post02-jessica-yu

Jessica Yu

12/3/2017

A Novice's Guide to One-Way ANOVA

Introduction

Perhaps the most commonly used and well-known measure of center is the mean. ANOVA, or "one-way analysis of variance", analyzes the means of multiple independent groups for statistical significance. The usefulness of ANOVA becomes clear when we consider what happens when we analyze the means of more than two independent groups.

When there are just two groups, you can examine the difference between their means using a t-test. When the number of groups increases to three, it's still possible to compare each of the three different pairings of groups by doing three separate t-tests. However, this can quickly result in the need for many t-tests…when studying five groups, you would have ten different pairings and the need to perform ten t-tests. Who wants to do ten t-tests? Besides, this labor-intensive process is especially undesirable as each additional t-test increases the chance of a Type I error, that is, falsely concluding significance when there is none.

Ideally, you would compare the means of multiple groups using only one test that produces a single p value indicating if the groups have consistent means or significantly different means. ANOVA does precisely this: rather than looking for differences between multiple pairs of group means, **ANOVA finds how the entire collection of group means is spread out and compares the result to how the means would be spread out if all of the groups were sampled from the same population; i.e. if there were no significant differences between the means.**

ANOVA is useful in many different contexts, such as for research experiments in economics, medicine, and other fields that require **hypothesis testing**. This post will walk through the steps of using ANOVA in R for these broadly applicable and highly useful purposes.

The **objectives** of the post:

> - Understand F-tests and the F ratio
> - Identify useful R functions for ANOVA and multiple testing
> - Walk through a simple ANOVA example
> - Identify contemporary applications of ANOVA

1. Important Points: F Ratio, P Value, and Degrees of Freedom

To produce a p value, ANOVA uses the F ratio and the degrees of freedom of the data.

Some quick reminders:

p value: The p value is the probability of observing the given events under the null hypothesis. It evaluates how consistent the data is with the null hypothesis.

If the p value from the ANOVA is significant (less than the alpha level), then you can reject the null and accept the alternative hypothesis.

F tests: The result of an ANOVA test is calculated using the F ratio, a test statistic that compares the amount of variability between and within groups.

F = variation between sample means / variation within the samples

where df1 is the degrees of freedom in the numerator and df2 the degrees of freedom in the denominator.

df1 is one less than the number of groups.

df2 is given by subtracting the number of groups from the number of observations.

Degrees of freedom: Intuitively, the degrees of freedom represent the amount of information your data provide that you can "spend" to estimate the values of unknown population parameters and their variability. The number of degrees of freedom is based on:

- the number of observations in your sample and
- the number of parameters in your model.

Example of determining the degrees of freedom: Consider a one-sample t-test that estimates one parameter, the population mean. The sample size of n constitutes the n pieces of information used in estimating the population mean. One degree of freedom is "used up"" to estimate the mean, leaving n-1 degrees of freedom. Thus, a one-sample t-test follows a t-distribution with n-1 degrees of freedom.

If you attempt to estimate population parameters using too few observations or for too many parameters, you reduce the degrees of freedom and thus the reliability of your estimations.

Key point: ANOVA functions in R use the F ratio and the degrees of freedom to calculate the p value.

2. The PlantGrowth Dataset and Initial Observations

We will use the PlantGrowth data set available in R. The dataset contains results from an experiment that compares the effect of different treatments upon the dried weight of plants. There are three groups: a control group and two experimental groups.

```
# load packages
library(ggplot2)
```

```
# load the dataset in R using data()
data("PlantGrowth")
head(PlantGrowth)
```

```
##    weight group
## 1    4.17   ctrl
## 2    5.58   ctrl
## 3    5.18   ctrl
## 4    6.11   ctrl
## 5    4.50   ctrl
## 6    4.61   ctrl
```

"weight" is plant weight and "group" indicates the treatment applied: none for the control group, or treatment 1 ("trt1") or treatment 2 ("trt2").

Quick Observations: To get a quick idea of the variation of the data for each group, create a boxplot.

```
plantgrowth_boxpl <- ggplot(PlantGrowth, aes(x = group, y = weight)) +
  geom_boxplot(fill = "cornflowerblue") +
  ggtitle("PlantGrowth Weight Boxplots")

plantgrowth_boxpl
```


PlantGrowth Weight Boxplots
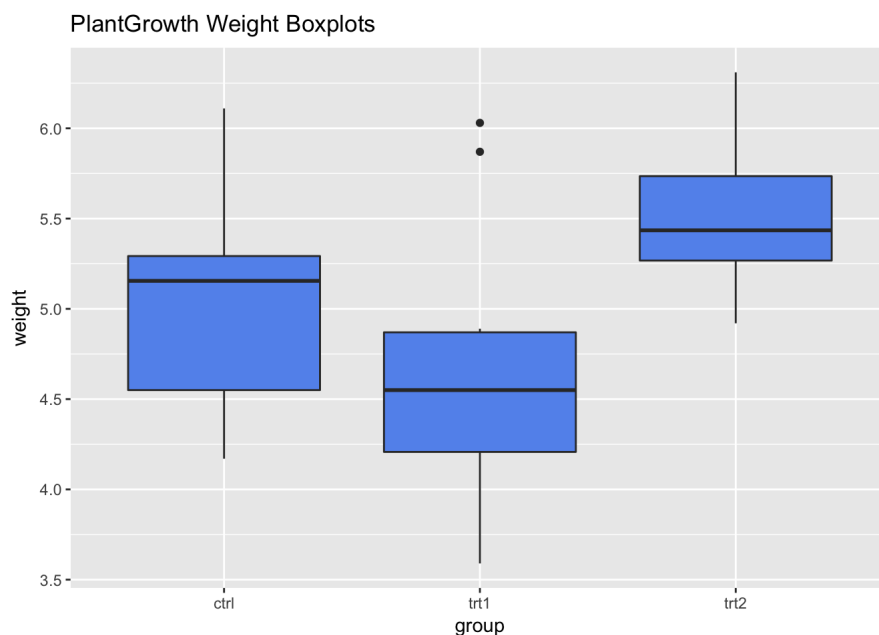
```
#ggsave("../images/plantgrowth_boxplot.pdf", plantgrowth_boxpl)
```

We see that there are differences between the minimum, maximum, and median values across the three groups. trt1 has the lowest plant weight, a very narrow third quartile range, and two outliers.

Right away, we also see that the biggest differences appear to exist between trt1 and trt2. In fact, trt1 appears to be performing worse than the control while trt2 appears to be performing better than the control. After first performing an ANOVA F-test, we can use pairwise analyses to ascertain if there are statistically significant differences between trt1 and trt2.

3. ANOVA

aov()

The function aov() is used to fit an analysis of variance model by fitting a linear model for each stratum of data. A stratum is one of the non-overlapping data subgroups used in stratified random sampling of the total population.

aov(formula, data = NULL)

- **formula**: a formula specifying the model, given in the form response ~ factor. Response is the dependent variable and factor the variable that divides the data into groups.

- **data**: the data frame in which the variables are located

```
# call aov(), with response variable plant weight and factor (treatment) group
aov_results <- aov(weight ~ group, PlantGrowth)

# examine the results using summary(), which produces the ANOVA table
summary(aov_results)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## group        2  3.766  1.8832   4.846 0.0159 *
## Residuals   27 10.492  0.3886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: The p value is .012, which is lower than the default alpha value of .05. Thus, we reject the null hypothesis that there are equal means for the weight of plants subjected to different treatments. We conclude that there are significant differences between the means of plant weight.

Residual Plot: To quickly gauge how well aov() fitted the original data, we can obtain the residuals and plot them against the fitted values. There are 3 aov fitted values, one for each group or stratum (control, trt1, trt2).

```
#resid() extracts model residuals for the inputted object
aov_resid <- resid(aov_results)

# calculate fitted values by subtracting residuals from actual values (resid = actual - fitted, so fitted = actual
- resid by rearranging the equation)
aov_fit <- PlantGrowth$weight - aov_resid

# create a new data frame with fitted values and residual values
resid_fit_df <- data.frame(aov_fit, aov_resid)

resid_plot <- ggplot(resid_fit_df, aes(x = aov_fit, y = aov_resid)) +
  geom_point(col = "tomato") +
  geom_hline(yintercept = 0) +
  ggtitle("ANOVA Residual Plot", subtitle = "PlantGrowth Weights") +
  xlab("ANOVA fitted values") +
  ylab("Residuals")

resid_plot
```

ANOVA Residual Plot
PlantGrowth Weights
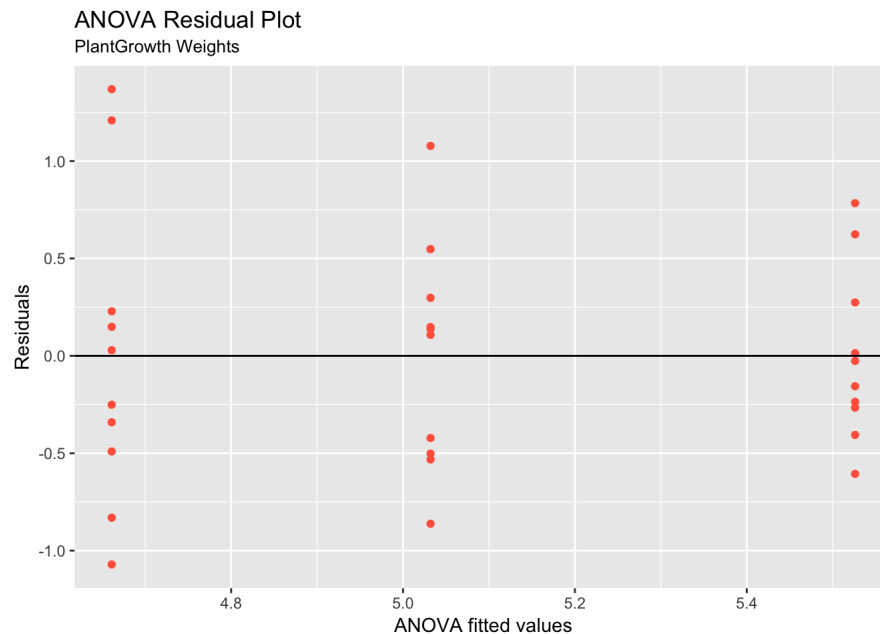

```
#ggsave("../images/resid_plot.pdf", resid_plot)
```

The plot verifies that the residuals are randomly distributed, with points falling randomly and evenly on both sides of 0.

4. Pairwise Comparisons

What if we wanted to dig deeper and find out why the means of plant weight differ? Further analyses can be done to help determine the source of the differences. Perhaps we'll find out if it is trt1 or trt2 that drives the difference in plant weight.

pairwise.t.test()

The function pairwise.t.test() is used to produce p values for pairwise t-tests between the groups, with corrections for the errors produced by multiple testing.

Simplified, it allows us to determine if there are statistically significant differences between A-B, A-C, and B-C, in the case of three groups. This can help pinpoint where the difference indicated by an ANOVA F-test is coming from.

A word on corrections for multiple testing: When multiple hypotheses are tested such as through multiple t-tests, the likelihood that a rare event will be observed increases with each test. This boosts the likelihood of making a Type I error by incorrectly rejecting the null hypothesis that all alternative hypotheses are false.

The **Bonferroni correction** is one correction method that counters this increase by lowering the significance level (the alpha level). It divides the overall alpha level by the number of hypotheses being tested. If we are testing 3 hypotheses at an alpha level of .05, the Bonferonni correction would test each individual hypothesis at an alpha level of .0167.

pairwise.t.test(x, g, p.adjust.method)

- **x**: the response vector

- **g**: the grouping vector or factor

- **p.adjust.method**: the method for adjusting p values to correct for multiple testing

```
# attach() attaches a database so that R will search it when evaluating a variable; after using attach(), objects
in the database can be accessed by simply providing their names as seen in the second line of code
attach(PlantGrowth)

# response variable is weight, factor is weight, and adjustment method used is the Bonferroni correction
pairwise.t.test(weight, group, p.adjust = "bonferroni")
```

```
## 
##  Pairwise comparisons using t tests with pooled SD
## 
## data:  weight and group
## 
##      ctrl  trt1
## trt1 0.583 -
## trt2 0.263 0.013
## 
## P value adjustment method: bonferroni
```

```
#detach()
detach()
```

There are no significant differences between the control group and trt1 or the control group and trt2. Thus, it appears that applying either treatment did not have a significant effect on the weight of the plants. However, there is a statistically significant difference between trt1 and trt2 with p = .013, implying that the treatments produced significantly different mean plant weights. Thinking back to the boxplot that showed large differences between trt1 and trt2, this is not too surprising.

TukeyHSD()

Another way for performing multiple comparisons is using Tukey's Honest Significance Test. The function TukeyHSD() creates a set of confidence intervals on the differences between the means, using the specified confidence level.

TukeyHSD(x, conf.level = 0.95)

- **x**: an object fitted usually through aov()

- **conf.level**: a numeric value between 0 and 1 establishing the confidence level to use; default is 0.95

```
# aov_results used aov() to fit the data, as seen at the beginning of 3. ANOVA
TukeyHSD(aov_results, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = weight ~ group, data = PlantGrowth)
## 
## $group
##            diff        lwr       upr     p adj
## trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
## trt2-ctrl  0.494 -0.1972161 1.1852161 0.1979960
## trt2-trt1  0.865  0.1737839 1.5562161 0.0120064
```

Using the last column showing p values adjusted for multiple comparisons, we arrive at significance conclusions consistent with those produced using the p values of pairwise.t.test(), although numerical values differ.

Conclusion/TLDR

Rather than performing three separate t-tests, we used an ANOVA F-test to evaluate if there were significant differences in the means of plant weight in three groups. After determining through the p value that there were indeed statistically significant differences, we used pairwise.t.test() and Tukey's Honest Significance Test to identify the driver of this difference. Since neither treatment produced weights that were significantly different from the control, it appears that neither treatment is useful in improving plant weight. However, there were significant differences between the weights produced by the two treatments. Treatment 2 seems to be headed in the right direction by increasing plant weight while Treatment 1 seems to have reduced plant weight. Shifting resources from Treatment 1 to Treatment 2 may allow the further development of Treatment 2. One day, this may lead to the successful creation of a plant treatment that causes significant gain in plant weight.

In this way, ANOVA can help shape and answer research questions in a wide array of fields. While this post focused on One-Way ANOVA, there is also Two-Way ANOVA in which an ANOVA is done using two independent variables. More generally, for n independent variables, an N-Way ANOVA is performed. In the case of dependent variables, MANOVA (multivariate analysis of variance) may be used. For analyses involving covariate variables, ANCOVA (analysis of covariance) is used. All analysis forms are highly useful in suitable research contexts and will continue to help drive research progress.

References

One-way ANOVA, Columbia University

The Basic Idea of an Analysis of Variance (ANOVA), Dummies

Understanding Analysis of Variance (ANOVA) and the F-test, Minitab

One-way Analysis of Variance (ANOVA), R-bloggers

What are degrees of freedom?, Minitab

Residual Plots for One-Way ANOVA, Minitab

Residual Plot, R Tutorial

pairwise.t.test, R Documentation

To attach() or not attach(): that is the question, R-bloggers

Bonferroni Correction, Wolfram MathWorld