

Exercícios em computador

C1 Use os dados do arquivo WAGE1 para este exercício.

```
library(wooldridge)
data("wage1")
```

(i) Encontre o nível de escolaridade médio da amostra. Quais são os menores e os maiores valores de anos de educação?

```
medio <- mean(wage1$educ)
menor <- min(wage1$educ)
maior <- max(wage1$educ)
```

- Nível de escolaridade médio: 13
- Menor valor de educação: 0
- Maior valor de educação: 18

(ii) Encontre o salário-hora médio da amostra. Ele parece alto ou baixo?

```
salario_medio <- (mean(wage1$wage))
```

(iii) Os dados salariais são reportados em dólares de 1976. Pesquisando na Internet ou em uma fonte impressa, encontre o Índice de Preços do Consumidor (IPC) para os anos de 1976 e 2013.

| Ano | IPC |
|------|-------|
| 1976 | 56.9 |
| 2013 | 233.0 |

(iv) Use os valores do IPC da parte (iii) para encontrar o salário-hora médio em dólares de 2013. Agora o salário-hora médio parece razoável?

```
salario_medio * (233 / 56.9)
```

```
## [1] 24.14397
```

(v) Quantas mulheres existem na amostra? E quantos homens?

```
tabela <- table(wage1$female)
names(tabela) <- c("Homem", "Mulher")
tabela
```

```
## Homem Mulher
##      274    252
```

C2 Use os dados do arquivo BWGHT para responder a essas questões.

```
data("bwght")
```

(i) Quantas mulheres existem na amostra e quantas relataram fumar durante a gravidez?

```
qtde_mulheres <- nrow(bwght)
qtde_nao_fumantes <- sum(bwght$cigs > 0)
```

Existem 1388 na amostra e 212 relataram fumar durante a gravidez.

(ii) Qual é o número médio de cigarros consumidos por dia? A média é uma boa medida da mulher “típica” neste caso? Explique.

```
# media arredondada
round(mean(bwght$cigs))
```

```
## [1] 2
```

O número médio de cigarros consumidos por dia não é condizente com a amostra. Cerca de 85% das mulheres não fumaram durante a gravidez. Isso ocorre pois o valor atribuído quando a mulher não é fumante é 0.

(iii) Entre mulheres que fumaram durante a gravidez, qual é o número médio de cigarros consumidos por dia? De que forma isso se compara com sua resposta ao item (ii) e por quê?

```
# media
mean(subset(bwght, cigs > 0)$cigs)
```

```
## [1] 13.66509
```

Esse número é muito maior do que no caso anterior, porque excluímos 85% das observações da amostra que eram 0.

(iv) Encontre a média de fatheduc na amostra. Por que somente 1.192 observações são usadas para calcular essa média?

```
# subconjunto da amostra que contem as observacoes que possuem NA na variavel fatheduc
nrow(subset(bwght, is.na(fatheduc)))
```

```
## [1] 196
```

```
# media
mean(bwght$fatheduc, na.rm = TRUE)
```

```
## [1] 13.18624
```

Parte das observações na variável de educação dos pais possui NA, ou seja, não foi possível obter essa informação no momento da captura dos dados

(v) Relate a renda média familiar e seu desvio padrão em dólares.

```
# media
mean(bwght$faminc)
```

```
## [1] 29.02666
```

```
# desvio-padrao
sd(bwght$faminc)
```

```
## [1] 18.73928
```

C3 Os dados existentes no arquivo MEAP01 são do estado de Michigan no ano de 2001. Use estes dados para responder às seguintes questões.

```
data("meap01")
```

(i) Encontre os maiores e os menores valores de math4. Essa variação faz sentido? Explique.

```
# valor maximo
max(meap01$math4)
```

```
## [1] 100
```

```
# valor minimo
min(meap01$math4)
```

```
## [1] 0
```

(ii) Quantas escolas têm uma taxa de aprovação perfeita no teste de matemática? Que porcentagem da amostra total isso representa?

```
# aprovacao perfeita = 100
round(100 * mean(meap01$math4 == 100), 2) # porcentagem arredondada
```

```
## [1] 2.08
```

(iii) Quantas escolas têm taxas de aprovação em matemática de exatamente 50%?

```
# quantidade escolas com 50% de aprovacao em matematica
sum(meap01$math4 == 50)
```

```
## [1] 17
```

(iv) Compare as taxas médias de aprovação em matemática e leitura. Qual teste tem a aprovação mais difícil?

```
apply(meap01[, c("math4", "read4")], FUN = mean, MARGIN = 2)
```

```
##      math4      read4
## 71.90900 60.06188
```

(v) Encontre a correlação entre math4 e read4. O que você conclui?

```
cor(meap01$math4, meap01$read4)
```

```
## [1] 0.8427281
```

As duas variáveis são positivamente e fortemente correlacionadas.

```
# gastos por aluno medio
mean(meap01$exppp)
```

(vi) A variável exppp são os gastos por aluno. Encontre o exppp médio e seu desvio padrão. Você diria que há uma variação ampla nos gastos por aluno?

```
## [1] 5194.865
```

```
# desvio-padrao
sd(meap01$exppp)
```

```
## [1] 1091.89
```

O desvio-padrão mostra que existe uma ampla variação no gasto médio por aluno.

(vii) Suponha que a Escola A gaste US\$ 6.000 por estudante e a Escola B gaste US\$ 5.500 por aluno. Com que percentual os gastos da Escola A superam os da Escola B? Compare isso a $100 \cdot [\log(6.000) - \log(5.500)]$, que é a diferença percentual aproximada baseada na diferença dos logs naturais. Ver Seção A.4, no Apêndice A (Disponível no site da Cengage.)

```
round(((log(6000) - log(5500)) * 100), 2)
```

```
## [1] 8.7
```

C4 Os dados contidos em JTRAIN2 são provenientes de um experimento de capacitação profissional direcionado para homens de baixa renda durante 1976-1977; ver Lalonde (1986).

```
data("jtrain2") # carregando base de dados do pacote do wooldridge
```

(i) Use a variável indicadora `train` para determinar a proporção de homens que recebeu treinamento profissional.

```
round((mean(jtrain2$train) * 100), 1)
```

```
## [1] 41.6
```

(ii) A variável `re78` são os ganhos de 1978, medidos em milhares de dólares de 1982. Encontre as médias de `re78` para a amostra de homens que recebeu capacitação profissional e para aquela que não recebeu. A diferença é economicamente grande?

```
# sem treinamento
(sem_treino <- mean(subset(jtrain2, train == 0)$re78))
```

```
## [1] 4.554802
```

```
# com treinamento
(com_treino <- mean(subset(jtrain2, train == 1)$re78))
```

```
## [1] 6.349145
```

A diferença percentual entre quem fez o treinamento e quem não fez é de 33.2%.

(iii) A variável `unem78` é um indicador de um homem estar desempregado ou não em 1978. Que proporção dos homens que receberam treinamento profissional está desempregada? E entre aqueles que não receberam treinamento? Comente sobre a diferença.

```
treino <- factor(jtrain2$train, labels = c("sem treino", "com treino"))
unem <- factor(jtrain2$unem78, labels = c("empregado", "desempregado"))

( tabela <- table(unem, treino, dnn = c("DESEMPREGO", "TREINAMENTO")) )
```

```
##                TREINAMENTO
## DESEMPREGO      sem treino com treino
## empregado         168      140
## desempregado       92       45
```

```
round(prop.table(tabela, margin = 2)[2,] * 100, 2)
```

```
## sem treino com treino
##      35.38      24.32
```