Lecture 14
MATH-57091 Probability and Statistics for High-School Teachers.

Artem Zvavitch

Department of Mathematical Sciences, Kent State University

December 5, 2016.

Error Random Variable

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate

Error Random Variable

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0.

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e:

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2.

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: **e is a normal random variable with expectation 0 and variance σ^2.** Another way to make this assumption is that

$$\mathbb{E}Y = \alpha + \beta x \text{ (i.e. } \mathbb{E}e = 0)$$

and

$$\mathrm{Var}(Y) = \sigma^2 \text{ (i.e. } \mathrm{Var}(e) = \sigma^2).$$

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2. Another way to make this assumption is that

$$\mathbb{E}Y = \alpha + \beta x \text{ (i.e. } \mathbb{E}e = 0)$$

and

$$\mathrm{Var}(Y) = \sigma^2 \text{ (i.e. } \mathrm{Var}(e) = \sigma^2).$$

Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have.

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e,$$

where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2. Another way to make this assumption is that

$$\mathbb{E}Y = \alpha + \beta x \text{ (i.e. } \mathbb{E}e = 0)$$

and

$$\mathrm{Var}(Y) = \sigma^2 \text{ (i.e. } \mathrm{Var}(e) = \sigma^2).$$

Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have. We will also assume that the response we get about x is independent one from another, i.e. x_1 is independent of x_2, is independent of x_3....

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e.$$

Where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2.
Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have.

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e.$$

Where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2.
Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have. So HOW can we do it?

Error Random Variable

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e.$$

Where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2.
Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have. So HOW can we do it? Consider our set of values (x_i, Y_i), note that for each fixed i

$$\frac{Y_i - \mathbb{E}Y_i}{\sqrt{\mathrm{Var}(Y_i)}} = \frac{Y_i - (\alpha + \beta x_i)}{\sigma}$$

is $N(0,1)$ (standard normal distribution).

Error Random Variable

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e.$$

Where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: e **is a normal random variable with expectation** 0 **and variance** σ^2.

Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have. So HOW can we do it? Consider our set of values (x_i, Y_i), note that for each fixed i

$$\frac{Y_i - \mathbb{E}Y_i}{\sqrt{\text{Var}(Y_i)}} = \frac{Y_i - (\alpha + \beta x_i)}{\sigma}$$

is $N(0,1)$ (standard normal distribution). Moreover, those random variables are independent, (because of our assumption on x_i)

In the previous class we have defined the linear regression model by the relationship

$$Y = \alpha + \beta x + e.$$

Where α and β are unknown parameters, that we must estimate and e is a random error, with mean 0. To provide more statistical information about α and β and to use methods we developed before, we need to have additional assumption on e: **e is a normal random variable with expectation 0 and variance σ^2.**

Note that the quantity σ^2 is unknown and we will need to estimate it from the data we have. So HOW can we do it? Consider our set of values (x_i, Y_i), note that for each fixed i

$$\frac{Y_i - \mathbb{E}Y_i}{\sqrt{\mathrm{Var}(Y_i)}} = \frac{Y_i - (\alpha + \beta x_i)}{\sigma}$$

is $N(0,1)$ (standard normal distribution). Moreover, those random variables are independent, (because of our assumption on x_i) thus

$$\sum_{i=1}^{n} \frac{(Y_i - (\alpha + \beta x_i))^2}{\sigma^2}$$

is a chi-squared random variable (χ^2) with n degrees for freedom (See Lecture 10.1).

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

Error Random Variable

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

The quantities $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$ are called **residuals**. They represent the difference between the actual and the predicted responses.

Error Random Variable

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

The quantities $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$ are called **residuals**. They represent the difference between the actual and the predicted responses. Define

$$SS_R = \sum_{i=1}^{n} (Y_i - (\alpha + \beta x_i))^2.$$

Error Random Variable

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

The quantities $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$ are called **residuals**. They represent the difference between the actual and the predicted responses. Define

$$SS_R = \sum_{i=1}^{n} (Y_i - (\alpha + \beta x_i))^2.$$

Thus SS_R/σ^2 is a χ^2 with $n-2$ degrees of freedom and (please, check Lecture 10.1 to refresh info about χ^2 distribution):

$$\frac{\mathbb{E}(SS_R)}{\sigma^2} = n-2 \text{ and } \mathbb{E}\left(\frac{SS_R}{n-2}\right) = \sigma^2$$

and we can use $SS_R/(n-2)$ to estimate σ^2.

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

The quantities $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$ are called **residuals**. They represent the difference between the actual and the predicted responses. Define

$$SS_R = \sum_{i=1}^{n} (Y_i - (\alpha + \beta x_i))^2.$$

Thus SS_R/σ^2 is a χ^2 with $n-2$ degrees of freedom and (please, check Lecture 10.1 to refresh info about χ^2 distribution):

$$\frac{\mathbb{E}(SS_R)}{\sigma^2} = n-2 \text{ and } \mathbb{E}\left(\frac{SS_R}{n-2}\right) = \sigma^2$$

and we can use $SS_R/(n-2)$ to estimate σ^2. Note that in most cases $SS_R/(n-2)$ is computed with a help of special program / computer.

Next, when we substitute the estimators $\hat{\alpha}$ and $\hat{\beta}$ instead of α and β, we will "loose" two degrees of freedom and the new random variable will be again χ^2 random variable with $n-2$ degrees of freedom:

$$\sum_{i=1}^{n} \frac{(Y_i - (\hat{\alpha} + \hat{\beta} x_i))^2}{\sigma^2}$$

The quantities $Y_i - (\hat{\alpha} + \hat{\beta} x_i)$ are called **residuals**. They represent the difference between the actual and the predicted responses. Define

$$SS_R = \sum_{i=1}^{n} (Y_i - (\alpha + \beta x_i))^2.$$

Thus SS_R/σ^2 is a χ^2 with $n-2$ degrees of freedom and (please, check Lecture 10.1 to refresh info about χ^2 distribution):

$$\frac{\mathbb{E}(SS_R)}{\sigma^2} = n-2 \text{ and } \mathbb{E}\left(\frac{SS_R}{n-2}\right) = \sigma^2$$

and we can use $SS_R/(n-2)$ to estimate σ^2. Note that in most cases $SS_R/(n-2)$ is computed with a help of special program / computer. Still, if we wish to do it by hands, one (just direct substitution, of definition for $\hat{\alpha}$ and $\hat{\beta}$) can use the following formula

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}},$$

where

$$S_{xY} = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}), \; S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ and } S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

Very simple example (check lecture 13.2)

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

Then $\bar{x} = 7/3$ and $\bar{Y} = 2$,

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

Then $\bar{x} = 7/3$ and $\bar{Y} = 2$, moreover,

$$S_{xY} = (1 - 7/3)(2 - 2) + (2 - 7/3)(3 - 2) + (4 - 7/3)(1 - 2) = -2$$

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

Then $\bar{x} = 7/3$ and $\bar{Y} = 2$, moreover,

$$S_{xY} = (1 - 7/3)(2 - 2) + (2 - 7/3)(3 - 2) + (4 - 7/3)(1 - 2) = -2$$

$$S_{xx} = (1 - 7/3)^2 + (2 - 7/3)^2 + (4 - 7/3)^2 \approx 4.7$$

Very simple example (check lecture 13.2)

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

Then $\bar{x} = 7/3$ and $\bar{Y} = 2$, moreover,

$$S_{xY} = (1 - 7/3)(2 - 2) + (2 - 7/3)(3 - 2) + (4 - 7/3)(1 - 2) = -2$$

$$S_{xx} = (1 - 7/3)^2 + (2 - 7/3)^2 + (4 - 7/3)^2 \approx 4.7$$

and

$$S_{YY} = (2 - 2)^2 + (2 - 3)^2 + (2 - 1)^2 = 2$$

An area manager in a department store wants to study the relationship between the number of workers on duty at a certain department and the value (in hundreds) of merchandise lost to shoplifters:

Number of Workers	Loss
1	2
2	3
4	1

Then $\bar{x} = 7/3$ and $\bar{Y} = 2$, moreover,

$$S_{xY} = (1-7/3)(2-2) + (2-7/3)(3-2) + (4-7/3)(1-2) = -2$$

$$S_{xx} = (1-7/3)^2 + (2-7/3)^2 + (4-7/3)^2 \approx 4.7$$

and

$$S_{YY} = (2-2)^2 + (2-3)^2 + (2-1)^2 = 2$$

and thus

$$SS_R \approx \frac{4.7 * 2 - (-2)^2}{4.7} \approx 1.15$$

and $\delta^2 = 1.15/(3-2) = 1.15$.

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$,

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent.

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise).

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed.

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed. This gives (using the properties of normal distribution, and summation of normally distributed independent random variables) that $\hat{\beta}$ is normally distributed with

$$\mathbb{E}\hat{\beta} = \beta \text{ and } \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}.$$

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed. This gives (using the properties of normal distribution, and summation of normally distributed independent random variables) that $\hat{\beta}$ is normally distributed with

$$\mathbb{E}\hat{\beta} = \beta \text{ and } \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}.$$

Thus we can consider standardized variable

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} = \sqrt{\frac{S_{xx}}{\sigma^2}}(\hat{\beta} - \beta),$$

has a standard normal distribution.

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed. This gives (using the properties of normal distribution, and summation of normally distributed independent random variables) that $\hat{\beta}$ is normally distributed with

$$\mathbb{E}\hat{\beta} = \beta \text{ and } \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}.$$

Thus we can consider standardized variable

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} = \sqrt{\frac{S_{xx}}{\sigma^2}}(\hat{\beta} - \beta),$$

has a standard normal distribution. Unfortunately, life is not so easy — σ^2 is an unknown parameter. Which we need to estimate.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed. This gives (using the properties of normal distribution, and summation of normally distributed independent random variables) that $\hat{\beta}$ is normally distributed with

$$\mathbb{E}\hat{\beta} = \beta \text{ and } \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}.$$

Thus we can consider standardized variable

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} = \sqrt{\frac{S_{xx}}{\sigma^2}}(\hat{\beta} - \beta),$$

has a standard normal distribution. Unfortunately, life is not so easy — σ^2 is an unknown parameter. Which we need to estimate. We need to replace it with $SS_R/(n-2)$, which is χ^2 - random variable with $(n-2)$ degrees of freedom.

Testing the Hypothesis that $\beta = 0$.

When we consider a simple linear regression model

$$Y = \alpha + \beta x + e.$$

it is really important to understand how to test the hypothesis that $\beta = 0$, notice that in this case we would get $Y = \alpha + e$ or that Y has more or less no relation to x, i.e. Y and x are independent. So we would like to test

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Our goal is to study "what we can study", i.e. $\hat{\beta}$ - the estimator of β and to reject H_0 when $\hat{\beta}$ is far from zero (and not to reject it otherwise). Reminder: we have assumed before that e is normally distributed. This gives (using the properties of normal distribution, and summation of normally distributed independent random variables) that $\hat{\beta}$ is normally distributed with

$$\mathbb{E}\hat{\beta} = \beta \text{ and } \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}.$$

Thus we can consider standardized variable

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} = \sqrt{\frac{S_{xx}}{\sigma^2}}(\hat{\beta} - \beta),$$

has a standard normal distribution. Unfortunately, life is not so easy — σ^2 is an unknown parameter. Which we need to estimate. We need to replace it with $SS_R/(n-2)$, which is χ^2 -random variable with $(n-2)$ degrees of freedom. It turns out that the resulting quantity

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}}(\hat{\beta} - \beta)$$

has a t distribution (again, please, check lecture 10.1) with $(n-2)$ degrees of freedom (denoted by T_{n-2}).

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}}(\hat{\beta} - \beta)$$

has a t distribution with $(n-2)$ degrees of freedom.

Testing the Hypothesis that $\beta = 0$.

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}}(\hat{\beta} - \beta)$$

has a t distribution with $(n-2)$ degrees of freedom.
Notice that we would like to check that $\beta = 0$ substituting this in to the above formula we get

A significance-level-γ test of H_0 is to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |TS| \geq t_{n-2,\gamma/2}, \\ \text{Not Reject } H_0 & \text{otherwise} \end{array}$$

Where

$$TS = \sqrt{\frac{(n-2)S_{xx}}{SS_R}}\hat{\beta}.$$

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}}(\hat{\beta} - \beta)$$

has a t distribution with $(n-2)$ degrees of freedom.
Notice that we would like to check that $\beta = 0$ substituting this in to the above formula we get

A significance-level-γ test of H_0 is to

$$\begin{array}{cc} \text{Reject } H_0 & \text{if } |TS| \geq t_{n-2,\gamma/2}, \\ \text{Not Reject } H_0 & \text{otherwise} \end{array}$$

Where

$$TS = \sqrt{\frac{(n-2)S_{xx}}{SS_R}}\hat{\beta}.$$

Note, that, as in many cases before, a possible algorithm to perform the test is to first compute the value of Test Statistics (TS).

Testing the Hypothesis that $\beta = 0$.

$$\sqrt{\frac{(n-2)S_{xx}}{SS_R}}(\hat{\beta} - \beta)$$

has a t distribution with $(n-2)$ degrees of freedom.
Notice that we would like to check that $\beta = 0$ substituting this in to the above formula we get

A significance-level-γ test of H_0 is to

Reject H_0	if $\|TS\| \geq t_{n-2,\gamma/2}$,
Not Reject H_0	otherwise

Where

$$TS = \sqrt{\frac{(n-2)S_{xx}}{SS_R}}\hat{\beta}.$$

Note, that, as in many cases before, a possible algorithm to perform the test is to first compute the value of Test Statistics (TS). The H_0 should then be rejected if $\gamma \geq p$, where the p value is given by

$$p = P(|T_{n-2}| \geq |TS|) = 2P(T_{n-2} \geq |TS|).$$

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed.

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim,

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim, i.e. to check

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim, i.e. to check

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

The goal is to compute the value of test statistics. We first compute

$$S_{xx} = 700, \quad S_{YY} = 21.757 \quad \text{and } S_{xY} = -119.$$

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim, i.e. to check

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

The goal is to compute the value of test statistics. We first compute

$$S_{xx} = 700, \quad S_{YY} = 21.757 \quad \text{and } S_{xY} = -119.$$

Next

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}} = \frac{700 * 21.757 - (-119)^2}{700} = 1.527 \text{ and } \hat{\beta} = \frac{S_{xY}}{S_{xx}} = \frac{-119}{700} = -0.17$$

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim, i.e. to check

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

The goal is to compute the value of test statistics. We first compute

$$S_{xx} = 700, \quad S_{YY} = 21.757 \quad \text{and } S_{xY} = -119.$$

Next

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}} = \frac{700 * 21.757 - (-119)^2}{700} = 1.527 \text{ and } \hat{\beta} = \frac{S_{xY}}{S_{xx}} = \frac{-119}{700} = -0.17$$

Finally

$$TS = \sqrt{\frac{(n-2)S_{xx}}{SS_R}}\hat{\beta} = \sqrt{\frac{5 * 700}{1.527}}(-0.17) = -8.139$$

Example.

An individual claims that the fuel consumption of his automobile does not depend on how fast the car is driven, To test the plausibility of this hypothesis, the car was tested at various speeds between 45 and 75 miles per hour. With the following resulting data:

Speed	Miles per gallon
45	24.2
50	25.0
55	23.3
60	22.0
65	21.5
70	20.6
75	19.8

Do these data refute the claim that the milage per gallon of gas is unaffected by the speed at which the car is being driven?

Solution: We will model this question using simple linear regression $Y = \alpha + \beta x + e$, where Y will be miles per gallon of the car and x the speed. The claim made is that $\beta = 0$, and we would like to check if we can reject this claim, i.e. to check

$$H_0 : \beta = 0 \text{ against } H_1 : \beta \neq 0.$$

The goal is to compute the value of test statistics. We first compute

$$S_{xx} = 700, \quad S_{YY} = 21.757 \quad \text{and } S_{xY} = -119.$$

Next

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}} = \frac{700 * 21.757 - (-119)^2}{700} = 1.527 \text{ and } \hat{\beta} = \frac{S_{xY}}{S_{xx}} = \frac{-119}{700} = -0.17$$

Finally

$$TS = \sqrt{\frac{(n-2)S_{xx}}{SS_R}}\hat{\beta} = \sqrt{\frac{5 * 700}{1.527}}(-0.17) = -8.139$$

Using tables for t-distribution (see lecture 10.1 or google it) we find out that $t_{5,0.005} = 4.032$ we can reject the hypothesis that $\beta = 0$ at 1 percent level of significance. So we reject the claim. **Save money – do not drive to fast!!!**