Advanced Statistics
DS2003 (BDS-4A)
Lecture 16

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

14 April, 2022

Previous Lecture

- Difference in two means (t or z test)
  - Example of diamonds (99 carat and 100 carat)
- Difference in more than two means (ANOVA)
  - Example 1: The Wolf River in Tennessee → aldrin concentration at three levels (surface, mid-depth and bottom)
    - Research Question: Is there a difference between the mean aldrin concentrations among the three levels? → To compare means of 3+ groups we use a new test called *ANOVA* and a new statistic called *F*
  - Example 2: Hours worked vs Education

		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: $df_E = df_T - df_G$

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

		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:

$$SSE = SST - SSG$$

$$SSE = 54.29 - 16.96 = 37.33$$

		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 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

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

		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

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

		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:

$$SSE = SST - SSG$$

$$SSE = 54.29 - 16.96 = 37.33$$

		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 error

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

$$MSG = 16.96/2 = 8.48$$
$$MSE = 37.33/27 = 1.38$$

		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

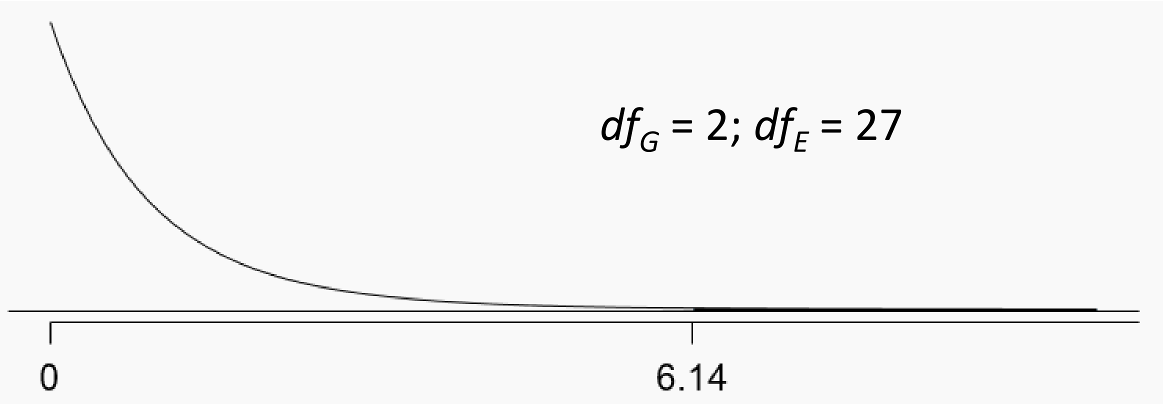
$$F = \frac{MSG}{MSE}$$

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

		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

p-value is 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)

GSS - Hours worked vs Education (ANOVA table)

	DF	Sum Sq	Mean Sq	F value	Prob (> F)
Degrees	*4*	*2006.16*	501.54	*2.189*	0.0682
Residuals	*1167*	267382	*229.12*		
Total	*1171*	*269388.16*			

Cannot reject H_0.

	Educational attainment					
	Less than HS	HS	Jr College	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	1172

Lecture For Today

- Difference in more than two means (ANOVA), continued
- Which group is different?

(1) Independence

Does this condition appear to be satisfied?

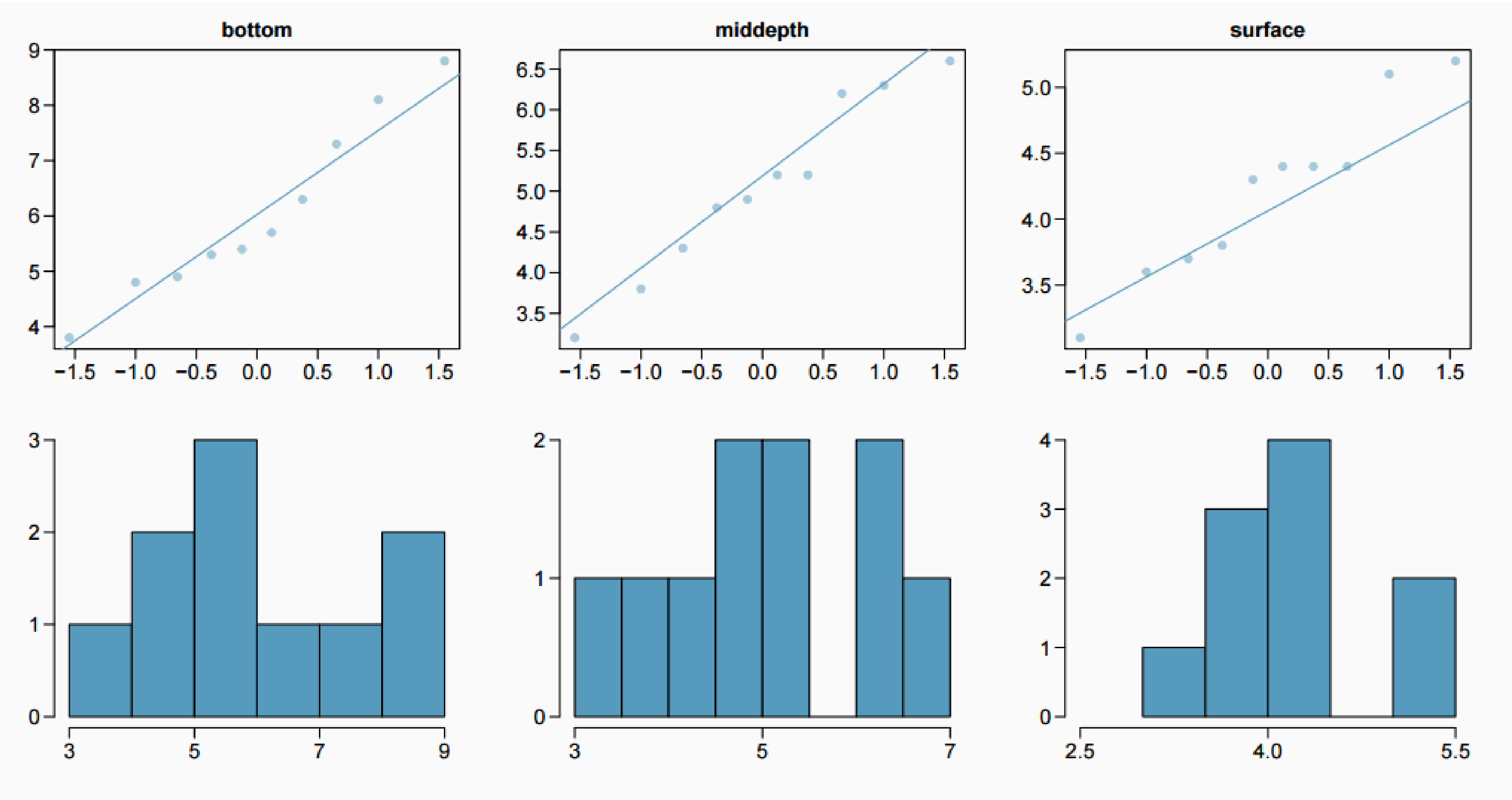
(1) Independence

Does this condition appear to be satisfied?

In this study the we have no reason to believe that the aldrin concentration won't be independent of each other

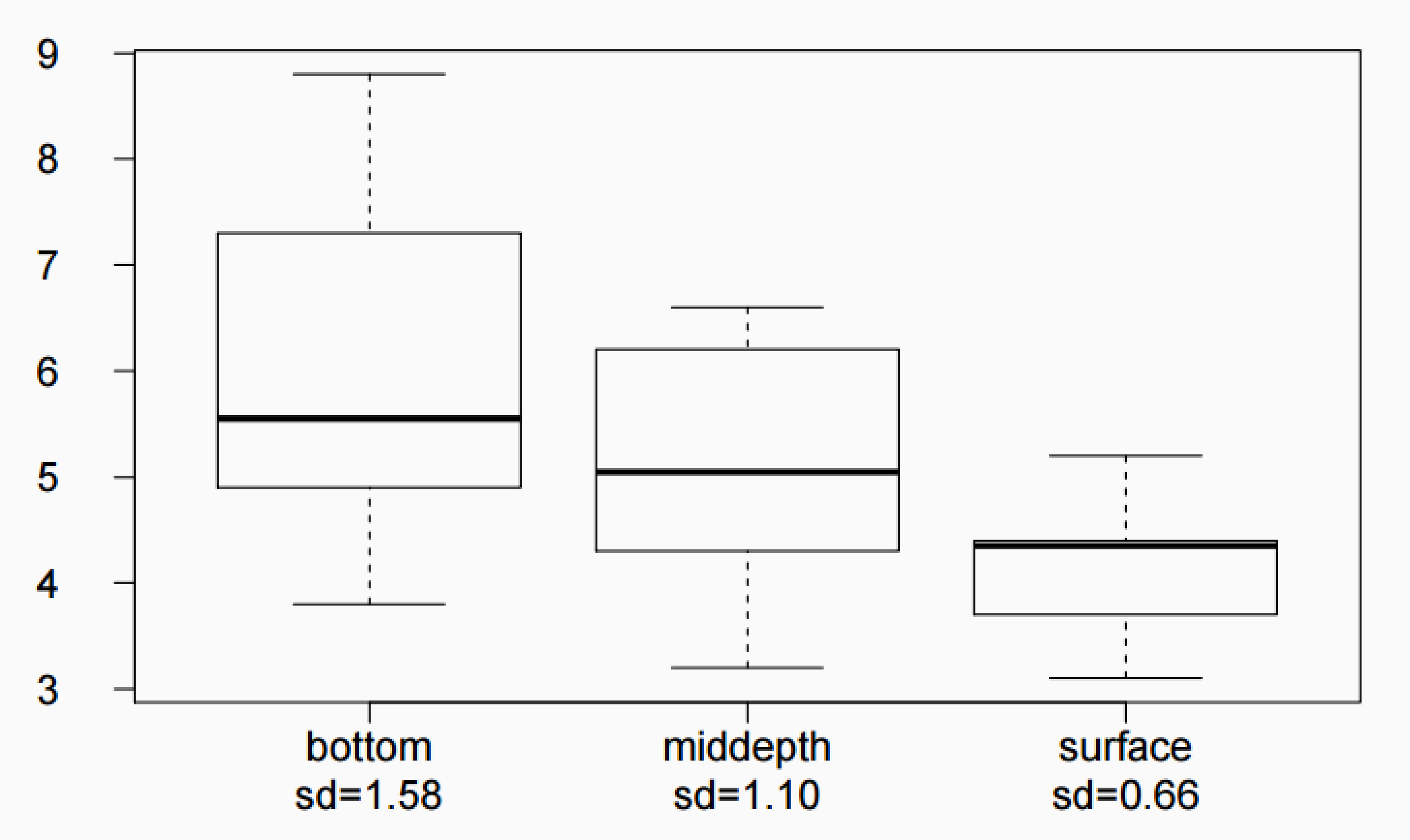
(2) Approximately normal

Does this condition appear to be satisfied?



(3) Constant variance

Does this condition appear to be satisfied?



Which means differ?

- Earlier we concluded that at least one pair of means differ. The natural question that follows is “which ones?”

Which means differ?

- Earlier we 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

Which means differ?

- Earlier we 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?

Which means differ?

- Earlier we 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

Remember:** a type 1 error **means rejecting the null hypothesis when it's actually true

Multiple comparisons

- The scenario of testing many pairs of groups is called multiple comparisons

Multiple comparisons

- The scenario of testing many pairs of groups is called multiple comparisons
- The Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests:

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

where K is the number of comparisons being considered

Multiple comparisons

- The scenario of testing many pairs of groups is called multiple comparisons
- The Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests:

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

where K is the number of comparisons being considered

- If there are k groups, then usually all possible pairs are compared and

$$K = \frac{k(k-1)}{2}$$

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 for two sample t tests for determining which pairs of groups have significantly different means?

A. $\alpha^* = 0.05$
B. $\alpha^* = 0.05/2 = 0.025$
C. $\alpha^* = 0.05/3 = 0.0167$
D. $\alpha^* = 0.05/6 = 0.0083$

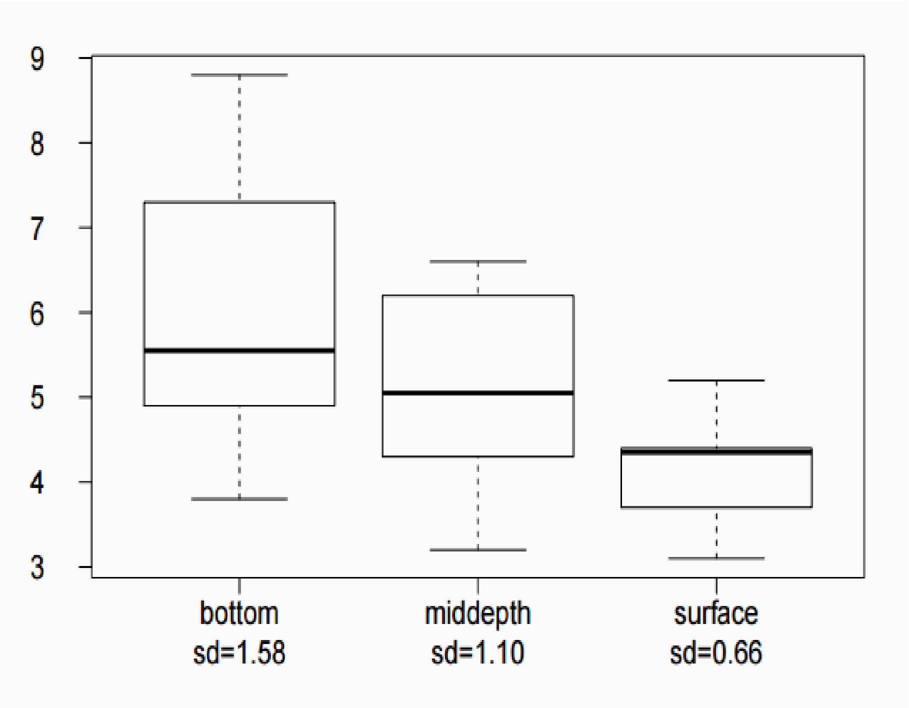
Determining the modified α

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

A. $\alpha^* = 0.05$
B. $\alpha^* = 0.05/2 = 0.025$
C. *$\alpha^* = 0.05/3 = 0.0167$*
D. $\alpha^* = 0.05/6 = 0.0083$

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?

	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

	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			

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

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

	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

	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			

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

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

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

	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

	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			

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

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

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

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

	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

	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			

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

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

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

$$\alpha^* = \frac{0.05}{3} = 0.0167$$

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

	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

	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			

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

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

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

$$\alpha^* = \frac{0.05}{3} = 0.0167$$

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

Pairwise comparisons

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

Pairwise comparisons

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

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

Pairwise comparisons

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

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

$$T_{27} = \frac{(6.04 - 4.2)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{1.84}{0.53} = 3.47$$

Pairwise comparisons

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

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

$$T_{27} = \frac{(6.04 - 4.2)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{1.84}{0.53} = 3.47$$

$$p - value < 0.01 \quad (two - sided)$$

Pairwise comparisons

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

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

$$T_{27} = \frac{(6.04 - 4.2)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{1.84}{0.53} = 3.47$$

$$p - value < 0.01 \quad (two - sided)$$

$$\alpha^* = \frac{0.05}{3} = 0.0167$$

Pairwise comparisons

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

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

$$T_{27} = \frac{(6.04 - 4.2)}{\sqrt{\frac{1.38}{10} + \frac{1.38}{10}}} = \frac{1.84}{0.53} = 3.47$$

$$p - value < 0.01 \quad (two - sided)$$

$$\alpha^* = \frac{0.05}{3} = 0.0167$$

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

More 1-Way ANOVA Tests

Group1	Group2	Group3
48	49	47
51	55	51
49	53	55
45	54	47
53	48	57
46	47	52
51	54	48
45	55	52
53	48	48
53	45	52
50	47	47
45	46	51
48	48	58
45	51	47
45	50	49

Three different set of males, and we are measuring the height they jump in a vertical jump in centimeters

I want to perform a 1-way ANOVA to determine if there is a significant difference between the average height measures of my 3 groups

https://1drv.ms/x/s!Apc0G8okxWJ1y2TfLO0Cj1M7B9G1?e=Y2syQo

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

[One-Way ANOVA: Definition, Formula, and Example - Statology]

Sources

- openintro.org/os (Chapter 7, Section 7.5)

Helpful Links:

- 1-Way ANOVA Tests using Excel:
  - https://www.youtube.com/watch?v=ZvfO7-J5u34
- Jbstatistics:
  - https://www.youtube.com/watch?v=QUQ6YppWCeg
  - https://www.youtube.com/watch?v=WUoVftXvjiQ