

Case 4

Camille Peixoto Almeida

30/04/23

1 Intervalo de confiança para um parâmetro - σ conhecido

A seguir são apresentados os intervalos de confiança para a média dos retornos para cada uma das criptomoedas (Dogecoin, Stellar e XRP) do banco de dados "cripto.rds" com 95% de confiança e com desvios padrão populacionais conhecidos. Portanto, usa-se a seguinte expressão analítica para encontrar esses intervalos:

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (1)$$

Em que \bar{X} é a média dos retornos, σ é o desvio padrão conhecido e o $z_{\frac{\alpha}{2}}$ são os valores da tabela de distribuição normal de probabilidades.

Criptomoeda	Desvio padrão	Média amostral	n	Intervalo - 95% de confiança (z = 1,96)
Dogecoin	0,02	0,00166	1811	$0,00075 < \mu < 0,00259$
Stellar	0,015	0,00330	1578	$0,00256 < \mu < 0,00404$
XRP	0,012	0,00266	1944	$0,00212 < \mu < 0,00319$

1.1 Para 90% de confiança:

Criptomoeda	Desvio padrão	Média amostral	n	Intervalo - 90% de confiança (z = 1,645)
Dogecoin	0,02	0,00166	1811	$0,00089 < \mu < 0,00244$
Stellar	0,015	0,00330	1578	$0,00268 < \mu < 0,00393$
XRP	0,012	0,00266	1944	$0,00221 < \mu < 0,00311$

1.2 Para 99% de confiança:

Criptomoeda	Desvio padrão	Média amostral	n	Intervalo - 99% de confiança (z = 2,575)
Dogecoin	0,02	0,00166	1811	$0,00046 < \mu < 0,00288$
Stellar	0,015	0,00330	1578	$0,00233 < \mu < 0,00428$
XRP	0,012	0,00266	1944	$0,00196 < \mu < 0,00336$

Discussão: Vê-se que para níveis de confiança maiores o intervalo de confiança aumenta. Isso ocorre, porque o nível de confiança representa a probabilidade de que o intervalo de confiança contenha o valor do parâmetro estimado.

Portanto, quando se aumenta o nível de confiança exige-se uma maior certeza de que o parâmetro esteja no intervalo. Para garantir isso temos que aumentar a amplitude do intervalo.

2 Intervalo de confiança para σ desconhecido

Para quando os desvios padrão populacionais desconhecidos utilizamos a expressão analítica seguinte:

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \quad (2)$$

Em que, novamente, \bar{X} e σ são, respectivamente, a média amostral e o desvio padrão que, neste caso, é desconhecido e, por esse motivo, **utilizamos os valores t da tabela de distribuição t-Student**.

2.1 Para 90% de confiança:

Criptomoeda	Desvio amostral	Média amostral	n	G.L	t (90% - confiança)	Intervalo
Dogecoin	0,08514	0,00166	1811	1810	1,645	$-0,00162 < \mu < 0,00496$
Stellar	0,08518	0,00330	1578	1577	1,645	$-0,00022 < \mu < 0,00683$
XRP	0,08110	0,00266	1944	1943	1,645	$-0,00037 < \mu < 0,00568$

2.2 Para 95% de confiança:

Criptomoeda	Desvio amostral	Média amostral	n	G.L	t (95% - confiança)	Intervalo
Dogecoin	0,08514	0,00166	1811	1810	1,96	$-0,00226 < \mu < 0,00559$
Stellar	0,08518	0,00330	1578	1577	1,96	$-0,00090 < \mu < 0,00751$
XRP	0,08110	0,00266	1944	1943	1,96	$-0,00095 < \mu < 0,00626$

2.3 Para 99% de confiança:

Criptomoeda	Desvio amostral	Média amostral	n	G.L	t (99% - confiança)	Intervalo
Dogecoin	0,08514	0,00166	1811	1810	2,575	$-0,00349 < \mu < 0,00682$
Stellar	0,08518	0,00330	1578	1577	2,575	$-0,00222 < \mu < 0,00883$
XRP	0,08110	0,00266	1944	1943	2,575	$-0,00208 < \mu < 0,00739$

Observação: Os valores de t e z coincidem uma vez que o número de graus de liberdade para as amostras é alto. Isso implica que a distribuição t-Student tenda à distribuição normal de probabilidades. Em consequência disso os valores t tendem a se aproximar dos valores de z.

Os valores de t e z coincidirem é uma condição que favorece que os intervalos de confiança se aproximem, pois a diferença será dada somente em função do valor do desvio amostral.

Analisando a diferença entre os intervalos de σ conhecido para σ desconhecido:

Confiança(%)	Criptomoeda	Amplitude do intervalo		Diferença relativa das amplitudes (%)
		Para σ conhecido	Para σ desconhecido	
90	Dogecoin	0.00155	0.00658	324.516
	Stellar	0.00125	0.00705	464
	XRP	0.0009	0.00605	572.222
95	Dogecoin	0.00184	0.00785	326.630
	Stellar	0.00148	0.00841	468.243
	XRP	0.00107	0.00721	573.832
99	Dogecoin	0.00242	0.01031	326.033
	Stellar	0.00195	0.01105	466.667
	XRP	0.0014	0.00947	576.429

Em que a coluna chamada "**Diferença relativa das amplitudes (%)**" significa o quanto, em porcentagem, a amplitude para σ desconhecido é maior que para σ conhecido, ou seja:

$$D = 100 \cdot \frac{A_2 - A_1}{A_1} \quad (3)$$

Em que:

1. D é a diferença relativa das amplitudes
2. A_1 é a amplitude para σ conhecido
3. A_2 é a amplitude para σ desconhecido

Da tabela acima podemos observar diretamente que, para um específico nível de confiança, todos os os intervalos de σ desconhecido foram maiores que os intervalos de confiança quando σ é conhecido.

O quanto o intervalo de σ desconhecido foi maior em relação ao intervalo de σ conhecido pode ser visto na diferença relativa das amplitudes.

Assim, é possível analisar que para a Dogecoin o intervalo aumenta em torno de 326% quando σ não é conhecido. Para a criptomoeda Stellar, o aumento é de, aproximadamente, 466% e, para a XRP, aproximadamente, 574%.

Essa diferença ocorre, uma vez que os desvios amostrais das criptomoedas são maiores que os desvios padrão populacionais dados e em torno das mesmas influências de porcentagem nos aumentos.

	Desvios padrão populacionais	Desvios amostrais	Diferença relativa de desvios (%)
Dogecoin	0,02	0,08514	325.7
Stellar	0,015	0,08518	467.867
XRP	0,012	0,08110	575.833

3 Expressões analíticas da diferença de médias

3.1 Com variâncias conhecidas

$$\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{(\sigma_{X_1})^2}{n_1} + \frac{(\sigma_{X_2})^2}{n_2}} < d < \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{(\sigma_{X_1})^2}{n_1} + \frac{(\sigma_{X_2})^2}{n_2}} \quad (4)$$

Em que:

1. d é a diferença de médias (parâmetro estimado)
2. \bar{X}_1 é a média amostral para a população 1 (criptomoeda 1).
3. \bar{X}_2 é a média amostral para a população 2 (criptomoeda 2).
4. σ é o desvio populacional conhecidos para cada criptomoeda.
5. n é o tamanho de cada amostra.

3.2 Com variâncias desconhecidas e iguais

$$\bar{X}_1 - \bar{X}_2 - t_{\frac{\alpha}{2}} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < d < \bar{X}_1 - \bar{X}_2 + t_{\frac{\alpha}{2}} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (5)$$

Em que a distribuição do parâmetro mudará para a distribuição **t-Student** e S_p é o estimador ponderado para σ^2 e é dado por:

$$S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \quad (6)$$

Sendo S_1^2 e S_2^2 são as variâncias amostrais de cada criptomoeda 1 e 2.

3.3 Com variâncias desconhecidas e não necessariamente iguais

$$\bar{X}_1 - \bar{X}_2 - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < d < \bar{X}_1 - \bar{X}_2 + t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad (7)$$

4 Intervalo de confiança para a diferença das médias de retorno de XRP e Stellar

4.1 Com desvios padrão populacionais conhecidos

Para 95% de confiança com desvios padrão populacionais conhecidos (utiliza-se valores de z da tabela de distribuição normal de probabilidades) e iguais a $\sigma_{XRP} = 0.02$ e $\sigma_{Stellar} = 0.012$, calcula-se o intervalo de confiança por meio da equação 4. Resulta-se em :

$$-0,000456 < d < 0,001715$$

com $\bar{X}_{XRP} = 0,002658$, $\bar{X}_{Ste} = 0,003304$, $n_{XRP} = 1944$ e $n_{Ste} = 1578$.

4.2 Com desvios padrão populacionais desconhecidos e iguais

Para 95% de confiança com desvios padrão populacionais desconhecidos e iguais (utiliza-se valores de t da tabela de distribuição t-Student de probabilidades), calcula-se por meio da expressão analítica 5. Resulta-se em:

$$-0,004863 < d < 0,006156$$

com :

1. $\bar{X}_{XRP} = 0,002658$ e $\bar{X}_{Ste} = 0,003304$ (médias amostrais)
2. $n_{XRP} = 1944$ e $n_{Ste} = 1578$ (números amostrais)
3. $S_{XRP}^2 = 0,006578$ e $S_{Ste}^2 = 0,007255$ (variâncias amostrais)
4. $S_p = 0,082952$ (estimador ponderado para σ)

4.3 Com desvios padrão populacionais desconhecidos e não necessariamente iguais

Para 95% de confiança com desvios populacionais e não necessariamente iguais (utiliza-se valores de t da tabela de distribuição t-Student de probabilidades), calcula-se por meio da expressão analítica 7. Resulta-se em:

$$-0,004891 < d < 0,006184$$

com :

1. $\overline{X}_{XRP} = 0,002658$ e $\overline{X}_{Ste} = 0,003304$ (médias amostrais)
2. $n_{XRP} = 1944$ e $n_{Ste} = 1578$ (números amostrais)
3. $S^2_{XRP} = 0,006578$ e $S^2_{Ste} = 0,007255$ (variâncias amostrais)

5 Script - Case 4

```
##### Camille Peixoto Almeida 12702259 #####
# importação de bibliotecas
library(tidyverse)
library(ggplot2)

# selecionar a base de dados
df <- readRDS("cripto.rds")

# retirar os espacos nulos (NA)
df <- na.omit(df)

#embaralhar o conjunto de dados
df_embaralhado <- df[sample(1:nrow(df)), ]

#criei 3 banco de dados separados para cada moeda
df_Dogecoin <- subset(df, df$name == "Dogecoin")
df_Stellar <- subset(df, df$name == "Stellar")
df_XRP <- subset(df, df$name == "XRP")

# criei coluna chamada retorno para cada um dos dados para que o dado da
# dogecoin não seja considerado no retorno da stellar e o dado da stellar não
# atrapalhe o retorno da XRP

df_Dogecoin$retornoDogecoin = 1
for (i in 2:1811) {
  df_Dogecoin$retornoDogecoin[i] <- log(df_Dogecoin$close[i]/df_Dogecoin$close[i-1],
    base = exp(1))
}

df_Stellar$retornoStellar = 1
for (i in 2:1578) {
  df_Stellar$retornoStellar[i] <- log(df_Stellar$close[i]/df_Stellar$close[i-1],
    base = exp(1))
}

df_XRP$retornoXRP = 1
```

```

for (i in 2:1944) {
  df_XRP$retornoXRP[i] <- log(df_XRP$close[i]/df_XRP$close[i-1], base = exp(1))
}

# medias amostrais dos retornos de cada moeda

media_doge <- mean(df_Dogecoin$retornoDogecoin)
media_ste <- mean(df_Stellar$retornoStellar)
media_XRP <- mean(df_XRP$retornoXRP)

# DESVIOS PADRÃO CONHECIDOS - NORMAL

# intervalo de confiança
sigma_doge <- 0.02
sigma_ste <- 0.015
sigma_XRP <- 0.012

# 95 % de confiança
z95<- 1.96
# 90% de confiança
z90 <- 1.645
# 99% de confiança
z99 <- 2.575

z <- z99

ndoge <- 1811
extremoD_doge <- media_doge + z*sigma_doge/(ndoge)^0.5
extremoE_doge <- media_doge - z*sigma_doge/(ndoge)^0.5

nste<- 1578
extremoD_ste <- media_ste + z*sigma_ste/(nste)^0.5
extremoE_ste <- media_ste - z*sigma_ste/(nste)^0.5

nXRP <- 1944
extremoD_XRP <- media_XRP + z*sigma_XRP/(nXRP)^0.5
extremoE_XRP <- media_XRP - z*sigma_XRP/(nXRP)^0.5

#DESVIOS PADRÃO DESCONHECIDOS - T-STUDENT

desv_amostral_doge <- sd(df_Dogecoin$retornoDogecoin)
desv_amostral_ste <- sd(df_Stellar$retornoStellar)
desv_amostral_XRP <- sd(df_XRP$retornoXRP)

```

```

# 95 % de confiança
t95<- 1.96
# 90% de confiança
t90 <- 1.645
# 99% de confiança
t99 <- 2.575

t <- t95

ndoge <- 1811
extremoD_doge_desc <- media_doge + t*desv_amostral_doge/(ndoge)^0.5
extremoE_doge_desc <- media_doge - t*desv_amostral_doge/(ndoge)^0.5

nste <- 1578
extremoD_ste_desc <- media_ste + t*desv_amostral_ste/(nste)^0.5
extremoE_ste_desc <- media_ste - t*desv_amostral_ste/(nste)^0.5

nXRP <- 1944
extremoD_XRP_desc <- media_XRP + t*desv_amostral_XRP/(nXRP)^0.5
extremoE_XRP_desc <- media_XRP - t*desv_amostral_XRP/(nXRP)^0.5

#####
# Intervalo de Confiança para Dois Parâmetros : o intervalo de confiança
#para a diferença da média dos retornos entre as moedas XRP e Stellar

# sigmas conhecidos
diferenca_media <- media_ste - media_XRP
sigma_XRP_part2 <- 0.02
sigma_ste_part2 <- 0.012

extremoD_dif <- diferenca_media + z95*(((sigma_XRP_part2)^2)/nXRP
+((sigma_ste_part2)^2)/nste)^0.5
extremoE_dif <- diferenca_media - z95*(((sigma_XRP_part2)^2)/nXRP
+((sigma_ste_part2)^2)/nste)^0.5

var_XRP <- var(df_XRP$retornoXRP)
var_ste <- var(df_Stellar$retornoStellar)

# sigmas desconhecidos e iguais

Sp <- (((nste-1)*var_ste+(nXRP -1)*var_XRP)/(nste + nXRP - 2))^0.5

extremoD_dif_desc_iguais <- diferenca_media + Sp*t*(nste^(-1)+ nXRP^(-1))^0.5
extremoE_dif_desc_iguais <- diferenca_media - Sp*t*(nste^(-1)+ nXRP^(-1))^0.5

```



```
# sigmas desconhecidos e não necessariamente iguais

extremoD_dif_desc_diferentes <- diferenca_media +
t*(var_ste/nste+ var_XRP/nXRP)^0.5
extremoE_dif_desc_diferentes <- diferenca_media -
t*(var_ste/nste+ var_XRP/nXRP)^0.5
```

6 Referências

1. HO, Linda Lee; RIBEIRO, Celma de Oliveira. Intervalo de confiança.PRO3200 - Estatística, Departamento de Engenharia e Produção, Universidade de São Paulo.2022.
2. RAMOS, Alberto. Apostila de Estatística-PRO3200. Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Produção, São Paulo.2021