Contrasts in R

April 24, 2011

Abstract

Categorical explanatory variables can be included in a regression model by representing them by dummy variables. A *contrast* matrix determines the coding scheme of the dummies. Warning: this note is addressed to those who want to understand the details.

1 Contrasts

A *contrast matrix* for a categorical variable with K categories is an $K \times (K-1)$-matrix. It is used to define $K-1$ dummy variables that represent the categorical variable in a linear regression.

We represent the value of measurement (or individual) i on the variable (in the ith row of the regression matrix) by a horizontal vector $B_{i,\cdot}$ of length K, which has $B_{i,j} = 1$ if the individual is of category j and $B_{i,j} = 0$ otherwise. If B is the matrix with rows $B_{i,\cdot}$, then `R` adds the $K-1$ columns BC to the regression matrix, if C is the current value of the contrast matrix. Thus R creates a parameter for each column in the contrast matrix.

If the $K-1$ columns of C are linearly independent and linearly independent of the constant vector 1, then any vector in $\mathbb{R}^K$ can be written as

$$\mu 1 + C\beta,$$

for some $\mu \in \mathbb{R}$ and $\beta \in \mathbb{R}^{K-1}$. In particular, we could think of representing the "mean vector" of a set of K individuals that take different levels on the categorical variable, but are identical otherwise. We would like to choose the contrasts such that the hypotheses $H_0\colon \beta_j = 0$ express properties of interest.

If $m = \mu 1 + C\beta$ is the mean vector of the set of K individuals and we would like to test a linear hypothesis $H_0\colon d^T m = 0$ for some $d \in \mathbb{R}^K$ with $d^T 1 = 0$, then it is convenient to choose the contrast matrix C to have first

column d and $K-2$ other columns that are orthogonal to d. Then

$$d^T m = d^T(\mu 1 + C\beta) = e_1^T \beta = \beta_1.$$

Thus $H_0: d^T m = 0$ is equivalent to testing $H_0: \beta_1 = 0$. If in R the categorical variable is a column with the name `catvar` in the `data.frame`, then the contrast is set to the matrix C by

```
> contrasts(data$catvar)=C
```

A call to R's `lm` function will next produce this test automatically.

2 Standard choices

There are several standard choices of contrast matrices, none of which are orthogonal. We show them below for $K = 6$.

Treatment is the default. You can change the default with the `options` function.

Treatment: all categories relative to the first one.

```
> contrasts(eggs$Lab)=contr.treatment
> contrasts(eggs$Lab)
    2 3 4 5 6
I   0 0 0 0 0
II  1 0 0 0 0
III 0 1 0 0 0
IV  0 0 1 0 0
V   0 0 0 1 0
VI  0 0 0 0 1
```

For treatment contrasts

$$m = \mu 1 + C\beta = \begin{pmatrix} \mu \\ \mu+\beta_1 \\ \mu+\beta_2 \\ \mu+\beta_3 \\ \mu+\beta_4 \\ \mu+\beta_5 \end{pmatrix} \text{ iff } \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2-m_1 \\ m_3-m_1 \\ m_4-m_1 \\ m_5-m_1 \\ m_6-m_1 \end{pmatrix}.$$

Thus $H_0: \beta_j = 0$ asserts that the mean at the jth level of the category is equal to the mean at the first level: $m_j = m_1$.

By default R takes the alphabetically first level as the base level. The SAS contrast is the treatment contrast, but with the last level as the base level. We can also change the ordering of the levels explicitly.

```
> contrasts(eggs$Lab)=contr.SAS
> contrasts(eggs$Lab)
    1 2 3 4 5
I   1 0 0 0 0
II  0 1 0 0 0
III 0 0 1 0 0
IV  0 0 0 1 0
V   0 0 0 0 1
VI  0 0 0 0 0
> contrasts(eggs$Lab)=contr.treatment
> relevel(eggs$Lab,ref="VI")
# This is suppose to put level VI first, but it does not seem
# to work..
```

Sum: give all categories its own dummy variable with parameter. Keep dummies/parameters 1 to $K-1$, and express the last one as minus the sum of the others.

```
> contrasts(eggs$Lab)=contr.sum
> contrasts(eggs$Lab)
    [,1] [,2] [,3] [,4] [,5]
I      1    0    0    0    0
II     0    1    0    0    0
III    0    0    1    0    0
IV     0    0    0    1    0
V      0    0    0    0    1
VI    -1   -1   -1   -1   -1
```

For sum contrasts,

$$m = \mu 1 + C\beta = \begin{pmatrix} \mu+\beta_1 \\ \mu+\beta_2 \\ \mu+\beta_3 \\ \mu+\beta_4 \\ \mu+\beta_5 \\ \mu-\beta_1-\beta_2-\beta_3-\beta_4-\beta_5 \end{pmatrix} \text{ iff } \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \bar{m}_6 \\ (5/6)(m_1-\bar{m}_{-1}) \\ (5/6)(m_2-\bar{m}_{-2}) \\ (5/6)(m_3-\bar{m}_{-3}) \\ m(5/6)(m_4-\bar{m}_{-4}) \\ (5/6)(m_5-\bar{m}_{-5}) \end{pmatrix}.$$

Here $\bar{m}_6$ is the average of $m_1, \ldots, m_6$ and $\bar{m}_{-j}$ is the average of $(m_s : s \neq j)$. Thus the hypothesis $H_0 : \beta_j = 0$ asserts that m_j is equal to the mean of the other m's. (There is no test for m_6.)

Helmert: kth relative to the average of $1, \ldots, k-1$, scaled by K.

```
> contrasts(eggs$Lab)=contr.helmert
> contrasts(eggs$Lab)
    [,1] [,2] [,3] [,4] [,5]
I     -1   -1   -1   -1   -1
II     1   -1   -1   -1   -1
III    0    2   -1   -1   -1
IV     0    0    3   -1   -1
V      0    0    0    4   -1
VI     0    0    0    0    5
```

For Helmert contrasts, with $\bar{m}_j$ the average of $m_1, m_2, \ldots, m_j$,

$$m = \mu 1 + C\beta = \begin{pmatrix} \mu - \beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5 \\ \mu + \beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5 \\ \mu + 2\beta_2 - \beta_3 - \beta_4 - \beta_5 \\ \mu + 3\beta_3 - \beta_4 - \beta_5 \\ \mu + 4\beta_4 - \beta_5 \\ \mu + 5\beta_5 \end{pmatrix} \text{ iff } \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \bar{m}_6 \\ (m_2 - m_1)/2 \\ (m_3 - \bar{m}_2)/3 \\ (m_4 - \bar{m}_3)/4 \\ (m_5 - \bar{m}_4)/5 \\ (m_6 - \bar{m}_5)/6 \end{pmatrix}.$$

Thus H_0: $\beta_j = 0$ asserts that m_j is equal to the average of $m_1, m_2, \ldots, m_{j-1}$.

Poly:

```
> contrasts(eggs$Lab)=contr.poly
> contrasts(eggs$Lab)
             .L         .Q         .C         ^4          ^5
I   -0.5976143  0.5455447 -0.3726780  0.1889822 -0.06299408
II  -0.3585686 -0.1091089  0.5217492 -0.5669467  0.31497039
III -0.1195229 -0.4364358  0.2981424  0.3779645 -0.62994079
IV   0.1195229 -0.4364358 -0.2981424  0.3779645  0.62994079
V    0.3585686 -0.1091089 -0.5217492 -0.5669467 -0.31497039
VI   0.5976143  0.5455447  0.3726780  0.1889822  0.06299408
>  round(contrasts(eggs$Lab),2)
       .L    .Q    .C    ^4    ^5
I   -0.60  0.55 -0.37  0.19 -0.06
II  -0.36 -0.11  0.52 -0.57  0.31
III -0.12 -0.44  0.30  0.38 -0.63
IV   0.12 -0.44 -0.30  0.38  0.63
V    0.36 -0.11 -0.52 -0.57 -0.31
VI   0.60  0.55  0.37  0.19  0.06
```

The columns of the `poly` contrasts are orthogonal and are (presumably) the values of a set of first, second, etc.. order polynomials that are orthogonalized relative to the empirical measure on a grid of points. As explained in general hypotheses $H_0: \beta_j = 0$ are equivalent to $C_j^T m = 0$ for C_j the jth column of C, corresponding to the jth degree polynomial.

Difference: (my name; it is not standard in R).

```
> C=matrix(1,6,5)
> for (i in 1:5) for (j in i:5) C[i,j]=0
> C
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    0
[2,]    1    0    0    0    0
[3,]    1    1    0    0    0
[4,]    1    1    1    0    0
[5,]    1    1    1    1    0
[6,]    1    1    1    1    1
> contrasts(eggs$Lab)=C
> contrasts(eggs$Lab)
    [,1] [,2] [,3] [,4] [,5]
I      0    0    0    0    0
II     1    0    0    0    0
III    1    1    0    0    0
IV     1    1    1    0    0
V      1    1    1    1    0
VI     1    1    1    1    1
```

For difference contrasts

$$m = \mu 1 + C\beta = \begin{pmatrix} \mu \\ \mu + \beta_1 \\ \mu + \beta_1 + \beta_2 \\ \mu + \beta_1 + \beta_2 + \beta_3 \\ \mu + \beta_1 + \beta_2 + \beta_3 + \beta_4 \\ \mu + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 \end{pmatrix} \text{ iff } \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 - m_1 \\ m_3 - m_2 \\ m_4 - m_3 \\ m_5 - m_4 \\ m_6 - m_5 \end{pmatrix}.$$

So $H_0: \beta_j = 0$ means that there is no difference betwen m_{j+1} and m_j.

3 Some theory

The first lemma shows that any matrix whose columns together with the intercept column span the full space can indeed do what a contrast is sup-

posed to do: represent all possible values of individuals. It also shows that the coding BC is correct.

Lemma 3.1 *Suppose that the columns of C together with the constant vector 1 span $\mathbb{R}^K$. For every vector Y in $\mathbb{R}^n$ such that $Y_i = Y_j$ if $B_{i,\cdot} = B_{j,\cdot}$, there exist unique $\mu \in \mathbb{R}$ and $\beta \in \mathbb{R}^{K-1}$ such that $Y = \mu 1_n + BC\beta$.*

Proof By assumption the $K \times K$-matrix $(1, C)$ has full range. Thus there exists $\beta \in \mathbb{R}^K$ so that $(1, C)\beta$ is equal to the vector that has jth coordinate equal to Y_i for some individual of class j, i.e. $B_{i,\cdot} = e_j^T$, if there is one, and arbitrary if there is no such individual. Then $Y_i = \big((1, C)\beta\big)_j = e_j^T(1, C)\beta = B_{i,\cdot}(1, C)$ for every i, j as in the construction, or $Y = B(1, C)\beta = \beta_1 B1 + (BC)(\beta_2, \ldots, \beta_K)^T$. ∎

The second lemma addresses orthogonality of the columns that are added to the design matrix through contrasts.

Lemma 3.2 *If $a, b \in \mathbb{R}^K$ with $a \perp b$ and the design is balanced (i.e. $n_j := \#(i : B_{i,\cdot} = e_j^T$ is the same for every j), then $Ba \perp Bb$.*

Proof Because $B^T B = \operatorname{diag}(n_1, \ldots, n_K)$, we have $(Ba)^T(Bb) = n_1 a^T b$ in the balanced case. ∎

In particular, if the columns of C are orthogonal to 1_K, and the design is balanced, then the added columns are orthogonal to 1_n. In general, this seems not the case.