# RMS2 Week 7: Factorial ANOVA in lm with Dummy Codes

## Problem Set & Lab

This week’s problem set and lab ask you to follow through the same steps as we did in lecture for a larger factorial ANOVA analysis (3x2) with dummy coded variables. There is comparatively little R analysis in this lab, but a lot of thinking about the structure of the model, and using an example to understand the meaning of the beta coefficients. **Hint: You may well have some exam questions of this type.**

Data

**The data comes from a study into patient care in a paediatric wards. A researcher was interested in whether the subjective well-being of patients differed dependent on the post-operation treatment schedule they were given, and the hospital in which they were staying.**

**Three treatments were evaluated (Factor 1: TreatA, TreatB, TreatC) in two hospitals (Factor 2: Hosp1, Hosp2). Thirty patients were tested in each of the resulting 6 groups. The total sample was 180. The outcome was subjective well-being rating (SWB) as an average of multiple raters (the patient, a member of their family, and a friend). The SWB score ranged from 0 to 20.**

Q1: Classification table of means

**Read in the data from LEARN and compute the descriptive statistics split by factors in order to gain the cell means. Populate the classification table below (note the size of the table gives you a clue here).**

# Read data

lab <- read.csv("C:/Users/Tom Booth/Desktop/Weeks 6 to 8/lab7.csv", header = T)

# Descriptives by Group

library(psych)

describeBy(lab$SWB, group = list(lab$Treatment, lab$Hospital), mat = T)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Hospital 1 | Hospital 2 |  |
| Treatment A | 10.80 | 7.87 | 9.34 |
| Treatment B | 9.43 | 13.11 | 11.27 |
| Treatment C | 10.11 | 7.98 | 9.05 |
|  | 10.11 | 9.65 | 9.88 |

Q2: Table of dummy codes

**When running the analysis, R will produce dummy codes for you, but it is important to practice setting these up. Complete the table below with a full set of dummy coded variables. Again, the size of the table should give you a clue as to how many you need (remember (r-1), (c-1) and (r-1)(c-1)). The first two columns of the table should contain the levels of your two factors.**

OK, so we need (r-1) for treatments. So 3-1 = 2 dummies.

We need (c-1) for hospital. So 2-1 = 1 dummy.

And we need (r-1)(c-1) for the interactions. So (3-1)(2-1) = 2\*1 = 2 dummies.

We will set Treatment A and Hospital 1 as the references.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Column** | **T1** | **T2** | **H1** | **T1H1** | **T2H1** |
| Treat A | Hosp 1 | 0 | 0 | 0 | 0 | 0 |
| Treat B | Hosp 1 | 1 | 0 | 0 | 0 | 0 |
| Treat C | Hosp 1 | 0 | 1 | 0 | 0 | 0 |
| Treat A | Hosp 2 | 0 | 0 | 1 | 0 | 0 |
| Treat B | Hosp 2 | 1 | 0 | 1 | 1 | 0 |
| Treat C | Hosp 2 | 0 | 1 | 1 | 0 | 1 |

Q3: Specification of the full linear model

**Write out the full linear model using the labels you gave for your dummy variables above.**

Q4: Expressing cell means in terms of beta parameters

**OK, now things get trickier. Using your specification in Q3, and the weights you gave to the dummies in Q2, express each of the cell means in terms of the parameters in the model.**

**Hint: Look again at the examples in the lecture. Start with plugging in ones and zeros, and go from there.**

|  |  |  |
| --- | --- | --- |
|  | = |  |
|  | = |  |
|  | = |  |
|  |  |  |
|  | = |  |
|  | = |  |
|  | = |  |
|  |  |  |
|  | = |  |
|  | = |  |
|  | = |  |
|  |  |  |
|  | = |  |
|  | = |  |
|  | = |  |
|  |  |  |
|  | = |  |
|  | = |  |
|  | = |  |
|  |  |  |
|  | = |  |
|  | = |  |
|  | = |  |

Q5: Null hypotheses and “in words” descriptions of the interpretation of each beta coefficient

**Now think about the model in terms of the beta coefficients.**

1. **Solve the expressions in Q4 for the set of parameters in the model. (Hint: This should tricky but it is just a bit of subtracting from sides of the equation: if you really get stuck here, look at the solution and skip to the next step).**
2. **For each coefficient in the model, write the formal H0, and explain this in words.**

**Remember: Each H0 is going to be of the form that the difference between two means is 0. So to do this, you simply need to work out which pair of means is being compared for each beta. This is a little trickier for the interactions. Here you will be asking with the difference between two differences in means is equal to 0.**

A:

|  |  |
| --- | --- |
| **Solved** | **Null (H0)** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

B:

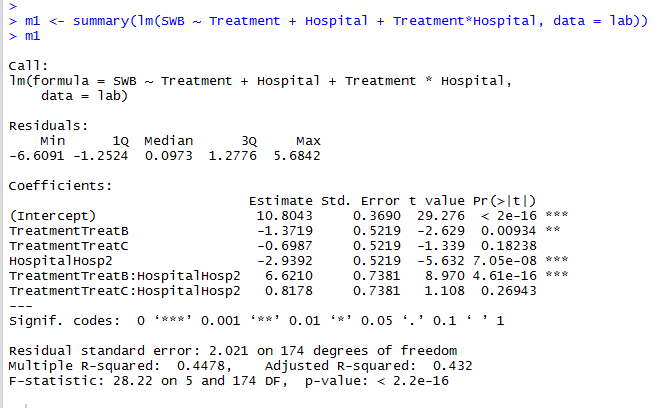
|  |  |
| --- | --- |
|  | **Null (H0) in words** |
|  | Mean of treatment A hospital 1 is equal to 0. |
|  | Difference between Treatment B and Treatment A in Hospital 1 is 0. |
|  | Difference between Treatment C and Treatment A in Hospital 1 is 0. |
|  | Difference between Treatment A in Hospital 1 and Hospital 2 is 0. |
|  | Difference between Treatment A and Treatment B, is different in Hospital 1 and Hospital 2 |
|  | Difference between Treatment A and Treatment C, is different in Hospital 1 and Hospital 2 |

Q6: Run the model

**Specify and run the model as lm() in R.**

***[I bet you never thought you would be relieved to finally be asked to do something in R!!]***

**Hint: Remember to check what level of your factors R is using as the reference, you may need to change this to match your dummy coding above. Look back at the coding working example for how to do this.**



OK, so the running of the model is as usual. We can see that R is treating our factors as factors (of course we could have double checked this before running the model using is.factor() ), and we can also see that it has appropriately used Treatment A and Hospital 1 as the reference points. So we are good.

Q7: Match Q5 to the model results

**Now, using all the necessary steps above, look at your R output. Follow the same procedure as we did at the end of the lecture, insert the various cell means into your results from Q5A, and check they match the values of the coefficient in the model. And lastly, for each coefficient, make a formal statement with respect to the null.**

For ease I am going to copy and paste the cell means back down here. Remember there will be a little bit of rounding error here as the table is only presented to 2 decimal places. If you want to be more precise, you can do all this in R.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Hospital 1 | Hospital 2 |  |
| Treatment A | 10.80 | 7.87 | 9.34 |
| Treatment B | 9.43 | 13.11 | 11.27 |
| Treatment C | 10.11 | 7.98 | 9.05 |
|  | 10.11 | 9.65 | 9.88 |

b0 = 10.80

*(Mean Treatment A, Hospital 1).*

b1 = 9.43 – 10.80 = -1.37

*(Difference between Treatment B Hospital 1 and Treatment A Hospital 1)*

b2 = 10.11 – 10.80 = -0.69

*(Difference between Treatment C Hospital 1 and Treatment A Hospital 1)*

b3 = 7.87 – 10.80 = -2.93 (Difference between Treatment A in Hospital 1 and Treatment A Hospital 2)

b4 = 13.11 – 9.43 – 7.87 + 10.80 = 6.61 (within rounding error!)

*(Difference between Treatment A and Treatment B, is different in Hospital 1 and Hospital 2)*

b5 = 7.98 – 10.11 – 7.87 + 10.80 = 0.80 (within rounding error!)

*(Difference between Treatment A and Treatment C, is different in Hospital 1 and Hospital 2)*

*Formals about the null and an interpretation – all with alpha – 0.05*

*IMPORTANT: For sake of fullness we discuss the interpretation of the lower order effects here, but not we have an interaction, and so generally speaking, we would not be concerned with the interpretation of the dummy coefficients.*

b1: Reject the null that the difference in means is 0 (b = -1.37, p < .05). SWB was significantly higher under Treatment A than Treatment B in Hospital 1.

b2: Fail to reject the null that the difference in means is 0 (b = -0.70, p > .05). There was no significant difference in SWB between Treatment A and Treatment C in Hospital 1.

b3: Reject the null that the difference in means is 0 (b = -2.94, p < .05). SWB was significantly higher under Treatment A in Hospital 1 than in Hospital 2.

b4: Reject the null that the difference in the mean differences is 0 (b = 6.62, p < .05). The difference in SWB between Treatment A and Treatment B was greater in Hospital 2 versus Hospital 1.

b5: Fail to reject the null that the difference in the mean differences is 0 (b = 0.82, p > .05). The difference in SWB between Treatment A and Treatment C did not significantly differ in Hospital 2 versus Hospital 1.

## What would I need in a report?

Clearly for a practical report or use of this model we would not want all of the above steps. If you were to use this approach in your dissertations, you would want to include (not strictly in this order):

* Whole sample descriptive statistics and the same statistics split by cell.
* Details of your coding scheme
* Details of your model specification
* Which parameters test your key hypotheses
* A results table and an interpretation.
* A means plot (see lecture for example – look back at ANOVA notes on how to produce these in R. There is an example using ggplot2 in this week’s lecture code).
* Assumption checks.

## Extension exercises for week 8

If you have time, and wish, you can complete the same set of steps as in week 7 for effects coded variables discussed in lecture 15.

**Note:** The main lab for week 8 is a repeated measures lab, but this is made available as an additional exercise for those wanting to work more in depth with effects codes.

Answers for the additional exercise will be posted at the end of week 8.