

# STAT 215: Two-Way ANOVA

Colin Reimer Dawson

Last Revised December 7, 2016

In this worksheet you will work through some two-way ANOVA analyses by hand, to help you get a feel for what is going on. Then, you will do the same thing in R, to see how to do it, and also to check your hand calculations.

1. Do ants have a preference for different types of sandwich? A student in an introductory statistics class decided to find out. The student chose three types of sandwich fillings (vegemite, peanut butter, and ham & pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches.

To conduct the experiment the student randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes they placed a jar over the sandwich bit and counted the number of ants. They repeated the process, allowing time for ants to return to the hill after each trial, until they had two samples for each combination of the three factors.

Here are the mean numbers of ants based on bread type and whether or not the sandwich had butter:

	White	Wheat	Multi-Grain	Rye	Overall
Buttered	53.50	46.00	50.00	46.00	48.875
No Butter	35.50	42.50	36.00	38.50	38.125
Overall	44.50	44.25	43.00	42.25	43.5

- (a) Write down an additive two-way ANOVA model using these two factors as the predictors.

(b) Estimate the  $\alpha$  and  $\beta$  coefficients. (See the last page for formulas if you forget how to do this.)

(c) Using the fact that the total sum of squares for the Ants variable is 10786, fill in the ANOVA table, including the  $F$  statistics for each factor. Use the `xpf()` function in R or the  $F$ -distribution applet in StatKey to get  $P$ -values. What is your conclusion?

Source	df	SS	MS	F	P
Bread					
Butter					
Residuals				—	—
Total				—	—

(d) Based on the sample means, does the ants' preference for butter appear to be different across the different bread types?

(e) Write down a two-way model *with an interaction*. What does each interaction coefficient  $\gamma$  represent in context?

(f) To estimate the  $\gamma$ s, we can examine the prediction made by the additive model above for each combination, and then let  $\gamma$  be the necessary adjustment to the predicted mean so that the prediction equation gives the actual cell mean. Estimate the  $\gamma$ s for the eight combinations of bread and butter.

(g) With the interaction term in the model, some of the variance that was previously unexplained is now attributed to the interaction between bread type and butter type. Revise the ANOVA table using the  $\gamma$ s to reflect this. Note that the total sum of squares does not depend on the model; only on the response variable data. What is your conclusion?

Source	df	SS	MS	F	P
Bread					
Butter					
Bread:Butter					
Residuals				—	—
Total				—	—

2. Now let's see how to do all of that using R. The data is in `SandwichAnts`, in the `Stat2Data` package.

- (a) Read it in, and fit the two-way additive and interaction models (use `aov()` in place of `lm()`).
- (b) Type the following to get the  $\alpha$ ,  $\beta$  and  $\gamma$  estimates (replace `MODEL` with the name of your model). Note that the levels of each factor will be sorted in alphabetical order; not the order they appear in the data. Do the results match what you computed by hand?

```
model.tables(MODEL, type = "effects")
```

- (c) Use `plotModel()` (after loading the `mosaic` package) to get a plot of the means overlaid on the data. Does it match what you plotted by hand?
- (d) Use `summary()` to get the ANOVA table. Does it match what you computed by hand?

Sum of Squares formulas for ANOVA

$$\begin{aligned}SS_A &= \sum_j \sum_k \sum_{i=1}^{n_{jk}} \hat{\alpha}_j^2 \\SS_B &= \sum_j \sum_k \sum_{i=1}^{n_{jk}} \hat{\beta}_k^2 \\SS_{AB} &= \sum_j \sum_k \sum_{i=1}^{n_{jk}} \hat{\gamma}_{jk}^2 \\SS_{Error} &= SS_{Total} - SS_A - SS_B - SS_{AB} = \\SS_{Total} &= \sum_j \sum_k \sum_{i=1}^{n_{jk}} (y_{jki} - \bar{y})^2 = 1573\end{aligned}$$

Note that you can simplify the calculations a lot by using the fact that there are a lot of repeated values in the sums.