa normalmente por uma análise descritiva preliminar. A mais elementar análise que se pode fazer é simplesmente comparar as médias amostrais. Se estas se revelarem muito díspares, é sinal de que poderá haver um, ou mais, grupos de caraterísticas diferentes. Para tirar médias num data frame existe a função `aggregate` que permite tirar médias aos três grupos de uma vez só:

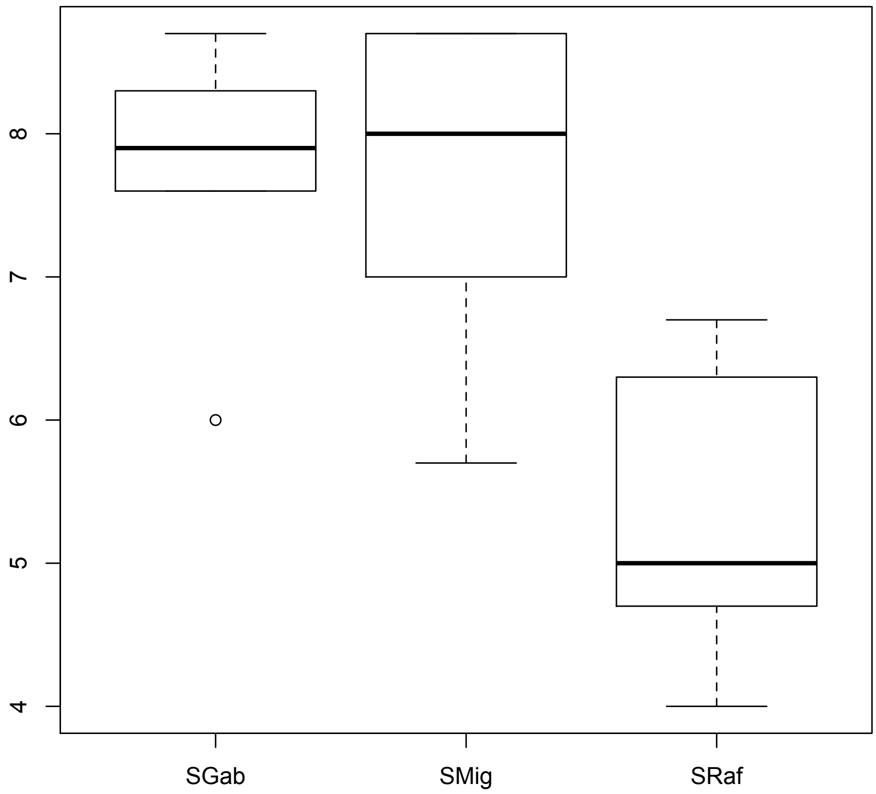
```
> aggregate(prod ~ vinha, data = dados, mean)
  vinha prod
1  SGab 7.70
2  SMig 7.62
3  SRaf 5.34
```

Em linguagem corrente, os parâmetros significam algo como "agregue as colunas `prod` e `vinha`, do data frame `dados`, e calcule a média amostral dessa agregação". Como vemos, temos aqui a suspeita de que a vinha de São Rafael está a produzir menos do que as outras duas.

Uma forma de visualizar graficamente a distribuição dos dados, separada por grupos, é usando o diagrama de extremos-e-quartis **[2]**. Desenhar este diagrama em R faz-se com a função `boxplot`:

```
> boxplot(prod ~ vinha, data = dados)
```

que nos devolve o gráfico abaixo:



As barras grossas dos diagramas representam a mediana das observações (valor tal que metade das observações são superiores a esse valor, e a outra metade inferior). Normalmente a mediana está perto da média amostral, pelo que também aqui temos a suspeita de baixa produção em São Rafael.

O teste ANOVA serve então para verificar se a média amostral baixa de São Rafael é indicadora de uma tendência de menor produção ou se é meramente o resultado de flutuações estatísticas.

Validação de pressupostos

Previamente à realização da ANOVA há que verificar dois pressupostos fundamentais: normalidade dos grupos (dentro de cada grupo as observações seguem uma distribuição normal) e homogeneidade das variâncias dos grupos (a dispersão dos dados em torno da média é igual para todos os grupos). Um terceiro pressuposto da ANOVA é a independência entre as observações, i.e. que uma observação não influencia a outras. Este pressuposto não pode ser verificado experimentalmente e tem de ser assegurado pelo investigador *a priori* **[1].**

Para verificar a normalidade podemos recorrer ao teste de Shapiro-Wilk novamente ao `aggregate`:

```
> aggregate(prod ~ vinha, data = dados, FUN = function(x)
shapiro.test(x)$p.value)
  vinha      prod
1  SGab 0.3912130
2  SMig 0.3220117
3  SRaf 0.6232539
```

Notar que foi preciso alterar ligeiramente o comando `aggregate`. Isto porque o comando `shapiro.test` devolve naturalmente não o valor de prova do teste (p-value) mas sim o valor de uma variável aleatória, *W*, que depois é usado para achar o p-value. Com o comando acima, os valores que aparecem sob a coluna prod são os p-values dos testes a cada grupo. Recordemos que p-values acima de 0,10 (10%) significam que a normalidade é plausível, valores entre 0,01 e 0,10 (1-10%) significam normalidade sob suspeita, e valores abaixo de 0,01 (1%) significam ausência de normalidade. No caso em mãos, todos os p-values são claramente acima de 10%, pelo que a normalidade dos grupos está validada.

Quanto à homogeneidade, aplica-se o teste de Bartlett. No R temos:

```
> bartlett.test(prod ~ vinha, data = dados)

     Bartlett test of homogeneity of variances

data:  prod by vinha
Bartlett's K-squared = 0.16277, df = 2, p-value = 0.9218
```

Com p-value de 92%, também a homogeneidade está validada.

Consequências da não-validação de pressupostos

A violação do pressuposto de normalidade é crítica se os grupos forem pequenos (abaixo de 30 observações) e não tão importante se forem grandes. A violação do pressuposto de homogeneidade pode levar a não sejam detetadas diferenças entre os grupos, que de facto existem. Seja como for, a ANOVA é um teste bastante robusto. Regra geral é apenas quando o resultado está na "zona cinzenta", i.e. de p-value 1-10%, que não se deve tirar conclusões se houver pressupostos não validados. O estudante é encorajado a reler **[1]** para compreender bem o significado do p-value de um teste estatístico.

Tabela ANOVA

Validados os pressupostos, procedemos então à construção da tabela ANOVA. No R a tabela é construída com o comando `anova`, aplicado ao modelo linear (`lm`) entre `prod` e `vinha`.

```
> anova(lm(prod ~ vinha, data = dados))
Analysis of Variance Table

Response: prod
          Df Sum Sq Mean Sq F value  Pr(>F)
vinha      2 17.957  8.9787  6.7593 0.01081 *
Residuals 12 15.940  1.3283
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A coluna `Pr(>F)` indica o p-value, que está muito perto do 1%. Confirma-se portanto a suspeita de que haja um grupo com média de população diferente dos outros.

Testes post-hoc

No caso acima é simples perceber que a vinha de São Rafael é o grupo desviante. Tal pode não ser tão evidente em casos mais complicados. Num caso geral, quando uma ANOVA indica que existem grupo(s) desviantes, existem testes complementares, ditos testes de comparação múltipla, ou testes *post-hoc*, que tentam identificar quais são esses grupos. Um dos mais usados é o teste de Tukey (*Tukey's honest significant difference*). No R esse teste é executado com o comando `TukeyHSD` sobre um objeto de classe `aov`:

```
> TukeyHSD(aov(lm(prod ~ vinha, data = dados)))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lm(prod ~ vinha, data = dados))

$vinha
           diff       lwr        upr     p adj
SMig-SGab -0.08 -2.024676  1.8646756 0.9933850
SRaf-SGab -2.36 -4.304676 -0.4153244 0.0180999
SRaf-SMig -2.28 -4.224676 -0.3353244 0.0220435
```

Notar que aqui tem mesmo de se usar `aov` e não `anova` (coisas do R...). As comparações revelam haver diferenças significativas (c.f. `p adj`, p-value, que estão perto de 1% em dois casos) entre São Rafael e as outras duas vinhas, e diferenças irrisórias entre São Miguel e São Gabriel. Assim, a conclusão que podemos tirar pode-se resumir formalmente por algo como

$$\mu_1 = \mu_2 > \mu_3$$

A partir desta conclusão o produtor de vinhos pode pensar em pesquisar razões para a menor produtividade de São Rafael.

Pressuspostos não-validados: teste de Kruskal-Wallis

Um teste alternativo quando não estão validados os pressupostos da ANOVA é o teste de Kruskal-Wallis. Este é um teste não-paramétrico, i.e. as suas hipóteses H_0 e H_1 não se referem a parâmetros da distribuição das variáveis estatísticas, como a média μ, mas sim a caraterísticas gerais das dessas distribuições:

H_0: os grupos têm a mesma distribuição vs.
H_1: há pelo menos um grupo com distribuição diferente dos restantes

O teste de Kruskal-Wallis pode se executado independentemente do tipo de dados, mas tem mais dificuldades em identificar diferenças entre grupos, i.e. regra geral devolve p-values mais elevados. No R este teste é muito simples de fazer:

```
> kruskal.test(prod ~ vinha, data = dados)

	Kruskal-Wallis rank sum test

data:  prod by vinha
Kruskal-Wallis chi-squared = 6.6676, df = 2, p-value = 0.03566
```

Como vemos, o teste de Kruskal-Wallis devolve um p-value maior do que a ANOVA, mas também ele perto de 1%, confirmando mais uma vez que (pelo menos) um dos grupos é diferente dos restantes. Também para este caso se podem realizar testes post-hoc, para tentar identificar o(s) grupo(s) desviantes. No entanto, o pacote base do R não tem nenhum comando para este efeito e é necessário importar um pacote (*package*) suplementar (p.ex. a package `PMCMR`).

Para finalizar esta secção, menciona-se que o teste de Kruskal-Wallis funciona transformando as observações em postos (*ranks*), após o que calcula um valor, *H*, e com ele obtém um p-value. Esta transformação de valores em postos chama-se *rank transform* e é uma das formas principais de relacionar a estatística paramétrica com a não-paramétrica **[3]**.

ANOVA de 2 fatores

Quando se suspeita que, além do tratamento, possam existir outros fator a influenciar os valores observados da variável estatística, entra-se no domínio da *ANOVA fatorial*. Esta é uma técnica que permite identificar eventuais diferenças entre grupos devidas a múltiplos fatores (e não apenas um) e que é adequada como ferramenta para análise de vários planeamentos experimentais.

Dada a vastidão do tema, neste breve texto exploraremos apenas com realizar em R uma ANOVA de 2 fatores e efeitos fixos. Regra geral, outros planeamentos podem ser efetuados recorrendo aos mesmos comandos R; apenas que depois a sua interpretação é ligeiramente diferente (como p.ex. é a diferença de interpretações entre a ANOVA 1 fator de efeitos fixos vs efeitos aleatórios). No final comentaremos brevemente o que fazer em caso de não-validação de pressupostos.

Para compreender as hipóteses em confrontação numa ANOVA de 2 fatores e efeitos fixos é conveniente reproduzir o modelo subjacente, que pode ser descrito da seguinte forma:

$$Y_{ijk} = \mu_{ij} + E_{ijk}, \qquad E \sim N(0, \sigma)$$

onde:

Y_{ijk} é a observação *ijk* da variável estatística em estudo, da qual vamos depois tirar amostras.
μ_{ij} é a média associada aos níveis *i* e *j* dos fatores 1 e 2, respetivamente.
E_{ijk} é o erro estatístico da observação *ijk*, i.e. a fonte de aleatoriedade das observações.

A equação acima é o que se chama um "modelo estatístico", ou seja, uma forma de tentar explicar matematicamente a origem de uma realidade observada (valores de Y_{ijk}). Normalmente μ_{ij} é decomposto adicionalmente da seguinte forma:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

μ é uma média global de Y_{ij} (na ANOVA de 1 fator a média era descrita por $\mu_i = \mu + \alpha_i$).
α_i é a influência do nível *i* do fator 1.
β_i é a influência do nível *j* do fator 2.
γ_{ij} é a influência combinada dos níveis *i* e *j* dos fatores 1 e 2, respetivamente.

Tendo em consideração a este modelo, a ANOVA de 2 fatores tenta então dar resposta a três testes de hipótese:

Teste ao fator 1: (não há influência do fator 1 vs. há influência do fator 1)

$$H_0: \forall_i \; \alpha_i = 0 \quad vs. \quad H_1: \exists_i \; \alpha_i \neq 0$$

Teste ao fator 2: (não há influência do fator 2 vs. há influência do fator 2)

$$H_0: \forall_j \; \beta_j = 0 \quad vs. \quad H_1: \exists_j \; \beta_j \neq 0$$

Teste à interação: (não há influência do fator 1 quando combinado com o fator 2, vs. há influência do fator 1 quando combinado com o fator 2)

$$H_0: \forall_{ij}\, \gamma_{ij} = 0 \quad vs. \quad H_1: \exists_{ij}\, \gamma_{ij} \neq 0$$

O termo de interação aparece para modelar situações em que os dois fatores, quando vistos isoladamente não influenciam a variável de resposta, mas que, quando combinados, já influenciam.

Vejamos agora estes conceitos em ação num problema concreto, novamente na área vitivinícula. Um produtor faz três tipos de vinho (Bairrada, verde, Porto), e guarda a sua produção em dois tipos de cascos (carvalho, pinho). O produtor deseja saber se o grau do vinho que produz depende, ou não, destes dois fatores: tipo de vinho e tipo de casco.

Entrada de dados

Para a entrada de dados, há que definir no R os dois fatores. Tal como para a ANOVA de 1 fator, a formatação em `data.frame` é a mais simples (embora haja outras formas de entrar e tratar os dados). Os dados que vamos usar deverão estar organizados num formato como o seguinte:

```
   grau tipoVinho tipoCasco
1    14  bairrada  carvalho
2    11  bairrada  carvalho
3    11  bairrada  carvalho
4    11  bairrada  carvalho
5     8  bairrada  carvalho
6    10  bairrada  carvalho
7    11  bairrada     pinho
8    11  bairrada     pinho
9    10  bairrada     pinho
10   17  bairrada     pinho
11   16  bairrada     pinho
12   14  bairrada     pinho
13   10     verde  carvalho
14   10     verde  carvalho
15    7     verde  carvalho
16    5     verde  carvalho
17    8     verde  carvalho
18    8     verde  carvalho
19    8     verde     pinho
20    7     verde     pinho
21    4     verde     pinho
22    6     verde     pinho
23    7     verde     pinho
24    8     verde     pinho
25   17     porto  carvalho
26   21     porto  carvalho
27   17     porto  carvalho
28   15     porto  carvalho
29   19     porto  carvalho
30   16     porto  carvalho
31   15     porto     pinho
32   21     porto     pinho
33   18     porto     pinho
```

```
34   16     porto     pinho
35   17     porto     pinho
36   19     porto     pinho
```

Temos portanto um data frame com três colunas: observação, nível do fator 1, nível do fator 2. Notar que temos 6 observações para cada combinação *ij* de fatores.

Entrar estes dados pode ser feito diretamente, mas não é muito simples fazê-lo dessa forma porque é preciso ter muuuuuita atenção a como se associa as observações aos níveis *ij* dos fatores. Assim, e para evitar erros, recomenda-se abrir a janela de edição e entrar os dados dessa forma. Para os temerários, os dados acima podem ser entrados diretamente com os seguintes comandos:

```
> grau = c(14, 11, 11, 11, 8, 10, 11, 11, 10, 17, 16, 14, 10, 10, 7, 5, 8, 8, 8, 7, 4, 6, 7, 8, 17, 21, 17, 15, 19, 16, 15, 21, 18, 16, 17, 19)
> tipoVinho = c(rep("bairrada",12),rep("verde",12),rep("porto",12))
> tipoCasco = c(rep(c(rep("carvalho",6),rep("pinho",6)),3))
> grauVinho = data.frame(grau,tipoVinho,tipoCasco)
```

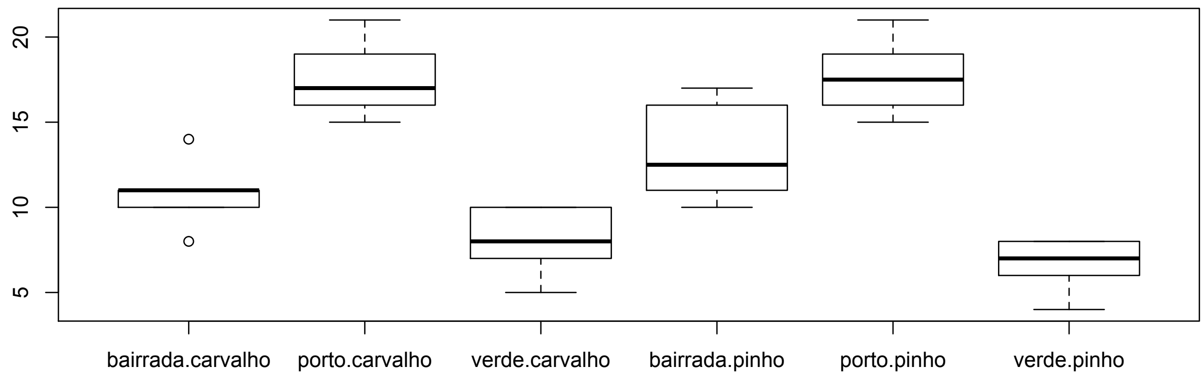
Existem outras formas de entrar sequencialmente os dados, recorrendo p.ex. ao comando `rbind`, mas o seu uso tem diversas subtilezas em que não vale a pena entrarmos agora.

Análise descritiva preliminar e interaction plots

A análise descritiva pode ser feita da forma usual com o comando `boxplot`, desta vez associado a dois fatores, separados pelo sinal de multiplicação `*`:

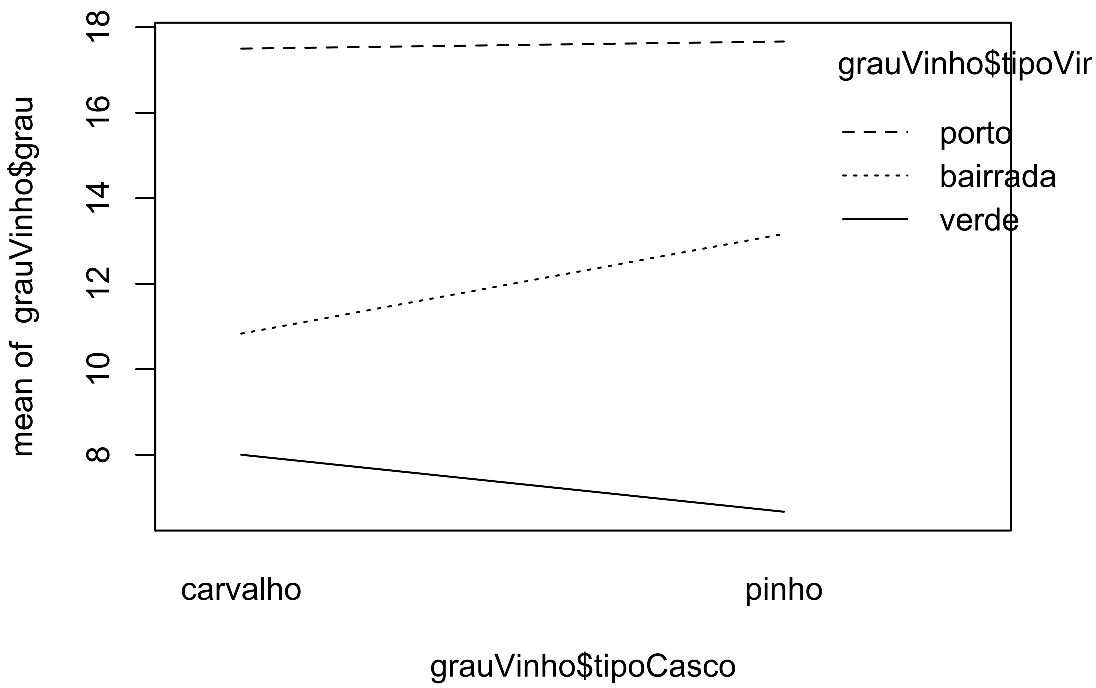
```
> boxplot(grau ~ tipoVinho * tipoCasco, data = grauVinho)
```

que nos devolverá



A representação em boxplot pode ajudar a identificar a influência dos dois fatores, mas isso nem sempre é fácil quando há dois fatores em jogo. Se for difícil interpretar o diagrama boxplot pode-se usar uma outra ferramenta gráfica para o efeito, o diagrama de interação (*interaction plot*). Vejamos na prática o que é e como funciona.

```
> interaction.plot(grauVinho$tipoCasco, grauVinho$tipoVinho,
grauVinho$grau)
```

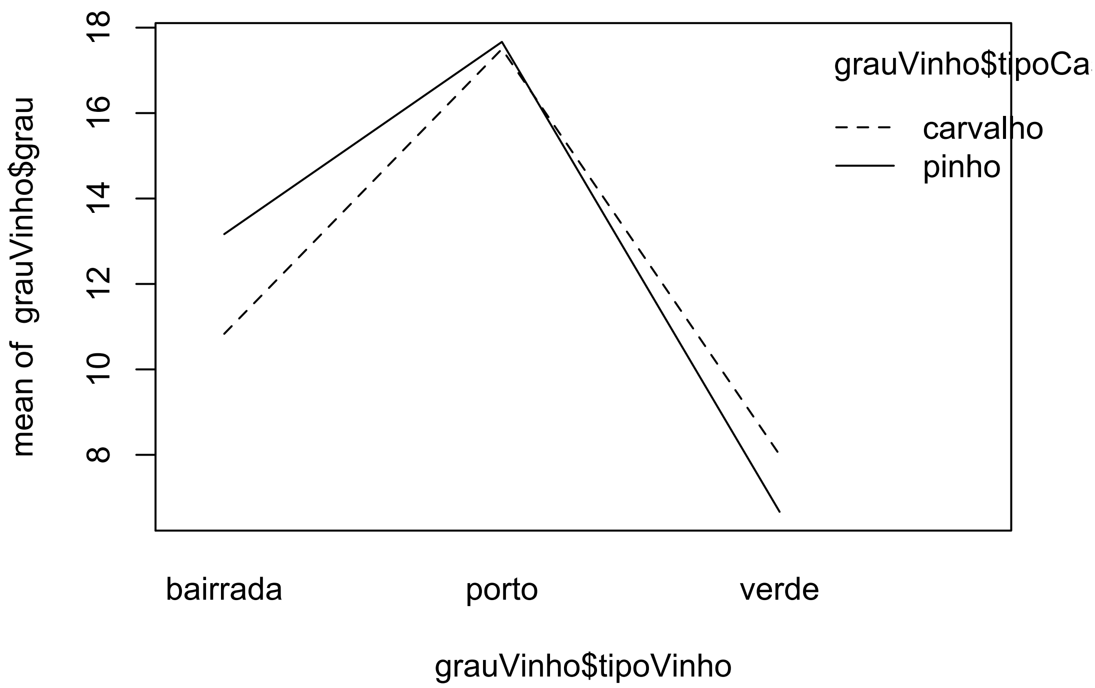


Há comandos para melhorar o aspeto do desenho, mas isso agora não interessa. Note-se também que o comando `interaction.plot` não permite a entrada de dados na forma mais simples `(tipoCasco,tipoVinho,Grau,data=grauVinho)`, pelo que é preciso invocar diretamente na forma `df$coluna`, em que `df` é o nome do data frame e `coluna` o nome da coluna pretendida.

Vejamos então o resultado do diagrama. Os pontos dos segmentos indicam os valores das médias dos grupos respetivos. P.ex. o grupo carvalho.verde (i.e. vinho verde maturado em casco de carvalho) tem média de cerca de 8 graus, o grupo pinho.porto média de cerca de 17,5 graus, etc. Claramente o que vemos é que apesar do tipo de casco aparentar ter ligeira influência na média da graduação dos vinhos, parece que o que o define o grau é mais o tipo de vinho.

O diagrama de interação pode também escrito com o tipo de vinho no eixo horizontal, bastando para isso trocar a ordem dos fatores:

```
> interaction.plot(grauVinho$tipoVinho, grauVinho$tipoCasco,
grauVinho$grau)
```



O diagrama novamente indicia que o tipo de vinho faz variar a média de graduação, ao passo que o tipo de casco parece não influenciar tanto. Ambos os diagramas apontam pois no sentido de que o tipo de vinho é um fator que influencia a graduação, enquanto que as diferenças quanto ao tipo de casco podem bem ser apenas flutuações estatísticas.

Quanto à interação, à partida não há grande evidência para sua existência. Haveria, se p.ex., as curvas no último gráfico tivessem perfis muito diferentes. Digamos, se as curvas fossem simétricas.

Notar que a ordem pela qual os fatores aparecem em `interaction.plot` é importante. Se tentarmos p.ex. colocar a variável estatística em 1º lugar, vai dar erro:

```
> interaction.plot(grauVinho$grau, grauVinho$tipoCasco,
grauVinho$tipoVinho)
Error in plot.window(...) : need finite 'ylim' values
In addition: There were 26 warnings (use warnings() to see them)
```

A ordem é estrita: primeiro vem o fator que queremos nos xx, segundo o fator a traçar e só depois a variável estatística em estudo (i.e. variável de resposta).

Munidos das conclusões descritivas preliminares, passemos então à ANOVA, que nos dirá quanto das diferenças de médias acima são indicadoras de tendências definidas ou de meras flutuações estatísticas.

Validação de pressupostos

Tal como para a ANOVA de 1 fator, há que validar pressupostos de normalidade e homogeneidade. Os comandos R são semelhantes aos do caso de 1 fator, agora com a fórmula de dados estendida.

Normalidade:

```
> aggregate(grau~tipoVinho*tipoCasco,data=grauVinho,FUN=function(x)
shapiro.test(x)$p.value)
  tipoVinho tipoCasco      grau
1  bairrada  carvalho 0.4326271
2     porto  carvalho 0.6590855
3     verde  carvalho 0.4659683
4  bairrada     pinho 0.3104755
5     porto     pinho 0.9636699
6     verde     pinho 0.2117055
```

Sem problemas aqui. Passemos à homogeneidade.

Homogeneidade:

Um pequeno problema aqui é que o comando `bartlett.test` não consegue lidar corretamente com a divisão de grupos em 2+ fatores:

```
> bartlett.test(grau ~ tipoVinho * tipoCasco, data = grauVinho)
Error in bartlett.test.formula(grau ~ tipoVinho * tipoCasco, data =
grauVinho) :
  'formula' should be of the form response ~ group
```

Assim sendo, temos primeiro de criar uma nova coluna (com nome p.ex. `bart`) no data frame `grauVinho` antes de invocar o teste de Bartlett. Há várias maneiras de fazer isso, mas a mais simples é talvez criar um n.º de grupo através de `rep` e inserir a nova coluna do data frame:

```
> grauVinho$bart=rep(1:6,each=6)
```

Este comando cria e preenche a nova coluna, `bart`. Note-se que `rep(1:6,each=6)` significa: "crie grupos de 1 a 6, cada um com 6 replicações". Num caso geral, se os *N* grupos tiverem todos a mesma dimensão *k*, o comando será `rep(1:N,each=k)`. Se os grupos tiverem dimensões diferentes, haverá que mudar o comando para algo como `c(rep(1,n1), rep(2,n2), ..., rep(N,nN))`. Uma alternativa será usar o comando `paste`:

```
> grauVinho$bart2=paste(grauVinho$tipoVinho,grauVinho$tipoCasco)
```

O comando `paste` é, no fundo, mais fácil de usar do que `rep`, mas a coluna gerada no data frame fica bastante confusa de ler (experimente ver data frame depois de `paste`...). Daí a opção por `rep`, que usaremos pelo resto deste documento.

Independentemente da forma como se define a coluna auxiliar, estando esta definida, podemos finalmente correr o teste de Bartlett:

```
> bartlett.test(grau ~ bart, data = grauVinho)
```

```
        Bartlett test of homogeneity of variances

data:  grau by bart
Bartlett's K-squared = 2.2628, df = 5, p-value = 0.8117
```

A homogeneidade é então plausível.

Validados os pressupostos, passemos então à execução da ANOVA.

Tabela ANOVA de 2 fatores e sua interpretação

Curiosamente, a ANOVA principal é bem mais simples de correr do que todos os preliminares! Basta apenas o simples comando:

```
> anova(lm(grau~tipoVinho*tipoCasco,data=grauVinho))
Analysis of Variance Table

Response: grau
                    Df Sum Sq Mean Sq F value    Pr(>F)    
tipoVinho            2 632.06 316.028 68.7848 6.221e-12 ***
tipoCasco            1   1.36   1.361  0.2963    0.5903    
tipoVinho:tipoCasco  2  20.39  10.194  2.2189    0.1263    
Residuals           30 137.83   4.594                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vejamos os resultados. A tabela contém três p-values: o primeiro é o resultado do teste ao fator tipo de vinho. Este p-value é extremamente baixo, pelo que é de rejeitar H0 (tipo de vinho não influencia) e considerar H1, i.e. que o tipo de vinho efetivamente influencia o grau.

O segundo p-value é o resultado do teste ao fator tipo de casco. O p-value está claramente acima de 10%, pelo que não devemos rejeitar H0, i.e. devemos considerar que o tipo de casco não influencia o grau.

Finalmente, o terceiro p-value é o resultado ao teste de interação. Também aqui o p-value acima de 10% indica que não devemos rejeitar H0, i.e. que não há interação. Atenção ao português nestes dois últimos p-values, que é algo traiçoeiro devido às duplas (ou triplas!!) negações: "não rejeitar a inexistência de influência de X" significa basicamente que podemos considerar que X não influencia os valores da variável de resposta.

Testes post-hoc

Havendo evidência de que há diferenças entre grupos, podemos querer identificar quais. O teste de Tukey permite-nos fazer isso com maior precisão do que a mera observação dos boxplots ou do diagrama de interação. O teste de Tukey completo pode ser invocado com o comando seguinte.

```
> TukeyHSD(aov(lm(grau~tipoVinho*tipoCasco,data=grauVinho)))
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

```
Fit: aov(formula = lm(grau ~ tipoVinho * tipoCasco, data = grauVinho))

$tipoVinho
                     diff        lwr       upr    p adj
porto-bairrada   5.583333   3.426058  7.740609 1.40e-06
verde-bairrada  -4.666667  -6.823942 -2.509391 2.65e-05
verde-porto    -10.250000 -12.407275 -8.092725 0.00e+00

$tipoCasco
                    diff       lwr      upr    p adj
pinho-carvalho 0.3888889 -1.070291 1.848069 0.590265

$`tipoVinho:tipoCasco`
                                         diff         lwr        upr     p adj
porto:carvalho-bairrada:carvalho    6.6666667   2.9026013 10.4307320 0.0001060
verde:carvalho-bairrada:carvalho   -2.8333333  -6.5973987  0.9307320 0.2296044
bairrada:pinho-bairrada:carvalho    2.3333333  -1.4307320  6.0973987 0.4300724
porto:pinho-bairrada:carvalho       6.8333333   3.0692680 10.5973987 0.0000729
verde:pinho-bairrada:carvalho      -4.1666667  -7.9307320 -0.4026013 0.0232836
verde:carvalho-porto:carvalho      -9.5000000 -13.2640653 -5.7359347 0.0000002
bairrada:pinho-porto:carvalho      -4.3333333  -8.0973987 -0.5692680 0.0167379
porto:pinho-porto:carvalho          0.1666667  -3.5973987  3.9307320 0.9999931
verde:pinho-porto:carvalho        -10.8333333 -14.5973987 -7.0692680 0.0000000
bairrada:pinho-verde:carvalho       5.1666667   1.4026013  8.9307320 0.0029420
porto:pinho-verde:carvalho          9.6666667   5.9026013 13.4307320 0.0000001
verde:pinho-verde:carvalho         -1.3333333  -5.0973987  2.4307320 0.8864597
porto:pinho-bairrada:pinho          4.5000000   0.7359347  8.2640653 0.0119507
verde:pinho-bairrada:pinho         -6.5000000 -10.2640653 -2.7359347 0.0001541
verde:pinho-porto:pinho           -11.0000000 -14.7640653 -7.2359347 0.0000000
```

Como vemos, o comando `TukeyHSD` devolve três tipos de p-values: um para comparações considerando apenas o tipo de vinho, outro para comparações do tipo de casco e um terceiro para comparações grupo-a-grupo, considerando todos os fatores.

Basicamente como só o teste ao tipo de vinho deu rejeição de H0, só é preciso olhar aos resultados dessas comparações, que estão nas primeiras três linhas de resultados. E para estas a conclusão é clara: todos os tipos de vinho são diferentes entre si e temos o ranking da graduação bem definido: Verde < Bairrada < Porto.

Caso de uma observação por célula e modelos sem interação

Por vezes a situação é tal que o investigador não vê razão para existência de interação entre os dois fatores principais. Também no caso de haver apenas uma observação por grupo *ij*, ou célula *ij*, não é possível incluír no modelo um termo de interação. Tanto num caso como no outro é possível elaborar uma ANOVA que não inclua interações. No R essa ANOVA é executada simplesmente trocando o sinal de multiplicação (*) por um de soma (+):

```
> anova(lm(grau~tipoVinho+tipoCasco,data=grauVinho))
Analysis of Variance Table

Response: grau
          Df Sum Sq Mean Sq F value    Pr(>F)
tipoVinho  2 632.06 316.028 63.9157 6.665e-12 ***
tipoCasco  1   1.36   1.361  0.2753    0.6034
```

```
Residuals 32 158.22   4.944
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nos casos sem replicação os testes de normalidade e homogeneidade de variância devem ser feitos não grupo-a-grupo (os grupos só têm uma observação), mas sim fator-a-fator, i.e.

```
> aggregate(grau~tipoVinho,data=grauVinho,FUN=function(x)
shapiro.test(x)$p.value)
  tipoVinho      grau
1  bairrada 0.1218285
2     porto 0.2485253
3     verde 0.4556177

> aggregate(grau~tipoCasco,data=grauVinho,FUN=function(x)
shapiro.test(x)$p.value)
  tipoCasco      grau
1  carvalho 0.3794905
2     pinho 0.3654139

> bartlett.test(grau~tipoVinho,data=grauVinho)

     Bartlett test of homogeneity of variances

data:  grau by tipoVinho
Bartlett's K-squared = 1.808, df = 2, p-value = 0.405

> bartlett.test(grau~tipoCasco,data=grauVinho)

     Bartlett test of homogeneity of variances

data:  grau by tipoCasco
Bartlett's K-squared = 0.25974, df = 1, p-value = 0.6103
```

Por ultimo, duas coisas: a primeira é que se se tentar incluir um termo de interação numa ANOVA de 2 fatores sem replicação, o que vai acontecer é que o R devolve uma mensagem de erro como a seguinte:

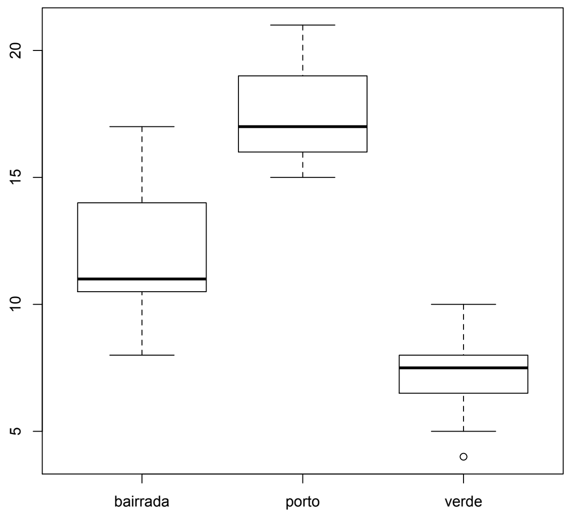
```
Warning message:
In anova.lm(lm(atest[, 1] ~ atest[, 2] * atest[, 3])) :
  ANOVA F-tests on an essentially perfect fit are unreliable
```

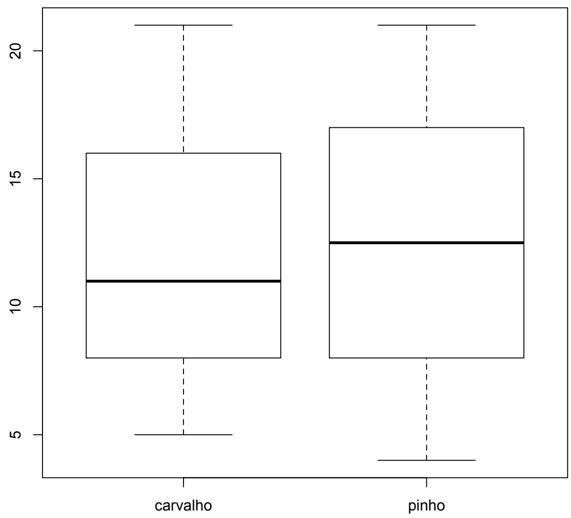
Isto acontece porque se está a tentar forçar a existência de uma interação num modelo estatístico que não a suporta. Substituindo o sinal * por + resolve-se este problema.

A segunda é que, dada a inexistência de interação em planeamentos sem interação, não faz sentido fazer diagramas de interação. Neste caso os estudos preliminares podem ser feitos fator-a-fator, de forma semelhante ao teste de Bartlett:

```
> boxplot(grau~tipoVinho,data=grauVinho)
> boxplot(grau~tipoCasco,data=grauVinho)
```

Note-se no entanto que estes comandos são meramente exemplificativos, uma vez que se referem a um planeamento com replicações sem interações. Por curiosidade, aqui se apresenta o resultado em termos de boxplots:





Não-validação de pressupostos – ANOVA por ranks

No caso de múltiplas violações grosseiras dos pressupostos haverá que recorrer a um método não-paramétrico para analisar o planeamento experimental elaborado. Ora acontece não há equivalente simples do teste de Kruskal-Wallis para dois fatores. Neste caso o que se pode fazer é seguir o procedimento sugerido por Conover e Iman **[3]** de substituir as observações da variável em estudo pelos seus postos (*ranks*) e correr uma ANOVA usual sobre eles. Este procedimento, denominado "transformação de posto" (*rank transform*), pode ser feito em R de uma forma extremamente simples: basta substituir, no comando da ANOVA, o `nome` da variável de resposta por `rank(nome)`! Vejamos o que acontece para o caso do grau dos vinhos:

```
> anova(lm(rank(grau)~tipoVinho*tipoCasco,data=grauVinho))
Analysis of Variance Table

Response: rank(grau)
                    Df  Sum Sq Mean Sq F value    Pr(>F)    
tipoVinho            2 3117.54 1558.77 74.2370 2.417e-12 ***
tipoCasco            1    3.36    3.36  0.1601    0.6919    
tipoVinho:tipoCasco  2   98.18   49.09  2.3379    0.1139    
Residuals           30  629.92   21.00                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A ANOVA sobre os postos pode depois ser complementada pelos testes post-hoc usuais, bastando para isso ter novamente o cuidado de usar sempre `rank(nome)` como variável de resposta. A interpretação destes testes não deve é ser em termos de médias da variável de resposta, mas sim, naturalmente, em termos do rank médio das observações dos grupos. I.e. "grupo A tem rank médio maior que grupo B" ou algo de semelhante.

A ANOVA sobre os postos também pode ser usada como uma alternativa ao teste de Kruskal-Wallis para o caso de 1 fator, com a vantagem de poder ser seguida de um banal teste de Tukey sobre os postos.

Crítica da ANOVA por ranks

A possibilidade de fazer uma ANOVA por ranks pode levar-nos a pensar que este método é uma boa solução sempre que os pressupostos da ANOVA não forem validados. Ora se por um lado é verdade que isso torna tudo muito mais simples, por outro lado há que notar duas coisas.

Em primeiro lugar, para o caso de planeamento de 1 fator, os p-values de uma ANOVA sobre os ranks são diferentes dos de um teste de Kruskal-Wallis, que é um teste exato. Apesar de, como Conover e Iman argumentaram, não haver grandes diferenças entre os dois testes, elas existem. Assim, é apenas recomendado recorrer a uma ANOVA RT se não houver na literatura um outro teste não-paramétrico para analisar o planeamento experimental em questão.

Em segundo lugar, foi descoberto (c.f. **[4]** p.104) que para os planeamentos fatoriais, como a ANOVA de 2 ou mais fatores, o p-value do teste de interação por ranks deflaciona artificialmente quando há rejeição de H0 nos fatores 1 e 2. Que quer isto dizer? Simplesmente que numa ANOVA de 2 fatores por ranks, se a evidência apontar para que ambos os fatores influenciem a variável de resposta, é possível que o teste à interação fique enviesado no sentido de rejeição. Ou seja, que detetamos interações quando elas na verdade não existem.

Este último problema pode ser resolvido da seguinte forma: se a ANOVA por ranks não der tripla rejeição de H0, então à partida está tudo bem e pode-se confiar nos resultados. Mas nos casos de tripla rejeição deve-se suspeitar do resultado ao teste de interação. Neste caso é ver o diagrama de interação para ajudar a decidir. Se não houver evidência clara de interação, a conclusão deve ser nesse sentido. E vice-versa. Em alternativa, se se quiser mesmo um p-value exato ao teste de interação, há que correr uma ANOVA específica para o efeito (ver **[4]** p.106), a "ANOVA-ART" (*aligned rank transform ANOVA*), disponível no pacote R `ARTool`.

Exercícios

1. **ANOVA de 1 fator: exercícios abstratos.** Para as alíneas abaixo elabore e execute uma ANOVA de 1 fator, e comente os resultados. Para esse efeito:

   - Entre os dados no R e realize um estudo descritivo preliminar.
   - Verifique os pressupostos da ANOVA e escolha o teste adequado; ANOVA ou Kruskal-Wallis.
   - Execute o teste acima escolhido e tire conclusões.
   - Se relevante, complemente as suas conclusões com testes post-hoc.

   Grupos e observações:

   a.
   ```
   A = c(2.2, 2.6, 2.3, 2.8, 3.6)
   B = c(3.2, 2.0, 4.1, 3.7, 4.0, 2.6)
   C = c(0.6, 2.9, 3.0, 2.2, 3.2, 2.5, 3.8)
   D = c(6.2, 3.6, 4.0, 3.3, 3.2)
   ```

   b.
   ```
   peras = c(32, 27, 31, 28, 34, 29, 26, 29, 30, 31)
   macas = c(43, 41, 39, 35, 39, 38, 36, 42, 43, 40)
   limoes= c(50, 53, 55, 49, 48, 50, 51, 50, 51, 50)
   ```

   c. Grupo G1: 9, 10, 7, 12
   Grupo G2: 22, 19, 20, 18
   Grupo G3: 21, 20, 20, 20, 21, 20
   Grupo G4: 29, 30, 30, 30, 29

   d. lisboa: 11 163 73 100 58
   porto: 193 199 210 191 199
   gaia: 172 201 240 209 213
   amadora: 307 309 291 293 282
   coimbra: 129 240 261 335 208
   braga: 143 153 158 150 152

   e. Tabela de dados:

I	J	K	L	M	N	P	Q	R	S
1.66	3.39	-1.73	0.69	1.21	-0.19	-0.05	3.68	1.36	-0.70
-0.10	1.33	-1.42	-1.84	2.16	-0.01	0.77	1.77	0.86	0.37
-0.84	2.04	0.88	-1.13	2.67	-0.52	1.20	0.80	1.62	-0.88

2. Resolva os exercícios do texto de apoio deste módulo recorrendo ao R, cujos enunciados são aqui resumidos para conveniência do leitor.

   a. Perdas de calor através de três tipos de portas cilindradas - planeamento experimental ANOVA de 1 fator e efeitos fixo (dados ao lado). Questões a responder:
   1. *"Haverá diferenças significativas nos três tipos de porta, no que refere à perda de calor?"*
   2. *"Se sim, que tipo de porta se destingue dos outros?"*

	Temperaturas	
A	**B**	**C**
10	12	12
11	12	12
8	11	15
10	8	10
8	10	14
	9	14
	10	

b. Máquinas de afinar telemóveis - planeamento experimental ANOVA de 1 fator e efeitos aleatórios (dados abaixo). Questão a responder: *"Levarão as máquinas analisadas igual tempo a afinar telemóveis?"*

Máquina Electrónica	Tempo de afinação (em minutos)				
1	20	17	18	21	20
2	21	20	18	20	21
3	23	21	20	20	18
4	21	18	21	22	20

3. Uma pessoa deseja comprar um automóvel. Para esse efeito analisa os consumos de 5 viaturas em diversos trajetos citadinos, tendo com isso chegado à seguinte tabela: (valores em litros/100 km)

Audi A3 1.2 T	6,1	7,9	7,1	6,0	8,5
Honda Insight híbrido	3,9	6,0	4,0	5,4	5,4
Renault Mégane 1.4 TCE	7,8	5,8	6,1	7,6	6,6
Volkswagen Golf 1.2 TSI	7,7	5,8	6,2	7,7	8,1
Toyota Auris híbrido	3,8	5,4	5,8	4,3	4,0

Indique a ANOVA mais adequada ao tratamento deste planeamento experimental. Em seguida, valide pressupostos, execute a ANOVA e, se possível, indique qual(is) a(s) viatura(s) estatisticamente mais económicas, a uma significância de 5%.

4. **ANOVA de 2 fatores: exercícios abstratos.** Para as alíneas abaixo elabore e execute uma ANOVA de 2 fatores adequada aos dados em causa e comente os resultados. Para esse efeito:

- Entre os dados no R e realize um estudo descritivo preliminar com diagramas de interação e boxplots; o que achar que ajude a compreender a distribuição dos dados.
- Verifique os pressupostos da ANOVA e escolha o teste adequado; ANOVA ou ANOVA RT.
- Execute o teste acima escolhido e tire conclusões quanto aos três tipos de hipótese em confronto.
- Se relevante, complemente as suas conclusões com testes post-hoc.

Grupos e observações:

a. Fatores f1f2, níveis f1=1234 f2=123, grupos de 3 observações.

```
f11=c(71, 60, 38)     f12=c(24, 70, 27)    f13=c(54, 81, 4)
f21=c(120, 63, 31)    f22=c(37, 9, 66)     f23=c(32, 56, 47)
f31=c(68, 33, 17)     f32=c(66, 117, 42)   f33=c(128, 42, 11)
f41=c(53, 37, 59)     f42=c(41, 60, 54)    f43=c(26, 53, 108)
```

b. Fatores AB, níveis A=12 B=12, grupos desiguais de 4/5/6/7 observações.

```
A1B1 = c(11.3, 10.4,  9.6,  7.0)
A1B2 = c(13.7, 11.3, 13.1, 13.8, 13.4)
A2B1 = c(9.6, 11.0, 11.5,  6.6,  6.7,  9.4)
A2B2 = c(16.3, 13.6, 14.6, 13.2, 17.3, 15.2, 16.7)
```

c. ANOVA de 2 fatores sem replicação

Indivíduo	manha	tarde	noite
Joao	1.3	2.2	3.1
Jose	4.6	5.5	6.4
Joaquim	7.9	8.8	9.7

d. Fatores f1f2, níveis f1=vermelho/verde/azul níveis f2=red/green/blue, grupos de 4 obs.

V.Estatística	Red	Green	Blue
Vermelho	5.41 7.93 6.81 6.47	2.97 3.00 2.12 3.60	3.31 3.11 2.58 1.61
Verde	1.90 3.87 3.81 2.66	7.55 7.36 9.41 6.46	2.29 2.57 3.69 3.30
Azul	2.99 2.93 2.23 1.20	3.48 3.05 1.99 3.30	7.33 7.82 5.91 5.95

e. Fatores XY, níveis X=12 Y=12, grupos de 4 observações.

```
Nr val X Y bart
1    1 1 1    1
2    1 1 1    1
3    1 1 1    1
4    2 1 1    1
5    1 1 2    2
6    2 1 2    2
7    2 1 2    2
8    2 1 2    2
9    3 2 1    3
10  19 2 1    3
11   4 2 1    3
12   3 2 1    3
13   5 2 2    4
14   4 2 2    4
15   4 2 2    4
16  12 2 2    4
```

5. Uma dona-de-casa formada em estatística decide testar se o tipo de detergente que usa nas lavagens e a temperatura da água são factores que influenciam a sujidade removida da roupa. Para tal, faz 6 grupos de 4 lavagens a cada uma das temperaturas “frio”, “quente” e “escaldar” e usando os detergentes “rapa-tudo” e “brancoso”, num total de 24 lavagens, tendo obtido as seguintes percentagens de sujidade removida:

Detergente\temp	Frio	Quente	Escaldar
Rapa-tudo	23, 43, 53, 45	74, 53, 93, 33	86, 58, 85, 93
Brancoso	25, 73, 63, 24	45, 74, 45, 83	93, 56, 89, 73

Que conclusões tirará a dona-de-casa destes dados?

6. Um professor de medicina descobre um papel suspeito no chão da sala no final de uma aula, contendo a tabela abaixo. Que consegue deduzir da informação na tabela?

Nome	**Matéria a estudar**	**Estudou?**	**Nota Anatomia**
Anselmo	Ossos	Sim	8
Bartolomeu	Ossos	Sim	8
Celestino	Ossos	Sim	5
Diana	Ossos	Sim	5
Eugénio	Ossos	Sim	6
Francisco	Músculos	Sim	11
Gertrudes	Músculos	Sim	11
Hélder	Músculos	Sim	13
Isabel	Músculos	Sim	13
Joaquim	Músculos	Sim	15
Kléber	Tecidos	Sim	9
Luís	Tecidos	Sim	13
Manuela	Tecidos	Sim	13
Nuno	Tecidos	Sim	10
Odete	Tecidos	Sim	15
Paulo	Ossos	Não	6
Quentino	Ossos	Não	10
Rui	Ossos	Não	9
Sandra	Ossos	Não	5
Tânia	Ossos	Não	8
Uriel	Músculos	Não	7
Vítor	Músculos	Não	8
Wilson	Músculos	Não	7
Xavier	Músculos	Não	4
Yolanda	Músculos	Não	8
Zélia	Tecidos	Não	8
Abel	Tecidos	Não	6
Bárbara	Tecidos	Não	8
Carla	Tecidos	Não	5
Dario	Tecidos	Não	6