

“How can the glm output in R be interpreted, with an example?”

Authored by
stats writer

July 2, 2024

RECOMMENDED CITATION

stats writer (2024). “How can the glm output in R be interpreted, with an example?”. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165520>

The glm (generalized linear model) output in R can be interpreted as a statistical model that is used to analyze the relationship between a dependent variable and one or more independent variables. The output provides information on the coefficients, standard errors, and p-values of the independent variables, as well as the overall model fit and significance. An example of interpreting the glm output in R would be if we are examining the relationship between a person's age (independent variable) and their income (dependent variable). The output would show the coefficient for age, indicating the magnitude of the effect of age on income, as well as the p-value, which would determine if the relationship is statistically significant. This information can be used to understand and make predictions about the relationship between age and income.

Interpret glm Output in R (With Example)

The glm() function in R can be used to fit generalized linear models.

This function uses the following syntax:

```
glm(formula, family=gaussian, data, ...)
```

where:

formula: The formula for the linear model (e.g. $y \sim x_1 + x_2$)
family: The statistical family to use to fit the model. Default is gaussian but other options include binomial, Gamma, and poisson among others.
data: The name of the data frame that contains the data

In practice, this function is used most often to fit by specifying the 'binomial' family.

The following example shows how to interpret the glm output in R for a logistic regression model.

Example: How to Interpret glm Output in R

For this example, we'll use the built-in dataset in R:

```
#view first six rows of mtcars dataset
```

```
head(mtcars)
```

```
mpg cyl disp hp drat wt qsec vs am gear carb
```

```
Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4
```

```
Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4
```

```
Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1
```

```
Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1
```

```
Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3
```

```
2
```

```
Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1
```

We will use the variables `disp` and `hp` to predict the probability that a given car takes on a value of 1 for the `am` variable.

The following code shows how to use the `glm()` function to fit this logistic regression model:

```
#fit logistic regression model
```

```
model <- glm(am ~ disp + hp, data=mtcars,  
family=binomial)
```

```
#view model summary
```

```
summary(model)
```

```
Call:
```

```
glm(formula = am ~ disp + hp, family = binomial, data =  
mtcars)
```

```
Deviance Residuals:
```

```
Min 1Q Median 3Q Max
```

```
-1.9665 -0.3090 -0.0017 0.3934 1.3682
```

```
Coefficients:
```

```
Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) 1.40342 1.36757 1.026 0.3048
```

```
disp -0.09518 0.04800 -1.983 0.0474 *
```

```
hp 0.12170 0.06777 1.796 0.0725 .
```

```
---
```

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

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 43.230 on 31 degrees of freedom
```

Residual deviance: 16.713 on 29 degrees of freedom
AIC: 22.713

Number of Fisher Scoring iterations: 8

Here's how to interpret each piece of the output:

Coefficients & P-Values

The coefficient estimate in the output indicate the average change in the log odds of the response variable associated with a one unit increase in each predictor variable.

For example, a one unit increase in the predictor variable disp is associated with an average change of -0.09518 in the log odds of the response variable am taking on a value of 1. This means that higher values of disp are associated with a lower likelihood of the am variable taking on a value of 1.

For example, the z value for the predictor variable disp is calculated as $-.09518 / .048 = -1.983$.

The p-value $\Pr(>|z|)$ tells us the probability associated with a particular z value. This essentially tells us how

well each predictor variable is able to predict the value of the response variable in the model.

For example, the p-value associated with the z value for the disp variable is .0474. Since this value is less than .05, we would say that disp is a statistically significant predictor variable in the model.

Depending on your preferences, you may decide to use a significance level of .01, .05, or 0.10 to determine whether or not each predictor variable is statistically significant.

Null & Residual Deviance

The null deviance in the output 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 the specific model that we fit 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.

For example, in our regression model we can observe the following values in the output for the null and residual deviance:

**Null deviance: 43.23 with df = 31
Residual deviance: 16.713 with df = 29**

We can use these values to calculate the X^2 statistic of the model:

**$X^2 = \text{Null deviance} - \text{Residual deviance}$
 $X^2 = 43.23 - 16.713$
 $X^2 = 26.517$**

There are $p = 2$ predictor variables degrees of freedom.

We can use the to find that a X^2 value of 26.517 with 2

degrees of freedom has a p-value of 0.000002.

Since this p-value is much less than .05, we would conclude that the model is highly useful.

AIC

The Akaike information criterion (AIC) is a metric that is used to compare the fit of different regression models. The lower the value, the better the regression model is able to fit the data.

It is calculated as:

$$\text{AIC} = 2K - 2\ln(L)$$

where:

K: The number of model parameters. $\ln(L)$: The log-likelihood of the model. This tells us how likely the model is, given the data.

The actual value for the AIC is meaningless.

However, if you fit several regression models, you can compare the AIC value of each model. The model with the lowest AIC offers the best fit.

Related:

Additional Resources

The following tutorials provide additional information on how to use the glm() function in R:

The following tutorials explain how to handle common errors when using the glm() function:

ARABPSYCHOLOGY.COM