Return

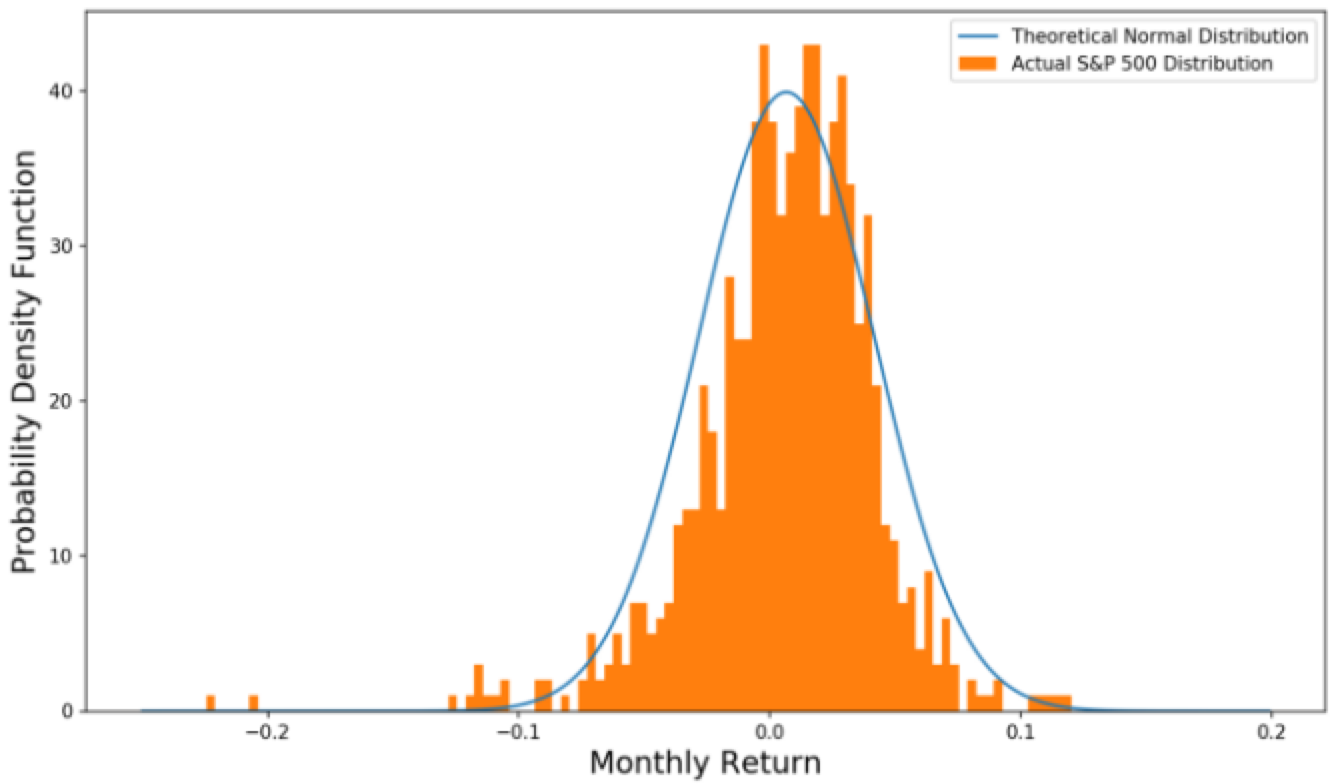
$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

R_t is called the *simple return* of the asset with price series P_t. However, in statistical analysis of financial data, we usually consider *log returns* r_t, which are defined as:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1}) = \log(1 + R_t)$$

Simple and log returns are not one and the same, although for small relative changes in the price series P_t, they do not differ much. For example, if $R_t = 0.00\%$, then $r_t = 0.00\%$, if $R_t = 1.00\%$, then $r_t = 0.995\%$ and if $R_t = 5.00\%$, then $R_t = 4.88\%$.

Normality in Financial Variables (A very strong Assumption)



Actual distribution vs. normal distribution

Deviation from Normalcy

In basic statistics and probability theory, we almost exclusively deal with the first and second central moment of a random variable, namely expectation and variance. The definitions are as follows:

k^{th} moment of X : $m_k = E[X^k]$, e.g. expectation $\mu = E[X]$

k^{th} central moment of X : $\mu_k = E[(X-\mu)^k]$, e.g. variance $Var(X) = E[(X-\mu)^2]$

In the statistical analysis of financial data, or better, in risk management, one is often also interested in the third and fourth central moments, which are the basis for skewness and kurtosis.

Skewness

The third central moment tells us how symmetrical a distribution gathers around its mean. Rather than working with the third central moment directly, it is, by convention, standardized. The definition of *skewness* is as follows:

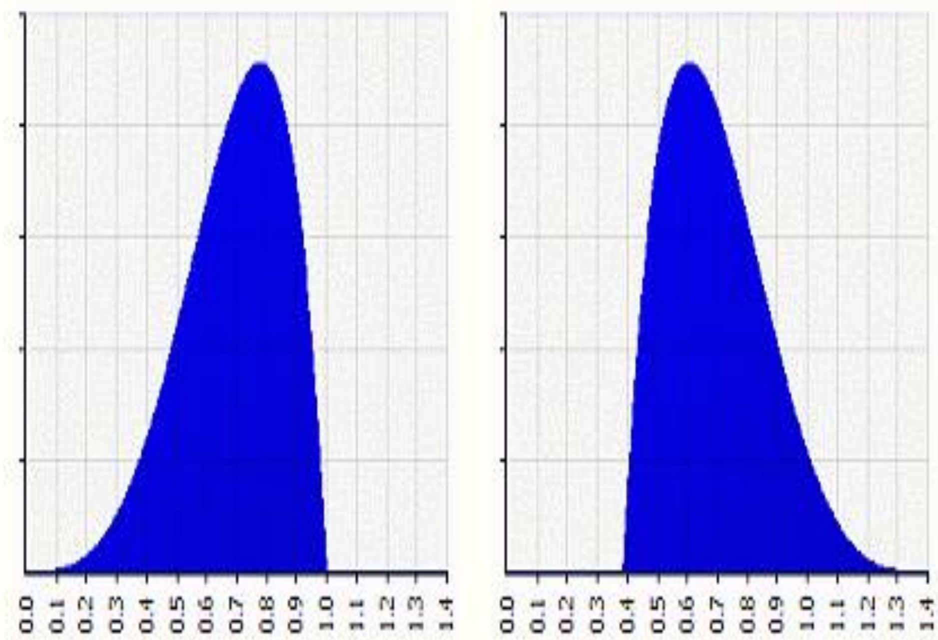
$$Skew = \frac{E[(X-\mu)^3]}{\sigma^3}.$$

Any random variable with a symmetric distribution will have $Skew = 0$. Values greater than zero indicate positive skewness, i.e. distributions that have a heavy tail on the right hand side. Conversely, $Skew < 0$ indicates a left-skewed distribution.

Let us consider a situation where two investments' return distributions have identical mean and variance, but different skewness parameters. Which one is to prefer? Typically, risk managers are wary of negative skew: in that situation, small gains are the norm, but big losses can occur, carrying the risk of going bankrupt. The *sample skewness* is usually estimated as follows:

$$\hat{S}kew = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\hat{\sigma}}\right)^3$$

Skewness


0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4

Kurtosis

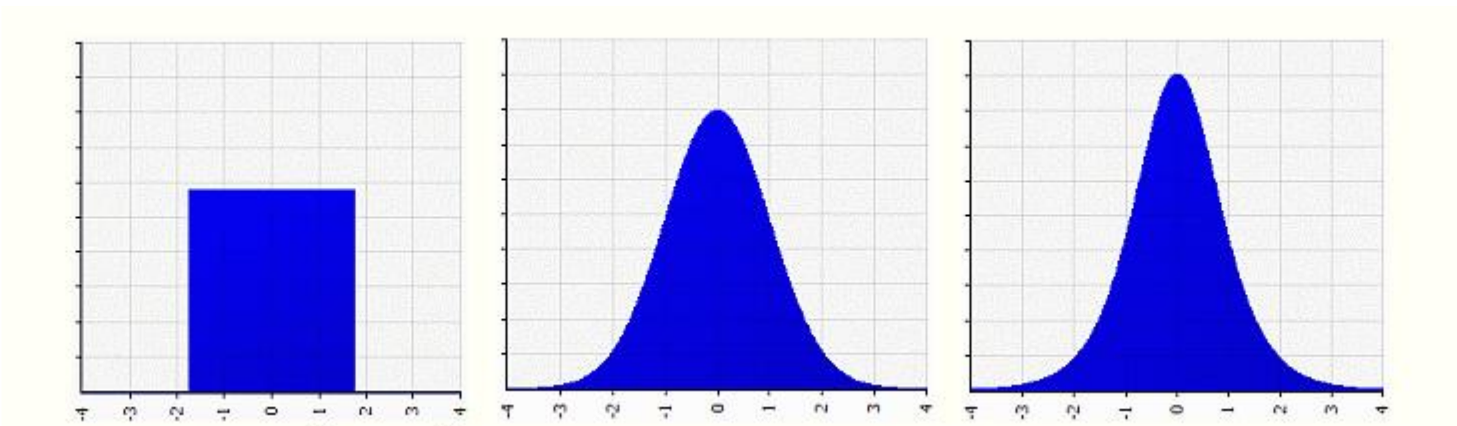
The *kurtosis* is the standardized fourth central moment. Similar to the variance it measures how spread out a distribution is, but it puts more weight on the tails. The exact definition is:

$$Kurt = \frac{E[(X-\mu)^4]}{\sigma^4}$$

financial analysis, an asset with leptokurtic log returns needs to be taken seriously. It means that big losses (as well as big gains) can occur, and one should be prepared for it. Estimation of the kurtosis happens by:

$$\hat{Kurt} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i-\overline{x})}{\sigma}\right)^4$$

Kurtosis



Hypothesis Testing

- One ‘ordinary” sample---> Inference about population

- One ‘properly chosen’ sample is a good representative--→ Inference about population

Null and Alternative Hypothesis

- A null hypothesis is a statement of the status quo or no effect. (H_0)
- An alternative hypothesis is one in which some difference or effect is expected. H_1

Concept of ‘p-value’

- *P*-value answer the question: What is the probability of the observed test statistic … **when H_0 *is true*?**
- Thus, smaller and smaller *P*-values provide stronger and stronger evidence against H_0
- Small *P*-value $\Rightarrow$ strong evidence

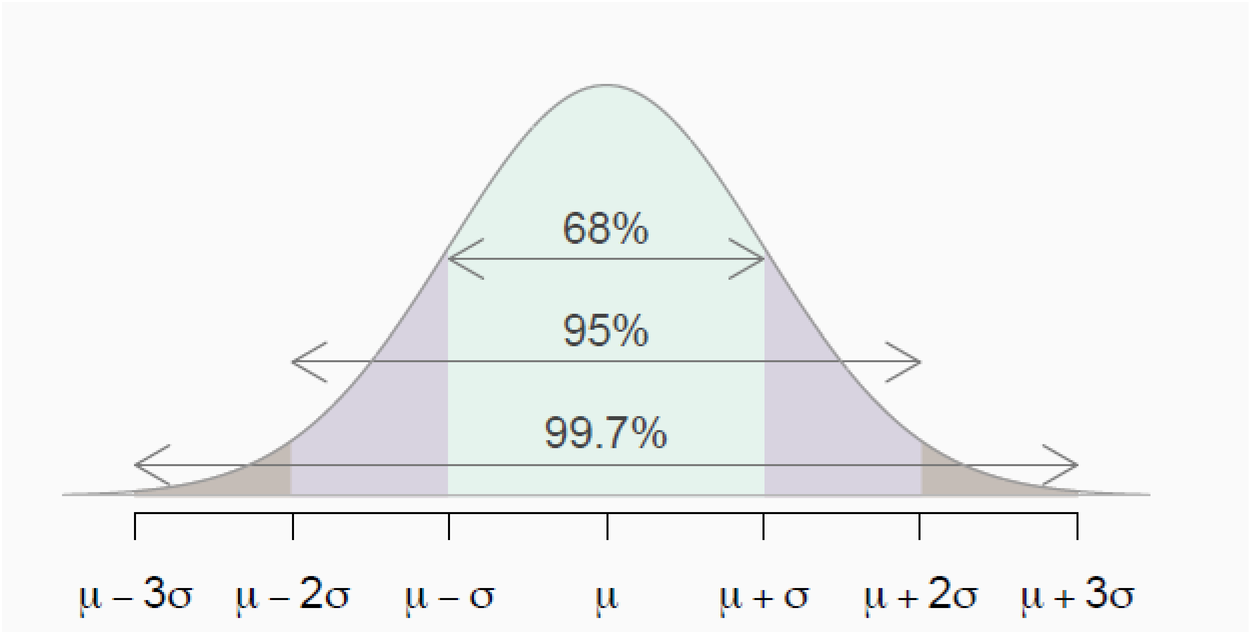
Examples

$P = .27 \Rightarrow$ non-significant evidence against H_0

$P = .01 \Rightarrow$ highly significant evidence against H_0

Benchmark value= 0.05 (5% level of significance)

Normal Distribution



Normal Distribution with mean μ and standard deviation σ

Z-statistic

$$Z = \frac{X - \mu}{\sigma}$$

Type of Errors in estimation

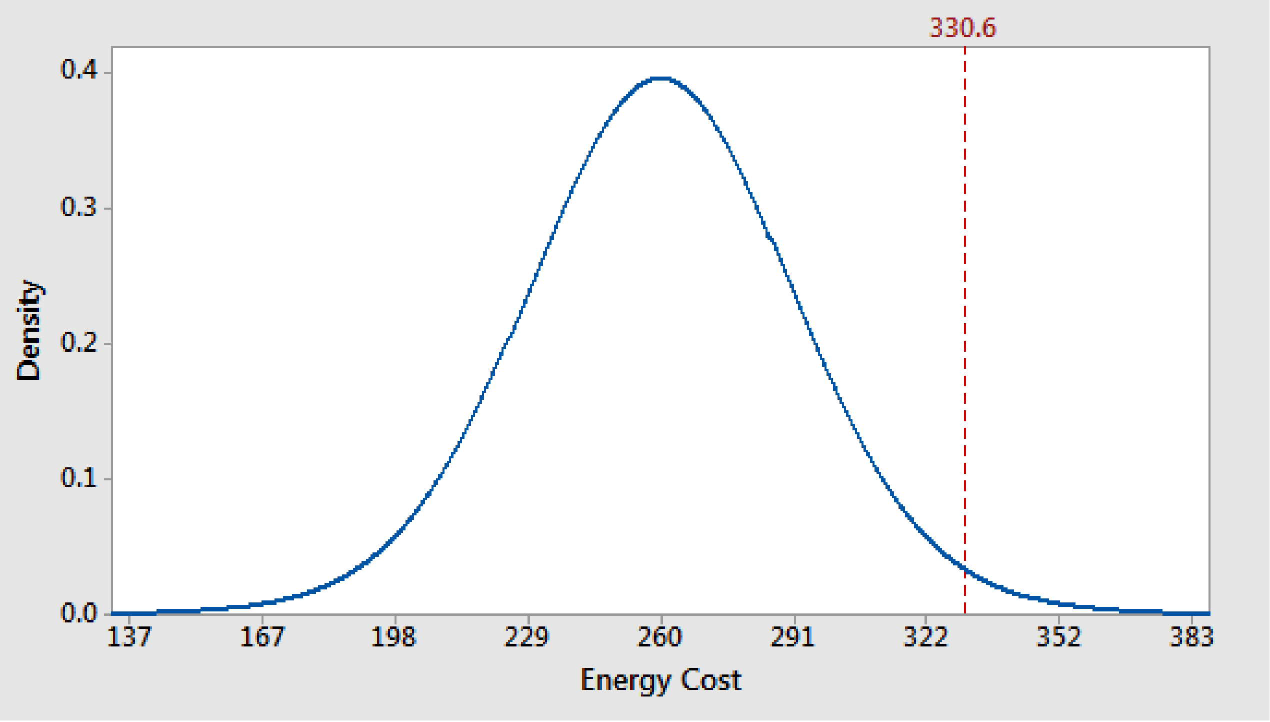
- Type-I error (You reject a true null hypothesis)

You punish a person who is not guilty!

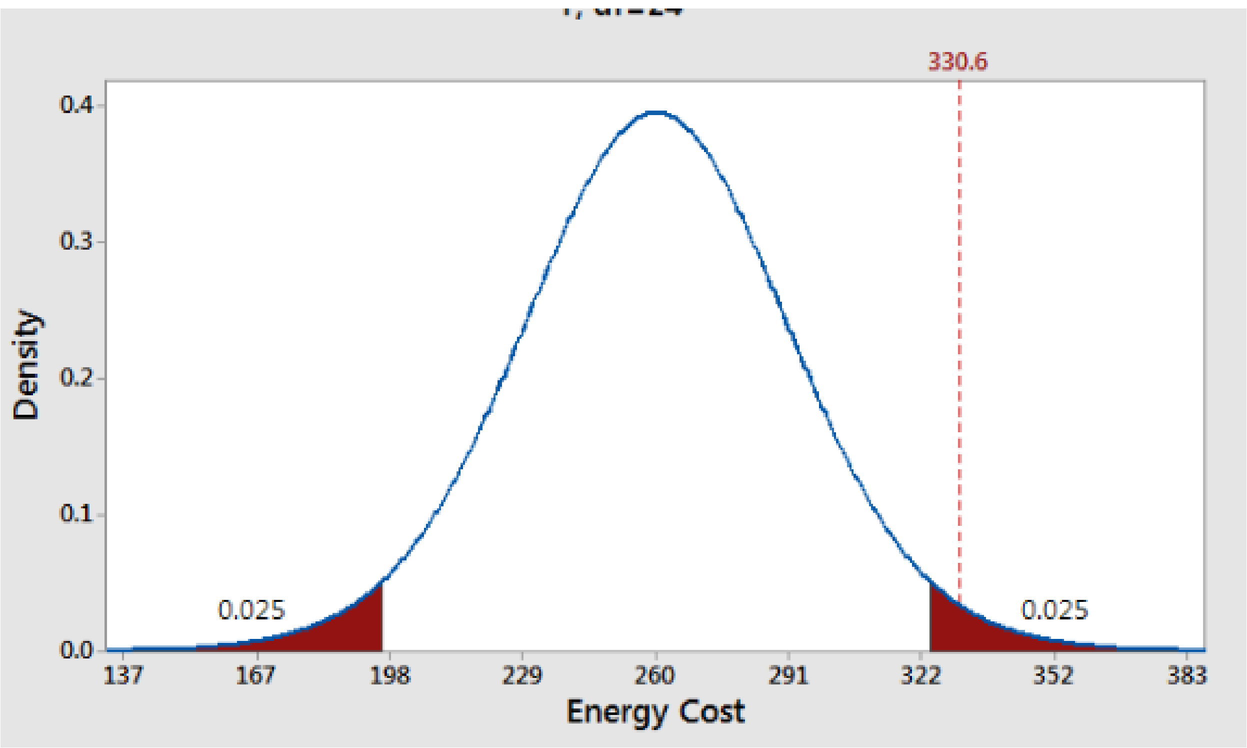
- Type-II error (You accept a false null hypothesis)

You release a truly guilty person!

Level of Significance in a test



Two tail test



t-test statistic

$$t = \frac{m - \mu}{s/\sqrt{n}}$$

t = Student's t-test

m = mean

μ = theoretical value

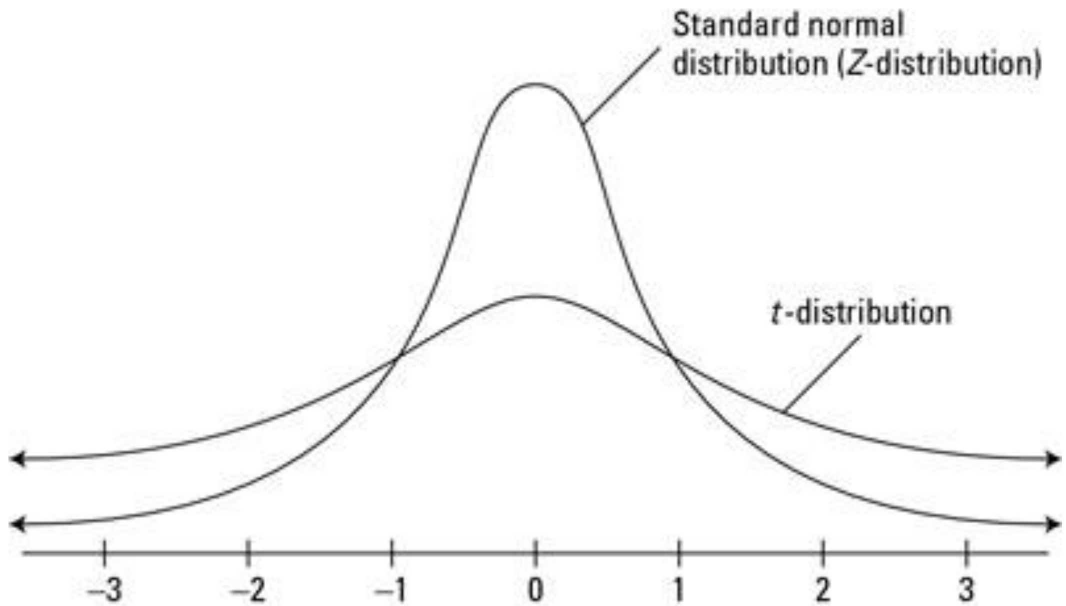
s = standard deviation

n = variable set size

2-t rule of thumb

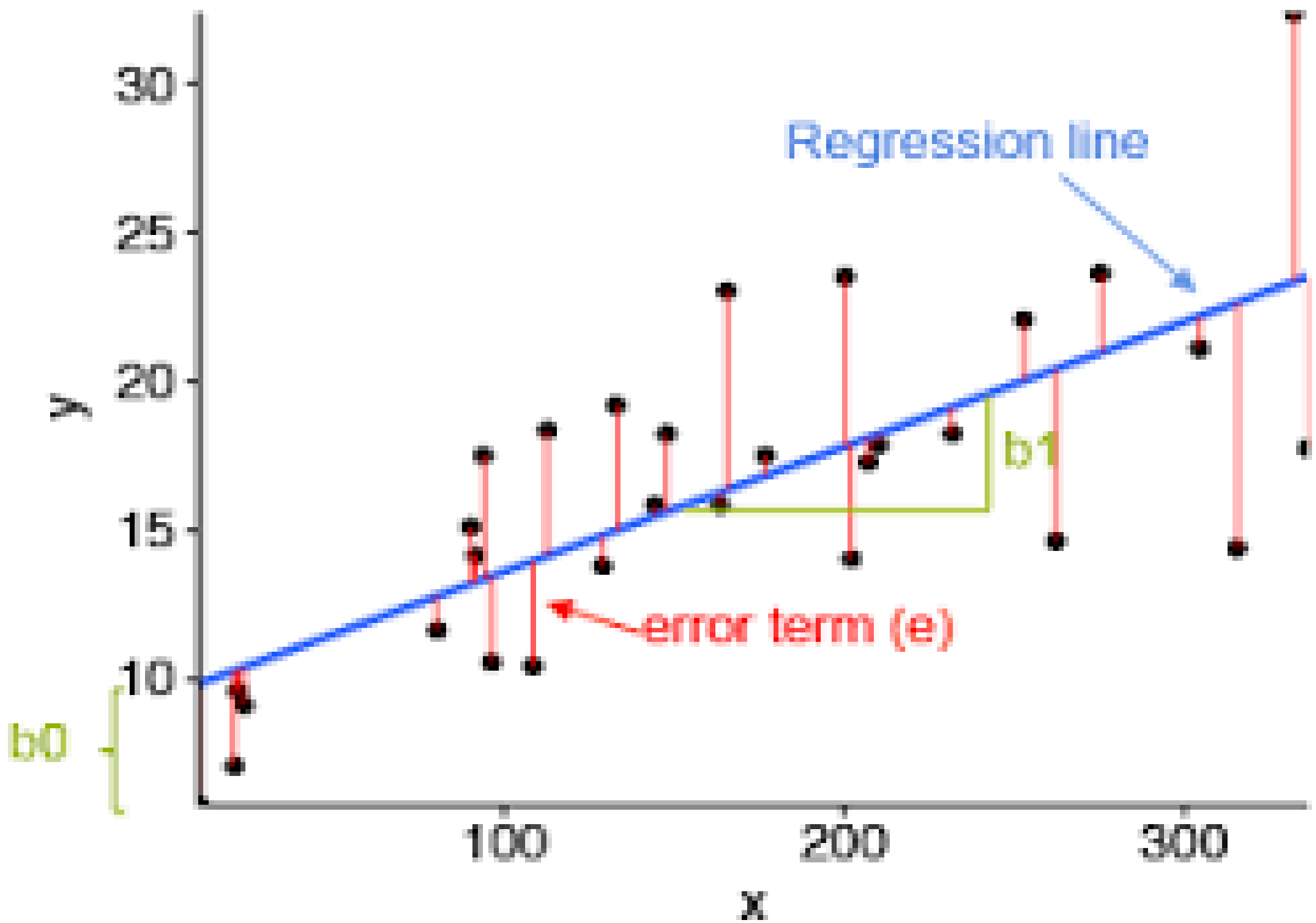
- If the sample size is more than or equal to 20 ($n>=20$) and level of significance is 5% then an absolute value of 2 ($|t|>=2$) will suffice to reject the null hypothesis.

t-distribution



Notion of linear model (Regression)

$$y = b0 + b1 * x$$



Variance and covariance

$$\sigma^2 = \frac{\sum(\chi - \mu)^2}{N}$$

$$\mathrm{Cov}\,(X, Y) = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_j - \overline{Y}\right)}{n}$$

Derivation

$$Q=\sum_{i=1}^{n}\left(Y_i-\hat{Y}\right)^2=\sum_{i=1}^{n}\left(Y_i-a-bX_i\right)^2$$

$$\frac{\partial Q}{\partial a}=\sum_{i=1}^{n}-2\left(Y_i-a-bX_i\right)=2\left(na+b\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}Y_i\right)=0$$

$$a=\overline{Y}-b\overline{X}$$

$$\frac{\partial Q}{\partial b}=\sum_{i=1}^{n}-2X_i\left(Y_i-a-bX_i\right)=\sum_{i=1}^{n}-2\left(X_iY_i-aX_i-bX_i^2\right)=0$$

$$\sum_{i=1}^{n}\left(X_iY_i-X_i\overline{Y}+bX_i\overline{X}-bX_i^2\right)=0$$

Cont..

$$b=\frac{\sum_{i=1}^{n}\left(X_iY_i-X_i\overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i^2-X_i\overline{X}\right)}=\frac{\sum_{i=1}^{n}\left(X_iY_i\right)-n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n}\left(X_i^2\right)-n\overline{X}^2}$$

$$b=\frac{\sum_{i=1}^{n}\left(X_iY_i-X_i\overline{Y}\right)+\sum_{i=1}^{n}\left(\overline{X}\,\overline{Y}-Y_i\overline{X}\right)}{\sum_{i=1}^{n}\left(X_i^2-X_i\overline{X}\right)+\sum_{i=1}^{n}\left(\overline{X}^2-X_i\overline{X}\right)}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}=\frac{Cov(X,Y)}{Var(X)}$$