

# Relatório - Avaliação Prática 1

Marco Tulio Alves de Barros

2022-09-23

**Ex 1:** Diferentes levantamentos de dados foram realizados sem a definição prévia do erro amostral, porém é necessária essa informação para a divulgação adequada. Encontre qual o erro para cada uma das situações abaixo.

Funções úteis para calcular erro:

```
erro_quali = function(n0, pi, z) {  
  # z = 1.96 para confiança de 95% ou 2.58 para confiança de 99%  
  # pi = incidência  
  # n0 = população amostral  
  e = sqrt(z^2 * pi*(1-pi) / n0);  
  round(e * 100, 2);  
}  
  
erro_quanti = function(n0, s, z) {  
  # z = 1.96 para confiança de 95% ou 2.58 para confiança de 99%  
  # s = desvio padrão  
  # n0 = população amostral  
  e = sqrt(z^2 * s^2 / n0);  
  round(e, 2);  
}  
  
n0_isolado = function(N, n) {  
  # Correção de n0 dada uma população N e tamanho amostral final n  
  n0 = (n * N) / (N - n);  
  trunc(n0)-1;  
}
```

Item a)  $n = 850$ ,  $pi = 0.45$ ,  $nc = 95\%$ .

```
a = erro_quali(850, 0.45, 1.96); a
```

```
## [1] 3.34
```

Item b)  $n = 700$ ,  $N = 40000$ ,  $nc = 99\%$ .

```
n0 = n0_isolado(40000, 700);  
# Como não foi passado pi, por padrão adota-se pi=0.5 e erro qualitativo  
b = erro_quali(n0, 0.5, 2.58); b
```

```
## [1] 4.84
```

Item c)  $n = 1000$ ,  $N = 20000$ ,  $S = 18$ ,  $nc = 95\%$ .

```
n0 = n0_isolado(20000, 1000);  
c = erro_quanti(n0, 18, 1.96); c
```

```
## [1] 1.09
```

Item d)  $n = 380$ ,  $N = 5000$ ,  $S = 45$ ,  $nc = 99\%$ .

```
n0 = n0_isolado(5000, 380);  
d = erro_quanti(n0, 45, 2.58); d
```

```
## [1] 5.73
```

OBS: Nos itens A e B o erro em variáveis qualitativas é dado em porcentagem (%). Enquanto nos itens C e D, por se tratar de um erro para variáveis quantitativas, o valor obtido deve ter alguma unidade de medida (ex: reais, kgs)

**Ex 2:** Um estudo foi realizado para avaliar os gastos médios de compras realizadas em uma rede e os resultados são apresentados em prova1.xlsx. Com base nesse conjunto retire uma amostra de tamanho 30 com base nos últimos seis números de sua matrícula e responda as seguintes questões.

```
require(readxl)
```

```
## Carregando pacotes exigidos: readxl
```

```
dados_brutos = read_excel("C:/Users/marco/OneDrive/Área de Trabalho/AULAS/Estatística/prova1.xlsx");  
set.seed(560105);  
amostra = sample(dados_brutos$Gasto, 30); amostra
```

```
## [1] 207.77 231.73 132.65 215.76 156.45 229.43 255.39 151.43 184.55 155.96  
## [11] 208.14 258.51 194.00 255.73 260.97 165.71 208.48 183.44 205.89 224.49  
## [21] 259.72 264.53 192.20 181.15 147.68 239.70 220.73 208.18 257.90 217.12
```

Item a) Encontre a média, desvio padrão e variância.

```
media = mean(amostra);
mediana = median(amostra);
desvio = sd(amostra);
variancia = var(amostra);
rbind(media, mediana, desvio, variancia);
```

```
##           [,1]
## media      209.17967
## mediana    208.33000
## desvio     38.41662
## variancia 1475.83668
```

Item b) Qual o coeficiente de variação? Comente sobre ele.

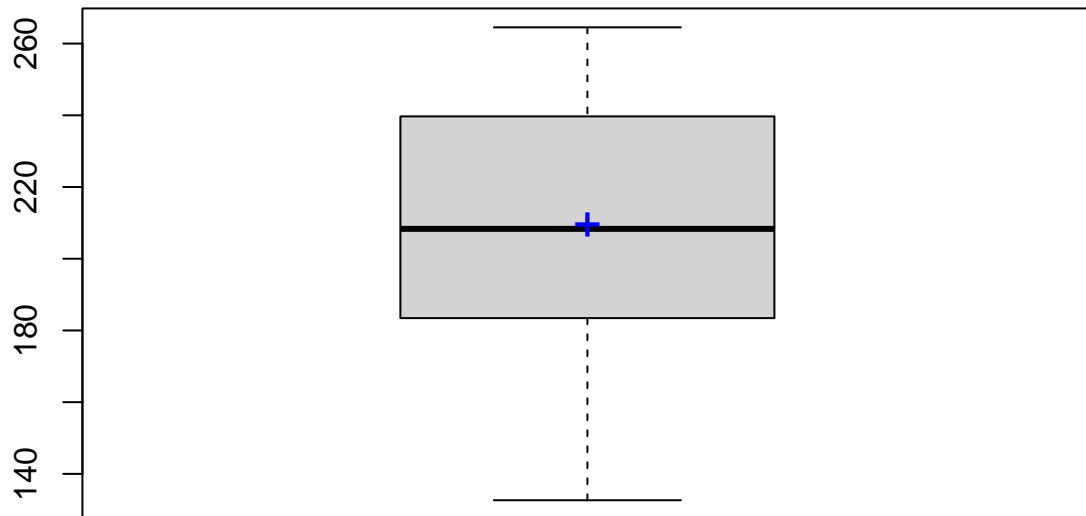
```
CV = round(desvio / media * 100, 2); CV
```

```
## [1] 18.37
```

Dado um coeficiente de variação de 18.37%, podemos concluir que a amostra possui uma *média dispersão* ( $15\% < CV < 30\%$ ), e portanto temos que o grupo selecionado é relativamente homogêneo, próximo do limite de baixa dispersão. Possivelmente apresenta poucos valores extremos, sendo que a maioria se mantém numa determinada faixa semelhante.

Item c) Construa o Box-plot e comente

```
boxplot(amostra);
points(media, pch="+", cex=1.5, col="blue")
```

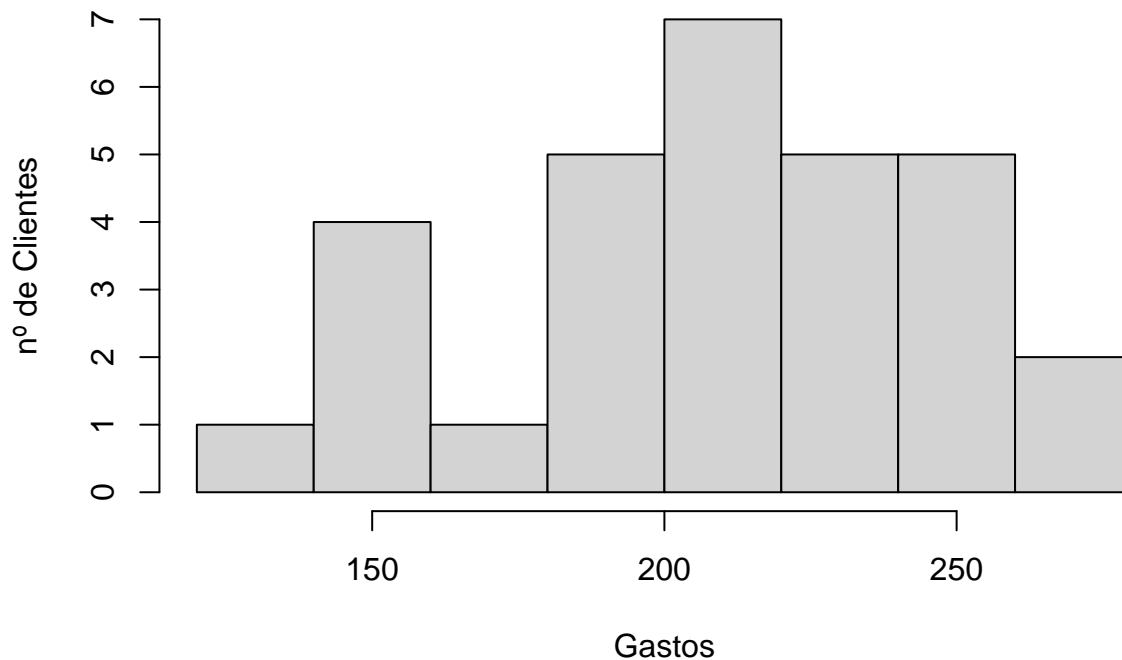


Analisando o box-plot podemos dizer que o primeiro quartil se encontra na faixa de gasto de 180, o segundo quartil (ou mediana) na faixa dos 210 e o terceiro quartil na faixa dos 240. Pode-se afirmar que a amostra se confirma relativamente simétrica, sem presença de elementos extremamente destoantes. Observa-se também que a maioria dos elementos se encontra na porção superior, visto que a caixa se encontra levemente deslocada para cima em relação ao eixo Y. A média e a mediana estarem tão próximas, cruz e linha central juntas, também é um bom indicativo da simetria no grupo.

**Item d) Construa o histograma.**

```
hist(amostra, main = "Histograma Gastos 560105", xlab = "Gastos", ylab = "nº de Clientes")
```

## Histograma Gastos 560105



Item e) Deseja-se realizar algumas promoções para alguns grupos de cliente levando-se em consideração os gastos. Indique quais as promoções serão feitas para cada tipo de cliente com base nas seguintes condições:

```
sorted = sort(amostra); sorted
```

```
## [1] 132.65 147.68 151.43 155.96 156.45 165.71 181.15 183.44 184.55 192.20
## [11] 194.00 205.89 207.77 208.14 208.18 208.48 215.76 217.12 220.73 224.49
## [21] 229.43 231.73 239.70 255.39 255.73 257.90 258.51 259.72 260.97 264.53
```

```
np = len * percentil ==> i (inteira) f (fracionária)
```

A promoção A será feita para os 20% clientes que menos gastam.

```
np = length(sorted) * 0.2; np # 6.0
```

```
## [1] 6
```

```
# Como i = 6 e f = 0, devemos fazer a soma da posição i com i+1 e dividir por 2
# (nesse caso i=6, logo s[6] e s[7])
yp = (sorted[6] + sorted[7]) / 2; yp
```

```
## [1] 173.43
```

Portanto, a promoção A deve se concentrar nos produtos que os clientes gastando abaixo da faixa dos 173.43 costumam consumir

**A promoção B será feita entre os 20% e 55% dos clientes que menos gastam.**

```
np = length(sorted) * 0.55; np # 16.5
```

```
## [1] 16.5
```

```
# Como f > 0, devemos pegar o valor na posição i+1 (nesse caso i=16, logo s[17])  
yp = sorted[17]; yp
```

```
## [1] 215.76
```

Assim, a promoção B deve focar nos produtos que os clientes gastando entre 173.43 e 215.76 consomem

**A promoção C será feita entre os 55% e 93% dos clientes que menos gastam.**

```
np = length(sorted) * 0.93; np # 27.9
```

```
## [1] 27.9
```

```
# Como f > 0, devemos pegar o valor na posição i+1 (nesse caso i=27, logo s[28])  
yp = sorted[28]; yp
```

```
## [1] 259.72
```

Desse modo, a promoção C deve abranger os produtos visados pelos clientes com gastos entre 215.76 e 259.72;

**A promoção D será feita para os 7% clientes que mais gastam.**

```
np = length(sorted) * 0.93; np # 27.9
```

```
## [1] 27.9
```

```
# Como f > 0, devemos pegar o valor na posição i+1 (nesse caso i=27, logo s[28])  
yp = sorted[28]; yp
```

```
## [1] 259.72
```

E por fim, a promoção D há de concentrar-se em atrair clientes gastando acima de 259.72, pois estes são os 7% que mais gastam

**Item f) Construa a tabela de frequências**

```

k = round(sqrt(length(sorted))); k # Quantidade de classes

## [1] 5

amp = (max(sorted) - min(sorted)) / k; amp # Intervalo das classes

## [1] 26.376

intervalos = c(min(sorted), min(sorted)+amp, min(sorted)+2*amp,
               min(sorted)+3*amp, min(sorted)+4*amp, min(sorted)+5*amp); intervalos

## [1] 132.650 159.026 185.402 211.778 238.154 264.530

# Definindo vetor de intervalos
tabela = hist(sorted, plot=F, breaks=intervalos, right=F);

GASTOS = c("132.650|-159.026", "159.026|-185.402", "185.402|-211.778",
           "211.778|-238.154", "238.154|-264.530");
           # Fazendo o nome das colunas a partir dos intervalos
fi = round(tabela$count); # Definindo a frequência absoluta simples
frp = round(fi/sum(fi)*100, 2); # Definindo a frequência relativa simples
Fi = cumsum(fi); # Definindo a frequência absoluta acumulada
Frp = cumsum(frp); # Definindo a frequência relativa acumulada

tabela_freq = data.frame(GASTOS, fi, frp, Fi, Frp); tabela_freq

##           GASTOS fi  frp Fi   Frp
## 1 132.650|-159.026  5 16.67  5  16.67
## 2 159.026|-185.402  4 13.33  9  30.00
## 3 185.402|-211.778  7 23.33 16  53.33
## 4 211.778|-238.154  6 20.00 22  73.33
## 5 238.154|-264.530  8 26.67 30 100.00

```

Item g) Compare os dados da amostra com os não selecionados quanto as suas médias e variabilidade. Em seguida escolha um dos dois grupos para serem os clientes especiais. Justifique sua escolha.

```

require(vecsets)

## Carregando pacotes exigidos: vecsets

resto = vecsets::vsetdiff(dados_brutos$Gasto, amostra); resto

## [1] 222.63 149.91 209.22 184.87 240.10 176.85 159.50 233.83 256.30 190.91
## [11] 221.59 276.78 235.59 254.39 191.89 192.14 193.66 175.00 143.30 221.82

```

```
# Função do pacote vecsets que vai ler o dados brutos e remover dela os
# elementos presentes na amostra, deixando assim um vetor resto com os 20 elementos
# não selecionados.
```

```
media_resto = mean(resto); # Média do resto
desvio_resto = sd(resto); # Desvio padrão do resto
variancia_resto = var(resto); # Variância do resto
```

```
# Montando colunas da tabela para facilitar comparação dos dados
```

```
Medidas = c("Média", "Desvio", "Variância");
Amostra = c(media, desvio, variancia);
Resto = c(media_resto, desvio_resto, variancia_resto);
```

```
comp = data.frame(Medidas, Amostra, Resto); comp
```

```
##      Medidas      Amostra      Resto
## 1      Média    209.17967    206.51400
## 2      Desvio     38.41662     36.80485
## 3 Variância   1475.83668   1354.59696
```

CONCLUSÃO: Visto que a variância e o desvio padrão do conjunto não selecionado na amostra são ligeiramente menores (+120 na variância e +2 no desvio), provavelmente escolheria o grupo de 20 elementos que não estão na amostra principal para serem clientes especiais. Apesar do resto apresentar uma média de gastos um pouco inferior (+3 unidades), uma variância e desvio menores são indicativos de um grupo mais homogêneo e menos disperso. Assim, acredito que seja que melhor trabalhar com um grupo mais homogêneo.

**Item h) Assumindo que esse estudo pode ser considerado como um estudo piloto, utilize as informações obtidas para calcular qual o tamanho amostral levando-se em consideração um erro amostral de +- R\$3.50, um nível de confiança de 95% em uma população de 45000 consumidores.**

```
N = 45000; # população total de consumidores
e = 3.5; # Erro
z = 1.96; # Valor para nível de confiança de 95%
```

```
## Variável quantitativa, e já temos o desvio padrão, portanto calcula-se n0 por:
```

```
n0 = z^2 * sd(sorted)^2 / e^2; n0
```

```
## [1] 462.8224
```

```
# Corrigindo para o tamanho da população
```

```
n = trunc(N * n0 / (N + n0)) + 1; n
```

```
## [1] 459
```

Portanto 459 é o tamanho amostral desse estudo piloto