

Log-Log Regression

April 14, 2024

Overview

Log-log regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables, where both the dependent and the independent variables undergo logarithmic transformations. This method is particularly effective for examining multiplicative relationships and elasticities between variables that exhibit exponential growth or decline.

Purpose

- **Elasticity Analysis:** Log-log regression is commonly employed to analyze the elasticity of a variable, which is useful for understanding how percentage changes in one variable are associated with percentage changes in another.
- **Scale Invariance:** The logarithmic transformation of both dependent and independent variables helps in handling data across different scales and makes the model scale-invariant.

Business Applications of Log-Log Regression

Marketing

The following are some marketing applications of log-log regression:

Price Elasticity: Log-log regression can be used to estimate the price elasticity of a product's demand. A linear relationship can be depicted between the **log of the quantity demanded** and the **log of the price** by taking the natural logarithm of both the dependent and independent variables. The estimated **price elasticity of demand** can be used to inform pricing strategies and maximize revenue for businesses.

Advertising Effectiveness: The effect of advertising on sales can be estimated using log-log regression. A linear relationship can be modeled between the **log of sales** and the **log of advertising** expenditure by taking the natural logarithm of both the dependent and independent variables. The estimated coefficient can be used to guide advertising expenditure decisions and assist businesses in optimizing their marketing campaigns.

Brand Loyalty: Log-log regression can be utilized to estimate the impact of brand loyalty on sales. A linear relationship can be modeled between the **log of sales** and the **log of brand loyalty** by taking the natural logarithm of both the dependent and independent variables. The estimated coefficient can be used to **inform brand strategy decisions** and assist businesses in identifying market share expansion opportunities.

Market Segmentation Analysis: Log-log regression can be used to identify and analyze market segments based on product attributes. A linear relationship can be modeled between the log of the product attribute and the log of market share by taking the natural logarithm of both the dependent and independent variables. Estimates of the resulting coefficients can be used to determine which product attributes are most essential to each market segment and to inform decisions regarding product development. [2]

Finance

The following are some finance applications of log-log regression:

Asset Pricing Models: Log-log regression can be used to estimate asset pricing models, such as the capital asset pricing model (CAPM). By taking the natural logarithm of both the dependent and independent variables, a linear relationship can be modeled between the **log of the expected return** and the **log of the risk premium**. The resulting coefficient estimate can be used to inform investment decisions and help investors evaluate the risk and return of a portfolio.

Risk Management: Log-log regression can be used to estimate risk models, such as **value at risk (VaR)**. By taking the natural logarithm of both the dependent and independent variables, a linear relationship can be modeled between the **log of the portfolio value** and the **log of the portfolio risk**. The resulting coefficient estimate can be used to estimate the level of risk that the portfolio is exposed to and inform risk management decisions.

Option Pricing: Log-log regression can be used to estimate option pricing models, such as the Black-Scholes model. By taking the natural logarithm of both the dependent and independent variables, a linear relationship can be modeled between the **log of the stock price** and the **log of the option price**. The resulting coefficient estimate can be used to inform option pricing decisions and help investors **evaluate the fair value of an option**.

Credit Risk Analysis: Log-log regression can be used to estimate credit risk models, such as the **credit default swap (CDS) pricing** model. By taking the natural logarithm of both the dependent and independent variables, a linear relationship can be modeled between the

log of the CDS spread and the **log of the credit risk**. The resulting coefficient estimate can be used to inform credit risk analysis and help investors evaluate the **creditworthiness** of a company. [3]

Organizational Behavior

The following are some applications of log-log regression in Organizational Behavior:

Analysis of Employee attrition: Log-log regression can be used to predict employee attrition rates based on a variety of variables, including compensation, job satisfaction, and work environment. A linear relationship can be modeled between the log of the employee turnover rate and the log of the various factors by taking the natural logarithm of both the dependent and independent variables. Estimates of the resulting coefficients can be used to identify the most influential factors influencing employee attrition and inform strategies for employee retention.

Analysis of Employee Performance: Log-log regression can be used to predict employee performance based on a number of variables, including job training, work experience, and job satisfaction. A linear relationship can be modeled between the log of employee performance and the log of the various factors by taking the natural logarithm of both the dependent and independent variables. The estimated coefficients can be used to determine the most influential employee performance factors and to inform training and development strategies.

Organizational Culture Analysis: Analyzing the influence of organizational culture on employee behavior and attitudes can be accomplished through the use of log-log regression. By taking the natural logarithm of both the dependent and independent variables, it is possible to construct a linear relationship between the log of employee behavior and attitudes and the log of the different aspects of organizational culture. The estimated coefficients can be used to determine the most influential aspects of organizational culture on employee conduct and attitudes.

Leadership Effectiveness Analysis: Analyzing the influence of leadership on employee behavior and performance can be accomplished using log-log regression. A linear relationship can be modeled between the log of employee behavior and performance and the log of the various leadership factors by taking the natural logarithm of both the dependent and independent variables. The estimated coefficients can be used to identify the most influential leadership factors on employee behavior and performance, as well as to inform leadership development strategies. [4]

Model

Model Form

The general form of a log-log regression model is represented by the following equation:

$$\log(Y) = \beta_0 + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \dots + \beta_n \log(X_n) + \epsilon \quad (0.1)$$

Where:

- Y is the dependent variable (after applying the logarithm).
- X_1, X_2, \dots, X_n are the independent variables (after applying logarithms to each).
- $\beta_0, \beta_1, \dots, \beta_n$ are the coefficients that the regression model aims to estimate.
- ϵ is the error term.

Assumptions

Similar to other regression analyses, log-log regression relies on several assumptions:

- **Linearity:** The relationship between the log-transformed dependent and independent variables is linear.
- **Independence:** Observations must be independent of each other.
- **Homoscedasticity:** The variance of error terms should be consistent across different values of independent variables.
- **Normal Distribution of Errors:** Errors should follow a normal distribution for valid hypothesis testing.

Advantages

- **Interpretability of Elasticities:** The coefficients in a log-log model represent the elasticity between variables, providing insights into how a percentage change in one variable affects another.
- **Handling Non-linear Patterns:** By transforming variables logarithmically, non-linear relationships can be modeled linearly, enhancing the analysis of complex patterns.

Disadvantages

- **Zero Values Problem:** Logarithmic transformation cannot be applied directly to zero or negative values, which might require additional data adjustments or transformations.
- **Interpretation Complexity:** Understanding and interpreting elasticities might be less straightforward compared to additive models, requiring a good grasp of logarithmic relationships. [1]

Estimation

1. The natural logarithm transformation of both variables can help to capture non-linear relationships, stabilize the data variance, and clarify the relationship between the variables. The coefficients β_0 and β_1 are elasticities that quantify the percentage change in the dependent variable for a one percent change in the independent variable.
2. The method of least squares is used to minimize the sum of squared residuals when estimating coefficients. This involves determining the values of β_0 and β_1 that minimize the difference between the observed and predicted values of the dependent variable. [5]

MODEL 1: Base Model

For reference and a point of comparison, we setup the same Base Model as the one used in the previous chapter on Log-Linear Regression.

```
data(mtcars)
attach(mtcars)
```

Simple Linear Regression

```
fit <- lm(mpg ~ wt, data = mtcars)
summary(fit)
```

Call:

```
lm(formula = mpg ~ wt, data = mtcars)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

```
-4.5432 -2.3647 -0.1252  1.4096  6.8727
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.2851      1.8776   19.858 < 2e-16 ***
wt           -5.3445      0.5591   -9.559 1.29e-10 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.046 on 30 degrees of freedom

Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446

F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

MODEL 2: Log-Log Model

Model 2 adopts a log-log regression approach to explore the relationship between the weight (`wt`) of cars and their fuel efficiency (`mpg`) using the `mtcars` dataset. In contrast to a simple linear regression model that predicts `mpg` directly from `wt`, this model predicts the logarithm of `mpg` from the logarithm of `wt`. This double logarithmic transformation is particularly useful for examining how percentage changes in vehicle weight influence percentage changes in fuel efficiency, reflecting an elasticity-based relationship between these variables.

```
fit2 <- lm(log(mpg) ~ log(wt), data = mtcars)
summary(fit2)
```

Call:

```
lm(formula = log(mpg) ~ log(wt), data = mtcars)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-0.18141 -0.10681 -0.02125  0.08109  0.26930
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.90181    0.08790   44.39 < 2e-16 ***
log(wt)        -0.84182    0.07549  -11.15 3.41e-12 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1334 on 30 degrees of freedom

Multiple R-squared: 0.8056, Adjusted R-squared: 0.7992
F-statistic: 124.4 on 1 and 30 DF, p-value: 3.406e-12

Model Output and Interpretation

The results from our log-log regression model are elucidated below:

Residuals

The residuals, or the differences between observed and predicted values of $\log(\text{mpg})$, show a median close to zero (-0.02125), indicating that, on average, the model's predictions closely match the actual log values. The spread of residuals from the minimum (-0.18141) to the maximum (0.26930) shows that most predictions fall within this range, suggesting a reasonable model fit.

Model Fit

- **Residual Standard Error (RSE): 0.1334** on 30 degrees of freedom indicates the typical deviation of the predicted values from the actual values.
- **Multiple R-squared (0.8056):** About 80.56% of the variability in $\log(\text{mpg})$ is explained by the model, indicating a strong fit.
- **Adjusted R-squared (0.7992):** This adjusted statistic accounts for the number of predictors and still explains about 79.92% of the variability in $\log(\text{mpg})$, supporting a robust model.
- **F-statistic (124.4):** This high value and the corresponding very low p-value (3.406e-12) affirm the overall significance of the model, demonstrating that the model fits the data well and the variables are appropriate.

Coefficients

- **Intercept (3.90181):** This value represents the expected value of $\log(\text{mpg})$ when $\log(\text{wt})$ equals zero. Since $\log(\text{wt})$ equals zero when wt is 1 (not zero due to the nature of logarithmic transformation), this intercept can be interpreted as the expected log of mpg for a car with a weight of 1 unit.

To find the fuel efficiency (mpg) of a car with a weight of 1 unit, we can use the intercept value from the log-log regression model. The calculation involves converting $\log(\text{mpg})$ back to mpg using the exponential function:

To ensure the given equations compile correctly in a Quarto (previously R Markdown) document when converting to HTML or PDF, you can format the equations using LaTeX syntax within a Quarto document. Here's how you can write it:

$$\text{mpg} = e^{\text{intercept}} = e^{3.90181}$$

Evaluating this expression yields:

$$\text{mpg} = 49.49$$

This result indicates that the mpg for a car weighing 1 unit is approximately 49.49 miles per gallon.

- **$\log(\text{wt})$ Coefficient (-0.84182):** This coefficient indicates the expected change in $\log(\text{mpg})$ for each one-unit increase in $\log(\text{wt})$. Specifically, a one-unit increase in $\log(\text{wt})$ is associated with a decrease of 0.84182 in $\log(\text{mpg})$. This negative coefficient reflects a strong inverse relationship, suggesting that as car weight increases on a logarithmic scale, fuel efficiency decreases.

In practical terms, a 1% increase in a car's weight leads to an approximate 0.84182% decrease in its fuel efficiency. This is interpreted as the elasticity of mpg to weight, indicating a highly elastic relationship where small percentage changes in weight lead to substantial percentage changes in mpg.

Therefore, the beta coefficient of -0.84182 indicates a strong negative elasticity, signifying that increases in car weight have a substantial negative impact on fuel efficiency, all else being constant.

In contrast, in the linear-linear model, the relationship is additive rather than multiplicative, with each unit increase in weight reducing the mpg linearly by 0.27178 units.

Calculating Percentage Change:

To translate this coefficient into more intuitive terms, we use the exponential function:

$$\text{Percentage Change} = e^{\text{Coefficient}}$$

For the given coefficient:

$$e^{-0.84182} \approx 0.431$$

This result means that for every 1% increase in weight, the fuel efficiency decreases to about 43.1% of its previous amount in terms of multiplicative change.

Interpreting the Percentage:: If fuel efficiency decreases to 43.1% of its previous amount with each 1% increase in weight, it effectively decreases by 56.9% (100% - 43.1%) of the additional amount that it would have been without the weight increase. This shows a significant decline in efficiency as weight increases.

The coefficient in a log-log regression tells us about the elasticity between variables. Here, a negative elasticity of about -0.84182 means that increasing the car's weight by 1% results in reducing its fuel efficiency by about 56.9%, indicating a strong negative impact of increased weight on fuel efficiency.

Significance

In the log-log regression model, we conduct hypothesis testing for each coefficient separately to assess their statistical significance in predicting the dependent variable, $\log(\text{mpg})$. Here's how we define and interpret the tests for the intercept and $\log(\text{wt})$:

- **For the Intercept:**

- **Null Hypothesis (H0):** The intercept is zero, suggesting it has no effect on $\log(\text{mpg})$ when all independent variables are zero (logarithmically speaking).
- **Alternative Hypothesis (H1):** The intercept is not zero, indicating it does affect $\log(\text{mpg})$ and provides a baseline level of mpg when wt equals 1 unit (since $\log(1) = 0$).
- **Result:** Given that the p-value for the intercept is significantly below 0.05, we reject the null hypothesis. This means there is strong statistical evidence that the intercept is a significant contributor to the model, affecting the baseline mpg calculation.

- **For $\log(\text{wt})$:**

- **Null Hypothesis (H0):** The coefficient for $\log(\text{wt})$ is zero, implying no relationship between weight and mpg on a logarithmic scale.
- **Alternative Hypothesis (H1):** The coefficient for $\log(\text{wt})$ is not zero, suggesting that changes in weight have a significant effect on mpg.
- **Result:** The very low p-value associated with the $\log(\text{wt})$ coefficient leads us to reject the null hypothesis. This rejection provides substantial evidence that an increase in weight significantly decreases mpg, affirming the predictive power of $\log(\text{wt})$ in the model.

The statistical tests confirm that both the baseline level of mpg and the effect of weight changes are significant factors in the model.

Another Log-Log Model

We present two regression models, Model 3a (Linear-Linear Model) and Model 3b (Log-Log Model), using the `mtcars` dataset. Both models aim to predict the miles per gallon (`mpg`) of cars based on their weight (`wt`) and transmission type (`am`, with levels “Automatic” and “Manual”).

MODEL 3a - Linear-Linear Model

```
# Convert am to a factor variable
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))

fit3a <- lm((mpg) ~ wt + am, data = mtcars)
summary(fit3a)
```

Call:

```
lm(formula = (mpg) ~ wt + am, data = mtcars)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.5295	-2.3619	-0.1317	1.4025	6.8782

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.32155	3.05464	12.218	5.84e-13 ***
wt	-5.35281	0.78824	-6.791	1.87e-07 ***
amManual	-0.02362	1.54565	-0.015	0.988

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.098 on 29 degrees of freedom

Multiple R-squared: 0.7528, Adjusted R-squared: 0.7358

F-statistic: 44.17 on 2 and 29 DF, p-value: 1.579e-09

MODEL 3b: Another Log-Log Model

```
fit3b <- lm(log(mpg) ~ log(wt) + am, data = mtcars)
summary(fit3b)
```

Call:

```
lm(formula = log(mpg) ~ log(wt) + am, data = mtcars)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.194110	-0.117056	-0.008833	0.071274	0.258052

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.04162	0.14480	27.911	< 2e-16 ***
log(wt)	-0.93629	0.10822	-8.652	1.58e-09 ***
amManual	-0.08329	0.06886	-1.210	0.236

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1324 on 29 degrees of freedom

Multiple R-squared: 0.815, Adjusted R-squared: 0.8022

F-statistic: 63.87 on 2 and 29 DF, p-value: 2.369e-11

Model Output and Interpretation

The output from Model 3b highlights the relationship between the logarithm of miles per gallon (`mpg`) and factors including the logarithm of car weight (`wt`) and transmission type (`am`). Below, we detail the interpretation of the model's results:

Residuals

- The residuals of the model show a median very close to zero (-0.008833) and are relatively tightly distributed, ranging from -0.194110 to 0.258052. This tight clustering indicates that the model predictions are generally close to the observed data, demonstrating a good fit.

Coefficients

- **Intercept (4.04162):** The intercept, statistically highly significant with a p-value much less than 0.05, represents the expected value of $\log(\text{mpg})$ when $\log(\text{wt})$ equals zero and when the car has an automatic transmission. The high t-value (27.911) underlines its statistical reliability. For practical purposes, this value can be interpreted as the expected log of mpg for a hypothetically light car with a weight of 1 unit (since $\log(1) = 0$) and an automatic transmission.
- **$\log(\text{wt})$ Coefficient (-0.93629):** The coefficient for $\log(\text{wt})$ is negative, indicating that as the weight of a car increases, its fuel efficiency decreases. The elasticity of mpg with respect to weight is about -0.936, meaning that a 1% increase in weight leads to roughly a 0.936% decrease in mpg. This relationship is statistically significant, with a very low p-value (1.58e-09) and a strong t-value (-8.652), suggesting a strong and reliable negative impact of weight on fuel efficiency.
- **amManual (-0.08329):** This coefficient is associated with cars having manual transmissions compared to the baseline of automatic transmissions. The negative sign suggests a decrease in fuel efficiency for manual compared to automatic, but this result is not statistically significant (p-value = 0.236), indicated by a t-value of -1.210. It implies that, after controlling for weight, the type of transmission (manual vs. automatic) does not have a significant effect on the mpg in this dataset.

Model Fit and Statistics

- **Residual Standard Error (0.1324):** The RSE is quite low, which reflects the small average distance of the data points from the fitted line, indicating a good fit.
- **Multiple R-squared (0.815):** Approximately 81.5% of the variation in $\log(\text{mpg})$ is explained by the combined effects of $\log(\text{wt})$ and transmission type. This high R-squared value shows that the model does a good job of capturing the relationship between these variables.
- **Adjusted R-squared (0.8022):** Adjusted for the number of predictors, it confirms that the model explains a significant amount of the variability in $\log(\text{mpg})$.
- **F-statistic (63.87):** This value tests the overall significance of the regression model, and the associated p-value (2.369e-11) indicates the model is statistically significant.

This analysis reveals that weight is a critical factor in predicting the fuel efficiency of cars, with significant impacts demonstrated through the log-log regression model. However, the transmission type does not show a statistically significant influence on mpg when considering this dataset.