USP – ICMC – SME0806 - Estatística Computacional
2º trabalho – 1º/2020

1. Os dados abaixo foram coletados de duas variáveis denotadas por X e Y. Deve ser estimada a razão $\theta = \mathrm{E}(X)/\mathrm{E}(Y)$, se $\mathrm{E}(Y) \neq 0$.

   (a) Para diferentes números de amostras *bootstrap* B ($B \leqq 5000$), represente graficamente uma estimativa de θ corrigida em relação ao viés. Compare com a estimativa sem correção de viés. O eixo horizontal do gráfico (B) pode ser representado em escala logarítmica (base 10).

   (b) Apresente uma estimativa do erro padrão do estimador do item 1a.

Observação	1	2	3	4	5	6	7	8
x	8406	2342	8187	8459	4795	3516	4796	10238
y	-1200	2601	-2705	1982	-1290	351	-638	-2719

2. Uma forma de quantificar a concordância entre medições obtidas por dois métodos diferentes (X e Y) é dada pelo coeficiente de correlação de concordância, que é definido como
$$\rho_c = 1 - \frac{\mathrm{DQM}}{\mathrm{DQM}|_{\sigma_{XY}=0}} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2},$$
em que DQM denota o desvio quadrático médio $\mathrm{DQM} = \mathrm{E}[(X-Y)^2] = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$ e $\mathrm{DQM}|_{\sigma_{XY}=0}$ significa que a esperança é calculada supondo que a covariância entre X e Y é nula. Pode ser provado que $\rho_c \in [-1, 1]$. Os valores extremos representam discordância e concordância perfeitas, respectivamente, enquanto que $\rho_c = 0$ representa ausência de associação.

   Um estimador para ρ_c é dado pela sua versão amostral, ou seja,
$$r_c = \frac{2S_{XY}}{\left(\overline{X} - \overline{Y}\right)^2 + S_X^2 + S_Y^2},$$
em que $S_{XY} = \sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})/n$, $S_X^2 = \sum_{i=1}^n (X_i - \overline{X})^2/n$ e $S_Y^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2/n$. Pode ser provado que $Z = \frac{1}{2}\log\left(\frac{1+r_c}{1-r_c}\right)$ tem distribuição normal assintótica com média $\mu_Z = \frac{1}{2}\log\left(\frac{1+\rho_c}{1-\rho_c}\right)$ e variância
$$\sigma_Z^2 = \frac{1}{n-2}\left[\frac{(1-\rho^2)\rho_c^2}{(1-\rho_c^2)\rho^2} + \frac{2v^2(1-\rho_c)\rho_c^3}{(1-\rho_c^2)^2\rho} - \frac{v^4\rho_c^4}{2(1-\rho_c^2)^2\rho^2}\right], \quad n > 2, \tag{1}$$

em que ρ denota o coeficiente de correlação linear de Pearson e $v^2 = (\mu_X - \mu_Y)^2/(\sigma_X \sigma_Y)$.

Os dados abaixo são de uma amostra de pares de determinação do teor de arsênio na água de um rio. As medições, em $\mu g/l$, foram realizadas com dois métodos, identificados como "tradicional" (X) e "novo" (Y).

```
Amostra de pares (x, y)
x: 8,71 7,01 3,28 5,60 1,55 1,75 0,73 3,66 0,90 9,39 4,39
3,69 0,34 1,94 2,07 1,38 1,81 1,27 0,82 1,88 5,66 0,00
0,40 0,00 1,98 10,21 4,64 5,66 19,25

y: 7,35 7,92 3,40 5,44 2,07 2,29 0,66 3,43 1,25 6,58 3,31
2,72 2,32 1,50 3,50 1,17 2,31 1,88 0,44 1,37 7,04 0,49
1,29 0,37 2,16 12,53 3,90 4,66 15,86
```

(a) Represente graficamente os dados e comente sobre a concordância entre os dois métodos.

(b) Utilizando reamostragem, avalie a aproximação assintótica da distribuição de Z.

(c) Compare as estimativas do erro padrão de Z utilizando o resultado assintótico da expressão (1) e reamostragem.

(d) Utilizando intervalos de confiança, discuta a concordância entre os dois métodos.