If all the values of a particular explanatory variable are set to zero for a given level of a particular factor, then that level is **intentionally aliased**. This sort of aliasing is a useful programming trick in ANCOVA when we wish a covariate to be fitted to some levels of a factor but not to others.

**Orthogonal polynomial contrasts:** `contr.poly`

Here are the data from a randomized experiment with four levels of dietary supplement:

```r
data <- read.table("c:\temp\poly.txt", header=T)
attach(data)
names(data)
[1] "treatment" "response"
```

We begin by noting that the factor levels are in alphabetical order (not in ranked sequence – none, low, medium, high – as we might prefer):

```r
tapply(response, treatment, mean)
```

<table>
<thead>
<tr>
<th></th>
<th>high</th>
<th>low</th>
<th>medium</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>response</td>
<td>4.50</td>
<td>5.25</td>
<td>7.00</td>
<td>2.50</td>
</tr>
</tbody>
</table>

The `summary.lm` table from the one-way analysis of variance looks like this:

```r
model <- lm(response ~ treatment)
summary(model)
```

```
Call:
  lm(formula = response ~ treatment)

Residuals:
  Min     1Q   Median     3Q    Max
-1.250e+00 -5.000e-01 -1.388e-16 5.000e-01 1.000e+00

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.50000   0.37500   12.000   4.84e-08 ***
treatmentlow 0.75000   0.53033    1.414    0.1827

treatmentmedium  2.50000   0.53033    4.714   0.000502 ***
treatmentnone   -2.00000   0.53033   -3.771   0.002666 **

Residual standard error: 0.75 on 12 degrees of freedom
Multiple R-Squared: 0.8606, Adjusted R-squared: 0.8258
F-statistic: 24.7 on 3 and 12 DF, p-value: 2.015e-05
```

The `summary.aov` table looks like this:

```r
summary.aov(model)
```

```
Df   Sum Sq  Mean Sq   F value Pr(>F)
treatment 3  41.687 13.986    24.704 2.015e-05 ***
Residuals 12   6.750   0.563
```

We can see that treatment is a factor but it is not ordered:

```r
is.factor(treatment)
```

[1] TRUE
is.ordered(treatment)

[1] FALSE

To convert it into an ordered factor, we use the ordered function like this:

treatment<-.ordered(treatment,levels=c("none","low","medium","high"))

levels(treatment)

[1] "none" "low" "medium" "high"

Now the factor levels appear in their ordered sequence, rather than in alphabetical order. Fitting the ordered factor makes no difference to the summary.aov table:

model2<-lm(response ~ treatment)

summary.aov(model2)

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>3</td>
<td>41.687</td>
<td>13.896</td>
<td>24.704</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>6.750</td>
<td>0.562</td>
<td></td>
</tr>
</tbody>
</table>

but the summary.lm table is fundamentally different when the factors are ordered. Now the contrasts are not contr.treatment but contr.poly (which stands for orthogonal polynomial contrasts):

summary(model2)

Call:

lm(formula = response ~ treatment)

Residuals:

    Min      1Q  Median      3Q     Max
-1.250e+00 -5.000e-01 -1.596e-16 5.000e-01 1.000e+00

Coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.81250    0.18750   25.667 7.45e-12 ***
treatment.L  1.73300    0.37500    4.621 0.000589 ***
treatment.Q -2.62500    0.37500   -7.000 1.43e-05 ***
treatment.C -0.72667    0.37500   -1.938 0.076520 .

Residual standard error: 0.75 on 12 degrees of freedom
Multiple R-Squared: 0.8606, Adjusted R-squared: 0.8258
F-statistic: 24.7 on 3 and 12 DF, p-value: 2.015e-05

The levels of the factor called treatment are no longer labelled low, medium, none as with treatment contrasts (above). Instead they are labelled L, Q and C, which stand for linear, quadratic and cubic polynomial terms, respectively. But what are the coefficients, and why are they so difficult to interpret? The first thing you notice is that the intercept 4.8125 is no longer one of the treatment means:

tapply(response,treatment, mean)

none   low  medium  high
 2.50   5.25    7.00    4.50

You could fit a polynomial regression model to the mean values of the response with the four levels of treatment represented by a continuous (dummy) explanatory variable (say, x<-c(1, 2, 3, 4)), then fitting terms for $x$, $x^2$ and $x^3$ independently. This is what it would look like: