



UNIVERSIDADE D  
**COIMBRA**

Faculdade de Ciências e Tecnologia da Universidade de Coimbra  
Departamento de Matemática  
Ano Letivo 2022/2023

# Estimador do quociente e estimador da regressão

Trabalho realizado por  
Ana Carina Carvalho  
Margarida Biscaia

Amostragem e Sondagens

# Conteúdo

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introdução</b>  | <b>2</b>  |
| <b>2</b> | <b>Conceitos Teóricos</b>                                    | <b>2</b>  |
| 2.1      | Amostragem: conceitos básicos . . . . .                      | 2         |
| 2.1.1    | Comparação de sistemas de amostragem . . . . .               | 2         |
| 2.1.2    | Intervalos de confiança . . . . .                            | 2         |
| 2.2      | Amostragem simples sem reposição . . . . .                   | 3         |
| 2.3      | Estimação com informação auxiliar . . . . .                  | 3         |
| 2.3.1    | Estimador do quociente . . . . .                             | 3         |
| 2.3.2    | Estimador da regressão . . . . .                             | 4         |
| <b>3</b> | <b>Simulação das populações e comparação dos estimadores</b> | <b>6</b>  |
| 3.1      | População da Bélgica . . . . .                               | 6         |
| 3.2      | População da Suíça . . . . .                                 | 9         |
| <b>4</b> | <b>Conclusão</b>   | <b>12</b> |
|          | <b>Referências</b>   | <b>12</b> |

# 1 Introdução

Para o plano SSR e para duas populações convenientemente escolhidas, pretende-se comparar pelo método de Monte Carlo o estimador do quociente,  $\hat{y}_q$ , e o estimador da regressão,  $\hat{y}_r$ . Para tal, foram utilizadas duas populações incluídas nas bases de dados **data(belgianmunicipalities)** e **data(swissmunicipalities)** no *package* “sampling” do R. Além disso, pretende-se comparar estes dois estimadores com a média amostral,  $\hat{\bar{y}}$ , e analisar a probabilidade de cobertura e margem de erro dos intervalos de confiança produzidos.

## 2 Conceitos Teóricos

### 2.1 Amostragem: conceitos básicos

Seja  $\mathcal{U}$  uma população finita. Admitiremos que cada uma das  $N$  unidades da população é identificável através de um número de ordem, podendo-se assim identificar  $\mathcal{U}$  com o conjunto dos  $N$  primeiros números naturais

$$\mathcal{U} = \{1, \dots, N\} .$$

A  $N$  chamamos tamanho da população. Representando por  $y$  a variável ou característica de interesse sobre a qual pretendemos obter informação, e por  $y_k$  o valor dessa variável para a  $k$ -ésima unidade da população, o vetor de  $\mathbb{R}^N$

$$\mathbf{y} = (y_1, \dots, y_N) = (y_k, k \in \mathcal{U})$$

contém o valor da característica  $y$  para todas as unidades da população.

#### 2.1.1 Comparação de sistemas de amostragem

Seja  $\theta$  o parâmetro de interesse e  $\hat{\theta}$  o seu estimador. A qualidade de um sistema de amostragem é normalmente avaliada através do erro quadrático médio definido por

$$\text{EQM}(\hat{\theta}) = E(\hat{\theta} - \theta) = \text{Var}(\hat{\theta}) + (\text{Viés}(\hat{\theta}))^2 .$$

**Definição 2.1.1** Dizemos que o sistema de amostragem  $(p_1, \hat{\theta}_1(\mathbf{y}))$  é pelo menos tão eficiente como  $(p_2, \hat{\theta}_2(\mathbf{y}))$  em  $\mathcal{Y} \subseteq \mathbb{R}^N$ , se

$$\text{EQM}(\hat{\theta}_1(\mathbf{y})) \leq \text{EQM}(\hat{\theta}_2(\mathbf{y})) , \text{ para todo o } \mathbf{y} \in \mathcal{Y} .$$

#### 2.1.2 Intervalos de confiança

Sendo  $\hat{\theta}$  um estimador centrado de  $\theta$ , um intervalo de confiança para  $\theta$  com nível de confiança aproximadamente igual a  $100(1 - \alpha)\%$  tem por extremidades

$$\hat{\theta} \pm z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})}$$

onde  $z_{1-\alpha/2}$  é o quantil de ordem  $1 - \alpha/2$  da distribuição normal standard.

Este intervalo de confiança para  $\theta$  pode ser interpretado da forma usual: se tomarmos sucessivas amostras segundo o plano de amostragem em causa e construirmos os respetivos intervalos com nível de confiança  $100(1 - \alpha)\%$ , aproximadamente  $100(1 - \alpha)\%$  deles conterão  $\theta$ . À proporção de vezes que tal ocorre designamos por **probabilidade de cobertura**. Quanto maior a probabilidade de cobertura do estimador mais eficiente ele será.

## 2.2 Amostragem simples sem reposição

Seja  $S = (S_1, S_2, \dots, S_n)$  uma amostra aleatória com plano de amostragem  $p$ .

**Teorema 2.2.1** *Num plano de amostragem SSR de tamanho  $n$  a **média empírica***

$$\hat{y} = \frac{1}{n} \sum_{k \in S} y_k$$

*é um estimador centrado de  $\bar{y}$  com variância*

$$\text{Var}(\hat{y}) = \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n}.$$

## 2.3 Estimação com informação auxiliar

Assumiremos que a informação auxiliar sobre a população está disponível do conhecimento duma variável auxiliar  $x$  associada à variável de interesse  $y$ .

### 2.3.1 Estimador do quociente

Tomando como motivação para a definição deste estimador uma possível relação de proporcionalidade entre a variável de interesse  $y$  e a variável auxiliar  $x$ , da qual supomos conhecer o respetivo total  $t_x$ , suponhamos então que para algum  $\lambda \in \mathbb{R}$  se tem

$$y_k \approx \lambda x_k, \quad k \in \mathcal{U}.$$

Neste caso,

$$t_y \approx \lambda t_x$$

e para toda a amostra aleatória  $S$  temos

$$\hat{t}_y = \frac{N}{n} \sum_{k \in S} y_k \approx \lambda \frac{N}{n} \sum_{k \in S} x_k = \lambda \hat{t}_x.$$

Assim,

$$t_y \approx \frac{t_x}{\hat{t}_x} \hat{t}_y = \frac{\hat{t}_y}{\hat{t}_x} t_x,$$

o que motiva estimar  $t_y$  através do estimador do quociente definido por

$$\hat{t}_q = \hat{r}t_x,$$

com

$$\hat{r} = \frac{\hat{t}_y}{\hat{t}_x}.$$

O estimador  $\hat{t}_q$  é assim um estimador de tipo rácio que sabemos não ser cêntrico.

É possível obter uma aproximação para a variância dada por

$$\text{Var}(\hat{t}_q) \approx N^2 \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} \left(1 - \frac{2rs_{yx} - r^2s_x^2}{s_y^2}\right),$$

com

$$r = \frac{t_y}{t_x}.$$

De forma semelhante, podemos estimar  $\bar{y}$  através do **estimador do quociente** definido por

$$\hat{\bar{y}}_q = \frac{\hat{\bar{y}}}{\hat{\bar{x}}} \cdot \bar{x},$$

com variância dada por

$$\text{Var}(\hat{\bar{y}}_q) \approx \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} \left(1 - \frac{2rs_{yx} - r^2s_x^2}{s_y^2}\right).$$

### 2.3.2 Estimador da regressão

Tomando como motivação para a definição deste estimador uma possível relação linear entre a variável de interesse  $y$  e a variável auxiliar  $x$ , da qual supomos conhecer o respetivo total  $t_x$ , suponhamos então que

$$y_k \approx ax_k + b, \quad k \in \mathcal{U},$$

com  $a, b \in \mathbb{R}$ . Neste caso,

$$t_y \approx at_x + Nb,$$

e para toda a amostra aleatória  $S$  temos,

$$\hat{t}_y = \frac{N}{n} \sum_{k \in S} y_k \approx a \frac{N}{n} \sum_{k \in S} x_k + Nb = \hat{at}_x + Nb$$

e

$$\hat{s}_{yx} = \frac{1}{n-1} \sum_{k \in S} (y_k - \bar{y})(x_k - \bar{x}) \approx a \frac{1}{n-1} \sum_{k \in S} (x_k - \bar{x})^2 = a\hat{s}_x^2.$$

Assim,

$$t_y \approx \hat{at}_x + (\hat{t}_y - \hat{at}_x) = \hat{t}_y - \hat{a}(\hat{t}_x - t_x),$$

com

$$\hat{a} = \frac{\hat{s}_{yx}}{\hat{s}_x^2},$$

o que motiva estimar  $t_y$  através do estimador da regressão definido por

$$\hat{t}_r = \hat{t}_y - \hat{a}(\hat{t}_x - t_x).$$

Uma vez que  $\hat{a} = \hat{s}_{yx}/\hat{s}_x^2$  é um estimador de tipo rácio, o estimador da regressão é um estimador enviesado de  $t_y$ .

É possível obter uma aproximação para a variância dada por

$$\text{Var}(\hat{t}_r) \approx N^2 \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} (1 - \rho_{xy}^2)$$

com

$$\rho_{xy} = \frac{s_{xy}}{s_y s_x}.$$

De forma semelhante, podemos estimar  $\bar{y}$  através do **estimador da regressão** definido por

$$\hat{y}_r = \hat{y} - \hat{a}(\hat{x} - \bar{x}),$$

com variância dada por

$$\text{Var}(\hat{y}_r) \approx \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} (1 - \rho_{xy}^2).$$

**Observação.** Atendendo a que

$$\begin{aligned} \text{Var}(\hat{y}) &= \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n}, \\ \text{Var}(\hat{y}_q) &\approx \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} \left(1 - \frac{2rs_{yx} - r^2 s_x^2}{s_y^2}\right), \end{aligned}$$

e

$$\text{Var}(\hat{y}_r) \approx \left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} (1 - \rho_{xy}^2),$$

esperamos que o estimador da regressão  $\hat{y}_r$  possa ser mais eficiente que os estimadores  $\hat{y}$  e  $\hat{y}_q$ , quando o tamanho da amostra é grande (caso em que o viés de  $\hat{y}_r$  pode ser desprezado). No entanto, é preciso não esquecer que os estimadores  $\hat{y}_q$  e  $\hat{y}_r$  são enviesados e que as fórmulas anteriores para as suas variâncias resultam da utilização de diversas aproximações que podem ter menor qualidade quando o tamanho da amostra não é grande.

### 3 Simulação das populações e comparação dos estimadores

#### 3.1 População da Bélgica

Considerámos a população  $y = \text{belgianmunicipalities\$TaxableIncome}$  (rendimento tributável total em euros em 2001) disponível na livraria ‘sampling’ do R, com  $N = 589$ . Tomámos  $n = 100$  e considerámos 1000 repetições do processo de amostragem para o plano SSR. Como variável auxiliar utilizámos  $x = \text{belgianmunicipalities\$Tot03}$  (população total no dia 1 de julho de 2003).

Para averiguarmos o tipo de relação entre  $x$  e  $y$ , construímos o correspondente gráfico de dispersão (figura 1).

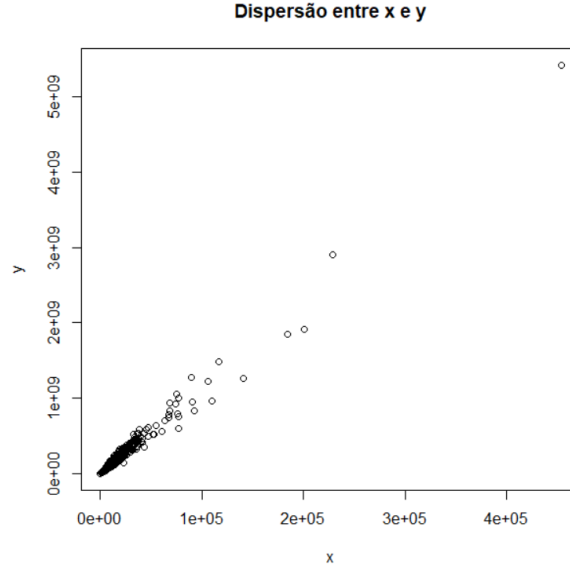


Figura 1: Gráfico de dispersão entre as variáveis  $x$  e  $y$

Como podemos observar, as variáveis  $x$  e  $y$  são fortemente correlacionadas ( $\rho_{xy} = 0.9883982$ ), pelo que é razoável considerar os estimadores do quociente e da regressão para estimar  $\bar{y}$ .

#### Variância empírica e EQM das $M$ amostras

Neste caso, obtivemos as seguintes variâncias empíricas das  $M$  amostras

$$\text{Var}(\hat{\bar{y}}) = 8.254811 \times 10^{14}, \quad \text{Var}(\hat{\bar{y}}_r) = 2.041012 \times 10^{13}, \quad \text{Var}(\hat{\bar{y}}_q) = 1.994624 \times 10^{13}$$

Podemos concluir que a variância do estimador da média amostral é superior à dos outros estimadores. A variância do estimador da regressão e do quociente são muito semelhantes, no entanto, a variância do estimador do quociente é ligeiramente inferior, diferença que não considerámos significativa. Como este resultado contrariou as nossas expectativas, decidimos avaliar o EQM das  $M$  amostras. Obtivemos

$$\text{EQM}(\hat{\bar{y}}) = 8.274905 \times 10^{14}, \quad \text{EQM}(\hat{\bar{y}}_r) = 2.041070 \times 10^{13}, \quad \text{EQM}(\hat{\bar{y}}_q) = 1.994625 \times 10^{13}$$

pelo que o vés dos estimadores não é significativo.

## Distribuição empírica

Vamos então analisar a distribuição empírica dos estimadores de  $\bar{y}$ . Nos gráficos seguintes, a linha a vermelho indica o verdadeiro valor do parâmetro de interesse, a média de  $y$ , e os círculos a azul indicam a média das estimativas obtidas. No caso dos estimadores cêntricos, estes círculos devem estar sobre a linha a vermelho.

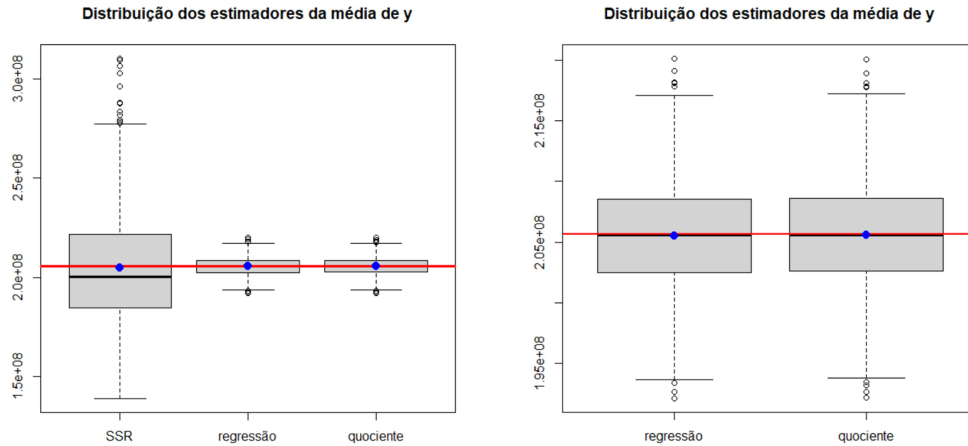


Figura 2: Distribuição empírica dos estimadores de  $\bar{y}$

Pela análise da figura 2, notamos que o estimador da média amostral tem uma maior dispersão em relação aos estimadores da regressão e do quociente, e apresenta mais outliers. Em geral, podemos deduzir que o estimador da regressão e o estimador do quociente parecem fornecer estimativas mais confiáveis e precisas da média da população em comparação com o estimador SSR. Além disso, notamos que os estimadores da regressão e do quociente têm distribuições muito semelhantes em relação à dispersão e outliers.

## Variância de cada amostra

Vamos agora comparar os erros quadráticos médios dos três estimadores. Tendo em conta que temos uma amostra de tamanho considerável, o viés é desprezável, pelo que podemos comparar apenas as variâncias. Obtivemos os seguintes valores para as variâncias, de algumas das  $M$  amostras retiradas:

| SSR                       | regressão                 | quociente                 |
|---------------------------|---------------------------|---------------------------|
| $1.565657 \times 10^{15}$ | $1.596242 \times 10^{13}$ | $1.615225 \times 10^{13}$ |
| $5.448249 \times 10^{15}$ | $3.070870 \times 10^{13}$ | $3.070939 \times 10^{13}$ |
| $2.826985 \times 10^{15}$ | $2.154282 \times 10^{13}$ | $2.220477 \times 10^{13}$ |
| $1.600661 \times 10^{15}$ | $2.015938 \times 10^{13}$ | $2.189999 \times 10^{13}$ |

Tabela 1: Comparação entre as variâncias dos estimadores de  $\bar{y}$

Pela tabela 1, podemos concluir que a variância do estimador da média amostral é superior à dos outros estimadores. A variância do estimador da regressão e do quociente são muito semelhantes, no entanto, aqui a variância do estimador da regressão já é ligeiramente inferior.



### Margem de erro e probabilidade de cobertura dos intervalos de confiança

Vamos analisar a probabilidade de cobertura e a margem de erro dos intervalos de confiança de níveis 90%, 95% e 99% . Obtivemos as seguintes margens de erro para algumas das  $M$  amostras retiradas:

| 825 | SSR       | regressão | quociente | 4   | SSR       | regressão | quociente |
|-----|-----------|-----------|-----------|-----|-----------|-----------|-----------|
| 90% | 67891647  | 6072928   | 6416582   | 90% | 119400957 | 9142871   | 9163490   |
| 95% | 80892175  | 7235829   | 7645289   | 95% | 142264971 | 10893633  | 10918201  |
| 99% | 106315430 | 9509946   | 10048094  | 99% | 186976818 | 14317346  | 14349635  |

Tabela 2: Comparação entre as margens de erro dos intervalos de confiança

Notamos que a margem de erro para o estimador SSR é muito superior à dos outros estimadores. Além disso, a margem de erro para o estimador da regressão e do quociente é muito semelhante, sendo que a da regressão é ligeiramente inferior. Podemos também deduzir que quanto maior o nível de significância, maior a margem de erro do intervalo.

Em relação às probabilidades de cobertura obtivemos os seguintes resultados.

|     | SSR   | regressão | quociente |
|-----|-------|-----------|-----------|
| 90% | 0.993 | 0.817     | 0.831     |
| 95% | 0.995 | 0.892     | 0.902     |
| 99% | 1     | 0.949     | 0.959     |

Tabela 3: Comparação entre as probabilidades de cobertura dos estimadores de  $\bar{y}$  para  $n = 100$

Pela tabela 3, concluímos que as probabilidades de cobertura são inferiores aos níveis exigidos. No entanto, se aumentarmos o valor de  $n$ , por exemplo para  $n = 450$ , obtemos valores mais próximos do desejado, como podemos observar na tabela 4.

|     | SSR   | regressão | quociente |
|-----|-------|-----------|-----------|
| 90% | 0.878 | 0.868     | 0.881     |
| 95% | 0.916 | 0.923     | 0.929     |
| 99% | 0.966 | 0.969     | 0.979     |

Tabela 4: Comparação entre as probabilidades de cobertura dos estimadores de  $\bar{y}$  para  $n = 450$

### QQ -plot

A probabilidade de cobertura do estimador SSR é mais alta do que a dos outros estimadores. Isto pode estar relacionado com a margem de erro dos intervalos de confiança do SSR (como é maior então é mais provável que o parâmetro pertença ao intervalo), mas também com a distribuição de cada estimador, mais propriamente com a distribuição de

$$\frac{\hat{\theta} - \theta}{\sqrt{\text{Var}(\hat{\theta})}}.$$

Assim, decidimos analisar os respectivos QQ-plots. Obtivemos os resultados da figura 3.

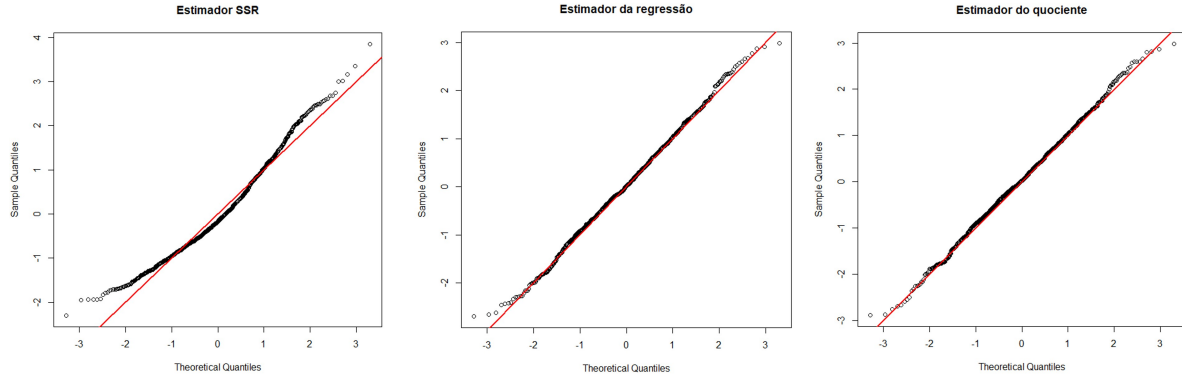


Figura 3: Comparação entre os QQ-plots dos estimadores

Como podemos observar, os estimadores da regressão e do quociente adaptam-se melhor a uma distribuição normal.

### 3.2 População da Suíça

Considerámos a população  $y = \text{swissmunicipalities\$H00PTOT}$  (número de domicílios) disponível na livraria ‘sampling’ do R, com  $N = 2896$ . Tomámos  $n = 800$  e considerámos 2000 repetições do processo de amostragem para o plano SSR. Como variável auxiliar utilizámos  $x = \text{swissmunicipalities\$POPTOT}$  (população total).

Para averiguarmos o tipo de relação entre estas duas variáveis, construímos o correspondente gráfico de dispersão (figura 4).

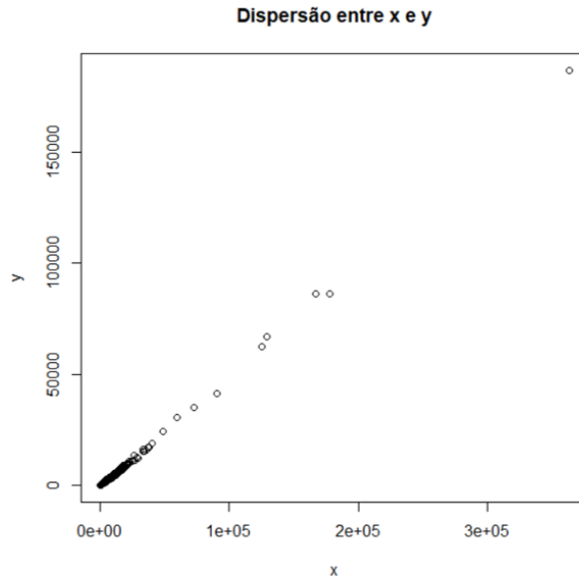


Figura 4: Gráfico de dispersão entre as variáveis  $x$  e  $y$

Como podemos observar, as variáveis  $x$  e  $y$  são fortemente correlacionadas ( $\rho_{xy} = 0.9981804$ ), pelo que

é razoável considerar os estimadores do quociente e da regressão para estimar  $\bar{y}$ .

### Variância empírica das $M$ amostras

Neste caso, obtivemos as seguintes variâncias empíricas das  $M$  amostras

$$\text{Var}(\hat{\bar{y}}) = 22582.7187 \quad \text{Var}(\hat{\bar{y}}_r) = 154.2267 \quad \text{Var}(\hat{\bar{y}}_q) = 504.5406$$

sendo de esperar um melhor desempenho dos estimadores do quociente e da regressão. Contrariamente à população da Bélgica, a variância do estimador da regressão é inferior à do estimador do quociente.

### Distribuição empírica

Vamos então analisar a distribuição empírica dos estimadores.

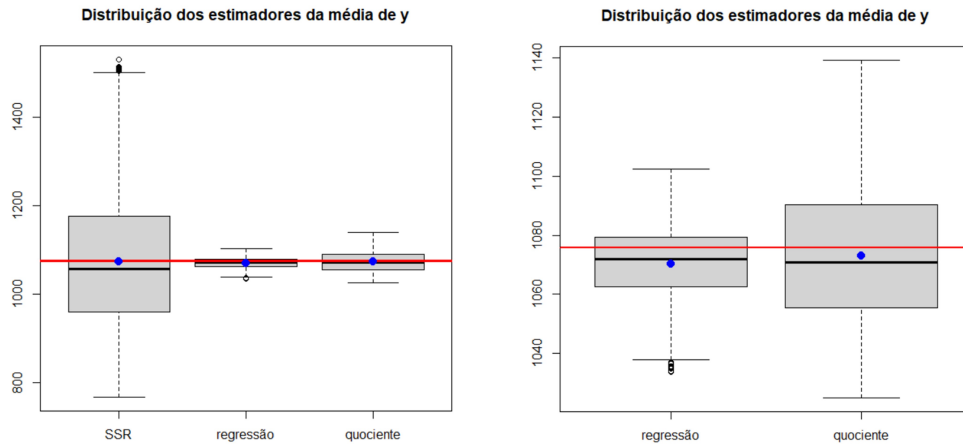


Figura 5: Distribuição empírica dos estimadores de  $\bar{y}$

Pela análise da figura 5, notamos que o estimador da média amostral apresenta novamente uma maior dispersão em relação aos estimadores da regressão e do quociente, e um maior número de outliers. Analogamente à população da Bélgica, podemos deduzir que o estimador da regressão e o estimador do quociente parecem fornecer estimativas mais confiáveis e precisas da média da população em comparação com o estimador SSR. No entanto, apesar dos estimadores da regressão e do quociente terem um comportamento muito semelhante, notamos agora uma maior dispersão no estimador do quociente, o que sugere que o estimador da regressão possa obter melhores resultados.

### Variância de cada amostra

Tendo novamente uma amostra de tamanho considerável, para comparar os EQM dos estimadores basta comparar as respectivas variâncias. Obtivemos os seguintes valores para as variâncias de algumas das  $M$  amostras retiradas:

| SSR       | regressão | quociente |
|-----------|-----------|-----------|
| 10921.411 | 72.38407  | 148.53761 |
| 9368.107  | 54.81317  | 97.02303  |
| 12582.125 | 112.67479 | 346.61465 |
| 21287.403 | 89.62623  | 262.96686 |

Tabela 5: Comparação entre as variâncias dos estimadores de  $\bar{y}$

Pela tabela 5, podemos concluir que a variância do estimador da média amostral é superior à dos outros estimadores. Embora a variância do estimador da regressão e do quociente sejam consideravelmente semelhantes, aqui é notório que a da regressão é inferior.

### Margem de erro e probabilidade de cobertura dos intervalos de confiança

Vamos analisar a probabilidade de cobertura e a margem de erro dos intervalos de confiança de níveis 90%, 95% e 99% . Obtivemos as seguintes margens de erro para algumas das  $M$  amostras retiradas:

| <b>1925</b> | SSR      | regressão | quociente | <b>500</b> | SSR      | regressão | quociente |
|-------------|----------|-----------|-----------|------------|----------|-----------|-----------|
| 90%         | 240.0088 | 15.57340  | 26.67574  | 90%        | 385.6569 | 22.71574  | 53.49887  |
| 95%         | 285.9680 | 18.55554  | 31.78386  | 95%        | 459.5061 | 27.06556  | 63.74333  |
| 99%         | 375.8436 | 24.38728  | 41.77307  | 99%        | 603.9224 | 35.57188  | 83.77695  |

Tabela 6: Comparação entre as margens de erro dos intervalos de confiança

Notamos que a margem de erro para o estimador SSR é bastante superior à dos outros estimadores. Além disso, a margem de erro para o estimador da regressão e do quociente é semelhante, sendo que o da regressão é ligeiramente inferior. Podemos também deduzir que quanto maior o nível de significância, maior a margem de erro do intervalo.

Em relação às probabilidades de cobertura obtivemos os seguintes resultados.

|     | SSR    | regressão | quociente |
|-----|--------|-----------|-----------|
| 90% | 0.823  | 0.755     | 0.76      |
| 95% | 0.8585 | 0.7995    | 0.7915    |
| 99% | 0.918  | 0.8405    | 0.826     |

Tabela 7: Comparação entre as probabilidades de cobertura dos estimadores de  $\bar{y}$  para  $n = 100$

Pela tabela 7, concluímos que as probabilidades de cobertura são inferiores aos níveis exigidos. No entanto, se aumentarmos o valor de  $n$ , por exemplo para  $n = 1500$ , obtemos valores mais próximos do desejado, como podemos observar na tabela 8.

|     | SSR    | regressão | quociente |
|-----|--------|-----------|-----------|
| 90% | 0.862  | 0.84      | 0.828     |
| 95% | 0.9065 | 0.885     | 0.865     |
| 99% | 0.9445 | 0.9305    | 0.913     |

Tabela 8: Comparação entre as probabilidades de cobertura dos estimadores de  $\bar{y}$  para  $n = 1500$

### QQ -plot

Analogamente à população da Bélgica decidimos analisar os QQ-plots. Obtivemos os resultados da figura 6.

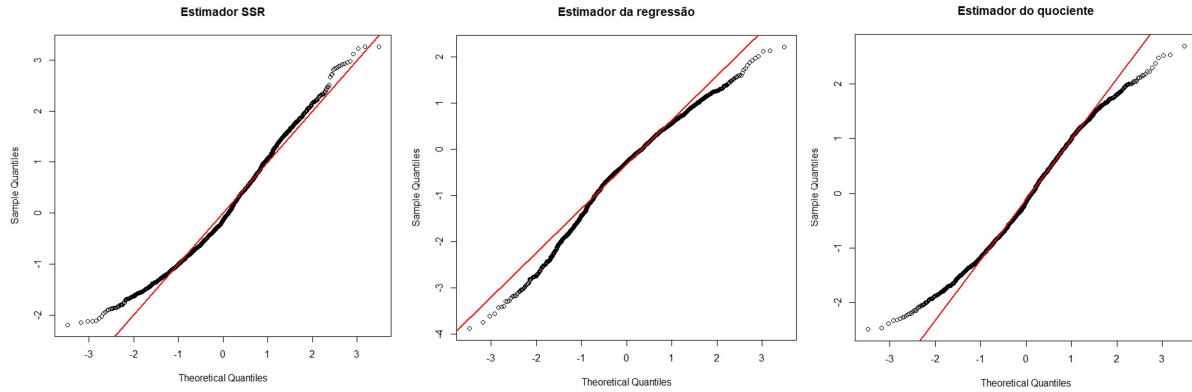


Figura 6: Comparação entre os QQ-plots dos estimadores

Como podemos observar, os estimadores SSR e da regressão adaptam-se melhor a uma distribuição normal.

## 4 Conclusão

Através destes dois exemplos, conseguimos perceber as diferenças existentes na estimação da média pelos três tipos de estimadores: SSR, regressão e quociente.

Em ambas as populações, o estimador da média amostral foi o que apresentou piores resultados. No entanto, as conclusões para o estimador da regressão e do quociente não foram tão evidentes.

Por um lado, o estudo feito da população da Bélgica revela que o estimador do quociente e o da regressão têm eficiências muito semelhantes. Em relação às variâncias de cada amostra, observou-se uma ligeira vantagem na utilização do estimador da regressão. No entanto a probabilidade de cobertura do estimador do quociente é ligeiramente superior. Assim, nesta população podemos concluir que o estimador da regressão e do quociente revelam um comportamento idêntico e mais eficiente que o estimador SSR.

Por outro lado, na população da Suíça os resultados foram mais conclusivos. A variância do estimador da regressão é inferior à dos outros estimadores, assim como a margem de erro dos intervalos de confiança.

Isto mostra que a eficiência destes dois estimadores pode depender da população, mas que em geral têm um comportamento muito semelhante.

## Referências

- [1] Tenreiro, C. (2023). *Notas do curso de Amostragem e Sondagens*.