

# Módulo 5 - Exercícios

Laboratório de Análise de Dados e Economia da Educação - educLAB

Fevereiro de 2021

## Sobre Probabilidade no R

- 1 - Qual e a probabilidade de 3 caras em 5 lançamentos de uma moeda honesta?
- 2 - Qual e a probabilidade de menos que 3 caras em 5 lançamentos de uma moeda honesta?
- 3 - Suponha que a probabilidade dos pais terem um filhos com cabelos loiros seja  $1/2$ . Se houverem 6 crianças na família, qual e a probabilidade de que metade delas terem cabelos loiros?
- 4 - Se a probabilidade de atingir um alvo num unico disparo é 0,3, qual é a probabilidade de que em 4 disparos o alvo seja atingido no mínimo 3 vezes?
- 5 - Plote, utilizando o R, os resultados de 2, 3 e 4.
- 6 - Suponha que o tempo necessario para atendimento de clientes em uma central de atendimento telefonico siga uma distribuição normal de média de 8 minutos e desvio padrão de 2 minutos.
  - Qual e a probabilidade de que um atendimento dure menos de 5 minutos?
  - E mais do que 9.5 minutos?
  - E entre 7 e 10 minutos?
  - 75% das chamadas telefonicas requerem pelo menos quanto tempo de atendimento?
- 7 - Considere  $z = 1.73$ 
  - Calcule  $P(Z < 1.73)$
  - Calcule  $P(0 < z < 1.73)$
- 8 - Seja uma distribuição normal  $X \sim N(3; 16)$ 
  - Calcule  $P(2 < X < 5)$
  - $P(X > 4)$
- 9 - Plote no R as distribuições de 6, 7 e 8

## Sobre Estatísticas usando o R

10 - Carregue o pacote ggplot2, utilize a base de dados "mpg" disponibilizada no pacote, identifique os tipos de variáveis e faça uma descrição dos dados. Para carregar a base de dados no R, após carregar o pacote ggplot2, utilize a função `data("mpg")`. Essa questão é livre para testar as mais diversas formas de descrever uma base de dados.

Dicionário da base de dados:

Variável	Significado
manufacturer	Marca
model	Modelo
displ	Cilindrada
year	Ano de fabricação
cyl	Número de Cilindros
trans	Tipo de marcha: Automática/ Manual
drv	Tração: f= frontal, r=traseira, 4=4x4
cty	Milhas por galão na cidade
hwy	Milhas por galão na estrada
fl	Tipo de combustível: r=regular, p=premium, d=diesel, e=ethanol, c=CNG(gás)
class	Tipo de Carro

Tabela 1: Dicionário\_mpg

## Sobre Regressão linear simples no R

Crie esses dois conjuntos de dados:

```
x <- c(1,2,3,4,5,6,7,8,9,10)
y <- c(2,5,8,5,3,9,10,12,9,10)
```

11 - Onde  $(x \sim y)$  (*variável resposta*  $\sim$  *variável preditora*). Após esse processo, Ajuste um modelo de regressão linear aos dados, estime dos parâmetros, **Erro** e **R2** do modelo e por fim expresse graficamente a função que ajusta a reta do modelo aos dados.

## Dicas

### Função Densidade (ou Probabilidade):

Calcula o valor da densidade, no caso de distribuições contínuas, ou a probabilidade  $P(X = x)$ , no caso de distribuições discretas, para cada elemento do vetor  $x$ . No R o nome dessa função é iniciado pela letra  $d$  mais o nome da distribuição (ex: dbinom, dpois etc.).

### Função Distribuição:

Calcula a distribuição acumulada:  $P(X \leq x)$ . O nome da função é iniciado pela letra  $p$  mais o nome da distribuição (ex: pnorm, pexp etc.).

### Função Probabilidade:

Calcula o valor de  $x$  correspondente a probabilidade  $p$  acumulada. É o inverso da função Distribuição. O nome da função é iniciado pela letra  $q$  mais o nome da distribuição (ex: qbeta, qcauchy etc.).

### Gerador aleatorio:

Gera numeros aleatórios baseados na distribuição definida. O nome da função é iniciado pela letra  $r$  mais o nome da distribuição (ex: rnorm, rbinom etc.).

Para mais distribuições, utilize o comando: `help.search("distribution")` e em seguida utilize: `?Nome da distribuição` para mais informações sobre a distribuição escolhida. Sendo assim, é importante revisar certas funções no R:

Distribuição	Sintaxe	Opções
-	<code>dbinom(x, n, p, opções)</code>	x: vetor contendo o n <sup>o</sup> de sucessos em n ensaios de Bernoulli
-	<code>pbinom(q, n, p, opções)</code>	p: probabilidade de sucesso
Binomial	<code>qbinom(prob, n, p, opções)</code>	q: vetor contendo os quantis em n ensaios de Bernoulli
-	<code>rbinom(obs, n, p)</code>	prob: vetor contendo as prob. em n ensaios de Bernoulli
-	-	obs: número de observações
-	-	n: número de ensaios de Bernoulli

Distribuição	Sintaxe	Opções
-	<code>dnorm(x, mean=0, sd=1, opções)</code>	x: vetor contendo os quantis
-	<code>pnorm(q, mean=0, sd=1, opções)</code>	mean: média da distribuição
Normal	<code>qnorm(p, mean=0, sd=1, opções)</code>	q: vetor contendo os quantis
-	<code>rnorm(n, mean=0, sd=1)</code>	p: vetor contendo as probabilidades
-	-	n: número de observações
-	-	sd: desvio padrão da distribuição