Project 3 Statistical Computing

Tyler Grimes

March 22, 2016

Introduction

Consider the experiment in which three treatments are compared. We label each treatment as "A", "B", and "C". The following data is collected.

Treatment	A	A	B	B	B	C	C	C	C
Response	6	8	9	11	9	17	15	16	16

Suppose our interest is in whether or not there is a difference in the means of the three treatment populations. That is, we want to test the hypothesis

$$\begin{aligned} &H_o : \mu_A = \mu_B = \mu_C \\ &H_a : \text{at least one of the means differs} \end{aligned}$$

In this project, we look at randomization tests based on the one-way ANOVA. Two tests will be performed, one using systematic permutations and the second using random permutations. A comparison of these is given at the end.

One-way ANOVA

Before we look at the permutation tests, let's first consider the one-way ANOVA. We have three different treatments, and we're interested in whether there is a difference in their means. The omnibus hypothesis test can be conducted using ANOVA. The test statistic is

$$F = \frac{MSB}{MSW} = \frac{\frac{SSB}{df_B}}{\frac{SSW}{df_W}} = \frac{\frac{\sum_{i=1}^{k}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}{k-1}}{\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{ij} - \bar{y}_{i\bullet})^2}{N-k}},$$

where $k = 3$ is the number of treatments levels, n_i is the number of observations for the i^{th} treatment, $N = \sum_{i=1}^{k} n_i$ is the total number of observations, $\bar{y}_{\bullet\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$ represents the grand mean, and $\bar{y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ is the i^{th} treatment mean.

If we assume that the observations are independent, and that the underlying populations are normally distributed and homoskedastic across the treatment groups, then under H_o, F has an $F(df_B, df_W)$-distribution.

Now, in this experiment there are only a handfull of observations for each treatment, so assessing the normality assumption is difficult. If this assumption is inappropriate, the ANOVA test would be invalid. The best way to procede if normality is a concern, is to drop the assumption altogether. We can do this by instead performing a randomization test.

Randomization test

Assumptions

Randomization tests allow us to loosen our assumptions about the treatment populations. However, in order to maintain the null hypothesis of equal treatment *means*, we still must assume that the observations are independent and that the treatment populations are homoskedastic and share the same distribution - but now *any* distribution, not necessarily normal. If we were to drop those assumptions as well, then the null hypothesis would be one of equal treatment *effects*. In that case, the alternative would be that the treatments differ in some way, be in in mean, dispersion, skewness, etc. If we were to reject the null, we may still infer that there is a difference in means, but this would only be reliable if the test statistic we chose was mostly sensitive to changes in means, and not to variance, skewness, and so on.

The best way to assess these assumptions is to use domain knowledge about the responses we're measuring, and determine if homoskedasticity and identical distributions are reasonable assumptions. We could use some statistical tests, but they will likely have very low power with such small samples. In any case, it is always good to visualize the data at hand. The box plots and histogram in figure 1 can help us visualize the dispersion and distribution of the data.

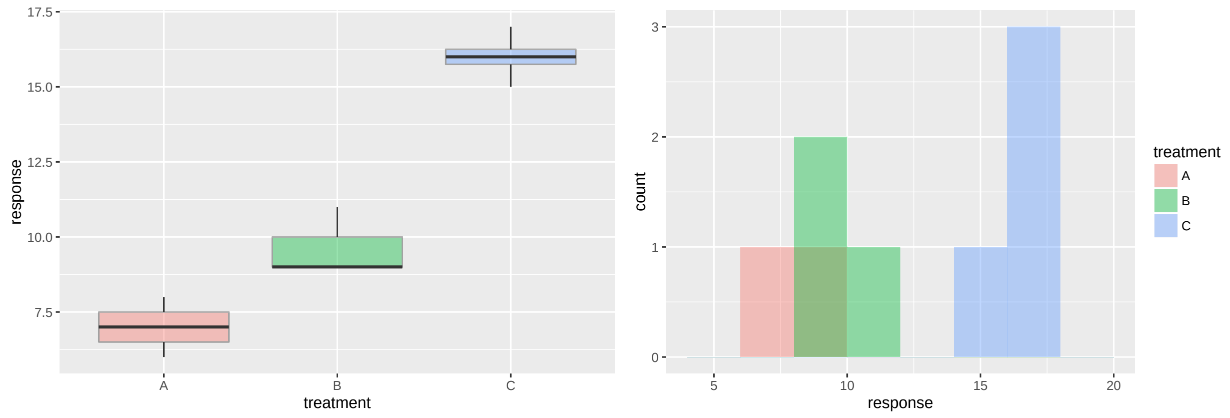

Figure 1: Box plot (left) and histogram (right) of the data from our experiment. These graphs don't reveal any obvious deviations from our assumptions of homoskedasticity and identical distributions.

Procedure

For this experiment, we assume subjects were randomly assigned to treatment groups. Under the null hypothesis, the responses do not depend on the treatment that was assigned; that is, if we could repeat the experiment, and a subject was assigned to a different treatment, the response would not change. We can simulate this act of repeating the experiment by simply rearranging the treatments amoung the subjects. In a randomization test, we use this idea of repeating the experiment to determine the probability of observing a test statistic as extreme as ours.

In total, there are $\frac{9!}{2!3!4!} = 1260$ unique ways to rearrange the treatments. To calculate a $P - value$, first evaluate the test statistic, F, for the initial data set. Then, calculate the statistic for each of the 1260 permutations; we'll denote these by F_p^*. The $P-value$ is the proportion of the F_p^*'s that are at least as extreme as F.

$$P - value = \frac{\text{number of} F_p^* \geq F}{\text{number of permutations}}$$

Equivalent test statistic

Consider the test statistic $F = \frac{\frac{SSB}{df_B}}{\frac{SSW}{df_W}}$. With the permutation test, we are only concerned with the inequality

$$F_p^* \geq F$$
$$\frac{SSB_p^*}{SSW_p^*}\frac{df_W}{df_B} \geq \frac{SSB}{SSW}\frac{df_W}{df_B}$$

where F_p^* is the statistic calculated on the p^{th} permuted data set. We can eliminate any constant factors that appear on boths sides. Since the degrees of freedom do not change with permutations, they can be cancelled out.

$$\frac{SSB_p^*}{SSW_p^*} \geq \frac{SSB}{SSW}$$

In addition, recall the identity $SSTO = SSB + SSW$, where $SSTO$ is the total sum of squares. This implies that, if SSB increases, then SSW necessarily decreases. Because of this relationship, we know SSB increases if and only if $\frac{SSB}{SSW}$ increases. Since we are only interested in whether or not a permuted data set satisfies the above inequality, it is sufficient to only compute SSB_p^* and compare it to SSB.

$$SSB_p^* \geq SSB$$

Furthermore, by rewriting SSB in its computational formula, we find more constants that can be removed.

$$SSB = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i} - \sum_{i=1}^{k} \frac{1}{N} \left(\sum_{l=1}^{k}\sum_{j=1}^{n} y_{lj}\right)^2$$

The second term is a constant, so it can be ignored. Finally, the test statistic is reduced to

$$F_r = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i}$$

The notation F_r will be used to denote this reduced, equivalent test statistic. For each permuted sample, $y_{(b)}^*$, we calculate $F_{r(b)}^* = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i}$. These will be compared to F_r computed from the original data set.

Systematic permutation test

In the systematic test, we produce all of the possible permutations of the data, compute the test statistic for each, and then evaluate the $P - value$.

Algorithm

It is assumed that a method for obtaining all permutations of a single vector is available; most programming languages provide functions for this procedure. In the implementation here, in R, we use the function "permn()" from the library "combinat". permn() returns a list of all permutations for a given vector.

To obtain a list of the 1260 different permutations of the treatment labels, we first use permn() to obtain all 9! permutations of the treatment labels, and then extract the 1260 unique permutations using the unique() function.

Now, to perform the systematic permutation test, we proceed as follows.

1. Compute F_r on the original data.
2. Obtain the set of permutations.
3. For each permutation:
4. `    Compute the test statistic $F^{*}_{r(b)}$.`
5. Set sum equal to the number of $F^{*}_{r(b)} \geq F_r$.
6. Set $P - value$ equal to sum divided by the number of permutations.

Code

```r
#Returns list of unique permutations.
systematic.permutations <- function(data) {
  treatments <- as.character(data[, 1])
  return(unique(permn(treatments)))
}

#Compute the test statistic for each permutation.
ts.list <- NULL
for(trt in systematic.permutations(data)) {
  ts.list <- c(ts.list, F.r(data.frame(treatment = trt, response = data[, 2])))
}

#Calculate the P-value.
p.val <- sum(ts.list >= ts)/length(ts.list)
```

Results

First we compute F_r on the original data.

```r
ts <- F.r(data)
```

This gives $F_r = 1402.3333$. Now, we calculate the test statistc for each of the permuted data sets, compare them to F_r, and calculate the percentage that are at least as large.

The original data set yields the highest value for the test statistic, so the $P - value$ comes out to be $\frac{1}{1260} = 0.0007937$. There's sufficient evidence to conclude that there is a difference in the treatment means.

Random permutation test

As the treatment sample sizes increases, the total number of permutations grows very quickly, making the task of systematic permutation computationally intractable. If this happens, we can instead use a randomization test, whereby we randomly sample from the set of permutations, rather than explicitly create each one. This gives an approximation to the actual $P - value$.

Algorithm

It is assumed that a method for random sampling from a vector without replacement in available. In R, we use the method "sample()". By using this method on the vector containing the 9 treatment levels, we obtain a new vector of the same length with the treatments shuffled around.

Now, to perform the random permutation test, with permutation size n, we proceed as follows.

1. Compute F_r on the original data.
2. Generate n random permutations
3. For each permutation:
4. `Compute the test statistic $F^{*}_{r(b)}$.`
5. Set sum equal to the number of $F^{*}_{r(b)} \geq F_r$.
6. Set $P - value$ equal to sum divided by n.

Code

```r
set.seed(12094)
#Return a list of random permutations.
random.permutations <- function(data, reps = 800) {
  treatments <- as.character(data[, 1])
  perm <- NULL
  for(i in 1:reps) {
    perm <- c(perm, list(sample(treatments, length(treatments))))
  }
  return(perm)
}

#Compute the test statistic for each permutation.
ts.list2 <- NULL
for(trt in random.permutations(data)) {
  ts.list2 <- c(ts.list2, F.r(data.frame(treatment = trt, response = data[, 2])))
}

#Calculate the P-value.
p.val2 <- sum(ts.list2 >= ts)/length(ts.list2)
```

Results

The $P - value$ for this randomization test is 0.00125. Remember, the $P - value$ is subject to change each time we run the test. To get a better approximation, we can increase the permutation size. For example, one run using $n = 10000$ gave a $P - value$ of 0.0008, which is close to the actual value.

Discussion

If the normality assumption to the treatment populations is suspect, a randomization test can be used to calculate a $P - value$. If we systematically compute all permutations, we get an exact answer. However, if the treatment sample sizes are large, or if we have many treatment levels, this systematic approach may be computationally intractable. In such a case, we can instead randomly draw from the set of permutations and get an approximate answer.

Since there are only 1260 possible permutations for this experiment, the systematic approach is tractable and we found the $P - value$ to be 0.0007937. We compared this to the randomization test by running one such test, which gave a $P - value$ of 0.00125. The relative error of this approximation is 57.5%.

We may not be satisfied with only one run of the randomization test. To get a better idea of how the randomization test performs, with a permutation size of $n = 800$, we can simulate the test many times and look at the distribution of $P - values$. The histogram in figure 2 shows the distribution of 1000 $P - values$ obtained from the randomization test.

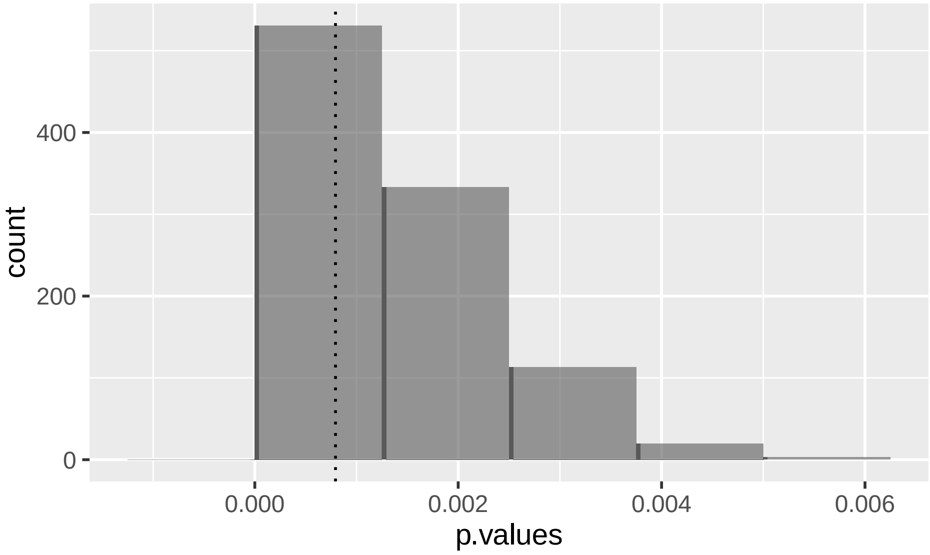

Figure 2: Histogram of 1000 P-values from the randomization test. Note, the possible values of the P-value are multiiples of 1/800, since the permutation size is 800. The dotted line marks the actual value of 0.0007937 (1/1260).

This distribution suggests that the randomization test will usually yield a $P - value$ that is within 0.00125 from the correct value; in this simulation, 86.4% of the p-values were either 0 or 0.00125. Based on this example, it appears that the randomization test is indeed a reliable approximation to the systematic test.