

Desvantagem salarial dos empregados na hotelaria por região

Glauber Santos

2023-02-06

Este exercício busca analisar a diferença entre o salário médio na hotelaria e nas demais atividades produtivas. Em geral, a remuneração na hotelaria é mais baixa do que nas demais atividades da economia. Contudo, essa desvantagem salarial não é homogênea no espaço, sendo mais acentuada em alguns lugares ou regiões.

Para analisar a variação regional da desvantagem salarial dos hoteleiros por região, iremos utilizar dados sobre a remuneração média dos trabalhadores da hotelaria e das demais atividades produtivas por município do Brasil. Compararemos a desvantagem salarial dos hoteleiros nos municípios das regiões Sul e Sudeste do Brasil (mais ricas) com aquela observada nos municípios das demais regiões do país. A ideia é entender se a desvantagem salarial dos hoteleiros é ainda maior nas regiões mais pobres do país.

Preparação de dados

O salário médio na hotelaria e na economia como um todo será computado a partir de dados do relatório de [Estatísticas do Cadastro Central de Empresas (CEMPRE)] (<https://www.ibge.gov.br/estatisticas/economicas/comercio/9016-estatisticas-do-cadastro-central-de-empresas.html>) do IBGE. Os dados do número de empregados e a massa salarial de cada município brasileiro por atividade econômica podem ser baixados através da Tabela 6450 do banco de dados SIDRA do IBGE.

Vamos baixar os dados do Sidra do IBGE. Na Tabela 6450, selecione as variáveis **Pessoal ocupado assalariado** e **Salários e outras remunerações**, a atividade econômica **55.10-8 Hotéis e similares**, bem como o **Total** das atividades econômicas, o ano de **2019** e **todos os municípios** do país em unidade territorial. Na definição das colunas, coloque **Ano**, **Variável** e **Classificação Nacional de Atividades**, e nas linhas coloque **Unidade Territorial**. O ano de 2019 foi selecionado por ser o mais recente representativo de um momento de não pandemia. Exporte a tabela do Sidra para o formato .csv (BR), salve na pasta do seu projeto e importe.

```
empregos <- read.csv("tabela6450.csv", skip=4, sep = ";")
empregos <- empregos[1:5570,]
```

```
head(empregos)
```

| ## | Cód. | Município | Total | X55.10.8.Hotéis.e.similares | Total.1 |
|------|-------------------------------|----------------------------|-------|-----------------------------|---------|
| ## 1 | 1100015 | Alta Floresta D'Oeste (RO) | 2443 | 8 | 58805 |
| ## 2 | 1100023 | Ariquemes (RO) | 16718 | X | 440332 |
| ## 3 | 1100031 | Cabixi (RO) | 536 | X | 13352 |
| ## 4 | 1100049 | Cacoal (RO) | 17058 | 222 | 436894 |
| ## 5 | 1100056 | Cerejeiras (RO) | 2022 | 25 | 50928 |
| ## 6 | 1100064 | Colorado do Oeste (RO) | 2110 | X | 47580 |
| ## | X55.10.8.Hotéis.e.similares.1 | | | | |
| ## 1 | | 101 | | | |
| ## 2 | | X | | | |
| ## 3 | | X | | | |

```
## 4          3303
## 5          333
## 6          X
```

```
names(empregos) <- c(
  "Município (Código)",
  "Município",
  "Total_Pessoal ocupado assalariado",
  "55.10-8 Hotéis e similares_Pessoal ocupado assalariado",
  "Total_Salários e outras remunerações",
  "55.10-8 Hotéis e similares_Salários e outras remunerações")

empregos[empregos=="-"] <- NA
empregos[empregos=="X"] <- NA
empregos <- type.convert(empregos)
```

Em seguida, vamos calcular o salário médio na hotelaria e na economia como um todo para cada município do país. Nos municípios em que não há registro de empregados no setor hoteleiro, o número de empregados e a massa salarial constam como NA (não disponível). O resultado do cálculo do salário médio nesses casos também será NA. Após o cálculo dos salários médios, vamos calcular a diferença salarial percentual.

```
empregos$salmed_hotel <-
  empregos$`55.10-8 Hotéis e similares_Salários e outras remunerações` /
  empregos$`55.10-8 Hotéis e similares_Pessoal ocupado assalariado`

empregos$salmed_geral <-
  empregos$`Total_Salários e outras remunerações` /
  empregos$`Total_Pessoal ocupado assalariado`

empregos$diferenca_salarial <- empregos$salmed_hotel/empregos$salmed_geral-1

head(empregos)
```

```
## Município (Código)      Município
## 1      1100015 Alta Floresta D'Oeste (RO)
## 2      1100023      Ariquemes (RO)
## 3      1100031      Cabixi (RO)
## 4      1100049      Cacoal (RO)
## 5      1100056      Cerejeiras (RO)
## 6      1100064      Colorado do Oeste (RO)
## Total_Pessoal ocupado assalariado
## 1                      2443
## 2                      16718
## 3                      536
## 4                      17058
## 5                      2022
## 6                      2110
## 55.10-8 Hotéis e similares_Pessoal ocupado assalariado
## 1                      8
## 2                      NA
## 3                      NA
## 4                      222
## 5                      25
```

```
## 6 NA
## Total_Salários e outras remunerações
## 1 58805
## 2 440332
## 3 13352
## 4 436894
## 5 50928
## 6 47580
## 55.10-8 Hotéis e similares_Salários e outras remunerações salmed_hotel
## 1 101 12.62500
## 2 NA NA
## 3 NA NA
## 4 3303 14.87838
## 5 333 13.32000
## 6 NA NA
## salmed_geral diferenca_salarial
## 1 24.07081 -0.4755059
## 2 26.33880 NA
## 3 24.91045 NA
## 4 25.61226 -0.4190916
## 5 25.18694 -0.4711546
## 6 22.54976 NA
```

Podemos ver o resultado na última coluna da tabela. Note que os valores estão registrados em decimais. No município 1100015, por exemplo, os empregados na hotelaria ganham, em média, 47,55% a menos do que os empregados da economia como um todo.

Por fim, vamos criar uma variável para identificar os municípios das regiões Sul e Sudeste do país. A região do país pode ser identificada pelo primeiro dígito do código do município. Municípios da região Norte começam com o dígito 1, Nordeste 2, Sudeste 3, Sul 4 e Centro Oeste 5. Sendo assim, vamos primeiro criar uma variável identificando a região e depois diferenciar Sul e Sudeste das demais regiões. Para tanto, será preciso considerar o código do município como variável numérica, embora originalmente ela tenha sido importada como texto.

```
empregos$regiao <- floor(as.numeric(empregos$`Município (Código)`)/1000000)

empregos$sulsudeste <- ifelse((empregos$regiao==3 | empregos$regiao==4), 1, 0)

head(empregos)
```

```
## Município (Código) Município
## 1 1100015 Alta Floresta D'Oeste (RO)
## 2 1100023 Ariquemes (RO)
## 3 1100031 Cabixi (RO)
## 4 1100049 Cacoal (RO)
## 5 1100056 Cerejeiras (RO)
## 6 1100064 Colorado do Oeste (RO)
## Total_Pessoal ocupado assalariado
## 1 2443
## 2 16718
## 3 536
## 4 17058
## 5 2022
## 6 2110
```

```
## 55.10-8 Hotéis e similares_Pessoal ocupado assalariado
## 1 8
## 2 NA
## 3 NA
## 4 222
## 5 25
## 6 NA
## Total_Salários e outras remunerações
## 1 58805
## 2 440332
## 3 13352
## 4 436894
## 5 50928
## 6 47580
## 55.10-8 Hotéis e similares_Salários e outras remunerações salmed_hotel
## 1 101 12.62500
## 2 NA NA
## 3 NA NA
## 4 3303 14.87838
## 5 333 13.32000
## 6 NA NA
## salmed_geral diferenca_salarial regioao sulsudeste
## 1 24.07081 -0.4755059 1 0
## 2 26.33880 NA 1 0
## 3 24.91045 NA 1 0
## 4 25.61226 -0.4190916 1 0
## 5 25.18694 -0.4711546 1 0
## 6 22.54976 NA 1 0
```

Análise de dados

Vamos começar a exploração dos dados com medidas descritivas da diferença salarial da hotelaria por região. Para calcular estatísticas, vamos utilizar o pacote `rstatix`.

```
library(rstatix)

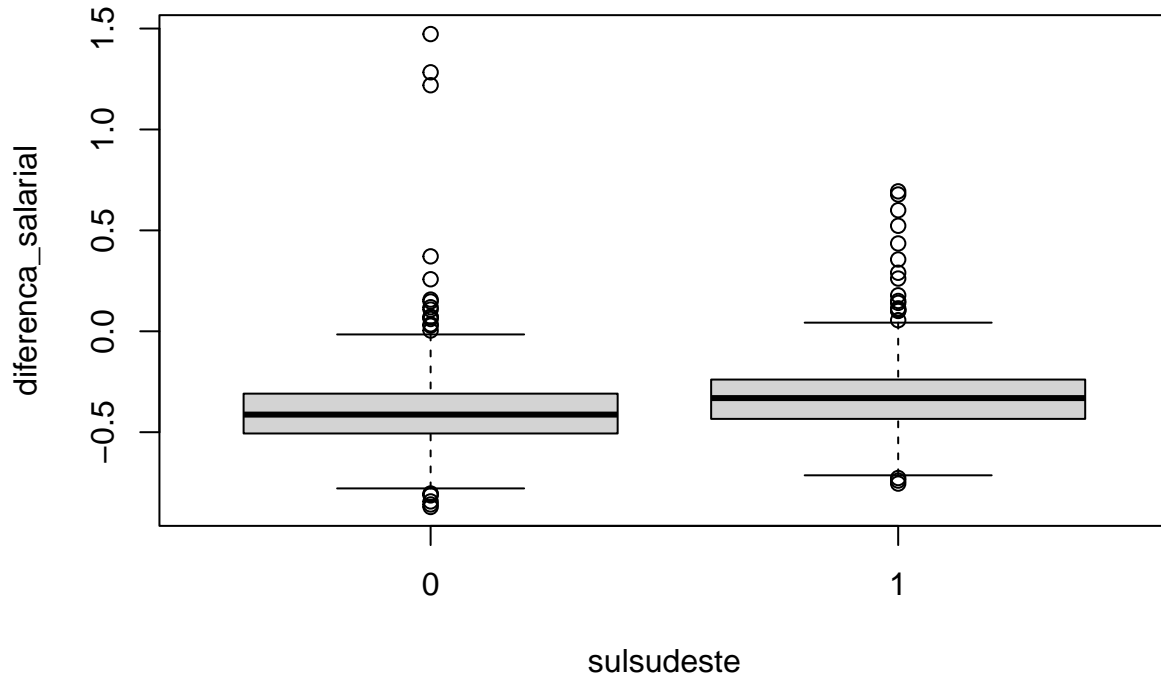
empregos %>%
  group_by(sulsudeste) %>%
  get_summary_stats(diferenca_salarial, type = "mean_sd")
```

```
## # A tibble: 2 x 5
##   sulsudeste variable      n   mean    sd
##   <dbl> <fct>         <dbl> <dbl> <dbl>
## 1      0 diferenca_salarial  548 -0.389 0.212
## 2      1 diferenca_salarial  746 -0.324 0.174
```

Os cálculos mostram que os empregados na hotelaria ganham cerca de 32,4% a menos do que a média dos trabalhadores na região Sul/Sudeste. No restante do país, a desvantagem salarial na hotelaria chega a 38,9%. A heterogeneidade da diferença salarial entre municípios é menor na região Sul/Sudeste, pois o desvio padrão (sd) é menor nessa parte do país.

Os dados da diferença salarial da hotelaria por região podem ser visualizados no gráfico boxplot apresentado a seguir.

```
boxplot(diferenca_salarial ~ sulsudeste, data = empregos)
```



Os dados observados revelam uma realidade concreta, permitindo-nos elaborar um enunciado particular. Contudo, na ciência, geralmente estamos interessados em enunciados gerais. A partir da análise concreta, não podemos saber se a desvantagem salarial é menor nas regiões Sul e Sudeste em razão de uma simples distribuição aleatória de valores entre municípios, ou se efetivamente podemos atribuir essa diferença à região. Para afirmar que a diferença é devida à região, temos que comparar a diferença de médias com aquela que seria esperada se as realidades municipais fossem aleatoriamente distribuídas no país. Em outras palavras, vamos testar a hipótese:

H1: A desvantagem salarial dos hoteleiros é maior nas regiões mais pobres do que nas regiões mais ricas.

A hipótese nula referente a H1 pode ser redigida como:

H0: A desvantagem salarial dos hoteleiros é a mesma nas regiões mais pobres e nas regiões mais ricas.

Para tanto, vamos começar testando a normalidade da distribuição de municípios segundo o valor da desvantagem salarial de hoteleiros.

```
empregos %>%
  group_by(sulsudeste) %>%
  shapiro_test(diferenca_salarial)
```

```
## # A tibble: 2 x 4
##   sulsudeste variable      statistic      p
##   <dbl> <chr>          <dbl>    <dbl>
## 1       0 diferenca_salarial  0.802 2.45e-25
## 2       1 diferenca_salarial  0.936 2.29e-17
```

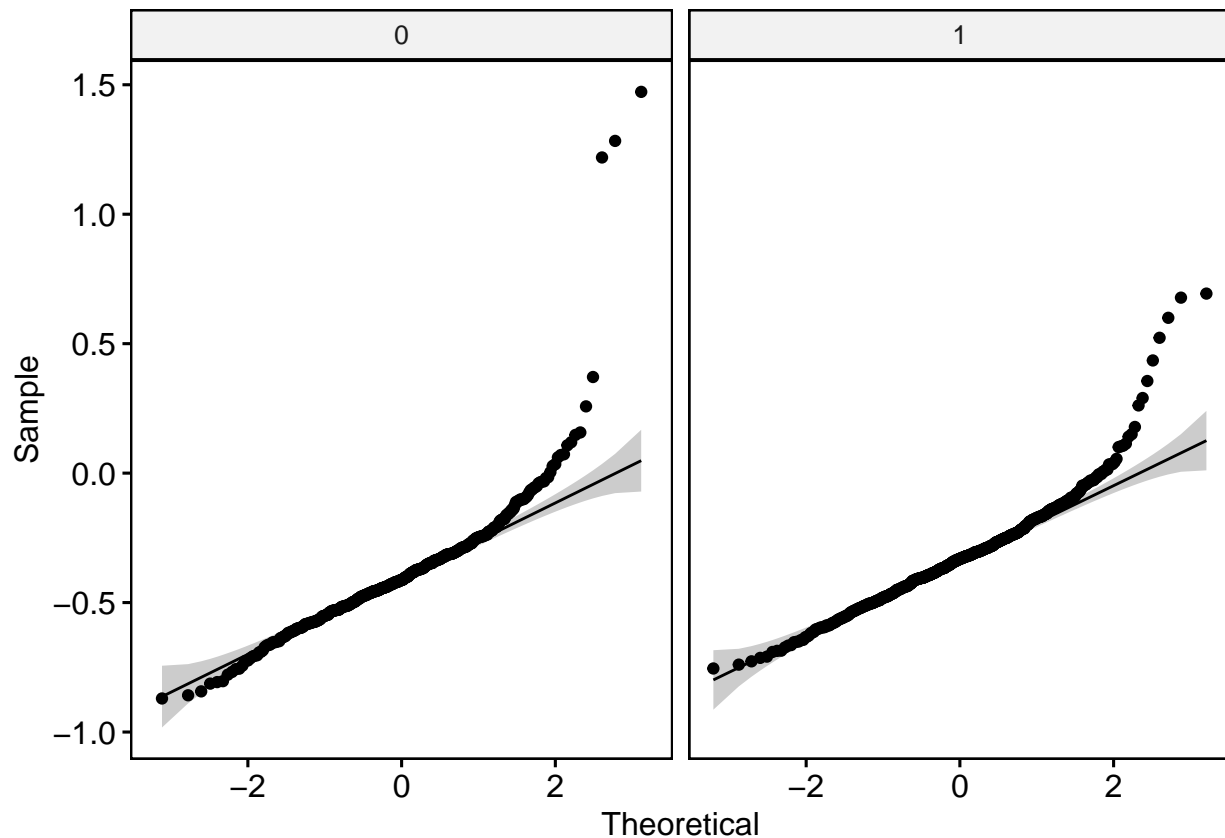
Os resultados indicam que as distribuições das duas partes do país se afastam da distribuição Normal, confirmando a impressão de assimetria positiva sugerida pelo boxplot. Uma alternativa a ser examinada neste caso é a consideração do logaritmo da diferença salarial.

```
empregos$log_diferenca_salarial <- log(empregos$diferenca_salarial)
empregos %>%
  group_by(sulsudeste) %>%
  shapiro_test(log_diferenca_salarial)
```

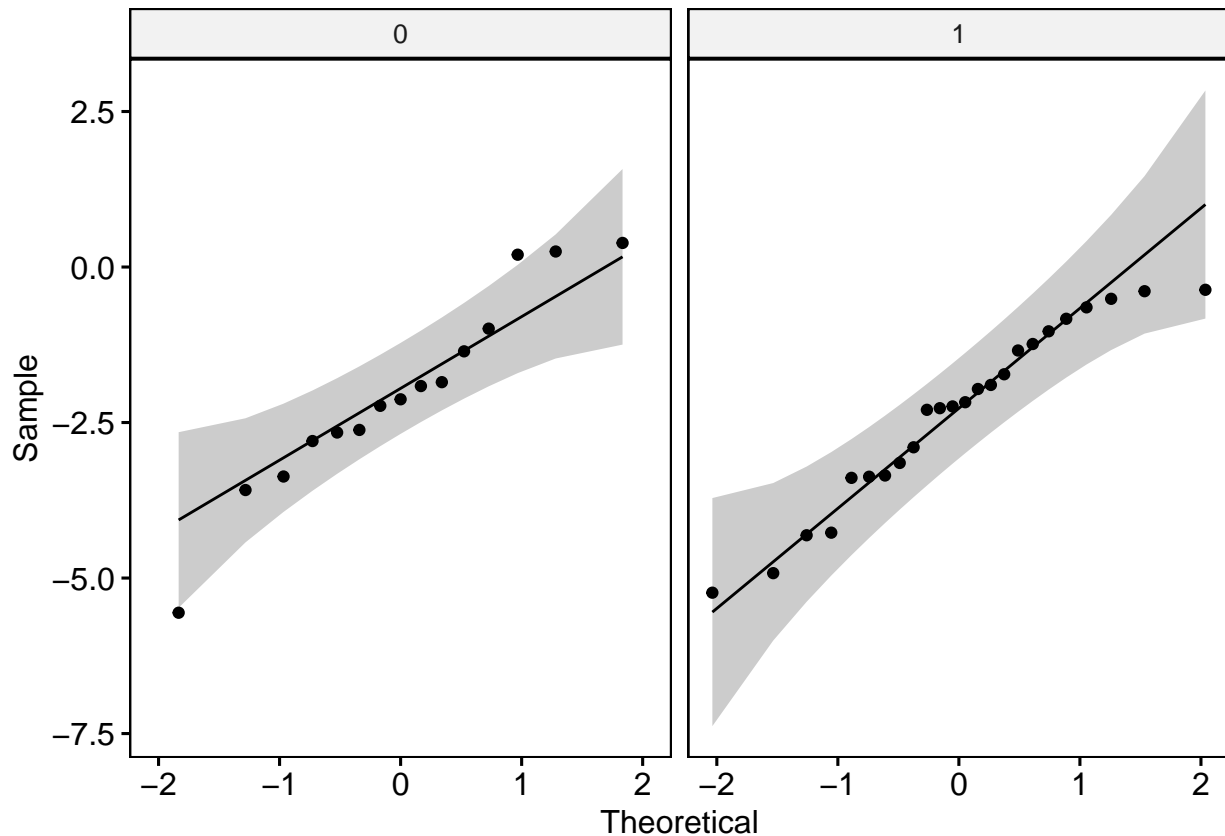
```
## # A tibble: 2 x 4
##   sulsudeste variable      statistic      p
##   <dbl> <chr>          <dbl> <dbl>
## 1     0 log_diferenca_salarial    0.950 0.520
## 2     1 log_diferenca_salarial    0.948 0.248
```

Refazendo o teste de normalidade Shapiro-Wilk para o logaritmo da diferença salarial, chega-se à conclusão de que pode-se assumir que os municípios seguem uma distribuição Normal. As distribuições com e sem transformação logarítmica são apresentadas a seguir. Para elaborar os gráficos, vamos utilizar o pacote `ggpubr`.

```
library(ggpubr)
ggqqplot(empregos, x = "diferenca_salarial", facet.by = "sulsudeste")
```



```
ggqqplot(empregos, x = "log_diferenca_salarial", facet.by = "sulsudeste")
```



Além da normalidade, a escolha do teste de médias requer a análise da igualdade de variâncias da desvantagem salarial para os dois grupos em questão. A igualdade das variâncias para as duas partes do país é analisada com auxílio do teste de Levene.

```
empregos$sulsudeste <- as.character(empregos$sulsudeste)
empregos %>% levene_test(log_diferenca_salarial ~ sulsudeste)
```

```
## # A tibble: 1 x 4
##   df1  df2 statistic    p
##   <int> <int>     <dbl> <dbl>
## 1     1    37   0.00426 0.948
```

O resultado do teste de igualdade de variâncias indica que a hipótese de igualdade não deve ser descartada. Sendo assim, para a comparação de médias, utiliza-se o teste t de Student, que assume variâncias iguais entre os dois grupos.

```
empregos %>%
  t_test(log_diferenca_salarial ~ sulsudeste, var.equal = TRUE) %>%
  add_significance()
```

```
## # A tibble: 1 x 9
##   .y.      group1 group2   n1   n2 statistic    df    p p.sig-1
```

```
##   <chr>                <chr> <chr> <int> <int>      <dbl> <dbl> <dbl> <chr>
## 1 log_diferenca_salarial 0      1      15    24      0.630    37 0.533 ns
## # ... with abbreviated variable name 1: p.signif
```

Os resultados apontam que a diferença entre regiões no logaritmo da diferença salarial não é significativa, sugerindo a rejeição de H1 e a aceitação de H0. Logo, conclui-se que não há evidências para descartar a hipótese de que a desvantagem salarial dos empregados do turismo é a mesma nas regiões Sul e Sudeste e no restante do país.

Por fim, apenas a título de exemplo, já que o teste t revelou a igualdade de médias, estima-se a dimensão da diferença encontrada entre as duas partes do país. Utilizando-se a estatística d de Cohen, conclui-se que o tamanho do efeito é 0,207, valor que é tradicionalmente classificado como pequeno.

```
empregos %>% cohens_d(log_diferenca_salarial ~ sulsudeste, var.equal = TRUE)
```

```
## # A tibble: 1 x 7
##   .y.      group1 group2 effsize    n1    n2 magnitude
## * <chr>    <chr> <chr>    <dbl> <int> <int> <ord>
## 1 log_diferenca_salarial 0      1      0.207    15    24 small
```

Referência <https://www.datanovia.com/en/lessons/t-test-in-r/>