

# Universidade Federal do Rio de Janeiro

## Centro de Ciências da Saúde

### Instituto de Biofísica Carlos Chagas Filho

**Disciplina:** CFB017 - Programação para Biociências

**Professor:** Dr. Vitor Lima Coelho

## Trabalho de Avaliação Contínua - nº 2

Em relação à aula “Manipulação de dados com Python - Parte 2 - pandas básico”, resolva o problema de programação.

1 - A pesquisadora Rosiane realizou um experimento de Real-Time qRT-PCR (Real-Time Quantitative Reverse Transcription PCR) para calcular a abundância ou número de cópias de um conjunto de RNAs de interesse (RNA\_1, RNA\_2, RNA\_3, ...) em diferentes condições (Condicao\_A, Condicao\_B, Condicao\_C, ...) e solicitou ao aluno Vitor que fizesse a quantificação absoluta. Para ajudá-lo nesta tarefa, a professora Rosiane forneceu o número de de moléculas do RNA controle e os respectivos valores de CT (*threshold cycle*) (tabela 1).

**Tabela 1. Número de cópias do RNA controle por CT.**

cópias de RNA controle/ $\mu$ g	Log cópias de RNA controle por micrograma	CT
2.50E+13	13.39794001	12.856
2.50E+12	12.39794001	16.1430
2.50E+10	10.39794001	22.9570
2.50E+09	9.397940009	26.3000
2.50E+08	8.397940009	29.8680

Vitor aprendeu que o primeiro passo para estimar a quantificação absoluta é obter a curva padrão a partir da Tabela 1, obtendo assim uma equação na forma  $y = mx + b$ . Fazendo o gráfico onde o eixo X é o logaritmo do número de cópias de RNA por micrograma e o eixo Y são os respectivos valores de CT, Vitor obteve a seguinte curva padrão (Figura 1):

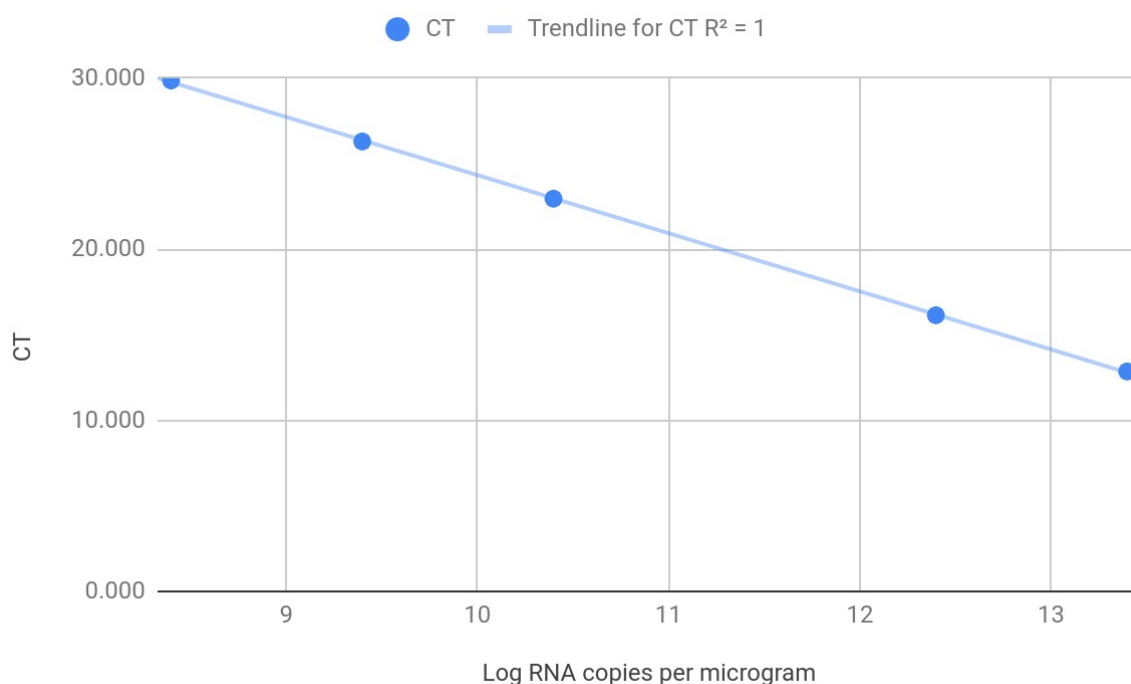


Figura 1: curva padrão.

Estimando os coeficientes da equação da curva padrão  $y = mx + b$  usando regressão linear, Vitor obteve a seguinte equação de curva padrão (Equação 1):

$$y = -3.397186047x + 58.53223295 \quad (\text{Equação 1})$$

Sabendo que a quantificação absoluta é dada pela Equação 2, dada por

$$Quantity = 10^{\left(\frac{C_t - b}{m}\right)} \quad (\text{Equação 2})$$

escreva um código em Python para calcular a quantificação absoluta dos RNAs de interesse (RNA\_1, RNA\_2, RNA\_3, ...) de acordo com os pontos a seguir:

1 - O programa deve ler a partir da linha de comando a planilha com os valores de CT para cada RNA de interesse em diferentes condições e os coeficiente  $m$  (*slope*) e o coeficiente  $b$  (*intercept-y*) da equação da curva padrão já calculada previamente. Por exemplo, suponha que seu código chama “calculaQuantAbsoluta.py”, você executaria da seguinte maneira:

```
python calculaQuantAbsoluta.py tabela.xlsx coeficiente_m coeficient_b
```

2 - Crie um dataframe nomeado como “df” para a tabela Excel passada como parâmetro;

3 - Crie um novo dataframe nomeado como “df\_q” que contém as colunas “Target\_Name”, “Stage” e “Quantity”, onde *Quantity* é calculado para cada RNA de interesse usando a

Equação 2.

4 - Junte os dataframes df e df\_q em um único dataframe df\_final.

5 - Salve df\_final como uma nova planilha.

6 - Imprima na tela o df\_final.



### Código 1:

```
## ITEM 1
# importação de bibliotecas
# o meu apenas funciona se eu importar pandas E pandas as pd.
import sys
import pandas
import pandas as pd
import xlrd
# para chamar no prompt de comando (terminal):
tabela = sys.argv[1]
coef_m = float(sys.argv[2])
coef_b = float(sys.argv[3])
## ITEM 2
df = pd.read_excel(tabela)
## ITEM 3
# criação de uma nova tabela:
# a coluna 'Stage e Target_Name' possuem valores repetidos. O que
# faria com que o output tivesse o TRIPLÔ de linhas.
df_q = df[['Target_Name', 'Stage']].drop_duplicates()
# adição de uma nova coluna:
df_q['Quantity'] = 10 ** ((df['CT']-coef_b)/coef_m)
## ITEM 4
# juntar as tabelas (deleção das colunas iguais):
df_final = pd.merge(df, df_q)
## ITEM 5
# salvar a nova tabela:
df_final.to_excel("/home/carol/Documents/CFB017/
Tabela_Qntd_Abs.xlsx", sheet_name="CT_Abs")
## ITEM 6
print(df_final)
# para chamar a função no prompt digite no terminal:
# python nome_desse_script.py caminho_até_a_sua_tabela coeficiente_m
coeficiente_b
#python                                     TAC2_Quantidade_Absoluta.py
~/Documents/CFB017/Valores_CT.xlsx -3.397186047 58.53223295
```

**Código 2 (Outra forma de fazer, mas pula itens):**

```
# importação de bibliotecas:
import sys
import pandas
import xlrd

# para chamar no prompt de comando (terminal):
tabela = sys.argv[1] #
coef_m = float(sys.argv[2])
coef_b = float(sys.argv[3])
CT_values = pandas.read_excel(tabela)
## adiciona uma nova coluna na tabela.
## os valores dessa coluna serão as quantidades absolutas calculadas
com a equação 2.
CT_values['Quantity'] = 10 ** ((CT_values['CT']-coef_b)/coef_m)
## salvar alterações em outro arquivo excel.
CT_values.to_excel("/home/carol/Documents/CFB017/
Tabela_Qntd_Abs_anna.xlsx",sheet_name="CT_Abs")
## impressão da tabela:
print(CT_values)
# para chamar a função no prompt digite no terminal:
# python nome_desse_script.py caminho_até_a_sua_tabela coeficiente_m
coeficiente_b
#python TAC2_Quantidade_Absoluta.py
~/Documents/CFB017/Valores_CT.xlsx -3.397186047 58.53223295
```