

# STAT 215: Two-Way ANOVA

Colin Reimer Dawson

Last Revised December 1, 2017

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.

Do ants have a preference for different types of sandwich bread? A student in an introductory statistics class in Australia decided to find out. The student made vegemite sandwiches using four types of bread (multigrain, rye, white, and whole wheat), some with butter on the bread, some without<sup>1</sup>.

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 twelve observations for each combination of bread type and butteredness.

Here are the mean numbers of ants based on bread type and whether or not the sandwich had butter. Each mean is computed from 12 observations.

	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

---

<sup>1</sup>I have altered the facts slightly here for simplicity; actually in addition to vegemite sandwiches, there were also peanut butter sandwiches, and ham and pickle sandwiches, but let's pretend the filling was the same to keep the number of factors to two.

1. Write down an additive two-way ANOVA model (that is, with no interaction) using these two factors as the predictors.
2. Estimate the  $\alpha$  and  $\beta$  coefficients. (See the last page for formulas if you forget how to do this.
3. 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. (Again, see the last page for formulas, and remember that the nested summations just mean that we have one term for every observation in the data. Hint: Take advantage of the fact that each  $\hat{\alpha}_j^2$ , occurs once for each observation associated with bread  $j$ , and similarly, each  $\hat{\beta}_k^2$  occurs once for each observation associated with “butteredness”  $k$ .)

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				—	—



7. 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  $\hat{\gamma}$ s to reflect this (see the formulas again, and again notice that each  $\hat{\gamma}_{jk}$  appears as many times as there are data points associated with the combination of bread  $j$  and butteredness  $k$ . Note also 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				—	—

8. 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

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

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