How can I test contrasts in R?

Authored by
stats writer

June 30, 2024

Testing contrasts in R is a statistical method used to compare the differences between two or more groups or levels within a variable. This can be done by creating a contrast matrix in R, which assigns coefficients to each level of the variable to be compared. These coefficients are then used to calculate the differences between the groups and determine if they are statistically significant. This method is commonly used in analysis of variance (ANOVA) and linear regression models. By testing contrasts in R, researchers can gain valuable insights into the relationship between variables and make informed decisions based on the results.

How can I test contrasts in R? | R FAQ

```
Version info: Code for this page was tested in R version 3.1.2 (2014-10-31)
On: 2015-06-15
With: knitr 1.8; Kendall 2.2; multcomp 1.3-8; TH.data 1.0-5; survival 2.37-7; mvtnorm 1.0-1
```

After fitting a model with categorical predictors, especially interacted categorical predictors, one may wish to compare different levels of the variables than those presented in the table of coefficients. We can start with a simple linear model with a continuous predictor and two interacted categorical predictors.

library(multcomp)hsb2<-read.csv("https://stats.idre.ucla.edu/stat/data/hsb2.csv") m1<-lm(read~socst+factor(ses)*factor(female),data=hsb2)summary(m1)

```
##
## Call:
## lm(formula = read ~ socst + factor(ses) * factor(female), data = hsb2)
##
## Residuals:
## Min 1Q Median 3Q Max
## -20.844 -5.581 0.238 4.754 18.429
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.9179 3.1844 7.51 2.1e-12 ***
## socst 0.5865 0.0563 10.42 < 2e-16 ***
## factor(ses)2 -1.5349 2.3900 -0.64 0.521
## factor(ses)3 -2.1245 2.6513 -0.80 0.424
## factor(female)1 -4.9856 2.5045 -1.99 0.048 *
## factor(ses)2:factor(female)1 2.3710 2.9752 0.80 0.426
## factor(ses)3:factor(female)1 7.3748 3.2662 2.26 0.025 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.93 on 193 degrees of freedom
```

Multiple R-squared: 0.419, Adjusted R-squared: 0.401
F-statistic: 23.2 on 6 and 193 DF, p-value:

The coefficients listed above provide contrasts between the indicated level
and the omitted reference level and have the following interpretations

The model output includes tests of the null hypotheses that these differences are equal to zero. However, we may be interested in comparing other combinations of ses and female.
We can manually compute these different combinations with some arithmetic, but what if we want to test these differences for siginficance?

We can do so by defining a contrast of interest and testing it with the
glht (generalized linear hypothesis test) command in the multcomp
package. To define the contrast, we can look at the order in which the
coefficients are presented in the output, then create a vector the length of the

coefficient list (including the intercept). To start, we can compare levels 2 and 3 of ses for female = 0. Thus, we want to test the difference between the third and fourth coefficients in our output. After we create our contrast vector, we pass it along with the model object to glht. Then, to see the result, we look at a summary of our glht object.

```
# difference between ses = 2 and ses =3 when female = 0K<-matrix(c(0,0,1,-1,0,0,0),1)t<-glht(m1,linfct=K)summary(t)
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = read ~ socst + factor(ses) * factor(female), data = hsb2)
##
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## 1 == 0 0.59 1.92 0.31 0.76
```

(Adjusted p values reported -- single-step method)

It seems the outcome is not significantly different between ses=2 and ses=3 when female=0. The estimate we see in this output is the same we would calculate by hand, but we get the significance test above:

m1$coef-m1$coef

factor(ses)2
0.58957

We can look at a slightly more complicated contrast, comparing levels 2 and 3
of ses for female = 1:

difference between ses = 2 and ses =3 when female = 1K<-matrix(c(0,0,1,-1,0,1,-1),1)t<-glht(m1,linfct=K)summary(t)

##
Simultaneous Tests for General Linear Hypotheses
##
Fit: lm(formula = read ~ socst + factor(ses) *

factor(female), data = hsb2)
##
Linear Hypotheses:
Estimate Std. Error t value Pr(>|t|)
1 == 0 -4.41 1.87 -2.35 0.02 *
---
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**
(Adjusted p values reported -- single-step method)

To test "differences of differences"-is the difference between ses =
2 and ses = 3 *different* for female = 0 vs. female = 1- we can define
our contrast as the difference in the vectors we defined above and test this
using glht:

 # looking at the difference of differences# ses = 2 vs. 3 for female = 0K1<-matrix(c(0,0,1,-1,0,0,0),1)# ses = 2 vs. 3 for female = 1K2<-matrix(c(0,0,1,-1,0,1,-1),1)# difference of differences(K<-K1-K2)

##

```
## 0 0 0 0 0 -1 1
```

Above is the resulting vector of contrast coefficients to test this difference of differences. We now test this contrast for significance

```
t<-glht(m1,linfct= K)summary(t)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = read ~ socst + factor(ses) * factor(female), data = hsb2)
##
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## 1 == 0 5.00 2.65 1.89 0.061 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Although approaching significance, the difference between ses=2 and ses=3 does not significantly differ between female=0 and female=1.

We can also test all possible pairwise combinations. To make this easier,
we will first create an "interaction" variable (using the function, interaction)
whose levels are created as a combination of the levels of ses and female.

 # all pairwise comparsions# creating a BIG group variablehsb2$tall<-with(hsb2,interaction(female, ses,sep="x"))head(hsb2$tall)

0x1 1x2 0x3 0x3 0x2 0x2
Levels: 0x1 1x1 0x2 1x2 0x3 1x3

All pairwise comparisons can then be calculated automatically by entering the interaction variable into the model as a single predictor.

m2<-lm(read~socst+tall,data= hsb2)l2<-glht(m2,linfct=mcp(tall="Tukey"))summary(l2)

##
Simultaneous Tests for General Linear Hypotheses
##

```
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = read ~ socst + tall, data = hsb2)
##
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## 1x1 - 0x1 == 0 -4.986 2.505 -1.99 0.34
## 0x2 - 0x1 == 0 -1.535 2.390 -0.64 0.99
## 1x2 - 0x1 == 0 -4.150 2.412 -1.72 0.51
## 0x3 - 0x1 == 0 -2.124 2.651 -0.80 0.97
## 1x3 - 0x1 == 0 0.265 2.630 0.10 1.00
## 0x2 - 1x1 == 0 3.451 1.821 1.90 0.40
## 1x2 - 1x1 == 0 0.836 1.825 0.46 1.00
## 0x3 - 1x1 == 0 2.861 2.091 1.37 0.74
## 1x3 - 1x1 == 0 5.250 2.075 2.53 0.12
## 1x2 - 0x2 == 0 -2.615 1.634 -1.60 0.59
## 0x3 - 0x2 == 0 -0.590 1.915 -0.31 1.00
## 1x3 - 0x2 == 0 1.800 1.901 0.95 0.93
## 0x3 - 1x2 == 0 2.025 1.884 1.08 0.89
## 1x3 - 1x2 == 0 4.414 1.875 2.35 0.17
## 1x3 - 0x3 == 0 2.389 2.085 1.15 0.86
## (Adjusted p values reported -- single-step method)
```

A few notes