

MBA⁺

**ARTIFICIAL INTELLIGENCE
& MACHINE LEARNING**

MBA⁺

PROGRAMANDO IA COM R

Prof. Elthon Manhas de Freitas

elthon@usp.br

2018

Revisão da última aula

- O que vimos na aula passada?
 - Concluir exercício Rmarkdown

Análise básica de um dataset

(usando o R como uma planilha Excel)

- Carregar o dataset BrFlights2.RData sites.google.com/usp.br/programando-ia-com-r
 - Trata-se de todos os voos comerciais brasileiros de 2016 e 2017
- Quais colunas podem ser transformadas em Fatores? Transforme-as.
- Crie coluna com o atraso da partida e o atraso da chegada.
- Crie coluna com distância euclidiana entre origem e destino.
- Crie coluna com tempo de viagem real.
- Como ver o primeiro quartil, média, mediana, etc. da coluna com o atraso na partida?
- Como ver o resumo da tabela toda?
- Qual companhia aérea com maior atraso médio?
- Qual estado de origem com maior atraso médio?
- Qual a relação média entre distância percorrida e tempo de vôo?
 - É possível identificar a companhia mais rápida?



Operações com Vetores e Matrizes

Aquecendo os motores – vetores



- Criar 6 vetores {a, b, c, d, e, ai}:
 - Vetor *a* de inteiros com 5 números:
 - Número da matrícula na FIAP (sem letras)
 - Data de nascimento, formato yyyyMMdd
 - 3 últimos dígitos do seu CPF
 - 6 primeiros dígitos do RG
 - 4 dígitos da placa do carro (pode inventar)
 - Vetor *b* com o dobro do vetor *a*
 - Vetor *c* com um terço de *b*
 - Vetor *d* com $c + a^2$
 - Vetor *e* com \sqrt{d}
 - Criar vetor *ai* com o inverso de *a*

```
Resposta dos exercicios de vetores FIAP
a <- c(3434, 19920929, 999, 341178, 408)
b <- a * 2
c <- b / 3
d <- c + a ** 2
e <- sqrt(d)
ai <- a[5:1]
```

```
• ai[1] <- a[5]
• ...
• ai[5] <- a[1]
```

Exercício – matrizes

```
## Criação de uma matriz aleatória
set.seed(123)
M1 <- matrix(rnorm(30), nrow = 6, ncol = 5)
M1
```

FIAP

- Criar matrizes:
 - $M1_{6 \times 5}$ com cada linha contendo o valor de um dos vetores
 - $M2_{5 \times 6}$ com cada coluna contendo o valor de um dos vetores
 - $M3_{6 \times 5}$ com cada elemento contendo 10% de $M1$, ou seja, $M1 \times 0,1$
- A tabela “attitude” é um data frame 30,7.
 - `class(attitude)`
 - `dim(attitude)`
 - Criar matriz $M4_{5 \times 6}$ sendo uma tabela com 5 linhas e 6 colunas quaisquer de `attitude`
 - Obs.: Converter de `data.frame` para `matrix` com o comando `as.matrix()`

```
## Resposta do exercício
M1 <- matrix(data = c(1,2,3,4,5), nrow = 5, ncol = 5)
M2 <- matrix(data = c(1,2,3,4,5), nrow = 5, ncol = 5)
M3 <- M1 * 0.1
M4 <- as.matrix(attitude[1:5, 1:6])
```

Operações com vetores e matrizes

- Todas as operações básicas, realizadas com número atômicos também podem ser feitas com vetores e matrizes (soma, subtração, etc.)

- `1 + 3`
- `1:5 + 3`
- `matrix(1:9, 3) + 3`
- `1 * 2`
- `1:5 * 2`
- `matrix(1:9, 3) * 2`

- Criar matriz (in memory) 19x19 com os valores 1 e 2 nos elementos, alternadamente.

- $$\begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

- Subtrair 1 de cada elemento desta matriz
- Somar 1 para cada elemento da linha 1, 2 na linha 2, ... e 19 na linha 19.
 - Dica: Use um vetor ;-)

```
Resposta
FIAP
> A = matrix(1:2, nrow = 19, ncol = 19)
> B = A + 1:19
```

Multiplicação de Matrizes

- $A_{la \times ca} \times B_{lb \times cb} = C_{la \times cb}$
- em que $ca = lb$

- Soma a multiplicação das linhas de A pelas colunas de B

- $c_{ij} = \sum_{k=1}^{ca} (a_{ik} \times b_{kj})$

- No R:

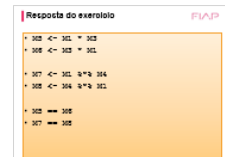
- `C <- A %*% B` ou `A %*% B -> C`

Exercício multiplicação de matrizes

- Criar matrizes:
 - $M5$ com cada elemento sendo a a multiplicação do elemento equivalente das matrizes $M1$ e $M3$. Ou seja $m5_{i,j} = m1_{i,j} \times m3_{i,j}$
 - $M6$ com cada elemento sendo a a multiplicação do elemento equivalente das matrizes $M3$ e $M1$. Ou seja $m6_{i,j} = m3_{i,j} \times m1_{i,j}$
 - Criar matriz $M7$ com a multiplicação das matrizes $M1$ e $M4$
 - Criar matriz $M8$ com a multiplicação das matrizes $M4$ e $M1$
- Comparar
 - $M5 == M6$?
 - $M7 == M8$?

Operador

$\oslash * \oslash$



- Matriz em que as linhas e colunas se invertem

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \Leftrightarrow A^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \dots & a_{m,1} \\ a_{1,2} & a_{2,2} & \dots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{bmatrix}$$

- Código no R:

`MT = t(M)`

- Obter a transposta $M1^T$ a $M8^T$ de cada uma das matrizes $M1$ a $M8$
- Comparar a multiplicação
 - $M8 \times M8^T$
 - $M8^T \times M8$

- diag
 - Cria uma matriz diagonal a partir de um vetor.
 - Pode ser usado para criar a matriz identidade
- solve
 - Obtém a matriz inversa M^{-1} de uma matriz quadrada
 - Também pode ser usada para obter $M^{-1} \times \alpha$

A conceptual image showing a business team of five people interacting with large, black, interlocking gears. One man on the right is standing on a small platform, holding a gear. A woman in the center is holding a gear up. Two men at the bottom are pushing or pulling gears. A man on the left is also interacting with a gear. The background is white, and the gears are large and prominent.

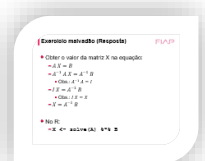
30 min reservados
para os exercícios
interativos + Desafio

Exercício malvado (Matemática, não R)

- Obter o valor da matriz X na equação:

$$- A X = B$$

- `A = as.matrix(attitude[21:25, 1:5])`
- `B = as.matrix(attitude[11:15, 1:5])`
- Primeiro resolver o sistema de equação matricial. Depois, como resolver no R.
- Voluntários?



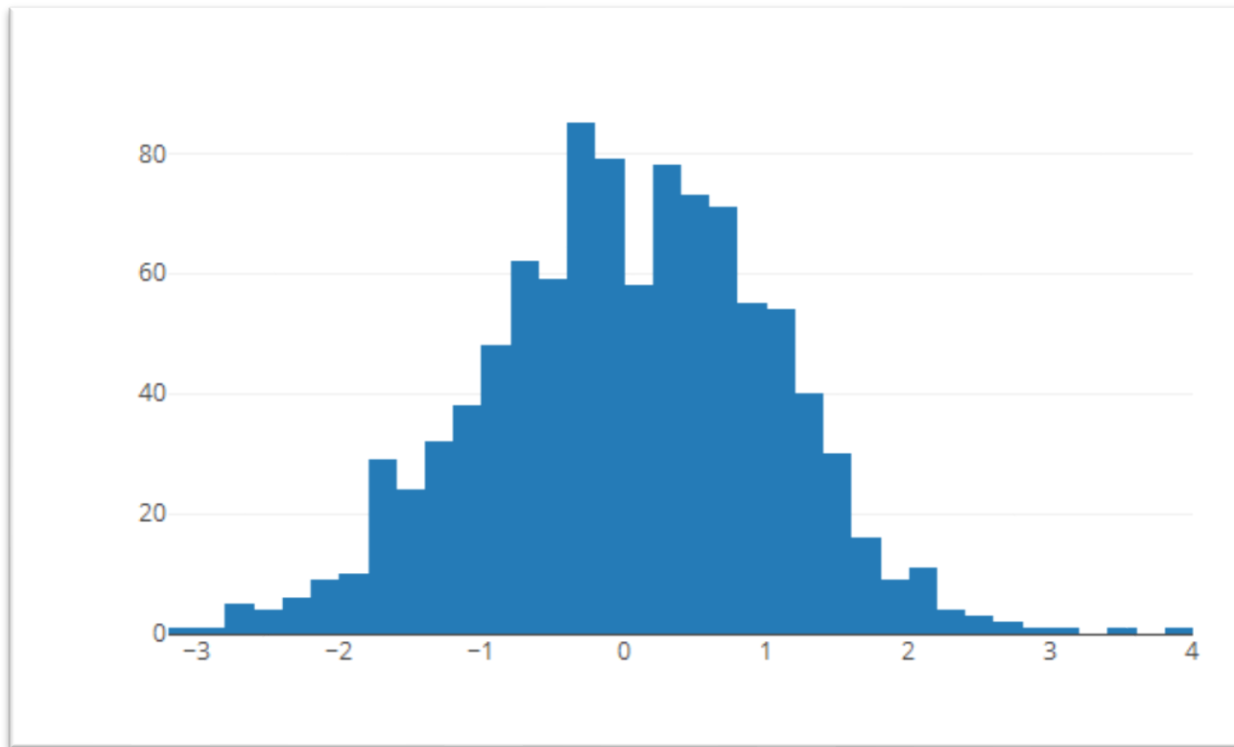


Amostras e Simulações

- Usado para quando se quer criar um conjunto fictício de dados amostrais

Gerando números aleatórios

- Família Normal



- Família normal

- `rnorm`: Para obter número aleatórios seguindo uma distribuição normal
- `dnorm` - Avalia a probabilidade da normal de um valor (dada a média μ e o desvio padrão σ)
- `pnorm` - Avalia a probabilidade ACUMULADA da normal de um valor (dada μ e σ)
 - Esta função deve formar uma curva sigmóide!

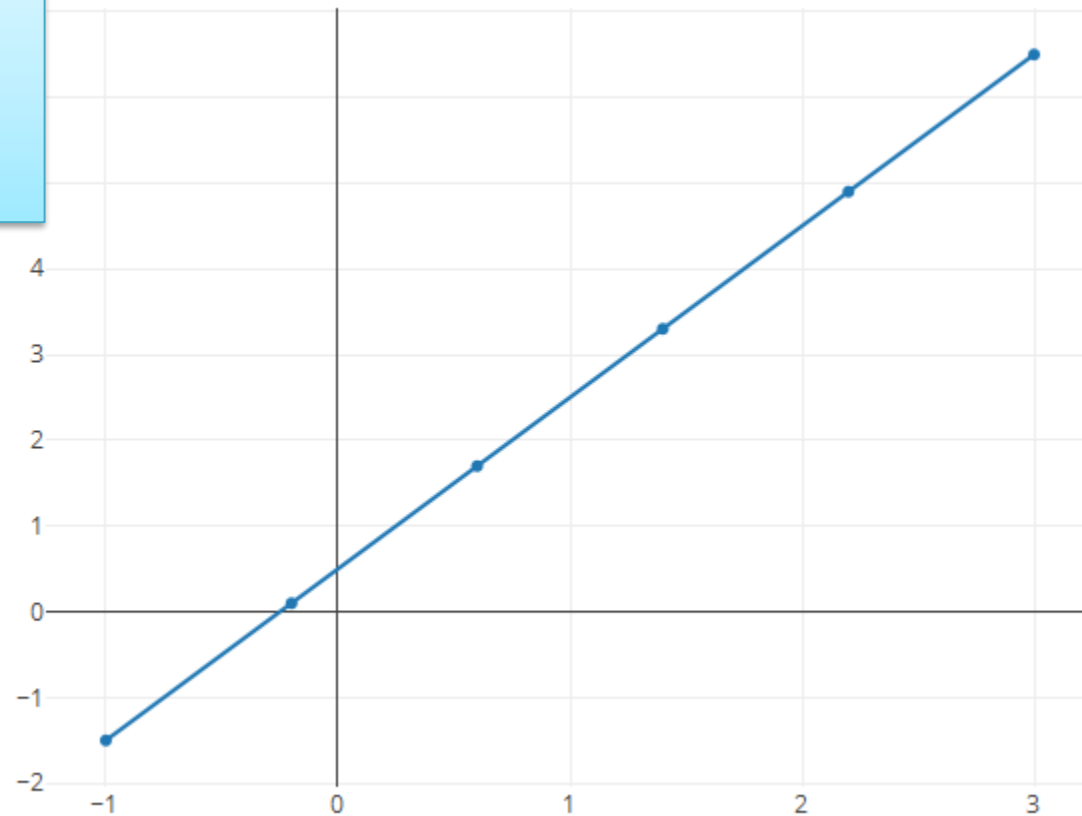
- Trabalhar com números aleatórios naturalmente faz com que se obtenha resultados igualmente aleatórios
- Quando se deseja que os resultados sejam reproduzidos (apesar da aleatoriedade), é preciso “plantar a semente” da aleatoriedade.
- para isso, se usa a instrução
 - `set.seed(seed)`

```
• set.seed(1)
• rnorm(5)
• rnorm(5)
• set.seed(1)
• rnorm(5)
```

Um modelo linear perfeito

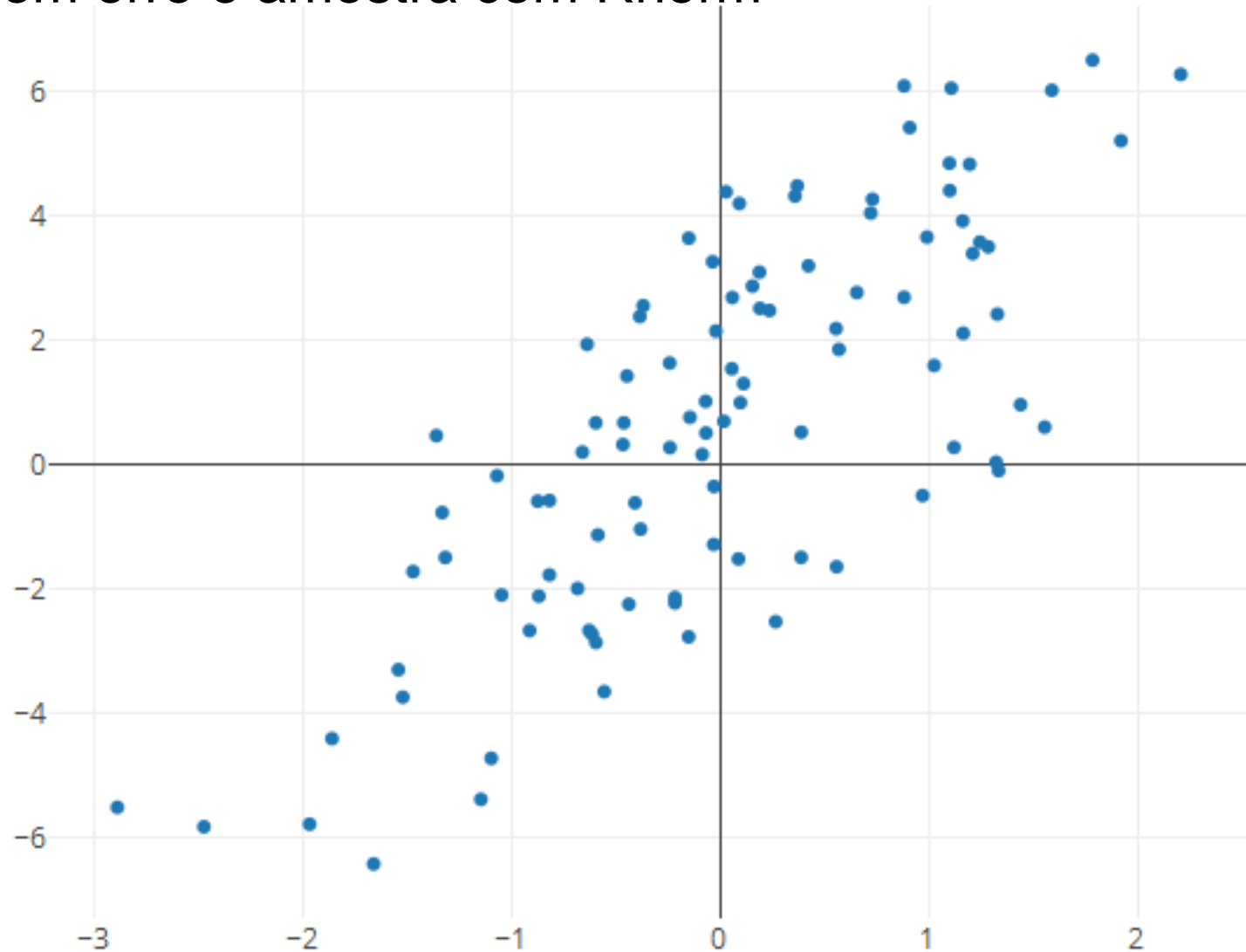
$$y = \beta_0 + \beta_1 x$$

- Em que
 - $\beta_0 = 0,5$
 - $\beta_1 = 2,0$
 - $x = 1$



Mas as amostras não são perfeitas

Plot com erro e amostra com Rnorm



Simulando um modelo linear

- Considere o seguinte modelo linear

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Em que

- $\beta_0 = 0,5$

- $\beta_1 = 2,0$

- $x \sim \mathcal{N}(0 ; 1^2)$

- $\varepsilon \sim \mathcal{N}(0 ; 2^2)$

Notação da função normal \mathcal{N} :

0 de média

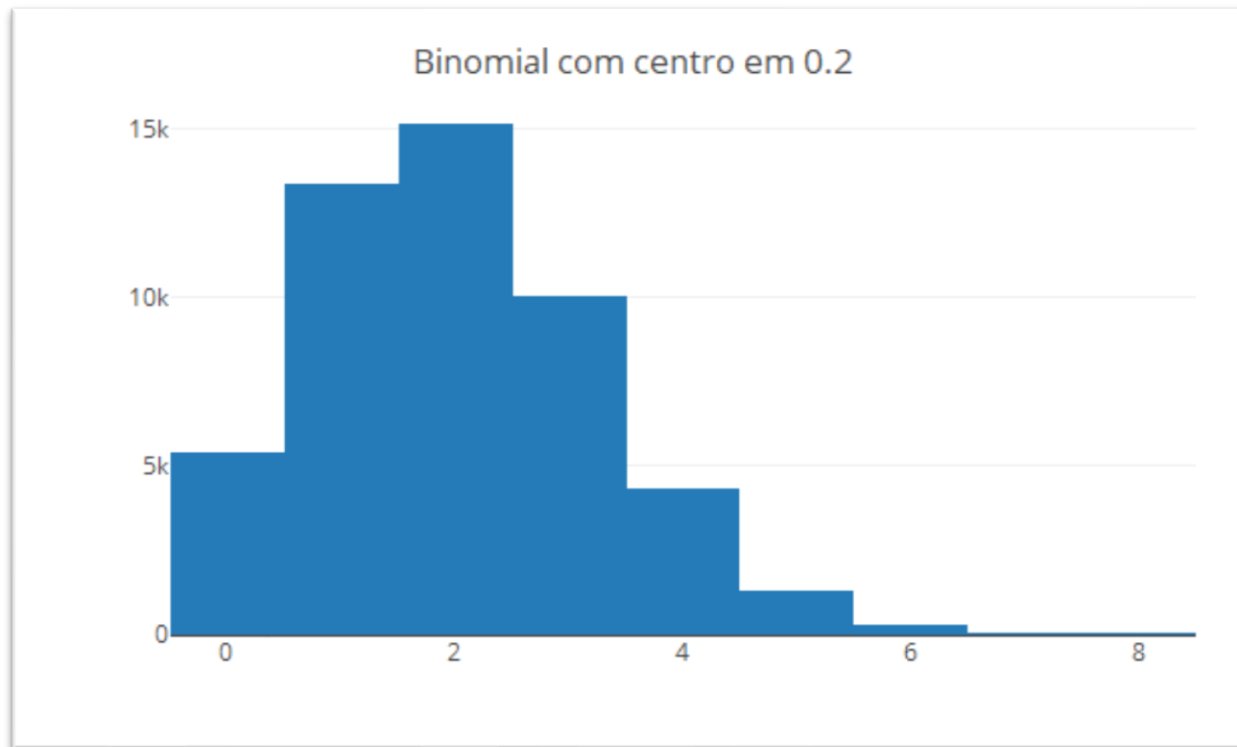
1² de variância
= 1 de desvio padrão

- Gerar os dados para o modelo linear descrito
 - Considerar 100 observações
 - Lembrar de “colocar a semente”
 - Plotar o resultado

```
• library(plotly)  
• plot_ly( x = x, y = y, type = "scatter" )
```

Gerando números aleatórios

- Família Binomial



- Família Binomial

- `rbinom`: Para obter número aleatórios seguindo uma distribuição binomial a partir de um tamanho t e dada uma distribuição de probabilidades p com $0 \leq p \leq 1$
- `dbinom` - Avalia a probabilidade binomial de um valor (dado um tamanho t e dada uma distribuição de probabilidades p)
- `pnorm` - Avalia a probabilidade ACUMULADA da binomial de um valor (dado t e p)
 - Esta função deve formar uma curva sigmóide!

Experimento binomial + normal

- A função binomial pode ser usada para se obter aleatórios binários

```
• set.seed(10)
• x <- rbinom(n = 100, size = 1, prob = 0.5)
```

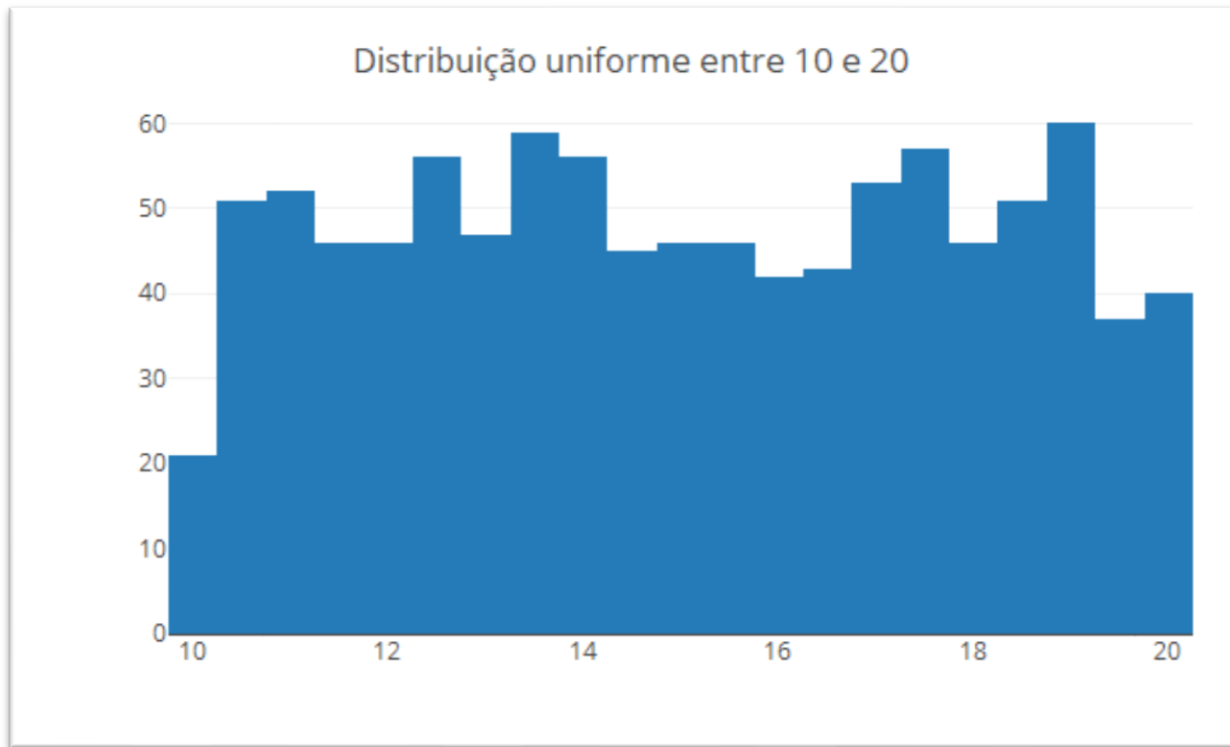
- Então seguimos com erros de amostragem como já fizemos antes

```
• e <- rnorm(n = 100, mean = 0, sd = 2)
• y <- 0.5 + 2 * x + e
```

- Poisson / Log-Linear deve ser visto nas aulas de estatística

Gerando números aleatórios

- Família Uniforme



- Família Uniforme

- `runif`: Para obter número aleatórios seguindo uma distribuição uniforme a partir de um mínimo *min* e um máximo *max*
- `dunif` - Avalia a probabilidade uniforme de um valor (dado um mínimo *min* e dado um máximo *max*)
- `punif` - Avalia a probabilidade ACUMULADA uniforme de um valor (dado *min* e *max*)
 - Esta função deve formar uma reta

Amostras a partir de um domínio

- comando `sample`

```
• set.seed(1)
• amostra = c( "T", "R", "I", "A", "N", "G", "U",
               "L", "O", "S")
• sample(x = amostra, replace = FALSE)
• sample(x = amostra, replace = TRUE)
• sample(x = amostra, size = 5)
• sample(x = amostra, size = 10, replace = TRUE,
         prob = c(1, 1, 5, 1, 1, 1, 1, 1, 1, 5))
```

Exercício: Amostra de uma base de dados

- Avaliar o data.frame “airquality”
 - Obter as dimensões deste data.frame
 - Obter o resumo de todas as colunas
 - exibir uma amostra dos 10 primeiros registros
 - exibir uma amostra com 10 linhas aleatórias
 - lembre-se de definir a semente para obter sempre o mesmo resultado

Resposta do exercício

```
dim(airquality)
summary(airquality)
head(airquality, n = 10)

set.seed(20)
linhas.idx <- seq_len(nrow(airquality))
linhas.sample <- sample(linhas.idx, 10)
airquality[linhas.sample, ]
```




Trabalho em grupo

Trabalho em grupo – Explorando em EDA

- Definição dos grupos
- Preencher formulário presente no site:
<https://sites.google.com/usp.br/programando-ia-com-r>
- Escolha dos 10 cases
 - (Agora, se tivermos 10 grupos)



Exercícios individuais

- Aprenda R no R**
- Portfólio individual**

lembrete: 2 pontos cada

MBA⁺

Copyright © **2018**

Prof. Elthon Manhas de Freitas

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).