CHAPTER 13 – TWO-WAY ANALYSIS OF VARIANCE

Two-way ANOVA has many of the same ideas as one-way ANOVA, with the main difference being the inclusion of another factor (or explanatory variable) in our model.

In the two-way ANOVA model, there are **two factors**, each with its own number of levels. When we are interested in the effects of two factors, it is much more advantageous to perform a two-way analysis of variance, as opposed to two separate one-way ANOVAs.

There are three main **advantages** of two-way ANOVA:

- It is more efficient to study two factors simultaneously rather than separately.
- We can reduce the residual variation in a model by including a second factor thought to influence the response.
- We can investigate interactions between factors.

The **interaction** between two variables is usually the most interesting feature of a two-way analysis of variance. When two factors interact, the effect on the response variable of one explanatory variable depends on the specific value or level of the other explanatory variable.

For example, the statement "being overweight caused greater increases in blood pressure for men than for women" is a statement describing interaction. The effect of weight (factor #1, categorical – overweight or not overweight) on blood pressure (response) depends on gender (factor #2, categorical – male or female).

The term **main effect** is used to describe the overall effect of a single explanatory variable. For our previous example, there would be two main effects: the effect of weight on blood pressure and the effect of gender on blood pressure.

The presence of a main effect might not necessarily be useful when an interaction effect exists. For example, it might be sensible to report the effect of being overweight on blood pressure without reporting that there is a difference in the effect of being overweight on blood pressure for men and women.

There are many types of interactions which can often be seen from patterns in graphs. We will study some types in the following examples.

Example #1(a)

The Bureau of Labor Statistics collects data on earnings of workers in the US classified according to various characteristics. Here are the mean weekly earnings in dollars of men and women in two age groups who were working full-time in the first quarter of 1997:

Age	Women	Men	Mean
16-19	239	264	251.5
20-24	302	333	317.5
Mean	270.5	298.5	284.5

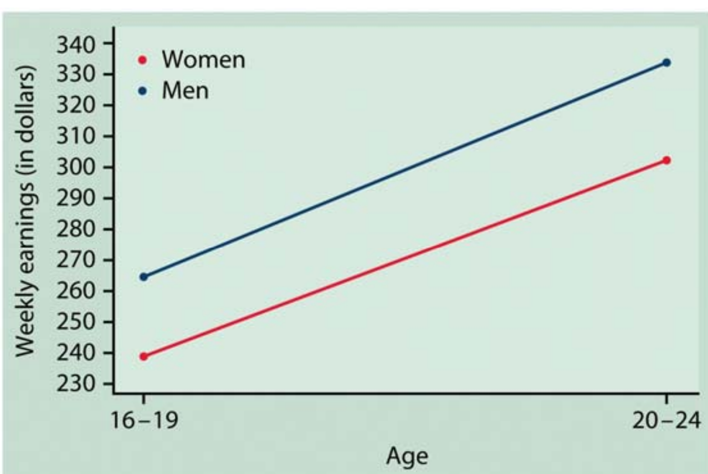

Figure 13.1

Figure 13.1 is a plot of the group means. It clearly shows an existence of main effects.

From the plot we can see that on average men earn more than women indicating an effect of gender. We can also see an effect of age since the older groups makes on average more than the younger age group.

What about the interaction between gender and age?

An **interaction** is present when the main effects cannot provide a **complete description** of the data.

In our graph, the only two effects that are present are an effect due to age and an effect due to gender. We know that men make, on average, more than women, but there is no dependence on age to determine the amount that men make more than women. Therefore, since there is no dependence between gender and age, there is no interaction effect.

Parallel lines in our plot usually imply little or **no interaction**.

Example #1 (b)

The survey described in 1(a) also gave weekly earnings for groups of older workers. Here is the complete table:

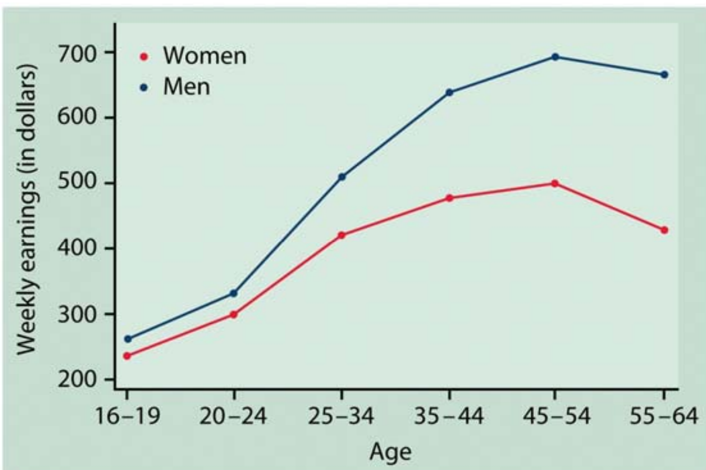

Figure 13.2

Age	Women	Men	Mean
16-19	239	264	251.5
20-24	302	333	317.5
25-34	422	514	468.0
35-44	475	639	557.0
45-54	496	692	594.0
55-64	426	660	543.0
Mean	393	517	455.0

Figure 13.2 is a plot of the group means. **It demonstrates both main effects and an interaction between gender and age.**

In this case, the main effects do not provide a complete description of the data since now the gender difference in earnings depends on which age group we examine. As the age of the workers increases, the earnings increase (except in the last age group). We also see a main effect of gender on earnings.

However, our interaction shows that the amount by which the men earn over the women depends on the age group.

Therefore, there is a very strong interaction present and our main effects don't provide us with enough information to get an accurate idea of the relationship between age and earnings as well as gender and earnings.

Example #2

A study of the energy expenditure of farmers in Burkina Faso collected data during the wet and dry seasons. The farmers grow millet during the wet season. In the dry season there is very little activity because the ground is too hard to grow crops. The **mean energy expended (in calories)** by men and women in Burkina Faso during the wet and dry seasons is given in the following table:

Season	Men	Women	**Mean**
Dry	2310	2320	**2315**
Wet	3460	2890	**3175**
Mean	**2885**	**2605**	**2745**

During the dry season both men and women use about the same amount of energy. However, during the wet season, both genders burn more calories. Therefore, there is an effect of season on the number of calories burned.

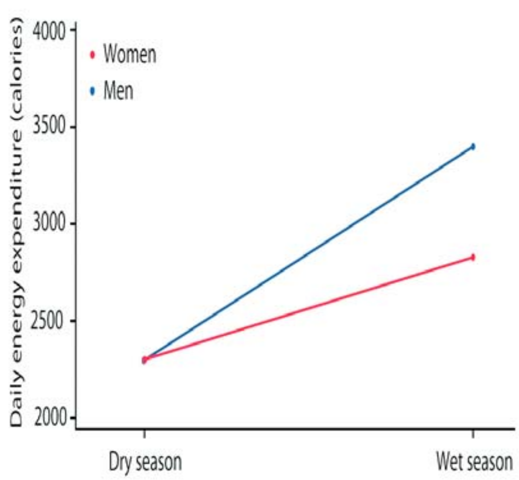

Figure 13.3

During the wet season, since it is the custom for men to do most of the field work, they use more energy on average than women. Therefore, there is an effect of gender on the number of calories burned, but **only** in the wet season.

So there is an interaction between gender and season, as well as main effects of both factors.

THE TWO-WAY ANOVA MODEL

When discussing two-way ANOVA models, we will use the labels A and B to represent our two factors. In examples, we will use the factor names so that we can easily see the meaning of each variable.

We use I to represent the number of levels of factor A and J to represent the number of levels of factor B. Therefore, we call the general two-way problem an I x J ANOVA, since every level of A is combined with every level of B, so that I x J groups are compared.

The sample size for level i of factor A and level j of factor B is n_{ij}. The total number of observations is $N = \Sigma n_{ij}$.

Example #3

A company wants to compare three different training programs for its new employees. Each of these programs takes 8 hours to complete. The training can be given for 8 hours on one day or for 4 hours on two consecutive days. The next 120 employees that the company hires will be the subjects for this study. After the training is completed, the employees are asked to evaluate the effectiveness of the program on a 7-point scale. Identify the response and explanatory variables, state the number of levels for each factor (I and J) and the total number of observations (N).

Solution:

Explanatory variables:
1) Training program (I = 3 levels)
2) The manner in which the employee does the training (J = 2 levels: 8 hrs f or one day, 4 hrs on two consecutive days).

Response variable: the 7-point scaled effectiveness score

Total number of observations N = 120

For our two-way ANOVA, we assume to have independent SRSs of size n_{ij} from each of I x J normal populations. The population means μ_{ij} may differ, but all populations must have the same standard deviation σ.

Let x_{ijk} be the k^{th} observation from the population having factor A at level i and factor B at level j. Since our observations differ from their mean by a value of ε_{ijk} ($x_{ijk} - \mu_{ij} = \varepsilon_{ijk}$), we can write our **two-way ANOVA model** in the form: $\mathbf{x_{ijk} = \mu_{ij} + \varepsilon_{ijk}}$.

In this case, $i = 1,2, \ldots, I$, $j = 1,2,\ldots, J$, $k = 1,2,\ldots,n_{ij}$ and the deviations ε_{ijk} are normally distributed with mean 0 and standard deviation σ.

As in the previous chapters, we can describe our model by **DATA = FIT + RESIDUAL**. In this case, the fit part of our model is the means μ_{ij}, (since we use each mean to describe each population) and the residual part is the deviations ε_{ijk} of the individual observations from their group means.

Our population means μ_{ij} and the population standard deviations are all unknown. Therefore, we pick simple random samples (SRSs) to learn about the relationship between our factors and the response in our samples and then use that to estimate the relationship in our population.

PARAMETER ESTIMATES

The population mean for the group with level i of factor A and level j of factor B is represented by μ_{ij}. It can be estimated by

$$\bar{x}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} .$$

Since every σ (population standard deviation) is assumed to be the same in two-way ANOVA, we can pool all of our estimates (sample standard deviations) to give one estimate of the population standard deviation.

$$s_p^2 = \frac{\sum (n_{ij} - 1) s_{ij}^2}{\sum (n_{ij} - 1)} \qquad \text{and} \qquad s_p = \sqrt{s_p^2}$$

Just like in one-way ANOVA, $\mathbf{s_p^2 = MSE}$, and so the numerator of s_p^2 is equal to SSE and the denominator is equal to DFE (since we have $\sum (n_{ij} - 1) =$ N - *IJ*, so we have N observations being compared to *IJ* sample means).

Example #4

Students in a statistics class gave information on their height and weight and also reported their perception of their weight (about right, underweight, overweight). The body mass index, BMI, was calculated for each combination of student gender and perception of weight category (BMI = Weight/Height). Calculate each sample mean based on the data.

	Perception of Weight		
	Underweight	**About Right**	**Overweight**
Female	20.2 20.6 18.1 21.4	22.5 23.1 20.9	24.5 26.0 22.7 24.6
Male	23.8 20.5	23.4 26.5 22.2 24.0	28.1 26.4 27.7

Solution:

Let factor A be the perception of weight and let factor B be gender. Within factor A, let level 1 be underweight, level 2 be about right, and level 3 be overweight. Within factor B (gender), let level 1 be female and let level 2 be male. So μ_{11} corresponds to the population mean BMI of underweight females, μ_{12} corresponds to the population mean BMI of underweight males, and so on. So to estimate these population values, we use x_{ij}(bar) to estimate μ_{ij} for i=1,2,3 and j=1,2.

Now,
x_{11}(bar) = (20.2 + 20.6 + 18.1 + 21.4) / 4 = 20.075
x_{12}(bar) = (23.8 + 20.5) / 2 = 22.15
x_{21}(bar) = (22.5 + 23.1 + 20.9) / 3 = 21.167
x_{22}(bar) = (23.4 + 26.5 + 22.2 + 24.0) / 4 = 24.025
x_{31}(bar) = (24.5 + 26 + 22.7 + 24.6) / 4 = 24.45
x_{32}(bar) = (28.1 + 26.4 + 27.7) / 3 = 27.4

Note: (bar) just means that the x has a bar over it (indicating a mean).

TWO-WAY ANOVA TABLE

In two-way ANOVA, there are three F statistics that are calculated and used in significance tests: two that test for the main effects and one that tests for an interaction.

Therefore, the sum of squares for our FIT (SSM) is made up of three parts:

SSA - sum of squares for factor A
SSB - sum of squares for factor B
SSAB - sum of squares for the interaction between factor A and factor B

Our total sum of squares and total degrees of freedom are still the sum of the sources of variation and degrees of freedom in our model.

SST = SSA + SSB + SSAB + SSE
DFT = DFA + DFB + DFAB + DFE

When our sample sizes are different, do not be alarmed if our sums of squares do not add up to our given SST. When our sample sizes are different, this can cause sums of squares which don't add.

Since we have I levels of factor A, DFA = $I - 1$.
Since we have J levels of factor B, DFB = $J - 1$.
Since SSM = SSA + SSB + SSAB and DFM = $IJ - 1$,
DFAB = $(IJ - 1) - (I - 1) - (J - 1) = (I - 1)(J - 1)$.
DFE = N – IJ (we have N observations and IJ sample means)
DFT = N – 1

TWO-WAY ANOVA TABLE

Source	Degrees of Freedom	Sums of Squares	Mean Square	F
A	$I-1$	SSA	SSA/DFA	MSA/MSE
B	$J-1$	SSB	SSB/DFB	MSB/MSE
AB	$(I-1)(J-1)$	SSAB	SSAB/DFAB	MSAB/MSE
Error	$N-IJ$	SSE	SSE/DFE	
Total	N - 1	SST		

Remember that when we do a two-way ANOVA, we need to assume that our **data is normally distributed** and that the **population standard deviations are equal.**

Therefore, we must make sure that **twice our smallest sample standard deviation is larger than our largest standard deviation**.

HYPOTHESES FOR TWO-WAY ANOVA

To test the main effect of A:

H_0: No main effect of A H_a: There exists a main effect of factor A

$F = \frac{MSA}{MSE}$ Compare to $F(I\text{-}1, N\text{-}IJ)$

To test the main effect of B:

H_0: No main effect of B H_a: There exists a main effect of factor B

$F = \frac{MSB}{MSE}$ Compare to $F(J\text{-}1, N\text{-}IJ)$

To test the interaction of A and B:

H_0: No interaction H_a: There exists an interaction between factor A and factor B

$F = \frac{MSAB}{MSE}$ Compare to $F((I\text{-}1)(J\text{-}1), N\text{-}IJ)$

Example #5

When a restaurant server writes a friendly note or draws a "happy face" on your restaurant check, is this just a friendly act or is there a financial incentive? Psychologists conducted a randomized experiment to investigate whether drawing a happy face on the back of a restaurant bill increased the average tip given to the server. One female server and one male server in a Philadelphia restaurant either did or did not draw a happy face on checks during the experiment. In all they drew happy faces on 45 checks and did not draw happy faces on 44 checks. The sequence of drawing the happy faces or not was random.

a) Identify the response and explanatory variables, state the number of levels for each factor (*I* and *J*) and the total number of observations (N).

Solution:

Response variable: tip
Explanatory variables: gender of server (2 levels), message (2 levels, yes or no)

Total number of observations: N = 89

b) Complete the following two-way ANOVA table and then **perform the appropriate F tests for main effects and interaction** and state your **conclusions**.

Try to fill the table in on your own (answer is below):

Source	DF	SS	MS	F
Message		14.7		
Gender			2602.0	
Interaction		438.7		
Error			109.8	
Total		12407.9		

Solution:

Source	DF	SS	MS	F
Message	1	14.7	14.7	0.134
Gender	1	2602.0	2602.0	23.7
Interaction	1	438.7	438.7	4.0
Error	85	9333.0	109.8	
Total	88	12407.9	---	

H_0: No main effect of message
H_a: A main effect of message exists

Test Statistic: $F_0 = 14.7/109.8 = 0.134$ with numerator degrees of freedom 1 and denominator degrees of freedom 85.

Decision Rule: This test statistic corresponds to a p-value of 0.7152. We do not have any evidence to reject the null hypothesis that there is main effect of message on the average amount a server gets tipped.

H_0: No main effect of gender
H_a: A main effect of gender exists

Test Statistic: $F_0 = 2602/109.8 = 23.7$ with numerator df = 1 and denominator df = 85.

Decision Rule: This test statistic corresponds to a p-value of less than .0001. We have very strong evidence that a main effect of gender does exist.

H_0: No interaction effect between gender and message
H_a: An interaction effect between gender and message exists.

Test Statistic: $F_0 = 438.7/109.8 = 4.0$ with numerator df = 1 and denominator df = 85.

Decision Rule: This test statistic corresponds to a p-value of .0487. We have evidence to reject the null hypothesis of no interaction at the $\alpha = 0.05$ level. We have reason to believe that there is an interaction effect.
