

Relatório

Douglas Cardoso

8/7/2021

Leitura dos dados

Antes, eu transformei os dados que você me deu em `csv` porque a manipulação fica mais fácil. Eu li desconsiderando as duas primeiras linhas (`skip = 2`); deu um bug aí eu desconsidere a primeira coluna (`col_select = -1`); coloquei o parâmetro que especifica qual o separador decimal dos dados, nesse caso a vírgula (`locale=...`); pedi para desconsiderar os nomes das colunas que veio junto com os dados, porque elas dificultam na manipulação dos dados depois, por possuírem espaços e acentos (`col_names = FALSE`); e depois especifiquei o tipo dos dados de cada coluna, se era `character` ou `double` (`double` é “numero com vírgula”, que não é inteiro). Por fim, eu atribui essas informações na variável `dados` (`dados <-`).

```
library(magrittr)
library(ggplot2)

(dados <- readr::read_csv('data/indicadores.csv', skip = 2, col_select = -1,
  locale = readr::locale(decimal_mark = ","),
  col_names = FALSE,
  col_types = list(
    X2 = readr::col_character(),
    X3 = readr::col_character(),
    X4 = readr::col_character(),
    X5 = readr::col_double(),
    X6 = readr::col_double(),
    X7 = readr::col_double()))

## # A tibble: 865 x 6
##   X2                                X3      X4      X5      X6      X7
##   <chr>                          <chr>   <chr>   <dbl> <dbl> <dbl>
## 1 Petróleo, Gás e Biocombustív~ 3R PETROLE~ Liquidez corre~ 4.76 3.72 -1.04
## 2 Comunicações                    ALGAR TELEC Liquidez corre~ 1.07 0.976 -0.0987
## 3 Saúde                          ALLIAR      Liquidez corre~ 1.04 1.04 0
## 4 Consumo Cíclico                ALPARGATAS Liquidez corre~ 2.06 2.06 0
## 5 Utilidade Pública              ALUPAR      Liquidez corre~ 3.26 3.26 0
## 6 Consumo não Cíclico            AMBEV S/A   Liquidez corre~ 1.00 1.00 0
## 7 Utilidade Pública              AMBIPAR     Liquidez corre~ 1.15 1.15 0
## 8 Utilidade Pública              AMPLA ENERG Liquidez corre~ 0.646 0.646 0
## 9 Consumo Cíclico               ANIMA       Liquidez corre~ 1.76 1.76 0
## 10 Consumo Cíclico              AREZZO CO   Liquidez corre~ 3.29 3.29 0
## # ... with 855 more rows
```

Renomeando as colunas

Aqui renomei as colunas para os nomes de forma que seja fácil manipular depois. Atribui isso a variável `da` por costume, não tem nenhum motivo especial. (`-> da`).

```
(dados %>%
  dplyr::rename(
    'setor_economico' = X2,
    'empresa' = X3,
    'descricao' = X4,
    'indice_pre_adocao' = X5,
    'indice_pos_adocao' = X6,
    'variacao' = X7) ->
da)
```

```
## # A tibble: 865 x 6
##   setor_economico empresa descricao indice_pre_adoc~ indice_pos_adoc~ variacao
##   <chr>          <chr>    <chr>          <dbl>          <dbl>      <dbl>
## 1 Petróleo, Gás e 3R PET~ Liquidez~      4.76          3.72      -1.04
## 2 Comunicações   ALGAR ~ Liquidez~      1.07          0.976    -0.0987
## 3 Saúde          ALLIAR Liquidez~      1.04          1.04       0
## 4 Consumo Cíclico ALPARG~ Liquidez~      2.06          2.06       0
## 5 Utilidade Públi ALUPAR Liquidez~      3.26          3.26       0
## 6 Consumo não Cíc AMBEV ~ Liquidez~      1.00          1.00       0
## 7 Utilidade Públi AMBIPAR Liquidez~      1.15          1.15       0
## 8 Utilidade Públi AMPLA ~ Liquidez~      0.646         0.646       0
## 9 Consumo Cíclico ANIMA  Liquidez~      1.76          1.76       0
## 10 Consumo Cíclico AREZZO~ Liquidez~      3.29          3.29       0
## # ... with 855 more rows
```

Estatísticas amostrais

Nesse tópico, eu imitei o artigo que você me mandou, fiz média, mediana e desvio padrão amostral, não sei se é preciso, mas fiz.

Média amostral

```
da %>%
  dplyr::group_by(descricao) %>%
  dplyr::rename('indicador' = descricao) %>%
  dplyr::summarise(
    media_pre = mean(indice_pre_adocao),
    media_pos = mean(indice_pos_adocao),
    variacao = media_pre - media_pos)
```

```
## # A tibble: 5 x 4
##   indicador          media_pre media_pos variacao
##   <chr>          <dbl>    <dbl>    <dbl>
## 1 Endividamento      1.69      1.89    -0.199
## 2 Imobilização de PL    1.12      1.36    -0.234
## 3 Liquidez corrente     1.77      1.73     0.0367
## 4 Liquidez Geral        1.08      0.926    0.153
## 5 Participação de Terceiros 2.99      3.37    -0.387
```

Mediana amostral

```
aaaaa aaaaa
aaaaaaa
```

aaaa

```
da %>%
  dplyr::group_by(descricao) %>%
  dplyr::rename('indicador' = descricao) %>%
  dplyr::summarise(
    mediana_pre = stats::median(indice_pre_adocao),
    mediana_pos = stats::median(indice_pos_adocao),
    variacao = mediana_pre - mediana_pos)

## # A tibble: 5 x 4
##   indicador          mediana_pre mediana_pos variacao
##   <chr>              <dbl>      <dbl>    <dbl>
## 1 Endividamento      1.35        1.62   -0.261
## 2 Imobilização de PL  1.04        1.19   -0.157
## 3 Liquidez corrente   1.37        1.35    0.0241
## 4 Liquidez Geral      0.865       0.718    0.146
## 5 Participação de Terceiros 1.16        1.39   -0.235
```

Desvio padrão amostral

```
da %>%
  dplyr::group_by(descricao) %>%
  dplyr::rename('indicador' = descricao) %>%
  dplyr::summarise(
    std_pre = sd(indice_pre_adocao),
    std_pos = sd(indice_pos_adocao),
    variacao = std_pre - std_pos)

## # A tibble: 5 x 4
##   indicador          std_pre std_pos variacao
##   <chr>              <dbl>  <dbl>    <dbl>
## 1 Endividamento      6.43    7.02   -0.595
## 2 Imobilização de PL  2.14    2.72   -0.576
## 3 Liquidez corrente   1.46    1.44    0.0238
## 4 Liquidez Geral      1.06    0.908   0.151
## 5 Participação de Terceiros 21.0   22.7   -1.73
```

Teste de Wilcoxon

O teste basicamente verifica se houve variação significativa nas medidas de centralidade entre dois “período”, nesse caso, é se houve variação significativa na mediana dos índices pré e pós adoção.

Aqui o código é complicado, mas não precisava, é um pouco da tal “programação funcional” do R. Vou explicar da melhor forma que consigo e por tópicos pra facilitar a didática. Cada tópico é uma linha do código.

- Eu agrupei os dados por índice (liquidez, tals, tals)
- Depois selecionei apenas as colunas que importa: a dos valores dos índices pré e pós adoção
- Eu aninhei as duas colunas pré e pós em uma só, apenas para facilitar uma manipulação depois
- Daqui pra baixo é o cálculo do `wilcox.test`, ele faz esse teste para cada índice, com o argumento `paired = TRUE` para especificar que é um *paired Wilcoxon test*.
- Aí depois eu só coloco na tela os resultados do teste, todos teram um valor `< p_value`, o que significa que não teve uma variação significativa.

```

da %>%
  dplyr::group_by(descricao) %>%
  dplyr::select(indice_pre_adocao, indice_pos_adocao) %>%
  dplyr::group_nest() %>%
  dplyr::mutate(
    wilcox = purrr::map(.x = data,
                        .f =
                          ~ stats::wilcox.test(.x$indice_pre_adocao,
                                                .x$indice_pos_adocao,
                                                paired = TRUE))) %$%

wilcox %>%
  purrr::set_names(unique(da$descricao))

```

```

## Adding missing grouping variables: `descricao`

## $`Liquidez corrente`
##
## Wilcoxon signed rank test with continuity correction
##
## data: .x$indice_pre_adocao and .x$indice_pos_adocao
## V = 1409, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##
## $`Liquidez Geral`
##
## Wilcoxon signed rank test with continuity correction
##
## data: .x$indice_pre_adocao and .x$indice_pos_adocao
## V = 1502, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##
## $Endividamento
##
## Wilcoxon signed rank test with continuity correction
##
## data: .x$indice_pre_adocao and .x$indice_pos_adocao
## V = 2080, p-value = 3.61e-12
## alternative hypothesis: true location shift is not equal to 0
##
##
## $`Participação de Terceiros`
##
## Wilcoxon signed rank test with continuity correction
##
## data: .x$indice_pre_adocao and .x$indice_pos_adocao
## V = 15051, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##
## $`Imobilização de PL`
##
## Wilcoxon signed rank test with continuity correction

```

```
##
## data: .x$indice_pre_adocao and .x$indice_pos_adocao
## V = 0, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

Gráfico extra

Aqui fiz uma tarefinha extra de forma gráfica que faz um boxplot por índice dos valores pré e pós, é uma forma visual de analisar se teve mudança significativa, infelizmente tem alguns valores outliers que literalmente *caga* todo o texto, mas ok, se você quiser e necessitar usar disso, eu posso ajudar a melhorá-lo.

```
da %>%
  dplyr::select(descricao, indice_pre_adocao, indice_pos_adocao) %>%
  tidyr::pivot_longer(c(indice_pre_adocao, indice_pos_adocao),
                      names_to = "periodo",
                      values_to = "indice") %>%
  ggplot(aes(x = periodo, y = indice, color = descricao)) +
  geom_boxplot() +
  theme_minimal() +
  labs(
    x = 'Período',
    y = 'Valor')
```

