In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \; i = 1, 2, ..., n$$
$$E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2 \quad and \quad Cov(\varepsilon_i, \varepsilon_j) = 0 \; for \; all \; i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad and \qquad Var(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$
$$S_{yy} = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$
$$S_{xy} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1\bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^{n} K_i Y_i, \quad K_i = \frac{\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

$$\hat{\beta}_0 = \sum_{i=1}^{n} L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$Var(\widehat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad S.E(\widehat{\beta}_1) = \sqrt{Var(\widehat{\beta}_1)}$$

and

$$Var(\widehat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right], \quad S.E(\widehat{\beta}_0) = \sqrt{Var(\widehat{\beta}_0)}$$

The unbiased estimate of σ^2 is

$$s^2 = MSE = \widehat{\sigma^2} = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^{n} e_i^2, \quad e_i = Y_i - \widehat{Y}_i, \quad i = 1, 2, \ldots$$

3- Sampling distribution

The normal error regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where:

Y_i is the observed response in the ith trial
X_i is a known constant, the level of the predictor variable in the ith trial
β_0 and β_1 are parameters
ε_i are independent $N(0, \sigma^2)$
$i = 1, \ldots, n$

3-1 Sampling distribution of $\frac{\widehat{\beta}_1-\beta_1}{S.E(\widehat{\beta}_1)}=\frac{b_1-\beta_1}{S.E(b_1)}$ and $\frac{\widehat{\beta}_0-\beta_0}{S.E(\widehat{\beta}_0)}=\frac{b_0-\beta_0}{S.E(b_0)}$

Lemma

Let be b_1 is the estimator of the slop of the simple linear regression model, then

$\frac{\widehat{\beta}_1-\beta_1}{S.E(\widehat{\beta}_1)}=\frac{b_1-\beta_1}{S.E(b_1)}$ has t distribution with (n-2) degrees of freedom.

Similarly,

Let be b_0 is the estimator of the intercept of the simple linear regression model, then

$\frac{\widehat{\beta}_0-\beta_0}{S.E(\widehat{\beta}_0)}=\frac{b_0-\beta_0}{S.E(b_0)}$ has t distribution with (n-2) degrees of freedom.

Interval Estimation

This distribution of $\frac{\widehat{\beta}_1-\beta_1}{S.E(\widehat{\beta}_1)}=\frac{b_1-\beta_1}{S.E(b_1)}$ can be used to construct $100(1-\alpha)\%$ confidence interval for β_1 as follows

$$\widehat{\beta_1} \pm t_{1-\alpha/2,n-2} S.E(\widehat{\beta_1})$$

Similarly, the distribution of $\frac{\widehat{\beta}_0-\beta_0}{S.E(\widehat{\beta}_0)}=\frac{b_0-\beta_0}{S.E(b_0)}$ can be used to construct $100(1-\alpha)\%$ confidence interval for β_1 as follows

$$\widehat{\beta_0} \pm t_{1-\alpha/2,n-2} S.E(\widehat{\beta_0})$$

Example

Consider the Toluca Company example, find 95% confidence intervals for both of β_1 and β_0.

For such data, we have calculated

$$Var(\hat{\beta}_1) = Var(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} = \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{2384}{19800} = .12040$$

Hence the squared error of $\hat{\beta}_1$ is

$$S.E(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{.12040} = .3470$$

Similarly,

$$S.E(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \sqrt{685.34} = 26.18$$

$t_{1-\alpha/2, n-2} = t_{0.975,23} =$ 2.068658=2.069

and

$\hat{\beta}_1 = 3.5702$, $\hat{\beta}_0 = 62.4$

Then 95% confidence interval for β_1 as follows

$$\widehat{\beta}_1 \pm t_{1-\alpha/2,n-2} S.E(\widehat{\beta}_1)$$

$$3.5702 \pm 2.069(0.3470)$$

Hence

$$2.85 < \beta_1 < 4.29$$

Similarly, then 95% confidence interval for β_0 as follows

$$\widehat{\beta}_0 \pm t_{1-\alpha/2,n-2} S.E(\widehat{\beta}_0)$$

$$62.4 \pm 2.069(26.18)$$

Hence

$$8.25 < \beta_0 < 116.57$$

This can be done easily using R as:

```
confint(model,level=0.95) #CIs for all parameters
```

Hypotheses Testing

This distribution of $\frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} = \frac{b_1 - \beta_1}{S.E(b_1)}$ and $\frac{\hat{\beta}_0 - \beta_0}{S.E(\hat{\beta}_0)} = \frac{b_0 - \beta_0}{S.E(b_0)}$ can be used to test some hypotheses concerning the coefficients of the simple linear regression model as follows:

Steps for testing β_1

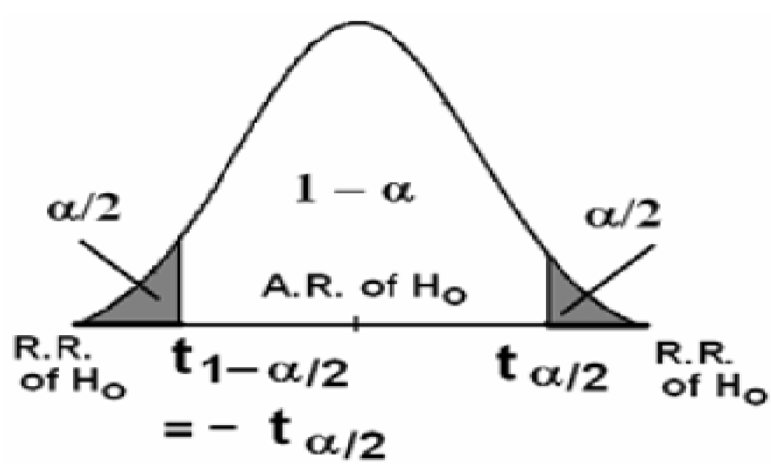
i) Setup the hypotheses

$$H_0 : \beta_1 = \beta_1^{(0)} \qquad vs \qquad H_1 : \beta_1 \neq \beta_1^{(0)}$$

ii) Test statistic under H0

$T_0 = \frac{b_1 - \beta_1^{(0)}}{S.E(b_1)}$ this statistic has t distribution with (n-2) d.f

iii) Critical regions



R.R: Rejection Region and A.R: Acceptance Region

iv) Decision

When the calculated T_0 belongs to the shaded areas, we reject the null hypothesis H0, otherwise Accept H0.

Steps for testing β_0

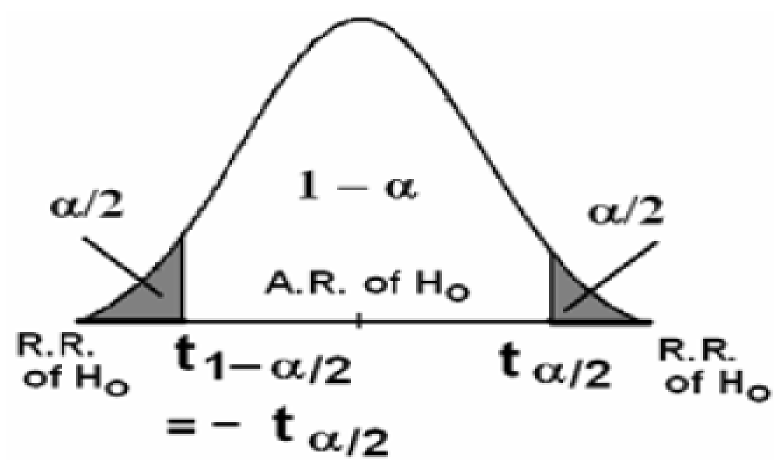
i) Setup the hypotheses

$$H_0 : \beta_0 = \beta_0^{(0)} \qquad vs \qquad H_1 : \beta_0 \neq \beta_0^{(0)}$$

ii) Test statistic under H0

$$T_0 = \frac{b_0 - \beta_0^{(0)}}{S.E(b_0)}$$ this statistic has t distribution with (n-2) d.f

iii) Critical regions



R.R: Rejection Region and A.R: Acceptance Region

iv) Decision

When the calculated T_0 belongs to the shaded areas, we reject the null hypothesis H0, otherwise Accept H0.

Remarks:

(1) In some applications, we may need to test

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

and

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$$

In these cases, you need to replace both of $\beta_1^{(0)}$ and $\beta_0^{(0)}$ by zeros in the test steps. This equivalents testing the significance of the linear term (β_1) or the intercept term β_0 . The Minitab output are designed for these cases.

(2) In some applications, we may need to test

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 > (<)\ 0$$

and

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 > (<) 0$$

In these cases, you need to replace the critical regions to one-sided critical regions.

(3) One may use the two-sided p-value approach: $p-value = 2P(T_0 > t_{\alpha/2})$, then reject H0 when $p-value \le \alpha$, otherwise accept H0. The one-sided p-value is $p-value = P(T_0 > t_{\alpha})$, then reject H0 when $p-value \le \alpha$, otherwise accept H0.

Example:

Consider the Toluca Company example, test the hypotheses

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \ne 0$$

and

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \ne 0$$

Testing β_1

i) Setup the hypotheses

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \ne 0$$

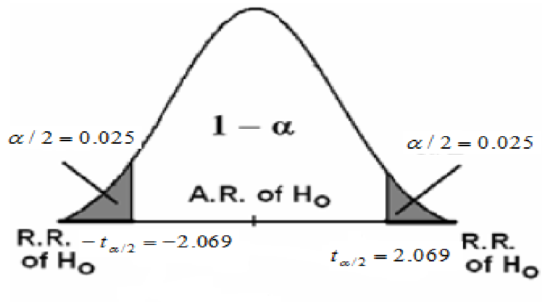
ii) Test statistics under H0

$$T_0 = \frac{b_1 - \beta_1^{(0)}}{S.E(b_1)} = \frac{3.5702 - 0}{0.3470} = 10.29$$

iii) Critical regions

$$\alpha = 0.05 --> \alpha/2 = 0.025$$

$$t_{\alpha/2} = 2.069 \quad and \ -t_{\alpha/2} = -2.069$$



R.R: Rejection Region and A.R: Acceptance Region

iv) Decision: The calculated $T_0 = 10.29$ belongs to the shaded areas, then we reject the null hypothesis H0

These calculations can be conducted using the R output as:

```
Coefficients:
            Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)  62.366    26.177     2.382   0.0259 *
x             3.570     0.347    10.290   4.45e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we can see from the results that, T=10.29. Also, the p-value=0.000<0.05, then reject H0.

Testing β_0

i) Setup the hypotheses

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$$

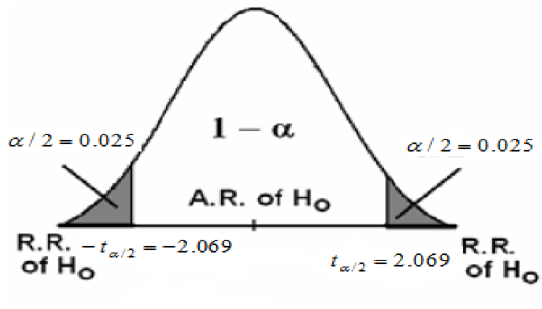
ii) Test statistics under H0

$$T_0 = \frac{b_0 - \beta_0^{(0)}}{S.E(b_1)} = \frac{62.4 - 0}{26.18} = 2.38$$

iii) Critical regions

$$\alpha = 0.05 -- > \alpha / 2 = 0.025$$

$$t_{\alpha/2} = 2.069 \quad and \ -t_{\alpha/2} = -2.069$$



R.R: Rejection Region and A.R: Acceptance Region

iv) Decision: The calculated $T_0 = 2.38$ belongs to the shaded areas, then we reject the null hypothesis H0.

As we can see from the results that, T=10.29. Also, the p-value=0.026<0.05, then reject H0.