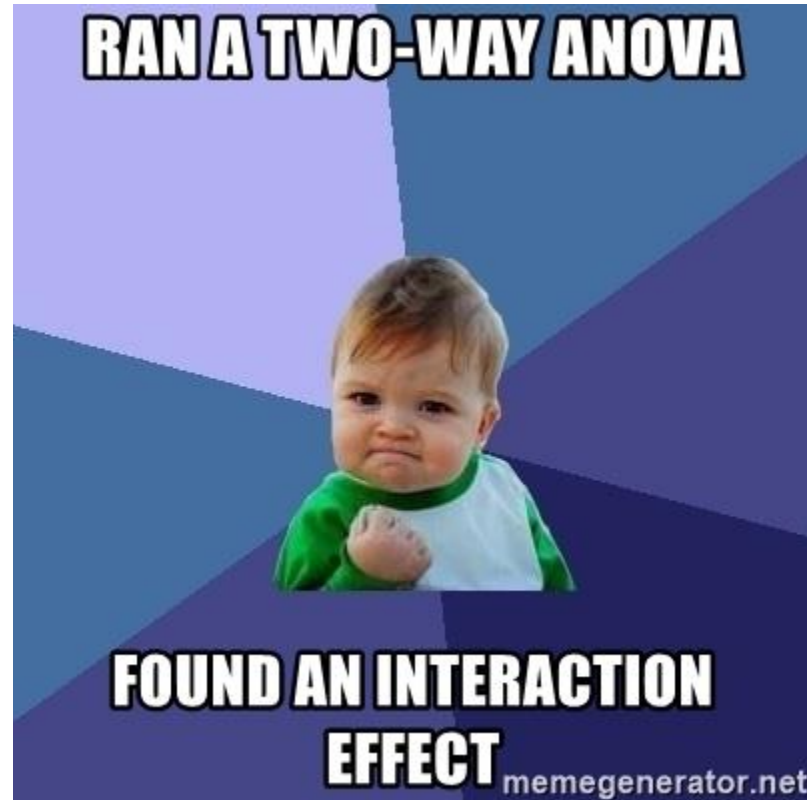


Analysis of Variance continued



Overview

Review/continuation of one-way ANOVA

Pairwise comparisons after running an ANOVA

Factorial ANOVAs and interaction effects

If there is time: unbalanced designs

Announcement: Workshop on practical data analysis for final project

Nathan will be holding a workshop on practical data analysis for the final project

Date: Monday, December 5th at 7:30pm

Location: Bass L06-A and on zoom <https://yale.zoom.us/j/99657825090>

The workshop is meant to supplement what you've learned in class with a more practical twist -- things like:

- Easy ways to format your plots, regressions, and code nicely,
- What working directories are, a bit on major data sources like the ACS and IPUMS
- (briefly) functional programming techniques for cleaner and more scalable code
- (briefly) frameworks for reproducible analyses and code
- This will probably not help you for the final exam in any way whatsoever, and most of the tips covered aren't strictly necessary to use for the final project. But I hope they will help you make better reports and analyses for this project and down the line.
- Please email Nathan <nathan.kim@yale.edu> for any topic suggestions or questions

One-way ANOVA

A **one-way analysis of variance (ANOVA)** is a parametric hypothesis test that can be used to examine if a set of means are all the same.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_A: \mu_i \neq \mu_j \text{ for some } i, j$$

The statistic we use for a one-way ANOVA is the F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1} \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

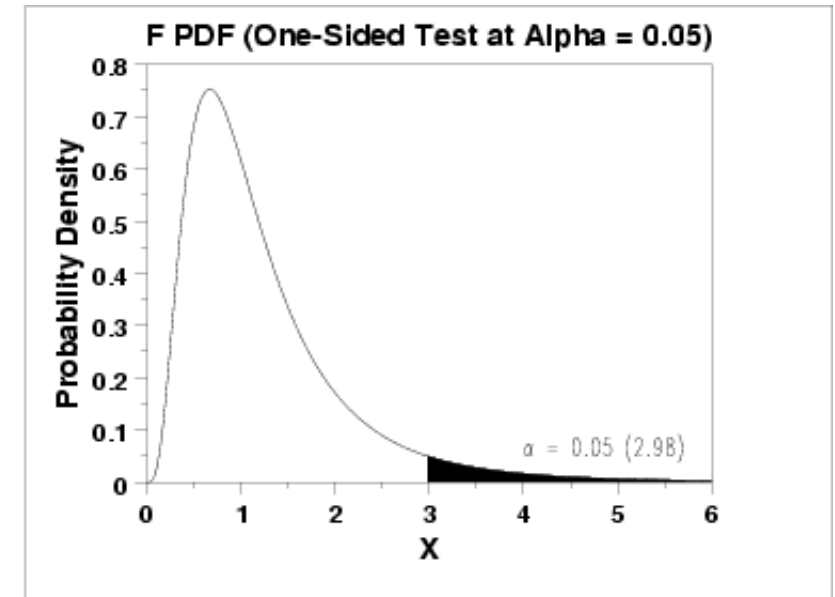
One-way ANOVA – the central idea

If H_0 is true, the F-statistic will come from an F distribution with parameters

- $df_1 = K - 1$
- $df_2 = N - K$

The F-distribution is valid if these conditions are met:

- The data in each group should follow a normal distribution
 - Check this with a Q-Q plot
- The variances in each group should be approximately equal
 - Check that $s_{\max}/s_{\min} < 2$



ANOVAs are robust to these assumptions, but what can we do if they are very badly violated?

Kruskal-Wallis (non-parametric) test

There are also **non-parametric** tests which don't make assumptions about normality

The **Kruskal-Wallis** test compares several groups to see if one of the groups 'stochastically dominates' another

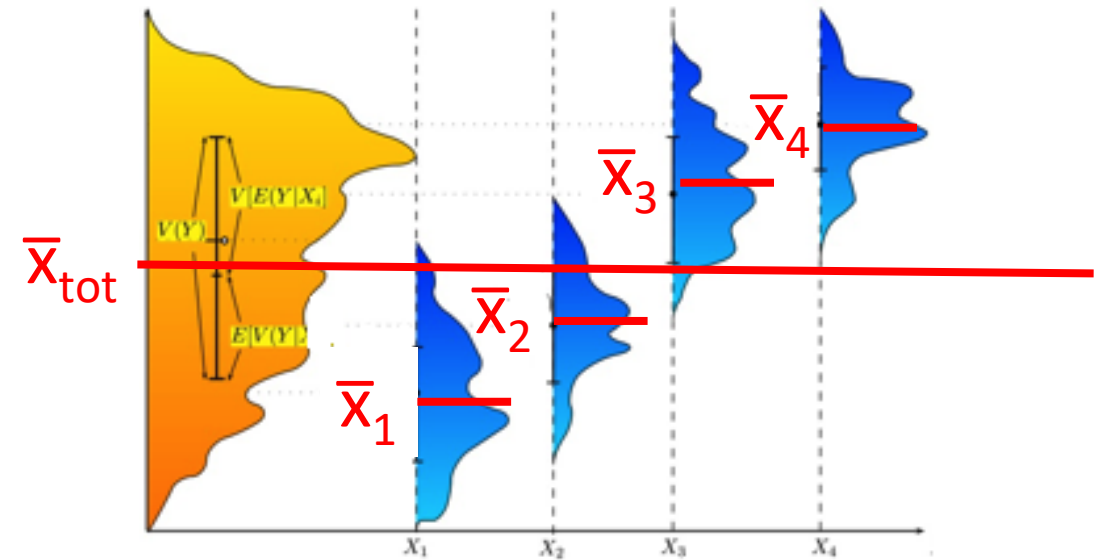
- Does not assume normality
- Tests if one group stochastically dominates another group
- Also tests whether the median for all the groups are the same
 - (if you assume groups have the same shaped and scale)
- The test is based on ranks so it is not influenced by outliers

The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1} \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\text{variability between group means}}{\text{variability within each group}}$$



ANOVA table

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1} \sum_{i=1}^K n_i (\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

Source	df	Sum of Sq.	Mean Square	F-statistic	p-value
Groups	$k - 1$	SSG	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSE}$	Upper tail $F_{k-1,n-k}$
Error	$n - k$	SSE	$MSE = \frac{SSE}{n-k}$		
Total	$n - 1$	$SSTotal$			

Where:

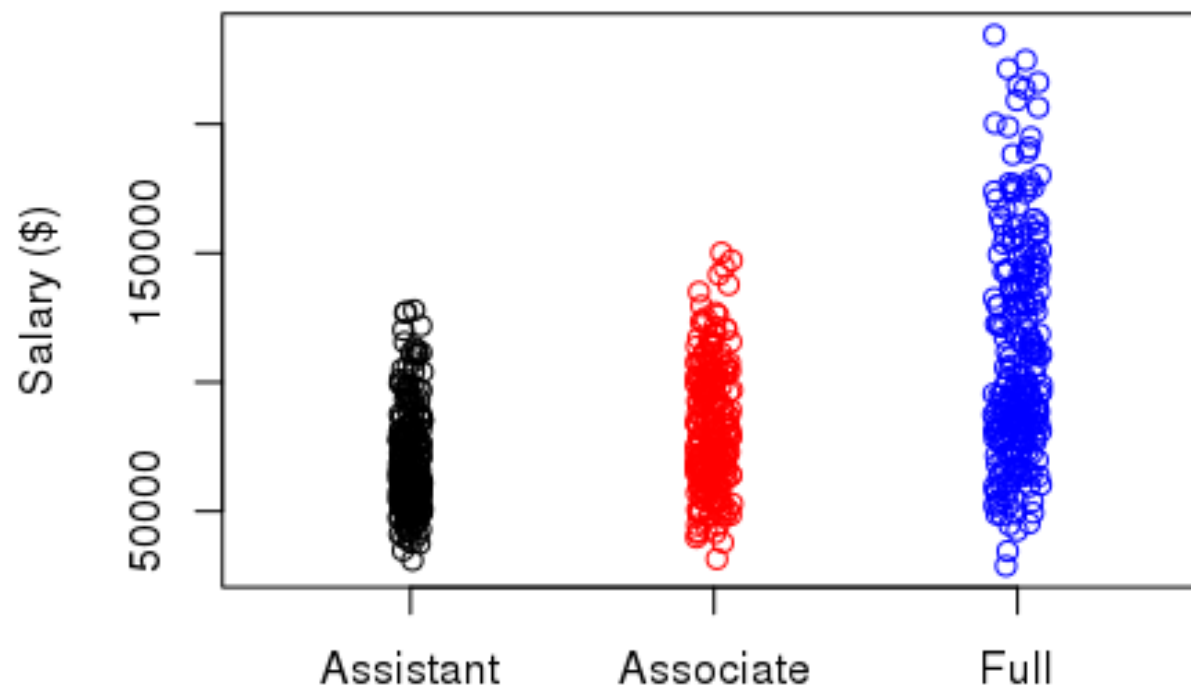
$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{tot})^2$$

$$SSG = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{tot})^2$$

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

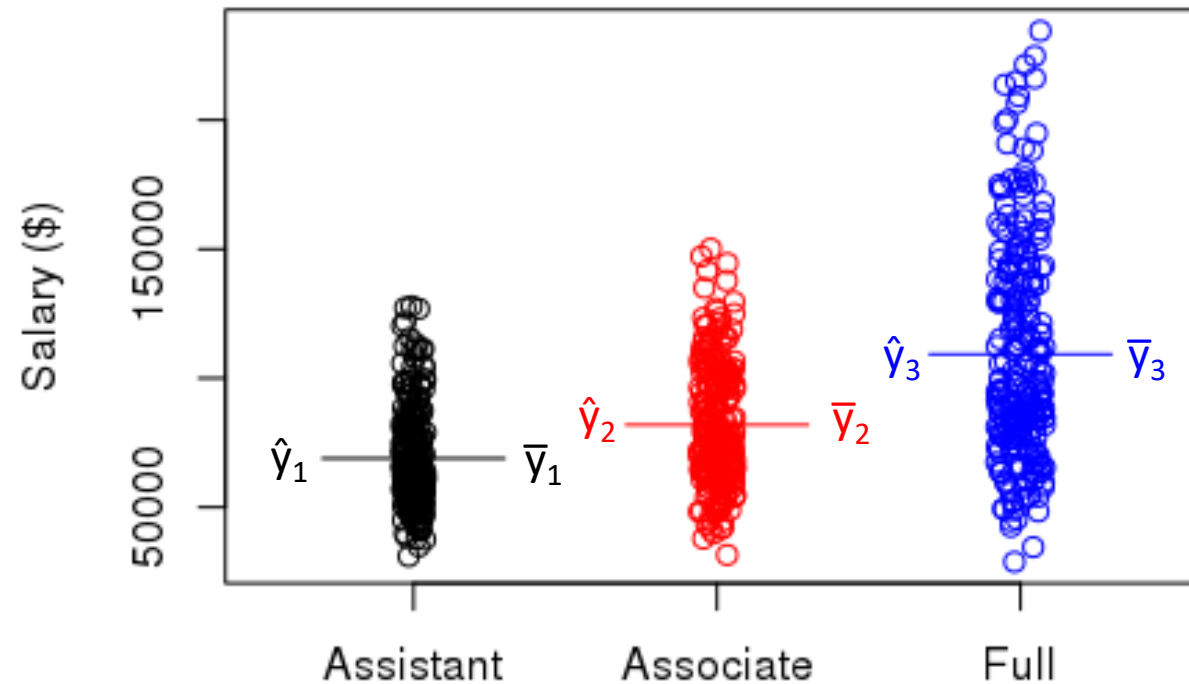
$SST = SSG + SSE$

ANOVA as regression with only categorical predictors



$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if Assistant Professor} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if Associate Professor} \\ \beta_0 + \epsilon_i & \text{if Full Professor} \end{cases}$$

Least squares prediction for \hat{y}_i is \bar{y}_k



$$\hat{y}_i = \bar{y}_k = \begin{cases} \bar{y}_1 & \text{if Assistant professor} \\ \bar{y}_2 & \text{if Associate professor} \\ \bar{y}_3 & \text{if Full} \end{cases}$$

Let's pick up where we left off last class exploring this in R...

Silly question: Do Assistant, Associate and Fully Professors get paid the same on average?



Planned comparisons/posthoc tests

Suppose we run a one-way ANOVA and we are able to reject the null hypothesis.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_A: \mu_i \neq \mu_j \text{ for some } i, j$$

Q: What else would we like to know?

Pairwise comparisons

There are several tests that can be used to examine which pairs of means differed; i.e., to test:

- $H_0: \mu_i = \mu_j$
- $H_A: \mu_i \neq \mu_j$

These tests include:

- Fisher's Least Significant Difference
- Bonferroni procedure/correction
- Tukeys Honest significantly different

Fisher's Least Significant Difference (LSD)

1. Perform the ANOVA
2. If the ANOVA F-test is not significant, stop
3. If the ANOVA F-test is significant, then you can test H_0 for a pairwise comparisons using:

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE \cdot (\frac{1}{n_i} + \frac{1}{n_j})}}$$

Estimate of the SE

Uses the MSE as a pooled estimate of the σ^2

Use a t-distribution with n-k degrees of freedom

Very 'liberal' tests

- Likely to make Type I errors (lots of false rejections of H_0)
- Less likely to make Type II errors (highest chance of detecting effects)

Bonferroni correction

Controls for the ***family-wise error rate***

- i.e., $\alpha = 0.05$ for making ***any*** Type I error ***over all pairs of comparisons***

1. Choose an α -level for the family-wise error rate α
2. Decide how many comparisons you will make. Call this m .
3. Reject any hypothesis tests that have p-values less than α/m
 - Pairwise tests typically done using a t-statistic, where the MSE is used in the estimate of the SE

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE \cdot (\frac{1}{n_i} + \frac{1}{n_j})}} \quad \text{Use a t-distribution with } n-k \text{ degrees of freedom}$$

Very 'conservative' tests

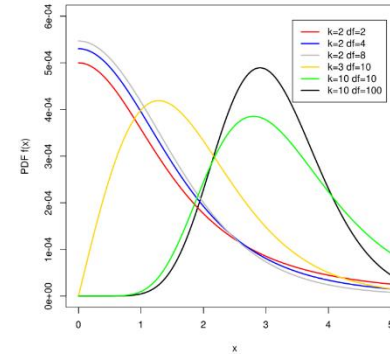
- Unlikely to make Type I errors (few false rejections of H_0)
- Likely to make Type II errors (insensitive at detecting real effects)

Tukey's Honest Significantly Different Test

Tukey's Honest Significantly Different test controls for the family-wise error rate but is less conservative than the Bonferroni correction

If the null hypothesis was true, q comes from a ***studentized range distribution***

$$q = \frac{\sqrt{2}(\bar{x}_{max} - \bar{x}_{min})}{\sqrt{MSE \cdot (\frac{1}{n_{max}} + \frac{1}{n_{min}})}}$$



We can compare $q = \frac{\sqrt{2}(\bar{x}_i - \bar{x}_j)}{\sqrt{MSE \cdot (\frac{1}{n_i} + \frac{1}{n_j})}}$ for a pair of means i, j , to a studentized range distribution with parameters k , and $N-k$, to get a p-value

- Still based on assumptions that the data in each group is normal with equal variance

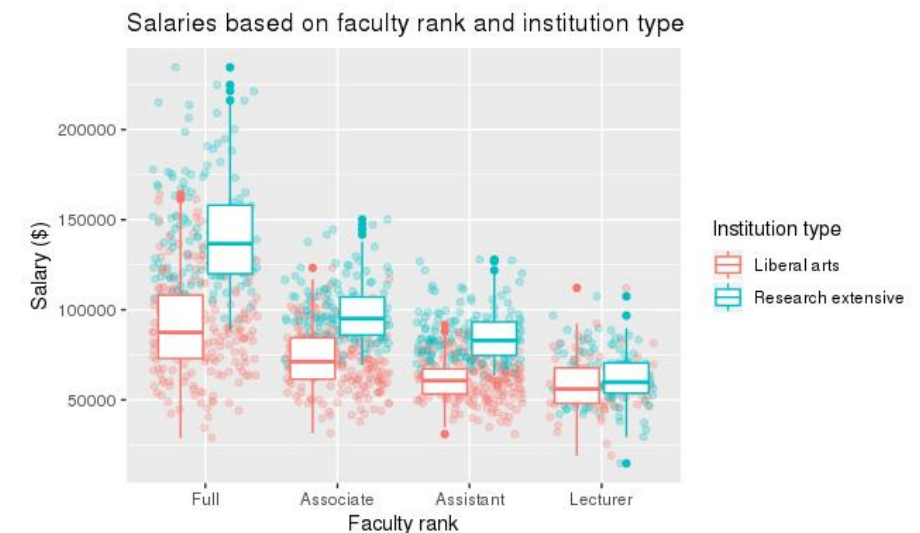
Let's try pairwise comparisons in R...

Factorial ANOVA

In a **factorial ANOVA**, we model the response variable y as a function of **more than one** categorical predictor

Example 1: Do faculty salaries depend on faculty rank, and the type of college/university

- Factors are:
 - **Rank:** Lecturer, Assistant, Associate, Full
 - **Institute:** liberal arts college, research university
 - 4 x 2 design



Factorial ANOVA

Example 2: A student at Queensland University of Technology conducted an experiment to determine what types of sandwiches ants prefer.

- Factors he looked at were:
 - **Bread:** rye, whole wheat multigrain, white
 - **Filling:** peanut butter, ham and pickle, and vegemite
 - 4 x 3 design

The student creating 4 sandwiches of all combinations of bread and filling (48 sandwiches total) and randomly left pieces in front of ant nests.

He then measured how many ants were on the sandwiches 5 minutes later.

Two-way ANOVA hypotheses

Main effect for A (bread type doesn't matter or institution type doesn't matter)

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_j = 0$$

$$H_A: \alpha_j \neq 0 \text{ for some } j$$

Where:

Main effect for B (filling doesn't matter)

$$H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$$

$$H_A: \beta_k \neq 0 \text{ for some } k$$

α_j : is the “effect” for factor A at level j

β_k : is the “effect” for factor B at level k

Interaction effect:

$$H_0: \text{All } \gamma_{jk} = 0$$

$$H_A: \gamma_{jk} \neq 0 \text{ for some } j, k$$

γ_{jk} : is the interaction between level j of factor A, and level k of factor B.

Two-way ANOVA in R with interaction

Source	df	Sum of Sq.	Mean Square	F-stat	p-value
Factor A	K - 1	SSA	$MSA = SSA/(K-1)$	MSA/MSE	$F_{K-1, KJ(c-1)}$
Factor B	J - 1	SSB	$MSB = SSB/(J-1)$	MSB/MSE	$F_{J-1, KJ(c-1)}$
A x B	(K-1)(J-1)	SSAB	$MSAB = SSAB/(K-1)(J-1)$	$MSAB/MSE$	$F_{(K-1)(J-1), KJ(c-1)}$
Error	KJ(c - 1)	SSE	$MSE = SSE/(K-1)(J-1)$		
Total	N - 1	SSTotal			

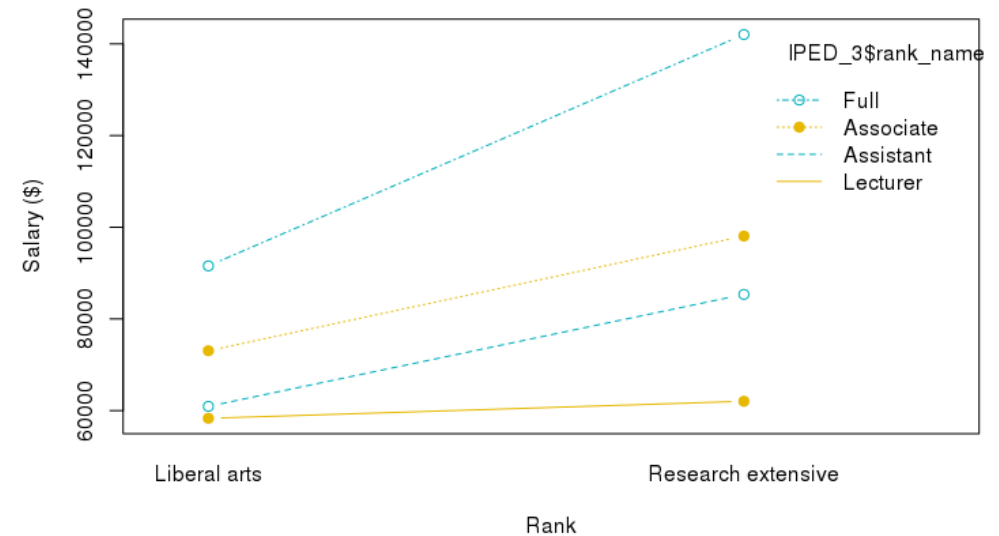
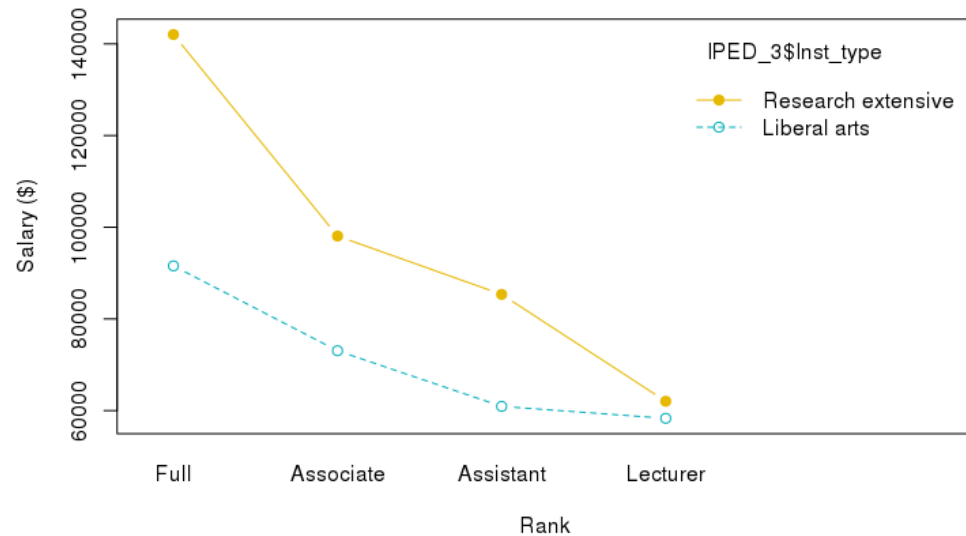
For balanced design: $SSTotal = SSA + SSB + SSAB + SSE$

ANOVA table for a balanced design with c replicates in each group

Interaction plots

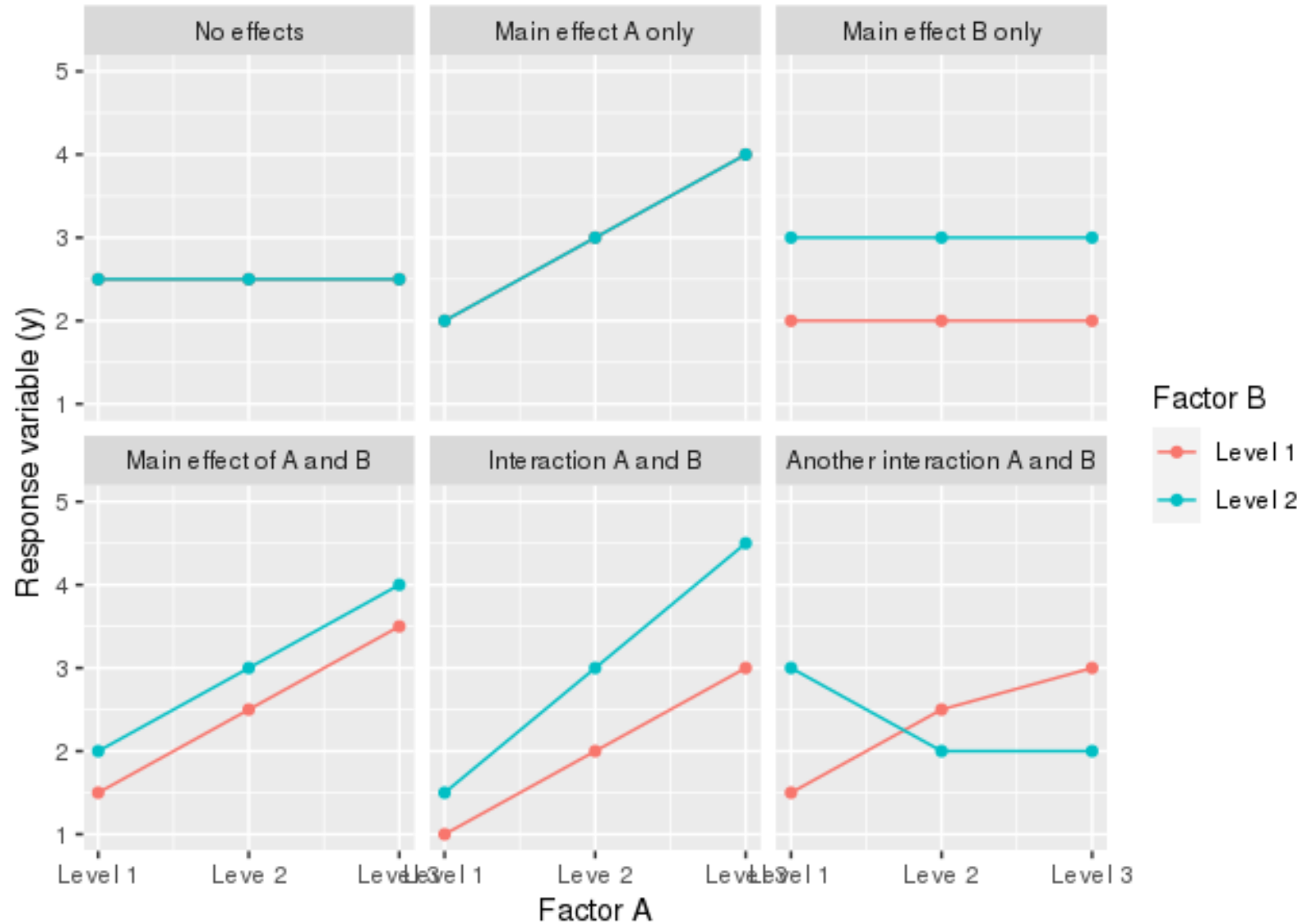
Interaction plots can help us visualize main effects and interactions

- Plot the levels of one of the factors on the x-axis
- Plot the levels of the other factor as separate lines



Either factor can be on the x-axis although sometimes there is a natural choice

Interpreting interaction plots



Interpreting interactions

When interactions are present, one must be careful interpreting main effects

- i.e., the value of one factor A, depends on the value of second factor B

For example, suppose you want to determine which condiment is the most enjoyable, chocolate sauce or mustard

- Run a 2 x 2 ANOVA, 20 people each condition
- Get rating of enjoyment

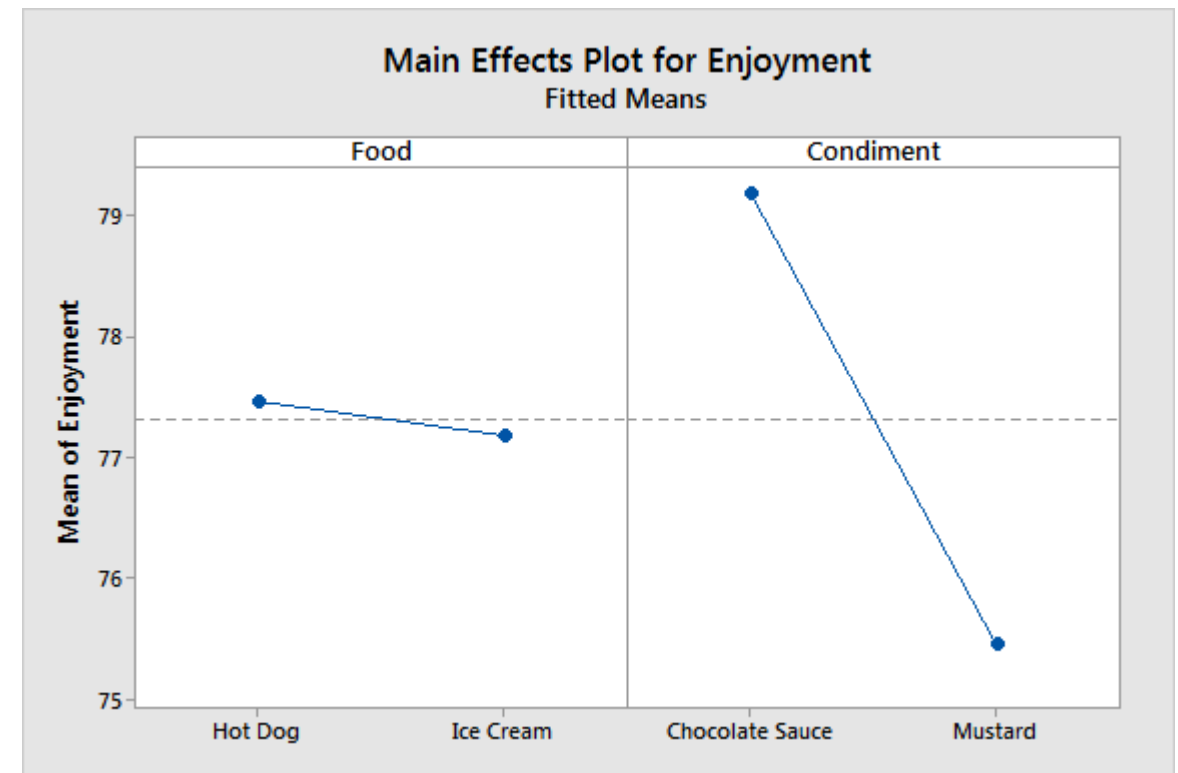
Number of participants	Ice cream	Hot dog
Chocolate sauce	20	20
Mustard	20	20



Model with only effects

```
> summary(aov(Enjoyment ~ Food + Condiment, data = condiments_food ))
```

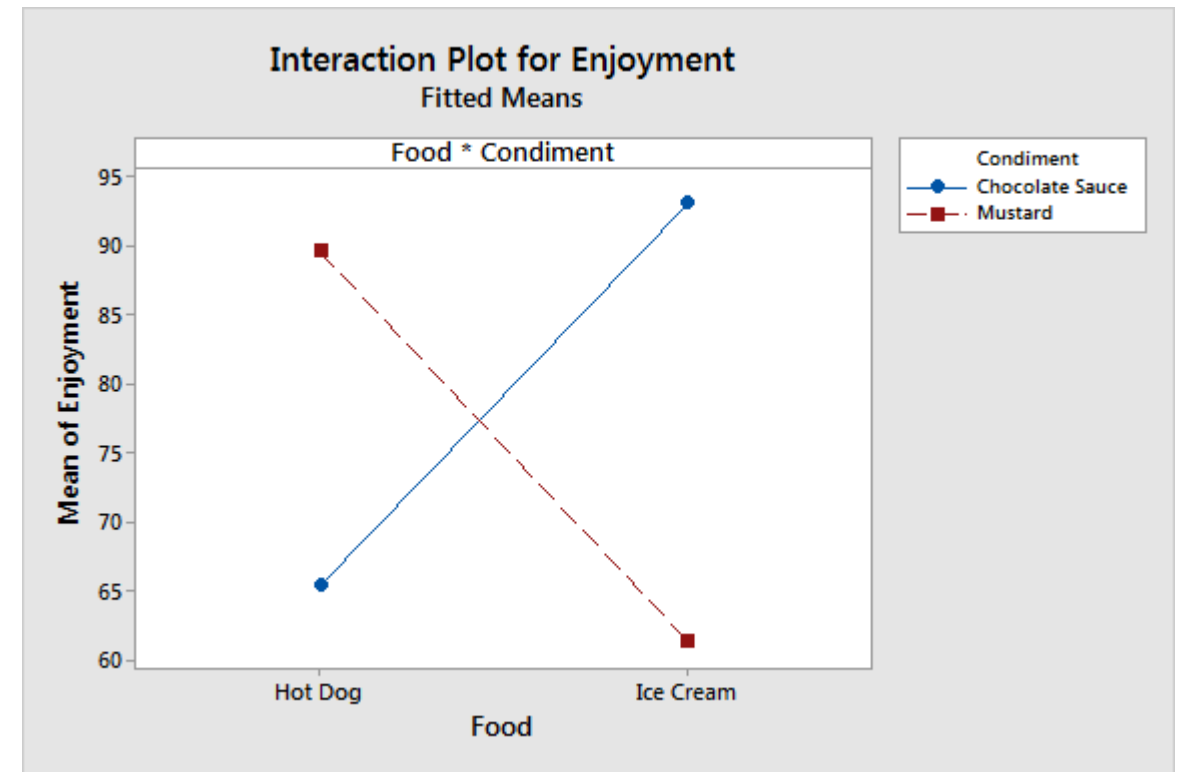
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Food	1	2	1.6	0.007	0.934
Condiment	1	278	277.5	1.214	0.274
Residuals	77	17601	228.6		



Model with interactions

```
> summary(aov(Enjoyment ~ Food*Condiment, data = condiments_food ))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Food	1	2	2	0.064	0.80136
Condiment	1	278	278	11.071	0.00135 **
Food:Condiment	1	15696	15696	626.153	< 0.0000000000000002 ***
Residuals	76	1905	25		



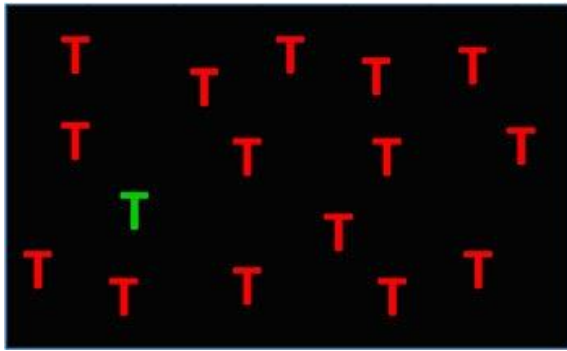
Let's examine two-way ANOVAs in R...

Homework 10

You will analyze data from a psychophysics experiment that explored popout attention

- Study done at Hampshire College by Jacob Prescott, Tapujit Debnath Tapu, Julian Oks, Kirsten Lydic

Background:

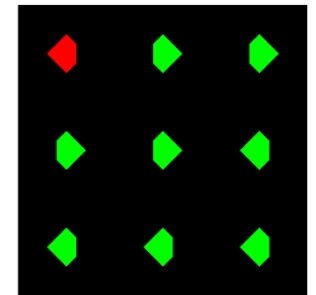
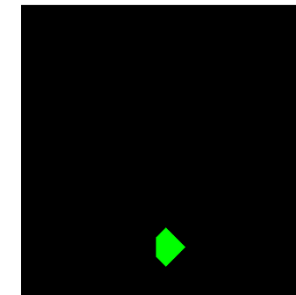


Exogenous attention



Endogenous attention

Single item Multiple items

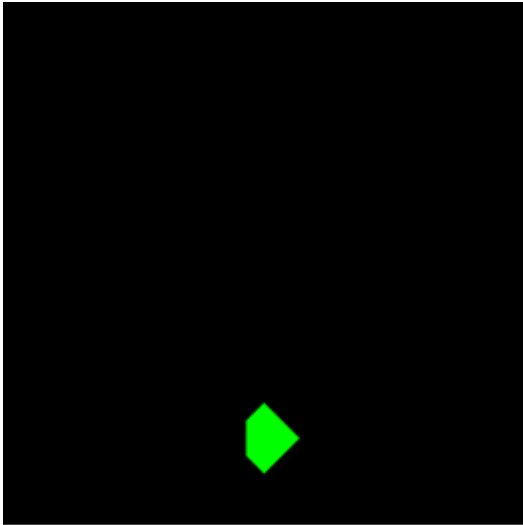


Do reaction times differ for:

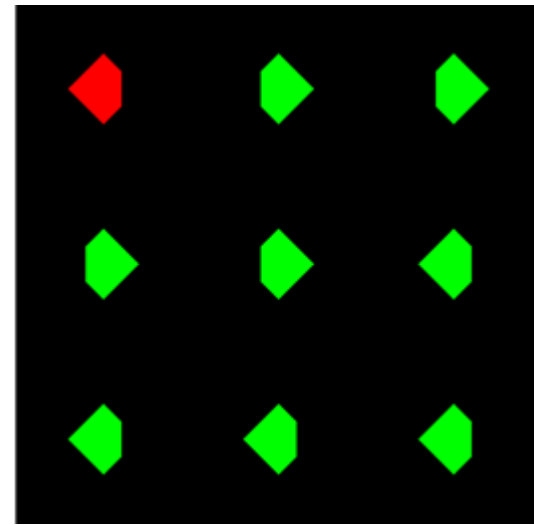
1. Position of target stimulus
2. Single vs. multiple item displays

Homework 10

Participants engaged in a reaction time task where they needed to respond as *quickly* and as *accurately* as possible



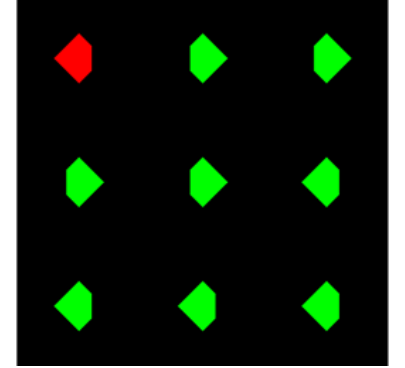
Press "z" because left side is cut off



Press "/" because right side is cut off

You can try it yourself at: https://run.pavlovvia.org/xfang/popout_exp1/html/

Homework 10



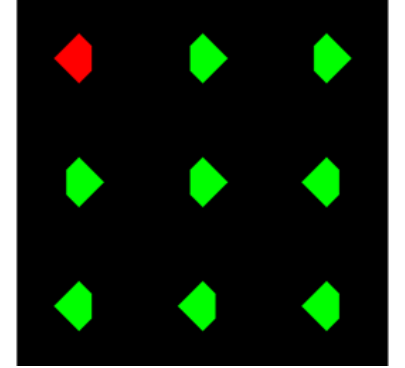
The experiment had a $9 \times 2 \times 2 \times 2$ **factorial** design:

1. Position (9 levels): 9 locations where the target stimulus could appear
2. Isolated/distractor condition (2 levels): isolated or cluttered display
3. Target color (2 levels): red or green target
 - For cluttered displays, the distractors always had the opposite color of the target
4. Cut direction (2 levels): left or right side of the target diamond was cut off
 - Corresponds to pressing the "z" or "/" key

The experiment had 10 blocks where all 72 ($9 \times 2 \times 2 \times 2$) stimuli were shown

8 volunteer participants participated in the experiment

Homework 10



On homework 10 you will run:

- A one-way ANOVA to see if the mean reaction time is the same at all target positions
- A two-way ANOVA to look at how both position and isolated/cluttered displays affect mean reaction times.
- Explore another question using this data

Questions?

Complete and balanced designs

Complete factorial design: at least one measurement for each possible combination of factor levels

- E.g., in a two-way ANOVA for factors A and B, if there are K levels for factor A, and J levels for factor B, then there needs to be at least one measurement for each of the KJ levels

Balanced design: the sample size is the same for all combination of factor levels

- E.g., there are the same number of samples in each of the KJ level combinations.
- The computations and interpretations for non-balanced designs are a bit harder.

Unbalanced designs

For unbalanced designs, there are different ways to compute the sum of squares, and hence one can get different p-values

- The problem is analogous to multicollinearity. If two explanatory variables are correlated either can account for the variability in the response data.

Type I sum of squares, (also called sequential sum of squares) the order that terms are entered in the model matters.

- `anova(lm(y ~ A*B))` gives different results than using `anova(lm(y ~ B*A))`
- $SS(A)$ is taken into account before $SS(B)$ is considered etc.

Type III sum of squares, the order that that terms are entered into the model does not matter.

- `Car::Anova(lm(y ~ A*B) , type = "III")` is the same as `car::Anova(lm(y ~ B*A) , type = "III")`
- For each factor, $SS(A)$, $SS(B)$, $SS(AB)$ is taken into account after all other factors are added

Let's examine it R...