The probrem is thus $\min_{\beta_0,\beta_1} \left\{ \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 = f(\beta_0,\beta_1) \right\}$

$$\frac{\partial f}{\partial \beta_0} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)]$$

$$\frac{\partial f}{\partial \beta_0} = 0 \iff \sum_{i=1}^{n} y_i - (\beta_0 + \beta_1 x_i) = 0$$

$$\iff \sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i$$

$$\frac{\partial f}{\partial \beta_1} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)]x_i$$

$$\frac{\partial f}{\partial \beta_1} = 0 \iff \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]x_i = 0$$

$$\iff \sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

After some algebra we get the following results for β_0 and β_1:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad \text{and}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n}$$

Example. 4 bulbs are tested for relationship between lifetime and operating voltage.

i	1	2	3	4
V, x_i	105	110	115	120
L, y_i	1400	1200	1120	950

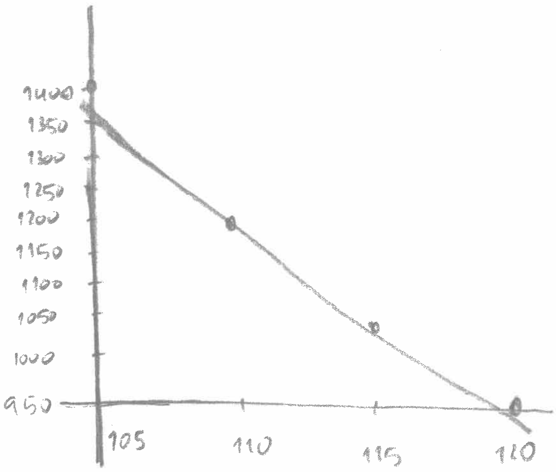
To find β_0 an β_1 we need:

$$\sum_{i=1}^{n} x_i = 450 \qquad \sum_{i=1}^{n} x_i^2 = 50750$$

$$\sum_{i=1}^{n} y_i = 4670 \qquad \sum_{i=1}^{n} x_i y_i = 521800$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{4 \cdot 521800 - 450 \cdot 4670}{4 \cdot 50750 - (450)^2} = -28.6$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{4670 + 28.6 \cdot 450}{4} = 4385$$



The regression line equation is

$$y = 4385 - 28.6X$$

As $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are the observed values of a sample, β_0 and β_1 are indeed random variables and estimators of the real population parameters $\hat{\beta}_0, \hat{\beta}_1$ of β_0, β_1 in $y = \beta_0 + \beta_1 x$ (the real population regression line). If we can find the distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$, then we can compute confidence intervals for β_0 and β_1 and test hypothesis on different values. To do that we need a probabilistic approach to the problem. Assume the sample $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and the model $y_i = \beta_0 + \beta_1 x_i + e_i$ where $i = 1, 2, \ldots n$ and the e_i are random noise such that they are independent, $E[e_i] = 0$ and $Var(e_i) = \sigma^2$. we consider the x_i as fixed but the y_i are random becau of the noise e_i

Observation: $y_i - (\beta_0 + \beta_1 x_i) = e_i$ so the least-squares method seeks to minimize the noise.

Theorem. Under the above assumptions $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$

Proof. Under the statistical model

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \quad \text{then}$$

$$E[\hat{\beta}_0] = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n E[y_i] - \sum_{i=1}^n x_i \sum_{i=1}^n x_i E[y_i]}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{\sum_{i=1}^n x^2 \sum_{i=1}^n (\beta_0 + \beta_1 x_i) - \sum_{i=1}^n x_i \sum_{i=1}^n x_i (\beta_0 + \beta_1 x_i)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{\sum_{i=1}^n x_i^2 \left(n\beta_0 + \beta_1 \sum_{i=1}^n x_i\right) - \sum_{i=1}^n x_i \left(\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{n\beta_0 \sum_{i=1}^n x_i^2 - \beta_0 \left(\sum_{i=1}^n x_i\right)^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$= \frac{\left[n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right]\beta_0}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$= \beta_0$ And similarly for β_1.

Theorem. Under the above assumptions

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \sigma_{\hat{\beta}_0}^2, \quad Var(\hat{\beta}_1) = \frac{n\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \sigma_{\hat{\beta}_1}^2$$

Theorem. If $e_i \sim N(0, \sigma^2)$ $i = 1, 2, \dots, n$ independent, then

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0}) \quad \text{and} \quad \hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

If we know σ^2 this would be enough to compute confidence intervals and test hypothesis, however that won't be always the case. We need to estimate σ^2 from the data.

Define the residual sum of squares RSS as:

$$RSS = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

Proposition. $s^2 = \frac{RSS}{n-2}$ is an unbiased estimate for σ^2

(that is $E[s^2] = \sigma^2$) and

$$s^2_{\hat{\beta}_0} = \frac{s^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad s^2_{\hat{\beta}_1} = \frac{n s^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

are estimates of $\hat{\sigma}^2_{\beta_0}$ and $\hat{\sigma}^2_{\beta_1}$ respectively

Proposition. Under the normality assumptions

$$\frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} \sim t_{(n-2)} \quad \text{and} \quad \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \sim t_{(n-2)}$$

An important test is that $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ because, if $\beta_1 = 0$ then $Y_i = \beta_0 + e_i$, that is, the dependent variable isn't really dependent on the prediction variable, but random noise about the horizontal line $y = \beta_0$.

In general, with help of the test statistic $T = \frac{\hat{\beta}_1 - \beta_1^0}{s_{\hat{\beta}_1}}$ we can test the null hypothesis $H_0: \beta_1 = \beta_1^0$ at α significance level as follows

H_1	Reject H_0 if
$\beta_1 \neq \beta_1^0$	$\lvert T_{obs} \rvert > t_{(n-2)}(1-\frac{\alpha}{2})$
$\beta_1 > \beta_1^0$	$T_{obs} > t_{(n-2)}(1-\alpha)$
$\beta_1 < \beta_1^0$	$T_{obs} < t_{(n-2)}(\alpha)$

A two-sided confidence interval at $100(1-\alpha)\%$ confidence level for $\hat{\beta}_1$ is given by $[\hat{\beta}_1 - t_{(n-2)}(1-\alpha/2)\, S_{\hat{\beta}_1},\ \hat{\beta}_1 \pm t_{(n-2)}(1-\frac{\alpha}{2})\, S_{\hat{\beta}_1}]$

Example. In our previous example of 4 bulbs tested for lifetimes against operating voltage we got: $\sum x_i = 450 \quad \sum y_i = 4670$

$$\sum x^2 = 50750 \quad \sum x_i y_i = 521800$$

Test $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

First we compute $s^2 = \frac{RSS}{n-2} = \frac{1}{2}\sum_{i=1}^{4}(y_i - 4385 + 28.6x_i)^2 = 1215$

$$S_{\hat{\beta}_1}^2 = \frac{ns^2}{n\sum_{i=1}^{4} x_i^2 - \left(\sum_{i=1}^{4} x_i\right)^2} = \frac{4 \cdot 1215}{4\ 50750 - (450)^2} = 9.72$$

$$T_{obs} = \frac{\hat{\beta}_1 - \beta_1^0}{S_{\hat{\beta}_1}} = \frac{-28.6}{\sqrt{9.72}} = -9.1735 \implies \lvert T_{obs} \rvert = 9.1735$$

At $\alpha = 0.05$ $t_{(n-2)}(1-\frac{\alpha}{2}) = t_{(2)}(0.975) = 4.303$

as $\lvert T_{obs} \rvert > 4.303$ we reject H_0. A 95% confidence interval is given by

$$[\hat{\beta}_1 - t_{(n-2)}(1-\tfrac{\alpha}{2})\, S_{\hat{\beta}_1},\ \hat{\beta}_1 + t_{(n-2)}(1-\tfrac{\alpha}{2})\, S_{\hat{\beta}_1}] = [-28.6 - 4.303\sqrt{9.72},\ -28.6 + 4.303\sqrt{9.72}]$$

$= [-42.02, -15.19]$

which confirms that $H_0: \beta_1 = 0$ must be rejected as 0 is not contained in the confidence interval.