What are some examples of interpreting log-likelihood values?

Authored by

stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *What are some examples of interpreting log-likelihood values?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=143415

Log-likelihood values are used to measure the goodness of fit of a statistical model to a set of data. This value is calculated by taking the logarithm of the likelihood function, which is a measure of how likely it is for the observed data to occur under the given model.

Interpreting log-likelihood values involves comparing them to other models or to a null model. A higher log-likelihood value indicates a better fit of the model to the data, while a lower value suggests a poorer fit. For example, if a model has a log-likelihood value of 100 and another model has a value of 80, it can be concluded that the first model is a better fit for the data.

Another way to interpret log-likelihood values is by comparing them to a null model, which assumes that there is no relationship between the variables in the model. A higher log-likelihood value than the null model suggests that the model has some predictive power and is a better fit for the data.

In summary, interpreting log-likelihood values involves comparing them to other models or to a null model in order to determine the goodness of fit of a statistical model to a set of data.

Interpret Log-Likelihood Values (With Examples)

The log-likelihood value of a regression model is a way to measure the goodness of fit for a model. The higher the value of the log-likelihood, the better a model fits a dataset.

The log-likelihood value for a given model can range from negative infinity to positive infinity. The actual log-likelihood value for a given model is mostly meaningless, but it's useful for comparing two or more models.

In practice, we often fit several regression models to a dataset and choose the model with the highest log-

likelihood value as the model that fits the data best.

The following example shows how to interpret log-likelihood values for different regression models in practice.

Example: Interpreting Log-Likelihood Values

Suppose we have the following dataset that shows the number of bedrooms, number of bathrooms, and selling price of 20 different houses in a particular neighborhood:

Bedrooms	Bathrooms	Price (thousands)
1	2	120
1	1	133
1	4	139
2	3	185
2	2	148
2	2	160
2	3	192
3	5	205
3	4	244
3	3	213
3	4	236
3	4	280
3	3	275
3	4	273
4	2	312
4	4	311
4	3	304
5	5	415
5	6	396
6	7	488

Suppose we'd like to fit the following two regression models and determine which one offers a better fit to the data:

Model 1: Price = β0 + β1(number of bedrooms)

Model 2: Price = β0 + β1(number of bathrooms)

The following code shows how to fit each regression model and calculate the log-likelihood value of each model in R:

```
#define data
df <- data.frame(beds=c(1, 1, 1, 2, 2, 2, 2, 3, 3, 3,
3, 3, 3, 3, 4, 4, 4, 5, 5, 6),
baths=c(2, 1, 4, 3, 2, 2, 3, 5, 4, 3,
4, 4, 3, 4, 2, 4, 3, 5, 6, 7),
price=c(120, 133, 139, 185, 148, 160, 192, 205, 244, 213,
236, 280, 275, 273, 312, 311, 304, 415, 396, 488))

#fit models
model1 <- lm(price~beds, data=df)
model2 <- lm(price~baths, data=df)

#calculate log-likelihood value of each model
logLik(model1)

'log Lik.' -91.04219 (df=3)

logLik(model2)

'log Lik.' -111.7511 (df=3)
```

The first model has a higher log-likelihood value (-91.04) than the second model (-111.75), which means the first model offers a better fit to the data.

Cautions on Using Log-Likelihood Values

When calculating log-likelihood values, it's important to note that adding more predictor variables to a model will almost always increase the log-likelihood value even if the additional predictor variables aren't statistically significant.

This means you should only compare the log-likelihood values between two regression models if each model has the same number of predictor variables.

To compare models with different numbers of predictor variables, you can perform a to compare the goodness of fit of two nested regression models.