

Lecture 07 - Analysis of Variance

Data Analysis in CJ (CJUS 6103)

Outline

- I. Logic of ANOVA
- II. Computation of ANOVA
- III. Computing ANOVA in R
- IV. Measures of Association
- V. Post-Hoc Tests

Logic of ANOVA

We have just encountered research problems in which we want to compare two sample means in order to make inferences about the two groups in the larger population. Now suppose that we want to compare means from three or more samples. In this case, we have a categorical independent variable with three or more categories, and a continuous dependent variable. When we have three or more sample means, the question we are interested in answering is this: Do the differences that we observe among the sample means indicate that there are significant differences across groups in the population? This is just a generalization of the two-sample case. There are several examples of such problems:

- 1) Sentence length as a function of offense type (violent, property, drug, other)
- 2) Fear of crime as a function of residential location (urban, suburban, rural)
- 3) Offending as a function of family structure (two-parents, single-parent, no biological parents).

One way to deal with this type of data is to carry out a series of two-sample hypothesis tests. An obvious disadvantage to this strategy is that the more groups that we have, the greater number of pairwise t-tests that we have to conduct. For example, say that we have an independent variable with three categories. This means that we would have to conduct (using the combination rule) $3!/2!(3-2)! = 3$ separate hypothesis tests. With four groups there are 6 hypothesis tests, with five groups there are 10 tests, with six groups there are 15 tests, with seven groups there are 21 tests, and so on. As you can see, the number of hypothesis tests quickly becomes unmanageable.

A more substantive disadvantage to conducting multiple hypothesis tests with three or more samples is that, since each of the tests is conducted on the same data, they are **not independent** of one another. The practical implication of this is that over multiple tests, the probability of committing a type I error (i.e., the probability of falsely rejecting the null hypothesis) on any given test is greater than α . As an example, suppose that we have three groups. We can determine using the binomial formula that the probability that we falsely reject the null hypothesis on at least one hypothesis test using a 5% criterion is $p(r > 0 | n = 3) = 1 - C_0^3(.05)^0(.95)^3 = .143$. If we increase the number of hypothesis tests, we find that $p(r > 0 | n = 4) = .185$, $p(r > 0 | n = 5) = .226$, $p(r > 0 | n = 6) = .265$, $p(r > 0 | n = 7) = .302$, and so on. Clearly, the probability that we would falsely reject in at least one hypothesis test becomes considerably greater than $\alpha = .05$.

As you already know, we want a statistical test that will help us decide whether these observed differences are the result of sampling variation or (presumably) real differences. Analysis of variance (ANOVA) is useful for

determining the extent to which there are statistically significant differences between three or more sample means. We refer to ANOVA as a **global** test, which means that it tests the **joint significance** of several sample means, rather than differences among specific pairs. The advantage to conducting a global test is that the probability of committing a type I error is constant.

Why is variance so important, when we are actually interested in comparing means? With ANOVA, we speak of two different kinds of variability: variability **within** and **between** groups. Let's consider each separately.

- 1) Variability **within** groups refers to how tightly clustered individual scores are from their group mean. When this variability is small, each of the cases within a group cluster tightly around their respective group means, indicating that more of the cases are similar within a particular group than are different.
- 2) Variability **between** groups (or across groups) refers to how tightly clustered the sample means are from each other, or from what we refer to as a **grand mean** (\bar{x}_G). When the variability between groups is large, the group means are only loosely clustered around the grand mean, indicating that the group means are more different than they are similar.

Now consider the ratio of **between-group** variability to **within-group** variability, and its implication for statistical inference. We refer to this as an **F-ratio**. When there is more variability within groups than between groups, the F-ratio will be less than one. This means that there is a great deal of overlap among the group distributions, and thus there is no relationship between group membership and the dependent variable. When there is more variability between groups than within groups, the F-ratio ratio will be greater than one. This means that there is little or no overlap among the groups, and thus group membership is associated with the dependent variable.

When we use ANOVA, we rely on a new probability distribution: the **F-distribution**. An F-ratio is simply the ratio of the variability between groups to the variability within groups. When we have a large F-ratio (i.e., one that is significantly greater than one), we will be led to reject the null hypothesis of no association between group membership and the outcome of interest.

Computation of ANOVA

In ANOVA terminology, we are interested in a measure of variability called the **sum of squared deviations about the mean**, or simply the **sum of squares**. You might remember that the sum of squares is simply the numerator of the variance formula. A quick tutorial on notation: n refers to the sample size of a particular group whereas N refers to the number of cases across all groups, i refers to a particular observation, and k refers to a group. The combination of i and k is interpreted to mean observation $\#i$ in group $\#k$. There are three different sums of squares that we want to know in ANOVA.

- 1) **Total sum of squares:** This is the sum of the squared deviations of each case around the grand mean.

$$SS_T = \sum (x_{ik} - \bar{x}_G)^2 = \sum x_{ik}^2 - N\bar{x}_G^2$$

- 2) **Between-groups sum of squares:** This is the sum of the squared deviations of each group mean around the grand mean.

$$SS_B = \sum n_k(\bar{x}_k - \bar{x}_G)^2 = \sum n_k\bar{x}_k^2 - N\bar{x}_G^2$$

- 3) **Within-groups sum of squares:** This is the sum of squared deviations of each case around its respective group mean.

$$SS_W = \sum (x_{ik} - \bar{x}_k)^2 = \sum x_{ik}^2 - \sum n_k\bar{x}_k^2 = SS_T - SS_B$$

The relationship among these three different sum of squares is straightforward:

$$SS_T = SS_B + SS_W$$

In addition to the sums of squares, we will also need to know our degrees of freedom:

- 1) **Total degrees of freedom:** $df_T = N - 1$
- 2) **Between-group degrees of freedom:** $df_B = k - 1$
- 3) **Within-group degrees of freedom:** $df_W = N - k$

Using this information, we can partition the variance into the **total* variance**, the **between-group** variance**, and the **within-group** variance. The variance is computed as the sum of squares divided by the respective degrees of freedom. The F-statistic is calculated as the ratio of the between-group variance to the within-group variance.

ANOVA Example - Sentence Length and Offense Type

Let's carry out a full hypothesis test with the data from a sentence length example. Suppose that we have a sample of 40 individuals that have committed one of four types of offenses. Our independent variable is offense type, and our dependent variable is sentence length in months.

Violent (x_1)	Property (x_2)	Drug (x_3)	Other (x_4)
6	4	6	1
18	6	3	3
20	3	3	1
15	10	4	1
20	12	6	6
30	8	9	9
25	6	10	3
12	10	3	6
24	8	2	2
20	15	3	4
190	82	49	36

Step 1. State Hypotheses

With ANOVA problems, the alternative hypothesis is always $\mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \mu_k$, and the null hypothesis is $\mu_1 = \mu_2 = \mu_3 = \dots \mu_k$. We make the assumption under the null hypothesis that the population means are equivalent. Under the alternative hypothesis, we make the assumption that *at least one* population mean is significantly different from *at least one* other population mean. ANOVA is what we call a **global test** in that it can tell us if there are significant differences in the means, but it cannot tell us exactly which means are significantly different.

Step 2. Obtain a Probability Distribution

In order to do an ANOVA, we have to resort to the F-distribution. The F-distribution is one-tailed, because we square the deviations. The shape of the F-distribution is defined by two sets of degrees of freedom. The df in the numerator is $df_B = k - 1 = 3$ (top row of a typical F-ratio table) and the df in the denominator is $df_W = N - k = 36$ (far left column of a typical F-ratio table).

Step 3. Make Decision Rules

Make decision rules. Let's use $\alpha = .01$. The critical value of F is 4.377, thus we will reject the null hypothesis if $F > 4.377$.

You can obtain the critical F-ratio value using the `qf()` function in R. The basic structure of the `qf()` function is:

`qf(p, df1, df2, lower.tail=TRUE/FALSE)`

Where `p` is your α -level, `df1` is the numerator degrees of freedom, and `df2` is the denominator degrees of freedom. We will exclusively be setting `lower.tail=FALSE` because the F-distribution, like the χ^2 distribution, only contains positive values (so no critical scores will belong on the left-hand side of the distribution).

Note: Although the `qf()` command accepts the arguments `df1=` and `df2=` when run in the console window, it does not appear to recognize them in inline code. If using the `qf()` function in inline code you can simply enter the calculation or value for `df1` and `df2` but be sure not to enter `df1=` or `df2=` (e.g., `4-1` and `40-4` are okay, `df1=4-1` and `df2=40-4` will NOT work in inline code). As a result of this, you must be especially attentive to df order - numerator first, denominator second!

Step 4. Calculate the Test Statistic

Violent		Property		Drug		Other	
x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2
6	36	4	16	6	36	1	1
18	324	6	36	3	9	3	9
20	400	3	9	3	9	1	1
15	225	10	100	4	16	1	1
20	400	12	144	6	36	6	36
30	900	8	64	9	81	9	81
25	625	6	36	10	100	3	9
12	144	10	100	3	9	6	36
24	576	15	225	3	9	4	16
190	4030	82	794	49	309	36	194

There are several pieces of information that we need to obtain to calculate the F-statistic. First, we need to compute the group means as well as the grand mean.

- $\bar{x}_1 = \sum x_{i1}/n_1 = 190/10 = 19$
- $\bar{x}_2 = \sum x_{i2}/n_2 = 82/10 = 8.2$
- $\bar{x}_3 = \sum x_{i3}/n_3 = 49/10 = 4.9$
- $\bar{x}_4 = \sum x_{i4}/n_4 = 36/10 = 3.6$
- $\bar{x}_G = \sum x_i/N = \frac{190+82+49+36}{10+10+10+10} = 8.925$

Second, we need to compute the sums of squares. This step only requires us to find the total sum of squares and between-group sum of squares, which we can use to solve for the within-group sum of squares.

- $SS_T = \sum x_i^2 - N\bar{x}_G^2 = (4030 + 794 + 309 + 194) - 40(8.925)^2 = 5327 - 3168.40 = 2140.775$
- $SS_B = \sum n_k \bar{x}_k^2 - N\bar{x}_G^2 = 10(19.0)^2 + 10(8.2)^2 + 10(4.9)^2 + 10(3.6)^2 - 40(8.925)^2 = 1465.875$
- $SS_W = \sum x_{ik}^2 - \sum n_k \bar{x}_k^2 = (4030 + 794 + 309 + 194) - (10 * (19.0^2)) + (10 * (8.2^2)) + (10 * (4.9^2)) + (10 * (3.6^2)) = 5327 - 4652.1 = 674.9$

Third, we use this information to calculate the F-statistic. It is convenient to put ANOVA data into the form of a table.

Source	SS	df	$MS = SS/df$	$F = MS_B/MS_W$
Between groups	1465.875	$k - 1 = 3$	488.62	$\frac{494.57}{18.75} = 26.06$
Within groups	674.9	$N - k = 36$	18.75	
Total	2140.775	$N - 1 = 39$	54.89	

Step 5. Make a Decision about the Null Hypothesis

Since $F = 26.38$, we reject the null hypothesis, and conclude that offense type is significantly associated with sentence length. Recall, however, that an F-test, since it is **global**, cannot tell us anything more substantive than this. We know that there are significant differences, but without conducting further tests (which we call **post-hoc tests**) we are unable to draw any further conclusions. Side note - many ANOVA functions in statistical programs contain options to report all combinations of **post-hoc tests**, generally with some kind of correction to the α level that avoids the increasing probability of Type I errors I described earlier. One popular correction is the Bonferroni method, which is kind of fun to say.

ANOVA Example - Fear of Crime and Residential Location

You collect data on fear of crime from a sample of 30 individuals, divided equally among urban, suburban, and rural areas. The research question is this: Is area of residence related to fear of crime?

Step 1. State Hypotheses

$$H_1 : \mu_U \neq \mu_S \neq \mu_R; H_0 : \mu_U = \mu_S = \mu_R$$

Step 2. Obtain a Probability Distribution

$$F\text{-distribution, } df_B = k - 1 = 3 - 1 = 2, df_W = N - k = 30 - 3 = 27$$

Step 3. Make Decision Rules

$$\alpha = .05 F_{crit} = 3.354; \text{reject } H_0 \text{ if } F > 3.354$$

Step 4. Calculate the Test Statistic

Urban		Suburban		Rural	
x_U	x_U^2	x_S	x_S^2	x_R	x_R^2
22	484	23	529	19	361
29	841	22	484	24	576
31	961	26	676	24	576
28	784	25	625	19	361
30	900	24	576	20	400
32	1024	25	625	24	576
32	1024	24	576	21	441
31	961	24	576	17	289
28	784	27	729	23	529
30	900	23	529	19	361
293	8663	243	5925	210	4470

First, obtain group means and the grand mean:

- $\bar{x}_1 = \sum x_{i1}/n_1 = 293/10 = 29.3$
- $\bar{x}_2 = \sum x_{i2}/n_2 = 243/10 = 24.3$
- $\bar{x}_3 = \sum x_{i3}/n_3 = 210/10 = 21$
- $\bar{x}_G = \sum x_i/N = \frac{293+243+210}{10+10+10} = 24.87$

The, compute sum of squares values:

- $SS_T = \sum x_i^2 - N\bar{x}_G^2 = (8663 + 5925 + 4470) - 30(24.87)^2 = 502.493$
- $SS_B = \sum n_k \bar{x}_k^2 - N\bar{x}_G^2 = 10(29.3)^2 + 10(24.3)^2 + 10(21.0)^2 - 30(24.87)^2 = 344.293$
- $SS_W = \sum x_i^2 - \sum n_k \bar{x}_k^2 = SS_T - SS_B = 502.493 - 344.293 = 158.2$

Then, create the table:

Source	SS	df	$MS = SS/df$	$F = MS_B/MS_W$
Between groups	344.293	$k - 1 = 2$	172.147	$\frac{172.147}{5.859} = 29.38$
Within groups	158.2	$N - k = 27$	5.859	
Total	502.493	$N - 1 = 29$	17.327	

Step 5. Make a Decision about the Null Hypothesis

Reject H_0 , conclude that area of residence is related to fear of crime. Alternatively, we conclude that at least one of these group means is different from one or more group means in the population.

ANOVA Example - Sentence Length and Offense Type (but this time using R)

In this short section, I am going to show you how to calculate the F-statistic using R and then how to validate that computation with an automatic method.

Input Data

```
violent<-c(6,18,20,15,20,30,25,12,24,20)
property<-c(4,6,3,10,12,8,6,10,8,15)
drug<-c(6,3,3,4,6,9,10,3,2,3)
other<-c(1,3,1,1,6,9,3,6,2,4)
```

Obtain Means

```
viol_xbar<-sum(violent)/length(violent)
viol_xbar
```

```
## [1] 19
```

```
prop_xbar<-sum(property)/length(property)
prop_xbar
```

```
## [1] 8.2
```

```
drug_xbar<-sum(drug)/length(drug)
drug_xbar
```

```
## [1] 4.9
```

```
oth_xbar<-sum(other)/length(other)
oth_xbar
```

```
## [1] 3.6
```

```
grand_xbar<-(sum(violent)+sum(property)+sum(drug)+sum(other))/
  (length(violent)+length(property)+length(drug)+length(other))
grand_xbar
```

```
## [1] 8.925
```

Note - when the sample sizes are the same for each group the grand mean is the mean of the group means $(19+8.2+4.9+3.6)/4 = 8.925$.

Obtain Sum of Squares


```

viol_sq<-violent^2
prop_sq<-property^2
drug_sq<-drug^2
oth_sq<-other^2

## SS Total
ss_ttl<-(sum(viol_sq)+sum(prop_sq)+sum(drug_sq)+sum(oth_sq))-
  ((length(viol_sq)+length(prop_sq)+length(drug_sq)+length(oth_sq))*
    (grand_xbar^2))
ss_ttl

```

```
## [1] 2140.775
```

```

## SS Between
ss_bet<-((length(violent)*viol_xbar^2)+(length(property)*prop_xbar^2) +
  (length(drug)*drug_xbar^2)+(length(other)*oth_xbar^2)) -
  ((length(viol_sq)+length(prop_sq)+length(drug_sq)+length(oth_sq))*
    (grand_xbar^2))
ss_bet

```

```
## [1] 1465.875
```

```

## SS Within
ss_with<-(sum(viol_sq)+sum(prop_sq)+sum(drug_sq)+sum(oth_sq)) -
  ((length(violent)*viol_xbar^2)+(length(property)*prop_xbar^2) +
    (length(drug)*drug_xbar^2)+(length(other)*oth_xbar^2))
ss_with

```

```
## [1] 674.9
```

Note: You can verify here that $ss_ttl - ss_bet = ss_with$.

Compute the F-ratio

```

## Mean Square Between
ms_bet<-ss_bet/(4-1)
ms_bet

```

```
## [1] 488.625
```

```

## Mean Square Within
ms_with<-ss_with/((length(violent)+length(property)+length(drug)+length(other))-4)
ms_with

```

```
## [1] 18.74722
```

```
## Mean Square Total
ms_ttl<-ss_ttl/((length(violent)+length(property)+length(drug)+length(other))-1)
ms_ttl
```

```
## [1] 54.89167
```

```
## Calculate F-Ratio
f_ratio<-ms_bet/ms_within
f_ratio
```

```
## [1] 26.06386
```

ANOVA Example - Automatic R Method for Sentence Length Data

```
## Create numeric vectors
violent<-c(6,18,20,15,20,30,25,12,24,20)
property<-c(4,6,3,10,12,8,6,10,8,15)
drug<-c(6,3,3,4,6,9,10,3,2,3)
other<-c(1,3,1,1,6,9,3,6,2,4)

## Create Factor Labels
viol_label<-c(rep("Violent",10))
prop_label<-c(rep("Property",10))
drug_label<-c(rep("Drug",10))
oth_label<-c(rep("Other",10))

## Create Individual Variables
senlen<-c(violent, property, drug, other)
off_type<-c(viol_label, prop_label, drug_label, oth_label)

## Bind Together in One Data Frame
df<-data.frame(senlen, off_type)

## Estimate and Print ANOVA Results
senlen_anova<-aov(senlen~off_type, data=df)
summary(senlen_anova)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## off_type    3 1465.9   488.6    26.06 0.00000000387 ***
## Residuals   36  674.9    18.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A lot easier! But, as before, you may only use this to check your work, not to answer any questions.

Measures of Association

While ANOVA can tell us whether there is a significant relationship between two variables, it cannot tell us anything about the strength of the relationship. We can, however, utilize what we know about variance to compute a measure of the strength of the association between the variables. We have available to us two measures of association called eta-square (η^2) and epsilon-square (ϵ^2), which are computed, respectively, as:

$$\eta^2 = \frac{SS_B}{SS_T}$$

$$\epsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{SS_W/df_W}{SS_T/df_T}$$

Both have an **explained variance** interpretation. They tell us what proportion of the total variance in the dependent variable is “explained” by the independent variable. Epsilon-square is a more conservative estimate, since it takes into account degrees of freedom.

Measures of Association - Sentence Length Example

$$\eta^2 = \frac{SS_B}{SS_T} = \frac{1465.875}{2140.775} = 0.685$$

$$\epsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{18.75}{54.89} = 0.658$$

We interpret these by saying that between 65.8% and 68.5% of the variance in sentence length is explained by offense type. Anything above 50% is generally a strong association.

Measures of Association - Fear of Crime Example

$$\eta^2 = \frac{SS_B}{SS_T} = \frac{344.293}{502.493} = 0.685$$

$$\epsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{5.859}{17.327} = 0.662$$

Between 66.2% and 68.5% of the variance in fear of crime is explained by residential location type.

Post-Hoc Tests

What if I am still interested in those individual group by group contrasts, though? That is, knowing there's significant between-group variation is helpful, but it still doesn't tell me which groups are different from one another. I need a way to maintain the alpha rate at my chosen level and to estimate all the potential group-by-group comparisons in the data.

Luckily, there are multiple commands allow for this and we generally refer to them as **Post-Hoc** tests, or tests we estimate after the estimation of some other model. The **Post-Hoc** tests for ANOVA models may also be referred to as multiple comparison corrections, as they *correct* for the alpha level issue we discussed earlier in lecture. In practice, there are many different multiple comparison correction methods to choose from. I'll just discuss the Bonferroni method in this lecture as it's one of the most commonly used.

Bonferroni Correction

The Bonferroni correction maintains what is known as a **Family-Wise Error Rate** at or below your selected alpha level through simple division. The correction is simply this:

$$\frac{\alpha}{m}$$

Where α is your selected alpha level and m is the number of multiple comparisons you are conducting.

Finding m

Remember the combination calculation we went over weeks ago? I hope so, because you'll need it here. To find the number of multiple comparisons I need to conduct to report all group-by-group contrasts I can use the following equation:

$$m = \frac{k!}{2! * (k - 2)!}$$

Where k is the total number of categories in the independent variable. Those who remember the combination equation will see that this is just a special case of it where $r=2$ and I have substituted k for n .

Sentence Length and Offense Type Example

There were four groups in this example (violent, property, drug, and other) so let's calculate m .

$$m = \frac{k!}{2! * (k - 2)!} = \frac{4!}{2! * (4 - 2)!} = 6$$

We have six total possible comparisons:

Violent Property	Property Drug	Other Drug
Violent Drug	Property Other	
Violent Other		

Now, let's figure out the necessary p-value for standard alpha level.

We could apply these new alpha levels manually, run individual t-tests, then obtain probabilities for each observed test statistic or... we could use a function in R that automatically reports the results from post-hoc multiple comparisons tests.

Alpha Level	Bonferroni Correction	Corrected Alpha Level
$\alpha = .05$	$\frac{.05}{6}$	0.0083
$\alpha = .01$	$\frac{.01}{6}$	0.00167
$\alpha = .001$	$\frac{.001}{6}$	0.000167

The function to estimate a Bonferroni correction looks like this:

```
pairwise.t.test(x, g, p.adjust.method="bonferroni")
```

Where **x** is the continuous outcome variable from the ANOVA, **g** is the group variable from the ANOVA, and **p.adjust.method="bonferroni"** tells R to use the Bonferroni method for multiple comparisons correction.

As noted before, the Bonferroni method is one of many, so this syntax could be used to estimate a variety of different multiple comparison corrections. Let's see what the results look like for this example.

```
with(df, pairwise.t.test(senlen, off_type, paired=FALSE,
                          p.adjust.method="bonferroni",
                          pool.sd=FALSE))
```

```
##
## Pairwise comparisons using t tests with non-pooled SD
##
## data:  senlen and off_type
##
##      Drug   Other   Property
## Other  1.00000 -      -
## Property 0.22121 0.03256 -
## Violent  0.00036 0.00016 0.00373
##
## P value adjustment method: bonferroni
```

The values in the table display the adjusted p-values for each comparison. The adjusted p-values represent the probability of observing the test statistic (or a more extreme test statistic) if the null hypothesis is true, after adjusting the probability values for the Bonferroni correction.

Unfortunately, the table does not display the actual group mean differences as well (nor can I find a function that neatly does so). The following code accomplishes this in a few steps. First, I need to revise my data frame so that each column represents a list of sentence lengths for the different types of crimes. The rows are not meaningful at first, but will be by the end of this first block of code.

```
obs<-c(1:10)
df<-data.frame(obs, violent, property, drug, other)

df<- df %>%
  gather(key="off_type", value="senlen", violent, property, drug, other) %>%
  convert_as_factor(obs, off_type)

head(df, 5)
```

```
##  obs off_type senlen
## 1    1  violent      6
## 2    2  violent     18
## 3    3  violent     20
## 4    4  violent     15
## 5    5  violent     20
```

The data are now in what is usually known as “long” format, but it’s not quite the right term for what we have here. Typically, “long” format data have multiple rows nested in some aggregated unit - for example, multiple time period observations for a single person. Here, each row is just a person with columns for the offense type they were convicted of and the sentence length they received. Terminology issues aside, we can now use a new function (`pairwise_t_tests`) to automatically detect all of the possible pairwise comparisons and group the results into one object. I use this new function because the other one (`pairwise.t.tests`) will not work with the pipe pipe operator I use below to compile the results.

```
tests <- df %>%
  pairwise_t_test(senlen~off_type,
    paired=FALSE, pool.sd=FALSE,
    p.adjust.method="bonferroni",
    detailed=TRUE)
```

I don’t want paired tests here (which is another way to refer to non-independent samples t-tests) so I set the paired option to FALSE. I also do not want the test to assume the variances are equal across groups, so I set `pool.sd=FALSE`. The `detailed=TRUE` option provides more detailed statistics in the output that you can look at after creating the option. I do not provide a summary of the object here because it does not fit properly on a slide or in a PDF - there are too many columns.

My next step is to display the results - I use a few new functions here to do so. The first is the `ggboxplot()` function, which is a more direct way of making a box plot that includes points for the individual values (it’d be like combining a box plot and a scatterplot with regular ggplot functions, but I can skip a step and use this one function).

Next, the `add_xy_positions()` function takes the `tests` object I just created and adds xy coordinates to plot p-values to the tests object. I can then add these coordinates to the existing plot with the `stat_pvalue_manual()` function, which just applies the significance labels (the stars) from the `tests` object to the comparison coordinates.

```
tests_plot<- ggboxplot(df, x = "off_type", y = "senlen", add = "point")
tests<-tests %>% add_xy_position(x="off_type")
tests_plot+stat_pvalue_manual(tests, label="p.adj.signif")
```

