Trabalho 2 - SME0806 - Estatística Computacional

Universidade de São Paulo

Diego G. de Paulo (10857040) Bruno H.da S Justino (11031621)
Douglas S. Souza (10733820) Caio H. M. Schiavo (11810602)
Vitor Gratiere Torres (10284952)

18/06/2020

Contents

Introdução

Este trabalho tem como principal objetivo mostrar estimativas pontuais e intervalares para o coeficiente de Gini que é um dos principais índices de desigualdade utilizados, como também para as variáveis associadas. O índice de Gini é uma medida de desigualdade comumente utilizada para calcular a desigualdade de distribuição de renda, mas pode ser usada também para qualquer distribuição, como concentração de terra e riqueza. Desenvolvida pelo estatístico italiano Corrado Gini e publicada no documento "Variabilità e Mutabilità" em 1912, a medida consiste em um número entre 0 e 1, onde 0 corresponde à completa igualdade de renda e 1 corresponde à completa desigualdade. Para a estimação, aplicaremos o método de bootstrap, baseado em um grande número de reamostras. Os métodos de reamostragem permitem quantificar incertezas calculando erros padrões e intervalos de confiança, bem como realizar testes de significância. Eles requerem menos suposições e geralmente fornecem respostas mais precisas do que os métodos tradicionais (MOORE,McCABE,DUCKWORTH,SCLOVE,1996).

Exercício 1

Motivação

Neste exercício a cargo dos alunos que realizam este trabalho, realizar estimativas pontuais e intervalares do coeficiente Gini, um indicador de desigualdade em relação ao PIB per Capita. Este coeficiente é definido por:

$$G = \frac{\sum_{i=1}^{n}\sum_{i=1}^{n}|x_i - x_j|}{2n^2\mu}$$

Metodologia

Para solução, será utilizado método bootstrap para amostrar valores $x_1^*, ..., x_n^*$ provenientes, com reposição dos valores observados da variável Pib per Capita. Após a obtenção desses valores será aplicada a função descrita acima para o Coeficiente de Gini a fim de obter uma estimação pontual e um intervalo de confiança bootstrap (neste caso de 95% de confiança).

Resolução

Para obter a estimação pontual será utilizado o seguinte resultado (baseado numa aproximação em simulações de Monte Carlo com B amostras bootstrap):

$$\frac{1}{B}\sum_{b=1}^{B} g(x_{b,1}^*, ..., x_{n,1}^*)$$

Já na obtenção das estimativas intervalares serão selecionados os quantis 2,5 e 97,5% do vetor de resultados gerados para o coeficiente.

```r
gini <- function(x, y) {
  z <- abs(x - y)
}

ex_1 <- function(df, r) {

  g <- c()
  n <- nrow(df)

  for (i in 1:r) {
    am <- sample(as.matrix(df[, 9]), n, replace = T)
    matriz <- outer(am, am, gini)
    mu <- mean(am)
    g[i] <- sum(matriz) / (2 * (n ^ 2) * mu)
  }

  cat("Estimativa pontual para o Coeficiente Gini = ", round(mean(g), 4),
      "com", r, "repetições", "\n")

  cat("Intervalo de 95% para o Coeficiente Gini:", "[",
      round(quantile(g, .025), 4), ";", round(quantile(g, .975), 4),
      "]", "com", r, "\nrepetições", "\n")
}

ex_1(df = df_fim, r = 500)
```

```
## Estimativa pontual para o Coeficiente Gini =  0,3254 com 500 repetições
## Intervalo de 95% para o Coeficiente Gini: [ 0,2967 ; 0,3566 ] com 500
## repetições
```

```r
ex_1(df = df_fim, r = 1000)
```

```
## Estimativa pontual para o Coeficiente Gini =  0,3259 com 1000 repetições
## Intervalo de 95% para o Coeficiente Gini: [ 0,2948 ; 0,3578 ] com 1000
## repetições
```

```r
ex_1(df = df_fim, r = 3000)
```

```
## Estimativa pontual para o Coeficiente Gini =  0,3257 com 3000 repetições
## Intervalo de 95% para o Coeficiente Gini: [ 0,2942 ; 0,3584 ] com 3000
## repetições
```

Como foi notado, o coeficiente teve um resultado apresentado próximo a 0,3, apresentando um índice de desigualdade relativamente baixo entre as cidades do estado em relação ao PIB per Capita.

Exercício 2

Motivação

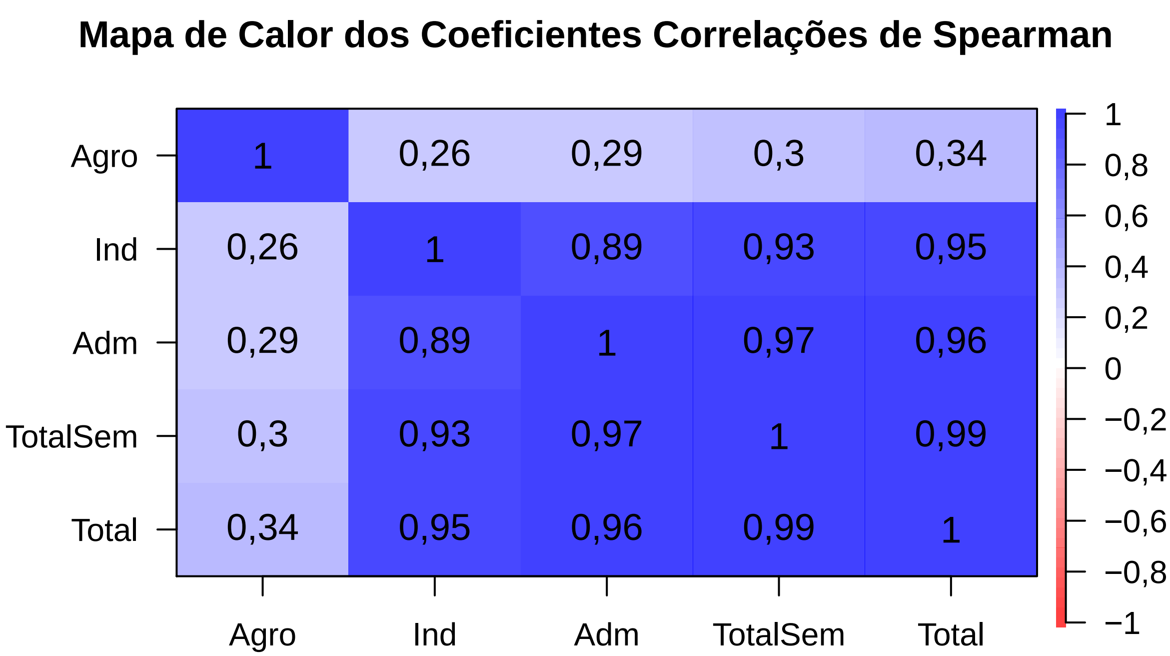
Para o exercício os resultados procurados são, também, uma estimação pontual e um intervalo de confiança para a associação entre as duas variáveis, que se enquadram em valor agregado, mais associadas entre si.

Metodologia

Para a avaliação desta associação foi selecionado o Coeficiente de Correlação de Spearman, apresentado em forma de mapa de calor conforme segue.

Resolução

```
corr_mat <- cor(df_fim[, c(2:6)], method = "s")
corPlot(corr_mat, cex = 1.2)
title("Mapa de Calor dos Coeficientes Correlações de Spearman")
```



Pelo o resultado obtido, as variáveis selecionadas foram Total (exclusive Administração Pública) e

Total Geral. Para a obtenção dos resultados de interesse deste exercício, um método semelhante ao método do exercício anterior foi utilizado com a diferença de que neste caso as amostras das váriaves selecionadas foram amostradas de forma pareada para a obtenção da medida de associação (vale lembrar que o Coeficiente de Correlação de Spearman leva em conta os ranks das observações para a obtenção do resultado).

```
ex_2 <- function(df, r) {

  g <- c()
  n <- nrow(df)

  for (i in 1:r) {
    index <- sample(c(1:n), n, replace = T)
    am1 <- df$TotalSem[index]
    am2 <- df$Total[index]
    g[i] <- cor(x = am1, y = am2, method = "s")
  }

  cat("Estimativa pontual para o Coeficiente de Correlação de Spearman = ",
      round(mean(g), 4), "com\n", r, "repetições", "\n")

  cat("Intervalo de 95% para o Coeficiente de Correlação de Spearman:", "\n[",
      round(quantile(g, .025), 4), ";", round(quantile(g, .975), 4),
      "]", "com", r, "repetições", "\n")
}

ex_2(df = df_fim, r = 500)
```

```
## Estimativa pontual para o Coeficiente de Correlação de Spearman =  0,9873 com
##  500 repetições
## Intervalo de 95% para o Coeficiente de Correlação de Spearman:
## [ 0,9837 ; 0,9902 ] com 500 repetições
```

```
ex_2(df = df_fim, r = 1000)
```

```
## Estimativa pontual para o Coeficiente de Correlação de Spearman =  0,9873 com
##  1000 repetições
## Intervalo de 95% para o Coeficiente de Correlação de Spearman:
## [ 0,9836 ; 0,9901 ] com 1000 repetições
```

```
ex_2(df = df_fim, r = 3000)
```

```
## Estimativa pontual para o Coeficiente de Correlação de Spearman =  0,9873 com
##  3000 repetições
## Intervalo de 95% para o Coeficiente de Correlação de Spearman:
## [ 0,9837 ; 0,9901 ] com 3000 repetições
```

Como é observado, de acordo com os resultados obtidos, a correlação está bem próxima de 1, que vai bem ao encontro do resultado visualizado no mapa de calor, ou seja, já era um resultado esperado.

Conclusão

Após a realização deste trabalho, o grupo de alunos responsável pode fixar o conteúdo de reamostragem com grande enfoque no método boostrap, principalmente pela utilização de dados reais obtidos a partir de um banco de dados obtidos disponibilizado pelo governo.

Referências

Notas de Aula de Estatística Computacional - Professor Dr. Mário de Castro

LETTIERI, M.; PAES, N. L. Medidas de Pobreza e Desigualdade: Uma Analise Teórica dos Principais Índices. Fortaleza, 2006. (Laboratório de Estudos da Pobreza (LEP)

Site: mathworld.wolfram.com - GiniCoefficient

Site: www.ipece.ce.gov.br - Entendendo Indice De Gini

Site: www.seade.gov.br - Dados PIB(2018)