Lecture 13 - Two way ANOVA

Sta102

June 6, 2016

Colin Rundel & Mine Çetinkaya-Rundel

One-way ANOVA

Example - Alfalfa

Researchers were interested in the effect that acid has on the growth rate of alfalfa plants. They created three treatment groups in an experiment: low acid, high acid, and control. The alfalfa plants were grown in a Styrofoam cups arranged near a window and the height of the alfalfa plants was measured after five days of growth. The experiment consisted of 5 cups for each of the 3 treatments, for a total of 15 observations.

	High Acid	Low Acid	Control
	1.30	1.78	2.67
	1.15	1.25	2.25
	0.50	1.27	1.46
	0.30	0.55	1.66
	1.30	0.80	0.80
$\bar{y}_i$	0.910	1.130	1.768
n	5	5	5
	$\mu = 1.269$		

Alfalfa Hypotheses

We would like to establish if the acid treatments are affecting the alfalfa's growth. Since we have a numerical response and categorical explanatory variable (>2 levels) we will use an ANOVA.

What should our hypotheses be?

Alfalfa Hypotheses

We would like to establish if the acid treatments are affecting the alfalfa's growth. Since we have a numerical response and categorical explanatory variable (>2 levels) we will use an ANOVA.

What should our hypotheses be?

H_0: $\mu_H = \mu_L = \mu_C$

H_A: At least one pair of means differ

Treatment Effect

Last time we mentioned that it is possible to write down a model for each data point using the form

$$y_{ij} = \mu_i + \epsilon_{ij}$$

where $i \in \{H, L, C\}$ is the treatment and $j \in \{1, 2, 3, 4, 5\}$ is the index of the observation within that treatment.

Treatment Effect

Last time we mentioned that it is possible to write down a model for each data point using the form

$$y_{ij} = \mu_i + \epsilon_{ij}$$

where $i \in \{H, L, C\}$ is the treatment and $j \in \{1, 2, 3, 4, 5\}$ is the index of the observation within that treatment.

We can rewrite this in terms of the grand mean μ as follows

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

where $\tau_i = \mu_i - \mu$ is the treatment effect of treatment i.

Treatment Effect

Last time we mentioned that it is possible to write down a model for each data point using the form

$$y_{ij} = \mu_i + \epsilon_{ij}$$

where $i \in \{H, L, C\}$ is the treatment and $j \in \{1, 2, 3, 4, 5\}$ is the index of the observation within that treatment.

We can rewrite this in terms of the grand mean μ as follows

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

where $\tau_i = \mu_i - \mu$ is the treatment effect of treatment i.

Using treatment effect we can rewrite our null hypothesis

$$H_0\colon \mu_H = \mu_L = \mu_C = \mu \quad \Rightarrow \quad H_0\colon \tau_H = \tau_L = \tau_C = 0$$

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr($>$F)
Treatment					
Residuals					
Total					

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment					
Residuals					
Total					

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

$$SSG = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2$$

$$SSE = SST - SSG =$$

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment					
Residuals					
Total					

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

$$= (1.3 - 1.27)^2 + (1.15 - 1.27)^2 + \cdots + (0.80 - 1.27)^2 = 5.88$$

$$SSG = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$$

$$SSE = SST - SSG =$$

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment					
Residuals					
Total					

$$
\begin{aligned}
SST &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 \\
&= (1.3-1.27)^2 + (1.15-1.27)^2 + \cdots + (0.80-1.27)^2 = 5.88 \\
SSG &= \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 \\
&= 5\times(0.91-1.269)^2 + 5\times(1.13-1.269)^2 + 5\times(1.768-1.269)^2 = 1.99 \\
SSE &= SST - SSG =
\end{aligned}
$$

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment					
Residuals					
Total					

$$
\begin{aligned}
SST &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 \\
&= (1.3-1.27)^2 + (1.15-1.27)^2 + \cdots + (0.80-1.27)^2 = 5.88 \\
SSG &= \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 \\
&= 5\times(0.91-1.269)^2 + 5\times(1.13-1.269)^2 + 5\times(1.768-1.269)^2 = 1.99 \\
SSE &= SST - SSG = 3.893
\end{aligned}
$$

Alfalfa ANOVA Table - Sum Sq

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		*1.99*			
Residuals		*3.89*			
Total		*5.88*			

$$
\begin{aligned}
SST &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 \\
&= (1.3-1.27)^2+(1.15-1.27)^2+\cdots+(0.80-1.27)^2 = 5.88 \\
SSG &= \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 \\
&= 5\times(0.91-1.269)^2+5\times(1.13-1.269)^2+5\times(1.768-1.269)^2 = 1.99 \\
SSE &= SST - SSG = 3.893
\end{aligned}
$$

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		1.986			
Residuals		3.893			
Total		5.879			

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		1.986			
Residuals		3.893			
Total		5.879			

$$df_T = n - 1$$

$$df_G = k - 1$$

$$df_E = n - k$$

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		1.986			
Residuals		3.893			
Total		5.879			

$$df_T = n - 1 = 15 - 1 = 14$$

$$df_G = k - 1$$

$$df_E = n - k$$

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		1.986			
Residuals		3.893			
Total		5.879			

$$df_T = n - 1 = 15 - 1 = 14$$

$$df_G = k - 1 = 3 - 1 = 2$$

$$df_E = n - k$$

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment		1.986			
Residuals		3.893			
Total		5.879			

$$df_T = n - 1 = 15 - 1 = 14$$
$$df_G = k - 1 = 3 - 1 = 2$$
$$df_E = n - k = 15 - 3 = 12$$

Alfalfa ANOVA Table - DF

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	*2*	1.986			
Residuals	*12*	3.893			
Total	*14*	5.879			

$$df_T = n - 1 = 15 - 1 = 14$$
$$df_G = k - 1 = 3 - 1 = 2$$
$$df_E = n - k = 15 - 3 = 12$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

$$\begin{aligned} MSG &= SSG/df_G \\ MSE &= SSE/df_E \\ F &= MSG/MSE \\ \text{P-value} &= P(>F) \end{aligned}$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E$$

$$F = MSG/MSE$$

$$\text{P-value} = P(> F)$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E = 3.907/12 = 0.324$$

$$F = MSG/MSE$$

$$\text{P-value} = P(> F)$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E = 3.907/12 = 0.324$$

$$F = MSG/MSE = 0.993/0.326 = 3.061$$

$$\text{P-value} = P(> F)$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986			
Residuals	12	3.893			
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E = 3.907/12 = 0.324$$

$$F = MSG/MSE = 0.993/0.326 = 3.061$$

$$\text{P-value} = P(> F) = 0.0843$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986	*0.993*	*3.061*	*0.0843*
Residuals	12	3.893	*0.324*		
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E = 3.907/12 = 0.324$$

$$F = MSG/MSE = 0.993/0.326 = 3.061$$

$$\text{P-value} = P(> F) = 0.0843$$

Alfalfa ANOVA Table - Mean Sq, F, P-value

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986	*0.993*	*3.061*	*0.0843*
Residuals	12	3.893	*0.324*		
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = 0.993$$

$$MSE = SSE/df_E = 3.907/12 = 0.324$$

$$F = MSG/MSE = 0.993/0.326 = 3.061$$

$$\text{P-value} = P(> F) = 0.0843$$

Based on these results we fail to reject H_0, there is not sufficient evidence to suggest that at least one pair of mean growth values are significantly different.

Randomized Block Design

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Imagine we are interested in exploring whether increasing the dosage of a Statin will reduce the risk of a heart attack. We randomly sample patients already on a Statin and randomly assign them to either maintain their current dosage or increase their dosage by 20%.

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Imagine we are interested in exploring whether increasing the dosage of a Statin will reduce the risk of a heart attack. We randomly sample patients already on a Statin and randomly assign them to either maintain their current dosage or increase their dosage by 20%.

- Possible that some of the patients in this sample may have had a previous heart attack,

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Imagine we are interested in exploring whether increasing the dosage of a Statin will reduce the risk of a heart attack. We randomly sample patients already on a Statin and randomly assign them to either maintain their current dosage or increase their dosage by 20%.

- Possible that some of the patients in this sample may have had a previous heart attack,
- Significant risk factor for a future heart attack

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Imagine we are interested in exploring whether increasing the dosage of a Statin will reduce the risk of a heart attack. We randomly sample patients already on a Statin and randomly assign them to either maintain their current dosage or increase their dosage by 20%.

- Possible that some of the patients in this sample may have had a previous heart attack,
- Significant risk factor for a future heart attack
- Their presence may alter our outcome

Random Sampling / Assignment

Random sampling removes nuisance factors/variables (things that affect your outcome that you are not interested in).

Imagine we are interested in exploring whether increasing the dosage of a Statin will reduce the risk of a heart attack. We randomly sample patients already on a Statin and randomly assign them to either maintain their current dosage or increase their dosage by 20%.

- Possible that some of the patients in this sample may have had a previous heart attack,
- Significant risk factor for a future heart attack
- Their presence may alter our outcome
- Control for this effect by excluding them

Controls

Exclusion

- Works if the number of patients with a previous heart attack is low
- Can only exclude so many nuisance factors before we run out of available population
- Restricts generalizability

Controls

Exclusion

- Works if the number of patients with a previous heart attack is low
- Can only exclude so many nuisance factors before we run out of available population
- Restricts generalizability

Blocking

- Samples grouped into *homogeneous* blocks where the nuisance factor(s) are held constant
- Variation within the block should be less than the variation between blocks
- Randomized treatment assignment within each block

Controls

Exclusion

- Works if the number of patients with a previous heart attack is low
- Can only exclude so many nuisance factors before we run out of available population
- Restricts generalizability

Blocking

- Samples grouped into *homogeneous* blocks where the nuisance factor(s) are held constant
- Variation within the block should be less than the variation between blocks
- Randomized treatment assignment within each block

“Block what you can; randomize what you cannot.”

Blocking and Alfalfa

In the description for the alfalfa acid rain experiment we are told that the Styrofoam cups are arranged next to a window.

What are some potential nuisance factors that could have affected the experiment's outcome?

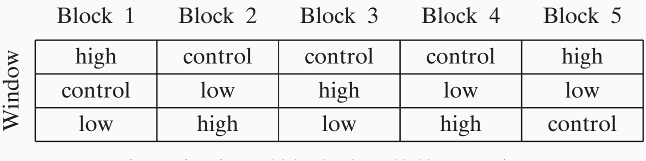
Do any of them lend themselves to blocking?

Blocking and Alfalfa

In the description for the alfalfa acid rain experiment we are told that the Styrofoam cups are arranged next to a window.

What are some potential nuisance factors that could have affected the experiment’s outcome?

Do any of them lend themselves to blocking?



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

Blocked Alfalfa

We will consider the simplest case of randomized block design where each block contains only one observation of each treatment.

	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
Trmt mean	0.910	1.130	1.768	
n	5	5	5	
		$\mu = 1.269$		

Block Data Model

When employing blocks we can think of each data point as

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

where

τ_i is the treatment effect for treatment i

β_j is the block effect of block j

ϵ_{ijk} is the residual of observation k in block j with treatment i

this is very similar to the one-way anova model we saw previous with the addition of the β_js.

Randomized Block ANOVA Table

With the introduction of the blocks there are now two hypotheses we would like to evaluate:

$$\begin{aligned} &H_0(\text{treatment}) : \tau_H = \tau_L = \tau_C = 0 && H_A : \text{At least one } \tau_i \neq 0 \\ &H_0(\text{block}) : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 && H_A : \text{At least one } \beta_i \neq 0 \end{aligned}$$

In order to test these hypotheses we will extend the ANOVA table we have been using.

Randomized Block ANOVA Table

With the introduction of the blocks there are now two hypotheses we would like to evaluate:

$$H_0(\text{treatment}) : \tau_H = \tau_L = \tau_C = 0 \qquad H_A : \text{At least one } \tau_i \neq 0$$

$$H_0(\text{block}) : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \qquad H_A : \text{At least one } \beta_i \neq 0$$

In order to test these hypotheses we will extend the ANOVA table we have been using.

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	df_G	SSG	MSG	F_G	
Block	df_B	SSB	MSB	F_B	
Error	df_E	SSE	MSE		
Total	df_T	SST			

Randomized Block ANOVA Table

	df	Sum Sq	Mean Sq	F value
Group	$k-1$	$\sum_{i=1}^{k} n_i(\bar{y}_{i\cdot} - \bar{y})^2$	SSG/df_G	MSG/MSE
Block	$b-1$	$\sum_{j=1}^{b} m_j(\bar{y}_{\cdot j} - \bar{y})^2$	SSB/df_B	MSB/MSE
Error	$n-k-b+1$	$SST - SSG - SSB$	SSE/df_E	
Total	$n-1$	$\sum_i \sum_j \sum_k (y_{ijk} - \bar{y})^2$		

- n - # observations
- k - # groups
- b - # blocks
- n_i - # observations in group i
- m_j - # observations in block j

- $\bar{y}$ - grand mean
- $\bar{y}_{i\cdot}$ - group mean for group i
- $\bar{y}_{\cdot j}$ - block mean for block j

Randomized Block ANOVA Table - Alfalfa

We already know some of the values from our previous one-way ANOVA, and it is easy to find the other *df* values.

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	MSG/MSE
Block	4	$\sum_{j=1}^{b} m_j(\bar{y}_{\bullet j} - \bar{y})^2$	SSB/df_B	MSB/MSE
Error	8	$SST - SSG - SSB$	SSE/df_E	
Total	14	5.879		

Sum of Squares Blocks

$$SSB = \sum_{j=1}^{b} m_j (\bar{y}_{\cdot j} - \bar{y})^2$$

Sum of Squares Blocks

$$SSB = \sum_{j=1}^{b} m_j(\bar{y}_{\cdot j} - \bar{y})^2$$

	High Acid	Low Acid	Control	Block Mean
Block 1	1.30	1.78	2.67	$1.917 = \bar{y}_{\cdot 1}$
Block 2	1.15	1.25	2.25	$1.550 = \bar{y}_{\cdot 2}$
Block 3	0.50	1.27	1.46	$1.077 = \bar{y}_{\cdot 3}$
Block 4	0.30	0.55	1.66	$0.837 = \bar{y}_{\cdot 4}$
Block 5	1.30	0.80	0.80	$0.967 = \bar{y}_{\cdot 5}$
Trmt mean	0.910	1.130	1.768	
n	5	5	5	
		$\bar{y} = 1.269$		

Sum of Squares Blocks

$$SSB = \sum_{j=1}^{b} m_j(\bar{y}_{\cdot j} - \bar{y})^2$$

	High Acid	Low Acid	Control	Block Mean
Block 1	1.30	1.78	2.67	$1.917 = \bar{y}_{\cdot 1}$
Block 2	1.15	1.25	2.25	$1.550 = \bar{y}_{\cdot 2}$
Block 3	0.50	1.27	1.46	$1.077 = \bar{y}_{\cdot 3}$
Block 4	0.30	0.55	1.66	$0.837 = \bar{y}_{\cdot 4}$
Block 5	1.30	0.80	0.80	$0.967 = \bar{y}_{\cdot 5}$
Trmt mean	0.910	1.130	1.768	
n	5	5	5	
		$\bar{y} = 1.269$		

$$\begin{aligned} SSB = {} & 3 \times (1.917 - 1.269)^2 + 3 \times (1.550 - 1.269)^2 \\ & + 3 \times (1.077 - 1.269)^2 + 3 \times (0.837 - 1.269)^2 \\ & + 3 \times (0.967 - 1.269)^2 \\ = {} & 1.260 + 0.237 + 0.111 + 0.560 + 0.274 = 2.441 \end{aligned}$$

Completing the table

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	MSG/MSE
Block	4	2.441	SSB/df_B	MSB/MSE
Error	8	$SST - SSG - SSB$	SSE/df_E	
Total	14	5.879		

Completing the table

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	MSG/MSE
Block	4	*2.441*	SSB/df_B	MSB/MSE
Error	8	*1.452*	SSE/df_E	
Total	14	5.879		

Completing the table

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	MSG/MSE
Block	4	*2.441*	*0.6103*	MSB/MSE
Error	8	*1.452*	*0.1815*	
Total	14	5.879		

Completing the table

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	*5.471*
Block	4	*2.441*	*0.6103*	*3.362*
Error	8	*1.452*	*0.1815*	
Total	14	5.879		

Calculating P-values

The two F values that we have calculated can be used to evaluate the two hypotheses we started with.

- Treatment effect

$$H_0 : \ \tau_H = \tau_L = \tau_G, \ \ H_A : \ \text{At least one pair of treatment effects differ}$$

- Block effect

$$H_0 : \ \beta_1 = \beta_2 = \ldots = \beta_5, \ \ H_A : \ \text{At least one pair of block effects differ}$$

To calculate the P-value for each hypothesis we use F_G and F_B respectively to find $P(> F)$ for an F distribution with the appropriate degrees of freedom.

Treatment Effect

We have calculated that $F_G = 5.471$, to find the P-value we need to the probability of observing a value equal to or larger than this from an F distribution with *2* and *8* degrees of freedom.

Using R we find that

```
pf(5.471, df1=2, df2=8, lower.tail=FALSE)

## [1] 0.03181681
```

Therefore, $P(> F_G) = 0.0318$, which leads us to reject H_0 - there is sufficient evidence to suggest that at least one pair of treatment effects differ.

Block Effect

Similarly, we have $F_B = 3.362$ and to find the P-value we need to the probability of observing a value equal to or larger than this from an F distribution with 4 and 8 degrees of freedom.

Using R we find that

```
pf(3.362, df1=4, df2=8, lower.tail=FALSE)

## [1] 0.06790077
```

Therefore, $P(> F_B) = 0.0679$, which leads us to fail to reject H_0 - there is not sufficient evidence to suggest that at least one pair of block effects differ.

How did blocking change our result?

- One-way ANOVA

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986	0.993	3.061	0.0843
Residuals	12	3.893	0.324		
Total	14	5.879			

- Randomized Block ANOVA

	df	Sum Sq	Mean Sq	F value	P(>F)
Group	2	1.986	0.993	5.471	0.0318
Block	4	2.441	0.6103	3.362	0.0679
Error	8	1.452	0.1815		
Total	14	5.879			

How did blocking change our result?

- One-way ANOVA

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	1.986	0.993	3.061	0.0843
Residuals	12	3.893	0.324		
Total	14	5.879			

- Randomized Block ANOVA

	df	Sum Sq	Mean Sq	F value	P(>F)
Group	2	1.986	0.993	5.471	0.0318
Block	4	2.441	0.6103	3.362	0.0679
Error	8	1.452	0.1815		
Total	14	5.879			

Blocking decreases df_E, which increases MSE (*bad*).
Blocking also decreases SSE, which decreases MSE (*good*).

Two-way ANOVA

From Randomized Block to Two-way ANOVA

All of the approaches we have just learned to handle blocking will also apply in the case where we would like to assess the effect of a second factor / predictor on our outcome variable.

Instead of examining treatment and block effects we instead examine two treatment effects.

None of the procedures or calculations change, only what we call things.

Two-way ANOVA Model

When employing two-way ANOVA we can think of each data point as

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

where

τ_i	is the effect of level *i* of treatment 1
β_j	is the effect of level *j* of treatment 2
ϵ_{ijk}	is the residual of observation *k* in with treatment 1 level *i* and treatment 2 level *j*

this is exactly the same as the randomized block ANOVA model except the β_js now refer to the effect of the second factor instead of a block effect.

Example - Spruce Moths

A scientist is interested in efficacy of various lure types in attracting Spruce moths to a trap. They are also interested in the effect of location of the trap on its efficacy as well.

Data on the right reflects the number of moths caught by each trap and location type.

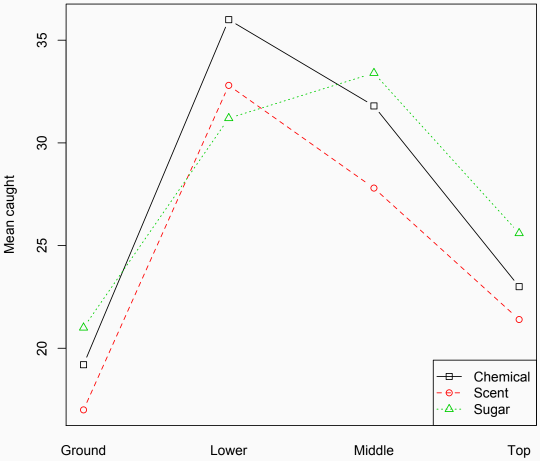
Factor 1 is the lure type (3 levels)
Factor 2 is the location (4 levels)
There are 5 observations per condition

	Scent	Sugar	Chemical
Top	28	35	32
	19	22	29
	32	33	16
	15	21	18
	13	17	20
Middle	39	36	37
	12	38	40
	42	44	18
	25	27	28
	21	22	36
Lower	44	42	35
	21	17	39
	38	31	41
	32	29	31
	29	37	34
Ground	17	18	22
	12	27	25
	23	15	14
	19	29	16
	14	16	1

Mean caught by Treatment

	Ground	Lower	Middle	Top	Lure Mean
Chemical	19.20	36.00	31.80	23.00	27.50
Scent	17.00	32.80	27.80	21.40	24.75
Sugar	21.00	31.20	33.40	25.60	27.80
Loc Mean	19.07	33.33	31.00	23.33	26.68

Mean caught by Treatment



Example - Spruce Moths - Hypotheses

Similar to the randomized block ANOVA, we have two hypothese to evaluate (one for each factor).

Lure effect:

$$H_0: \ \tau_{Ch} = \tau_{Sc} = \tau_{Su}, \quad H_A: \ \text{at least one pair of } \tau\text{s differ}$$

Location effect:

$$H_0: \ \beta_G = \beta_L = \beta_M = \beta_T, \quad H_A: \ \text{at least one pair of } \beta\text{s differ}$$

Example - Spruce Moths - ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lure					0.3859
Location		1981.38			0.0000
Residuals					
Total		5242.98			

Example - Spruce Moths - ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lure	2	113.03	56.52	0.97	0.3859
Location	3	1981.38	660.46	11.33	0.0000
Residuals	54	3148.57	58.31		
Total	59	5242.98			

Example - Spruce Moths - ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lure	2	113.03	56.52	0.97	0.3859
Location	3	1981.38	660.46	11.33	0.0000
Residuals	54	3148.57	58.31		
Total	59	5242.98			

Conclusions:

- Fail to reject H_0(Lure), there is not sufficient evidence to suggest the different lures have an effect.
- Reject H_0(Location), there is sufficient evidence to suggest the locations have an effect.

Difference between a blocking variable and a factor

We have just seen that computationally the two are treated the same when conducting an ANOVA.

What then is the difference?

- Factors are conditions we impose on the experimental units.
- Blocking variables are characteristics / innate properties of the experimental units.

Example - Lighting

A study is designed to test the effect of type of light on exam performance of students. 180 students are randomly assigned to three classrooms: one that is dimly lit, another with yellow lighting, and a third with white fluorescent lighting and given the same exam.

What are the factor(s) and/or block(s) for this experiment? What type of ANOVA would be appropriate?

Example - Lighting

A study is designed to test the effect of type of light on exam performance of students. 180 students are randomly assigned to three classrooms: one that is dimly lit, another with yellow lighting, and a third with white fluorescent lighting and given the same exam.

What are the factor(s) and/or block(s) for this experiment? What type of ANOVA would be appropriate?

The researcher also believes that light levels might have a different effect on males and females, so wants to make sure both genders are represented equally under the different light conditions.

After this modifications what are the factor(s) and/or block(s) for this experiment? What type of ANOVA would be appropriate?