# Statistics
## Categorical predictors (ANOVA) worksheet

**Maarten Speekenbrink** *UCL Experimental Psychology*
**November, 2016**

# 1 General Exercises

1. For the diet and exercise example discussed in the lecture, a researcher comes up with the following contrast codes $\lambda$:

| contrast | none | diet | exercise | both |
|---|---|---|---|---|
| $\lambda_1$ | -3 | 1 | 1 | 1 |
| $\lambda_2$ | 0 | -2 | 1 | 1 |
| $\lambda_3$ | 0 | 0 | -1 | 1 |

   (a) For each group, write down the resulting model for the observations (e.g., $Y_i = \beta_0 + \ldots$)

   *Answer:*

   You can answer this question by filling in the value of the contrast coded predictors in the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

   which are the same for each individual in a particular group. You then get the following predictions for each group:

   None: $Y_i = \beta_0 + \beta_1 \times (-3) + \beta_2 \times 0 + \beta_3 \times 0 = \beta_0 - 3\beta_1 + \epsilon_i$

   Diet: $Y_i = \beta_0 + \beta_1 - 2\beta_2 + \epsilon_i$

   Exercise: $Y_i = \beta_0 + \beta_1 + \beta_2 - \beta_3 + \epsilon_i$

   Both: $Y_i = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \epsilon_i$

   (b) Do the three contrast codes comprise a set of *orthogonal* contrast codes?

   *Answer:* Yes. First check whether the sums for the individual contrast variables are zero:

$$\sum_{k=1}^{m} \lambda_{1k} = -3 + 1 + 1 + 1 = 0$$
$$\sum_{k=1}^{m} \lambda_{2k} = 0 + (-2) + 1 + 1 = 0$$
$$\sum_{k=1}^{m} \lambda_{3k} = 0 + 0 + (-1) + 1 = 0$$

Then check whether the sums of the cross-products are all zero:

$$\sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} = (-3 \times 0) + (1 \times -2) + (1 \times 1) + (1 \times 1) = 0$$
$$\sum_{k=1}^{m} \lambda_{1k}\lambda_{3k} = (-3 \times 0) + (1 \times 0) + (1 \times -1) + (1 \times 1) = 0$$
$$\sum_{k=1}^{m} \lambda_{2k}\lambda_{3k} = (0 \times 0) + (-2 \times 0) + (1 \times -1) + (1 \times 1) = 0$$

(c) Which null hypothesis (in terms of the population means) can be tested by each contrast code?

*Answer:*

You can determine this with the formula

$$\beta_j = \frac{\sum_{k=1}^{m} \lambda_{jk}\mu_k}{\sum_{k=1}^{m} \lambda_{jk}^2}$$

which is the same as the one for $b_j$, but now with the population parameters $\mu$ rather than $\overline{Y}$.

Remember that a test of the slope is a test of the hypothesis $H_0 : \beta_j = 0$. So, for instance, the test of $\beta_1$, which tests the effect of the contrast coded predictor based on $\lambda_1$, is

$$\begin{aligned} 0 &= \frac{\sum_{k=1}^{m} \lambda_{1k}\mu_k}{\sum_{k=1}^{m} \lambda_{1k}^2} \\ 0 &= \frac{-3\mu_n + \mu_d + \mu_e + \mu_b}{(-3)^2 + 1^2 + 1^2 + 1^2} \\ \frac{3\mu_n}{12} &= \frac{\mu_d + \mu_e + \mu_b}{12} \\ 3\mu_n &= \mu_d + \mu_e + \mu_b \\ \mu_n &= \frac{\mu_d + \mu_e + \mu_b}{3} \end{aligned}$$

So the test using contrast code $\lambda_1$ tests whether the mean in the none group is equal to the mean of the other group means. Following the same procedure for the other contrasts gives the following list of tested hypotheses:

$\lambda_1$ tests $H_0 : 3\mu_n = \mu_d + \mu_e + \mu_b$, which is the same as $H_0 : \mu_n = \frac{\mu_d+\mu_e+\mu_b}{3}$

$\lambda_2$ tests $H_0 : 2\mu_d = \mu_e + \mu_b$, which is the same as $H_0 : \mu_d = \frac{\mu_e+\mu_b}{2}$

$\lambda_3$ tests $H_0 : \mu_e = \mu_b$

2. Another researcher is interested in how the effect of diet and exercise over time and uses four groups, one which has not started the diet and exercise programme and three groups which have been following the programme for different durations. To analyse the results, the researcher defines the following contrast codes:

| contrast | none | 4 weeks | 8 weeks | 12 weeks |
|---|---|---|---|---|
| $\lambda_1$ | -1 | 1 | 0 | 0 |
| $\lambda_2$ | 0 | -1 | 1 | 0 |
| $\lambda_3$ | 0 | 0 | -1 | 1 |

(a) Do these contrast codes comprise a set of orthogonal contrast codes?

*Answer:* No. Although the values of the individual variables sum to 0, the pairwise products do not, e.g. $\sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} = (-1\times 0)+(1\times -1)+(0\times 1)+(0\times 0) = -1$

(b) Which null hypotheses (in terms of the population means) do you think the researcher intends to test? How should these tests be performed? You don't actually have to perform the tests, but describe the procedure.

*Answer:* The researcher wants to know whether each additional four weeks makes a difference compared to the previous period. As these contrast codes do not form an orthogonal set, individual tests will be interdependent. As you hopefully know by now, testing the individual effects of a predictor in a model in which the predictor is correlated (partly redundant) with other predictors will give different results than testing the effect of that predictor when it is not correlated with the other predictors in the model. We're testing the unique effects of the predictors, and multi-collinearity usually reduces (but sometimes increases!) the unique effect of predictors. As we have only three contrast codes, the tests would be the same as usual. However, if we had more (dependent) contrast codes, Judd, McClelland & Ryan propose to test each contrast coded predictor as if it was part of an orthogonal set of contrast codes. To do this, we need to have the Sum of Squared Errors (SSE) for a model with three (orthogonal) contrast coded predictors (such as a e.g. a Helmert contrast). Then, for each of the three contrast, you can then compute the Sum of Squares Reduced (SSR), as

$$\text{SSR} = \frac{\left(\sum_{k=1}^{m} \lambda_k \overline{Y}_k\right)^2}{\sum_{k=1}^{m}(\lambda^2/n_k)}$$

You can then compute the $F$ statistic in the usual way, in this case as

$$F = \frac{\text{SSR}/1}{\text{SSE}/(n-4)}$$

as the model with the contrast coded predictors would have four parameters (one intercept, and three slopes). Another option is of course to use Tukey's HSD procedure.

(c) A final researcher believes each additional four weeks has the same effect, regardless of the number of weeks already spent in the programme. Specify a set of orthogonal contrast codes to test this hypothesis.

*Answer:* One option would be to use a polynomial contrast code

| contrast | none | 4 weeks | 8 weeks | 12 weeks |
|---|---|---|---|---|
| $\lambda_1$ | -3 | -1 | 1 | 3 |
| $\lambda_2$ | 1 | -1 | -1 | 1 |
| $\lambda_3$ | -1 | 3 | -3 | 1 |

The assumption then is that the effect of the linear contrast (i.e., that of a predictor $X_1$ with values assigned according to $\lambda_1$) will be significant, but the higher order effects not.

3. A company considers three possible designs for their website. To determine which one is best, they want to evaluate how quickly users can buy a product once they start the purchasing process. They collect the following data for each design:

| Design A Time (sec) | Design B Time (sec) | Design C Time (sec) |
|---|---|---|
| 92 | 86 | 81 |
| 86 | 93 | 80 |
| 87 | 97 | 72 |
| 76 | 81 | 82 |
| 80 | 94 | 83 |
| 87 | 89 | 89 |
| 92 | 98 | 76 |
| 83 | 90 | 88 |
| 84 | 91 | 83 |

The mean purchase times are 85.22, 91.00 and 81.56 for designs A, B and C respectively. To determine the differences between the designs, the company's researcher defines the following contrast codes:

| contrast | Design A | Design B | Design C |
|---|---|---|---|
| $\lambda_1$ | 2 | -1 | -1 |
| $\lambda_2$ | 0 | 1 | -1 |

To test for the effect of Design, the researcher compares a MODEL A with two contrast coded predictors $X_1$ and $X_2$,

$$\text{MODEL A: } \texttt{Time}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

to a MODEL C with only an intercept:

$$\text{MODEL C: } \texttt{Time}_i = \beta_0 + \epsilon_i$$

The Sum of Squared Error of MODEL C is SSE(C) = 1079.85.

Note that the Sum of Squares Reduced (SSR) for a particular contrast from a set of orthogonal contrast codes can be computed as

$$\text{SSR} = \frac{\left(\sum_{k=1}^{m} \lambda_k \overline{Y}_k\right)^2}{\sum_{k=1}^{m} \frac{\lambda_k^2}{n_k}}$$

where $n_k$ is the number of observations in group $k$ (see e.g. page 165 in Judd, McClelland, & Ryan, 2009).

(a) What is the value of SSE(A)?

*Answer:*

For the first contrast:

$$\begin{aligned}\text{SSR} &= \frac{\left(\sum_{k=1}^{m} \lambda_k \overline{Y}_k\right)^2}{\sum_{k=1}^{m}(\lambda_k^2/n_k)} \\ &= \frac{(2 \times 85.22 - 91.00 - 81.56)^2}{(2^2 + (-1)^2 + (-1)^2)/9} = \frac{4.4944}{2/3} = 6.7416\end{aligned}$$

For the second:

$$\begin{aligned}\text{SSR} &= \frac{\left(\sum_{k=1}^{m} \lambda_k \overline{Y}_k\right)^2}{\sum_{k=1}^{m}(\lambda_k^2/n_k)} \\ &= \frac{(91.00 - 81.56)^2}{(1^2 + (-1)^2)/9} = \frac{89.1136}{2/9} = 401.0112\end{aligned}$$

The total SSR equals $6.7416 + 401.0112 = 407.7528$, and SSR = SSE(C)−SSE(A), so SSE(A) $= 1079.85 - 407.7528 = 672.0972$.

(b) What is the value of the Proportional Reduction in Error (PRE) of MODEL A compared to MODEL C?

*Answer:* PRE $= \frac{\text{SSE(C)}-\text{SSE(A)}}{\text{SSE(C)}} = \frac{\text{SSR}}{\text{SSE(C)}} = \frac{407.7528}{1079.85} = 0.3776$

(c) Test whether this PRE is significant (use the conventional significance level of $\alpha = .05$)

*Answer:* $F = \frac{\text{SSR}/(\text{PA}-\text{PC})}{\text{SSE(A)}/(n-\text{PA})} = \frac{407.7528/(3-1)}{672.0972/(27-3)} = 7.28$. $P(F_{2,24} \geq 7.28) = .003$, so the result is significant.

(d) What is the null-hypothesis you have just tested (in terms of both the model parameters and the population means)?

*Answer:* $H_0 : \beta_1 = \beta_2 = 0$ and $H_0 : \mu_A = \mu_B = \mu_C$

(e) Estimate the parameters of MODEL A. How do you interpret the slopes?

*Answer:*

$b_0 = \frac{\sum_{k=1}^{m} \overline{Y}_k}{m} = \frac{85.22+91.00+81.56}{3} = 85.93$

$b_1 = \frac{\sum_{k=1}^{m} \lambda_k \overline{Y}_k}{\sum_{k=1}^{m} \lambda_k^2} = \frac{2\times 85.22-91.00-81.56}{2^2+1^2+1^2} = -0.35$

$b_2 = \frac{91.00-81.56}{1^2+1^2} = 4.72$

$b_0$ is the mean of the three group means. $b_1$ is one third of the difference between the mean of design A and the mean of the means of design B and C, e.g.,

$$b_1 = \frac{2 \times \bar{Y}_A - (\overline{Y}_B + \overline{Y}_C)}{2^2 + 1^2 + 1^2} = \frac{\bar{Y}_A - \frac{\overline{Y}_B + \overline{Y}_C}{2}}{3}$$

where the final expression was obtained by dividing the numerator and denominator by 2. $b_2$ is one half of the difference between the mean of design B and the mean of design C.

## 2 SPSS Exercises

1. Open the datafile `2016_Questionnaire_1_cleaned.sav`. This dataset contains the `Age` (in years), `Height` (in centimetres), `Weight` (in kilograms), and `Gender` (male or female) of those who filled in the Moodle questionnaire. For this exercise, we'll only look at weight and gender.

   (a) As these are likely to be outliers, filter out any cases with a weight equal to or larger than 150.

```
> # read the data
> library(foreign)
> dat <- as.data.frame(read.spss("2016_Questionnaire_1_cleaned.sav"))

Warning in read.spss("2016_Questionnaire_1_cleaned.sav"):  2016_Questionnaire
File-indicated value is different from internal value for at least one
of the three system values.  SYSMIS: indicated -1.79769e+308, expected
-1.79769e+308; HIGHEST: 1.79769e+308, 1.79769e+308; LOWEST: -1.79769e+308,
-1.79769e+308
Warning in read.spss("2016_Questionnaire_1_cleaned.sav"):  2016_Questionnaire
Unrecognized record type 7, subtype 18 encountered in system file

> # filter out large weights
> dat <- subset(dat,Weight < 150)
```

   (b) Compute the overall mean weight, as well as mean weight for males and females separately.

```
> # overall mean
> mean(dat$Weight)

[1] 63.04746

> # first let's redo the factor labels
> levels(dat$Gender) <- c("Female","Male")
> # mean for males
> mean(dat$Weight[dat$Gender == "Male"])

[1] 71.5

> # mean for females
> mean(dat$Weight[dat$Gender == "Female"])

[1] 58.53944

> ## another, somewhat simpler way to get this
> ## is to use the aggregate function:
> aggregate(Weight~Gender,FUN=mean,data=dat)
```

```
  Gender   Weight
1 Female 58.53944
2   Male 71.50000
```

*Answer:* The overall mean is 63.047. The mean for males is 71.5 and the mean for females is 58.539.

(c) Define a new predictor variable, **dummy**, with the value "0" for males, and value "1" for females. Estimate the parameters of the model

$$\texttt{weight}_i = \beta_0 + \beta_1 \texttt{dummy}_i + \epsilon_i$$

and determine whether the effect of **dummy** is significant. Look closely at the parameter estimates. What do these reflect?

```
> # create a dummy variable
> dat$dummy <- NA
> dat$dummy[dat$Gender == "Male"] <- 0
> dat$dummy[dat$Gender == "Female"] <- 1
> # estimate the model
> mod_dummy <- lm(Weight~dummy,data=dat)
> summary(mod_dummy)


Call:
lm(formula = Weight ~ dummy, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-21.500  -6.500  -1.020   4.961  37.461

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   71.500      1.503  47.564  < 2e-16 ***
dummy        -12.961      1.861  -6.963  1.3e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.41 on 136 degrees of freedom
Multiple R-squared:  0.2628,Adjusted R-squared:  0.2574
F-statistic: 48.48 on 1 and 136 DF,  p-value: 1.295e-10
```

*Answer:* The estimated model is

$$\texttt{weight}_i = 71.5 - 12.961 \times \texttt{dummy}_i + e_i$$

The effect of **dummy** is significant. The intercept represents the mean for males, while the slope $b_1$ represents the difference between the means of females and males, i.e., $b_1 = 58.539 - 71.5 = -12.961$.

(d) Define a new predictor variable, `effect`, with the value "-1" for males, and "1" for females. Estimate the parameters of the model

$$\texttt{weight}_i = \beta_0 + \beta_1 \texttt{effect}_i + \epsilon_i$$

and determine whether the effect of `effect` is significant. Look closely at the parameter estimates. What do these reflect?

```
> # effect coding is like dummy coding with a -1 for the reference group
> # so we can re-use the dummy code
> dat$effect <- dat$dummy
> dat$effect[dat$Gender == "Male"] <- -1
> # estimate the model
> mod_effect <- lm(Weight~effect,data=dat)
> summary(mod_effect)


Call:
lm(formula = Weight ~ effect, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-21.500  -6.500  -1.020   4.961  37.461

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  65.0197     0.9307  69.860  < 2e-16 ***
effect       -6.4803     0.9307  -6.963  1.3e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.41 on 136 degrees of freedom
Multiple R-squared:  0.2628,Adjusted R-squared:  0.2574
F-statistic: 48.48 on 1 and 136 DF,  p-value: 1.295e-10
```

*Answer:* The estimated model is

$$\texttt{weight}_i = 65.02 - 6.48 \times \texttt{effect}_i + e_i$$

The effect of `effect` is significant. The intercept represents the mean of the two group means, i.e., $b_0 = \frac{58.539+71.5}{2} = 65.02$. Note that this is not the same as the overall mean; this is because the groups have different sizes. The slope $b_1$ represents half of the difference between the means of males and females, i.e., $b_1 = (58.539 - 71.5)/2 = -6.48$. Note that, because in this case effect coding is an orthogonal contrast (it is the same as a Helmert contrast here), you can work this out using the general formula $b_j = \frac{\sum_k \lambda_{jk} \overline{Y}_k}{\sum_k \lambda_{jk}^2}$

(e) Define a new predictor variable, `effect2`, with the value "-0.5" for males, and "0.5" for females. Estimate the parameters of the model

$$\texttt{weight}_i = \beta_0 + \beta_1 \texttt{effect2}_i + \epsilon_i$$

and determine whether the effect of `effect2` is significant. Look closely at the parameter estimates. What do these reflect?

```
> # effect coding is like dummy coding with a -1 for the reference group
> # so we can re-use the dummy code
> dat$effect2 <- dat$effect/2
> # estimate the model
> mod_effect2 <- lm(Weight~effect2,data=dat)
> summary(mod_effect2)


Call:
lm(formula = Weight ~ effect2, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-21.500  -6.500  -1.020   4.961  37.461 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  65.0197     0.9307  69.860  < 2e-16 ***
effect2     -12.9606     1.8614  -6.963  1.3e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.41 on 136 degrees of freedom
Multiple R-squared:  0.2628,Adjusted R-squared:  0.2574 
F-statistic: 48.48 on 1 and 136 DF,  p-value: 1.295e-10
```

*Answer:* The estimated model is

$$\texttt{weight}_i = 65.02 - 12.961 \times \texttt{effect2}_i + e_i$$

The effect of `effect2` is significant. The interpretation of the intercept is the same as for the model with `effect`. The slope $b_1$ now represents the difference between the means of males and females, i.e., $b_1 = \frac{.5\overline{Y}_F - .5\overline{Y}_M}{.5^2 + (-.5)^2} = \frac{.5\overline{Y}_F - .5\overline{Y}_M}{.5} = \overline{Y}_F - \overline{Y}_M = 58.539 - 71.5 = -12.961$. So by using values of -.5 and .5, the slope now represents the difference between the group means (just like with dummy coding), not half the difference between the group means (as in the previous effect coding).

(f) When you look at the significance of the effects of `dummy`, `effect` and `effect2` in the three models above, what do you see?

*Answer:* The test results ($p$ and $F$ or $t$ values) are the same no matter what coding scheme is used. With `dummy` and `effect2` you tested whether the difference between the average female and male weight equals 0, while with `effect` you tested whether half the difference between the average female and male weight equals 0. That you get the same results with any (proper) form of contrast coding is always the case for two groups. More generally, it is also the case for any *omnibus* test of the slopes for all predictors that code for group membership, although in that case the tests for the slopes of the individual contrast coded predictors will differ depending on the specific contrasts, as they are testing specific differences between the means. Finally, for the case of two groups, there is only one contrast coded predictor and the test of the single slope is the same as the omnibus test.

2. Let's consider the data from General Exercise 3 again. You can find the data on Moodle as `WebDesign.sav`.

   (a) Use a oneway ANOVA to test whether there significant differences between the designs. If there are, perform a Tukey's HSD test to find exactly what the differences are.

```
> # load the data
> dat <- as.data.frame(read.spss("WebDesign.sav"))
> # let's fix the factor labels again
> levels(dat$design) <- c("A","B","C")
> # oneway ANOVA
> web_aov <- aov(time~design,data=dat)
> summary(web_aov)

            Df Sum Sq Mean Sq F value  Pr(>F)
design       2  408.1  204.04   7.289 0.00336 **
Residuals   24  671.8   27.99
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # Tukey HSD tests
> TukeyHSD(web_aov)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = time ~ design, data = dat)

$design
         diff         lwr       upr     p adj
B-A  5.777778  -0.4505207 12.006076 0.0725971
C-A -3.666667  -9.8949651  2.561632 0.3225832
C-B -9.444444 -15.6727429 -3.216146 0.0025028
```

*Answer:* The ANOVA gives an omnibus test of $F_{2,24} = 7.289$, $p = .003$, so there is a significant difference between the designs. Tukey's HSD shows the difference is between designs B and C.

(b) Specify the dummy coded variables needed to code for Design, and test for the effect of Design with multiple regression. Compare the $F$ value in the ANOVA table to the one obtained earlier. Are they the same?

```
> dat$dummy1 <- 0
> dat$dummy1[dat$design == "B"] <- 1
> dat$dummy2 <- 0
> dat$dummy2[dat$design == "C"] <- 1
> mod_web <- lm(time~design,data=dat)
> summary(mod_web)


Call:
lm(formula = time ~ design, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-10.0000  -2.1111   0.4444   2.5000   7.4444

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   85.222      1.764  48.324   <2e-16 ***
designB        5.778      2.494   2.317   0.0294 *
designC       -3.667      2.494  -1.470   0.1545
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.291 on 24 degrees of freedom
Multiple R-squared:  0.3779,Adjusted R-squared:  0.3261
F-statistic: 7.289 on 2 and 24 DF,  p-value: 0.00336
```

*Answer:* Yes, of course they are! The omnibus test (i.e., the test comparing the model with the two dummy coded predictors to a model with only an intercept) is testing the same null-hypothesis as an ANOVA. And, as you know, the Generalized Linear Model is basically what's going on behind the scenes in ANOVA (hence why SPSS now has a GLM routine).

(c) Look at the parameters of the model you have just estimated. How should you interpret the values?

*Answer:* I have used the following contrast variables:

| contrast | A | B | C |
|---|---|---|---|
| $\lambda_1$ | 0 | 1 | 0 |
| $\lambda_2$ | 0 | 0 | 1 |

So group A is the reference group and the value of the intercept $b_0 = 85.22$ is identical to the mean in group A. The value of $b_1 = 5.778$ is identical to the difference between group B and group A, and the value of $b_2 = -3.667$ is identical to the difference between the mean in group C and group A.

(d) Calculate the correlation between the dummy variables; are they independent predictors? Could tests of their individual significance replace something like Tukey's HSD test?

```
> cor(dat[,c("dummy1","dummy2")])

       dummy1 dummy2
dummy1    1.0   -0.5
dummy2   -0.5    1.0
```

*Answer:* The correlation is -.50, which is significant. So the predictors are dependent. But that isn't necessarily a huge problem. However, with dummy coding, we can't include a third dummy predictor to code for the difference between group B and C. If you want to test *all* pairwise differences, you will need something like Tukey's HSD. If your happy to test just two differences, then looking at the individual predictors is fine (and more powerful than Tukey's HSD). Of course, as you are performing two tests, the family-wise significance level (i.e., the probability of making at least one Type I error when performing a set of tests) will be higher than the significance level of each individual test. That is why people recommend performing the omnibus test first (which keeps the family-wise significance level at the desired value), and when this is significant, look for the significance of the individual predictors. This can provide some protection against inflation of the Type I error rate.

3. To test whether memory changes with age, a researcher has done a memory test on four different age groups. The results are:

| 30 yrs old | 40 yrs old | 50 yrs old | 60 yrs old |
|---|---|---|---|
| 14 | 12 | 17 | 13 |
| 13 | 14 | 14 | 10 |
| 15 | 16 | 14 | 7 |
| 17 | 11 | 9 | 8 |
| 12 | 12 | 13 | 6 |
| 10 | 18 | 15 | 9 |

You can also find this data on Moodle as `AgeMemory.sav`.

(a) Specify a set of appropriate contrast coded predictors and assess whether there is an effect of age using multiple regression.

```
> # load the data
> dat <- as.data.frame(read.spss("AgeMemory.sav"))
> # set-up Helmert coded contrasts
> dat$helmert1 <- -1
> dat$helmert1[dat$age == 30] <- 3
> dat$helmert2 <- 0
> dat$helmert2[dat$age == 40] <- 2
> dat$helmert2[dat$age > 40] <- -1
> dat$helmert3 <- 0
> dat$helmert3[dat$age == 50] <- 1
> dat$helmert3[dat$age > 50] <- -1
>
> mod_age_helmert <- lm(score~helmert1 + helmert2 + helmert3,data=dat)
> summary(mod_age_helmert)


Call:
lm(formula = score ~ helmert1 + helmert2 + helmert3, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6667 -1.8333  0.1667  1.3750  4.1667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4583     0.5254  23.712 4.09e-16 ***
helmert1      0.3472     0.3033   1.145  0.26586
helmert2      0.8611     0.4290   2.007  0.05842 .
helmert3      2.4167     0.7430   3.252  0.00399 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.574 on 20 degrees of freedom
Multiple R-squared:  0.4432,Adjusted R-squared:  0.3597
F-statistic: 5.306 on 3 and 20 DF,  p-value: 0.007442
```

*Answer:* We could for instance use a Helmert contrast:

| contrast | 30 years | 40 years | 50 years | 60 years |
|---|---|---|---|---|
| $\lambda_1$ | 3 | -1 | -1 | -1 |
| $\lambda_2$ | 0 | 2 | -1 | -1 |
| $\lambda_3$ | 0 | 0 | 1 | -1 |

We then get the model $Y_i = 12.458 + 0.347 X_{1i} + 0.861 X_{2i} + 2.417 X_{3i} + e_i$, with an $R^2 = .443$ and $F_{3,20} = 5.306$, $p = .007$ (this test compares this model to one with only an intercept). Neither the individual effect of $X_1$ or $X_2$ is significant (although the effect of $X_2$ is *marginally* significant). The individual effect of $X_3$

is significant, so there is a significant difference between 50 and 60 year olds.

(b) What is the value of the PRE of the model compared to one with only an intercept? What is the value of $\hat{\eta}^2$, the unbiased estimate of the population value of the PRE?

```
> # get the SS terms
> tmp <- anova(lm(score~1,data=dat),mod_age_helmert)
> tmp

Analysis of Variance Table

Model 1: score ~ 1
Model 2: score ~ helmert1 + helmert2 + helmert3
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     23 237.96
2     20 132.50  3    105.46 5.3061 0.007442 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> ## compute the PRE using this object
> PRE <- tmp$"Sum of Sq"[2]/tmp$RSS[1]
> PRE

[1] 0.4431798

> ## compute the unbiased estimate
> 1-((1-PRE)*(24-1))/(24-4)

[1] 0.3596568
```

*Answer:* PRE $= \frac{\text{SSE(C)}-\text{SSE(A)}}{\text{SSE(C)}} = \frac{\text{SSR}}{\text{SSE}} = \frac{105.458}{237.598} = 0.444$. Note that you can get the value of SSR from an ANOVA table as the SS for the effect of Age, and the value of SSE(C) as the "SS Total" (or "SS Corrected Total" if you use the GLM option (in GLM, the SS Total is the SSE of a model which fixes the intercept to $B_0 = 0$).

The unbiased estimate of $\eta^2$ is

$$\hat{\eta}^2 = 1 - \frac{(1 - \text{PRE})(n - \text{PC})}{n - \text{PA}} = 1 - \frac{(1 - .444)(24 - 1)}{24 - 4} = .361$$

. Up to rounding error, this is identical to the adjusted-$R^2$ shown in the output.

(c) Repeat the analysis above, now with polynomial contrast codes (if you used a polynomial contrast code above, you can use a Helmert contrast here).

```
> dat$poly1 <- dat$poly2 <- dat$poly3 <- NA
> # linear contrast
> dat$poly1[dat$age==30] <- -3
> dat$poly1[dat$age==40] <- -1
```

```
> dat$poly1[dat$age==50] <- 1
> dat$poly1[dat$age==60] <- 3
> # quadratic contrast
> dat$poly2[dat$age==30] <- 1
> dat$poly2[dat$age==40] <- -1
> dat$poly2[dat$age==50] <- -1
> dat$poly2[dat$age==60] <- 1
> # cubic contrast
> dat$poly3[dat$age==30] <- -1
> dat$poly3[dat$age==40] <- 3
> dat$poly3[dat$age==50] <- -3
> dat$poly3[dat$age==60] <- 1
> # estimate the model
> mod_age_poly <- lm(score~poly1+poly2+poly3,data=dat)
> summary(mod_age_poly)


Call:
lm(formula = score ~ poly1 + poly2 + poly3, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6667 -1.8333  0.1667  1.3750  4.1667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4583     0.5254  23.712 4.09e-16 ***
poly1        -0.7083     0.2350  -3.015  0.00685 **
poly2        -1.2917     0.5254  -2.458  0.02320 *
poly3        -0.2083     0.2350  -0.887  0.38580
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.574 on 20 degrees of freedom
Multiple R-squared:  0.4432,Adjusted R-squared:  0.3597
F-statistic: 5.306 on 3 and 20 DF,  p-value: 0.007442

> ## a quicker way to set up the desired contrast codes
> ## is to use the contrasts() function, and then extract
> ## the contrast codes using the model.matrix() function
> dat$age_factor <- factor(dat$age)
> contrasts(dat$age_factor) <- cbind(c(-3,-1,1,3),
+                                    c(1,-1,-1,1),
+                                    c(-1,3,-3,1))
> poly_contrast <- model.matrix(score~age_factor,data=dat)
```

```
> ## note that poly_contrast also has a column for the intercept:
> head(poly_contrast)

  (Intercept) age_factor1 age_factor2 age_factor3
1           1          -3           1          -1
2           1          -3           1          -1
3           1          -3           1          -1
4           1          -3           1          -1
5           1          -3           1          -1
6           1          -3           1          -1

> ## you can assign multiple variables in dat in one go
> dat[,c("poly1","poly2","poly3")] <- poly_contrast[,2:4]
> ## let's re-estimate the model to see whether the contrast
> ## variables were computed correctly
> summary(lm(score~poly1+poly2+poly3,data=dat))


Call:
lm(formula = score ~ poly1 + poly2 + poly3, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6667 -1.8333  0.1667  1.3750  4.1667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4583     0.5254  23.712 4.09e-16 ***
poly1        -0.7083     0.2350  -3.015  0.00685 **
poly2        -1.2917     0.5254  -2.458  0.02320 *
poly3        -0.2083     0.2350  -0.887  0.38580
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.574 on 20 degrees of freedom
Multiple R-squared:  0.4432,Adjusted R-squared:  0.3597
F-statistic: 5.306 on 3 and 20 DF,  p-value: 0.007442
```

*Answer:* The polynomial contrast variables are:

| contrast | 30 years | 40 years | 50 years | 60 years |
|---|---|---|---|---|
| $\lambda_1$ | -3 | -1 | 1 | 3 |
| $\lambda_2$ | 1 | -1 | -1 | 1 |
| $\lambda_3$ | -1 | 3 | -3 | 1 |

Using these, we get the model $Y_i = 12.458 - 0.708X_{1i} - 1.292X_{2i} - .208X_{3i} + e_i$, again with an $R^2 = .443$ and $F_{3,20} = 5.306$, $p = .007$. The effect of the linear contrast ($X_1$) is significant, as is the effect of the quadratic contrast ($X_2$). The effect of the cubic contrast is not significant.

(d) Treat age as a metric variable (i.e., include age "as is" rather than by a set of contrast codes). Test for the effect of age with linear regression. Compare the result to those obtained in the polynomial contrast analysis.

```
> mod_age_linear <- lm(score~age,data=dat)
> summary(mod_age_linear)


Call:
lm(formula = score ~ age, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5833 -2.2083 -0.4583  2.2917  5.2500 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  18.8333     2.4063   7.827 8.48e-08 ***
age          -0.1417     0.0519  -2.730   0.0122 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.842 on 22 degrees of freedom
Multiple R-squared:  0.253,Adjusted R-squared:  0.2191 
F-statistic: 7.452 on 1 and 22 DF,  p-value: 0.01223
```

*Answer:* The model is $Y_i = 18.833 - 0.142\texttt{age}_i + e_i$, with an $R^2 = .253$, and $F_{1,22} = 7.452$, $p = .012$. Note that this test result is different than that for the linear contrast variable. The reason for this is that in contrast analysis, we also included a quadratic and cubic contrast in the model. As a result, the SSE of the contrast model is smaller than the SSE of the model estimated here. Also, you can see that the slope for `age` is different than the slope for the linear contrast, which is of course due to the fact that `age` has a different scale than the linear contrast.

4. To test whether learning interferes with memory, 36 students were divided into six groups and asked to memorize 16 nonsense syllable pairs. They are rehearsed either 4, 8 or 12 times before they learn some new material. The new material is either some more nonsense pairs or some number pairs. They are then tested for recall of the original data. The results are:

| Other Material | Repetitions 4 | | 8 | | 12 | |
|---|---|---|---|---|---|---|
| **Numbers** | 10 | 11 | 16 | 12 | 16 | 14 |
| | 12 | 15 | 11 | 15 | 16 | 13 |
| | 14 | 10 | 13 | 14 | 15 | 16 |
| **Syllables** | 8 | 7 | 11 | 13 | 14 | 12 |
| | 4 | 5 | 9 | 10 | 16 | 15 |
| | 5 | 6 | 8 | 9 | 12 | 13 |

You can also find the data on Moodle as `LearnMemory.sav`.

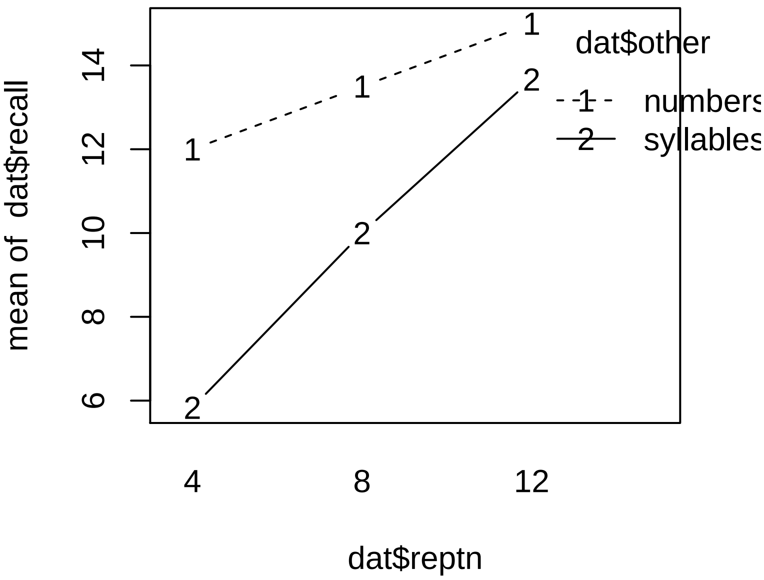
(a) Perform a factorial ANOVA. What are the null hypotheses you test? Interpret the results.

```
> # load the data
> dat <- as.data.frame(read.spss("LearnMemory.sav"))
> # turn reptn into a factor
> dat$reptn <- factor(dat$reptn)
> library(car)
> aov_learn <- aov(recall~other*reptn,data=dat)
> Anova(aov_learn,type="III")

Anova Table (Type III tests)

Response: recall
            Sum Sq Df  F value    Pr(>F)
(Intercept) 864.00  1 295.6654 < 2.2e-16 ***
other       114.08  1  39.0399 6.982e-07 ***
reptn        27.00  2   4.6198   0.01782 *
other:reptn  35.17  2   6.0171   0.00635 **
Residuals    87.67 30
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # a simple plot to interpret the interaction
> interaction.plot(dat$reptn,dat$other,dat$recall,type="b")
```




*Answer:*
In terms of population means, you can state the null hypotheses as follows. First, define the average in the Numbers groups as $\mu_{N\cdot} = \frac{\mu_{N,4}+\mu_{N,8}+\mu_{N,12}}{3}$ and the average in the Syllables group as $\mu_{S\cdot} = \frac{\mu_{S,4}+\mu_{S,8}+\mu_{S,12}}{3}$. We can then state the null hypothesis of no main effect of Other Material as

$$H_0 : \mu_{N\cdot} = \mu_{S\cdot}$$

For the main effect of Repetitions, we can first define similar means as $\mu_{\cdot 4} = \frac{\mu_{N,4}+\mu_{S,4}}{2}$, $\mu_{\cdot 8} = \frac{\mu_{N,8}+\mu_{S,8}}{2}$ and $\mu_{\cdot 12} = \frac{\mu_{N,12}+\mu_{S,12}}{2}$, and then state the null hypothesis of no main effect of Repetitions as

$$H_0 : \mu_{\cdot 4} = \mu_{\cdot 8} = \mu_{\cdot 12}$$

The test for the interaction test is a little trickier to define in terms of group means. Effectively, if there is no interaction, then a group mean would be the sum of the main effects of Other Material and Repetition. We can define a main effect as a deviation from the overall ("grand") mean. This overall mean can be defined as $\mu_{\cdot\cdot} = \frac{\mu_{N,4}+\mu_{N,8}+\mu_{N,12}+\mu_{S,4}+\mu_{S,8}+\mu_{S,12}}{6}$. A main effect of e.g. Other Material would indicate that $\mu_{N\cdot}$ differs from $\mu_{\cdot\cdot}$. So we can define a main effect as e.g.

$$\mu_{N\cdot} - \mu_{\cdot\cdot}$$

So there is no interaction if e.g.

$$\mu_{N,4} - \mu_{\cdot\cdot} = (\mu_{N\cdot} - \mu_{\cdot\cdot}) + (\mu_{\cdot 4} - \mu_{\cdot\cdot})$$
$$\mu_{N,4} = \mu_{N\cdot} + \mu_{\cdot 4} - \mu_{\cdot\cdot}$$

The null hypothesis of no interaction covers all group means, so

$$\begin{aligned} H_0: \quad \mu_{N,4} &= \mu_{N\cdot} + \mu_{\cdot 4} - \mu_{\cdot\cdot} \\ \mu_{N,8} &= \mu_{N\cdot} + \mu_{\cdot 8} - \mu_{\cdot\cdot} \\ \mu_{N,12} &= \mu_{N\cdot} + \mu_{\cdot 12} - \mu_{\cdot\cdot} \\ \mu_{S,4} &= \mu_{S\cdot} + \mu_{\cdot 4} - \mu_{\cdot\cdot} \\ \mu_{S,8} &= \mu_{S\cdot} + \mu_{\cdot 8} - \mu_{\cdot\cdot} \\ \mu_{S,12} &= \mu_{S\cdot} + \mu_{\cdot 12} - \mu_{\cdot\cdot} \end{aligned}$$

Performing a factorial ANOVA gives the following results:

| Source | SS | *df* | MS | $F$ | $p$ |
|---|---|---|---|---|---|
| Other | 121 | 1 | 121 | 41.407 | $< .001$ |
| Repetitions | 176.167 | 2 | 88.083 | 30.143 | $< .001$ |
| Interaction | 35.167 | 2 | 17.583 | 6.017 | .006 |
| Error | 87.667 | 30 | 2.922 | | |
| Total | 420 | 35 | | | |

So there is a difference in recall for numbers and syllables as other material, there is a difference in recall for at least two levels of repetitions, and there is an interaction between other material and repetitions (the effect of repetitions is different for syllables and numbers).

(b) Specify the contrast coded predictors needed to test for the main effects and interaction effect. Repeat the analysis with multiple regression. Compare the results to those obtained above.

```
> # I'll use the trick I showed earlier to do this
> contrasts(dat$other) <- c(-1,1)
> contrasts(dat$reptn) <- cbind(c(-1,0,1),
+                               c(-1,2,-1))
> dat[,c("c_other","lin_rep","quad_rep","other_lin","other_quad")] <-
+   model.matrix(recall~other*reptn,data=dat)[,2:6]
> mod_learn <- lm(recall ~ c_other + lin_rep + quad_rep + other_lin +
+                   other_quad,data=dat)
> summary(mod_learn)


Call:
lm(formula = recall ~ c_other + lin_rep + quad_rep + other_lin +
    other_quad, data = dat)

Residuals:
   Min     1Q Median     3Q    Max
-2.500 -1.125  0.000  1.042  3.000
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.66667    0.28491  40.949  < 2e-16 ***
c_other     -1.83333    0.28491  -6.435 4.16e-07 ***
lin_rep      2.70833    0.34894   7.762 1.17e-08 ***
quad_rep     0.04167    0.20146   0.207  0.83755
other_lin    1.20833    0.34894   3.463  0.00163 **
other_quad   0.04167    0.20146   0.207  0.83755
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.709 on 30 degrees of freedom
Multiple R-squared:  0.7913,Adjusted R-squared:  0.7565
F-statistic: 22.75 on 5 and 30 DF,  p-value: 2.21e-09

> ## get the omnibus tests
> ## I'm using the update function to delete terms from mod_learn
> # main effect of other:
> anova(update(mod_learn,.~.-c_other),mod_learn)

Analysis of Variance Table

Model 1: recall ~ lin_rep + quad_rep + other_lin + other_quad
Model 2: recall ~ c_other + lin_rep + quad_rep + other_lin + other_quad
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     31 208.667
2     30  87.667  1       121 41.407 4.163e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # main effect of reptn:
> anova(update(mod_learn,.~.-lin_rep - quad_rep),mod_learn)

Analysis of Variance Table

Model 1: recall ~ c_other + other_lin + other_quad
Model 2: recall ~ c_other + lin_rep + quad_rep + other_lin + other_quad
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     32 263.833
2     30  87.667  2    176.17 30.143 6.646e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # interaction effect of other:
> anova(update(mod_learn,.~.-other_lin - other_quad),mod_learn)
```

```
Analysis of Variance Table

Model 1: recall ~ c_other + lin_rep + quad_rep
Model 2: recall ~ c_other + lin_rep + quad_rep + other_lin + other_quad
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     32 122.833
2     30  87.667  2    35.167 6.0171 0.00635 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*Answer:* I used a Helmert contrast for Other ($\lambda_1$). As the levels of Repetitions are ordered, I used a polynomial contrast for Repetitions ($\lambda_2$ and $\lambda_3$). To get the 2 contrasts for the interaction effect ($\lambda_4$ and $\lambda_5$), you just multiply the values of the main effect contrasts (e.g., $\lambda_4 = \lambda_1 \times \lambda_2$, $\lambda_5 = \lambda_1 \times \lambda_3$). You then end up with the following set of orthogonal contrast variables:

| contrast | N,4 | N,8 | N,12 | S,4 | S,8 | S,12 |
|---|---|---|---|---|---|---|
| $\lambda_1$ | -1 | -1 | -1 | 1 | 1 | 1 |
| $\lambda_2$ | -1 | 0 | 1 | -1 | 0 | 1 |
| $\lambda_3$ | -1 | 2 | -1 | -1 | 2 | -1 |
| $\lambda_4$ | 1 | 0 | -1 | -1 | 0 | 1 |
| $\lambda_5$ | 1 | -2 | 1 | -1 | 2 | -1 |

Of course, the results are the same when using multiple regression, but we get more specific tests as well. Putting the results in an ANOVA source table gives:

| Source | $b$ | SS | $df$ | MS | $F$ | $p$ |
|---|---|---|---|---|---|---|
| Model | | 332.333 | 5 | 66.467 | 22.745 | $< .001$ |
| Other | | 121 | 1 | 121 | 41.407 | $< .001$ |
| $X_1$ | -1.833 | 121 | 1 | 121 | 41.407 | $< .001$ |
| Repetitions | | 176.167 | 2 | 88.083 | 30.143 | $< .001$ |
| $X_2$ | 2.708 | 176.042 | 1 | 176.042 | 60.242 | $< .001$ |
| $X_3$ | 0.042 | 0.125 | 1 | 0.125 | 0.043 | .838 |
| Interaction | | 35.167 | 2 | 17.583 | 6.017 | .006 |
| $X_4$ | 1.208 | 35.042 | 1 | 35.042 | 11.991 | .002 |
| $X_5$ | 0.042 | 0.125 | 1 | 0.125 | 0.043 | .838 |
| Error | | 87.667 | 30 | 2.922 | | |
| Total | | 420 | 35 | | | |

Note that it is handy to use the SPSS GLM option to get (part of) these results (in particular the SSR terms; you can actually also obtain the estimates of the slopes by asking for parameter estimates). If you use the SPSS GLM option with your own contrast coded predictors, you have to enter these as "covariates" rather than fixed factors. As before, we find significant main effects and a significant interaction. The nice thing here is that we get more information. For instance, we see that the slope of $X_1$ is negative. Looking at the values of $\lambda_1$, we then know that recall is higher when the other material is numbers as compared to syllables (in fact, when the other material is numbers, participants recall on average $2 \times 1.833 = 3.666$ more pairs than when the other material is syllables). So there is more interference when the new material has the same format as the old material. Another interesting thing is the significant effect of $X_2$, which is the linear effect of Repetitions, and the non-significant effect of $X_3$, which is the quadratic effect of Repetitions. So each additional 4 repetitions seems to have the same effect on recall (resulting in 2.708 more recalled syllables). Also, the effect of $X_4$ is significant, so that the slope of the linear effect of Repetitions is moderated by Other Material. In this case, the parameter $b_4$ is positive. Looking at the values of $\lambda_4$, you should notice that these are the same as those of $\lambda_2$ for the syllables group, but the opposite of $\lambda_2$ for the numbers group. As the interaction parameter is positive, this means that the slope is larger for the syllables group than for the numbers group (in the numbers group, the model effectively subtracts something from the slope, while something is added for the syllables group). Try creating a plot of recall by repetitions, with different lines for the other Material conditions.

In summary, not only have we exactly replicated the results of the factorial ANOVA, we now also know how to explain the significant effects we found there.