

Chapter 6 Worksheet

David Gerard

2018-12-07

The `sardine` data frame from Clarke (1936) contains the number of vertebrae and the location of 12,858 sardines. The data are available on Blackboard (as “sardine.rdata”) and you can load it into R using (you will probably have to change the path):

```
load("../data/sardine.rdata")
head(sardine)
```

The `location` variable is coded by

Code	Location
1	Alaska
2	British Columbia
3	San Francisco
4	Monterey
5	San Pedro
6	San Diego

1. Do you think the assumptions are satisfied enough that ANOVA is an appropriate model to use?
2. Regardless of your answer to part 1, run a test to see if the means are the same.
3. If we wanted to compare all pairs of locations, how many tests would we be performing?
4. For what pairs of locations do we have evidence that the sardines have different numbers of vertebrae? (use Tukey here)
5. If you would have used Holm-adjusted p -values instead, would the evidence for the pairwise differences in part 4 be as strong?
6. If the researchers would have suspected that vertebrae differ with colder climate *prior to looking at the data*, it might be reasonable to test if Alaskan and British Columbian sardines have different mean numbers of vertebrae. What model and hypotheses would you use to answer this question?
7. Test the hypotheses from part 6. State your conclusion.

References

Clarke, F.N. 1936. “Variations in the Number of Vertebrae in Sardine, *Sardinops Caerulea* (Girard).” *Copeia* 1936 (3): 147–50.