Intro to Two-Way ANOVA

Keegan Korthauer
Department of Statistics
University of Wisconsin

Recap: One-Way ANOVA

- Allows us to determine if significant differences of a response variable exist across 3 or more levels (treatments) of a factor
- Null hypothesis: all treatment means are equal
- Assume normality and equality of variance
- F test statistic: ratio of MSTr over MSE
- Organize results in an ANOVA table
- If the results allow us to reject the null, we know to perform pairwise comparisons to find which treatments are different from the rest

Recap: Pairwise Comparison

- If we are only interested in one specific comparison of treatments i and j:
  - Use Fisher's Least Significant Difference method for CI/HT

- If we are interested in all pairwise combinations:
  - Use Bonferroni's simultaneous CIs or HTs if there is a small number of comparisons
  - Use Tukey-Kramer simultaneous CIs or HTs if there is a large number of comparisons

Two-Way ANOVA

- One-way ANOVA only allow us to evaluate a single factor variable (with 3+ levels) with a response variable

- What if we have a **two** factor variables and a single response variable?

- For example, what if we wanted to examine how music affects the productivity of our employees? Specifically, we want to examine the **type of music** (rock, country, jazz) as well as the **loudness of the music** (soft or loud).

- To do this, we would need to run a **two-way ANOVA**

Terminology

- Refer to the two factors as the **row factor** (I levels) and **column factor** (J levels)
  - In the table below, Catalyst is the row factor and Reagent is the row factor

TABLE 9.2 Yields for runs of a chemical process with various combinations of reagent and catalyst

	Reagent		
Catalyst	**1**	**2**	**3**
A	86.8 82.4 86.7 83.5	93.4 85.2 94.8 83.1	77.9 89.6 89.9 83.7
B	71.9 72.1 80.0 77.4	74.5 87.1 71.9 84.1	87.5 82.7 78.3 90.1
C	65.5 72.4 76.6 66.7	66.7 77.1 76.7 86.1	72.7 77.8 83.5 78.8
D	63.9 70.4 77.2 81.2	73.7 81.6 84.2 84.9	79.8 75.7 80.5 72.9

- Design is **complete** if every treatment combination is used
- Design is **balanced** if each treatment combination has the same number of replicates (K)

We'll restrict our discussion to this setting

Main vs. Interaction Effects

- Two-way ANOVA will provide us with not only the main effects (of the row and column factors) but also the **interaction effects** between the factors
  - Interaction effects are commonly present, so we need to check for them

- Here are some examples:
  - Diet Plan A resulted in a greater weight loss for women than for men, but Diet Plan B resulted in a greater weight loss for men than for women.
  - Fertilizer A was best in high sunlight areas but Fertilizer B was best in low light areas.

Two-Way ANOVA Assumptions

Assumptions which need to be checked:

1. Design is complete
2. Design is balanced
3. Number of replicates per treatment is at least 2
4. Within any treatment, observations are a simple random sample from a normal population
5. Population variance σ^2 is the same for each treatment group

Hypotheses

In a Two Way ANOVA, we have three sets of null hypotheses:

H_{0a}: The population means of the **row factor** are equal

H_{0b}: The population means of the **column factor** are equal

H_{0c}: There is no **interaction effect** between the row and column factors

Each has its own test statistic

Steps to Perform Test

1. Test whether all **interactions** are equal to zero (H_{0c})
2. Test whether the **row effects** are equal to zero (H_{0a})
3. Test whether the **column effects** are equal to zero (H_{0b})

Why in this order?

Cannot interpret the test of main effects
when interactions are present

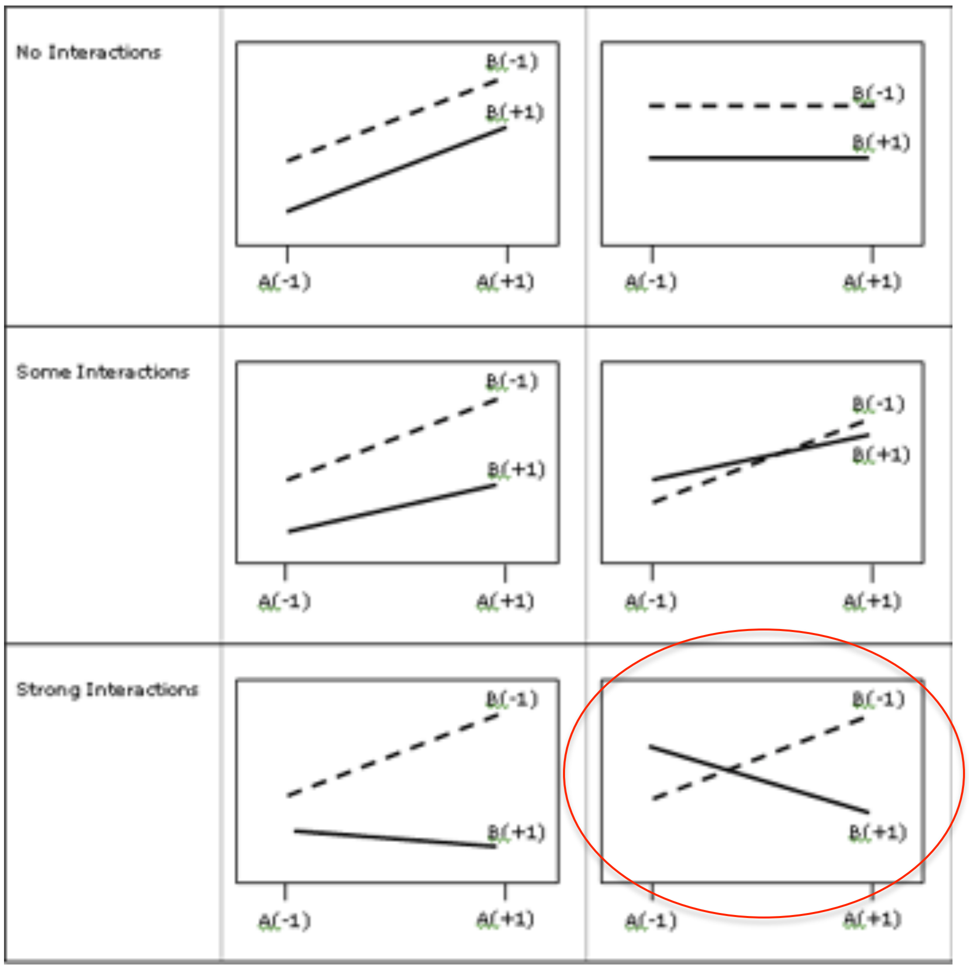
The effect of the row factor depends on
the column factor (and vise versa)

Possible Outcomes

Possible outcomes for a two-way ANOVA:

1. No significant effects
2. No significant interaction, one significant main effect
3. No significant interaction, two significant main effects
4. Significant interaction

Interaction Plot



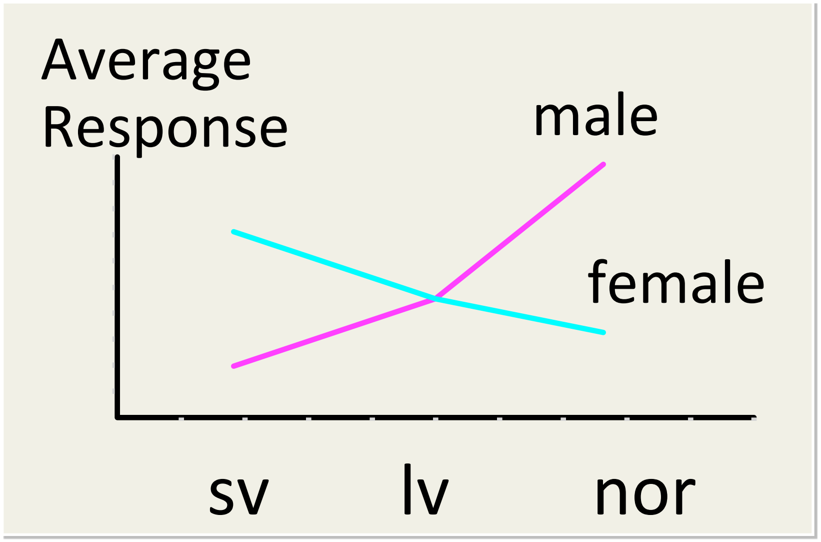
Used to qualitatively interpret main effects when interactions are present

e.g. A has a decreasing effect when B is 1 and an increasing effect when B is -1

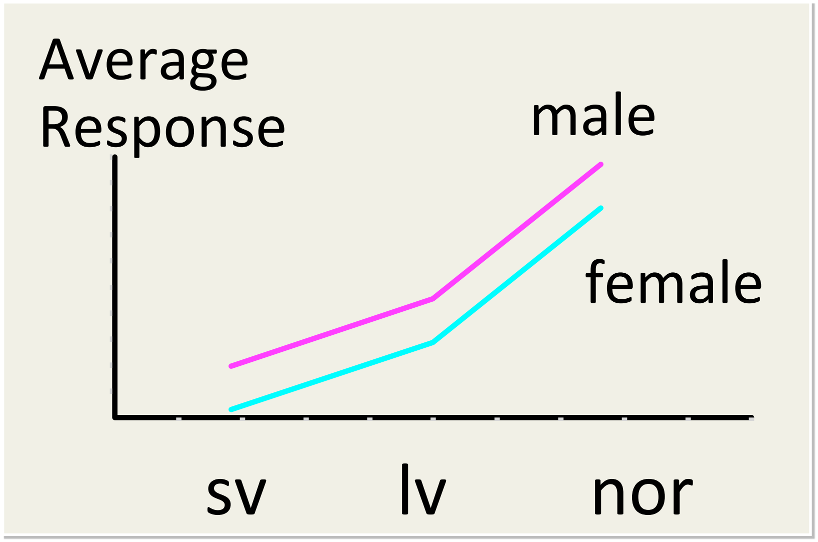
More Interaction Plots

Effects of Gender (male or female) and dietary group (sv: strict vegans, lv: lactovegetarians, nor: normal) on systolic blood pressure

Interaction



No Interaction



Sums of Squares

TABLE 9.5 ANOVA table for two-way ANOVA

Source	Degrees of Freedom	Sum of Squares
Rows (SSA)	$I-1$	$JK\sum_{i=1}^{I}\widehat{\alpha}_i^2 = JK\sum_{i=1}^{I}\overline{X}_{i..}^2 - IJK\overline{X}_{...}^2$
Columns (SSB)	$J-1$	$IK\sum_{j=1}^{J}\widehat{\beta}_j^2 = IK\sum_{j=1}^{J}\overline{X}_{.j.}^2 - IJK\overline{X}_{...}^2$
Interactions (SSAB)	$(I-1)(J-1)$	$K\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\gamma}_{ij}^2 = K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}^2 - JK\sum_{i=1}^{I}\overline{X}_{i..}^2$
		$-IK\sum_{j=1}^{J}\overline{X}_{.j.}^2 + IJK\overline{X}_{...}^2$
Error (SSE)	$IJ(K-1)$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk}-\overline{X}_{ij.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}^2$
Total (SST)	$IJK-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk}-\overline{X}_{...})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - IJK\overline{X}_{...}^2$

Test Statistics

1. Test whether all **interactions** are equal to zero (H_{0c}):

$$F_{AB} = MSAB/MSE \sim F_{(I-1)(J-1),\ IJ(K-1)}$$

2. Test whether the **row effects** are equal to zero (H_{0a}):

$$F_A = MSA/MSE \sim F_{J-1,\ IJ(K-1)}$$

3. Test whether the **column effects** are equal to zero (H_{0b}):

$$F_B = MSB/MSE \sim F_{I-1,\ IJ(K-1)}$$

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F statistic	P-value
A	I-1	SSA	MSA = SSA/(I-1)	F_A = MSA/MSE	$P(F_{I-1,IJ(K-1)} > F_A)$
B	J-1	SSB	MSB = SSB/(J-1)	F_B = MSB/MSE	$P(F_{J-1,IJ(K-1)} > F_B)$
AB	(I-1)(J-1)	SSAB	MSAB = $\frac{SSAB}{[(I-1)(J-1)]}$	F_{AB} = MSAB/MSE	$P(F_{(I-1)(J-1), IJ(K-1)} > F_{AB})$
Error	IJ(K-1)	SSE	MSE = SSE/[IJ(K-1)]		
Total	IJK-1 = N-1	SST			

Example

Source	SS	df	MS	F	*p* - value
A	5.0139	1	5.0139	100.28	0
B	2.1811	2	1.0906	21.81	.0001
AB	0.1344	2	0.0672	1.34	.298
Error	0.6000	12	0.0500		
Total (Corr)	7.9294	17			

Example – Reaction Yield

- Response: yield of desired product
- Factors: Catalyst (A,B,C,D) and Reagent (1,2,3)

Two-way ANOVA: Yield versus Catalyst, Reagent

Source	DF	SS	MS	F	P
Catalyst	3	877.56	292.521	9.36	0.000
Reagent	2	327.14	163.570	5.23	0.010
Interaction	6	156.98	26.164	0.84	0.550
Error	36	1125.33	31.259		
Total	47	2487.02			

Interpreting Two-Way ANOVAs

- Be able to interpret the results of an ANOVA given the ANOVA table
  - Which factor(s) have an effect on the response
  - Whether an interaction exists and if so how to interpret it
- Be able to fill in missing values from an ANOVA table
- Be able to interpret interaction plots

Example – Pesticide Absorption

- Response: Pesticide absorption level
- Factors: Concentration (A,B,C) and Duration (1,2,3)
- Fill in the missing quantities:

Two-way ANOVA: Absorbed versus Concentration, Duration

Source	DF	SS	MS	F	P
Concent	2	49.991		107.99	0.000
Duration	2	19.157	9.579		0.000
Interaction	4	0.337	0.084	0.36	
Error	27	6.250	0.231		
Total	35	75.735			

Example – Pesticide Absorption

- Response: Pesticide absorption level
- Factors: Concentration (A,B,C) and Duration (1,2,3)
- Fill in the missing quantities:

Two-way ANOVA: Absorbed versus Concentration, Duration

Source	DF	SS	MS	F	P
Concent	2	49.991	24.996	107.99	0.000
Duration	2	19.157	9.579	41.38	0.000
Interaction	4	0.337	0.084	0.36	0.832
Error	27	6.250	0.231		
Total	35	75.735			

Example – Silicone Thickness

- Response: Thickness of silicone dioxide layer on a semiconductor
- Factors: Furnace location (1,2,3) and Wafer Type (A,B,C)
- Fill in the missing quantities:

Two-way ANOVA for Thickness versus Wafer, Location

Source	DF	SS	MS	F	P
Wafer	2	5.8756		2.07	0.155
Location	2	4.1089	2.0544		0.262
Interaction		21.349	5.3372	3.76	0.022
Error	18		1.4207		
Total	26	56.907			

Example – Silicone Thickness

- Response: Thickness of silicone dioxide layer on a semiconductor
- Factors: Furnace location (1,2,3) and Wafer Type (A,B,C)
- Fill in the missing quantities:

Two-way ANOVA for Thickness versus Wafer, Location

Source	DF	SS	MS	F	P
Wafer	2	5.8756	2.9378	2.07	0.155
Location	2	4.1089	2.0544	1.45	0.262
Interaction	4	21.349	5.3372	3.76	0.022
Error	18	25.573	1.4207		
Total	26	56.907			

Next

- Last Homework due Friday – Solutions posted to Learn@UW immediately after they are turned in
- Review class Friday; Practice finals posted on Learn@UW (solutions posted tomorrow)
- Extra office hours – Friday 11am-1pm
- Final Exam (cumulative) Sunday, May 11 from 2:45-4:45 in B130 Van Vleck
  - Calculator
  - 2 pages hand-written notes (front and back)