Assignment 4 - ANOVA

Problem 1

(a) Sketch two boxplots with data from an experiment with 5 treatment groups: one in which you are likely to reject the null hypothesis from an overall F-test, and one in which you are likely to fail to reject the null. Both boxplots should appear consistent with the assumptions of ANOVA. Fill in the table below with the appropriate models for this test.

	Group 1	Group 2	Group 3	Group 4	Group 5
Full model					
Reduced model					

(b) Suppose you had reason to expect Groups 1 and 3 to be different from the others. Sketch a boxplot of a dataset in which you are likely to reject the overall null hypothesis that all groups are equal, but fail to reject a reduced model in which mean 2, 4, and 5 are equal. Fill in the table below with the appropriate models for this test.

	Group 1	Group 2	Group 3	Group 4	Group 5
Full model					
Reduced model					

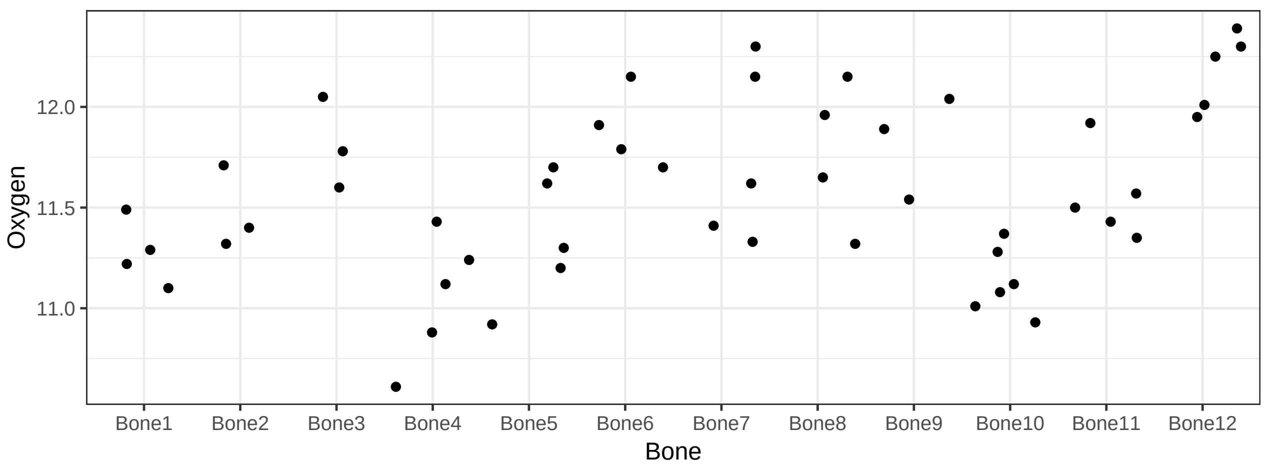
Problem 2 Was Tyrannosaurus Rex Warm-Blooded? (Based on Exercise 23 from Sleuth, Ch 5, p146)
Warm-blooded animals are expected to have fairly constant temperatures throughout their bodies. From the fossil records, oxygen isotopic composition of vertebrate bone phosphate is related to the temperature at which the bone developed. By collecting oxygen isotopic concentrations of bone phosphate in different bone specimens from a single T-Rex skeleton, researchers were able to look for indications of non-constant temperature throughout the body, thus inferring whether the T-Rex was warm-blooded (relatively constant body temperature, no difference in oxygen concentrations) or cold-blooded (non-constant body temperatures, differences in oxygen concentrations). We will use ANOVA to ask whether there is a difference in mean oxygen concentrations among the different bone specimens.

(a) State the null and alternative hypotheses for an overall F-test in the context of this problem.
(You can use words rather than parameters to represent hypotheses.)

H_0 :

H_A :

(b) For each bone speciment in the plot below, draw a dashed line indicating the predicted response for the reduced model of our ANOVA. Draw a solid line to indicate the predicted response for the full model of the ANOVA.



(c) What is "reduced" about the reduced model in an ANOVA compared to the full model?

(d) Describe the distribution of the data as it relates to the assumptions of an ANOVA. Inspect plots in R, using the provided R code, but answer only in text.

[NOTE: This is good practice for you! Even when you CAN include figures with your write-up, your text should be a thorough and standalone assessment of the assumptions: Enough to convince a reader that the data are consistent with model assumptions.]

(e) Write the formula for the test statistic in an ANOVA. Fill in the values of the formula using the ANOVA output from R. Finally, specify the specific reference distribution for your test statistic.

Test statistic: Reference distribution:

(f) Write an "evidence" statement interpreting the results of the ANOVA in the context of this problem.

Problem 3 Write a statistical report using the cloud seeding case study data from Chapter 3, following the guidelines in the "Writing a statistical report" document (D2L - Course Info). Data for this problem are available using `Sleuth3::case0301`.

> The data were collected in southern Florida between 1968 and 1972 to test a hypothesis that massive injection of silver iodide into cumulus clouds can lead to increased rainfall. On each of 52 days that were deemed suitable for cloud seeding, a random mechanism was used to decide whether to seed the target cloud on that day or to leave it unseeded as a control. An airplane flew through the cloud in both cases, since the experimenters and the pilot were themselves unaware of whether on any particular day the seeding mechanism in the plane was loaded or not (that is, they were blind to the treatment). Precipitation was measured as the total rain volume falling from the bloud base following the airplane seeding run, as measured by radar.

Data from J. Simpson, A. Olsen, and J. Eden, "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification," *Technometrics 17 (1975): 161-166)*

Use plots and think about the design to assess the assumptions of your chosen statistical analysis and to justify decisions you make about the analysis. Follow guidelines in Writing a Statistical Report (as well as examples and guidelines for each section), making sure to use correct wording in your Summary of Statistical Findings and Scope of Inference sections. If you use a log-transformation, please use a backtransformation for your Summary of Statistical Findings.

The qqplot below provides some indication that our data might have a consistent pattern of deviation from normality, in violation of the assumptions of the ANOVA.

```
## Error in residuals(lm.full):  object 'lm.full' not found
## Error in residuals(lm.full):  object 'lm.full' not found
## Error in residuals(lm.full):  object 'lm.full' not found
## Error in sigma(lm.full):  object 'lm.full' not found
```

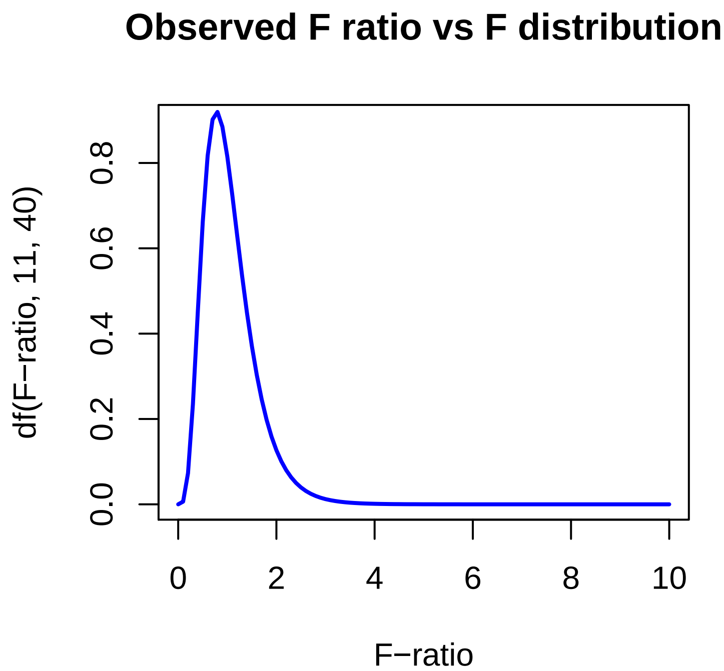
By using the residuals, we can compare the distribution of all the error taken as a single population. The above figure does not show a strong left or right skew, but rather some possible multimodality (multiple peaks) and shortened tails when compared to a normal distribution; leading to deviations between empirical and theoretical quantiles throughout the distributions. One might argue that due to lack of consistent pattern, this is more likely to be due to sample error, than a true signal that the data come from a non-normal distribution; although there is a possibility that each bone specimen follows a different population distribution as well.

Given that the data seem fairly consistent with the equal variance assumption, we could use a permutation test, in place of an F-test to look for a difference among group means. The results of this test are given below:

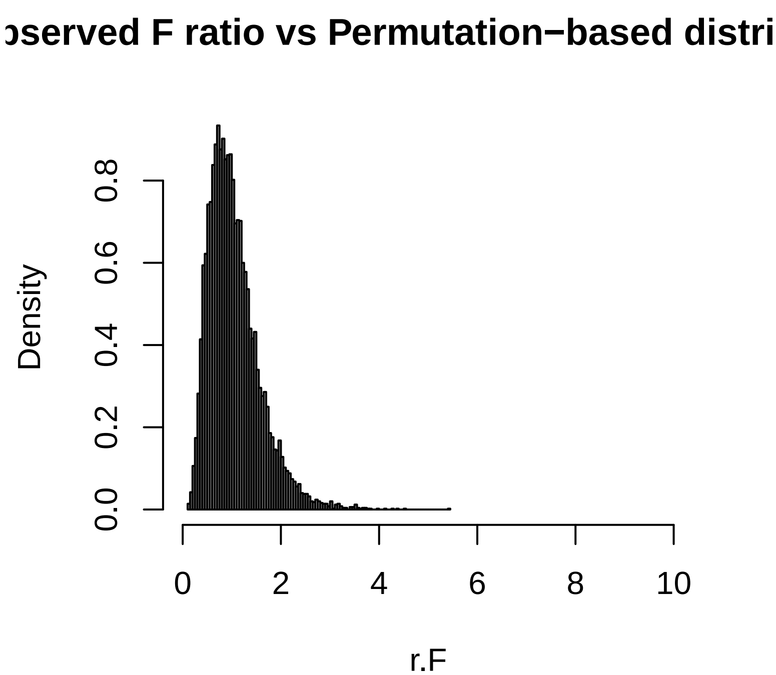
```
F.obs <- anova(lm.full)$F[1]  # observed F-ratio from data

## Error in anova(lm.full):  object 'lm.full' not found

n.sims<-10^4         # Lots of replicates to derive reference distribution
r.F<-numeric(n.sims) # Placeholder to save random F-ratios

for(i in 1:n.sims){
  # Step 1: Shuffle observations to randomly reassign groups
  randomized.data<-data.frame(Oxygen=sample(TRex$Oxygen), Bone=TRex$Bone)
  # Step 2: Calculate F-ratio for difference among groups
  r.F[i]<-anova(lm(Oxygen~Bone, data=randomized.data))$F[1]
}
```

```
F.obs  #Observed F-ratio

## Error in eval(expr, envir, enclos):  object 'F.obs' not found

mean(r.F >= F.obs) # p-value: number of random F-ratios that were

## Error in mean(r.F >= F.obs):  object 'F.obs' not found

                   #  more extreme than observed
```

```
## Error in rect(xleft = F.obs, xright = 10, ybottom = 0, ytop = 1, col = "lightblue", :  object 'F.obs'
not found
```




```
## Error in paste("F =", round(F.obs, 3), "\n p =", signif(anova(lm.full)[[5]][1], :  object 'F.obs'
not found
```




```
## Error in rect(xleft = F.obs, xright = 10, ybottom = 0, ytop = 1.5, col = "lightblue", :  object
'F.obs' not found
## Error in paste("F =", round(F.obs, 3), "\n p =", signif(mean(r.F >= F.obs), :  object 'F.obs' not
found
```