



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
Departamento de Engenharia Industrial
Rua Marquês de São Vicente, 225
22453-900 – Rio de Janeiro
Brasil

ENG1536 – Inferência Estatística

Laboratório – Guia de Estudo 3

Neste estudo, você aprenderá alguns gráficos introdutórios em R. Apesar da sua simplicidade, esses gráficos são amplamente utilizados em análise exploratória de dados e na produção de documentos científicos. Com eles, você poderá enriquecer os relatórios e as apresentações que fizer na faculdade e no estágio. Isso aumentará o seu poder de comunicação e a sua influência.

A primeira função gráfica que você deve aprender chama-se `plot`. No exemplo abaixo, desenhamos com ela um segmento de parábola, aproveitando para conhecer apenas alguns dos parâmetros mais úteis da função. No console do R, digite a expressão abaixo.

```
x <- seq(0,3,by=.01);  
plot(x,x^2,type="l",lwd=5,col="blue",main="Uma parábola",xlim=c(-1,4),ylim=c(-2,12))
```

Dentro da função, os dois primeiros argumentos `x` e `x^2` são as coordenadas horizontal e vertical dos pontos do gráfico. O parâmetro `type="l"` define o tipo de gráfico que desejamos, neste caso, um gráfico de linha; outros exemplos de uso são `type="p"` para um gráfico de pontos (esse é o padrão da função) e `type="h"` para um gráfico de barras verticais (*histogram*). O parâmetro `lwd=5` define a espessura (*line width*) da linha do gráfico. O parâmetro `col="blue"` define a cor do gráfico. O parâmetro `main="Uma parábola"` define um título principal para o gráfico. Os parâmetros `xlim=c(-1,4)` e `ylim=c(-2,12)` servem para definir os limites do eixo horizontal e vertical, respectivamente (observe que os limites são dados por vetores de duas dimensões, construídos com a função `c`). Você não precisa usar esses parâmetros se quiser apenas visualizar o gráfico de uma função, mas eles e muitos outros ajudam a produzir gráficos num padrão profissional.

Quando a função `plot` é chamada, a janela gráfica é criada do zero, apagando gráficos anteriores que você tenha criado. Você pode superpor um gráfico de linhas sobre um gráfico já pronto usando a função `lines`. Experimente digitar a expressão abaixo, Ainda com o gráfico da parábola na tela:

```
lines(x, 3*sqrt(x), lwd=12, col="red")
```

A próxima função que você deve aprender é extremamente útil em estatística. Quando estudamos uma amostra de números gerados por um mesmo processo estocástico, a primeira ferramenta que usamos para tentar visualizar a distribuição de probabilidade da variável aleatória que gerou os números é o **histograma**. É um gráfico que particiona o eixo horizontal em faixas de valores (tecnicamente chamadas de “células” do histograma) e representa no eixo vertical a quantidade de observações que caem dentro de cada faixa. Em R, a função que produz histogramas chama-se `hist`. Um exemplo:

```
amostra <- rnorm(10000); hist(amostra)
```

A primeira expressão no exemplo acima gera uma amostra de dez mil números normalmente distribuídos, com valor esperado zero e variância unitária (como aprendemos no Guia de Estudo 2). Na segunda expressão, a função `hist` constrói o histograma dessa amostra aleatória. Ele sugere fortemente a forma clássica de sino da função densidade de probabilidade normal. A função `hist` dispõe de um excelente algoritmo para determinar automaticamente um bom particionamento do eixo horizontal, mas o usuário pode modificá-lo.

Uma função interessante que complementa o histograma é `rug`. Ela acrescenta pequenos fios verticais (como aqueles de um tapete) ao eixo horizontal, para representar as observações individuais que são agrupadas no histograma. Veja numa amostra normal menor:

```
amostra <- rnorm(50); hist(amostra); rug(amostra, col="red", lwd=2)
```

Observe, no exemplo acima, que vários parâmetros que você aprendeu antes com a função podem ser usados dentro de outras funções gráficas. O valor padrão desses parâmetros gráficos pode ser alterado pelo usuário através da função `par`.

Assim como o histograma nos serve para vislumbrar a função densidade de probabilidade que gerou a amostra, a função `ecdf` (*empirical cumulative distribution function*) ajuda a construir um gráfico não decrescente que esboça a verdadeira função distribuição acumulada:

```
amostra <- rnorm(50); plot(ecdf(amostra))
```

O gráfico gerado pela expressão acima é uma escada com degraus localizados nos valores sorteados em `amostra`. A altura de cada degrau é igual a $1/50$, pois são 50 observações. Entre duas observações, a altura do gráfico é constante. À medida que o número de observações na amostra aleatória aumente, a forma desse gráfico tende para a curva suave da verdadeira distribuição normal acumulada. (Qual função em R você usaria para obter essa curva suave?)

Ao longo dos próximos estudos, você aprenderá outras funções e recursos gráficos em R. Agora resolva os exercícios na lista da próxima página.

Lista de exercícios 3

- (1) Experimente modificar os valores dos parâmetros que você aprendeu junto com a função `plot`, para observar o seu funcionamento. Depois, consulte a documentação da função e descubra o nome dos parâmetros que permitem modificar o nome dos eixos do gráfico.
- (2) Você aprendeu que o valor do parâmetro `col` em funções gráficas pode ser especificado com o nome da cor em inglês. Existem formas mais sofisticadas de especificar esse valor e ter acesso a uma paleta de cores muito mais variada. Uma dessas formas é a função `rgb`, que especifica qualquer cor através dos seus componentes vermelho (*red*) , verde (*green*) e azul (*blue*). Estude a documentação dessa função e modifique o exemplo da função `plot` de forma que o gráfico da parábola tenha cor amarela. (Existem inúmeros sites na Internet que fornecem o código RGB de qualquer cor, pesquise.)
- (3) Estude a documentação da função `hist` e descubra o parâmetro que faz com que o eixo vertical represente o percentual de observações da amostra dentro de cada célula (ao invés do número de observações). Depois, refaça o histograma de dez mil observações normais, usando essa nova representação e superponha a ele o gráfico da função densidade de probabilidade normal padrão, obtida com a função `dnorm` (Estudo 2). Use uma sequência de valores que começa em -3, termina em +3 e salta a cada 0,1.
- (4) Estude a documentação da função `hist` e descubra o parâmetro que permite alterar o número de células do histograma. Depois, construa o histograma de 1.000 sorteios normais, usando 4, 20, 50, 100 e 1000 células. Compare os gráficos.
- (5) Reconstrua o gráfico da função distribuição acumulada empírica de 50 observações normais. Depois, superponha a ele o gráfico da verdadeira função distribuição acumulada normal padrão, obtida com a função `pnorm` (Estudo 2). Use uma sequência de valores que começa em -3, termina em +3 e salta a cada 0,1.