

# STAT 213: Two-Way ANOVA (Additive Model)

Colin Reimer Dawson

Last Revised May 2, 2018

In this handout 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. There were six sandwiches for each combination of bread type and “butteredness”.

	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. Calculate the estimated  $\alpha$  and  $\beta$  coefficients (that is, calculate  $\hat{\alpha}_A$  for each bread type  $A$ , and  $\hat{\beta}_B$  for each butteredness  $B$ ).
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. (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; the simplified versions take advantage of the repetition of identical  $\hat{\alpha}$  and  $\hat{\beta}$  terms. Caution: Remember that each mean is computed from *six* sandwiches)

Use the `xpf()` function in R (requires `mosaic`) or any other  $F$ -distribution utility you are familiar with (e.g., StatKey) to get  $P$ -values. What is your conclusion?

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

4. 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 model (use `aov()` in place of `lm()`).
- (b) Type the following to get the  $\alpha$  and  $\beta$  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()` (requires `mosaic`) to get a plot of the means overlaid on the data.
- (d) Use `summary()` to get the ANOVA table. Does it match what you computed by hand?
- (e) Load `DescTools` and use `PostHocTest(MODEL, method = "hsd")` (replacing `MODEL` with the name of your model) to get pairwise tests and confidence intervals, adjusted for multiple comparisons using Tukey's HSD. What do you conclude?

Sum of Squares formulas for (additive) two-way ANOVA

$$\begin{aligned}SS_A &= \sum_A \sum_B \sum_{i=1}^{n_{AB}} \hat{\alpha}_A^2 = \sum_A n_{A\cdot} \hat{\alpha}_A^2 \\SS_B &= \sum_A \sum_B \sum_{i=1}^{n_{AB}} \hat{\beta}_B^2 = \sum_B n_{\cdot B} \hat{\beta}_B^2 \\SS_{Error} &= \sum_A \sum_B \sum_{i=1}^{n_{AB}} \hat{\epsilon}_{ABi}^2 = SS_{Total} - SS_A - SS_B \\SS_{Total} &= \sum_A \sum_B \sum_{i=1}^{n_{AB}} (y_{ABi} - \bar{y})^2\end{aligned}$$

Notation:

- $n_{A\cdot}$  = number of observations at level  $A$  of the first variable (combining levels of the second variable)
- $n_{\cdot B}$  = number of observations at level  $B$  of the second variable (combining levels of the first variable)
- $n_{AB}$  = number of observations with the combination of level  $A$  and level  $B$  of the two variables.