

Estatística univariada

Os testes e estatísticas univariadas mais conhecidos e comumente aplicados são as regressões, testes- t e anovas. Embora haja uma infinidade de nomes e derivações para estas análises (e.g. ancova, anova em blocos, anova de medidas repetidas, teste-t pareado...), todas são apenas modelos lineares. Veremos que embora haja uma função para fazer teste-t (`?t.test`) e outra para fazer anova (`?aov`) todas essas

análises podem ser feitas usando apenas uma função, chamada `lm()`, que significa “linear models”. Portanto, **não se preocupe com os nomes de análises**, pois são todos modelos lineares. **Preocupe-se apenas em especificar o seu modelo corretamente.**

Em geral, a maioria das análises é feita especificando-se o modelo que se deseja testar. No R, para especificar um modelo é preciso usar a notação de formulas. Por exemplo, para um modelo de regressão com uma variável resposta **Y** (ou dependente) e uma variável preditora **X** (ou independente) a notação de formula a ser usada no R é: $Y \sim X$. Onde Y é a variável resposta e X é a variável preditora. O \sim (“til”) significa “em relação a” ou “modelado por”. Ou seja, **Y em relação a X**, ou **Y modelado por X**. Caso haja duas variáveis preditoras em seu modelo a formula será $Y \sim X1 + X2$. Isso medirá o efeito das duas variáveis preditoras. Caso queira avaliar também a interação entre X1 e X2 a formula será $Y \sim X1 * X2$. Veremos outras possibilidades mais adiante.

Para facilitar o entendimento da abordagem usando modelos lineares iremos começar por análises de regressão.

Regressão Linear Simples

A regressão linear simples é utilizada para analisar relações entre variáveis contínuas. Para exemplificar como fazer uma regressão no R vamos usar dados do número de espécies de plantas em diversas ilhas de Galápagos em relação ao tamanho das ilhas (figura 8.5 e 8.6 do livro de Gotelli e Ellison). Como feito no livro, usaremos os dados transformados em log10 da riqueza de espécies e o log10 da área das ilhas para fazer a regressão. Para isso importe o arquivo de dados galapagos.txt.

Para fazer a regressão no R a função é `lm()`, para linear models. Mais adiante vamos ver que testes-t, anovas também são todos modelos lineares e que podem ser analisados usando apenas a função `lm()`.

```
> galap<-read.table...
> riqueza<-galap[,3]
> area<-galap[,2]

> plot(area,riqueza) # Figura 8.5 do livro
```

Note que a relação não é linear e que um ponto assemelha-se a um outlier. Para linearizar os dados e reduzir o efeito do outlier transformamos os dados em log.

```
> riqueza.log<-log10(riqueza)
> area.log<-log10(area)
> plot(area.log,riqueza.log) # Figura 8.6 do livro
```

Agora vamos fazer a regressão usando a função `lm()`.

```
> resultado<-lm(riqueza.log ~ area.log) > resultado
```

Veja que o resultado da regressão mostra os coeficientes a (intercepto) e b (inclinação) da regressão. Para ver mais detalhes sobre o resultado da regressão precisamos usar a função `summary()`.

```
> summary(resultado)
```

Agora, além dos coeficientes da regressão, também são apresentadas outras estatísticas, como os

valores de t , probabilidades, os graus de liberdade e o r^2 .

Para apresentar os resultados da mesma forma que Gotelli e Ellison apresentam no livro (tabela 9.2) precisamos pedir uma tabela de anova do resultado da regressão. Para isso use:

```
> summary.aov(resultado)
```

Para inserir a linha de tendência da regressão use a função `abline()`. >

```
plot(area.log, riqueza.log)
```

```
> abline(resultado)
```

Regressão Múltipla

A regressão linear com uma única variável preditora X pode ser facilmente estendido a duas ou mais variáveis preditoras (regressão múltipla). O exemplo dado no livro de Gotelli e Ellison sobre quando duas ou mais variáveis preditoras são medidas em cada réplica. Por exemplo, em um estudo da variação na densidade de espécies de formigas em florestas de pântanos da Nova Inglaterra, foram medidas a latitude e a altitude em cada sítio amostral. Essas duas variáveis podem ser usadas em uma regressão múltipla. Para fazer, importe o conjunto de dados `formigas.txt`.

```
> formigas<-read...
```

```
> lat<-formigas[,1] # latitude
```

```
> alt<-formigas[,2] # altitude
```

```
> dens.formig<-formigas[,3] # densidade de formigas
```

```
> lm(log10(dens.formig) ~ lat+alt)
```

```
> reg.mult<-lm(log10(dens.formig) ~ lat+alt)
```

```
> reg.mult
```

```
> summary(reg.mult)
```

```
> summary.aov(reg.mult)
```

Caso queira ver o gráfico dos parciais desta regressão múltipla você precisará do pacote “car” e usar a função `avPlots()`.

```
> avPlots(reg.mult)
```