Задача 9.2

In [1]:

```
import numpy as np
import scipy.stats as sps
import matplotlib.pyplot as plt
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

$X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \ldots \varepsilon_i$, где ε_i независимы, $\varepsilon_i \sim N(0, \sigma^2)$ и $\varepsilon_i = \varepsilon_i^t \beta_2$, $i = 1, .., n$. Тогда

$$\begin{pmatrix} X_0 \\ X_1 \\ .. \\ X_n \end{pmatrix} = \begin{pmatrix} \beta_1 + \varepsilon_0 \\ \beta_1 + \beta_2 + \varepsilon_0 + \varepsilon_1 \\ .. \\ \beta_1 + \beta_2 + \varepsilon_0 + \ldots + \varepsilon_n \end{pmatrix}$$

Вычтем из каждой строчки матрицы предыдущую, введём столбец из $Y_0, \ldots Y_n$:

$$\begin{pmatrix} Y_0 \\ Y_1 \\ .. \\ Y_n \end{pmatrix} = \begin{pmatrix} X_0 \\ X_1 - X_0 \\ .. \\ X_n - X_{n-1} \end{pmatrix} = \begin{pmatrix} \beta_1 + \varepsilon_0 \\ \beta_2 + \varepsilon_1 \\ .. \\ \beta_2 + \varepsilon_n \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ .. \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ .. \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ .. & .. \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ .. \\ \varepsilon_n \end{pmatrix}$$

Значит, в построенной гаусовской линейной модели $Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ .. & .. \\ 0 & 1 \end{pmatrix}$, $Y = \begin{pmatrix} X_0 \\ X_1 - X_0 \\ .. \\ X_n - X_{n-1} \end{pmatrix}$, оценки наименьших квадратов для β_1, β_2 - $\begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \end{pmatrix} = (Z^T Z)^{-1} Z^T Y$

Найдем $\widehat{\beta_1}, \widehat{\beta_2}$:

In [2]:

```
file_obj = open('Regression.csv')
data = np.array([float(line.strip()) for line in file_obj])
```

In [3]:

```
n = data.size
y = np.zeros(n)
y[1:] = data[:n - 1]
y = data - y
z = np.empty((n, 2))
z[:, 0] = np.zeros(n)
z[:, 1] = np.ones(n)
z[0, 0] = 1
z[0, 1] = 0
y.reshape(y.size, 1)
theta_est = np.linalg.inv(z.T.dot(z)).dot(z.T).dot(y)
```

In [4]:

```
print theta_est
```

```
[ 63.5725       9.96734144]
```

Несмещённая оценка для σ^2 - $\frac{\|Y-Z\widehat{\theta}\|^2}{n+1-k}$, так как здесь $k = 2$, то - $\frac{\|Y-Z\widehat{\theta}\|^2}{n-1} = \frac{\sum_{i=0}^{n}(Y_i-(Z\widehat{\theta})_i)^2}{n-1}$:

In [5]:

```
sigma_2_est = ((y - z.dot(theta_est))**2).sum() / (n - 1)
```

In [6]:

```
print sigma_2_est
```

```
4.21857003554
```

$\varepsilon = \varepsilon_i^t \beta_2 \sim N(0, \sigma^2)$, значит, $\varepsilon_i^t \sim N\left(0, \frac{\sigma^2}{\beta_2^2}\right)$. Оценка диспресии отсчета времени - $\frac{\widehat{\sigma^2}}{\widehat{\beta_2}^2}$:

In [7]:

```
print sigma_2_est / ((theta_est[1]) ** 2)
```

```
0.0424626009249
```

Заметим, что получившаяся оценка дисперсии отсчета времени мала.