

# Pairwise Comparisons in ANOVA

Keegan Korthauer

Department of Statistics

University of Wisconsin

# Recap: One-Way ANOVA

Question of Interest: Are the treatment means different?

# Terminology

- **Response variable:** continuous measurement, also called the outcome variable or dependent variable
- **Factor:** categorical variable that can take on several different values, often called **levels** or **treatments** of the factor
- **Experimental units:** objects upon which measurements are made
- **Replicates:** the experimental units assigned to a given treatment

# Sums of Squares

Quantities we need to compute the test statistic:

**Treatment sum of squares (SSTr)** 
$$SSTr = \sum_{i=1}^I J_i \bar{X}_i^2 - N\bar{X}_{..}^2$$

Indicates how different the treatment means are from each other

**Error sum of squares (SSE)** 
$$SSE = \sum_{i=1}^I \sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^I J_i \bar{X}_i^2$$

Measures the variation around the treatment means

**Total sum of squares (SST)** 
$$SST = \sum_{i=1}^I \sum_{j=1}^{J_i} X_{ij}^2 - N\bar{X}_{..}^2$$

Measures the total variation in the response

# ANOVA Test Statistic

The following F statistic measures evidence against the null hypothesis that all treatment means are equal:

$$F = \frac{SSTr / (I - 1)}{SSE / (N - I)} = \frac{MSTr}{MSE}$$

Get the p-value from the F table with I-1 and N-I degrees of freedom

# ANOVA Table

Often the sums of squares are presented in a table like this:

<b>Source</b>	<b>d.f.</b>	<b>SS</b>	<b>MS</b>	<b>F</b>	<b>P</b>
<b>Treatment</b>	$I - 1$	$SSTr$	$MSTr = SSTr / (I - 1)$	$MSTr / MSE$	
<b>Error</b>	$n - I$	$SSE$	$MSE = SSE / (n - I)$		
<b>Total</b>	$n - 1$	$SST$			

Note that  $SST = SSE + SSTr$

# CI for Treatment Mean

- We can formulate a  $100(1-\alpha)\%$  confidence interval for an individual treatment mean:

**Point estimate  $\pm$  critical value \* SD of point estimate**

- Point estimate for mean of treatment  $i$ :  $\bar{X}_i$
- Reasonable guess for the standard deviation of the point estimate:  $s_i / \sqrt{J_i}$

BUT...

Assumption 2 of ANOVA says that all treatments have equal variance, so we can use the pooled average of all sample standard deviations which turns out to be:

$$\sqrt{MSE / J_i}$$

# CI for Treatment Mean Cont'd

- All that's left is the critical value
- Assumption 1 - Normal populations:
  - Use Z/t depending on sample size
  - For simplicity we'll always use t (this is conservative for large  $J_i$ )
- Final form of the CI for mean of treatment i:

$$\bar{X}_i \pm t_{N-I, \alpha/2} \sqrt{MSE / J_i}$$



# Example – Compressive Strength of Concrete

Find a 95% confidence interval for the mean compressive strength of concrete with curing temperature of 20 degrees Celsius.

T (°C)	Strengths				
0	31.2	29.6	30.8	30.0	31.4
10	30	27.7	31.1	31.3	30.6
20	35.9	36.8	35.0	34.6	36.5
30	38.3	37.0	37.5	36.1	38.4

# Now What?

- If we can reject  $H_0$  then we can conclude that at least two treatment means are different from each other
- We can construct CIs for each individual treatment mean
- How do we tell exactly which treatments are different from the others?
- We need to perform pairwise comparisons

# Pairwise Comparisons

- The F test does not tell us *which* treatments are different from the rest.
- Sometimes an experimenter has in mind two specific treatments,  $i$  and  $j$ , and wants to study the difference:

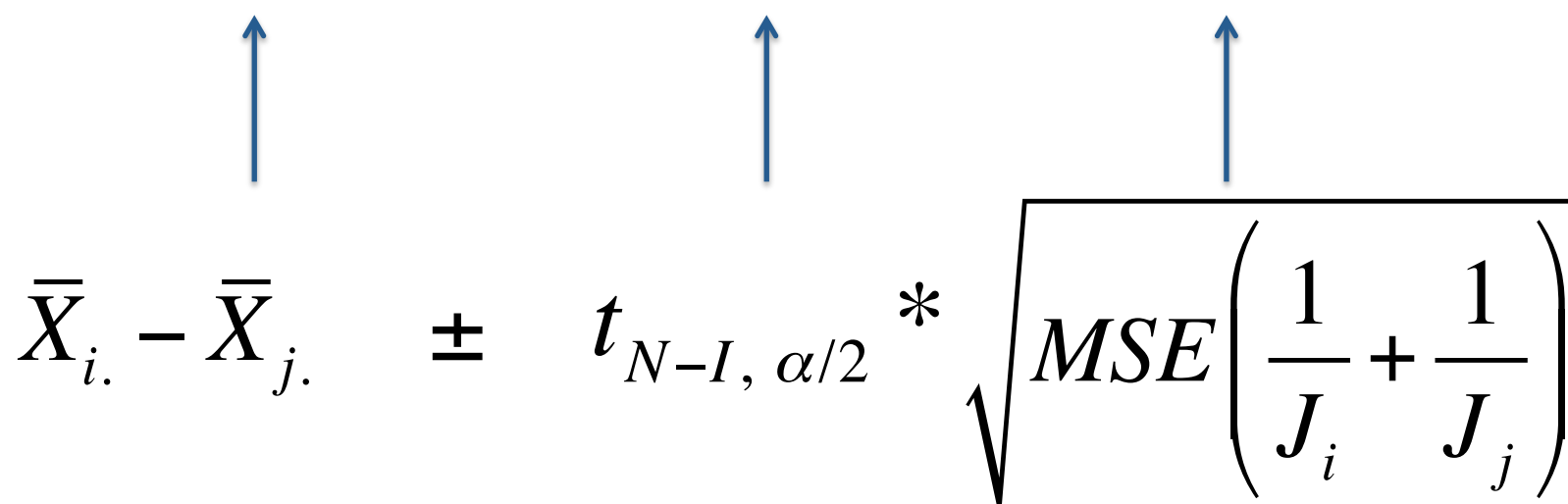
$$\mu_i - \mu_j$$

- Simplest possible solution: CI or HT for difference in means using the method of **Fisher's Least Significant Difference**
  - Appropriate if we are only going to examine ONE pre-specified difference

# Fisher's Least Significant Difference (LSD)

100(1- $\alpha$ )% CI for difference in means of treatment i and j:

**Point estimate  $\pm$  critical value \* SD of point estimate**


$$\bar{X}_{i.} - \bar{X}_{j.} \pm t_{N-1, \alpha/2} * \sqrt{MSE \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$$

# Example – Compressive Strength of Concrete

Find a 95% confidence interval for the mean difference in compressive strength for concrete cured at 20 degrees versus 30 degrees.

T (°C)	Strengths				
0	31.2	29.6	30.8	30.0	31.4
10	30	27.7	31.1	31.3	30.6
20	35.9	36.8	35.0	34.6	36.5
30	38.3	37.0	37.5	36.1	38.4

# Fisher's Least Significant Difference (LSD)

HT for difference in means of treatment  $i$  and  $j$ :

$$H_0 : \mu_i - \mu_j = 0 \text{ vs. } H_1 : \mu_i - \mu_j \neq 0$$

**Test Statistic = (Point Estimate – Hypothesized Value)  
SD of Point Estimate**

$$\frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSE \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}} \sim t_{N-I}$$

“LSD”: smallest difference in means that will be significant

=> Reject  $H_0$  at the  $\alpha$  level if  $|\bar{X}_i - \bar{X}_j| > t_{N-I, \alpha/2} \sqrt{MSE \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$

# Simultaneous CIs and HTs

- Recall from Section 6.14 (Multiple Testing) that our confidence *decreases as the number of tests performed increases*
  - Cannot simply test each pairwise treatment combination individually with Fisher's LSD at desired level
  - We need to make adjustments for **multiple comparisons**
- We can do this with **simultaneous** CIs/HTs:
  - We are confident at the  $100(1-\alpha)\%$  level that *every* CI contains the true difference
  - We may reject, at level  $\alpha$ , *every*  $H_0$  whose p-value  $< \alpha$
- Two possible methods: Bonferroni adjustment and Tukey-Kramer method

# Bonferroni Simultaneous CI/HT

- Adjust the significance level by the number of comparisons we will make
- In an experiment with I treatments, there are  $C=I(I-1)/2$  possible pairwise comparisons
- *For all pairs of treatments i and j*, the simultaneous CI/HT uses the modified t critical value:

$$(\bar{X}_i - \bar{X}_j) \pm t_{N-I, \alpha/(2C)} \sqrt{MSE \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$$

$$|\bar{X}_i - \bar{X}_j| > t_{N-I, \alpha/(2C)} \sqrt{MSE \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$$



# Tukey-Kramer Method

- An alternative method to construct simultaneous CIs/HTs
- Also called “Honestly Significant Difference” (in contrast to Fisher’s “Least Significant Difference”)
- Based on a special distribution called the **Studentized range distribution** (different than student’s t)
  - The Studentized range distribution has two degrees of freedom parameters: **I and N-I** for the Tukey-Kramer method (Table A.9)
  - The critical value is denoted  $q_{I, N-I, \alpha}$

Not I-1 and N-I as in the ANOVA F test

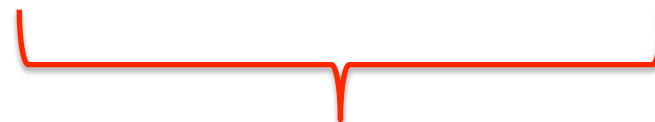
# TK Method Cont'd

- Simultaneous CIs for differences in treatment means (for any pair  $i$  and  $j$ ):

$$(\bar{X}_i - \bar{X}_j) \pm q_{I, N-I, \alpha} \sqrt{\frac{MSE}{2} \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$$

- To test all null hypotheses of the form  $H_0: \mu_i - \mu_j = 0$  simultaneously, reject the null hypothesis when

$$|\bar{X}_i - \bar{X}_j| > q_{I, N-I, \alpha} \sqrt{\frac{MSE}{2} \left( \frac{1}{J_i} + \frac{1}{J_j} \right)}$$



Honestly Significant Difference

# Example – Compressive Strength of Concrete

Conduct **simultaneous** hypothesis tests using the Bonferroni and Tukey-Kramer methods at the 0.05 level for the pairwise null hypotheses that the difference between treatment means is equal to zero.

T (°C)	Strengths				
0	31.2	29.6	30.8	30.0	31.4
10	30	27.7	31.1	31.3	30.6
20	35.9	36.8	35.0	34.6	36.5
30	38.3	37.0	37.5	36.1	38.4

# Bonferroni vs. Tukey-Kramer

- Both make adjustments for multiple comparisons
- When only a few comparisons will be made, Bonferroni can be more powerful
- If the number of comparisons is large, the Tukey-Kramer is usually superior
  - Takes advantage of the assumption of normal populations

# Next

- Intro to two-factor ANOVAs
- Last Homework due Friday – Solutions posted to Learn@UW immediately after they are turned in
- Review class Friday; Practice final posted tomorrow
- Final Exam (cumulative) Sunday, May 11 from 2:45-4:45 in B130 Van Vleck