

What is the definition of Regression Through the Origin and can you provide an example?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition of Regression Through the Origin and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165249>

Regression through the origin is a statistical method used to determine the relationship between two variables, where one variable is assumed to be a multiple of the other. This means that the regression line passes through the origin (0,0) on a graph, indicating that the intercept is equal to 0. This type of regression is also known as simple linear regression without an intercept.

For example, let's say we want to study the relationship between the number of hours spent studying and the grades obtained by students. If we assume that for every additional hour of studying, the grade increases by a fixed amount, then we can use regression through the origin to determine the exact relationship between these two variables. The regression line will pass through the origin, indicating that there is no fixed starting point for the relationship and that the grade is directly proportional to the number of hours spent studying. This can help us make predictions about a student's grade based on their study hours.

Regression Through the Origin: Definition & Example

Simple linear regression is a method that can be used to quantify the relationship between one or more predictor variables and a .

A simple linear regression model takes on the following form:

$$y = \beta_0 + \beta_1 x$$

where:

y: The value of the response variable
 β_0 : The value of the response variable when $x = 0$ (known as the "intercept" term)
 β_1 : The average increase in the response variable associated with a one unit increase in

xx: The value of the predictor variable

A modified version of this model is known as regression through the origin, which forces y to be equal to 0 when x is equal to 0.

This type of model takes on the following form:

$$y = \beta_1 x$$

Notice that the intercept term has been completely dropped from the model.

This model is sometimes used when researchers know that the response variable must be equal to zero when the predictor variable is equal to zero.

In the real world, this type of model is used most commonly in forestry or .

For example, researchers may use tree circumference to predict tree height. If a given tree has a circumference of zero, it must have a height of zero.

Thus, when fitting a regression model to this data it wouldn't make sense for the intercept term to be non-

zero.

The following example shows the difference between fitting an ordinary simple linear regression model compared to a model that implements regression through the origin.

Example: Regression Through the Origin

Suppose a biologist wants to fit a regression model using tree circumference to predict tree height. She goes out and collects the following measurements for a sample of 15 trees:

Circumference (inches)	Height (inches)
15	200
19	234
25	285
39	375
44	440
46	470
49	564
54	544
67	639
79	750
81	830
84	854
88	901
90	912
99	989

#create data frame

```
df <- data.frame(circ=c(15, 19, 25, 39, 44, 46, 49, 54, 67,  
79, 81, 84, 88, 90, 99),  
height=c(200, 234, 285, 375, 440, 470, 564, 544, 639, 750,  
830, 854,  
901, 912, 989))
```

#fit a simple linear regression model

```
model <- lm(height ~ circ, data = df)
```

#fit regression through the origin

```
model_origin <- lm(height ~ 0 + ., data = df)
```

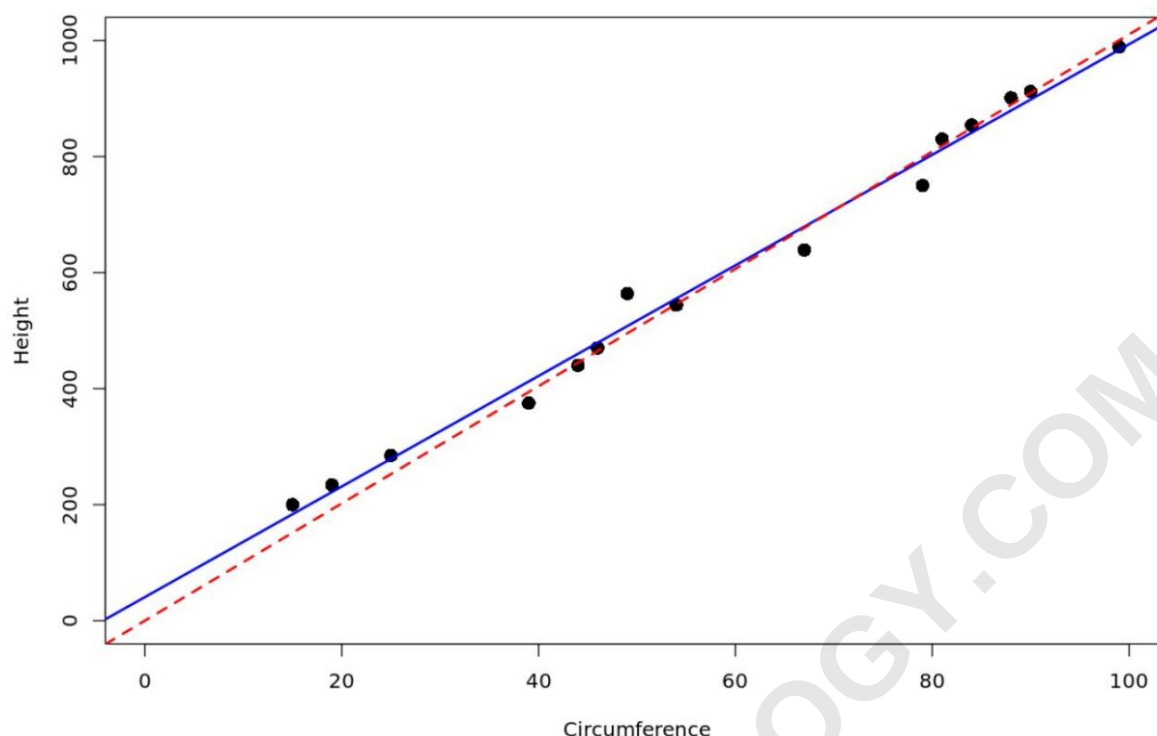
#create scatterplot

```
plot(df$circ, df$height, xlab='Circumference',  
ylab='Height',  
cex=1.5, pch=16, ylim=c(0,1000), xlim=c(0,100))
```

#add the fitted regression lines to the scatterplot

```
abline(model, col='blue', lwd=2)
```

```
abline(model_origin, lty='dashed', col='red', lwd=2)
```



The red dashed line represents the regression model that goes through the origin and the blue solid line represents the ordinary simple linear regression model.

We can use the following code in R to get the coefficient estimates for each model:

```
#display coefficients for simple linear regression model  
coef(model)
```

```
(Intercept) circ
```

```
40.696971 9.529631
```

```
#display coefficients for regression model through the
```

origin

coef(model_origin)

circ

10.10574

The fitted equation for the simple linear regression model is:

Height = 40.6969 + 9.5296(Circumference)

And the fitted equation for the regression model through the origin is:

Height = 10.1057(Circumference)

Notice that the coefficient estimates for the circumference variable are slightly different.

Cautions on Using Regression Through the Origin

Before using regression through the origin, you must be absolutely sure that a value of 0 for the predictor variable implies a value of 0 for the response variable. In many scenarios, it's almost impossible to know this for sure.

And if you're using regression through the origin to save one degree of freedom from estimating the intercept, this rarely makes a substantial difference if your sample size is large enough.

If you do choose to use regression through the origin, be sure to state your reasoning in your final analysis or report.

Additional Resources

The following tutorials provide additional information about linear regression: