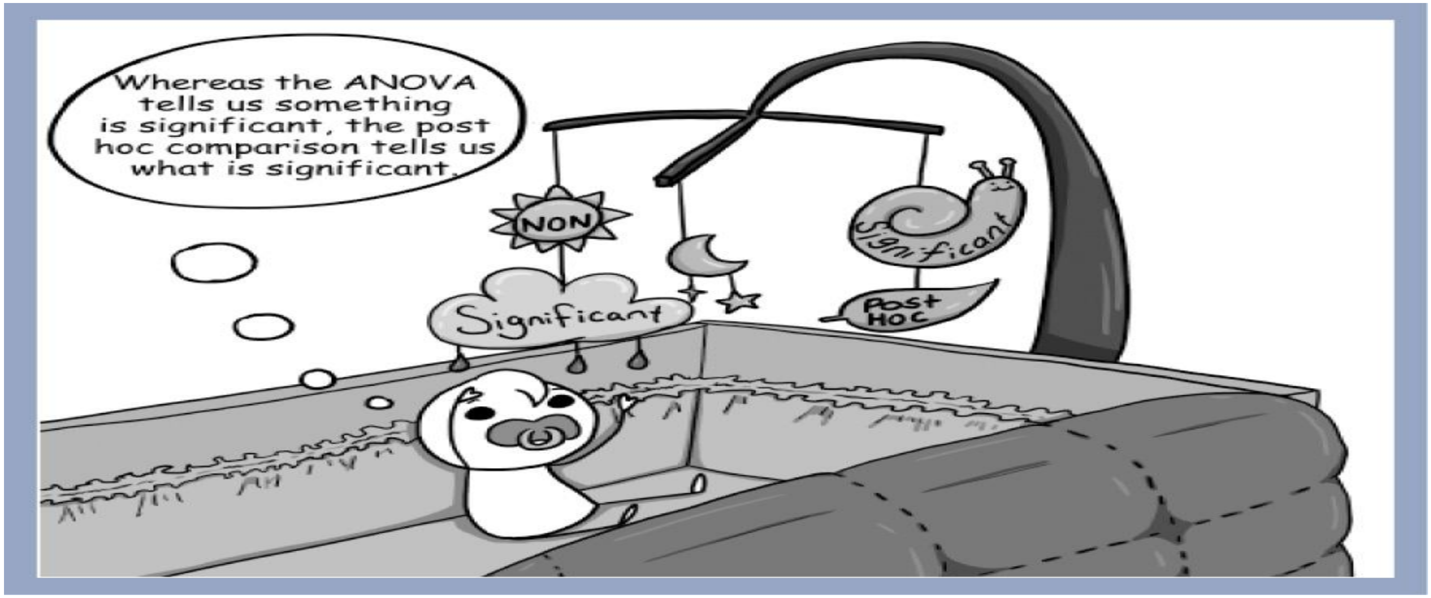
Introductory Statistics: A Problem-Solving Approach

by Stephen Kokoska

Chapter 11

The Analysis of Variance



Introduction

- One-way, or single-factor, ANOVA (ANalysis Of VAriance) involves the analysis of data sampled from more than two populations.
- For example, consider a study in which random samples of the amount of carbon dioxide in underground train tunnels are obtained in four different cities.
  - The data may be used to determine whether there is any difference in the mean amount of carbon dioxide in train tunnels among the four cities.
  - The single factor that varies among the populations is the city.
- Another example: A researcher is investigating techniques for controlling the amount of electricity lost during transmission over five utility lines.
  - Experimental results may be used to determine whether there is any difference in the mean amount of electricity lost for five differently designed lines.
  - The single factor here is the design of the electricity line. In theory, everything else is the same among the five populations.

ANOVA Notation

Population	1	2	$\cdots$	i	$\cdots$	k
Population mean	μ_1	μ_2	$\cdots$	μ_i	$\cdots$	μ_k
Population variance	σ_1^2	σ_2^2	$\cdots$	σ_i^2	$\cdots$	σ_k^2
Sample size	n_1	n_2	$\cdots$	n_i	$\cdots$	n_k
Sample mean	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_i$	$\cdots$	$\bar{x}_k$
Sample variance	s_1^2	s_2^2	$\cdots$	s_i^2	$\cdots$	s_k^2

k = the number of populations under consideration

$n = n_1 + n_2 + \cdots + n_k$ = the total number of observations in the entire data set

Hypotheses: The null and alternative hypotheses are stated in terms of the population means.

- H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ (All k population means are equal.)
- H_a: $\mu_i \neq \mu_j$ for some $i \neq j$ (At least two of the k population means differ.)

To denote observations, use a single letter with two subscripts: x_{ij}.

The first subscript indicates the sample number.

The second subscript denotes the observation number within the sample.

x_{ij} = the jth measurement taken from the ith population

X_{ij} = the corresponding random variable

One-Way ANOVA Assumptions

1. The *k* population distributions are normal.
2. The *k* population variances are equal.
3. The samples are selected randomly and independently from the respective populations.

The ANOVA *F* Statistic

- To determine statistical significance, we need a test statistic:

The **ANOVA *F* statistic** for testing the equality of several means has following form:

$$F = \frac{\text{variation among the sample means } (MSA)}{\text{variation among individuals in the same sample } (MSE)}$$

- *F* is always zero or positive
  - *F* is zero only when all sample means are the same
  - *F* gets larger as means move farther apart
- Large values of *F statistic* are evidence against H_0*: equal means*
- The *F* test is upper-one-sided

- The *F* distributions are a family of right-skewed distributions that take only values greater than 0. A specific *F* distribution is determined by the degrees of freedom of the numerator and denominator of the *F*-statistic.
- Our brief notation will be $F_{(df1,\ df2)}$ with df_1 degrees of freedom in the numerator and df_2 degrees of freedom in the denominator.

The ANOVA *F* Statistic

$$F = \frac{\text{variation among the sample means } (MSA)}{\text{variation among individuals in the same sample } (MSE)}$$

The measures of variation in the numerator and denominator are ***mean squares:***

$$\bar{x} = \bar{x}_{..}$$

- Numerator: **Mean Square for Factor A** (MSA)

$$MSA = \frac{SSA}{k-1} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2}{k-1}$$

SSA is used to denote the sum of squares due to factor (rather than SSF) because a two-way ANOVA includes two factors A and B.
The sample size is used as a weight in the expression for SSA.

- Denominator: **Mean Square for Error** (MSE)

$$MSE = \frac{SSE}{n-k} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n-k}$$

MSE is also called the **pooled sample variance**, written as s_p^2 (s_p is the ***pooled standard deviation***).

One-Way ANOVA Table

One-way ANOVA calculations are often presented in a table:

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Factor (Between groups/samples)	SSA	$k-1$	$\text{MSA} = \dfrac{\text{SSA}}{k-1}$	$\dfrac{\text{MSA}}{\text{MSE}}$
Error (Within groups/samples)	SSE	$n-k$	$\text{MSE} = \dfrac{\text{SSE}}{n-k}$	
Total	SST	$n-1$		

Total Variation Decomposition:

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_{..}\right)^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{k} n_i\left(\bar{x}_{i.}-\bar{x}_{..}\right)^2}_{\text{SSA}} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_{i.}\right)^2}_{\text{SSE}}$$

The mean of the observations in the *i*th sample:

$$\bar{x}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n_i}\left(x_{i1} + x_{i2} + \cdots + x_{in_i}\right)$$

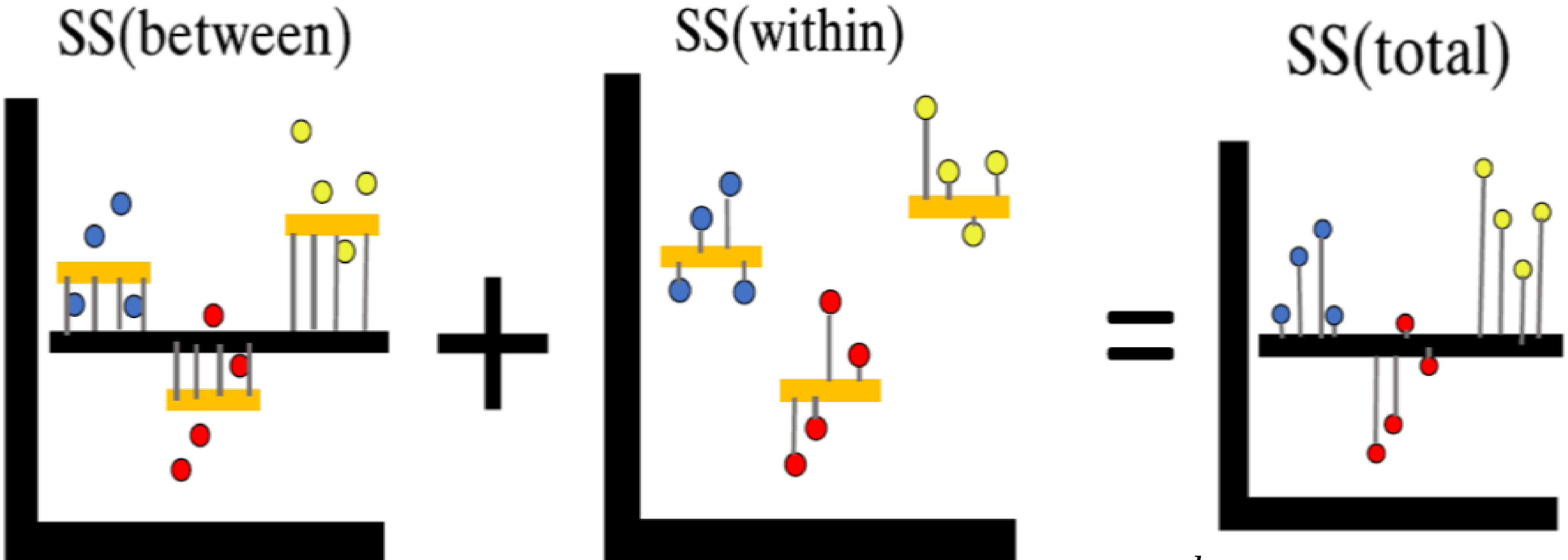
Grand mean: $\bar{x}_{..} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$

sum of all the observations $t_{..} = \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$

sum of all the observations in the *ith* sample $t_{i.} = \displaystyle\sum_{j=1}^{n_i} x_{ij}$

We use Sums of Squares (SS) to quantify these clues

Total variation in the data (the total sum of squares) is the variability of individual observations from the grand mean. It is decomposed into a sum of *between-samples* variation and *within-samples* variation.



$$\sum_{i=1}^{k} n_i\left(\bar{x}_{i.} - \bar{x}_{..}\right)^2 \qquad \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i.}\right)^2 \qquad \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{..}\right)^2$$

Variation is the variability between the sample means. It tells us how different the sample means are from each other

Variation is the variability of the observations from their sample mean

The Test Procedure

Consider the ratio ***F* = MSA/MSE.**

- If the value of F is close to 1, then the two estimates of *variances (MSA and MSE)*, are approximately the same. There is no evidence to suggest that the population means are different.
- If the value of F is much greater than 1, then the variation between samples (MSA) is greater than the variation within samples (MSE). This suggests that the alternative hypothesis is true.
- If the one-way ANOVA assumptions are satisfied and H_0 is true, then the statistic F = MSA/MSE has an F distribution with $k - 1$ and $n - k$ degrees of freedom.
- Because large values of F suggest that H_a is true, the rejection region is only in the right tail of the distribution.
- H_0: $\mu_1 = \mu_2 = \cdot\cdot\cdot = \mu_k$

  H_a: $\mu_i \neq \mu_j$ for some $i \neq j$

  $$\text{TS: } F = \frac{\text{MSA}}{\text{MSE}}$$

  RR: If $F > F_{\alpha,\, k-1,\, n-k}$ OR If $p < \alpha$, then we reject H_0.

Example: Take with a Grain of Salt

A recent research study suggests that a high-salt diet in older women increases the risk of breaking a bone. Although the biological mechanism is still unclear, there does appear to be an association between excessive sodium intake and bone fragility. One measure of bone health is the level of vitamin D in the blood. **As** determined by food questionnaires, independent random samples of older women in four different salt intake categories were obtained. The vitamin D blood level (in nmol/L) was measured in each. Is there any evidence to suggest that at least two of the population mean vitamin D blood levels are different? Use $\alpha = 0.05$. The data are summarized in the table (next slide):

$k = 4$ groups
$n = 20$ total observations

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_a: $\mu_i \neq \mu_j$ for some $i \neq j$

Sample	Observations				
Very high	91.5	77.5	94.5	77.5	92.0
High	89.0	92.0	98.2	80.0	86.7
Moderate	92.5	100.7	94.0	93.3	106.3
Low	100.1	98.0	99.1	103.9	97.6

$$\text{RR: } F > F_{\alpha,k-1,n-k} = F_{0.05,3,16} = \mathbf{3.24}$$

Example: Take with a Grain of Salt

Sample	Sample size	Sample total	Sample mean	Sample variance
Very high	$n_1 = 5$	$t_{1.} = 433.0$	$\bar{x}_{1.} = 86.60$	$s_1^2 = 70.30$
High	$n_2 = 5$	$t_{2.} = 445.9$	$\bar{x}_{2.} = 89.18$	$s_2^2 = 44.94$
Moderate	$n_3 = 5$	$t_{3.} = 486.8$	$\bar{x}_{3.} = 97.36$	$s_3^2 = 35.62$
Low	$n_4 = 5$	$t_{4.} = 498.7$	$\bar{x}_{4.} = 99.74$	$s_4^2 = 6.36$

$$\text{SST} = \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})^2}_{\text{definition}} = \underbrace{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2\right) - \frac{t_{..}^2}{n}}_{\text{computational formula}}$$

$$\text{SSA} = \underbrace{\sum_{i=1}^{k} n_i(\bar{x}_{i.} - \bar{x}_{..})^2}_{\text{definition}} = \underbrace{\left(\sum_{i=1}^{k} \frac{t_{i.}^2}{n_i}\right) - \frac{t_{..}^2}{n}}_{\text{computational formula}}$$

$$\text{SSE} = \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2}_{\text{definition}} = \underbrace{\text{SST} - \text{SSA}}_{\text{computational formula}}$$

Example: Take with a Grain of Salt

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Factor (Between groups/samples)	598.98	3	199.66	$\frac{199.66}{39.31}$
Error (Within groups/samples)	628.862	16	39.31	
Total	1227.872	19		

$$SST = \left(91.5^2 + 77.5^2 + \cdots + 97.6^2\right) - \frac{1864.4^2}{20}$$

$$= 175{,}027.24 - 173{,}799.368 = \mathbf{1227.872}$$

$$t_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} = 1864.4$$

$$SSA = \left(\frac{433.0^2}{5} + \frac{445.9^2}{5} + \frac{486.8^2}{5} + \frac{498.7^2}{5}\right) - \frac{1864.4^2}{20}$$

$$= 174{,}398.348 - 173{,}799.368 = \mathbf{598.98}$$

SSE = SST − SSA = 1227.872 − 598.88 = **628.862**

MSA = SSA/(k − 1)

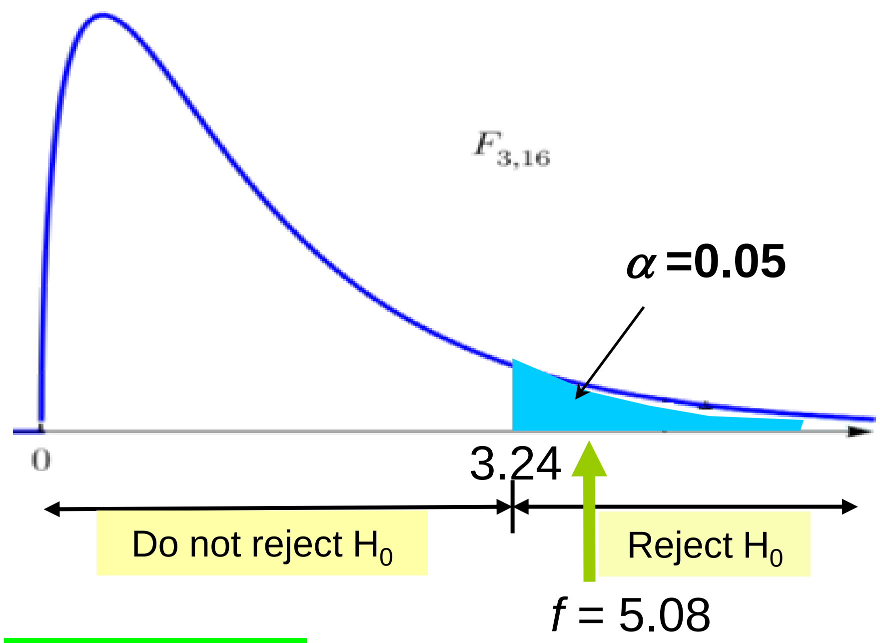
= 598.98/(4 − 1) = 199.66

MSE = SSE/(n − k)

= 628.892/(20 − 4) = 39.31

Example: Take with a Grain of Salt

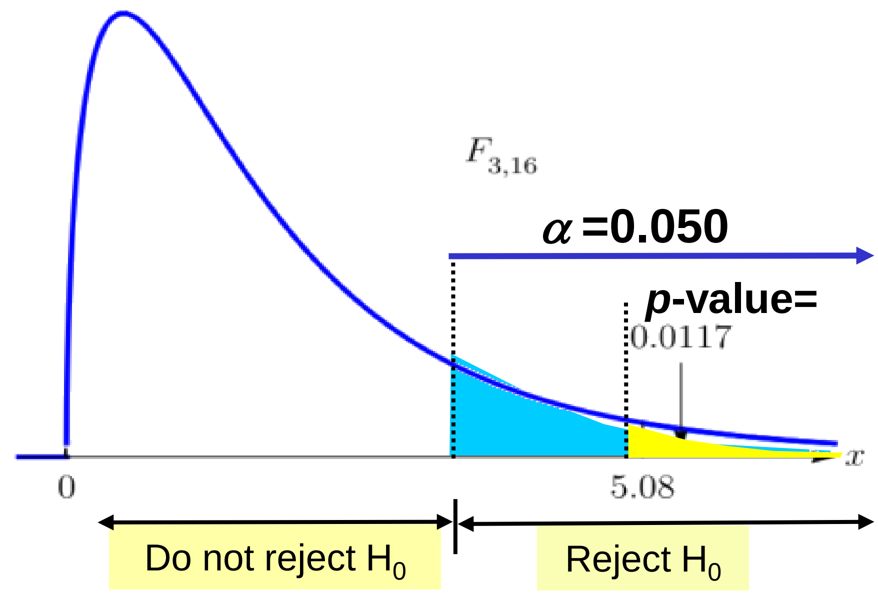
1st. approach using z-critical value



qf(0.95,3,16)
3.238872

Since the f-test statistic of 5.08 is > the critical value of 3.24, we reject the H_0

2nd. approach using p-value



pf(5.08,3,16)
0.9883462 → 1-0.9883462 = 0.0116538
3.24 ≤ 5.08 ≤ 5.29
$F_{0.05,3,16} \leq 5.08 \leq F_{0.01,3,16}$
Therefore, $0.01 \leq p \leq 0.05$

The H_0 would be rejected because the p-value = 1.17% is < α = 5%

Conclusion: There is evidence to suggest that at least two population means are different.

Multiple Comparison Procedures (Isolating Differences)

- If ANOVA overall test showed statistical significance, then a detailed follow-up analysis can examine all pair-wise parameter comparisons to determine which parameters differ from which and by how much.
- The probability of a type I error is set under the assumption that only one test is conducted per experiment. The more tests we conduct, the greater our chance of making an error is.
- Post hoc tests (or posttests) are additional hypothesis tests that are done after an ANOVA to determine exactly which mean differences are significant and which are not.
- In statistical terms, this is called making pairwise comparisons. This process involves performing a series of separate hypothesis tests, and each of these tests includes the risk of a Type I error.
- As you do more and more separate tests, the risk of a Type I error accumulates and is called the experimentwise alpha level.
- Three widely used techniques for making multiple comparisons of a set of treatment means.

Method	**Treatment Sample Sizes**
Bonferroni	Equal or Unequal
Tukey	Equal
Fisher LSD	Unequal

Multiple Comparison Procedures (Isolating Differences)

- Instead of hypothesis tests, we usually construct multiple confidence intervals for the difference between population means. The procedure is:
  - Recall: If a CI for $\mu_1 - \mu_2$ contains 0, there is no evidence to suggest that the population means are different.
  - If the CI does not include 0, there is evidence to suggest that the two population means are different.
  - We want a $100(1 - \alpha)\%$ CI for all possible paired comparisons (with overall probability α).
- If we *fail to reject* H_0 in a one-way ANOVA, there is no evidence to suggest any difference among population means. **The statistical analysis stops there.**
- However, if we *reject* H_0 in a one-way ANOVA, there is evidence to suggest an overall difference among population means. Next step: Try to isolate the difference(s).
- Find the pair(s) of means contributing to the overall significant difference. Use a **multiple comparison** procedure.

Bonferroni Confidence Intervals

- The general form of a Bonferroni CI is similar to a CI for the difference between two means based on a *t* distribution.
- Uses a pooled estimate of the common variance: MSE from ANOVA table.
- A *t* critical value is used to achieve a simultaneous, or family-wise, confidence level of $100(1 - \alpha)\%$.

Bonferroni Multiple Comparison Procedure:

In a one-way analysis of variance, suppose there are k groups, there are $n = n_1 + n_2 + \cdots + n_k$ total observations, and H_0 is rejected.

1. There are $c = k(k - 1)/2$ pairs of population means to compare.
2. The c simultaneous $100(1-\alpha)\%$ **Bonferroni confidence intervals** have the following values as endpoints:

$$\left(\overline{x}_{i.} - \overline{x}_{j.}\right) \pm t_{\alpha/(2c),n-k} \cdot \sqrt{\text{MSE}} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad \text{for all } i \neq j$$

Example: Small-Scale Farming

In Canada, most of the country's farms are small-scale, family-owned and -operated businesses. Only a small portion of the farms in Canada earn more than $1 million in annual revenue. Independent random samples of farms were obtained from four Canadian provinces, with 10 observations in each group. The size (in acres) was recorded for each farm. The resulting sample means and the ANOVA table are shown below:

Group number	Province (factor)	Sample mean
1	Prince Edward Island	402.3
2	New Brunswick	421.1
3	Quebec	326.1
4	British Columbia	314.3

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean square**	***F***	***p* value**
Factor	86185.9	3	28728.6	6.26	0.0016
Error	165086.0	36	4585.7		
total	251271.9	39			

Example: Small-Scale Farming

- The ANOVA test is significant at the p = 0.0016 level.
- There is evidence to suggest that at least one pair of population means is different (an overall difference).
- Construct the Bonferroni 95% confidence intervals and use them to isolate the pair(s) of means contributing to this overall experiment difference.
- The number of pairwise comparisons needed is:

$$c = \frac{(4)(3)}{2} = 6$$

$$95\% = 100(1-\alpha)\% \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2c} = \frac{0.05}{2(6)} = 0.0042$$

- This right-tail probability is not specified in Table 5 in the Appendix. However, we can use technology, with $n - k = 40 - 4 = 36$ degrees of freedom (df of Error from ANOVA table), to find

$t_{\alpha/(2c),\,36} = t_{0.0042,\,36} = -\ 2.7888.$

qt(0.0042,36)

- 2.788823

Example: Small-Scale Farming

The Bonferroni confidence interval for the difference $\mu_1 - \mu_2$ is:

$$(\bar{x}_{1.} - \bar{x}_{2.}) \pm t_{0.0042,36} \cdot \sqrt{\text{MSE}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(402.3 - 421.1) \pm (2.7888)\left(\sqrt{4585.7}\right)\sqrt{\frac{1}{10} + \frac{1}{10}} = (-103.3,\ 65.7)$$

The remaining five confidence intervals are found in the same manner. Each Bonferroni confidence interval is shown in the following table:

The initial ANOVA test indicates that there is an overall difference among the four population means. The simultaneous 95% Bonferroni confidence intervals suggest that this overall difference is due to a difference between μ_1 and μ_4, μ_2 and μ_3, and μ_2 and μ_4.

Difference	**Bonferroni confidence interval**	**Significantly different**
$\mu_1 - \mu_2$	(−103.3, 65.7)	No
$\mu_1 - \mu_3$	(−8.3, 160.7)	No
$\mu_1 - \mu_4$	(3.5, 172.5)	Yes
$\mu_2 - \mu_3$	(10.5, 179.5)	Yes
$\mu_2 - \mu_4$	(22.3, 191.3)	Yes
$\mu_3 - \mu_4$	(−72.7, 96.3)	No

Example: Small-Scale Farming

Visualizing the Comparison:

- There is another common, compact, graphical method for summarizing the results of a multiple comparison procedure.
- Write the sample means in order from smallest to largest.
- Use the results from a multiple comparison procedure to draw a horizontal line under the groups of means that are not significantly different.

$$\begin{array}{cccc} \bar{x}_{4.} & \bar{x}_{3.} & \bar{x}_{1.} & \bar{x}_{2.} \\ 314.3 & 326.1 & 402.3 & 421.1 \end{array}$$

$$\begin{array}{cccc} \bar{x}_{4.} & \bar{x}_{3.} & \bar{x}_{1.} & \bar{x}_{2.} \\ 314.3 & 326.1 & \multicolumn{2}{c}{\underline{402.3 \quad 421.1}} \end{array}$$

$$\begin{array}{cccc} \bar{x}_{4.} & \bar{x}_{3.} & \bar{x}_{1.} & \bar{x}_{2.} \\ \multicolumn{2}{c}{\underline{314.3 \quad 326.1}} & \multicolumn{2}{c}{\underline{402.3 \quad 421.1}} \end{array}$$

$$\begin{array}{cccc} \bar{x}_{4.} & \bar{x}_{3.} & \bar{x}_{1.} & \bar{x}_{2.} \\ \multicolumn{2}{c}{\underline{314.3 \quad 326.1}} & \multicolumn{2}{c}{\underline{402.3 \quad 421.1}} \end{array}$$

Those pairs of means not connected by a horizontal line are significantly different.

Example: Small-Scale Farming

An R plot of the confidence intervals for the difference between sample means, another way to visualize which pair(s) of mean(s) is(are) different.