Can contrasts be applied to factors with only one level?

Authored by

stats writer

April 30, 2024

Contrasts refer to the comparison of different levels within a categorical factor in statistical data analysis. In some cases, a factor may have only one level, meaning there is no variation or differentiation within that factor. Therefore, it may not be possible to apply contrasts to factors with only one level as there is no contrast or comparison to be made. This is because contrasts require at least two levels within a factor to determine any meaningful differences. Hence, in such scenarios, the application of contrasts may not be applicable or relevant.

Fix: contrasts can be applied only to factors with 2 or more levels

One common error you may encounter in R is:

Error in `contrasts<-`(`*tmp*`, value = contr.funs]) : contrasts can be applied only to factors with 2 or more levels

This error occurs when you attempt to fit a regression model using a predictor variable that is either a factor or character and only has one unique value.

This tutorial shares the exact steps you can use to troubleshoot this error.

Example: How to Fix 'contrasts can be applied only to factors with 2 or more levels'

Suppose we have the following data frame in R:

#create data frame

```
df <- data.frame(var1=c(1, 3, 3, 4, 5),
var2=as.factor(4),
var3=c(7, 7, 8, 3, 2),
var4=c(1, 1, 2, 8, 9))

#view data frame
df
```

```
var1 var2 var3 var4
1 1 4 7 1
2 3 4 7 1
3 3 4 8 2
4 4 4 3 8
5 5 4 2 9
```

Notice that the predictor variable var2 is a factor and only has one unique value.

If we attempt to fit a multiple linear regression model using var2 as one of the predictor variables, we'll get the following error:

```
#attempt to fit regression model
model <- lm(var4 ~ var1 + var2 + var3, data=df)
```

Error in `contrasts<-`(`*tmp*`, value = contr.funs]) : contrasts can be applied only to factors with 2 or more levels

We get this error because var2 only has one unique value: 4. Since there isn't any variation at all in this predictor variable, R is unable to effectively fit a regression model.

We can actually use the following syntax to count the number of unique values for each variable in our data frame:

```
#count unique values for each variable
sapply(lapply(df, unique), length)
```

```
var1 var2 var3 var4
4 1 4 4
```

And we can use the function to display each of the unique values for each variable:

```
#display unique values for each variable
lapply(df, unique)
```

$var1

1 3 4 5

$var2

4

Levels: 4

$var3

7 8 3 2

We can see that var2 is the only variable that has one unique value. Thus, we can fix this error by simply dropping var2 from the regression model:

#fit regression model without using *var2* as a predictor variable

model <- lm(var4 ~ var1 + var3, data=df)

#view model summary

summary(model)

Call:

lm(formula = var4 ~ var1 + var3, data = df)

Residuals:

1 2 3 4 5

0.02326 -1.23256 0.91860 0.53488 -0.24419

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.4070 3.6317 2.315 0.1466

var1 0.6279 0.6191 1.014 0.4172

var3 -1.1512 0.3399 -3.387 0.0772 .

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

Residual standard error: 1.164 on 2 degrees of freedom

Multiple R-squared: 0.9569, Adjusted R-squared: 0.9137

F-statistic: 22.18 on 2 and 2 DF, p-value: 0.04314