Intro to Factorial Experiments

Keegan Korthauer
Department of Statistics
UW Madison

Factorial Experiments

- Used to compare different **factors** or **treatments**
- For example – compare the mean crop yield for 4 different fertilizer types
- If there were just two factors, could apply methods in chapter 5 and 6 for comparing means in two different populations
- How do we compare across 3 or more groups?

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

Example – Compressive Strength of Concrete

- A study is conducted to investigate the **effect of curing temperature on the compressive strength** of a certain type of concrete.
- Five concrete specimens are cured at each of four temperatures, and compressive strength of each specimen is measured (in MPa):

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

Example – Compressive Strength of Concrete

- **Response variable:** compressive strength (in MPa)
- **Factor:** curing temperature
- **Levels:** all possible values of the factor (0, 10, 20, 30)
- **Experimental units:** each concrete specimen (20 total)
- **Replicates:** five specimens per temperature setting

Example – Removal Rate

- The removal rate of ammoniacal nitrogen (a toxic pollutant) is an important aspect in the treatment landfill runoff
- The rate of removal (in percent per day) is recorded for several days for each of several treatment methods:

Treatment	Rate of Removal			
A	5.21	4.65		
B	5.59	2.69	7.57	5.16
C	6.24	5.94	6.41	
D	6.85	9.18	4.94	
E	4.04	3.29	4.52	3.75

Example – Compressive Strength of Concrete

- **Response variable:** removal rate of ammoniacal nitrogen (percent per day)
- **Factor:** runoff treatment
- **Levels:** all possible treatments (A, B, C, D, E)
- **Experimental units:** each day measured (16 total)
- **Replicates:** number of days per treatment (varies 2 -4)

Completely Randomized Design

- All experimental units are assigned to a factor level (or treatment) **randomly**
  - Example: number each of the concrete specimens from 1 to 20, then generate a random ordering of the integers from 1 to 20
  - Units are **equally likely** to be assigned to any of the treatments

- Desirable property for testing difference in treatment means

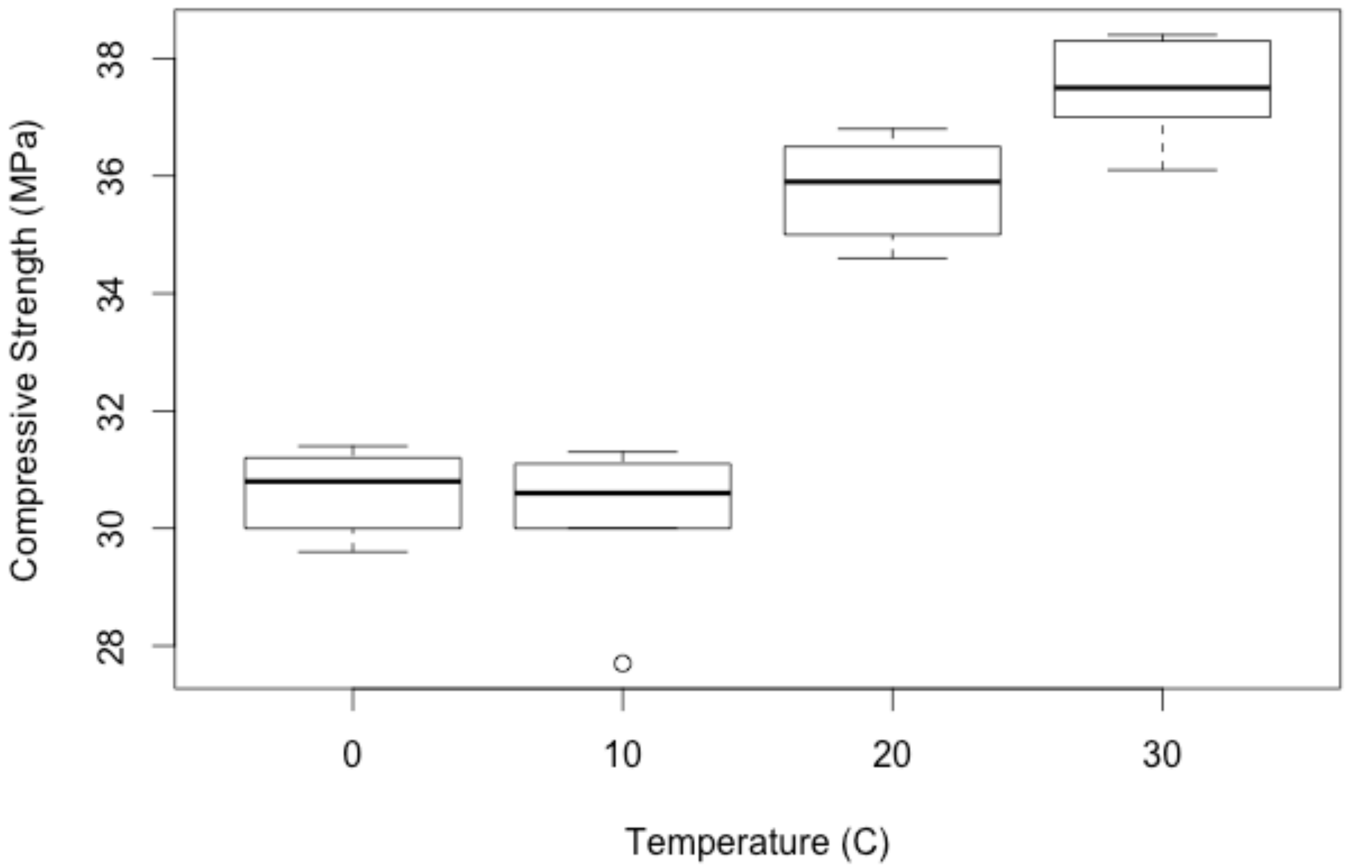
- Not a completely randomized design: if all the specimens cured at 0 degrees were run on Monday, the specimens run at 10 degrees were run on Tuesday, and so on…
  - Treatment effect may be confounded with day effect

Question of Interest

Are the treatment means different?

How to answer?

Example – Compressive Strength of Concrete



Formulating the Hypotheses

- Let the means for each of I treatments be $\mu_1, \mu_2, \ldots, \mu_I$
- Let the sample sizes in each treatment be $J_1, J_2, \ldots, J_I$
- The total sample size is $N = J_1 + J_2 + \ldots + J_I$
- Then we wish to test the hypotheses:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \text{ versus}$$

$$H_1 : \text{two or more of the } \mu_I \text{ are different}$$

One-Way Analysis of Variance (ANOVA)

- If there were only two treatments, we could use the two-sample t-test (section 6.7)

- Instead, we will use the **one-way analysis of variance**
  - "One-way" since there is only one factor

- The test statistic will involve computing sums of squares (similar to SLR/MLR)
  - The idea is to see if the treatment explains a substantial portion of the variation in the response variable

Some Notation

- Let X_{ij} denote the response measurement for the j^{th} observation in the i^{th} treatment

- Mean for the i^{th} treatment: $$\bar{X}_{i.} = \frac{\sum_{j=1}^{J_i} X_{ij}}{J_i}$$

- Sample grand mean:

$$\bar{X}_{..} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}}{N} = \frac{\sum_{i=1}^{I} J_i \bar{X}_{i.}}{N}$$

Sums of Squares

Quantities we'll need to compute the test statistic:

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

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

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

Sums of Squares

Quantities we'll need to compute the test statistic:

Treatment sum of squares (SSTr)

Indicates how different the treatment means are from each other

Error sum of squares (SSE)

Measures the variation around the treatment means

Total sum of squares (SST)

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:

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

Note that SST = SSE + SSTr

Assumptions for the ANOVA

For the test to be valid, we need to make the following assumptions:

1. The treatment populations must be normal
   - Check with QQ plots (each treatment separately or the entire sample)

2. The treatment populations must all have the same variance
   - Check the residual plot for differences in spread by treatment

Balanced Design

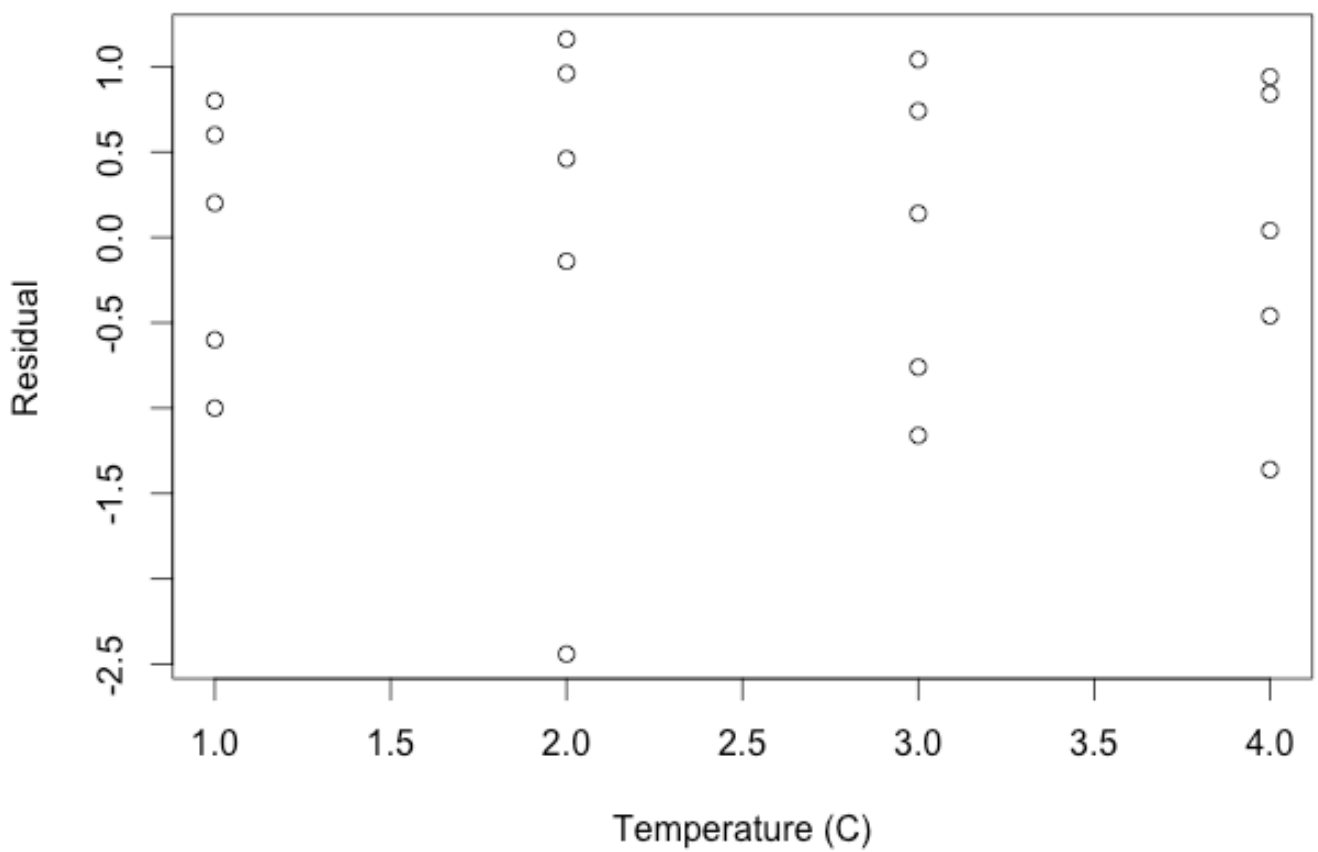
- The number of replicates is the same for each treatment
- Desirable property since the effect of unequal variances is not as severe for balanced designs
- The more unbalanced the design, the greater the effect of unequal variance

Example – Compressive Strength of Concrete

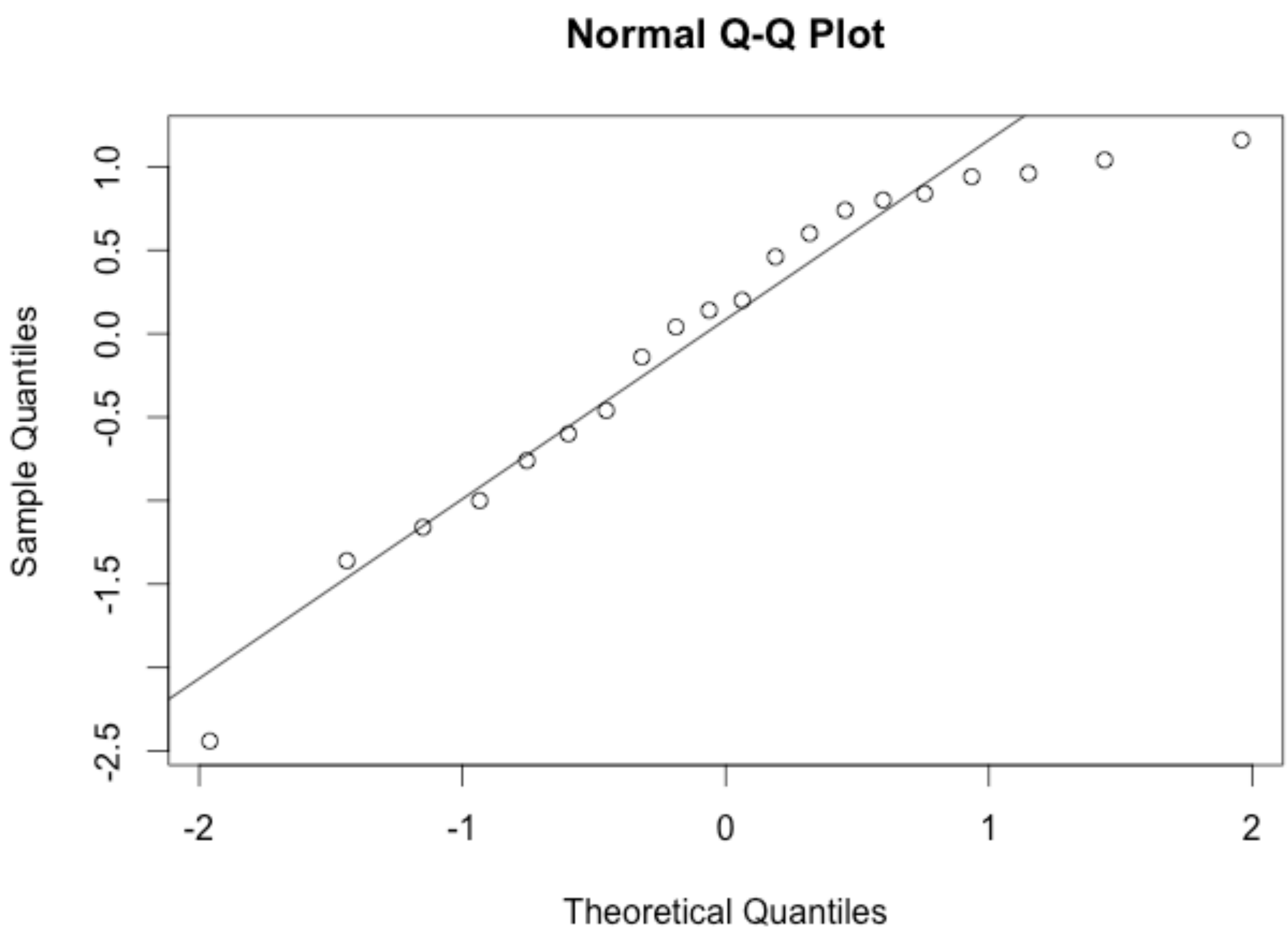
Check the assumptions for the ANOVA and if satisfactory, test the null hypothesis that the mean compressive strength is the same for curing temperatures 0, 10, 20 and 30 degrees Celsius

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

Residual Plot



QQ Plot



Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
1.0
0.5
0.0
-0.5
-1.5
-2.5
-2
-1
0
1
2

Now What?

- If we can reject H_0 then we can conclude that at least two treatment means are different from each other

- Exactly which ones are different from the rest?

- We'll answer that next time

Notes

- Note that this is just an intro to the subject of experimental design
  - Stat 424: Statistical Experimental Design for Engineers
- HW 12 (Last) posted later today
- Final Exam (cumulative) Sunday, May 11 from 2:45-4:45 in B130 Van Vleck