

Using contrasts to test trends: precautions

Jeanette Mumford

November 19, 2008

Sometimes a study will have multiple trial types where it might be expected that the BOLD signal will increase linearly, quadratically or with some other trend over trial types. For example, there were 3 tasks with increasing difficulty, you might expect the BOLD signal to increase linearly. One approach to studying this linear trend is to first model each trial type separately and then use a contrast to test the trend. For example, say you have 3 trial types, with 1 block of each trial. Ignoring the hemodynamic delay, the hypothesized data, regressors you would use in the model and hypothesized parameter estimates obtained using that model are illustrated in Figure 1. As shown in the third panel of Figure 1 the parameter estimates increase linearly with the difficulty if our hypothesis is true.

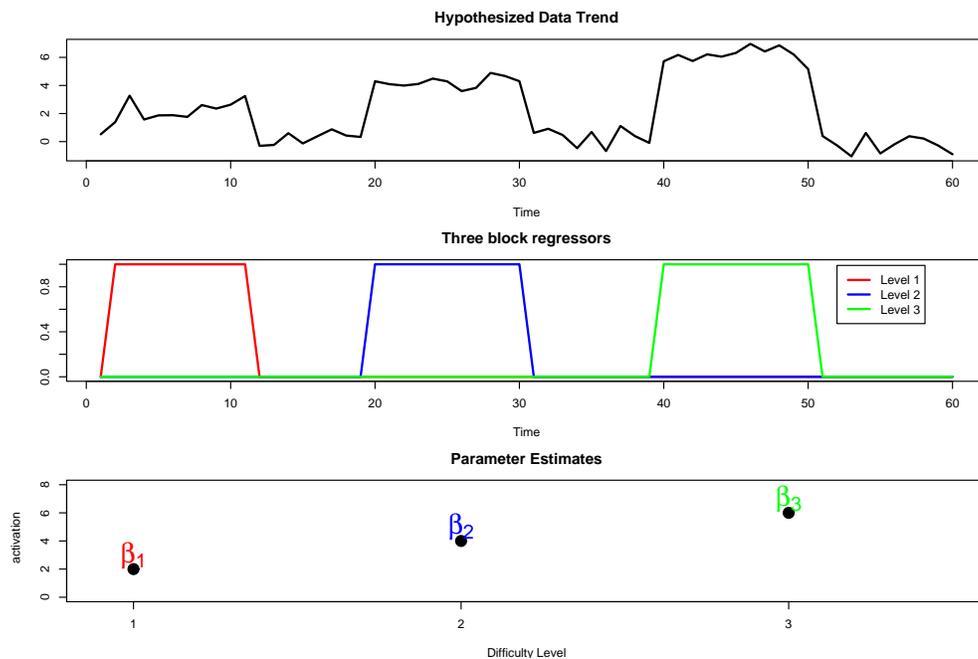


Figure 1: The top panel shows a cartoon of what data from a block design with 3 task levels and a linear increase in BOLD might look like. The middle panel shows the regressors that would be used in the model for each level. The third panel shows the parameter estimates corresponding to the model in the middle panel when describing the data in the top panel.

Assume that we are not only interested in testing if this linear increase exists, but we would also like to study the three tasks separately. A popular choice is to use a linear contrast to test whether there is a linear increase in the β 's. Although your intuition might tempt you

to use the contrast $c = [1, 2, 3]$ to test for a linear trend, the correct contrast is $c = [-1, 0, 1]$, which is obtained by taking $[1, 2, 3] - \text{mean}([1, 2, 3])$. The purpose of subtracting the mean is because we want this contrast to reflect the slope of the line going through the points and not the intercept of that line. In other words we're orthogonalizing the slope to the intercept.

If you have run a lot of models before and are familiar with constructing contrasts, you might be a little uncomfortable with the contrast $c = (-1, 0, 1)$ for two reasons. First, this is typically the contrast used to test the hypothesis $H_0 : \beta_3 = \beta_1$ versus $H_A : \beta_3 > \beta_1$. Second, the calculation of a line going through three points seems like it should consider all three points, right?

If you are a really familiar with statistics you might also find the use of the contrast $c = (-1, 0, 1)$ disturbing because, assuming this contrast is giving us the usual slope, the hypothesis test using this contrast in this way *never* actually compares the fit of this slope to the data. In linear regression we use the residuals to assess how good the model fit is by including the variance of the residuals in the denominator of our test statistic. If the fit is bad, then the residuals will be large and the residual variance, $\hat{\sigma}^2$ will be large causing our test statistic to be small. In this case $\hat{\sigma}^2$ is assessing how well our three block regressors fit the data, not how well the slope of the line fits the 3 parameter estimates. Even if we had 4 levels of difficulty, so 4 regressors and a contrast for a linear trend without zeros, $c = (1, 2, 3, 4) - \text{mean}((1, 2, 3, 4)) = (-1.5, -0.5, 0.5, 1.5)$, although all 4 parameter estimates go into the calculation of the slope, the hypothesis test again does not actually consider how well the points match the fitted line and hence the test statistic is not penalize.

I will now clarify the preceding points with examples. The questions I will answer are:

1. When there is a 0 in the contrast, how can the resulting estimate be the slope?
2. What is does this hypothesis test tell us?
3. Should we be using this? If not, what do we use?

1 Review of the test statistic

The easiest way to understand what is happening is to take a close look at the test statistic that is used. If X is our design matrix (middle panel Figure 1), c is our vector contrast and Y is our time series data (top panel Figure 1), then the contrast estimate is given by

$$c\hat{\beta} = c(X'X)^{-1}X'Y \tag{1}$$

and the test statistic is

$$t = \frac{c\hat{\beta}}{\sqrt{c(X'X)^{-1}c'\hat{\sigma}^2}}. \tag{2}$$

Note that $\hat{\beta}$ is a vector of the individual parameter estimates, $\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2, \dots]$. First we will focus on the numerator of the test statistic, $c\hat{\beta}$, and we will show that when using a linear contrast $c\hat{\beta}$ is actually proportional to the true slope through the parameter estimates. This occurs even when there are 0's in the contrast.

Second we will focus on the denominator of the test statistic. We aren't interested in $c(X'X)^{-1}c'$, this is typically the focus when looking at design efficiency or statistical power. Instead we will focus on $\hat{\sigma}^2$, which is the residual variance. The residual variance reflects the

variability in the data that the model does not describe and is high when the model doesn't fit the data well and low when the fit is good. The important question to ask is whether $\hat{\sigma}^2$ is reflecting how well our line fits our parameter estimates, or if it something else. We'd like it to be high when the line is a bad fit, but we'll show it isn't necessarily true when using linear contrasts.

2 The numerator

Having zeros in the contrast is initially the most disturbing part of this problem. As I said earlier the contrast $(-1, 0, 1)$ arises from demeaning $(1, 2, 3)$, since we don't want our estimate to reflect variability in the data that the intercept would model. It is easier to see that this works through an example. Using the 3 blocks from figure 1 you can simulate some data by adding a little noise and then model that data. Since this is a simulation, I can choose what my parameter estimates should be, so to force a violation of linearity I choose $\beta_1 = 10$, $\beta_2 = -5$, $\beta_3 = 5$ and so the simulated data and linear model fit in R are shown in the following code. Note that the data and design are demeaned, since this is a typical step done in fMRI analysis. Even if you are not familiar with R, this code should be rather straightforward. The function `rnorm` is used to generate random noise and the "-1" in the `lm` call is to remove the intercept, which is unnecessary since everything has been demeaned. You can also run this code yourself, but of course your results will be slightly different since the error is randomly generated.

```
> a = rep(0, 60)
> a[2:11] = 1
> b = rep(0, 60)
> b[20:30] = 1
> c = rep(0, 60)
> c[40:50] = 1
> y = 10 * a - 5 * b + 20 * c + rnorm(60, mean = 0, sd = 5)
> ydm = y - mean(y)
> adm = a - mean(a)
> bdm = b - mean(b)
> cdm = c - mean(c)
> mod1 = (lm(ydm ~ adm + bdm + cdm - 1))
> summary(mod1)
```

Call:

```
lm(formula = ydm ~ adm + bdm + cdm - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-12.8960	-2.9113	0.1982	2.5373	11.7679

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
adm	12.464	1.753	7.111	2.08e-09	***
bdm	-4.499	1.693	-2.658	0.0102	*

```

cdm    18.806      1.693  11.108 6.95e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 4.758 on 57 degrees of freedom
Multiple R-Squared:  0.7709,    Adjusted R-squared:  0.7588
F-statistic: 63.92 on 3 and 57 DF,  p-value: < 2.2e-16

```

As expected, the parameter estimates, $\hat{\beta}_1 = 12.464$, $\hat{\beta}_2 = -4.4992$, $\hat{\beta}_3 = 18.8058$ are somewhat close to the true values specified in our simulation of $\beta_1 = 10$, $\beta_2 = -5$, and $\beta_3 = 20$. Unfortunately R doesn't easily accept contrasts and estimate the corresponding test statistic, but this can be done using equation 2 for the test statistic. To explain what is done, I actually demean the intuitive contrast $c = (1, 2, 3)$ and then I divide by the sum of the squared values. This is just a scaling factor that, as we will see, will result in exactly the slope of the line going through our three points $\hat{\beta}_1 = 12.464$, $\hat{\beta}_2 = -4.4992$, $\hat{\beta}_3 = 18.8058$. If you look at the test statistic above in equation 2 if I were to replace the contrast c by a scaled version ac , then the a in the numerator, which is just a scalar, will cancel out with the a in the denominator. So scaling the contrast will result in exactly the same test statistic.

First will look at the test statistic and p-value using the linear contrast, note that by using `mod1$coeff` I can pull out the parameter estimates from the model I ran above and `%*%` is how matrix multiplication is done in R.

```

> con = c(1, 2, 3)
> con = con - mean(con)
> con = con/sum(con^2)
> con

[1] -0.5  0.0  0.5

> numerator = t(con) %*% mod1$coeff
> numerator

      [,1]
[1,] 3.170921

```

The `numerator` variable is the contrast estimate, $c\hat{\beta} = 3.170921$. Let's compare this value to the slope of the line we'd get by fitting a line through our parameter estimates, in a separate model. Note that the linear regression function in R automatically includes an intercept, which is fine in this case.

```

> x2 = c(1, 2, 3)
> y2 = mod1$coeff
> mod2 = lm(y2 ~ x2)
> summary(mod2)

```

```

Call:
lm(formula = y2 ~ x2)

```

Residuals:

adm	bdm	cdm
6.711	-13.423	6.711

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.582	25.112	0.103	0.935
x2	3.171	11.624	0.273	0.830

Residual standard error: 16.44 on 1 degrees of freedom

Multiple R-Squared: 0.06926, Adjusted R-squared: -0.8615

F-statistic: 0.07441 on 1 and 1 DF, p-value: 0.8305

Notice that the slope parameter in the above model, 3.170921 is the same as the contrast estimate we got above, $\text{numerator} = c\hat{\beta} = 3.170921$. So, it turns out that the numerator of our test statistic is exactly right, even if there is a 0 in the contrast, we get the correct value for the slope. The bottom of panel Figure 2 displays the three parameter estimates and the fitted line. Before we look at the test statistics from the two models above, intuitively, from looking at the figure, it does not appear that there is a linear trend. Although the line is somewhat close to the points, it is clear that these three points do not lie in a straight line. This issue will be discussed in the next section where we focus on the denominator. The main points from this section are

- The linear contrast gives us a value that is proportional to the slope.
- Scaling a contrast by a value results in the same test statistic.
- The numerator of our linear contrast test statistic is equivalent to the numerator of the test statistic obtained from fitting a separate model of a line through the parameter estimates.

3 Denominator

Now that we've established that the numerator of our linear contrast test statistic is perfectly fine, we'll change the focus to the denominator. Specifically the residual variance, $\hat{\sigma}^2$. Generally, this value is related to how close the predicted values are to the data. So, in the case of the test statistic given above in equation 2, where the design X was modeling the 3 blocks shown in figure 1, this residual variance reflects how well the individual blocks fit the data. To reiterate, no matter what contrast you specify, $\hat{\sigma}^2$ in equation 2 reflects how well the design X fits the data. So in the case of fitting a linear contrast $\hat{\sigma}^2$ does *not* reflect how well the fitted line fits your parameter estimates. This is a problem because if your model fit really well, $\hat{\sigma}^2$ will be small and your test statistic, even for slopes close to 0, can be quite large. To illustrate this point, I'll now use R to calculate the p-value for the linear contrast, which we can compare to the p-value in the separate linear model, `mod2`, given above. Since our data had 60 points and 3 regressors, our degrees of freedom are 57, so that's where the 57 comes from below. Also note that `solve` is how you invert matrices in R and `t(X)` is the transpose of a matrix, X .

```

> X = cbind(adm, bdm, cdm)
> denom = t(con) %*% solve(t(X) %*% X) %*% con * sum(mod1$residuals^2)/57
> T = numerator/sqrt(denom)
> pval = pt(T, 57, lower.tail = FALSE)
> T

      [,1]
[1,] 3.050734
> pval

      [,1]
[1,] 0.001730873

```

So, you can see the p-value from using the linear contrast is 0.001730873. This seems highly unlikely when looking at the second panel of Figure 2, since the line doesn't seem very close to the data. The reason the test statistic is so large (p-value is so small) is because the variance in the denominator contains the residual variance, $\hat{\sigma}^2$, from the model fit in the top panel of Figure 2, *not* the residual variance from the fit shown in the bottom panel of Figure 2. Looking back at the linear fit called `mod2` calculated previously to verify the slope, you can see that the p-value using the residual variance based on the linear fit is much larger, $p = 0.83$. Note that although `mod2` is okay for illustrating this point, typically it is a bad idea to estimate a model with so few degrees of freedom. So, if this isn't a possibility, how should we model linear trends in fMRI data analysis?

Luckily in fMRI we typically have many subjects over which we are interested in studying this possible linear trend. The standard modeling approach when there are multiple subjects, each with multiple runs of data, is to use a hierarchical modeling approach where each run is modeled independently in a first level analysis, then runs are combined over subjects using a fixed effects approach in a second level analysis, followed by a mixed effects third level model to combine the data over subjects. So far we have established that the linear trend cannot be directly studied at the first level, due to the possibility of underestimating the p-value. It might be tempting to estimate the linear trend for each subject at the second level, but since a fixed effects model is typically used, the residual variance isn't even calculated. I suggest modeling the linear trend at the third level, making sure to adjust for the repeated measures within subject. For example, if I had 3 subjects my 3rd level model would have 4 regressors: The first 3 would be subject indicators, such as the ones used in a paired t test model, and the last regressor would be the demeaned slope as shown in the following, where the row labels correspond to subject number and level of the stimulus. In this case the contrast you would test to look at the linear effect would be $c = (0, 0, 0, 1)$. Note the dependent variable in this model would be the second level contrast estimates averaging over runs within subject.

$$\begin{array}{l}
S1L1 \\
S1L2 \\
S1L3 \\
S2L1 \\
S2L2 \\
S2L3 \\
S3L1 \\
S3L2 \\
S3L3
\end{array}
\begin{pmatrix}
1 & 0 & 0 & -1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 \\
0 & 1 & 0 & -1 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & -1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 1
\end{pmatrix}$$

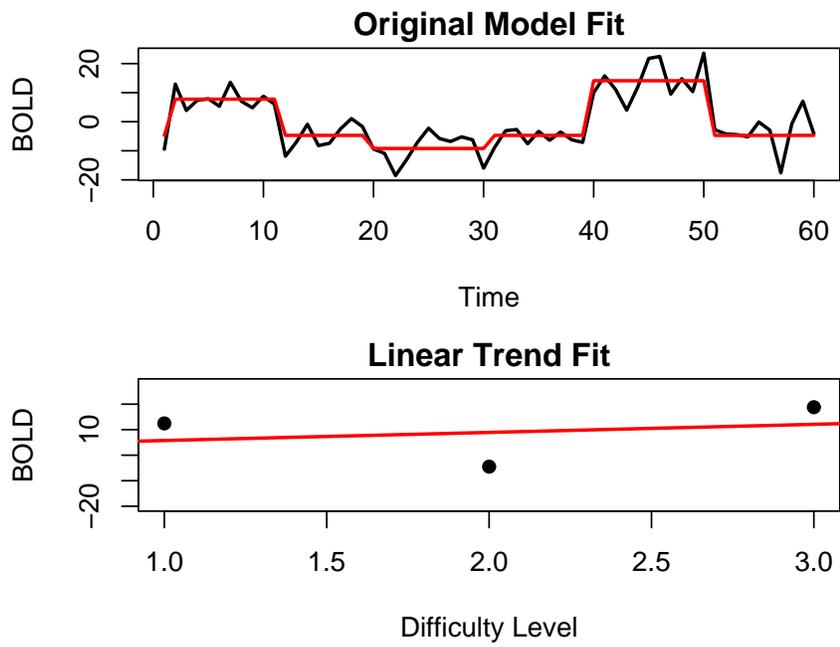


Figure 2: The top panel shows how close the original model fit is to the data. The bottom panel shows that the fit of a line through the three parameter estimates is not a good fit. When using linear contrasts the residual variance from the top panel is used and does not reflect the poor fit seen in the bottom panel.

This is probably the best way to guarantee that your p-values for the linear trends are not under estimated. It could very well be the case that running the model both ways will give similar results, I have yet to find some data to run both of these models for comparison. It might be the case that even if you use the first level linear contrast and take that to higher levels, when the linear fit is poor, presumably it will be poor and variable across subjects and so the between subject variance in the third level will be inflated and will reflect this. Although, it could be the case that the data across subjects consistently look like the example illustrated above, in which case the between-subject variance wouldn't be inflated at the group level and the hypothesis test results using a linear contrast would be misleading.

The main points of this section were

- The denominator of a test statistic contains two variance contributions, where one reflects the variance of the residuals of the original model fit.
- When testing linear contrasts the variance in the denominator of the statistic reflects the fit of your original model, not the linear fit you are interested in. P-values may be artificially low due to this.
- It is safest to wait until the group model to test linear trends using the data across all subjects.