In [1]:
```
import numpy as np
import math as math
import random
import csv
import matplotlib.pyplot as plt

%matplotlib inline
```

8.Линейная регрессия. Задача 2.

Условие: Пусть $X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \ldots + \varepsilon_i$, где $i = 0, 1, \ldots, n$ -- расстояния, которое проехал трамвай за i секунд по показанию датчика. Здесь β_1 - начальное раастояние, β_2 - скорость трамвая, ε_0 - ошибка начального показания датчика. Трамвай едет с постоянной скоростью, и через каждую секунду датчик фиксирует расстояние, которое проехал трамвай. Отсчет времени идет от предыдущего замера, причем отсчет идет с ошибкой. Для $i = 0, 1, \ldots, n$ величина ε_i, есть ошибка приращения растояния, то есть $\varepsilon_i = \varepsilon_i^t \beta_2$, где ε_i^t - ошибка отсчета времени. Все ошибки ε_i независимы и распределены по закону $N(0, \sigma^2)$.

Сведите задачу к линейной модели и найдите оценки наименьших квадратов для начального расстояния β_1 и скорости β_2, а также несмещенную оценку для σ^2, из которой выразите оценку дисперсии отсчета времени.

Данные возьмите из файла Regression.csv. Сделайте выводы.

Из теоретической задачи 8.2.

$$\widehat{\beta_1} = X_0$$

$$\widehat{\beta_2} = \frac{X_n - X_0}{n}$$

$$\widehat{\sigma^2} = \frac{1}{n-2} \sum_{i=1}^{n} \left(X_i - X_{i-1} - \frac{X_n - X_0}{n} \right)^2$$

Оценим дисперсию отсчета времени:

$\varepsilon_i \sim N(0, \sigma^2),$
$\varepsilon_i^t = \frac{\varepsilon_i}{\beta_2}$, значит, $\varepsilon_i^t \sim N(0, \frac{\sigma^2}{\beta_2^2})$

$$\hat{\sigma_t^2} = \frac{\hat{\sigma^2}}{\hat{\beta_2^2}}$$

In [2]:
```
#Считаем данные
with open('Regression.csv', 'r') as file:
    data = list(map(float, file))
n = len(data) - 1
```

In [3]:
```
#Оценки начального расстояния и скорости из теор.задачи
beta_1 = data[0]
beta_2 = (data[n] - data[0]) / n

sigma = 0
for i in range(1, len(data)):
    sigma += (data[i] - data[i - 1] - (data[n] - data[0])/n)**2
sigma /= (n-2)

sigma_t = sigma / (beta_2**2)
```

In [4]:
```
print("beta_1:  ", beta_1)
print("beta_2:  ", beta_2)
print("sigma:   ", sigma)
print("sigma_t: ", sigma_t)
```

```
beta_1:   82.0053
beta_2:   11.970782982982982
sigma:    1.5267747059886494
sigma_t:  0.01065442069716372
```

Вывод:

Результаты показывают, что линейная модель подходит для данной выборки. Это можно объяснить тем, что движение трамвая близко к равномерному, и скорость β_2 можно оценить. Видно, что показания датчика довольно точные (дисперсия мала), а потому линейная модель дает хорошее приближение.