

How can I calculate the standardized regression coefficients in R?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the standardized regression coefficients in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=147803>

Calculating standardized regression coefficients in R involves using the "lm" function to create a linear regression model and then using the "coef" function to extract the regression coefficients. These coefficients can then be standardized by dividing each coefficient by the standard deviation of its respective predictor variable. This process allows for a direct comparison of the impact of each predictor variable on the outcome variable, regardless of the scale or units of the predictor variables. Standardized regression coefficients are useful for identifying the most influential predictors in a model and for making comparisons between different models.

Calculate Standardized Regression Coefficients in R

Typically when we perform , the resulting regression coefficients in the model output are unstandardized, meaning they use the raw data to find the line of best fit.

```
model <- lm(price ~ age + sqfeet, data=df)
```

However, it's possible to standardize each predictor variable and the response variable (by subtracting the mean value of each variable from the original values and then dividing by the variables standard deviation) and then perform regression, which results in standardized regression coefficients.

The easiest way to calculate standardized regression coefficients in R is by using the function to standardize each variable in the model:

```
model <- lm(scale(price) ~ scale(age) + scale(sqfeet),
```

data=df)

The following example shows how to calculate standardized regression coefficients in practice.

Example: How to Calculate Standardized Regression Coefficients in R

Suppose we have the following dataset that contains information about the age, square footage, and selling price of 12 houses:

#create data frame

```
df <- data.frame(age=c(4, 7, 10, 15, 16, 18, 24, 28, 30, 35, 40, 44),  
sqfeet=c(2600, 2800, 1700, 1300, 1500, 1800,  
1200, 2200, 1800, 1900, 2100, 1300),  
price=c(280000, 340000, 195000, 180000, 150000,  
200000,  
180000, 240000, 200000, 180000, 260000, 140000))
```

#view data frame

df

age sqfeet price

1 4 2600 280000

2 7 2800 340000
3 10 1700 195000
4 15 1300 180000
5 16 1500 150000
6 18 1800 200000
7 24 1200 180000
8 28 2200 240000
9 30 1800 200000
10 35 1900 180000
11 40 2100 260000
12 44 1300 140000

Suppose we then perform multiple linear regression using age and square footage as the predictor variables and price as the response variable:

```
#fit regression model  
model <- lm(price ~ age + sqfeet, data=df)
```

```
#view model summary  
summary(model)
```

Call:

```
lm(formula = price ~ age + sqfeet, data = df)
```

Residuals:**Min 1Q Median 3Q Max****-32038 -10526 -6139 21641 34060****Coefficients:****Estimate Std. Error t value Pr(>|t|)****(Intercept) 34736.54 37184.32 0.934 0.374599****age -409.83 612.46 -0.669 0.520187****sqfeet 100.87 15.75 6.405 0.000125 *******---****Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1****Residual standard error: 24690 on 9 degrees of freedom****Multiple R-squared: 0.8508, Adjusted R-squared: 0.8176****F-statistic: 25.65 on 2 and 9 DF, p-value: 0.0001916**

From the model output we can see the unstandardized regression coefficients:

Intercept: 34736.54 Age: -409.83 Sq Feet: 100.87

Upon first glance, it appears that age has a much larger effect on house price since its coefficient in the regression table is -409.833 compared to just 100.866 for the predictor variable square footage.

However, the standard error is much larger for age compared to square footage, which is why the corresponding p-value is actually large for age ($p=0.520$) and small for square footage ($p=0.000$).

The reason for the extreme differences in regression coefficients is because of the extreme differences in scales for the two variables:

The values for age range from 4 to 44. The values for square footage range from 1,200 to 2,800.

#standardize each variable and fit regression model

```
model_std <- lm(scale(price) ~ scale(age) +  
scale(sqfeet), data=df)
```

#turn off scientific notation

```
options(scipen=999)
```

#view model summary

```
summary(model_std)
```

Call:

```
lm(formula = scale(price) ~ scale(age) + scale(sqfeet),  
data = df)
```

Residuals:**Min 1Q Median 3Q Max****-0.5541 -0.1820 -0.1062 0.3743 0.5891****Coefficients:****Estimate Std. Error t value Pr(>|t|)****(Intercept) -0.00000000000000000002253****0.1232881457926768426 0.000 1.000000****scale(age) -0.0924421263946849786****0.1381464029075653854 -0.669 0.520187****scale(sqfeet) 0.8848591938302141635****0.1381464029075653577 6.405 0.000125****(Intercept)****scale(age)****scale(sqfeet) *******---****Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1****Residual standard error: 0.4271 on 9 degrees of freedom****Multiple R-squared: 0.8508, Adjusted R-squared: 0.8176****F-statistic: 25.65 on 2 and 9 DF, p-value: 0.0001916****The regression coefficients in this table are**

standardized, meaning they used standardized data to fit this regression model.

The way to interpret the coefficients in the table is as follows:

A one standard deviation increase in age is associated with a 0.092 standard deviation decrease in house price, assuming square footage is held constant. A one standard deviation increase in square footage is associated with a 0.885 standard deviation increase in house price, assuming age is held constant.

Now we can see that square footage has a much larger effect on house price than age.

Note: The p-values for each predictor variable are the exact same as the previous regression model.

When deciding on the final model to use, we now know that square footage is much more important for predicting the price of a house compared to age.

The following tutorials provide additional information about regression models:

How to Read and Interpret a Regression Table

How to Interpret Regression Coefficients

ARABPSYCHOLOGY.COM