Lesson 21:
Chapter 13

Caleb Moxley

BSC Mathematics

18 November 15

§13.1 The Two-Way ANOVA Model

We've looked at a one-way ANOVA which compared the means of populations separated by a single variable. Our next test will separate populations based a two variables!

§13.1 The Two-Way ANOVA Model

We've looked at a one-way ANOVA which compared the means of populations separated by a single variable. Our next test will separate populations based a two variables! It's a two-way ANOVA.

§13.1 The Two-Way ANOVA Model

We've looked at a one-way ANOVA which compared the means of populations separated by a single variable. Our next test will separate populations based a two variables! It's a two-way ANOVA. But it doesn't just test for effects of two variables separately. It also tests for effects of **interactions** between variables.

§13.1 The Two-Way ANOVA Model

We've looked at a one-way ANOVA which compared the means of populations separated by a single variable. Our next test will separate populations based a two variables! It's a two-way ANOVA. But it doesn't just test for effects of two variables separately. It also tests for effects of **interactions** between variables. You can think of it conceptually as a combination of a two-way table test and two one-way ANOVA tests.

§13.1 The Two-Way ANOVA Model

Being able to use a two-way ANOVA test has some advantages.

§13.1 The Two-Way ANOVA Model

Being able to use a two-way ANOVA test has some advantages.

1. It's efficient to study two factors simultaneously.

§13.1 The Two-Way ANOVA Model

Being able to use a two-way ANOVA test has some advantages.

1. It's efficient to study two factors simultaneously.
2. Residual variance can be reduced in a model by including a second factor thought to influence the response.

§13.1 The Two-Way ANOVA Model

Being able to use a two-way ANOVA test has some advantages.

1. It's efficient to study two factors simultaneously.
2. Residual variance can be reduced in a model by including a second factor thought to influence the response.
3. We can investigate interactions between variables.

§13.1 The Two-Way ANOVA Model

Being able to use a two-way ANOVA test has some advantages.

1. It's efficient to study two factors simultaneously.
2. Residual variance can be reduced in a model by including a second factor thought to influence the response.
3. We can investigate interactions between variables.

Note: Each sample from the two-factor populations must be of the same size! It's possible to conduct the test with samples of different size, but it causes a slight problem in the requirement of how similar the standard deviations of these samples must be. It's safest to conduct the test when the sample sizes are all the same.

§13.1 The Two-Way ANOVA Model

Definition (assumptions for two-way ANOVA)

A SRS of size n_{ij} is taken from teach $I \times J$ normal population. The means μ_{ij} may differ but the standard deviations σ are all the same.

Let x_{ijk} be the k^{th} observation from the population having Factor A at level i and Factor B at level j. The statistical model is

$$x_{ijk} = \mu ij + \epsilon_{ijk}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, n_{ij}$, where the deviations ϵ_{ijk} are from a $N(0, \sigma)$ distribution.

§13.1 The Two-Way ANOVA Model

To conduct the test, you'll need the means of all the cells of data and their variances. You compute the pooled variance by taking the weighted average of all the cell variances where the weight for each cell is one fewer than the number of data points in the cell.

§13.1 The Two-Way ANOVA Model

To conduct the test, you'll need the means of all the cells of data and their variances. You compute the pooled variance by taking the weighted average of all the cell variances where the weight for each cell is one fewer than the number of data points in the cell. It looks like this:

To conduct the test, you'll need the means of all the cells of data and their variances. You compute the pooled variance by taking the weighted average of all the cell variances where the weight for each cell is one fewer than the number of data points in the cell. It looks like this:

$$s_p^2 = \frac{\sum (n_{ij} - 1)s_{ij}^2}{\sum (n_{ij} - 1)}.$$

To conduct the test, you'll need the means of all the cells of data and their variances. You compute the pooled variance by taking the weighted average of all the cell variances where the weight for each cell is one fewer than the number of data points in the cell. It looks like this:

$$s_p^2 = \frac{\sum(n_{ij} - 1)s_{ij}^2}{\sum(n_{ij} - 1)}.$$

Also, MSE $= s_p^2$, where the numerator is the SSE and the denominator is the DFE.

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

$$\text{SSM}=\text{SSA}+\text{SSB}+\text{SSAB}.$$

We compute the SSA and SSB as we would for a regular one-way ANOVA, ignoring Factor B for the SSA and Factor A for the SSB.

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

$$\text{SSM} = \text{SSA} + \text{SSB} + \text{SSAB}.$$

We compute the SSA and SSB as we would for a regular one-way ANOVA, ignoring Factor B for the SSA and Factor A for the SSB. Finally, We can calculate the SST as we would with a one-way ANOVA and find the SSAB based on the formula

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE}$$

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

$$\text{SSM}=\text{SSA}+\text{SSB}+\text{SSAB}.$$

We compute the SSA and SSB as we would for a regular one-way ANOVA, ignoring Factor B for the SSA and Factor A for the SSB.

§13.1 The Two-Way ANOVA Model

So we've discussed the portion of the SST made up by the SSE. We can break down the SSM into three component parts now since we have two factors/variables and their interactions:

$$\text{SSM}=\text{SSA}+\text{SSB}+\text{SSAB}.$$

We compute the SSA and SSB as we would for a regular one-way ANOVA, ignoring Factor B for the SSA and Factor A for the SSB. Finally, We can calculate the SST as we would with a one-way ANOVA and find the SSAB based on the formula

$$\text{SST}=\text{SSA}+\text{SSB}+\text{SSAB}+\text{SSE}$$

§13.2 Inference for Two-Way ANOVA

We have following table as the summary of our two-way ANOVA test.

We have following table as the summary of our two-way ANOVA test.

Source	DF	SS	MS	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		

§13.2 Inference for Two-Way ANOVA

We have following table as the summary of our two-way ANOVA test.

Source	DF	SS	MS	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		

You can get the P-values using the F distribution, the degrees of freedom of the numerator and denominator of the F test statistic and a right-tail.

§13.2 Inference for Two-Way ANOVA

We have following table as the summary of our two-way ANOVA test.

Source	DF	SS	MS	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		

You can get the P-values using the F distribution, the degrees of freedom of the numerator and denominator of the F test statistic and a right-tail. These P-values correspond to the tests for the A factor, the B factor, and the AB interaction (from top to bottom).

§13.2 Inference for Two-Way ANOVA

Definition (significance tests in two-way ANOVA)

The F test statistics computed in the previous table are

$$F_A = \frac{\text{MSA}}{\text{MSE}}, \ \ F_B = \frac{\text{MSB}}{\text{MSE}}, \ \text{and} \ F_{AB} = \frac{\text{MSB}}{\text{MSE}}.$$

The first two are used to conduct one-way ANOVAs, and so their null and alternative hypotheses are the same as for one-way ANOVAs. The third is used to test the null hypothesis that there is no effect due to an interaction between the two factors against the alternative that there is an effect due to the interaction of the two factors.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

§13.2 Inference for Two-Way ANOVA

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, $\text{SSE} = (1)(8) + (1)(50) + (1)(72) + (1)(2) = 132$. And $\text{DFE} = 4$.

§13.2 Inference for Two-Way ANOVA

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, SSE $= (1)(8) + (1)(50) + (1)(72) + (1)(2) = 132$. And DFE $= 4$. Thus, MSE $= 33$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, $\text{SSE} = (1)(8) + (1)(50) + (1)(72) + (1)(2) = 132$. And $\text{DFE} = 4$. Thus, $\text{MSE} = 33$. Also,

$$\text{SSA} = (4)(62 - 60.75)^2 + (4)(59.5 - 60.75)^2 = 12.5.$$

And $\text{DFA} = 1$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, SSE $= (1)(8) + (1)(50) + (1)(72) + (1)(2) = 132$. And DFE $=$ 4. Thus, MSE $= 33$. Also,

$$\text{SSA} = (4)(62 - 60.75)^2 + (4)(59.5 - 60.75)^2 = 12.5.$$

And DFA $= 1$. Thus MSA $= 12.5$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, SSE $= (1)(8) + (1)(50) + (1)(72) + (1)(2) = 132$. And DFE $= 4$. Thus, MSE $= 33$. Also,

$$\text{SSA} = (4)(62 - 60.75)^2 + (4)(59.5 - 60.75)^2 = 12.5.$$

And DFA $= 1$. Thus MSA $= 12.5$. Also,

$$\text{SSB} = (4)(61.5 - 60.75)^2 + (4)(60 - 60.75)^2 = 4.5$$

DFB $= 1$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

Well, $\text{SSE} = (1)(8)+(1)(50)+(1)(72)+(1)(2) = 132$. And $\text{DFE} = 4$. Thus, $\text{MSE} = 33$. Also,

$$\text{SSA} = (4)(62 - 60.75)^2 + (4)(59.5 - 60.75)^2 = 12.5.$$

And $\text{DFA} = 1$. Thus $\text{MSA} = 12.5$. Also,

$$\text{SSB} = (4)(61.5 - 60.75)^2 + (4)(60 - 60.75)^2 = 4.5$$

$\text{DFB} = 1$. Thus $\text{MSB} = 4.5$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

And finally, we have

$$\text{SST} = (56-60.75)^2 + 3(60-60.75)^2 + \cdots + (53-60.75)^2 = 329.5$$

. And DFT $= 7$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

And finally, we have

$$\text{SST} = (56 - 60.75)^2 + 3(60 - 60.75)^2 + \cdots + (53 - 60.75)^2 = 329.5$$

. And DFT $= 7$. Thus, we have

$$SSAB = 329.5 - 12.5 - 4.5 - 132 = 180.5.$$

And DFAB $= 1$.

Example (age & gender's effects on heart rate)

Below are samples size two from of four populations.

	< 30	> 30
F	56, 60	60, 70
M	60, 72	55, 53

And finally, we have

$$\text{SST} = (56-60.75)^2 + 3(60-60.75)^2 + \cdots + (53-60.75)^2 = 329.5$$

. And DFT $= 7$. Thus, we have

$$SSAB = 329.5 - 12.5 - 4.5 - 132 = 180.5.$$

And DFAB $= 1$. Thus, MSAB $= 180.5$.

Example (age & gender's effects on heart rate)

Source	DF	SS	MS	F	P
A	1	12.5	12.5	0.37879	0.7306
B	1	4.5	4.5	0.13636	0.5715
AB	1	180.5	180.5	5.44697	0.0795
Error	4	132	33		
Total	7	329.5			

Example (age & gender's effects on heart rate)

Source	DF	SS	MS	F	P
A	1	12.5	12.5	0.37879	0.7306
B	1	4.5	4.5	0.13636	0.5715
AB	1	180.5	180.5	5.44697	0.0795
Error	4	132	33		
Total	7	329.5			

If we conducted all three hypothesis tests at a 10% significance, what would we conclude?

Example (age & gender's effects on heart rate)

Source	DF	SS	MS	F	P
A	1	12.5	12.5	0.37879	0.7306
B	1	4.5	4.5	0.13636	0.5715
AB	1	180.5	180.5	5.44697	0.0795
Error	4	132	33		
Total	7	329.5			

If we conducted all three hypothesis tests at a 10% significance, what would we conclude? Well, it looks like there would be an effect due to interaction but no effect due to the factors separately!

§13.2 Inference for Two-Way ANOVA

Example (age & gender's effects on heart rate)

Source	DF	SS	MS	F	P
A	1	12.5	12.5	0.37879	0.7306
B	1	4.5	4.5	0.13636	0.5715
AB	1	180.5	180.5	5.44697	0.0795
Error	4	132	33		
Total	7	329.5			

If we conducted all three hypothesis tests at a 10% significance, what would we conclude? Well, it looks like there would be an effect due to interaction but no effect due to the factors separately! This is because all the P-values are larger than 5%.

Example (age & gender's effects on heart rate)

Source	DF	SS	MS	F	P
A	1	12.5	12.5	0.37879	0.7306
B	1	4.5	4.5	0.13636	0.5715
AB	1	180.5	180.5	5.44697	0.0795
Error	4	132	33		
Total	7	329.5			

If we conducted all three hypothesis tests at a 10% significance, what would we conclude? Well, it looks like there would be an effect due to interaction but no effect due to the factors separately! This is because all the P-values are larger than 5%. Note: Remember the P-values are computed using a right-tailed F distribution with the degrees of freedom of the numerator and denominator selected appropriately.

Example (graduation & age's effect on credit score)

The two-way table for the data is given below in a three-column format.

Data Set 1

§13.2 Inference for Two-Way ANOVA

Example (graduation & age's effect on credit score)

The two-way table for the data is given below in a three-column format.

Data Set 1

Example (graduation & age's effect on credit score)

The two-way table for the data is given below in a three-column format.

Data Set 1

Performing a two-way ANOVA on this data using technology, we get

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Row	1	27300.833	27300.833	7.8384974	0.0099
Column	2	80645	40322.5	11.577222	0.0003
Interaction	2	2321.6667	1160.8333	0.33329346	0.7198
Error	24	83590	3482.9167		
Total	29	193857.5			

Example (graduation & age's effect on credit score)

Thus, there seems to be no effect due to interaction between age and graduation status but there does seem to be separate effects due to age and graduation status.