Lecture 07 - Analysis of Variance

Samuel DeWitt

General Outline

For this lecture, we will be discussing hypothesis testing with three or more population means, also known as analysis of variance (ANOVA).

1) Logic of ANOVA
2) Computation of ANOVA
3) Computing ANOVA in R
4) Measures of Association
5) Post-Hoc Tests

Inference with Three or More Sample Means

Inference with three or more sample means is a fairly simple generalization of the two-sample setup.

Now, instead of examining the difference between two specific means, we will instead ask if there is a significant amount of variability **between** the set of three or more samples means as compared to the distribution of values **about** these sample means within each sample.

Inference with Three or More Sample Means

- Research Question
  - Do the differences we observe among the sample means indicate that there are significant differences across groups in the population?

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

Inference with Three or More Sample Means

You might ask - why not simply run t-tests between each unique group pairing?

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{(n_1-1)} + \frac{s_2^2}{(n_2-1)}}}$$

Here's why:

- It's cumbersome
  - 3 groups: 3 tests; 4 groups: 6 tests; 5 groups: 10 tests
- Probability of committing a type I error on any given test is greater than α (.05, for example)
  - 3 groups: p=.143; 4 groups: p=.185; 5 groups: p=.226

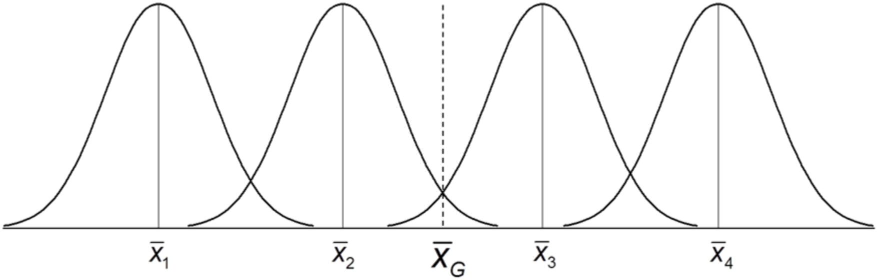
Logic of ANOVA

We want a statistical test that will help us decide whether the observed differences are the result of sampling variation or real differences.

- **An**alysis **of** **va**riance = ANOVA
  - Three or more sample means
  - Global test = joint significance of several means
  - Constant prob. of type I error (α)

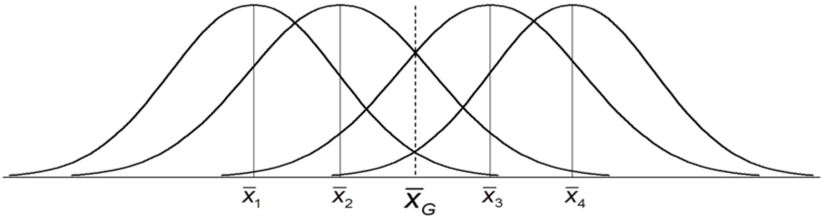
Why Variance?

- Analysis of Variance
  - Ratio of the variability **between** groups to the variability **within** groups
  - New test statistics - the F-ratio



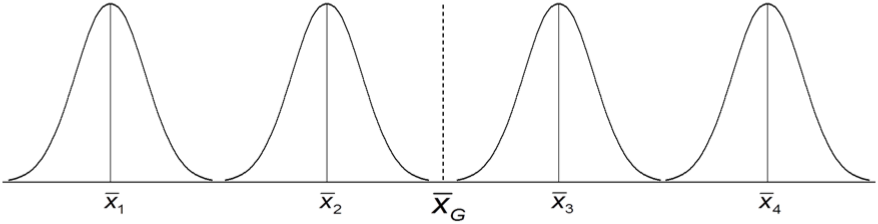
Why Variance? (cont)

- More variability **within** groups than between
  - F-ratio < 1.0
  - Too much overlap, there is no relationship between group membership and the outcome



Why Variance? (cont)

- More variability **between** groups than within.
  - F-ratio > 1.0
  - Little or no overlap, there is a relationship between group membership and the outcome



New Probability Distribution

- The F-distribution
  - The ratio of variability **between** groups to variability **within** groups.

Large F-ratios (i.e., significantly greater than 1.0) will lead us to reject the null hypothesis of no association between group membership and the outcome of interest.

New Calculations - Sum of Squares

- Sum of squares = numerator of variance
  - Total variation about the **grand mean** ($df_T = N - 1$)

$$SS_T = \sum (\mathrm{x}_{ik} - \overline{\mathrm{x}}_G)^2 = \sum \mathrm{x}_{ik}^2 - N\overline{\mathrm{x}}_G^2$$

- Between-group sum of squares ($df_B = k - 1$)

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

- Within-group sum of squares ($df_W = N - k$)

$$SS_W = \sum (\mathrm{x}_{ik} - \overline{\mathrm{x}}_k)^2 = \sum \mathrm{x}_{ik}^2 - \sum n_k\overline{\mathrm{x}}_k^2 = SS_T - SS_B$$

New Calculations - Sum of Squares

- Total sum of squares (SS_T)
  - $SS_T = SS_B + SS_W$
- Mean square (i.e., variance)
  - Mean square between: $MS_B = \frac{SS_B}{df_B}$
  - Mean square within: $MS_W = \frac{SS_W}{df_W}$
- F-ratio
  - $F = \frac{MS_B}{MS_W}$

Practical Example - Offense Type and Sentence Length

Violent	Property	Drug	Other
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

Quick Note - How to Find F_{crit}

Much like prior distributions, there are functions in R to calculate critical scores for the F-ratio. The function takes the following form:

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

Where **p** is your alpha level, **df1** is the between group degrees of freedom, **df2** is the within group degrees of freedom, and **lower.tail=TRUE/FALSE** returns the upper or lower-tail F-score that leaves that cumulative probability below or above it.

Offense Type and Sentence Length - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $\text{H}_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$
  - $\text{H}_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

- **Step 2 - Obtain a probability distribution**:
  - F-distribution; $df_B = k - 1 = 4 - 1 = 3$, $df_W = N - k = 40 - 4 = 36$
- **Step 3 - Make decision rules**:
  - $\alpha = .01$, $F_{crit} = 4.377$, Reject H_0 if TS > 4.377

Offense Type and Sentence Length - Hypothesis Test

- **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

Offense Type and Sentence Length - Hypothesis Test

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.

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

Offense Type and Sentence Length - Hypothesis Test

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.

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

Offense Type and Sentence Length - Hypothesis Test

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	
Within groups	674.9	$N-k=36$	18.747	$\frac{488.62}{18.747} = 26.06$
Total	2140.775	$N-1=39$	54.892	

- **Step 5 - Make a decision about the null hypothesis**:
  - Reject H_0, conclude that offense type is significantly associated with sentence length.

Another Example - Residential Location and Fear of Crime

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

Residential Location and Fear of Crime - Hypothesis Test

- **Step 1: State hypotheses**
  - $\text{H}_1 : \mu_U \neq \mu_S \neq \mu_R$; $\text{H}_0 : \mu_U = \mu_S = \mu_R$
- **Step 2: Obtain a probability distribution**
  - F-distribution, $df_B = 3 - 1 = 2$, $df_W = 30 - 3 = 27$
- **Step 3: Make decision rules**
  - $\alpha = .05$ $F_{crit} = 3.354$; reject H_0 if $F > 3.354$

Residential Location and Fear of Crime - Hypothesis Test

- **Step 4: Calculate the test statistic**
- $\overline{x}_1 = \sum \mathrm{x}_{i1}/n_1 = 293/10 = 29.3$
- $\overline{x}_2 = \sum \mathrm{x}_{i2}/n_2 = 243/10 = 24.3$
- $\overline{x}_3 = \sum \mathrm{x}_{i3}/n_3 = 210/10 = 21$
- $\overline{x}_G = \sum \mathrm{x}_i/N = \frac{293+243+210}{10+10+10} = 24.87$

Residential Location and Fear of Crime - Hypothesis Test

- **Step 4: Calculate the test statistic**
- $SS_T = \sum \mathrm{x}_i^2 - N\overline{\mathrm{x}}_G^2 = (8663 + 5925 + 4470) - 30(24.87)^2 = 502.493$
- $SS_B = \sum \mathrm{n}_k \overline{\mathrm{x}}_k^2 - N\overline{\mathrm{x}}_G^2 = 10(29.3)^2 + 10(24.3)^2 + 10(21.0)^2 - 30(24.87)^2 = 344.293$
- $SS_W = \sum \mathrm{x}_i^2 - \sum \mathrm{n}_k \overline{\mathrm{x}}_k^2 = SS_T - SS_B = 502.493 - 344.293 = 158.2$

Residential Location and Fear of Crime - Hypothesis Test

- **Step 4: Calculate the test statistic**

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

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

```r
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)
prop_xbar<-sum(property)/length(property)
drug_xbar<-sum(drug)/length(drug)
oth_xbar<-sum(other)/length(other)
grand_xbar<-(sum(violent)+sum(property)+sum(drug)+sum(other))/
  (length(violent)+length(property)+length(drug)+length(other))
```

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 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 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))
```

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_with
f_ratio
```

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

ANOVA Example - Automatic R Method for Sentence Length Data

```r
## 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
```

ANOVA and Measures of Association

ANOVA tells us whether there is a significant relationship between two variables, but it cannot tell us about the strength of the relationship

- Two measures of association
  - Eta square (η^2)
  - Epsilon square (ϵ^2)

ANOVA and Measures of Association

- Eta square (η^2)

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

- Epsilon square (ϵ^2)

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

- Explained variance interpretation
  - Proportion of the total variance in the dependent variable that is **explained** by variation in the independent variable.

Measures of Association - Offense Type and Sentence Length

- Eta square (η^2)

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

- Epsilon square (ϵ^2)

$$\epsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{SS_W/df_W}{SS_T/df_T} = 1 - \frac{18.75}{54.89} = 0.658$$

- Interpretation
  - Between 65.8% and 68.5% of the variance in sentence length is explained by offense type.

Measures of Association - Residential Location and Fear of Crime

- Eta square (η^2)

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

- Epsilon square (ϵ^2)

$$\epsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{SS_W/df_W}{SS_T/df_T} = 1 - \frac{5.859}{17.327} = 0.662$$

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

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.

Post-Hoc Tests

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.

Post-Hoc Tests - 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 just this:

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

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

Post-Hoc Tests - 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.

Post-Hoc Tests - Sentence Length and Offense Type

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		

Post-Hoc Tests - Sentence Length and Offense Type

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

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

Post-Hoc Tests - Sentence Length and Offense Type

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.

Post-Hoc Tests - Sentence Length and Offense Type

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.

Post-Hoc Tests - Sentence Length and Offense Type

```
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
```

Post-Hoc Tests - Sentence Length and Offense Type

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.

Post-Hoc Tests - Sentence Length and Offense Type

Unfortunately, the table does not display the actual group mean differences as well (nor can I find a function that neatly does so).

The code on the following slides accomplishes this in a few steps.

Post-Hoc Tests - Sentence Length and Offense Type

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
```

Post-Hoc Tests - Sentence Length and Offense Type

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.

Post-Hoc Tests - Sentence Length and Offense Type

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 operator (%>%) I use on the next slide to compile the results.

Post-Hoc Tests - Sentence Length and Offense Type

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

Post-Hoc Tests - Sentence Length and Offense Type

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.

Post-Hoc Tests - Sentence Length and Offense Type

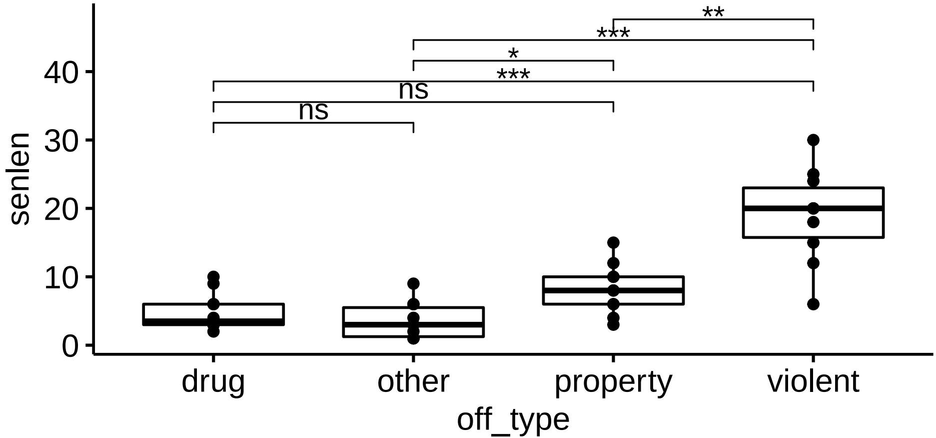
My next step is to display the results - I use a few new functions here to do so.

- The first is the ggboxplot() function,
  - A more direct way of making a box plot that includes points for the individual values.
- Next, the add_xy_positions() function
  - This takes the *tests* object I just created and adds xy coordinates to plot p-values.

```
tests_plot<- ggboxplot(df, x = "off_type", y = "senlen", add = "point")
tests<-tests %>% add_xy_position(x="off_type")
```

Post-Hoc Tests - Sentence Length and Offense Type

```
tests_plot+stat_pvalue_manual(tests, label="p.adj.signif")
```



Two Questions

What are your two questions today?