商業數量方法

Summer 2026

政治大學統計系余清祥

2026年7月6日及7日

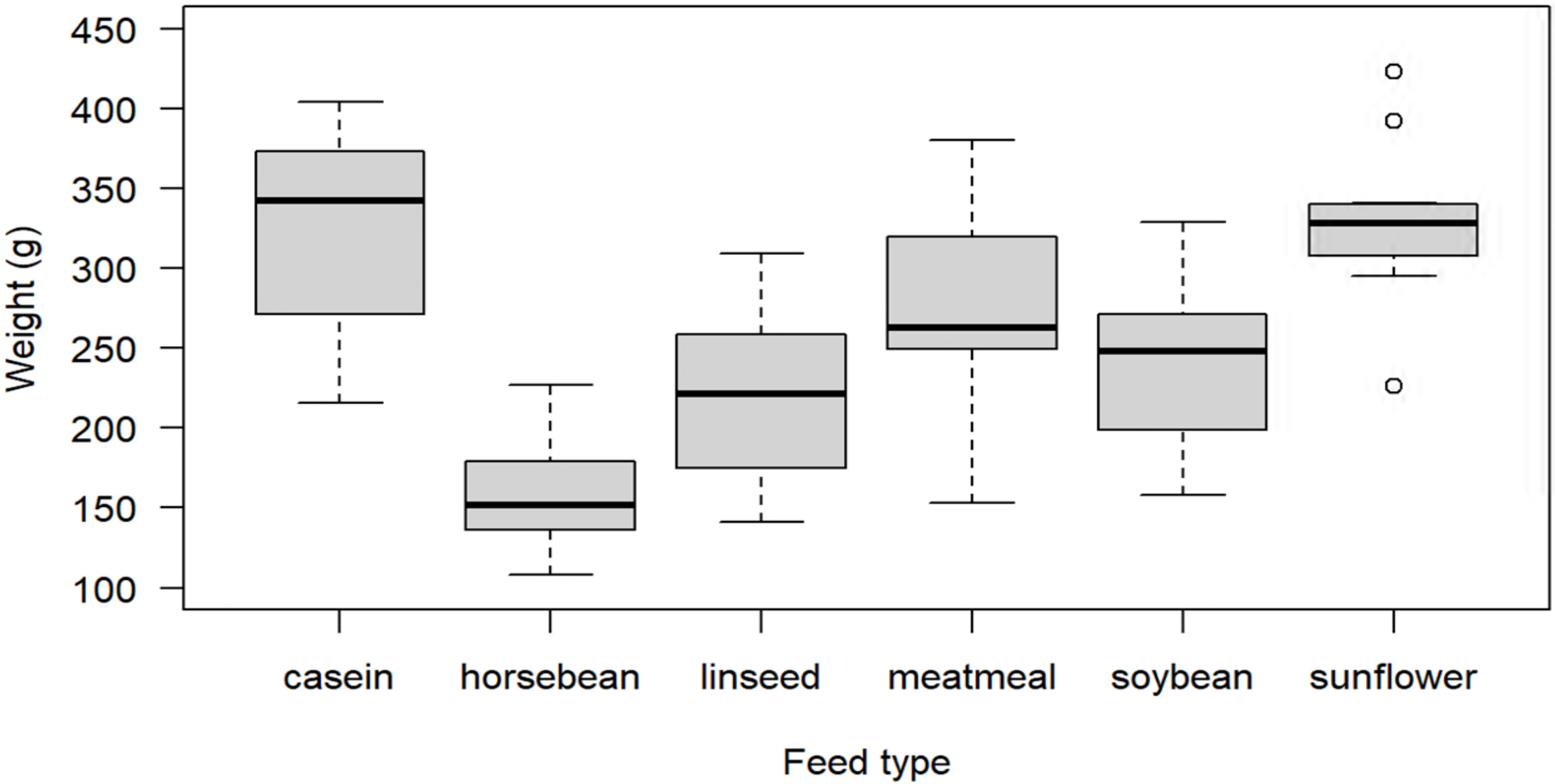
第十三章：變異數分析

http://csyue.nccu.edu.tw

Chapter Contents

CENGAGE

R data “chickwts”, weights of chickens fed



https://saeappliedstats.tech/GitHub-Bookdown_files/figure-html/unnamed-chunk-237-1.png

Introduction

Statistical studies can be classified as either experimental or observational.

- In an experimental study, a variable of interest is identified, one or more other variables are identified and controlled, and data are collected about how those variables influence the variable of interest.
- In an observational study, data are usually obtained through sample surveys that use good design principles, but the lack of rigorous controls associated with an experimental study prevents to establish cause-and-effect relationships.

In this chapter, we introduce three types of experimental designs:

- a completely randomized design
- a randomized block design
- a factorial experiment.

For each design we show how analysis of variance (ANOVA) can be used to analyze the data available. ANOVA can also be used to analyze data obtained through observational studies.

13.1 An Introduction to Experimental Design

Managers at Chemitech, Inc. want to implement an experimental study to determine which of three assembly methods can produce the greatest number of filtration systems per week.

In this example, the assembly method is an independent categorical variable, called the **factor** of the study, and it consists of three outcomes: method A, method B, and method C.

Each outcome is called a **treatment**. Each treatment defines one of the three populations of interest in the experimental study.

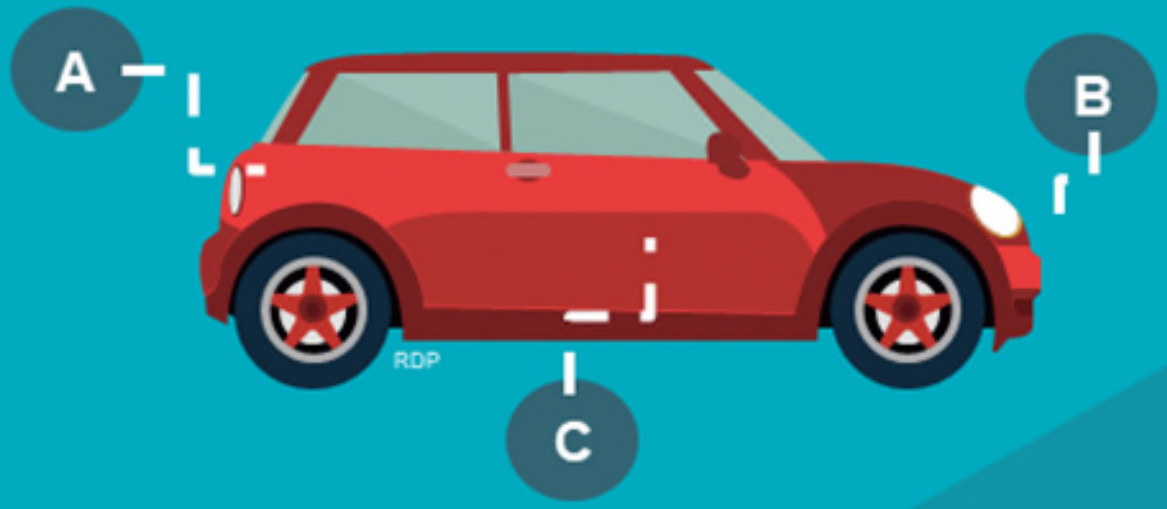
The Chemitech problem is an example of a **single-factor experiment**, as it involves one categorical factor (method of assembly.) More complex problems consisting of multiple factors are possible.

The dependent or **response** variable is the number of filtration systems assembled per week.

The primary statistical objective of the experiment is to determine whether the mean number of units produced per week is the same for all three populations (methods).

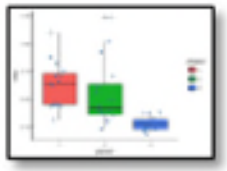
Analysis of Variance (ANOVA)

Bundling Campaign Contributions



https://miro.medium.com/max/1750/1*EhVFonWEfo9BrytCdcVV6Q.png

Analysis of Variance



A method to determine whether there are any statistically significant differences between the population means of three or more independent (unrelated) groups

Definition

Characteristic

- *respondent / dependent variable is a continuous variable*
- *independent variable is categorical variable with two or more group*
- *the data of all groups are normally distributed and homogeneous*
- *two or more groups are not related*

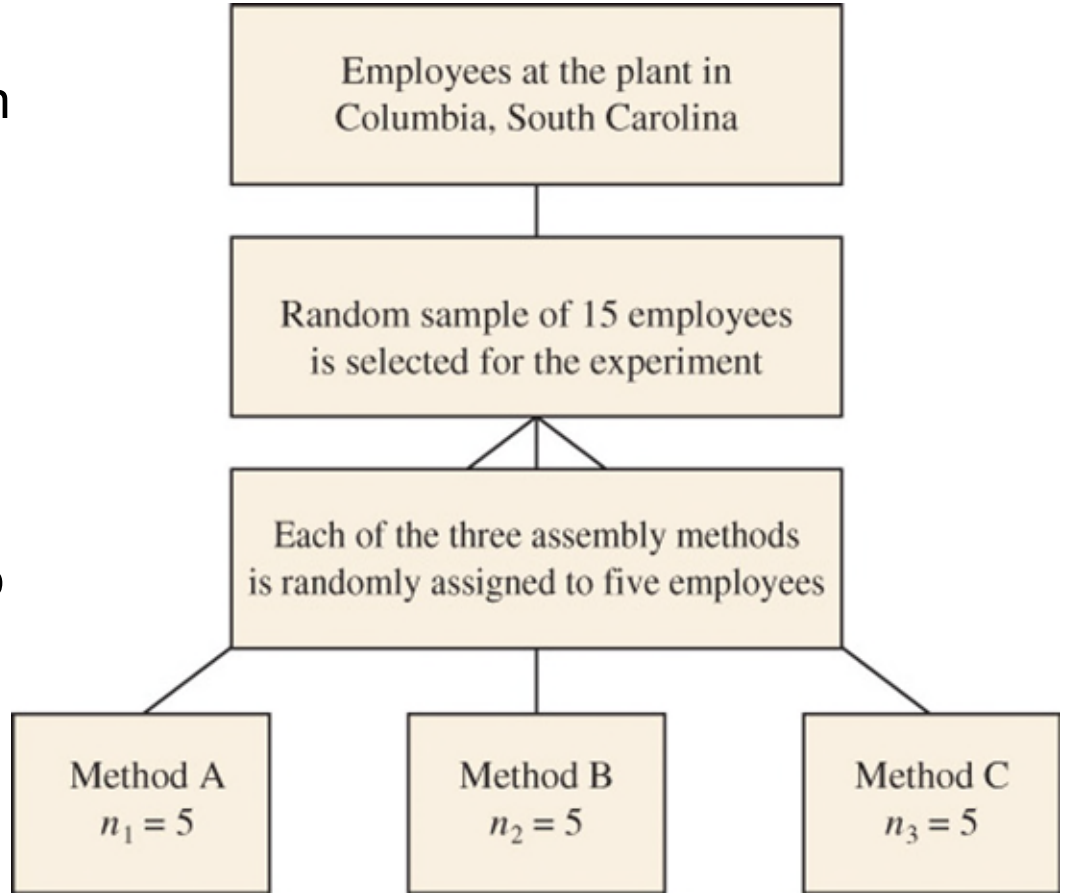
13.1 Application of a Completely Randomized Design

A random sample of 15 workers is selected from all assembly workers at the Chemitech production facility and each of the three assembly methods is randomly assigned to five employees.

In experimental design terminology, the 15 randomly selected workers are the **experimental units**, and the experimental design used is called a **completely randomized design**.

Because each method of assembly is assigned to five workers, we say that five *replicates* have been obtained.

The process of **replication** is another important principle of experimental design.



CENGAGE

13.1 Hypothesis Test for a Completely Randomized Design

In the Chemitech case, each worker assembles for one week the new filtration systems using the assigned method after being trained, and the number of units assembled is recorded.

DATAFile: *Chemitech*

The table to the right shows the sample mean, sample variance, and sample standard deviation for the three methods.

Method	A	B	C
	58	58	48
	64	69	57
	55	71	59
	66	64	47
	67	68	49
$\bar{x}$	62	66	52
s^2	27.5	26.5	31.0
s	5.244	5.148	5.568

The objective is to determine whether the differences between the sample means are large enough to make a conclusion about differences existing between the underlying population means.

In statistical terms, we write the hypotheses as follows, with index 1 corresponding to method A, index 2 to method B, and index 3 to method C.

H_o: $\mu_1 = \mu_2 = \mu_3$

H_a: Not all population means are equal

13.1 Assumptions for Analysis of Variance

Analysis of variance (ANOVA) is the statistical procedure used to determine whether the observed differences in the three sample means are large enough to reject the null hypothesis.

Three assumptions are required to conduct an analysis of variance.

1. **For each population, the response variable is normally distributed**.

   Implication: In the Chemitech experiment, the number of units produced per week (response variable) must be normally distributed for each assembly method.

2. **The variance of the response variable, σ^2, is the same for all of the populations**.

   Implication: In the Chemitech experiment, the variance of the number of units produced per week must be the same for each assembly method.

3. **The observations must be independent**.

   Implication: In the Chemitech experiment, the number of units produced per week for each worker must be independent of the number of units produced per week for any other worker.

13.1 Analysis of Variance: a Conceptual Overview of the Sampling Distribution of $\bar{x}$ when H_0 Is True

if we assume the null hypothesis true, we can think of each of the three sample means, $\bar{x}_1 = 62$, $\bar{x}_2 = 66$, and $\bar{x}_3 = 52$ as values drawn at random from the same sampling distribution of $\bar{x}$.

In this case, we can use the three sample means to estimate mean and variance of the sampling distribution of $\bar{x}$ as follows.

The overall sample mean: $\bar{\bar{x}} = (62 + 66 + 52)/3 = 60$

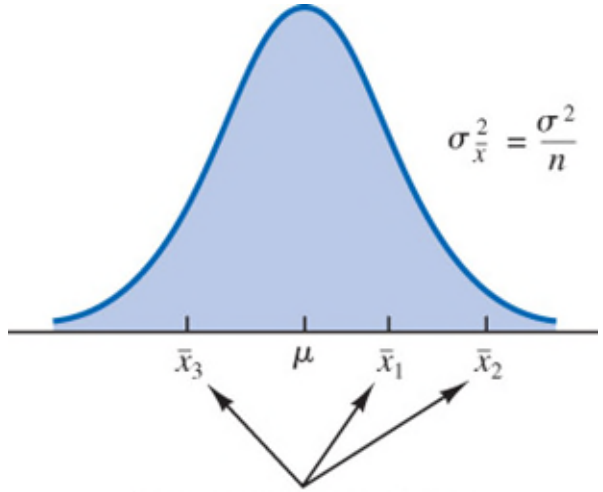
Variance of the sample means:

$$s_{\bar{x}}^2 = [(62 - 60)^2 + (66 - 60)^2 + (52 - 60)^2]/(3 - 1) = 52$$

Because $\sigma_{\bar{x}}^2 = \sigma^2/n$, solving for σ^2 gives

Estimate of $\sigma^2 = n(\text{Estimate of } \sigma_{\bar{x}}^2) = ns_{\bar{x}}^2 = 5(52) = 260$

$ns_{\bar{x}}^2$ is referred to as the **between-treatments** estimate of σ^2.



Sample means are "close together" because there is only one sampling distribution when H_0 is true

13.1 Analysis of Variance: a Conceptual Overview of the Sampling Distribution of $\bar{x}$ when H_0 Is False

If we assume the null hypothesis false, and suppose that the three population means all differ, then the three sample means must be drawn from distinct sampling distributions.

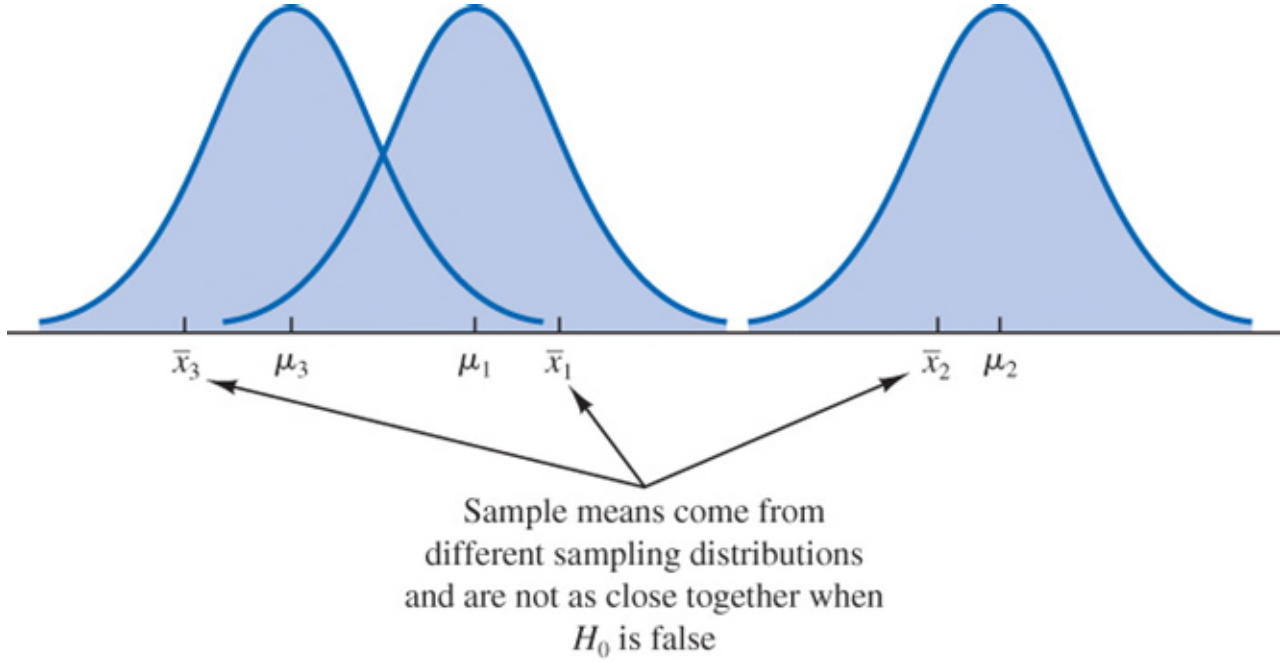
In this case, we provide an unbiased estimate of σ^2 by pooling the three sample variances, s_1^2, s_2^2, and s_3^2, into one overall estimate.

Estimate of σ^2:

$$= (27.5 + 26.5 + 31.0)/3 = 28.33$$

This estimate of σ^2 is called the pooled or **within-treatments** estimate of σ^2.

In general, when the population means are not equal, the between-treatments estimate seen in the previous slide will overestimate the population variance σ^2.



CENGAGE

13.2 Between-Treatments Estimate of Population Variance

The between-treatments estimate of the population variance, σ^2, is called the **mean square due to treatments** and is denoted MSTR.

The general formula for computing MSTR is

$$\boldsymbol{MSTR = \frac{SSTR}{k-1}}$$

Where *k* is the number of treatments and $k-1$ represents the numerator degrees of freedom

$\boldsymbol{SSTR = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}$ is the *sum of squares due to treatments*

n_j is the sample size of treatment *j*, with $j = 1 \ldots k$

$\bar{x}_j$ is the sample mean of treatment *j*, with $j = 1 \ldots k$

$\bar{\bar{x}}$ is the overall sample mean

CENGAGE

13.2 Within-Treatments Estimate of Population Variance

The within-treatments estimate of the population variance, σ^2, is called the **mean square due to error** and is denoted MSE.

Because MSE is based on the variation within each of the treatments, and it is not influenced by whether the null hypothesis is true, MSE always provides an unbiased estimate of σ^2.

The general formula for computing MSE is

$$\boldsymbol{MSE = \frac{SSE}{n_T - k}}$$

Where n_T is the total number of observations and $n_T - k$ the denominator degrees of freedom.

$\boldsymbol{SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2}$ is the *sum of squares due to error*

Where s_j^2 is the sample variance of treatment *j*, with $j = 1 \ldots k$

13.2 Comparing the Variance Estimates: The *F* Test

We can write the hypothesis test for the equality of *k* population means as

H_o: $\mu_1 = \mu_2 = \cdots = \mu_k$

H_a: Not all population means are equal

If H_o is true, MSTR and MSE provide two independent, unbiased estimates of σ^2. Thus, the test statistic for the equality of *k* population means is written as the ratio of the two mean squares.

$$\boldsymbol{F = \frac{MSTR}{MSE}}$$

The test statistic follows an *F* distribution with $k - 1$ degrees of freedom in the numerator and $n_T - k$ degrees of freedom in the denominator (*see notes.)

The following rejection rules apply

p-value approach: Reject H_0 if $\text{p-value} \leq \alpha$

critical value approach: Reject H_0 if $F \geq F_\alpha$

13.2 Test Statistic for the Chemitech Completely Randomized Design

The calculations of the *F* statistic for the results of the Chemitech experiment are as follows.

$$SSTR = \sum_{j=1}^{3} n_j(\bar{x}_j - \bar{\bar{x}})^2 = 5(62-60)^2 + 5(66-60)^2 + 5(52-60)^2 = 520$$

$$MSTR = SSTR/(k-1) = 520/(3-1) = 260$$

$$SSE = \sum_{j=1}^{3} (n_j - 1)s_j^2 = (5-1)27.5 + (5-1)26.5 + (5-1)31.0 = 340$$

$$MSE = SSE/(n_T - k) = 340/(15-3) = 28.33$$

The test statistic for the equality of the population means for the three assembly methods is

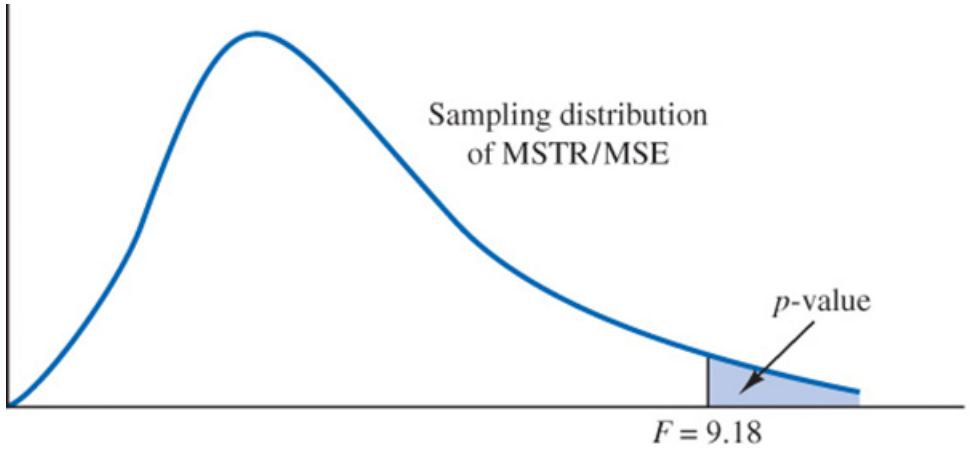
$$F = MSTR/MSE = 260/28.33 = 9.18$$

13.2 *p*-Value Approach to an *F* Test

In the Chemitech example, the *p*-value is the upper tail of an F distribution with $k - 1 = 3 - 1 = 2$ and $n_T - k = 15 - 3 = 12$ degrees of freedom.

Using Table 4 in Appendix B, the F distribution with 2 and 12 degrees of freedom provides the following.

Area in upper tail	**0.10**	**0.05**	**0.025**	**0.01**
***F* Value ($df_1 = 2, df_2 = 12$)**	2.81	3.89	5.10	6.93



Because $F = 9.18$ is greater than 6.93, the values in the "Area in Upper Tail" row show that the *p*-value range must be: $p\text{−value} \leq 0.01$ (*see notes.)

Because p−value $\leq \alpha = 0.05$, we reject the null hypothesis and conclude that the population mean number of units produced per week for the three assembly methods are not equal.

13.2 Critical Value Approach to an *F* Test

With $\alpha = 0.05$, $F_\alpha = F_{0.05}$ provides the critical value for the upper tail *F* test. Using Table 4 in Appendix B, with 2 degrees of freedom at the numerator, and 12 degrees of freedom at the denominator, we obtain the critical value, $F_{0.05} = 3.89$.

Area in upper tail	**0.10**	**0.05**	**0.025**	**0.01**
***F* Value ($df_1 = 2, df_2 = 12$)**	2.81	3.89	5.10	6.93

We can also use Excel to compute the critical value, $F_{0.05}$ =F.INV.RT(0.05,2,12) = 3.89

Thus, the rejection rule for the *F* test is:

Reject H_0 if $F \geq F_{0.05} = 3.89$

Because $F = 9.18$, we reject the null hypothesis and conclude that the population mean number of units produced per week for the three assembly methods are not equal.

CENGAGE

13.2 ANOVA Table for a Completely Randomized Design

The ANOVA table for a completely randomized design is formatted as follows

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*	*p*-value
Treatments	SSTR	$k - 1$	MSTR = SSTR / $(k - 1)$	MSTR / MSE	$Pr(F \geq$ MSTR / MSE$)$
Error	SSE	$n_T - k$	MSE = SSE / $(n_T - k)$		
Total	SST	$n_T - 1$			

The *total sum of squares*, SST, when divided by $n_T - 1$, provides the overall sample variance obtained by treating the entire set of n_T observations as one data set.

Also, because

$$SST = SSTR + SSE$$

and

$$n_T - 1 = (k - 1) + (n_T - k)$$

ANOVA is viewed as the process of **partitioning** SST into its sources of treatments and error.

13.2 ANOVA Table for the Chemitech Experiment

In the ANOVA table for the Chemitech experiment, we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*	*p*-value
Treatments	520	2	260	9.18	0.004
Error	340	12	28.33		
Total	860	14			

Where

$$SST = SSTR + SSE = 520 + 340 = 860$$

$$n_T - 1 = (k - 1) + (n_T - k) = (3 - 1) + (15 - 3) = 2 + 12 = 14$$

$$MSTR = SSTR/(k - 1) = 520/2 = 260$$

$$MSE = SSE/(n_T - k) = 340/12 = 28.33$$

$$F = MSTR/MSE = 260/28.33 = 9.18$$

13.2 Interval Estimate of the Population Means

The square root of MSE provides the best estimate of the population standard deviation σ.

$$s = \sqrt{MSE} = \sqrt{28.33} = 5.323$$

This estimate of σ is a pooled standard deviation because it is derived from MSE.

From our study of interval estimation in Chapter 8, we know that we can write the interval estimate of the population mean for method A of the Chemitech example as

$$\bar{x}_1 \pm t_{\alpha/2}\frac{s}{\sqrt{n_1}}$$

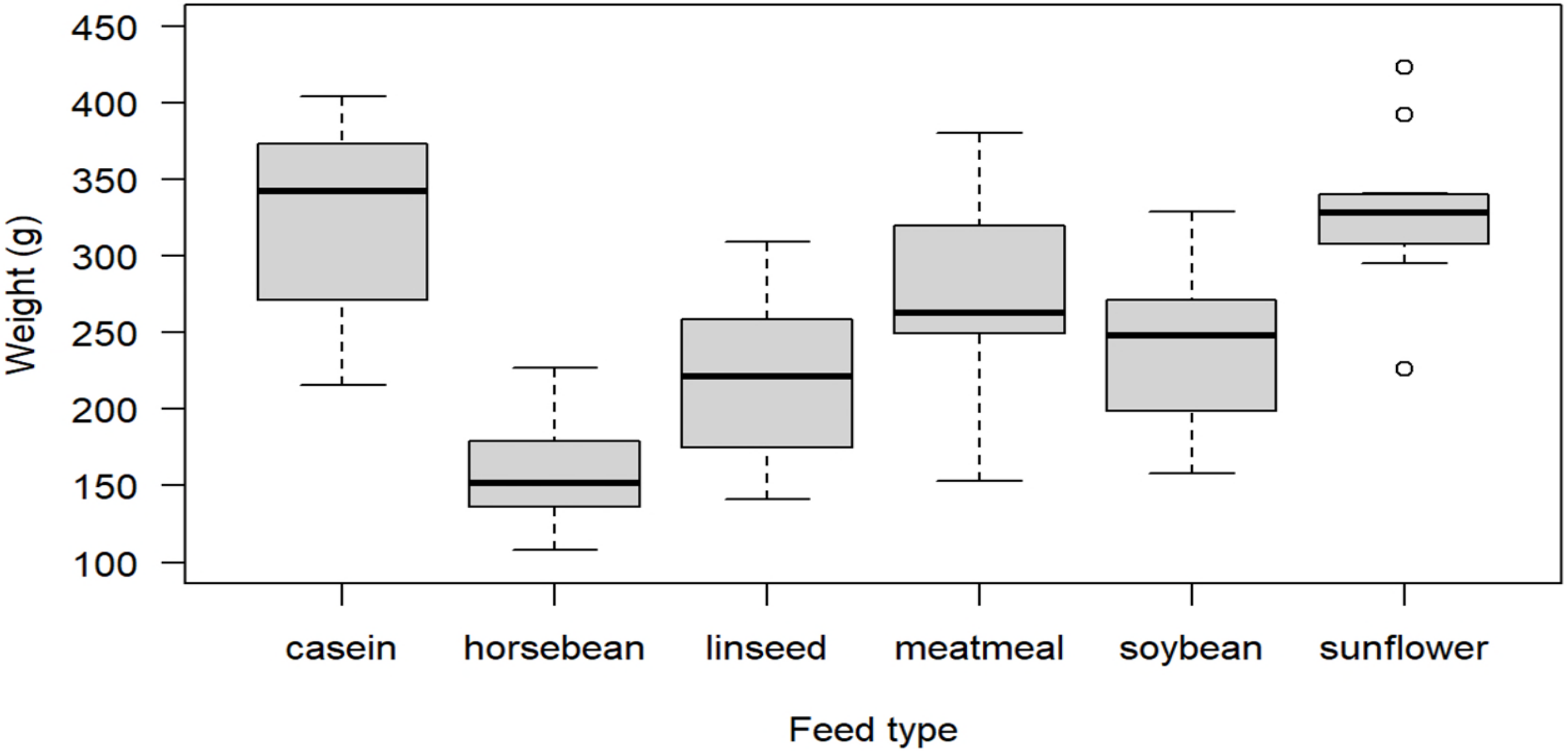
The degrees of freedom for $t_{\alpha/2}$ is $n_T - k = 12$. Using $1 - \alpha = 0.95$, we have

$$\bar{x}_1 \pm t_{.025}\frac{s}{\sqrt{n_1}} = 62 \pm 2.179\frac{5.323}{\sqrt{5}} = 62 \pm 5.19$$

The 95% confidence interval for method A goes from $62 - 5.19 = 56.81$ to $62 + 5.19 = 67.19$.

Similar calculations for methods B and C lead to $(60.81, 71.19)$ and $(46.81, 57.19)$.

Weights of Chickens Fed Different Types of Feed



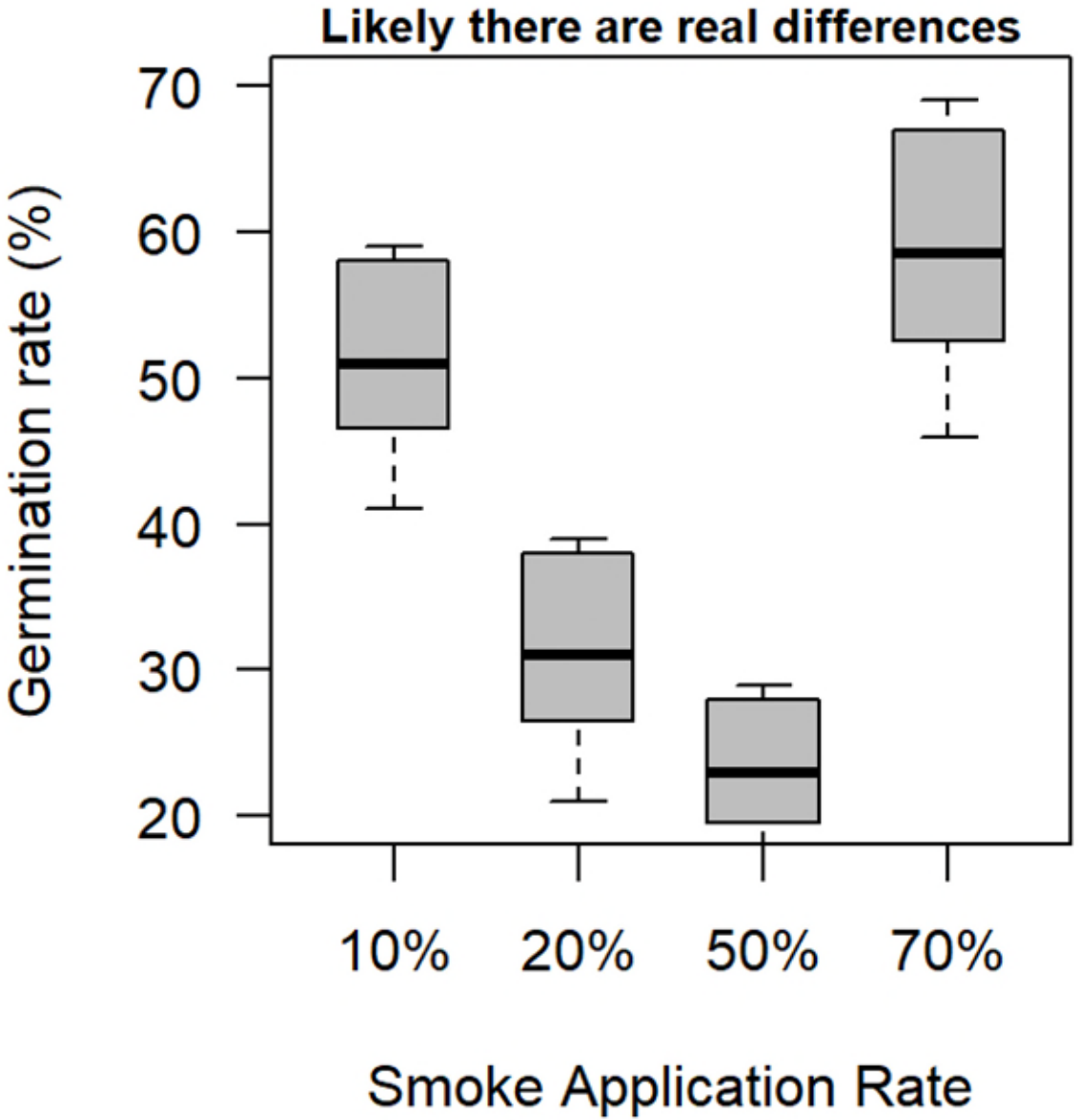
https://saeappliedstats.tech/GitHub-Bookdown_files/figure-html/unnamed-chunk-237-1.png

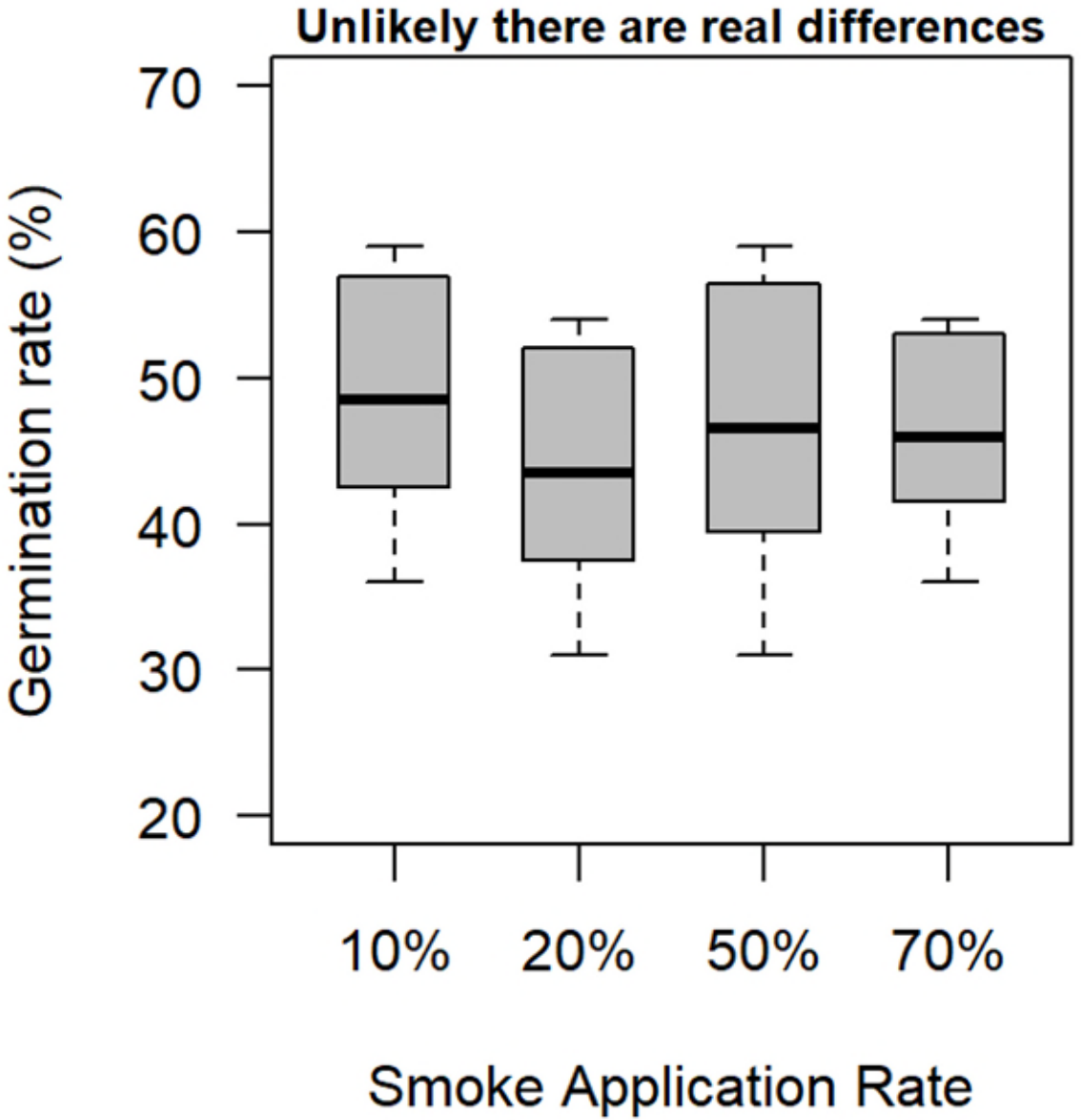
Pairwise t-test p-values: effect of feed on chicken weights

	Casein	Horsebean	Linseed	Meatmeal	Soybean
Horsebean	<0.001	-	-	-	-
Linseed	<0.001	0.094	-	-	-
Meatmeal	0.182	<0.001	0.094	-	-
Soybean	0.005	0.003	0.518	0.518	-
Sunflower	0.812	<0.001	<0.001	0.132	0.003

https://saeappliedstats.tech/analysis-of-variance.html

Boxplot can be useful in detecting between variation!





https://saeappliedstats.tech/analysis-of-variance.html

13.3 Fisher's LSD Procedure

A **multiple comparison procedure** such as **Fisher's Least Significant Difference (LSD)** can be used to conduct statistical comparisons between pairs of population means.

The Fisher's LSD procedure for the comparisons of μ_i and μ_j, can be summarized as follows.

$$H_o: \mu_i = \mu_j$$

$$H_a: \mu_i \neq \mu_j$$

Test statistic:

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

The following rejection rules apply

p-value approach: Reject H_0 if $\text{p}-\text{value} \leq \alpha$

critical value approach: Reject H_0 if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$

13.3 Fisher's LSD Procedure Based on the Test Statistic $\bar{x}_i - \bar{x}_j$

Many practitioners find easier to determine how large the difference between the sample means must be to reject H_0. Thus, this version of the Fisher's LSD procedure is preferred.

H_o: $\mu_i = \mu_j$

H_a: $\mu_i \neq \mu_j$

Rejection rule at a level of significance, α

Reject H_0 if $\left|\bar{x}_i - \bar{x}_j\right| \geq LSD$

Where

$\bar{x}_i - \bar{x}_j$ = test statistic

$$\boldsymbol{LSD = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

CENGAGE

13.3 Application of the Fisher's LSD Procedure

Let us apply the Fisher's LSD procedure to determine whether the population means of method A and method C are different, at a significance level, $\alpha = 0.05$.

The test statistic is $\bar{x}_1 - \bar{x}_3 = 62 - 52 = 10$

For a two-tail test, Table 2 in Appendix B provides $t_{0.025} = 2.179$ with 12 degrees of freedom.

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 12$)**	0.873	1.356	1.782	2.179	2.681	3.055

Thus, we have

$$LSD = t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_3}\right)} = 2.179\sqrt{28.33\left(\frac{1}{5}+\frac{1}{5}\right)} = 7.34$$

Because $|\bar{x}_1 - \bar{x}_3| = 10 \geq 7.34$, we reject H_0 and conclude that the population means for method A and method C differ.

CENGAGE

13.3 Fisher's LSD Procedure for the Complete Chemitech Example

Because the sample sizes for the three treatments in the Chemitech example are equal, $n_1 = n_2 = n_3 = 5$, it follows that only one value for LSD needs to be computed.

The computations for the remaining pairwise test statistics are

method A vs. method B: $\bar{x}_1 - \bar{x}_2 = 62 - 66 = -4$

method B vs. method C: $\bar{x}_2 - \bar{x}_3 = 66 - 52 = 14$

Using $LSD = 7.34$ calculated for the method A vs. method C comparison in the previous slide, we have the following rejection rules for the remaining pairwise comparisons.

method A vs. method B: $|\bar{x}_1 - \bar{x}_2| = 4 < 7.34$ Do not reject H_0

method B vs. method C: $|\bar{x}_2 - \bar{x}_3| = 14 \geq 7.34$ Reject H_0

Thus, we conclude that the population means for method A and method B both differ from the population mean for method C.

13.3 Confidence Interval Estimate of $\mu_1 - \mu_2$ Using Fisher's LSD Procedure

The general procedure to develop a confidence interval estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_i - \bar{x}_j \pm LSD$$

Where $t_{\alpha/2}$ is based on a *t* distribution with $n_T - k$ degrees of freedom.

For the Chemitech experiment, we found $LSD = 7.34$.

The 95% confidence interval estimates for the difference between each pairs of population means in the Chemitech example follow.

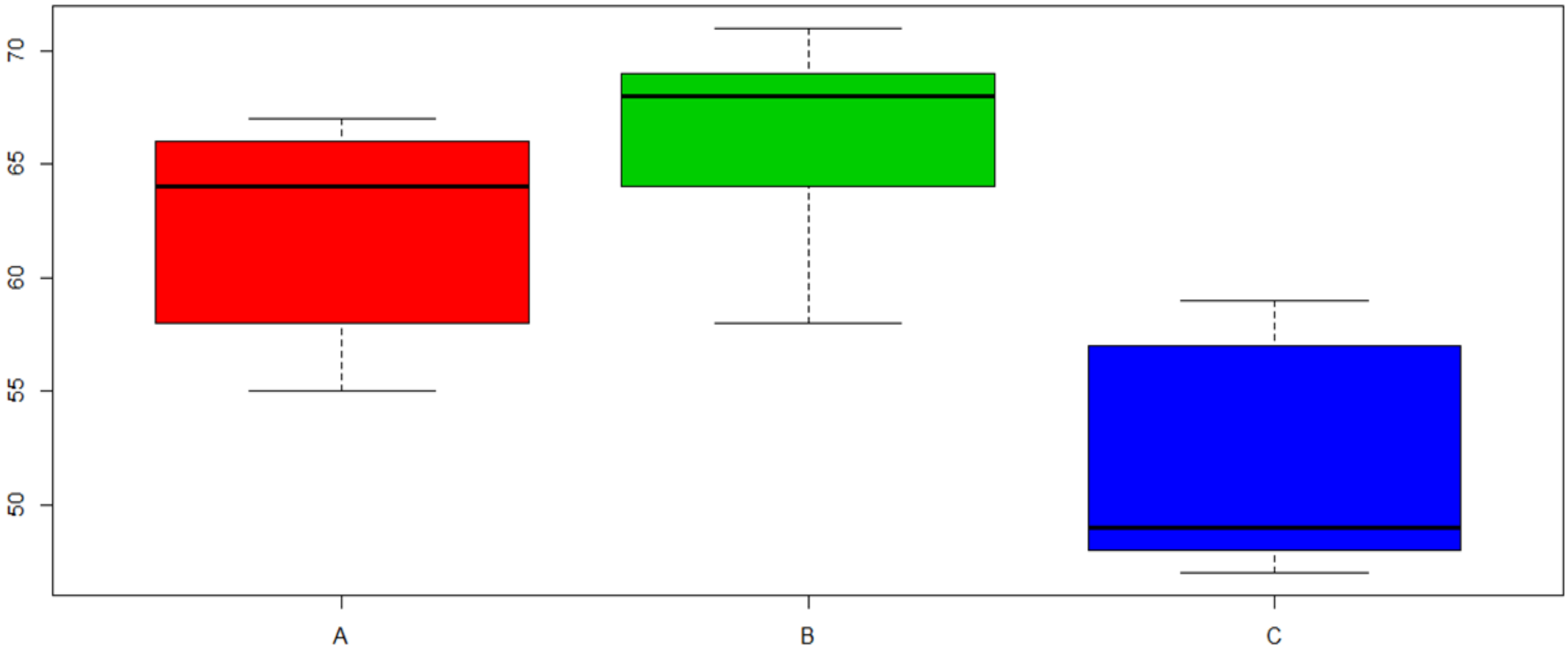
method A vs. method B: $\bar{x}_1 - \bar{x}_2 \pm LSD = -4 \pm 7.34 = (-11.34, 3.34)$

method A vs. method C: $\bar{x}_1 - \bar{x}_3 \pm LSD = 10 \pm 7.34 = (2.66, 17.34)$

method B vs. method C: $\bar{x}_2 - \bar{x}_3 \pm LSD = 14 \pm 7.34 = (6.66, 21.34)$

Because the interval estimate for $\mu_1 - \mu_2$ includes zero, we cannot reject H_0 for the method A vs method B pair of treatment means.

Boxplots of Chemitech Data



13.3 Type I Error Rates

In the Chemitech experiment, we used Fisher's LSD procedure to make three pairwise comparisons at a level of significance, $\alpha = 0.05$.

Thus, for each pairwise comparison, if H_0 is true, the probability to make a Type I error is 0.05.

In multiple comparison procedures, we refer to the level of significance α as the **comparisonwise Type I error rate** associated with a single pairwise comparison.

However, in an experiment with k treatments, the number of C pairwise comparisons is given by the number of combinations of k treatments taken 2 at a time, as seen in Chapter 4.

$$C = C_2^k = \binom{k}{2} = \frac{k!}{2!\,(k-2)!}$$

Thus, the probability of making at least a Type I error for one of the pairwise comparisons is

$$\alpha_{ew} = 1 - (1-\alpha)^C$$

We refer to α_{ew} as the **experimentwise Type I error rate**.

13.3 Type I Error Rates for the Chemitech Experiment

In the Chemitech example, with $k = 3$ treatments, we have

$$C = \binom{3}{2} = \frac{3!}{2!\,(3-2)!} = \frac{6}{(2)(1)} = 3 \text{ combinations}$$

and

$$\alpha_{ew} = 1 - (1-\alpha)^C = 1 - (1-0.05)^3 = 0.143$$

It should be noted that the experimentwise Type I error rate gets considerably larger for problems with more populations.

For example, with $\mathrm{k} = 4$ treatments, we would have $C = 6$ combinations, and

$$\alpha_{ew} = 1 - (1-0.05)^6 = 0.265$$

And with $\mathrm{k} = 5$ treatments, we would have $C = 10$ combinations, and

$$\alpha_{ew} = 1 - (1-0.05)^{10} = 0.401$$

CENGAGE

13.3 Bonferroni Adjustment

The Bonferroni adjustment allows to control the overall experimentwise error rate using a smaller comparisonwise error rate for each test, calculated as α/C, where *C* is the number of pairwise comparisons.

For example, in the Chemitech example, $\alpha/C = 0.05/3 = 0.017$. Thus, the experimentwise error rate would decrease to: $\alpha_{ew} = 1 - (1 - 0.05/3)^3 = 0.049$.

For $k = 4$, $C = 6$, and $\alpha/C = 0.05/6 = 0.0083$. Thus, $\alpha_{ew} = 1 - (1 - 0.05/6)^6 = 0.049$.

And for $k = 5$, $C = 10$, and $\alpha/C = 0.05/10 = 0.005$. Thus, $\alpha_{ew} = 1 - (1 - 0.05/10)^{10} = 0.049$.

However, as we learned in Chapter 9, for a fixed sample size, any decrease in the probability of making a Type I error will result in an increase in the probability of making a Type II error.

As a result, many practitioners are reluctant to apply the Bonferroni adjustment to control the probability of making a Type I error and adopt instead other procedures such as the Tukey's HSD procedure or Duncan's multiple range test.

13.4 Randomized Block Design

Recollect how in the completely randomized design, the treatments are randomly assigned to the experimental units.

A problem can arise whenever differences due to extraneous factors (ones not considered in the experiment) cause the experimental units to be heterogeneous.

When that happens, the MSE term becomes large, making the *F* statistic smaller, and signaling no difference among treatment means when in fact such a difference exists.

A **randomized block design** controls some of the extraneous sources of variation by **blocking** the experimental units into homogenous group, thus removing the sources of variation from the MSE term.

A randomized block design provides a better estimate of the true error variance and leads to a more powerful hypothesis test in terms of its ability to detect differences among treatment means.

13.4 ANOVA Table for a Randomized Block Design

The ANOVA procedure for the randomized block design partitions the total sum of squares (SST) into three groups: sum of squares due to treatments (SSTR), sum of squares due to blocks (SSBL), and sum of squares due to error (SSE).

$$SST = SSTR + SSBL + SSE$$

The total degrees of freedom, $n_T - 1$, are partitioned so that

$$n_T - 1 = (k - 1) + (b - 1) + (k - 1)(b - 1)$$

Where *k* is the number of treatments, *b* the number of blocks, and $n_T = kb$ the total number of observations.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*	*p*-value
Treatments	SSTR	$k - 1$	MSTR = SSTR / $(k - 1)$	MSTR / MSE	$Pr(F \geq$ MSTR / MSE$)$
Blocks	SSBL	$b - 1$	MSBL = SSBL / $(b - 1)$		
Error	SSE	$(k - 1)(b - 1)$	MSE = SSE / $[(k - 1)(b - 1)]$		
Total	SST	$n_T - 1$			

CENGAGE

13.4 Air Traffic Controller Stress Test

Three workstation alternatives (system A, system B, and system C) have been selected as having the best potential for reducing fatigue and stress of air traffic controllers.

To account for the within-group source of variation (MSE) due to individual controller differences, a randomized block design was selected.

In experimental design terminology:

- the workstation is the *factor of interest*
- the three workstation alternatives (systems A, B, and C) are the three treatments
- the sample of six controllers assigned to each treatment are the *blocks*

The *randomized* aspect of the randomized block design is the random order in which the treatments (systems) are assigned to the controllers.

The *response variable* is the stress level on each system measured for each participating controller during a follow-up interview and medical examination (DATAfile: *AirTraffic.)*

13.4 Summary of Stress Data for the Air Traffic Controller Stress Test

		Treatments			Row or Block Totals	Block Means
		System A	System B	System C		
Blocks	Controller 1	15	15	18	48	$\bar{x}_{1\cdot} = 48/3 = 16.0$
	Controller 2	14	14	14	42	$\bar{x}_{2\cdot} = 42/3 = 14.0$
	Controller 3	10	11	15	36	$\bar{x}_{3\cdot} = 36/3 = 12.0$
	Controller 4	13	12	17	42	$\bar{x}_{4\cdot} = 42/3 = 14.0$
	Controller 5	16	13	16	45	$\bar{x}_{5\cdot} = 45/3 = 15.0$
	Controller 6	13	13	13	39	$\bar{x}_{6\cdot} = 39/3 = 13.0$
Column or Treatment Totals		81	78	93	252	$\bar{\bar{x}} = \frac{252}{18} = 14.0$
Treatment Means		$\bar{x}_{\cdot 1} = \frac{81}{6}$	$\bar{x}_{\cdot 2} = \frac{78}{6}$	$\bar{x}_{\cdot 3} = \frac{93}{6}$		
		$= 13.5$	$= 13.0$	$= 15.5$		

CENGAGE

13.4 Computations for the Air Traffic Controller Stress Test

Step 1. Compute the total sum of squares (SST)

$$SST = \sum_{i=1}^{b}\sum_{j=1}^{k}\left(x_{ij} - \bar{\bar{x}}\right)^2 = (15-14)^2+(15-14)^2+(18-14)^2+\ldots+(13-14)^2 = 70$$

Step 2. Compute the sum of squares due to treatments (SSTR)

$$SSTR = b\sum_{j=1}^{k}\left(\bar{x}_{.j} - \bar{\bar{x}}\right)^2 = 6[(13.5-14)^2+(13.0-14)^2+(15.5-14)^2] = 21$$

Step 3. Compute the sum of squares due to blocks (SSBL)

$$SSBL = k\sum_{i=1}^{b}(\bar{x}_{i.} - \bar{\bar{x}})^2 = 3[(16.0-14)^2+(14.0-14)^2+(12.0-14)^2+\ldots+(13.0-14)^2] = 30$$

Step 4. Compute the sum of squares due to error (SSE)

$$SSE = SST - SSTR - SSBL = 70 - 21 - 30 = 19$$

CENGAGE

13.4 ANOVA Table for the Air Traffic Controller Stress Test

The sums of squares computed in the four steps and divided by their degrees of freedom, provide the corresponding mean square values shown in the ANOVA table below.

The ratio of MSRT and MSE produces the test statistic: $F = MSRT/MSE = 10.5/1.9 = 5.53$

Table 4 in Appendix B, with degrees of freedom $k - 1 = 2$ for the numerator, and $(k - 1)(b - 1) = 10$ for the denominator, indicates that $F = 5.53$ is between $F_{.025} = 5.46$ and $F_{.01} = 7.56$.

Area in Upper Tail	**0.10**	**0.05**	**0.025**	**0.01**
F **Value** $(df_1 = 2, df_2 = 10)$	2.92	4.10	5.46	7.56

The *p*-value is the area in the upper tail: $0.01 \leq p\text{−value} \leq 0.025$.

Thus, we reject H_0 and conclude that at least one population mean stress level differs.

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Squares**	***F***	***p*-value**
Treatments	21	2	10.5	10.5 / 1.9 = 5.53	0.024
Blocks	30	5	6.0		
Error	19	10	1.9		
Total	70	17			

CENGAGE

13.5 Factorial Experiment

The completely randomized design and the randomized block design we have considered so far enable us to draw statistical conclusions about one factor.

A **factorial experiment** allows simultaneous conclusions about two or more factors.

In this section, we will show the analysis for a two-factor experiment. The basic approach can be extended to experiments involving more than two factors.

The term *factorial* is used because the experimental conditions include all possible combinations of the factors.

- For example, for a levels of factor A and *b* levels of factor B, the experiment will involve collecting data on *ab* treatment combinations.

In experimental design terminology, the sample size for each treatment combination indicates the number of *replications* r, so that the total number of observations for a two-factor factorial experiment is abr.

CENGAGE

13.5 ANOVA Table for a Two-Factor Experiment

The ANOVA procedure for a two-factor experiment partitions the total sum of squares (SST) into four groups: sum of squares for factor A (SSA), sum of squares for factor B (SSB), sum of squares for interaction (SSAB), and sum of squares due to error (SSE).

$$SST = SSA + SSB + SSAB + SSE$$

The total degrees of freedom, $n_T - 1$, are partitioned so that

$$n_T - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + ab(r - 1)$$

Where a is the number of levels for factor A, b the number of levels for factor B, r the number of replications, and $n_T = abr$ the total number of observations.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*	*p*-value
Factor A	SSA	$a - 1$	MSA = SSA / $(a - 1)$	MSA / MSE	$Pr(F \geq$ MSA / MSE$)$
Factor B	SSB	$b - 1$	MSB = SSB / $(b - 1)$	MSB / MSE	$Pr(F \geq$ MSB / MSE$)$
Interaction	SSAB	$(a - 1)(b - 1)$	MSAB = SSAB / $[(a - 1)(b - 1)]$	MSAB / MSE	$Pr(F \geq$ MSAB / MSE$)$
Error	SSE	$ab(r - 1)$			
Total	SST	$n_T - 1$			

CENGAGE

13.5 The Two-Factor GMAT Experiment

To improve students' performance on the Graduate Management Admissions Test (GMAT), a Texas university is considering offering the following three GMAT preparation programs: (1) a three-hour review session, (2) a one-day program, or (3) an intensive 10-week course.

The second factor of interest is whether a student's undergraduate college affects the GMAT score. Three colleges are being considered: Business, Engineering, and Arts and Sciences.

Assume that a sample of two students from each college will be randomly assigned to each of the programs, for a total of 3 x 3 = 9 treatment combinations and 18 observations.

The table to the right shows the GMAT scores for the 18 students, the response variable for this two-factor experiment.

	Factor B: College		
Factor A: Preparation Program	**Business**	**Engineering**	**Arts and Sciences**
Three-hour review	500 580	540 460	480 400
One-day program	460 540	560 620	420 480
10-week course	560 600	600 580	480 410

CENGAGE

13.5 The Interaction Effect in a Two-Factor Experiment

The analysis of variance computations with the data (DATAfile: *GMATStudy)* will provide answers to the following questions.

- **Main effect (factor A)**: do the preparation programs differ in terms of effect on GMAT scores?
- **Main effect (factor B)**: do the undergraduate colleges differ in terms of effect on GMAT scores?
- **Interaction effect (factors A and B)**: does a student's undergraduate college affects the impact of the different types of preparation program?

The **interaction** refers to a new effect that we can study because we use a factorial experiment that considers all the treatment combinations.

If the interaction effect has a significant impact on the GMAT scores, we can conclude that the effect of the type of preparation program depends on the student's undergraduate college.

13.5 Summary of the Two-Factor GMAT Experiment

	Factor B: College	Business	Engineering	Arts and Science	Row Totals	Factor A means
Factor A: Preparation Program	Three-hour review	$\begin{array}{r}500\\580\\\hline 1080\end{array}$ $\bar{x}_{11} = \frac{1080}{2} = 540$	$\begin{array}{r}540\\460\\\hline 1000\end{array}$ $\bar{x}_{12} = \frac{1000}{2} = 500$	$\begin{array}{r}480\\400\\\hline 880\end{array}$ $\bar{x}_{13} = \frac{880}{2} = 440$	2960	$\bar{x}_{1.} = \frac{2960}{6} = 493.33$
	One-day program	$\begin{array}{r}460\\540\\\hline 1000\end{array}$ $\bar{x}_{21} = \frac{1000}{2} = 500$	$\begin{array}{r}560\\620\\\hline 1180\end{array}$ $\bar{x}_{22} = \frac{1180}{2} = 590$	$\begin{array}{r}420\\480\\\hline 900\end{array}$ $\bar{x}_{23} = \frac{900}{2} = 450$	3080	$\bar{x}_{2.} = \frac{3080}{6} = 513.33$
	10-week course	$\begin{array}{r}560\\600\\\hline 1160\end{array}$ $\bar{x}_{31} = \frac{1160}{2} = 580$	$\begin{array}{r}600\\580\\\hline 1180\end{array}$ $\bar{x}_{32} = \frac{1180}{2} = 590$	$\begin{array}{r}480\\410\\\hline 890\end{array}$ $\bar{x}_{33} = \frac{890}{2} = 445$	3230	$\bar{x}_{3.} = \frac{3230}{6} = 528.33$
	Column Totals	3240	3360	2670	9270	
	Factor B Means	$\bar{x}_{.1} = \frac{3240}{6} = 540$	$\bar{x}_{.2} = \frac{3360}{6} = 560$	$\bar{x}_{.1} = \frac{2670}{6} = 445$	$\bar{\bar{x}} = \frac{9270}{18} = 515$	

CENGAGE

13.5 Computations for the Two-Factor GMAT Experiment

Step 1. Compute the total sum of squares (SST)

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(x_{ijk} - \bar{\bar{x}}\right)^2 = (500-515)^2+(580-515)^2+(540-515)^2+\ldots+(410-515)^2 = 82{,}450$$

Step 2. Compute the sum of squares due to factor A (SSA)

$$SSA = br\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{\bar{x}})^2 = (3)(2)[(493.33-515)^2 + (513.33-515)^2 + (538.33-515)^2] = 6{,}100$$

Step 3. Compute the sum of squares due to factor B (SSB)

$$SSB = ar\sum_{j=1}^{b}\left(\bar{x}_{.j} - \bar{\bar{x}}\right)^2 = (3)(2)[(540-515)^2 + (560-515)^2 + (445-515)^2] = 45{,}300$$

Step 4. Compute the sum of squares for interaction (SSAB)

$$SSAB = r\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}\right)^2 = 2[(540-493.33-540+515)^2 + \ldots = 11{,}200$$

CENGAGE

13.5 ANOVA Table for the Two-Factor GMAT Experiment

Step 5. Compute the sum of squares due to error (SSE)

$$SSE = SST - SSA - SSB - SSAB = 82{,}450 - 6{,}100 - 45{,}300 - 11{,}200 = 19{,}850$$

The sums of squares computed in the previous five steps and divided by their degrees of freedom, provide the corresponding mean square values shown in the ANOVA table below.

The ratios of MSA / MSE, MSB / MSE, and MSAB / MSE produce the test statistics for Factor A, Factor B, and the interaction term.

We can estimate *p*-value ranges using the F values from Table 4 of Appendix B or use statistical software to obtain the exact *p*-values listed below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*	*p*-value
Factor A	6,100	2	3.050	1.38	0.299
Factor B	45,300	2	22,650	10.27	0.005
Interaction	11,200	4	2,800	1.27	0.350
Error	19,850	9	2,206		
Total	82,450	17			

13.5 Conclusions of the GMAT Two-Factor Experiment

Let us use the *p*-values in the ANOVA table and a level of significance, $\alpha = 0.05$, to conduct the hypothesis tests for the two-factor GMAT study.

- **Main effect (factor A)**: because *p*-value = $0.299 > 0.05$, there is no significant difference in the mean GMAT test scores for the three preparation programs.
- **Main effect (factor B)**: because *p*-value = $0.005 \leq 0.05$, there is a significant difference in the mean GMAT test scores among the three undergraduate colleges.
- **Interaction effect (factors A and B)**: because *p*-value = $0.350 > 0.05$ there is no significant interaction effect.

Therefore, because no significant interaction effect is reported, the study provides no reason to believe that the three preparation programs differ in their ability to prepare students from the different colleges for the GMAT.

The only significant effect is due to a student's undergraduate college. Analysis of the treatment means for factor B reveal that students in the Arts and Sciences college appear to be significantly less prepared for the GMAT than students in the other colleges.

Summary

- In this chapter, we showed how analysis of variance can be used to test for differences among means of several populations or treatments.
- We introduced the completely randomized design, the randomized block design, and the two-factor factorial experiment.
- The completely randomized design and the randomized block design are used to draw conclusions about differences in the means of a single factor.
- The primary purpose of blocking in the randomized block design is to remove extraneous sources of variation from the error term, providing a better estimate of the true error variance and a better test to determine whether the treatment means differ significantly.
- In all the experimental designs considered, the analysis of variance is performed by first partitioning the sum of squares and degrees of freedom into their various sources.
- We also showed how Fisher's LSD procedure and the Bonferroni adjustment can be used to perform pairwise comparisons to determine which means are different.