

What is the interpretation of null and residual deviance, and how can they be applied in real-world scenarios? Can you provide examples to illustrate their significance?

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the interpretation of null and residual deviance, and how can they be applied in real-world scenarios? Can you provide examples to illustrate their significance?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143427>

Null and residual deviance are statistical measures used in the analysis of binary or categorical data. In simpler terms, they help us understand how well our model fits the data.

The null deviance represents the amount of variability in the data that cannot be explained by the model. It is calculated by comparing the model's predicted values with the actual values. A lower null deviance indicates a better fit of the model to the data.

On the other hand, residual deviance measures the amount of variability in the data that is not explained even after accounting for the variables in the model. A lower residual deviance indicates a better fit of the model to the data.

In real-world scenarios, null and residual deviance can be applied to assess the goodness of fit of a model and to compare different models. They can also be used to identify outliers or influential data points that may affect the model's performance. For example, in a medical study, null and residual deviance can be used to determine the effectiveness of a new treatment compared to a standard treatment. In marketing, they can be used to evaluate the effectiveness of a new advertising campaign.

In summary, null and residual deviance provide valuable insights into the accuracy and reliability of statistical models, making them an essential tool in various fields such as healthcare, marketing, finance, and social sciences.

Interpret Null & Residual Deviance (With Examples)

Whenever you fit a general linear model (like logistic regression, Poisson regression, etc.), most statistical software will produce values for the null deviance and residual deviance of the model.

The null deviance tells us how well the response variable can be predicted by a model with only an intercept term.

The residual deviance tells us how well the response

variable can be predicted by a model with p predictor variables. The lower the value, the better the model is able to predict the value of the response variable.

To determine if a model is "useful" we can compute the Chi-Square statistic as:

$X^2 = \text{Null deviance} - \text{Residual deviance}$

with p degrees of freedom.

We can then find the p-value associated with this Chi-Square statistic. The lower the p-value, the better the model is able to fit the dataset compared to a model with just an intercept term.

The following example shows how to interpret null and residual deviance for a logistic regression model in R.

Example: Interpreting Null & Residual Deviance

For this example, we'll use the Default dataset from the ISLR package. We can use the following code to load and view a summary of the dataset:

```
#load dataset
```

```
data <- ISLR::Default
```

```
#view summary of dataset
```

```
summary(data)
```

```
default student balance income
```

```
No :9667 No :7056 Min. : 0.0 Min. : 772
```

```
Yes: 333 Yes:2944 1st Qu.: 481.7 1st Qu.:21340
```

```
Median : 823.6 Median :34553
```

```
Mean : 835.4 Mean :33517
```

```
3rd Qu.:1166.3 3rd Qu.:43808
```

```
Max. :2654.3 Max. :73554
```

This dataset contains the following information about 10,000 individuals:

default: Indicates whether or not an individual defaulted.
student: Indicates whether or not an individual is a student.
balance: Average balance carried by an individual.
income: Income of the individual.

We will use student status, bank balance, and income to build a that predicts the probability that a given individual defaults:

#fit logistic regression model

```
model <- glm(default~balance+student+income,  
family="binomial", data=data)
```

#view model summary

```
summary(model)
```

Call:

```
glm(formula = default ~ balance + student + income,  
family = "binomial",  
data = data)
```

Deviance Residuals:

```
Min 1Q Median 3Q Max  
-2.4691 -0.1418 -0.0557 -0.0203 3.7383
```

Coefficients:

```
Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.087e+01 4.923e-01 -22.080 < 2e-16 ***  
balance 5.737e-03 2.319e-04 24.738 < 2e-16 ***  
studentYes -6.468e-01 2.363e-01 -2.738 0.00619 **  
income 3.033e-06 8.203e-06 0.370 0.71152
```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2920.6 on 9999 degrees of freedom

Residual deviance: 1571.5 on 9996 degrees of freedom

AIC: 1579.5

Number of Fisher Scoring iterations: 8

We can observe the following values in the output for the null and residual deviance:

**Null deviance: 2920.6 with df = 9999
Residual deviance: 1571.5 with df = 9996
 $X^2 = \text{Null deviance} - \text{Residual deviance}$
 $X^2 = 2920.6 - 1579.0 = 1341.6$**

There are $p = 3$ predictor variables degrees of freedom.

We can use the to find that a X^2 value of 1331.6 with 3 degrees of freedom has a p-value of 0.000000.

Since this p-value is much less than .05, we would conclude that the model is highly useful for predicting the probability that a given individual defaults.

The following tutorials explain how to perform logistic

regression in practice in both R and Python:

ARABPSYCHOLOGY.COM