STA 3100 Programming with Data: Assignment 060

(90 points)

Tinned Tuna

The file `tuna.csv` contains data on 106 different tins of tuna available for purchase from Rainbow Tomatoes Garden. For each available item, the data include the name of the item, the price in dollars, the type of tuna and its Latin name, the country of origin, the brand, the size in grams and in "servings", and a variety of other descriptors (packing liquid (oil or water), contains salt (yes or no), contains sugar (yes or no), smoked (yes or no), grilled (yes or no), organic (yes or no)).

Note on Estimates, Confidence Intervals, and Tests of Hypotheses

In problems which require a point estimate, a confidence interval, and/or a test of hypotheses, you must express your results correctly and coherently **in words** in order to receive full credit.

Exercises

1. (15 pts) Import `tuna.csv` into R. Give a table and a bar plot showing the number of tins of tuna of each type.
2. Shoppers usually expect to get a "better deal" if they purchase a larger quantity of any good, so it may be interesting to examine the relationship between per-gram price and tin size.
   a. (7 pts) Fit a linear model regressing the per-gram price (`Price/Size`) on the size of the tin in grams (`Size`). Save the fitted model as `lm_s` and print a summary of the fit.
   b. (10 pts) Using the model `lm_s` from part a, what would you estimate to be the change in mean per-gram price corresponding to a 100 gram increase in tin size? Express your result to the nearest penny (1/100 of a dollar) and give a 95% confidence interval.
   c. (10 pts) Using the model `lm_s` from part a, give a point estimate and a 95% confidence interval for the expected per-gram price of a tin weighing 120 grams. Express your results to the nearest penny (1/100 of a dollar). (Hint: Use `predict()` with arguments `newdata = data.frame(Size = 120)` and `interval = "confidence"`. See the R help for `predict.lm()` for more information.)
3. We might suspect that the average per-gram price also depends on the type of tuna.
   a. (7 pts) Extend the model from the previous problem by adding `Type` as a predictor. In this model, the effects of `Size` and `Type` on per-gram price should not depend on each other (i.e., this should be an "additive model" with "no interactions"). Save the fitted model as `lm_st` and print a summary of the fit.
   b. (8 pts) Controlling for tin size, is there sufficient evidence to conclude that the per-gram price varies by tuna type? Test at the $\alpha = 0.05$ level of significance.
   c. (10 pts) Using the model `lm_st` from part a, give a point estimate and a 95% confidence interval for the expected difference in per-gram price of a tin bluefin tuna versus a tin of albacore tuna of the same size. Express your results to the nearest penny (1/100 of a dollar).

d. (15 pts) Provide a scatter plot of per-gram price (vertical axis) versus size in grams (horizontal axis). Use color to representing the different type of tuna and include the fitted lines implied by model `lm_st`. (Note: you should "jitter" the points to better reveal overlapping observations. This can be done with `geom_point(position = "jitter")` or more succinctly with `geom_jitter()`.)

4. (8 pts) Is there sufficient evidence to conclude that the effect of tin size on per-gram price depends on the type of tuna (or equivalently, that the effect of type of tuna depends on tin size)? Test at the $\alpha = 0.05$ level.