Lecture 16 - ANOVA

Sta102 / BME102

March 28, 2016

Colin Rundel

ANOVA

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).

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.

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.

Given that these compounds are denser than water, researchers believe that their molecules are likely to be deposited in sediment.

Wolf River - Data

Aldrin concentration (ng / L) 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

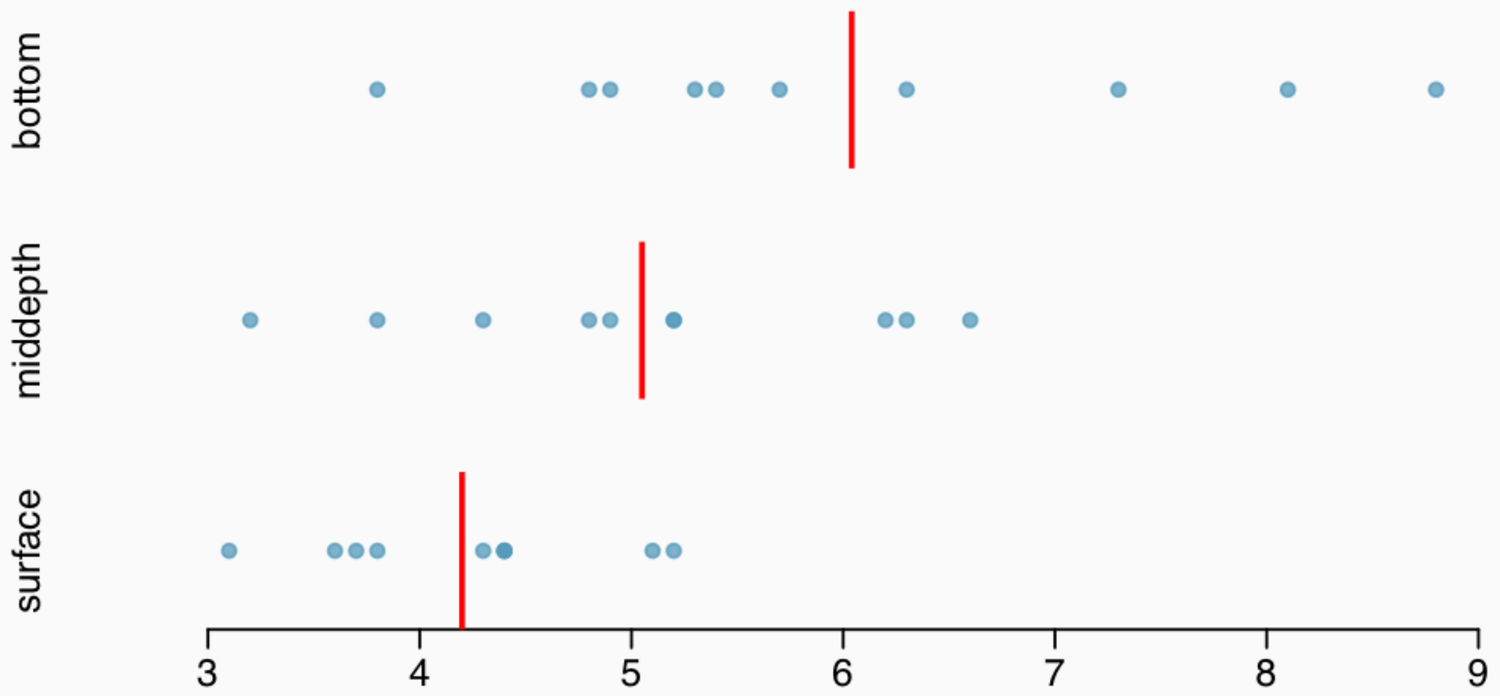
Exploratory analysis

Aldrin concentration (ng / L) 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

Exploratory analysis

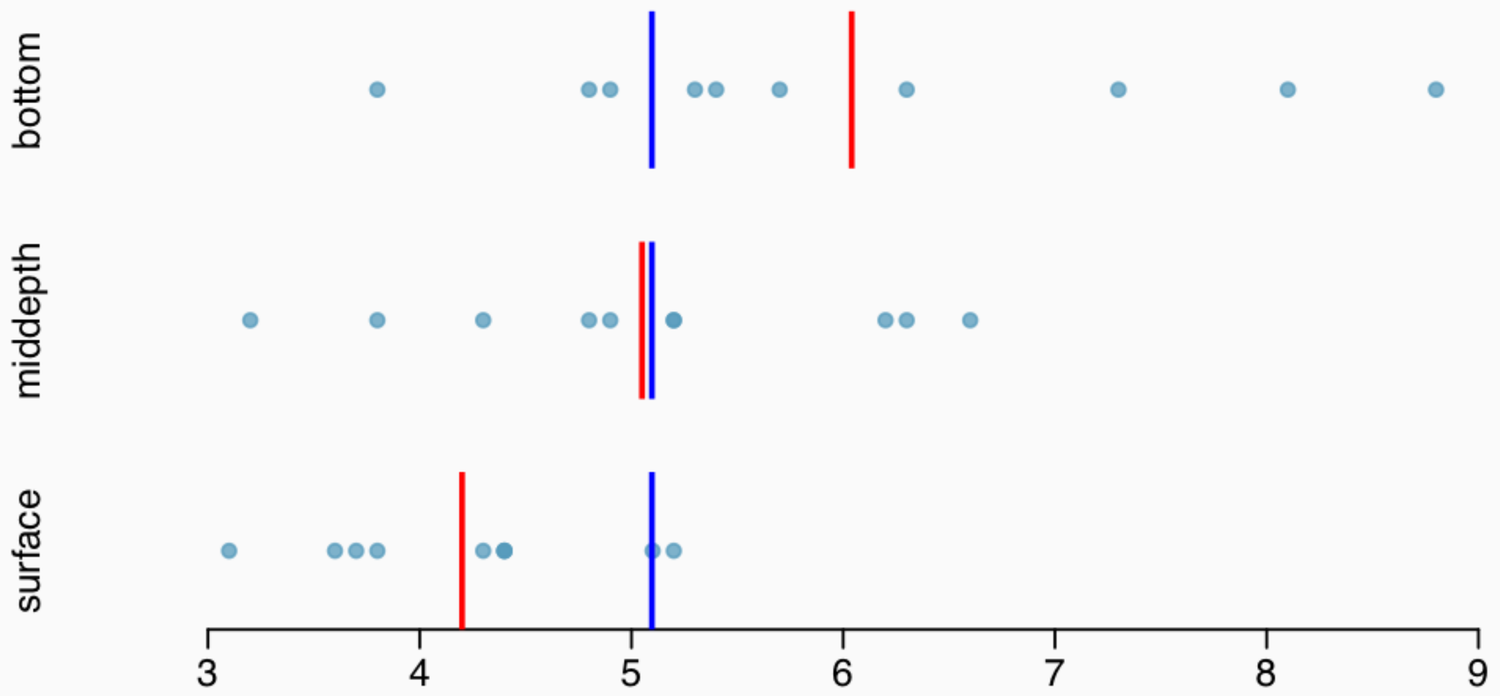
Aldrin concentration (ng / L) at three levels of depth.



	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

Exploratory analysis

Aldrin concentration (ng / L) 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?

Research question

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

- To compare means of 2 groups we use a T distribution.

Research question

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

- To compare means of 2 groups we use a *T* distribution.

- To compare means of 3 or more groups we use a new test called *ANOVA* (analysis of variance) and a new test statistic / sampling distribution - *F*.

$$F(df_1, df_2)$$

All pairwise tests?

Instead of ANOVA why can we not just do *t* tests for differences in each possible pair of groups?

All pairwise tests?

Instead of ANOVA why can we not just do t tests for differences in each possible pair of groups?

- The total number of tests increases rapidly, if there are k levels then $\binom{k}{2} = \frac{k(k-1)}{2}$ t tests are needed.
- When we run too many tests we increase our *overall* Type 1 Error rate.
- This issue is referred to as *multiple comparisons* or *multiple testing*.
- More on possible solutions later.

ANOVA

ANOVA is used to assess whether (some of) the means are different between the levels of the (categorical) predictor variable.

ANOVA

ANOVA is used to assess whether (some of) the means are different between the levels of the (categorical) predictor variable.

H_0 : The group means are all equal, $\mu_1 = \mu_2 = \cdots = \mu_k$, where μ_i represents the mean of the outcome for observations in category i.

H_A : At least one pair of group means are different.

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

Conditions

1. *Independence* - The observations should be independent *within* and *between* groups

2. *Nearly Normal* - The observations within each group should be nearly normal.

   - Larger sample does not fix this if non-normal

3. *Constant Varaince* - The variance (σ_i^2) across the groups should be equal.

(1) Independence

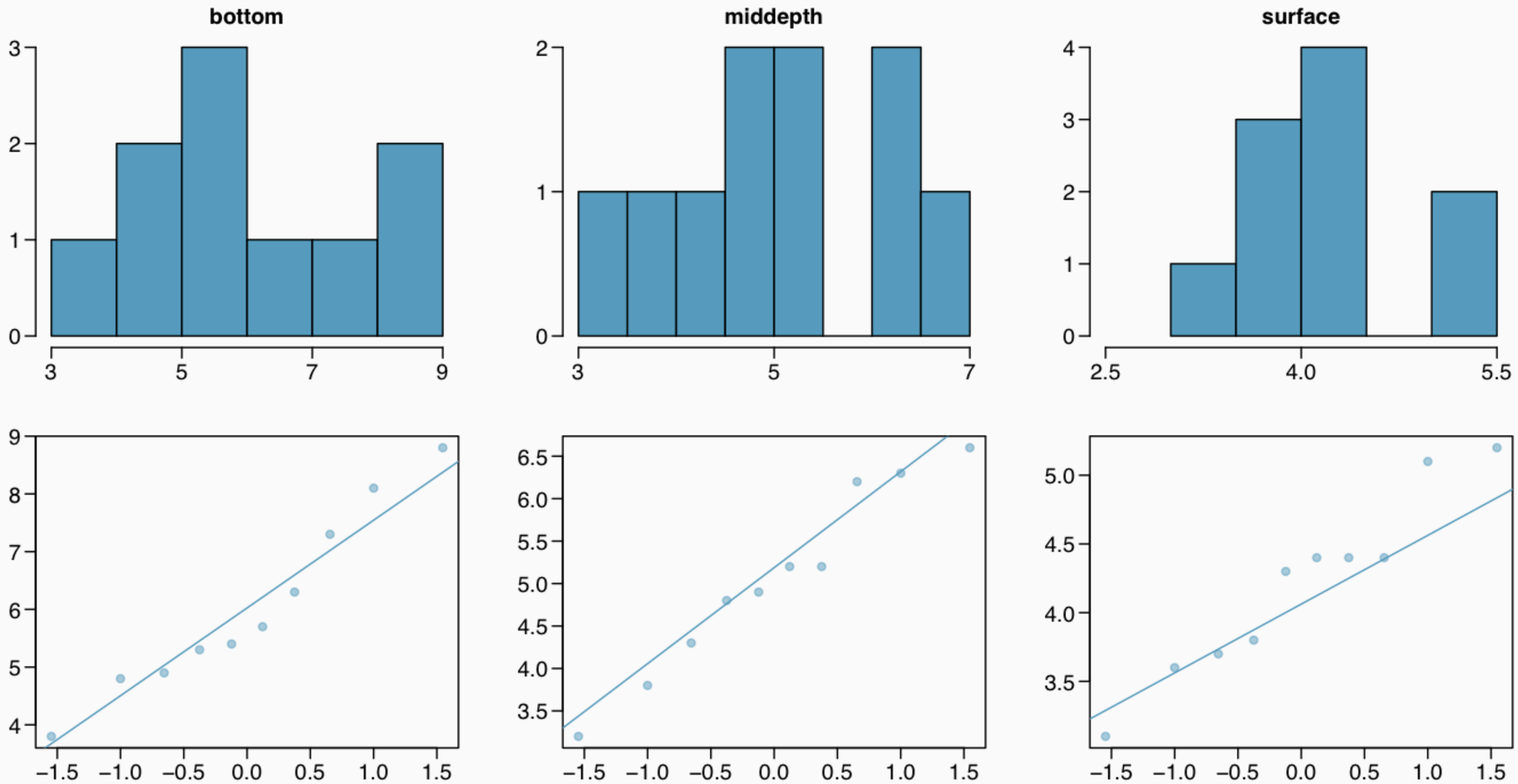
Does this condition appear to be satisfied for the Wolf River data?

- 30 diff. samples (not paired) locs.

- < 10%

(2) Approximately normal

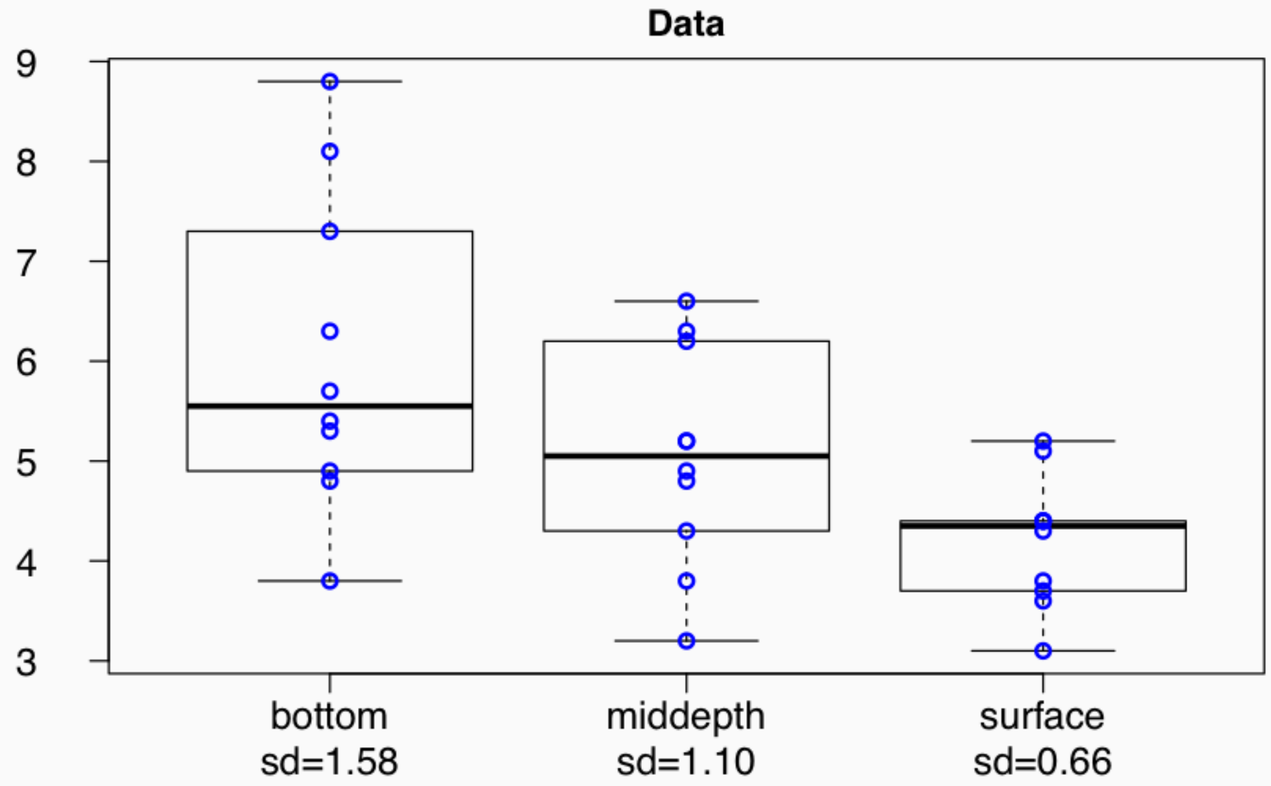
Does this condition appear to be satisfied?



(3) Constant variance

Does this condition appear to be satisfied?

	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



(3) Constant variance - Residuals

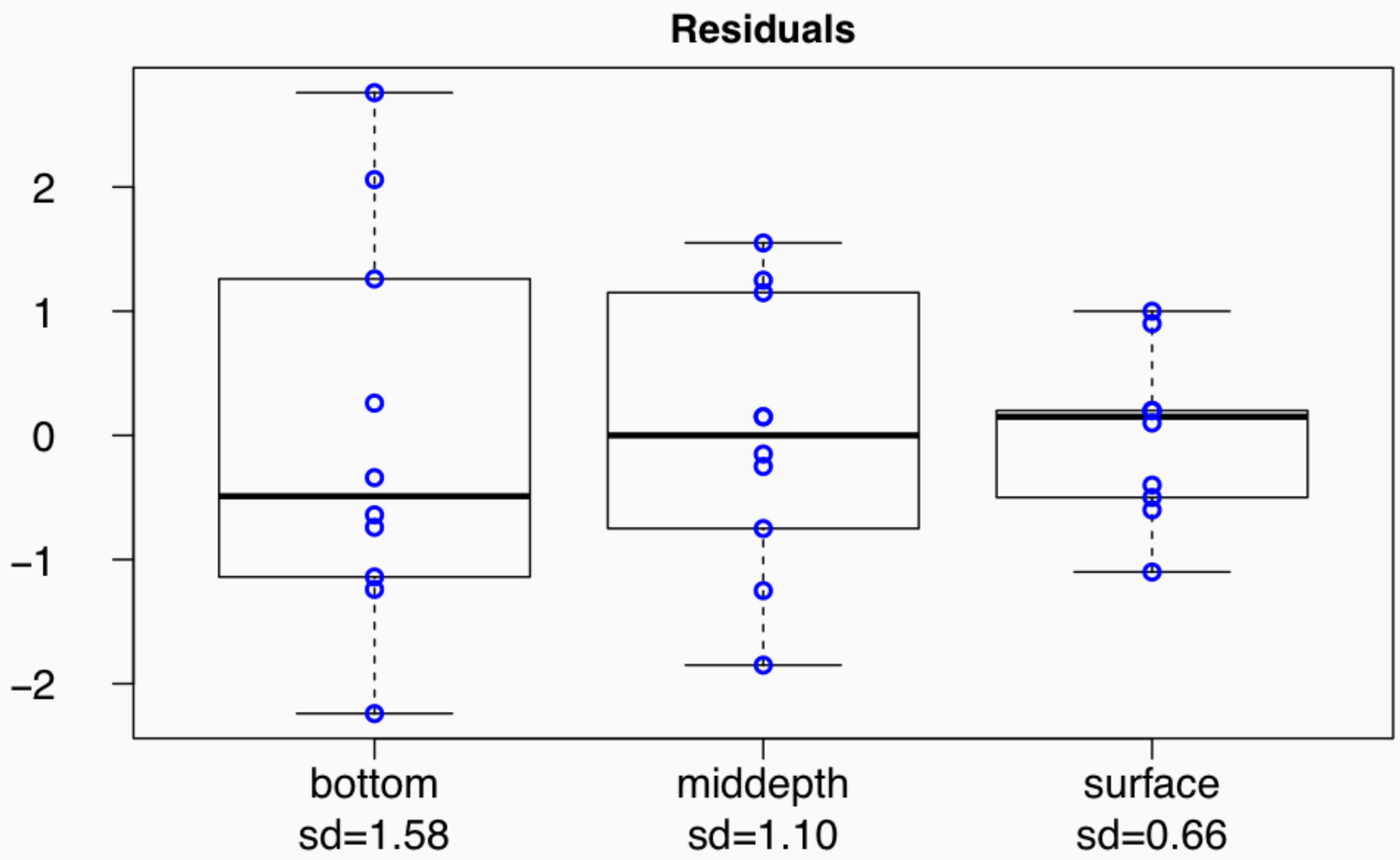
Another way to think about each data point (observations) is as follows:

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

↗ group mean of group i

↳ Resid obs j of group i

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



t test vs. ANOVA - Purpose

t test

Compares the means from *two* groups to see if they are so far apart that the observed difference cannot reasonably be attributed to sampling uncertainty.

$$H_0 : \mu_1 = \mu_2$$

ANOVA

Compares 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 uncertainty.

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

Note - When there are only two groups the t-test and ANOVA are *exactly* equivalent as long as we use a *pooled variance* for the t-test.

t test vs. ANOVA - Method

t test

Compute a test statistic (a ratio).

$$T = \frac{\text{difference btw. groups}}{\text{variability of groups}}$$

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

As T ↑ then the p-value ↓

ANOVA

Compute a test statistic (a ratio).

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

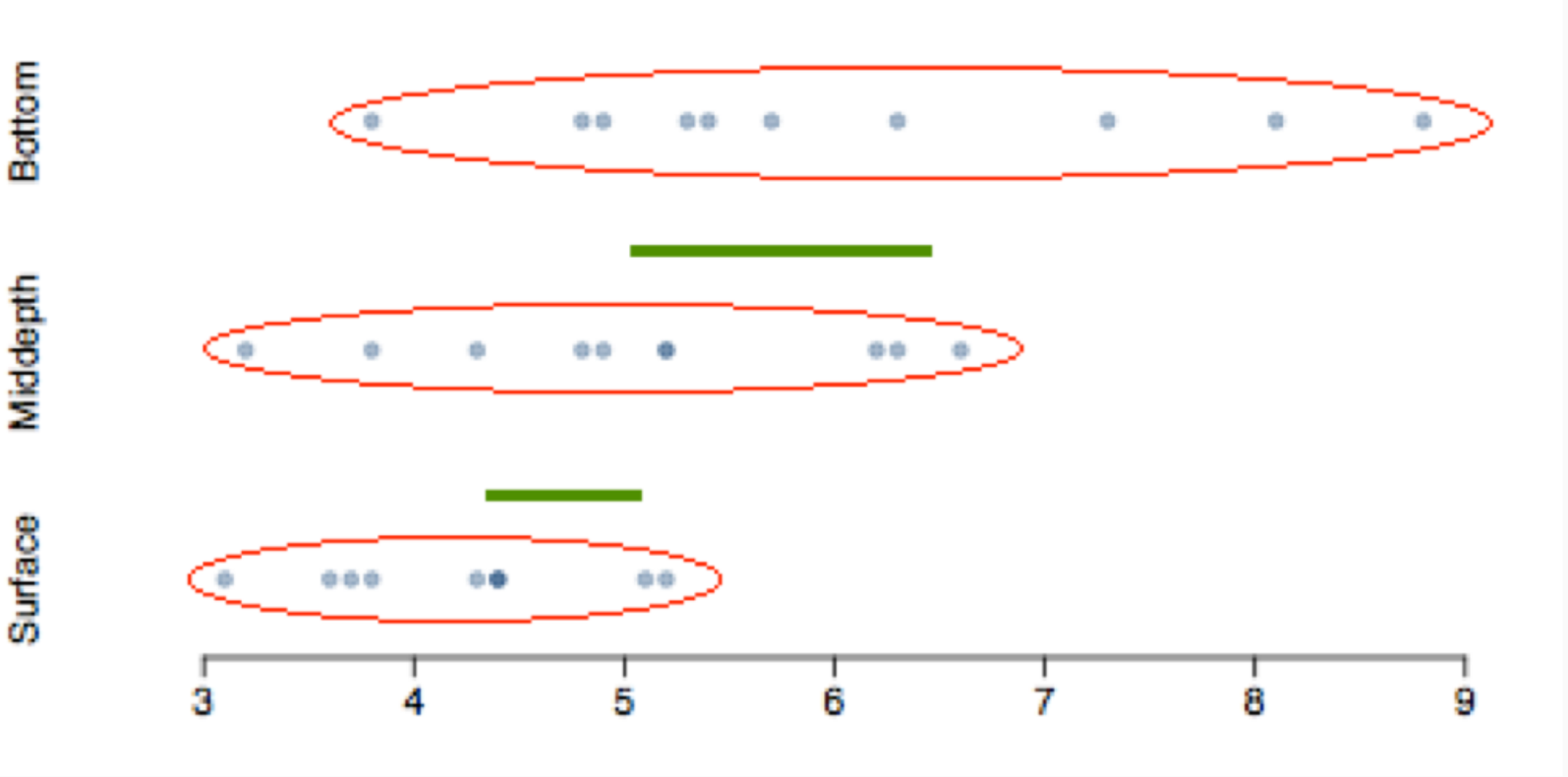
As F ↑ then the p-value ↓

Test statistic

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

How about between groups?

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



Types of Variability

For ANOVA we think of our variability (uncertainty) in terms of three separate quantities:

- *Total variability* - all of the variability in the data, ignoring any explanatory variable(s).

- *Group variability* - variability between the group means and the grand mean.

- *Error variability* - the sum of the variability within each group.

Sum of squares and Variability

Mathematically, we can think of the following measures of variability:

- Total variability - Sum of Squares Total

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

- Group variability - Sums of Squares Group

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

- Error variability - Sum of Squares Error

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

Partitioning Sums of Squares

With a little bit of careful algebra we can show that:

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

Total Variability = Group Variability (~~within~~) + Error Variability (~~between~~)

Sum of Squares Total = Sum of Squares Group + Sum of Squares Error

ANOVA Output

The results of an ANOVA is usually summarized in a tabular form that includes these measures of uncertainty as well as the calculation of the F test statistic.

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

$$F = \frac{\text{var btw}}{\text{var w/in}} = \frac{MSG}{MSE}$$

$$F_{df_1 = df_g} \quad df_2 = df_E$$

ANOVA output - SSG

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	*16.96*	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

Sum of squares between groups, SSG - Measures the variability between groups

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

where n_i is the size of group i, $\bar{y}_i$ is the average of group i, and $\bar{y}$ is the overall (grand) mean.

ANOVA output - SSG

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	*16.96*	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

Sum of squares between groups, SSG - Measures the variability between groups

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

where n_i is the size of group i, $\bar{y}_i$ is the average of group i, and $\bar{y}$ is the overall (grand) mean.

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

ANOVA output - SSG

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	*16.96*	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

Sum of squares between groups, SSG - Measures the variability between groups

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

where n_i is the size of group i, $\bar{y}_i$ is the average of group i, and $\bar{y}$ is the overall (grand) mean.

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

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

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	*54.29*			

Sum of squares total, SST - Measures the variability between groups

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

where $x_{i,j}$ is observation j of group i.

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	*54.29*			

Sum of squares total, SST - Measures the variability between groups

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

where $x_{i,j}$ is observation j of group i.

$$SST = (3.8 - 5.1)^2 + (4.8 - 5.1)^2 + (4.9 - 5.1)^2 + \cdots + (5.2 - 5.1)^2$$

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	*54.29*			

Sum of squares total, SST - Measures the variability between groups

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

where $x_{i,j}$ is observation j of group i.

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

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	*54.29*			

Sum of squares total, SST - Measures the variability between groups

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

where $x_{i,j}$ is observation j of group i.

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

ANOVA output (cont.) - SST

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	*54.29*			

Sum of squares total, SST - Measures the variability between groups

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

where $x_{i,j}$ is observation j of group i.

$$\begin{aligned} SST &= (3.8 - 5.1)^2 + (4.8 - 5.1)^2 + (4.9 - 5.1)^2 + \cdots + (5.2 - 5.1)^2 \\ &= (-1.3)^2 + (-0.3)^2 + (-0.2)^2 + \cdots + (0.1)^2 \\ &= 1.69 + 0.09 + 0.04 + \cdots + 0.01 \\ &= 54.29 \simeq s^2(n-1) \end{aligned}$$

ANOVA output (cont.) - SSE

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	*37.33*	1.38		
	*T*otal	29	54.29			

Sum of squares error, SSE - Measures the variability within groups:

$$SSE = SST - SSG$$

ANOVA output (cont.) - SSE

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	*37.33*	1.38		
	*T*otal	29	54.29			

Sum of squares error, SSE - Measures the variability within groups:

$$SSE = SST - SSG$$

$$SSE = 54.29 - 16.96 = 37.33$$

ANOVA output

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

ANOVA output

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	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: $df_E = df_T - df_G = n - k$

ANOVA output

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	6.13	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	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: $df_E = df_T - df_G = n - k$

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

$$df_T = n - 1 = 30 - 1 = 29$$

$$df_E = 29 - 2 = 27$$

ANOVA output (cont.) - MS

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	*8.48*	6.13	0.0063
(*E*rror)	Residuals	27	37.33	*1.38*		
	*T*otal	29	54.29			

Mean square

Mean square values are calculated as sum of squares divided by the degrees of freedom - these values represent the *normalized* measures of the variability between and variability within the groups respectively.

ANOVA output (cont.) - MS

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	*8.48*	6.13	0.0063
(*E*rror)	Residuals	27	37.33	*1.38*		
	*T*otal	29	54.29			

Mean square

Mean square values are calculated as sum of squares divided by the degrees of freedom - these values represent the *normalized* measures of the variability between and variability within the groups respectively.

$$MSG = SSG/df_G = 16.96/2 = 8.48$$

ANOVA output (cont.) - MS

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	*8.48*	6.13	0.0063
(*E*rror)	Residuals	27	37.33	*1.38*		
	*T*otal	29	54.29			

Mean square

Mean square values are calculated as sum of squares divided by the degrees of freedom - these values represent the *normalized* measures of the variability between and variability within the groups respectively.

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

ANOVA output (cont.) - F

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	*6.14*	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

Test statistic, F value

The F statistic is the ratio of the between group and within group variability.

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

ANOVA output (cont.) - P-value

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	*6.14*	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	29	54.29			

P-value

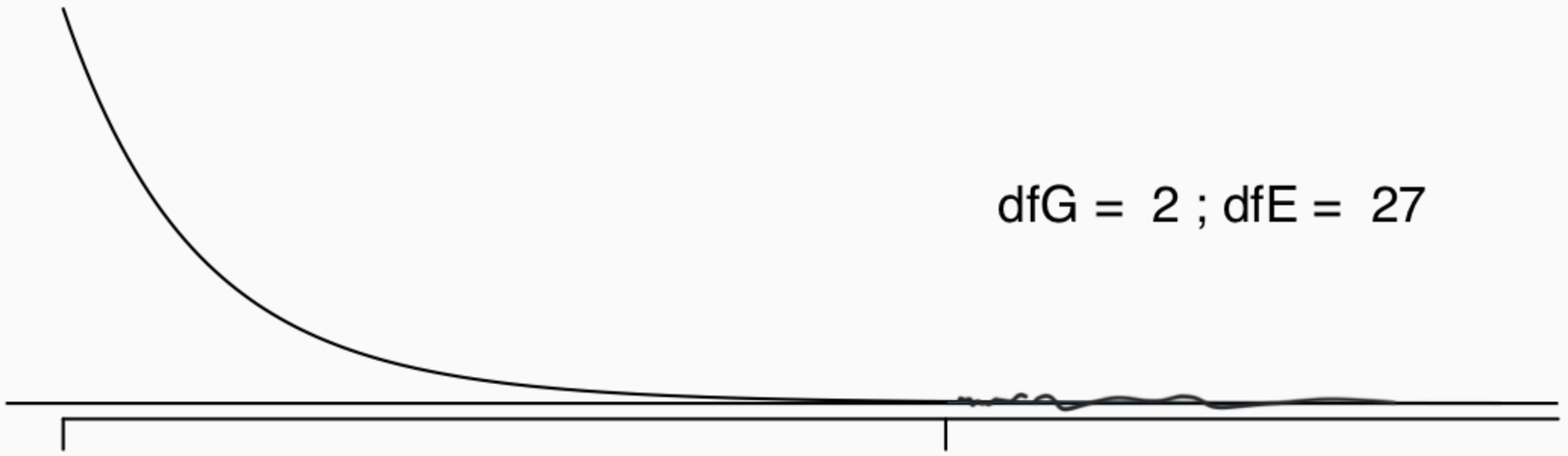
The probability of at least as large a ratio between the "between group" and "within group" variability, if in fact the means of all groups are equal. It's calculated as the area under the F distribution, with degrees of freedom df_G and df_E, above the observed F statistic.

ANOVA output (cont.) - P-value

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
(*G*roup)	depth	2	16.96	8.48	*6.14*	0.0063
(*E*rror)	Residuals	27	37.33	1.38		
	*T*otal	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 distribution, 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 pair of means differ (but we say specifically which pair).

Conclusion

If p-value is small (less than α), reject H_0. The data provide convincing evidence that at least one pair of means differ (but we say specifically which pair).

If the 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).

Conclusion

If p-value is small (less than α), reject H_0. The data provide convincing evidence that at least one pair of means differ (but we say specifically which pair).

If the 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 our hypothesis test for aldrin concentration in the Wolf river?

Conclusion

If p-value is small (less than α), reject H_0. The data provide convincing evidence that at least one pair of means differ (but we say specifically which pair).

If the 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 our hypothesis test for aldrin concentration in the Wolf river?

- The data provide convincing evidence that the average aldrin concentration is different for at least one pair.

Multiple comparisons/testing

Which means differ?

We've concluded that at least one pair of means differ. The natural question that follows is "which ones?"

Which means differ?

We've concluded that at least one pair of means differ. The natural question that follows is "which ones?"

We can perform two sample *t* tests for differences in each possible pair of groups (3 total in this case).

Which means differ?

We've concluded that at least one pair of means differ. The natural question that follows is "which ones?"

We can perform two sample *t* tests for differences in each possible pair of groups (3 total in this case).

As we mentioned previously, this presents a multiple testing issue - when we run too many tests, the Type 1 Error rate increases.

Which means differ?

We've concluded that at least one pair of means differ. The natural question that follows is "which ones?"

We can perform two sample *t* tests for differences in each possible pair of groups (3 total in this case).

As we mentioned previously, this presents a multiple testing issue - when we run too many tests, the Type 1 Error rate increases.

- If we were to conduct all three post-hoc tests, what would our overall Type 1 error rate be?

Overall Type I ER

$$= 1 - P(\text{no errors})$$

$$= 1 - 0.95\,(0.95)(0.95)$$

$$> 0.05$$

Correcting for Multiple testing

One common approach to address multiple testing is the *Bonferroni correction*

- For each individual test, use $\alpha^* = \alpha/K$ where K is the number of tests and α is the desired *overall* Type 1 error rate.
- This is a very stringent / conservative correction, assumes each decision is independent

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?

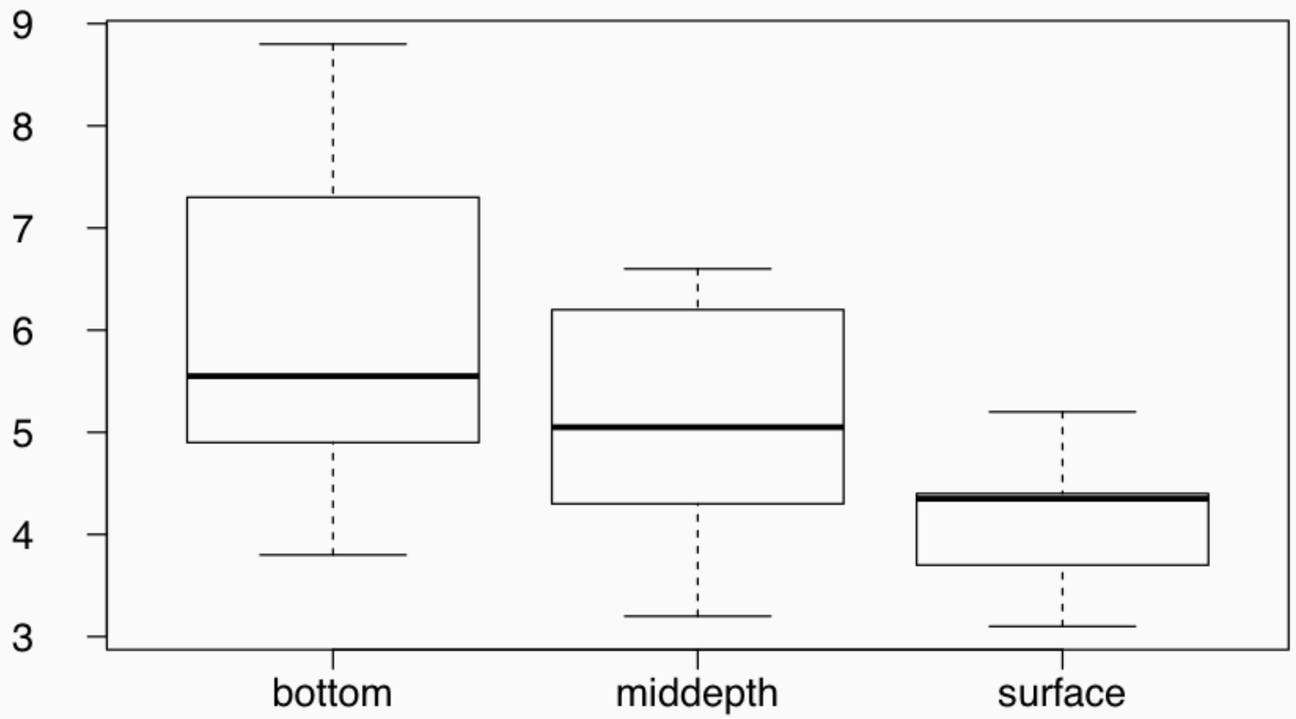
Determining the modified α

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?

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

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.)

For an ANOVA we make have an assumption that all the groups have equal variance, this is not a part of a normal t-test. When performing a posthoc test we should maintain this assumption and use a pooled estimate of variability and the appropriate degrees of freedom associated with this estimate for our t distribution.

- Replace within-group sample standard deviations with MSE, which is s^2_{pooled}
- Use the error degrees of freedom $(n - k)$ for t-distributions

Difference in two means - ANOVA posthoc test

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \approx \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

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

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m) - 0}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

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

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m) - 0}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

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

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

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m) - 0}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

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

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

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

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m) - 0}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

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

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

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

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

$$T_{df_E} = \frac{(\bar{x}_b - \bar{x}_m) - 0}{\sqrt{\frac{MSE}{n_b} + \frac{MSE}{n_m}}}$$

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

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

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

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?

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

$$T_{df_E} = \frac{(\bar{x}_{bottom} - \bar{x}_{surface})}{\sqrt{\frac{MSE}{n_{bottom}} + \frac{MSE}{n_{surface}}}}$$

$$T_{27} = \frac{(6.04 - 4.02)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{2.02}{0.53} = 3.81$$

$$p - value = P(T_{27} > 3.81 \text{ or } T_{27} < -3.81)$$

$$< 0.01$$

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

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

$$T_{df_E} = \frac{(\bar{x}_{bottom} - \bar{x}_{surface})}{\sqrt{\frac{MSE}{n_{bottom}} + \frac{MSE}{n_{surface}}}}$$

$$T_{27} = \frac{(6.04 - 4.02)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{2.02}{0.53} = 3.81$$

$$p - value = P(T_{27} > 3.81 \text{ or } T_{27} < -3.81)$$

$$< 0.01$$

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

Reject H_0, the data provide convincing evidence of a difference between the average aldrin concentrations at bottom and surface.

Practice Problem

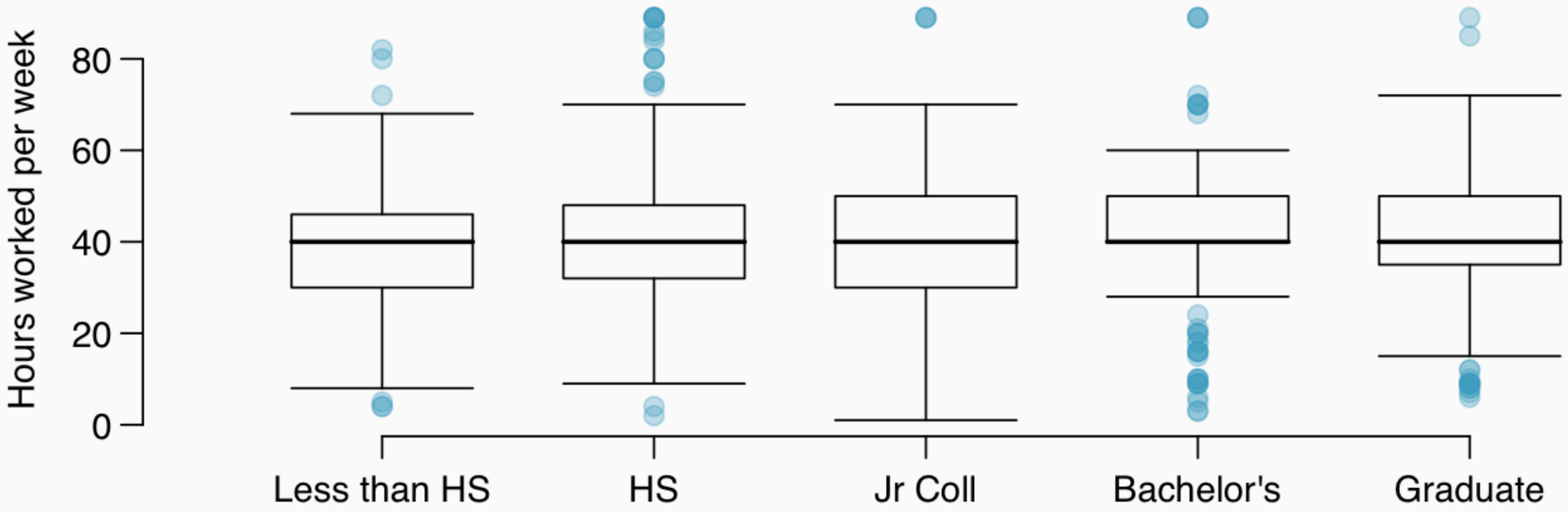
GSS - Hours worked vs Education

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.

GSS - Hours worked vs Education (data)

	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



GSS - Hours worked vs Education (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

GSS - Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree	*4*	*2006.16*	501.54	*2.189*	0.0682
Residuals	*1167*	267382	*229.12*		
Total	*1171*	*269388.16*			