Задача 8.2

In [1]:

```
%matplotlib inline
import scipy.stats as st
import numpy as np
import matplotlib.pyplot as plt
import math
import statsmodels.api as sm
import csv
```

In [5]:

```
data = np.genfromtxt('Regression.csv')
```

Задача аналогична теоретической задачи. Сведем к линейной модели. Рассмотреть разности вида $Y_i = X_{i+1} - X_i = \beta_2 + \varepsilon_i$ для $i > 0$. А также $Y_0 = X_0 = \beta_1 + \varepsilon_0$. Значение Y_i это расстояние, которое проезжает трамвай за i-ую секунду. Найдем оценку коэффициентов линейной регрессии методом наименьших квадратов:

- $\hat{\theta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$
- $\hat{\theta} = (Z^T Z)^{-1} Z^T Y$, где Y - вектор данных
- $Z = \begin{pmatrix} 1\ 0 \\ 0\ 1 \\ \dots \\ 0\ 1 \end{pmatrix}$

Решением является:

- $\overset{\wedge}{\beta_1} = Y_0$
- $\overset{\wedge}{\beta_2} = \overline{Y_i}$, где $i = \overline{1, n}$

Оптимальная оценка σ^2: $\overset{\wedge}{\sigma^2} = \frac{1}{n-k} ||Y - Z\hat{\theta}||^2$

- $\overset{\wedge}{\sigma^2} = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \frac{n}{n-2} s^2$

In [21]:

```
y = np.array([data[0]])
for i in range(1, len(data)):
    y = np.append(y, data[i] - data[i - 1])

beta_1 = data[0]
beta_2 = np.mean(y)
sigma = (len(data)-1) * st.moment(y, moment=2) / (len(data) - 2)
print("Оценка начального положения " + str(beta_1))
print("Оценка скорости "+ str(beta_2))
```

```
Оценка начального положения 20.6165
Оценка скорости 12.0410442
```

In [22]:

```
print("Дисперсия скорости " + str(sigma))
print("Дисперсия времени "+ str(sigma/beta_2**2))
```

```
Дисперсия скорости 4.332610261354912
Дисперсия времени 0.029882802380412566
```

Вывод

Дисперсия времени мала, следовательно показания датчика довольно точные. Линейная модель дает хорошее приближение.

In [ ]:

```

```