MATH1. Part II

Probability and Statistics

Chapter 11

Comparing the Means of Many Independent Samples

11.1 Introduction

Comparing the Means of <u>Two</u> Independent Samples (Review of Chapter 7)

- **Hypothesis testing for 2 independent samples: t test**
  - Null hypothesis H_0: $\mu_1 = \mu_2$
    - H_0: the hypothesis that μ_1 and μ_2 are <u>equal</u> (no difference)
  - Alternative hypothesis H_A: $\mu_1 \neq \mu_2$
    - H_A : the hypothesis that μ_1 and μ_2 are <u>NOT equal</u> (differ)
  - t test: $t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}}$ → P-value → P-value vs. α

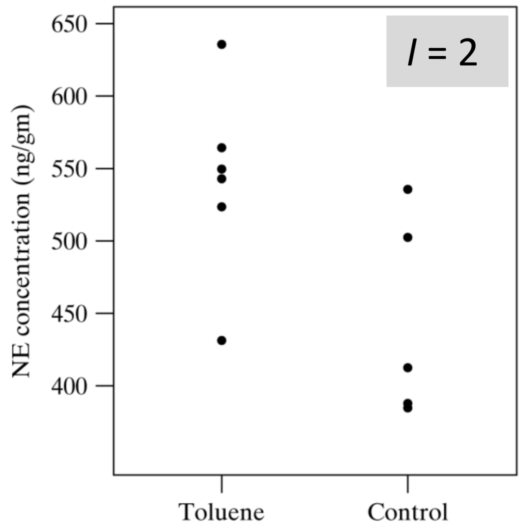

Figure 7.2.1 Parallel dotplots of NE concentration

What if we need to compare more than two independent samples?

11.1 Introduction

Comparing the Means of Many Independent Samples

- **Hypothesis testing for *I* independent samples:**
  - Global null hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$
    - H_0: all the population means are equal (no difference)
  - Alternative hypothesis H_A: $\mu_m \neq \mu_n$
    - H_A: at least one pair of the population means are NOT equal (differ)

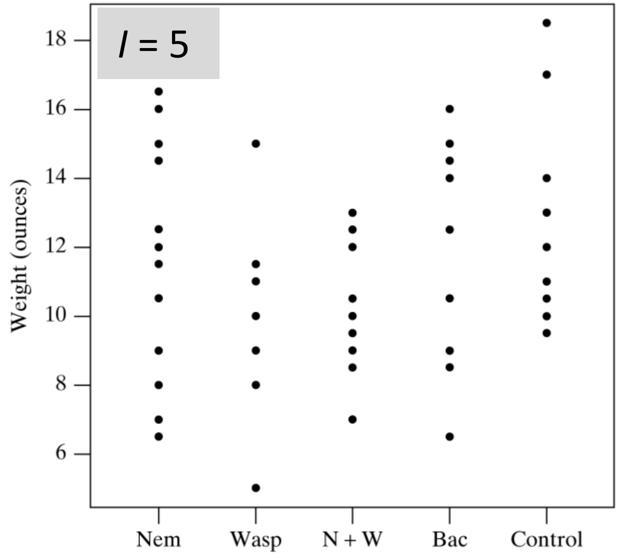

Figure 11.1.1 Weights of ears of corn receiving five different treatments

What statistical test should I use?

11.1 Introduction

A Graphical perspective on ANOVA

- In order to find compelling evidence for a difference in population means (H_A),
  - (1) not only must there be variation among the group means,
  - (2) but variation among the group means must be large relative to the inherent variability in the groups

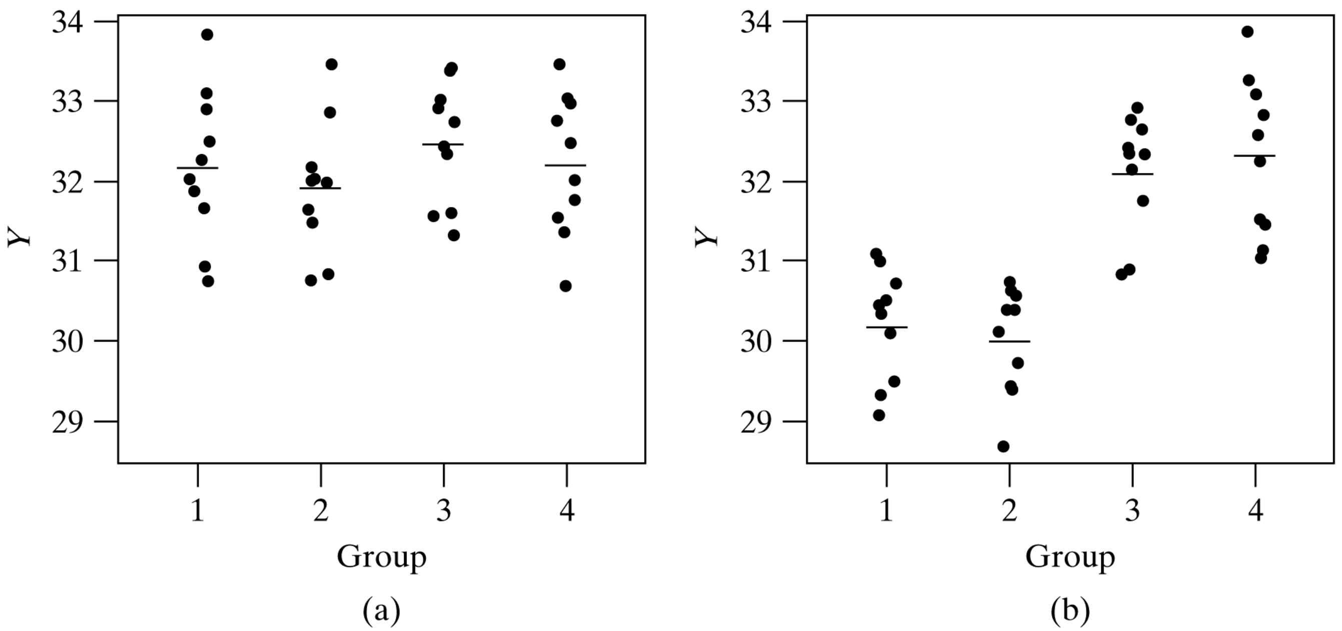

Figure 11.1.3

(a) H_0 true, (b) H_0 false, with small SDs for the groups

11.1 Introduction

A Graphical perspective on ANOVA

- “analysis of variance” → make an inference about means.
- If the between-group mean variability is large relative to within-group variability, we will take this as evidence against the null hypothesis (the population means are all equal).

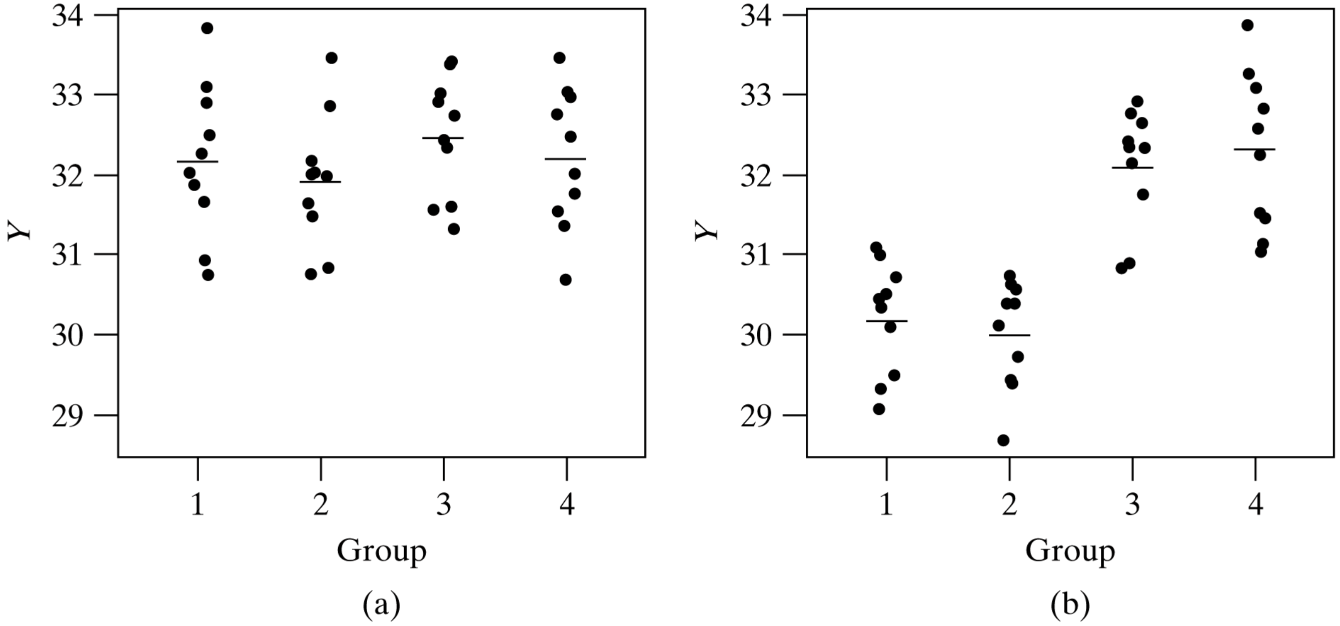

Figure 11.1.3
(a) H_0 true, (b) H_0 false, with small SDs for the groups

11.2 The Basic One-Way Analysis of Variance

One-way ANOVA

- The classical method of analyzing data from three or more (I>2) independent samples is called an ANalysis Of VAriance, or ANOVA.
- In applying analysis of variance (ANOVA), the data are regarded as random samples from I populations.
- The term “one-way” refers to the fact that there is one variable that defines the groups or treatments
  - e.g. in the sweet corn example the treatments were based on the type of harmful insect/bacteria.
    - Treatment 1: Nematodes
    - Treatment 2: Wasps
    - Treatment 3: Nematodes and wasps
    - Treatment 4: Bacteria
    - Treatment 5: Control

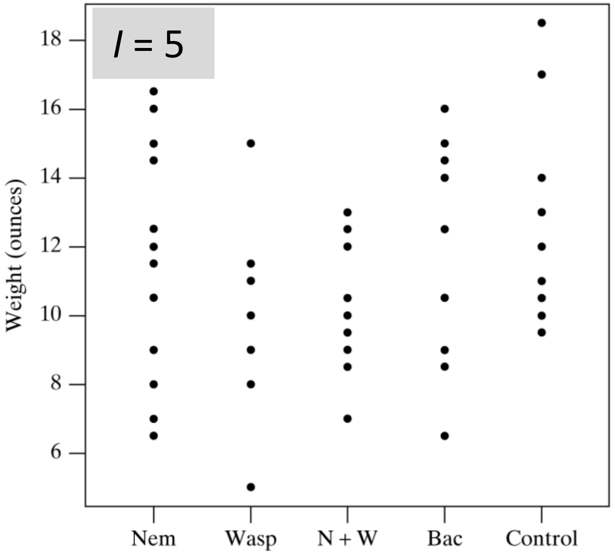

Figure 11.1.1 Weights of ears of corn receiving five different treatments

11.2 The Basic One-Way Analysis of Variance

One-way ANOVA

- **Notation**
  - Population means: μ_1 , μ_2 , … , μ_I
  - Population standard deviations: σ_1 , σ_2 , … , σ_I
  - To describe several groups of quantitative observations,
    - y_{ij} = observation j in group i
    - I = number of groups
    - n_i = number of observations in group i
    - $\bar{y}_i$ = mean for group i
    - s_i = standard deviation for group i

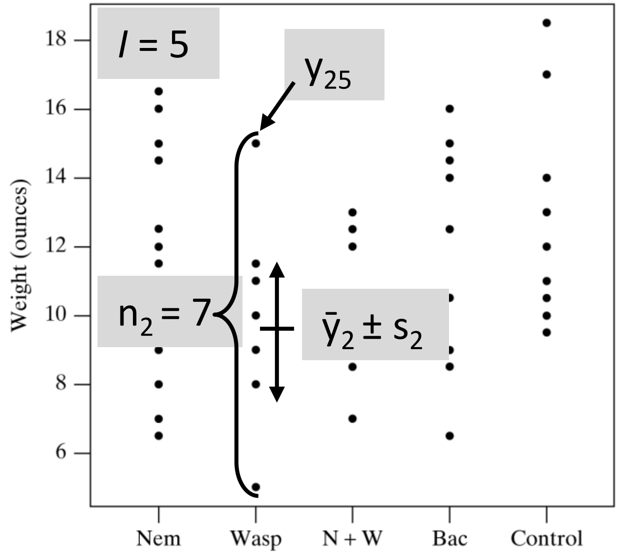

Figure 11.1.1 Weights of ears of corn receiving five different treatments

11.2 The Basic One-Way Analysis of Variance

One-way ANOVA

- **Notation**

  To describe several groups of quantitative observations,

  – The **total number of observations**

  $$n. = \sum_{i=1}^{I} n_i$$

  – The **grand mean** (the mean of all the observations)

  $$\bar{\bar{y}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n.}$$

  - The grand mean can be expressed as a weighted average of the group means

  $$\bar{\bar{y}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n.} = \frac{\sum_{i=1}^{I} n_i\,\bar{y}_i}{n.}$$

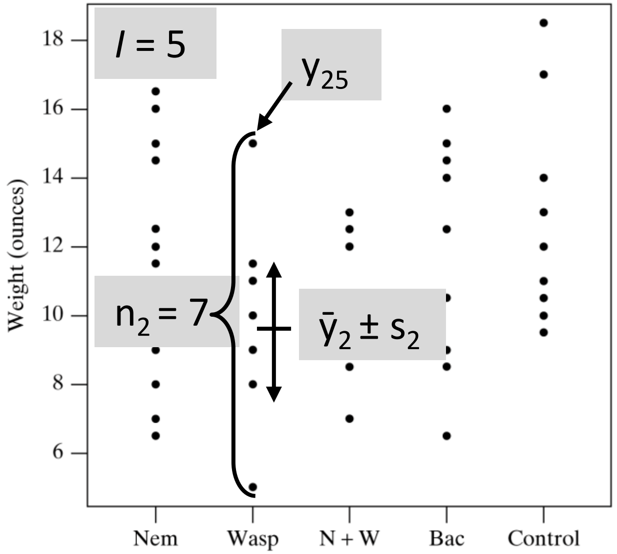

Figure 11.1.1 Weights of ears of corn receiving five different treatments

11.2 The Basic One-Way Analysis of Variance

One-way ANOVA

- **Notation**

Example 11.2.1 Weight Gain of Lambs

- What is the total number of observations?
- What is the grand mean?

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

One-way ANOVA

- **Notation**

Example 11.2.1 Weight Gain of Lambs

- What is the total number of observations?
- What is the grand mean?
  - The total number of observations is $n_{\cdot} = \sum_{i=1}^{I} n_i$ = 3+5+4 =12
  - The grand mean is
    $$\bar{\bar{y}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n_{\cdot}} = \frac{33+75+48}{12} = 13$$

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Within Groups

- **Pooled standard deviation**: a combined measure of variation within the I groups is the pooled standard deviation s_{pooled}, often simply denoted as just s, which is computed as follows.

Pooled Standard Deviation

$$s_{\text{pooled}} = s = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{\sum_{i=1}^{I}(n_i - 1)}} = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{n_{\bullet} - I}}$$

- **Pooled variance**: we call $s^2_{pooled} = s^2$ the pooled variance

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Within Groups

Example 11.2.1 Weight Gain of Lambs

- What is the pooled variance?
- What is the pooled standard deviation?

Pooled Standard Deviation

$$s_{\text{pooled}} = s = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{\sum_{i=1}^{I}(n_i - 1)}} = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{n_{\bullet} - I}}$$

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Within Groups

Example 11.2.1 Weight Gain of Lambs

- What is the pooled variance?
- What is the pooled standard deviation?
  - pooled variance $s^2_{pooled} = s^2$

$$s^2 = \frac{(3-1)4.3592 + (5-1)4.9502 + (4-1)4.9672}{3+5+4-3}$$

$$= 23.336$$

  - pooled standard deviation s_{pooled}

$$s = \sqrt{23.336} = 4.831$$

** This estimate depends only on the variability within the groups and not on their mean values.*

Pooled Standard Deviation

$$s_{\text{pooled}} = s = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{\sum_{i=1}^{I}(n_i - 1)}} = \sqrt{\frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{n_{\bullet} - I}}$$

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Within Groups - ANOVA notation

- **Pooled variance**:

$$s^2_{pooled} = s^2 = \frac{\sum_{i=1}^{I}(n_i - 1)s_i^{\,2}}{\sum_{i=1}^{I}(ni - 1)} = \frac{\sum_{i=1}^{I}(n_i - 1)s_i^{\,2}}{n. - I}$$

  - **SS (within)**: the numerator of s^2_{pooled} is known as the sum of squares within groups
  - **df (within)**: the denominator of s^2_{pooled} is known as the degrees of freedom within groups
  - **MS (within)**: their ratio is defined as the mean square within groups
    - Note that MS(within) is just another name for the pooled variance.

Mean Square Within Groups

$$\text{MS(within)} = \frac{\text{SS(within)}}{\text{df(within)}}$$

Sum of Squares and df Within Groups

$$\text{SS(within)} = \sum_{i=1}^{I}(n_i - 1)s_i^2$$

$$\text{df(within)} = n. - I$$

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Between Groups

- **MS(between)**
  - the mean square between groups, or MS(between), describes between-group variability for more than two groups.
  - In fact, the MS(between) would indeed be the sample variance of the group means.

Mean Square Between Groups

$$\text{MS(between)} = \frac{\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2}{I - 1} = \frac{\text{SS(between)}}{\text{df(between)}}$$

Sum of Squares and df Between Groups

$$\text{SS(between)} = \sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$$

$$\text{df(between)} = I - 1$$

11.2 The Basic One-Way Analysis of Variance

Measuring Variation <u>Between</u> Groups

Example 11.2.1 Weight Gain of Lambs (continued)

- What is the variation within groups?
- What is the variation between groups?

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Between Groups

Example 11.2.1 Weight Gain of Lambs (continued)

- What is the variation within groups?
- What is the variation between groups?
  - variation within groups

$$s^2 = \frac{(3-1)4.359^2 + (5-1)4.950^2 + (4-1)4.967^2}{3+5+4-3}$$

$$= \frac{210.025}{9} = 23.336$$

  - Same as pooled variance $s^2_{pooled} = s^2$
  - SS(within) = 210.025, df(within) = 9, and
  - MS(within) = 23.336

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

Mean Square Within Groups

$$\text{MS(within)} = \frac{\sum_{i=1}^{I}(n_i - 1)s_i^2}{n. - I} = \frac{\text{SS(within)}}{\text{df(within)}}$$

11.2 The Basic One-Way Analysis of Variance

Measuring Variation Between Groups

Example 11.2.1 Weight Gain of Lambs (continued)

- What is the variation within groups?
- What is the variation between groups?
  - variation between groups
    - $\bar{\bar{y}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n.} = \frac{33+75+48}{12} = 13$
    - SS(between) = 3 x (11 - 13)^2 + 5 x (15 - 13)^2 + 4 x (12 - 13)^2 =36
    - Since I = 3, we have df(between) = 3 - 1 = 2
    - MS(between) = 36/2 = 18

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

Mean Square Between Groups

$$\text{MS(between)} = \frac{\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2}{I-1} = \frac{\text{SS(between)}}{\text{df(between)}}$$

11.2 The Basic One-Way Analysis of Variance

A Fundamental Relationship of ANOVA Between Groups

- Consider an individual observation y_{ij}, the deviation of an observation from the grand mean is

$$y_{ij} - \bar{\bar{y}} = \underbrace{(y_{ij} - \bar{y}_i)}_{\text{1. a within-group deviation}} + \underbrace{(\bar{y}_i - \bar{\bar{y}})}_{\text{2. a between-group deviation}}$$

- It is also true that the analogous relationship holds for the corresponding sums of squares; that is,

$$\sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{\bar{y}})^2$$

$$= \sum_{i=1}^{I} (n_i - 1)\, si^2 + \sum_{i=1}^{I} n_i\, (\bar{y}_i - \bar{\bar{y}})^2$$

= SS(within) + SS(between)

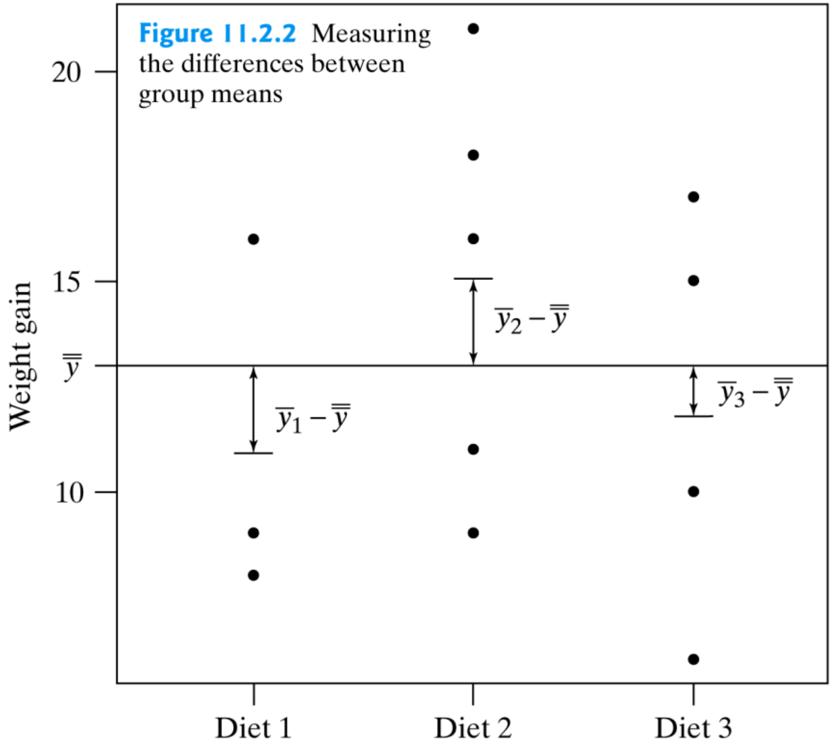

Figure 11.2.2 Measuring the differences between group means

- The quantity on the left-hand side of formula is called the total sum of squares, or **SS(total)**.

11.2 The Basic One-Way Analysis of Variance

A Fundamental Relationship of ANOVA Between Groups

- The total sum of squares, or SS(total) is defined as

Definition of Total Sum of Squares

$$\text{SS(total)} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$$

Relationship between Sums of Squares

$$\text{SS(total)} = \text{SS(between)} + \text{SS(within)}$$

- The total degrees of freedom, or df(total), is defined as follows:

Total df

$$\text{df(total)} = n_{\bullet} - 1$$

Relationship between df

$$\text{df(total)} = \text{df(between)} + \text{df(within)}$$

11.2 The Basic One-Way Analysis of Variance

The ANOVA table

- When working with the ANOVA quantities, it is customary to arrange them in a table.

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between groups	$I - 1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Within groups	$n_{\bullet} - I$	$\sum_{i=1}^{I} (n_i - 1)s_i^2$	SS/df
Total	$n_{\bullet} - 1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

- Comment on terminology
  - While the terms “between-groups” and “within-groups” are not technical terms, they are useful in describing and understanding the ANOVA model.
  - Computer software and other texts commonly refer to these sources of variability as treatment (between groups) and error (within groups).

11.2 The Basic One-Way Analysis of Variance

The ANOVA table

Example 11.2.1 Weight Gain of Lambs (continued)

- Construct the ANOVA table.

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

11.2 The Basic One-Way Analysis of Variance

The ANOVA table

Example 11.2.1 Weight Gain of Lambs (continued)

- Construct the ANOVA table.
  - SS(within) = 210.025, df(within) = 9
  - MS(within) = 23.336
  - SS(between) =36, I=3, df(between) = 3 - 1 = 2
  - MS(between) = 18

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

Table 11.2.3 ANOVA table for lamb weight gains

Source	df	SS	MS
Between diets	2	36	18.00
Within diets	9	210	23.33
Total	11	246	

11.3 The Analysis of Variance Model

- It can be helpful to think of ANOVA in terms of the following model.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

  - y_{ij}: jth observation in group i.
  - μ: represents the grand population mean
  - τ_i: represents the effect of group i
  - ε_{ij}: represents random error associated with observation j in group i
- The null hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$ is equivalent to H_0: $\tau_1 = \tau_2 = \ldots = \tau_I$
- Estimate the overall average, μ, with the grand mean of the **data** $\hat{\mu} = \bar{\bar{y}}$
  - estimate the population average for group i $\quad \hat{\mu}_i = \bar{y}_i$; $\hat{\mu} = \bar{\bar{y}}$
  - estimate the group effect $\quad \hat{\tau}_i = \bar{y}_i - \bar{\bar{y}}$; SS(between) = $\sum_{i=1}^{I} n_i \hat{\tau}_i^{\,2}$
  - estimate the random error $\quad \hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_i$; SS(within) = $\sum_{i=1}^{I} \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^{\,2}$
- We have $y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i) = \hat{\mu} + \hat{\tau}_i + \hat{\varepsilon}_{ij}$

11.4 The Global F Test

Hypothesis

- The global null hypothesis is H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$
  - H_0: all the population means are <u>equal</u> (no difference)
- The nondirectional alternative hypothesis H_A: The μi's are not all equal
  - H_A : at least one pair of the population means are <u>NOT equal</u> (differ)

The F distributions

- The F distributions named after the statistician and geneticist R. A. Fisher.
- F distribution depends on two parameters:
  - the numerator degrees of freedom: Numerator df = df(between)
  - the denominator degrees of freedom: Denominator df = df(within)
- Critical values for the F distribution are given in **Table 10** at the end of this book.

11.4 The Global F Test

The F test

- The F test is a classical test of the global null hypothesis.
- The test statistic, the **F statistic**, is calculated as follows:

$$F_s = \frac{\text{MS(between)}}{\text{MS(within)}}$$

  – Thus, large values of Fs tend to provide evidence against H_0.

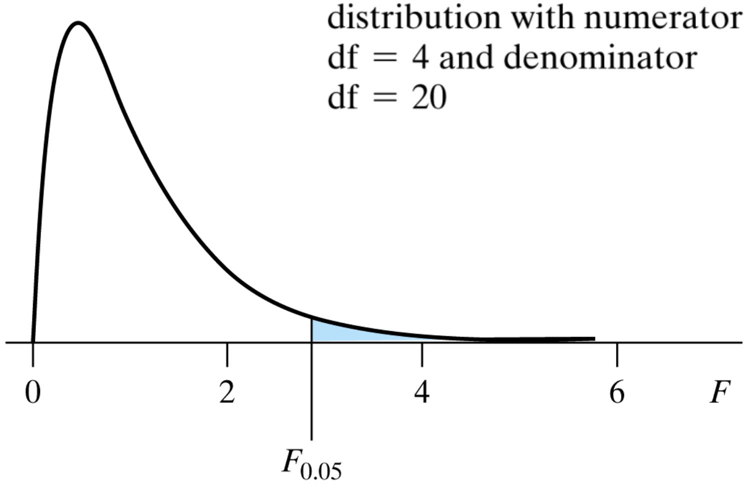

Figure 11.4.1 The F distribution with numerator df = 4 and denominator df = 20

Relationship Between F test And t test

- If only two groups are to be compared (I = 2), use either the F test or the t test
  – It can be shown that the F test and this "pooled" t test are actually equivalent procedures.
  – The test statistics is $t_s^2 = F_s$
  – Because of the equivalence of the tests, the application of the F test to compare the means of two samples will always give exactly the same P-value as the pooled t test applied to the same data.

11.4 The Global F Test

The F test

Example 11.4.1 Weight Gain of Lambs (continued)

- Is there any difference among the diets with respect to population mean weight gain?

Table 11.2.3 ANOVA table for lamb weight gains

Source	df	SS	MS
Between diets	2	36	18.00
Within diets	9	210	23.33
Total	11	246	

11.4 The Global F Test

The F test

Example 11.4.1 Weight Gain of Lambs (continued)

- Is there any difference among the diets with respect to population mean weight gain?
  - The global null hypothesis and alternative can be stated as
    - H_0: Mean weight gain is the same on all three diets

      H_0: $\mu_1 = \mu_2 = \mu_3$
    - H_A: Mean weight gain is NOT the same on all three diets

      H_A: The μ_i's are not all equal
  - From the ANOVA table (Table 11.2.3), $F_s = 18.00/23.33 = 0.77$, Numerator df = 2, Denominator df = 9
  - From Table 10 we find $F(2, 9)_{0.20} = 1.93$, so $P > 0.20$.
  - Thus, there is a lack of significant evidence against H_0;
  - there is insufficient evidence to conclude that there is any difference among the diets with respect to population mean weight gain.

Table 11.2.3 ANOVA table for lamb weight gains

Source	df	SS	MS
Between diets	2	36	18.00
Within diets	9	210	23.33
Total	11	246	

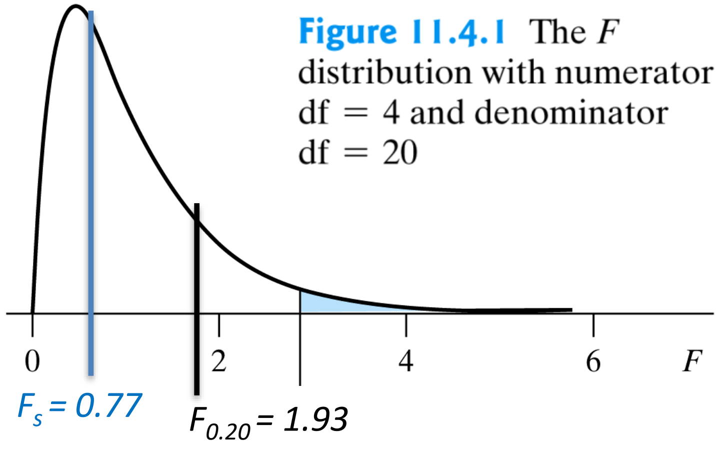

Figure 11.4.1 The F distribution with numerator df = 4 and denominator df = 20

11.5 Applicability of Methods

Standard Conditions

The ANOVA techniques, including the global F test, are valid if the following conditions hold.

1. Design conditions
   - (a) It must be reasonable to regard the groups of observations as random samples from their respective populations.
   - (b) The I samples must be independent of each other.

2. Population conditions
   - The I population distributions must be (approximately) normal with equal standard deviations:*

$$\sigma_1 = \sigma_2 = \dots = \sigma_I$$

   - These conditions are extensions of the conditions given in Chapter 7 for the independent-samples t test with the added condition that the standard deviations be equal. The condition of normal populations with equal standard deviations is less crucial if the sample sizes (n_i) are large and approximately equal.

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

- In the same way we cannot use a two-sample t test when data are paired, when an experiment has been blocked, we no longer can use our ANOVA methods of Section 11.4.
- Instead, we will use a randomized blocks ANOVA model.

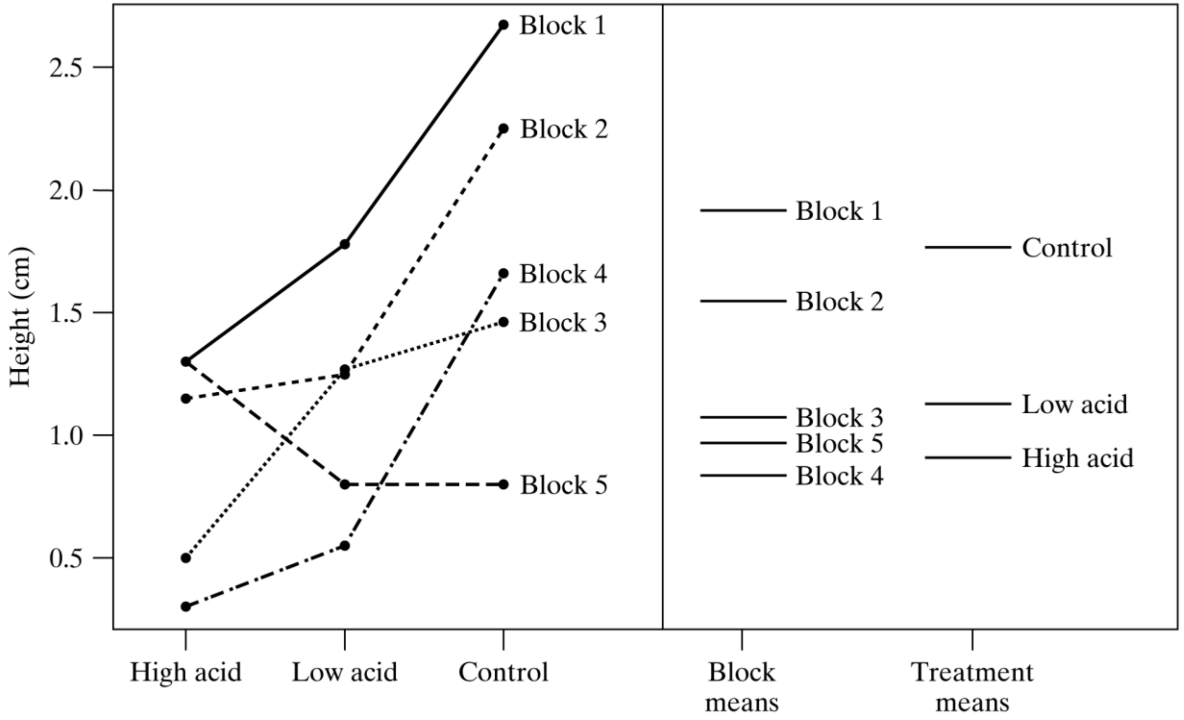

Figure 11.6.3 Dotplots of the alfalfa growth data with a summary of block and treatment means

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

- Creating the blocks
  - create blocks that are as homogeneous within themselves as possible, so that the inherent variation between blocks becomes as far as possible
  - variation between blocks rather than within blocks.

- We extend the ANOVA model presented in Section 11.3 to the following model:

$$y_{ijk} = \mu + \tau_i + \boldsymbol{\beta_j} + \varepsilon_{ijk}$$

  - y_{ijk}: the kth observation when treatment i is applied in block j.
  - μ: represents the grand population mean
  - τ_i: represents the effect of group I
  - **β_j: represents the effect of the jth block.**
  - ε_{ijk}: represents random error associated with observation k in block j and group i

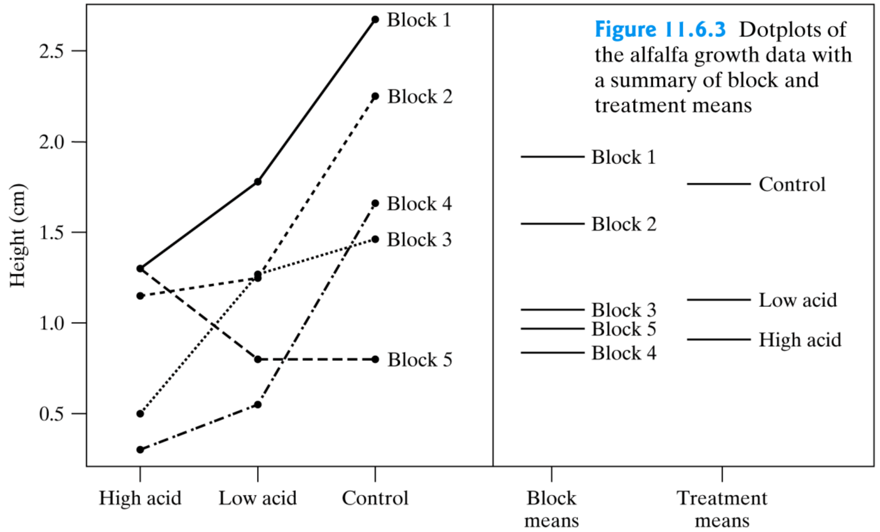

Figure 11.6.3 Dotplots of the alfalfa growth data with a summary of block and treatment means

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

- The one-way randomized complete block F test
  - One-way ANOVA:

    SS(total) = SS(within) + **SS(between)**

  - One-way ANOVA with blocks:

    SS(total) = SS(within) + SS(blocks) + **SS(treatments)**

    - The F test split the one-way ANOVA SS(within) into two parts: SS(blocks) - variability among the block means; SS(within) - the remaining unexplained variation in the data.
    - write SS(treatments) rather than SS(between) to describe the variability between treatment
  - The F test
    - F_s = MS(treatments)/MS(within)
    - Numerator df = df(treatments), Denominator df = df(within)

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

- Mean squares between blocks

Mean Squares Between Blocks

$$\text{MS(blocks)} = \frac{\sum_{j=1}^{J} m_j(\bar{y}_{\bullet j} - \bar{\bar{y}})^2}{J - 1}$$

- The sum of squares, or SS(blocks), and the total degrees of freedom, or df(blocks), are defined as follows:

Sum of Squares and df Between Blocks

$$\text{SS(blocks)} = \sum_{j=1}^{J} m_j(\bar{y}_{\bullet j} - \bar{\bar{y}})^2$$

$$\text{df(blocks)} = J - 1$$

m_j :the number of observations in block j; $\bar{y}_{\cdot j}$: average of the observations in block j

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

ANOVA Table – One-way ANOVA with blocks

– The F statistic: F_s = MS(treatments)/MS(within)

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between treatments	$I - 1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Between blocks	$J - 1$	$\sum_{j=1}^{J} m_j(\bar{y}_{\cdot j} - \bar{\bar{y}})^2$	SS/df
Within groups	$n. - I - J + 1$	**SS(within)** = SS(total) – SS(treatment) – SS(blocks)	SS/df
Total	$n. - 1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$	

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

Example 11.6.5 Agricultural Field Study

- Effect of acid on the growth rate of alfalfa plants.
- 3 treatment groups: low acid, high acid, and control.
- 5 cups for each of the 3 treatments, for a total of 15 observations.
- However, the cups were arranged near a window and researchers wanted to account for the effect of differing amounts of sunlight.
- Will acid affect the growth of alfalfa plants?

Figure 11.6.1 Design of the alfalfa experiment

Window	Block 1	Block 2	Block 3	Block 4	Block 5
	high	control	control	control	high
	control	low	high	low	low
	low	high	low	high	control

Organization of blocks for alfalfa experiment

Table 11.6.3 Alfalfa plant height after 5 days (cm)

	High acid	Low acid	Control	Block mean
Block 1	1.30	1.78	2.67	1.917
Block 2	1.15	1.25	2.25	1.550
Block 3	0.50	1.27	1.46	1.077
Block 4	0.30	0.55	1.66	0.837
Block 5	1.30	0.80	0.80	0.967
Treatment mean $= \bar{y}_i$	0.910	1.130	1.768	
n	5	5	5	

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

Example 11.6.5 Agricultural Field Study

- Will acid affect the growth of alfalfa plants?
  - The grand mean: $\bar{\bar{y}}$ = (1.30 + 1.15 + … + 0.80)/15 = 19.04/15 = 1.269
  - SS(treatments) = $5(0.910 - 1.269)^2 + 5(1.130 - 1.269)^2 + 5(1.768 - 1.269)^2 = 1.986$
  - df(treatments) = I-1 = 3 - 1 = 2
  - SS(blocks) = $3(1.917 - 1.269)^2 + 3(1.550 - 1.269)^2 + 3(1.077 - 1.269)^2 + 3(0.837 - 1.269)^2 + 3(0.967 - 1.269)^2 = 2.441$
  - df(blocks) = J-1 = 5 - 1 = 4
  - The total sum of squares is found as $(1.30 - 1.269)^2 + … + (0.80 - 1.269)^2 = 5.879$.
  - SS(within) = SS(total) - SS(treatments) - SS(blocks) = 5.879 - 1.986 - 2.441 = 1.452
  - df(within) = df(total) - df(treatments) - df(blocks) = 14 - 2 - 4 = 8.

Table 11.6.3 Alfalfa plant height after 5 days (cm)

	High acid	Low acid	Control	Block mean
Block 1	1.30	1.78	2.67	1.917
Block 2	1.15	1.25	2.25	1.550
Block 3	0.50	1.27	1.46	1.077
Block 4	0.30	0.55	1.66	0.837
Block 5	1.30	0.80	0.80	0.967
Treatment mean $= \bar{y}_i$	0.910	1.130	1.768	
n	5	5	5	

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between treatments	$I - 1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Between blocks	$J - 1$	$\sum_{j=1}^{J} m_j(\bar{y}_{\cdot j} - \bar{\bar{y}})^2$	SS/df
Within groups	$n. - I - J + 1$	SS(within) = SS(total) – SS(treatment) – SS(blocks)	SS/df
Total	$n. - 1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$	

Table 11.6.4 ANOVA table for alfalfa experiment

Source	df	SS	MS	F ratio
Between treatments	2	1.986	0.993	5.47
Between blocks	4	2.441	0.610	
Within groups	8	1.452	0.182	
Total	14	5.879		

11.6 One-Way Randomized Blocks Design

Analyzing Data from a Randomized Block Experiment

Example 11.6.5 Agricultural Field Study

- Will acid affect the growth of alfalfa plants?
  - H_0: Mean 5-day growth is the same for all three treatments (high acid, low acid, and control).

$$\mu_1 = \mu_2 = \mu_3.$$

  - H_A : Mean 5-day growth is NOT the same on all three treatments. The µi 's are not all equal.
  - F_s = MS(treatments)/ MS(within) = 0.993/0.182 = 5.47
  - with df = 2 for the numerator and 8 for the denominator.
  - From Table 10 we bracket the P-value as 0.02 < P-value < 0.05.
  - There is significant evidence that acid affects the growth of alfalfa plants.

Table 11.6.3 Alfalfa plant height after 5 days (cm)

	High acid	Low acid	Control	Block mean
Block 1	1.30	1.78	2.67	1.917
Block 2	1.15	1.25	2.25	1.550
Block 3	0.50	1.27	1.46	1.077
Block 4	0.30	0.55	1.66	0.837
Block 5	1.30	0.80	0.80	0.967
Treatment mean $= \bar{y}_i$	0.910	1.130	1.768	
n	5	5	5	

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between treatments	$I-1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Between blocks	$J-1$	$\sum_{j=1}^{J} m_j(\bar{y}_{\cdot j} - \bar{\bar{y}})^2$	SS/df
Within groups	$n. - I - J + 1$	SS(within) = SS(total) – SS(treatment) – SS(blocks)	SS/df
Total	$n. - 1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$	

Table 11.6.4 ANOVA table for alfalfa experiment

Source	df	SS	MS	*F* ratio
Between treatments	2	1.986	0.993	5.47
Between blocks	4	2.441	0.610	
Within groups	8	1.452	0.182	
Total	14	5.879		

11.7 Two-Way ANOVA

Factorial ANOVA

- Some analysis of variance settings involve the simultaneous study of <u>two or more factors.</u>
  - two factors do NOT interact
    - two factors are additive in their effects, if the joint influence of two factors is equal to the sum of their separate influences.
  - two factors interact
    - the effect that one factor has on a response variable depends on the level of a second factor.

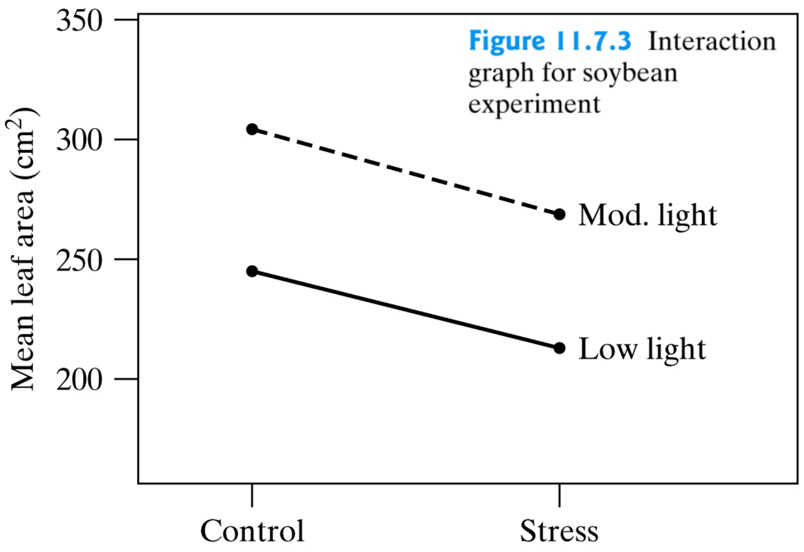

Figure 11.7.3 Interaction graph for soybean experiment

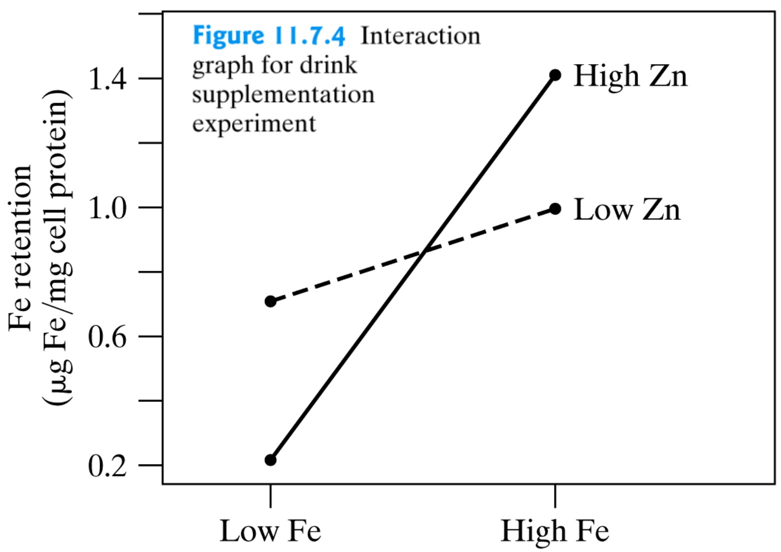

Figure 11.7.4 Interaction graph for drink supplementation experiment

11.7 Two-Way ANOVA

Factorial ANOVA

- We extend the ANOVA model presented in Section 11.3 to the following model:

$$y_{ijk} = \mu + \tau_i + \beta_j + \boldsymbol{\gamma_{ij}} + \varepsilon_{ijk}$$

  - where y_{ijk} is the kth observation of level i of the first factor and level j of the second factor.
  - τ_i represents the effect of level i of the first factor
  - β_j represents the effect of level j of the second factor
  - **γ_{ij} is the effect of the <u>interaction</u> between level i of the first factor and level j of the second factor**

Hypothesis

- The global null hypothesis is H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{IJ} = 0$
- The nondirectional alternative hypothesis H_A: The γ_{ij}'s are not all equal to 0

11.7 Two-Way ANOVA

The F test

- The F statistic: F_s = MS(**interaction**)/MS(within)

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between i treatments	$I - 1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Between j treatments	$J - 1$	$\sum_{j=1}^{J} m_j(\bar{y}_{\cdot j} - \bar{\bar{y}})^2$	SS/df
Interaction	$(I - 1) \times (J - 1)$	*Can be calculated by computer*	SS/df
Within groups	$n. - IJ$	**SS(within)** = SS(total) – SS(treatment) – SS(interaction)	SS/df
Total	$n. - 1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_j}(y_{ij} - \bar{\bar{y}})^2$	

11.7 Two-Way ANOVA

Factorial ANOVA

Example 11.7.3 Iron Supplements in Milk-Based Fruit Beverages

- Effects of drink fortification on the cellular retention of iron
- Researchers conducted an experiment by fortifying milk-based fruit drinks with low and high levels of iron (Fe) and zinc (Zn).
- There were 8 observations at each combination of Fe and Ze supplementation level.
- Whether Fe and Zn supplementation levels interact?

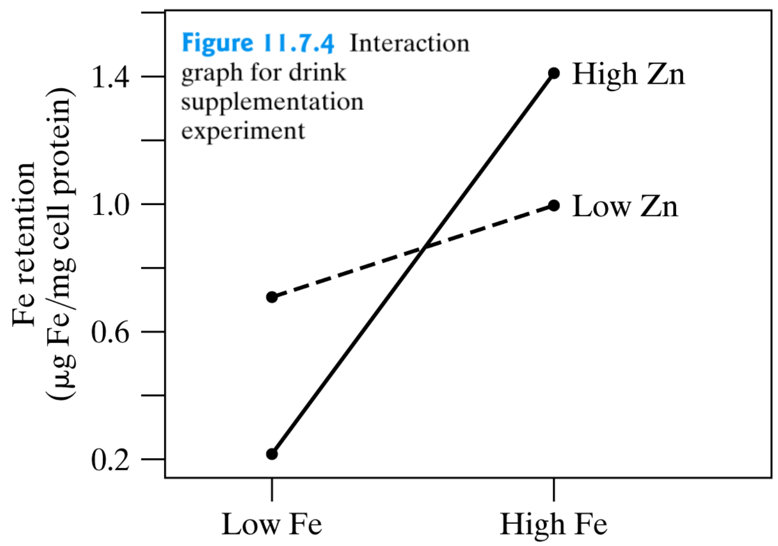

Figure 11.7.4 Interaction graph for drink supplementation experiment

Table 11.7.4 ANOVA table for drink supplement experiment

Source	df	SS	MS	*F* ratio
Between Fe levels	1	4.4023	4.4023	2317.0
Between Zn levels	1	0.0109	0.0109	5.74
Interaction	1	1.6555	1.6555	871.3
Within groups	28	0.0523	0.0019	
Total	31	6.1210		

11.7 Two-Way ANOVA

Factorial ANOVA

Example 11.7.3 Iron Supplements in Milk-Based Fruit Beverages

- Whether Fe and Zn supplementation levels interact?
  - H_0: $\gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = 0$
  - n. = 4X8 = 32 and df(total) = 32-1 = 31.
  - I = J = 2, df(Fe levels) = df(Zn levels) = 1.
  - df(interaction) = (I - 1) x (J - 1) = 1x1 = 1
  - df(within) = df(total) - df(Fe levels) - df(Zn levels) - df(interaction) = 31 - 1 - 1 - 1 = 28.
  - To test whether Fe and Zn supplementation levels interact, F_s = MS(Interaction)/ MS(within) = 1.6555/0.0019 = 871.3, which has degrees of freedom 1 for the numerator and 28 for the denominator.
  - From Table 10 we bracket the P-value as P-value < 0.0001.
  - The P-value is extremely small, indicating that the interaction pattern seen in Figure 11.7.4 is more pronounced than would be expected by chance alone. Thus, we reject H_0.

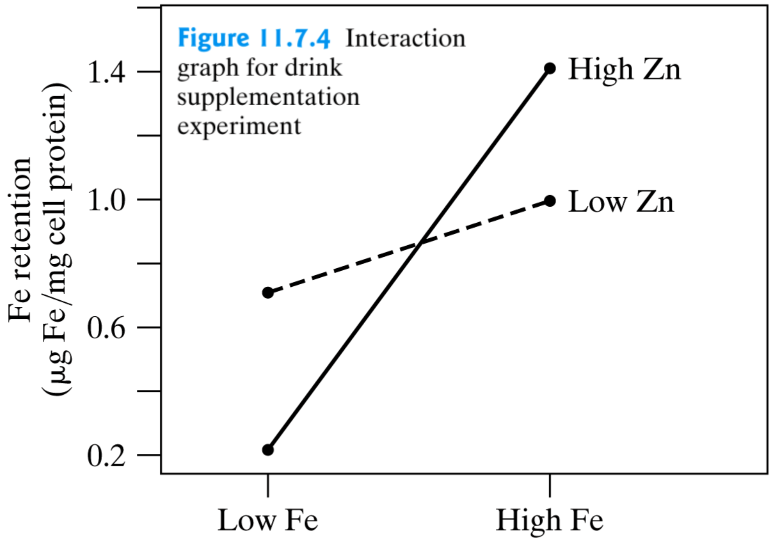

Figure 11.7.4 Interaction graph for drink supplementation experiment

Table 11.7.4 ANOVA table for drink supplement experiment

Source	df	SS	MS	*F* ratio
Between Fe levels	1	4.4023	4.4023	2317.0
Between Zn levels	1	0.0109	0.0109	5.74
Interaction	1	1.6555	1.6555	871.3
Within groups	28	0.0523	0.0019	
Total	31	6.1210		

Summary

Chapter 11. Comparing the Means of Many Independent Samples

- 11.1 Introduction
- 11.2 The Basic One-Way Analysis of Variance
- 11.3 The Analysis of Variance Model
- 11.4 The Global F Test
- 11.5 Applicability of Methods
- 11.6 One-Way Randomized Blocks Design
- 11.7 Two-Way ANOVA

Homework

Chapter 11

- 11.2.1; 11.2.5
- 11.4.4; 11.4.7
- 11.6.10;
- 11.7.4