Lecture 19 - ANOVA

Sta 111

Colin Rundel

June 11, 2014

Wolf River

- The Wolf River in Tennessee flows past an abandoned site once used by the pesticide industry for dumping wastes, including chlordane (pesticide), aldrin, and dieldrin (both insecticides).
- These highly toxic organic compounds can cause various cancers and birth defects.
- These compounds are denser than water and their molecules tend to become stuck in sediment, and are more likely to be found in higher concentrations near the bottom than near mid-depth.

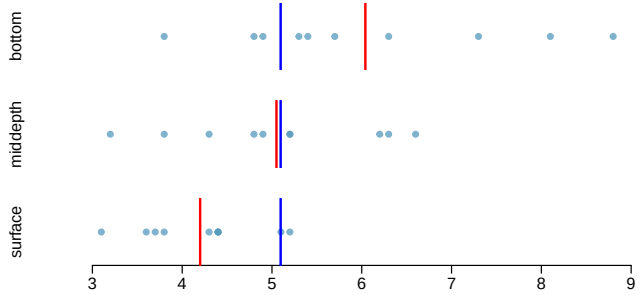
Wolf River - Data

Aldrin concentration (nanograms per liter) at three levels of depth.

	aldrin	depth
1	3.80	bottom
2	4.80	bottom
⋮	⋮	⋮
10	8.80	bottom
11	3.20	middepth
12	3.80	middepth
⋮	⋮	⋮
20	6.60	middepth
21	3.10	surface
22	3.60	surface
⋮	⋮	⋮
30	5.20	surface

Exploratory analysis

Aldrin concentration (nanograms per liter) at three levels of depth.



	n	mean	sd
bottom	10	6.04	1.58
middepth	10	5.05	1.10
surface	10	4.20	0.66
overall	30	5.10	1.37

Research question

Is there a difference between the mean aldrin concentrations among the three levels?

- To compare means of 2 groups we use a Z or a T statistic.
- To compare means of 3 or more groups we use a new test called *ANOVA* (analysis of variance) and a new test statistic, *F*.

ANOVA

ANOVA is used to assess whether the mean of the outcome variable is different for different levels of a categorical variable.

H_0 : The mean outcome is the same across all categories,

$$\mu_1 = \mu_2 = \cdots = \mu_k,$$

where μ_i represents the mean of the outcome for observations in category i.

H_A : At least one pair of means differ.

Note - this hypothesis test does not tell us if all the means are different or only if one pair is different, more on how to do that later.

Conditions

1. The observations should be independent within and between groups
   - If the data are a simple random sample from less than 10% of the population, this condition is satisfied.
   - Carefully consider whether the data may be independent (e.g. no pairing).
   - Always important, but sometimes difficult to check.
2. The observations within each group should be nearly normal.
   - Particularly important when the sample sizes are small.

   How do we check for normality?
3. The variability across the groups should be about equal.
   - Particularly important when the sample sizes differ between groups.

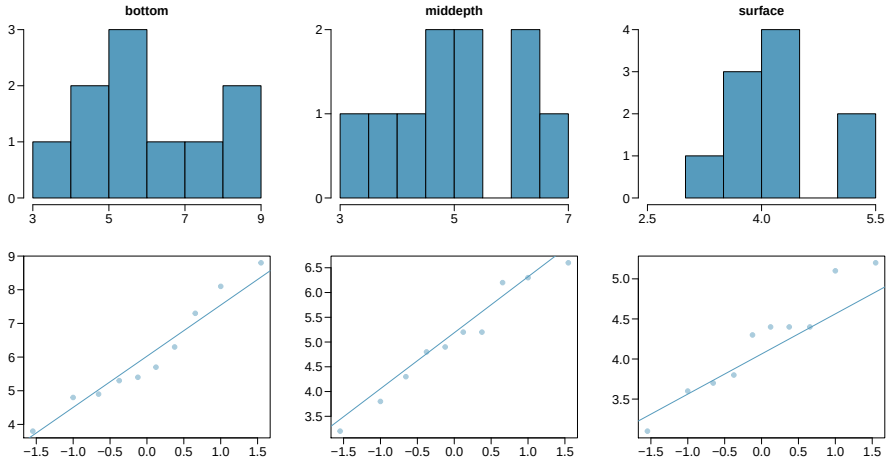
   How can we check this condition?

(1) Independence

Does this condition appear to be satisfied for the Wolf River data?

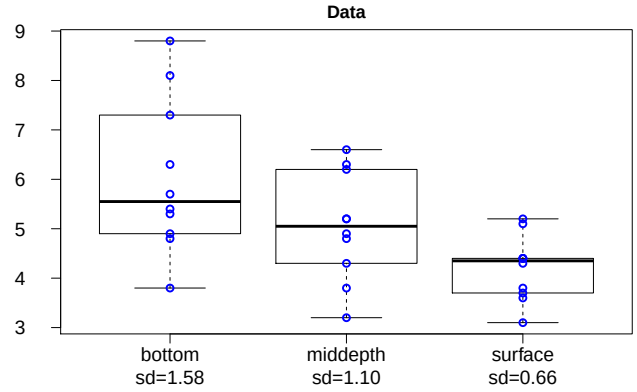
(2) Approximately normal

Does this condition appear to be satisfied?



(3) Constant variance

Does this condition appear to be satisfied?



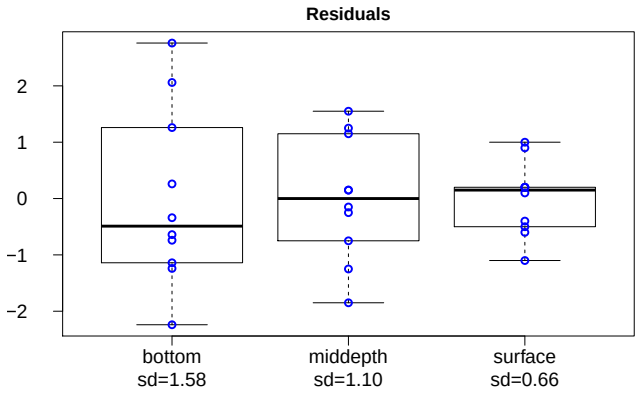
In this case it is somewhat hard to tell since the means are different.

(3) Constant variance - Residuals

One of the ways to think about each data point is as follows:

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

where ϵ_{ij} is called the residual ($\epsilon_{ij} = y_{ij} - \mu_i$).



z/t test vs. ANOVA - Purpose

z/t test

Compare means from *two* groups to see whether they are so far apart that the observed difference cannot reasonably be attributed to sampling variability.

$$H_0 : \mu_1 = \mu_2$$

ANOVA

Compare the means from *two or more* groups to see whether they are so far apart that the observed differences cannot all reasonably be attributed to sampling variability.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

z/t test vs. ANOVA - Method

z/t test

Compute a test statistic (a ratio).

$$z/t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE(\bar{x}_1 - \bar{x}_2)}$$

ANOVA

Compute a test statistic (a ratio).

$$F = \frac{\text{variability btw. groups}}{\text{variability w/in groups}}$$

- Large test statistics lead to small p-values.
- If the p-value is small enough H_0 is rejected, and we conclude that the population means are not equal.

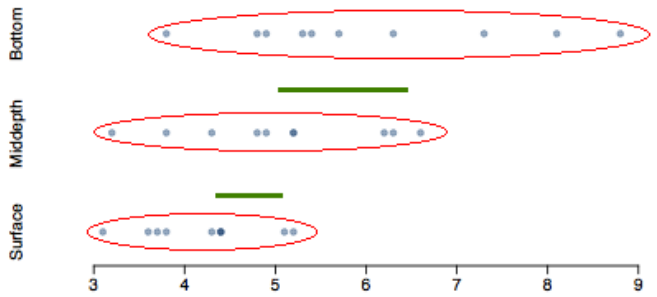
z/t test vs. ANOVA

- With only two groups t-test and ANOVA are equivalent, but only if we use a pooled standard variance in the denominator of the test statistic.
- With more than two groups, ANOVA compares the sample means to an overall *grand mean*.

Test statistic

Does there appear to be a lot of variability within groups? How about between groups?

$$F = \frac{\text{variability btw. groups}}{\text{variability w/in groups}}$$



Test statistic (cont.)

$$F = \frac{\text{variability btw. groups}}{\text{variability w/in groups}} = \frac{MSG}{MSE}$$

- *MSG* is mean square between groups

$$df_G = k - 1$$

where k is number of groups

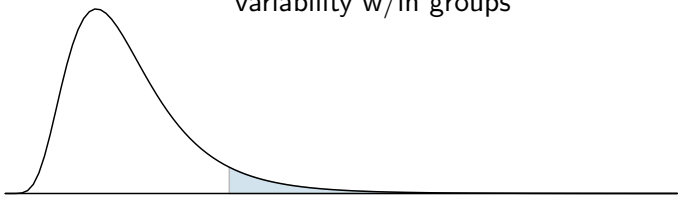
- *MSE* is mean square error - variability in residuals

$$df_E = n - k$$

where n is number of observations.

F distribution and p-value

$$F = \frac{\text{variability btw. groups}}{\text{variability w/in groups}}$$



- In order to be able to reject H_0, we need a small p-value, which requires a large F statistic.
- In order to obtain a large F statistic, variability between sample means needs to be greater than variability within sample means.

ANOVA output

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	6.13	0.0063
(Error)	Residuals	27	37.33	1.38		
	Total	29	54.29			

Degrees of freedom associated with ANOVA

- groups: $df_G = k - 1$, where k is the number of groups
- total: $df_T = n - 1$, where n is the total sample size
- error: $df_E = df_T - df_G$

- $df_G = k - 1 = 3 - 1 = 2$
- $df_T = n - 1 = 30 - 1 = 29$
- $df_E = 29 - 2 = 27$

ANOVA output (cont.)

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	6.13	0.0063
(Error)	Residuals	27	37.33	1.38		
	Total	29	54.29			

Sum of squares between groups, SSG

Measures the variability between groups

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

where n_i is each group size, $\bar{x}_i$ is the average for each group, $\bar{x}$ is the overall (grand) mean.

	n	mean
bottom	10	6.04
middepth	10	5.05
surface	10	4.2
overall	30	5.1

$$\begin{aligned} SSG &= \left(10 \times (6.04 - 5.1)^2\right) \\ &\quad + \left(10 \times (5.05 - 5.1)^2\right) \\ &\quad + \left(10 \times (4.2 - 5.1)^2\right) \\ &= 16.96 \end{aligned}$$

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	6.13	0.0063
(Error)	Residuals	27	37.33	1.38		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

$$SST = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

where x_i represent each observation in the dataset.

$$\begin{aligned} SST &= (3.8 - 5.1)^2 + (4.8 - 5.1)^2 + (4.9 - 5.1)^2 + \cdots + (5.2 - 5.1)^2 \\ &= (-1.3)^2 + (-0.3)^2 + (-0.2)^2 + \cdots + (0.1)^2 \\ &= 1.69 + 0.09 + 0.04 + \cdots + 0.01 = 54.29 \end{aligned}$$

ANOVA output (cont.) - SSE

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	6.13	0.0063
(Error)	Residuals	27	*37.33*	1.38		
	Total	29	54.29			

Sum of squares error, SSE

Measures the variability within groups:

$$SSE = SST - SSG = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

$$SSE = 54.29 - 16.96 = 37.33$$

ANOVA output (cont.) - MS

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	*8.48*	6.13	0.0063
(Error)	Residuals	27	37.33	*1.38*		
	Total	29	54.29			

Mean square

Mean square is calculated as sum of squares divided by the degrees of freedom.

$$\begin{aligned} MSG &= 16.96/2 = 8.48 \\ MSE &= 37.33/27 = 1.38 \end{aligned}$$

ANOVA output (cont.) - F

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	*6.14*	0.0063
(Error)	Residuals	27	37.33	1.38		
	Total	29	54.29			

Test statistic, F value

As we discussed before, the F statistic is the ratio of the between group and within group variability.

$$F = \frac{MSG}{MSE}$$

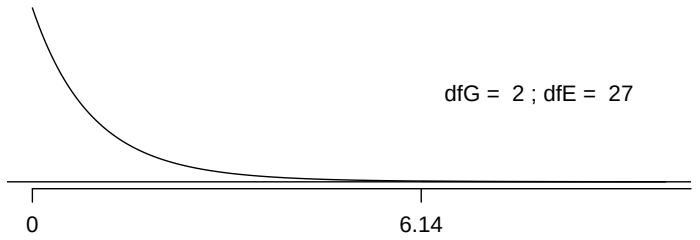
$$F = \frac{8.48}{1.38} = 6.14$$

ANOVA output (cont.) - P-value

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Group)	depth	2	16.96	8.48	*6.14*	0.0063
(Error)	Residuals	27	37.33	1.38		
	Total	29	54.29			

P-value

The probability of at least as large a ratio between the "between group" and "within group" variability, if in fact the means of all groups are equal. It's calculated as the area under the F curve, with degrees of freedom df_G and df_E, above the observed F statistic.



Conclusion

- If p-value is small (less than α), reject H_0. The data provide convincing evidence that at least one mean is different from (but we can't tell which one).
- If p-value is large, fail to reject H_0. The data do not provide convincing evidence that at least one pair of means are different from each other, the observed differences in sample means are attributable to sampling variability (or chance).

What is the conclusion of the hypothesis test for Wolf river?

Which means differ?

- We've concluded that at least one pair of means differ. The natural question that follows is "which ones?"
- We can do two sample t tests for differences in each possible pair of groups.

Can you see any pitfalls with this approach?

- When we run too many tests, the Type 1 Error rate increases.
- This issue is resolved by using a modified significance level.

Multiple comparisons

- The scenario of testing many pairs of groups is called *multiple comparisons*.
- If there are k groups, then there are $K = \binom{k}{2} = \frac{k(k-1)}{2}$ possible pairs.
- One common approach is the *Bonferroni correction* that uses a more *stringent* significance level for each test:

$$\alpha^* = \alpha / K$$

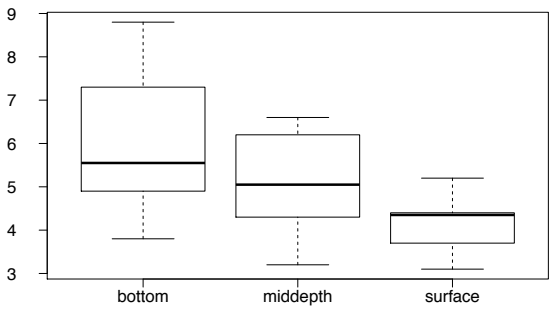
where K is the number of comparisons being considered.

Determining the modified α

In the aldrin data set depth has 3 levels: bottom, mid-depth, and surface. If $\alpha = 0.05$, what should be the modified significance level or two sample t tests for determining which pairs of groups have significantly different means?

Which means differ?

Based on the box plots below, which means would you expect to be significantly different?



(a) bottom & surface

(b) bottom & mid-depth

(c) mid-depth & surface

(d) bottom & mid-depth; mid-depth & surface

(e) bottom & mid-depth; bottom & surface; mid-depth & surface

Which means differ? (cont.)

If the ANOVA assumption of equal variability across groups is satisfied, we can use the data from all groups to estimate variability:

- Estimate any within-group standard deviation with $\sqrt{MSE}$, which is s_{pooled}
- Use the error degrees of freedom, $n - k$, for t-distributions

Difference in two means: after ANOVA

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

Is there a difference between the average aldrin concentration at the bottom and at mid depth?

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
depth	2	16.96	8.48	6.13	0.0063
Residuals	*27*	37.33	*1.38*		
Total	29	54.29			

	n	mean	sd
bottom	10	*6.04*	1.58
middepth	10	*5.05*	1.10
surface	10	4.2	0.66
overall	30	5.1	1.37

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m)}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

$$T_{27} = \frac{(6.04 - 5.05)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{0.99}{0.53} = 1.87$$

$$0.05 < p - value < 0.10 \quad \text{(two-sided)}$$

$$\alpha^\star = 0.05/3 = 0.0167$$

Fail to reject H_0, the data do not provide convincing evidence of a difference between the average aldrin concentrations at bottom and mid depth.

Is there a difference between the average aldrin concentration at the bottom and at surface?

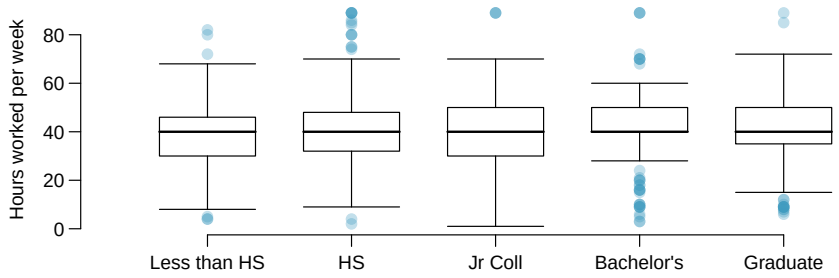
Work Hours and Education Level

Previously we have seen data from the General Social Survey in order to compare the average number of hours worked per week by US residents with and without a college degree. However, this analysis didn't take advantage of the original data which contained more accurate information on educational attainment (less than high school, high school, junior college, Bachelor's, and graduate school).

Using ANOVA, we can consider educational attainment levels for all 1,172 respondents at once instead of re-categorizing them into two groups. On the following slide are the distributions of hours worked by educational attainment and relevant summary statistics that will be helpful in carrying out this analysis.

Work Hours and Education Level

	Educational attainment					
	Less than HS	HS	Jr Coll	Bachelor's	Graduate	Total
Mean	38.67	39.6	41.39	42.55	40.85	40.45
SD	15.81	14.97	18.1	13.62	15.51	15.17
n	121	546	97	253	155	1,172



Work Hours and Education Level (ANOVA table)

Given what we know, fill in the unknowns in the ANOVA table below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree	*???*	*???*	501.54	*???*	0.0682
Residuals	*???*	267,382	*???*		
Total	*???*	*???*			

	Educational attainment					
	Less than HS	HS	Jr Coll	Bachelor's	Graduate	Total
Mean	38.67	39.6	41.39	42.55	40.85	40.45
SD	15.81	14.97	18.1	13.62	15.51	15.17
n	121	546	97	253	155	1,172

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 alfalfas growth. Since we have a numerical response and categorical explanatory variable we will use an ANOVA.

What should our hypotheses be?

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

H_A: At least one mean is different

Treatment Effect

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

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

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 known as the treatment effect.

Thinking in terms of the treatment effect we can rewrite our null hypothesis

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

Alfalfa ANOVA Table - Sum Sq

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

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \mu)^2$$
$$= (1.3 - 1.269)^2 + (1.15 - 1.269)^2 + \cdots + (0.80 - 1.269)^2 = 5.879$$
$$SSG = \sum_{i=1}^{k} n_i(\mu_i - \mu)^2$$
$$= 5 \times (0.91 - 1.269)^2 + 5 \times (1.13 - 1.269)^2 + 5 \times (1.768 - 1.269)^2 = 1.986$$
$$SSE = SST - SSG = 3.893$$

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	*0.993*	*3.061*	*0.0843*
Residuals	12	3.893	*0.324*		
Total	14	5.879			

$$MSG = SSG/df_G = 1.986/2 = \textit{0.993}$$
$$MSE = SSE/df_E = 3.907/12 = \textit{0.324}$$
$$F = MSG/MSE = 0.993/0.326 = \textit{3.061}$$
$$\text{P-value} = P(> F) = \textit{0.0843}$$

Based on these results we fail to reject H_0, and there is not sufficient evidence to suggest that at least one of the mean growth values is significantly different (or that at least one of the treatment effects is not zero).

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

However, random sampling / assignment ensure that in the long run these nuisance factors show up with equal frequency in all treatment levels and as such their effect(s) will cancel out.

Blocking

Why do we bother with controls then? Because they help reduce noise/uncertainty in the data.

Types of Controls

- Exclusion
  - Works if the number of patients with a previous heart attack is low
  - Can only exclude so many nuisance factors
  - 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
  - Previous heart attack block and a no previous heart attack block
  - 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?

Blocked Alfalfa

We will only 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:

$$H_0(\text{treatment}) : \tau_H = \tau_L = \tau_C = 0$$
$$H_0(\text{block}) : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

In order to test these hypotheses we will build on 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	$k-1$	$\sum_{i=1}^{k} n_i(\mu_{i\bullet} - \mu)^2$	SSG/df_G	MSG/MSE
Block	$b-1$	$\sum_{j=1}^{b} m_j(\mu_{\bullet j} - \mu)^2$	SSB/df_B	MSB/MSE
Error	$n-k-b+1$	$SST - SSG - SSB$	SSE/df_E	
Total	$n-1$	$\sum_i \sum_j \sum_k (y_{ijk} - \mu)^2$		

- n - # observations
- k - # groups
- b - # blocks
- n_i - # observations in group i
- m_j - # observations in block j
- $\mu_{i\bullet}$ - group mean for group i
- $\mu_{\bullet j}$ - block mean for group 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*	$\sum_{j=1}^{b} m_j(\mu_{\bullet j} - \mu)^2$	SSB/df_B	MSB/MSE
Error	*8*	$SST - SSG - SSB$	SSE/df_E	
Total	*14*	*5.879*		

Sum of Squares Blocks

$$SSB = \sum_{j=1}^{b} m_j(\mu_{\bullet j} - \mu)^2$$

	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	
		$\mu = 1.269$		

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

Completing the table

	df	Sum Sq	Mean Sq	F value
Group	2	1.986	0.993	*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

$$H_0: \ \tau_H = \tau_L = \tau_G, \quad H_A: \ \text{At least one treatment effect is not zero}$$

- Block effect

$$H_0: \ \beta_1 = \beta_2 = \ldots = \beta_5, \quad H_A: \ \text{At least one block effect is not zero}$$

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] 0.03182
```

Therefore, $P(> F_G) = 0.0318$, which leads us to reject H_0 - there is sufficient evidence to suggest that at least one treatment effect is not 0.

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] 0.0679
```

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 block effect is not 0.

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*).

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 if a second factor 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 factor 1
- β_j is the effect of level j of factor 2
- ϵ_{ijk} is the residual of observation k in block j with treatment i

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

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 to the right reflects the number of moths caught.

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

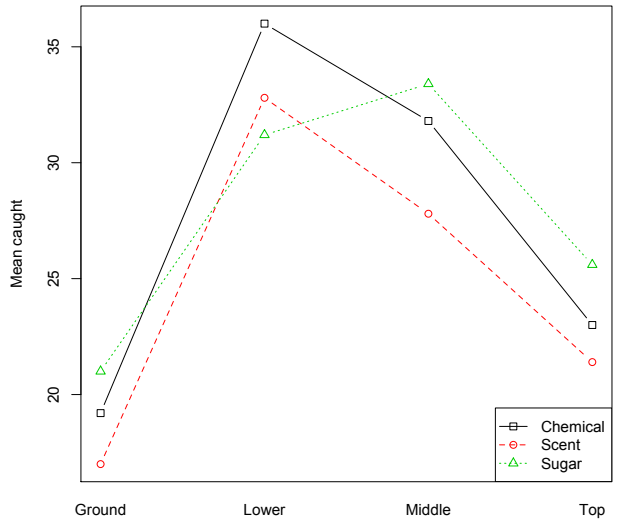
From Understandable Statistics, 7e

	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:

$$H_0: \ \tau_{Ch} = \tau_{Sc} = \tau_{Su}, \quad H_A: \ \text{at least one } \tau \text{ is not zero}$$

Location effect:

$$H_0: \ \beta_G = \beta_L = \beta_M = \beta_T, \quad H_A: \ \text{at least one } \beta \text{ is not zero}$$

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			

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 that the experimental units come with.

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?