# Tutorial 9
ST2137-2420

Material

This tutorial covers the topics and concepts from chapter 8. Think of this topic as a generalisation of the approaches in chapter 7. In chapter 9, we will proceed to linear regression. Take note that 2-sample models, ANOVA models and linear regression models are all linear models.

Question 1

The retailing manager of a supermarket chain wants to determine whether product location has any effect on the sale of pet toys. Three different aisle locations are considered: front, middle, and rear. A random sample of 18 stores is selected with 6 stores randomly assigned to each aisle location. The size of the display area and price of the products are constant for all stores. At the end of a one-month trial period, the sales volumes (in thousands of dollars) of the product in each store were recorded in the file `locate.txt`.

1. Assuming that the observations are Normally distributed, use SAS to assess if there is any evidence of a significant difference in average sales among the various aisle locations, at 5% significance level.
2. Boxplots are typically used to assess the distribution within each group. However when we have so few observations, it is sometimes useful to plot every single point, by group. Use `dotplot` from the lattice package in R to create the following plot:
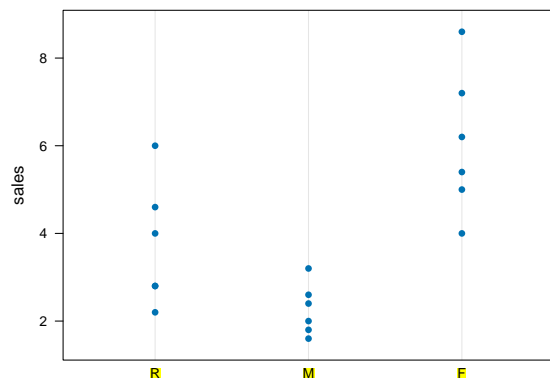
3. In R and Python, set the reference level to be "rear". Compute the confidence interval for the differences between (i) front and rear, and (ii) middle and rear. Use a Bonferroni correction to adjust for the multiple tests so that overall, the error rate is 5%.

Question 2

In earlier topics we noticed that, in the student performance dataset from `student-mat.csv`, G3 scores seem to be different for different `Medu` groups. Remove the group corresponding to `Medu=0` since there are so few observations. Use the following rule to remove outliers from *each group*: X_i is declared an outlier if

$$\frac{|X_i - \text{median}(X)|}{MAD(X)/0.6745} > 2.24$$

Perform the appropriate statistical test(s) to assess the following questions of interest:

4. Is there a significant difference between the 4 groups, at 5% significance level?
5. Estimate the confidence interval for a contrast comparing higher education to non-higher education (i.e. Medu = 4 vs. Medu = 1|2|3).
6. Use Tukey's HSD method to identify which pairs of groups are significantly different from one another at 5% family-wise error level.
7. Repeat the Tukey procedure with all outliers reinstated. How do the results differ?