Announcements

- Homework
  - Homework and Lab 8 due tonight!
  - Next homework should be posted tomorrow
- In-class lab on Wednesday
- Still working on write up for final projects
- Polling: `rembold-class.ddns.net`

Review Question

What is the appropriate $t^\star$ for a 99% confidence interval for the average difference between two means if you have 25 observations in one sample and 23 in the other?

(a) 1.72

(b) 2.49

(c) 2.51

(d) 2.82

	one tail	0.100	0.050	0.025	0.010	0.005
	two tails	0.200	0.100	0.050	0.020	0.010
df	21	1.32	1.72	2.08	2.52	2.83
	22	1.32	1.72	2.07	2.51	2.82
	23	1.32	1.71	2.07	2.50	2.81
	24	1.32	1.71	2.06	2.49	2.80
	25	1.32	1.71	2.06	2.49	2.79

Review Question

What is the appropriate $t^\star$ for a 99% confidence interval for the average difference between two means if you have 25 observations in one sample and 23 in the other?

(a) 1.72

(b) 2.49

(c) 2.51

(d) 2.82

	one tail	0.100	0.050	0.025	0.010	0.005
	two tails	0.200	0.100	0.050	0.020	0.010
df	21	1.32	1.72	2.08	2.52	2.83
	22	1.32	1.72	2.07	2.51	2.82
	23	1.32	1.71	2.07	2.50	2.81
	24	1.32	1.71	2.06	2.49	2.80
	25	1.32	1.71	2.06	2.49	2.79

Contaminants

- 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.
- The standard methods to test whether these substances are present in a river is to take samples at six-tenths depth.
- But since these compounds are denser than water and their molecules tend to stick to particles of

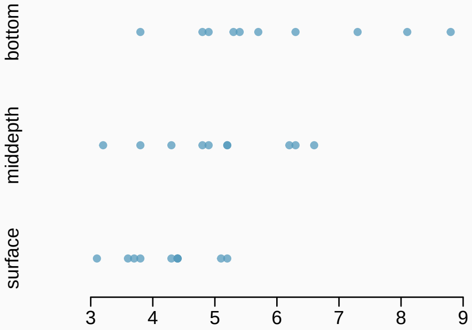
Data

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

	1	2	...	10	11	12	...	20	21	22	...	30
aldrin	3.80	4.80	...	8.80	3.20	3.80	...	6.60	3.10	3.60	...	5.20
depth	bot	bot	...	bot	mid	mid	...	mid	surf	surf	...	surf

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	5.1 0	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 Z or a T statistic.
- To compare means of 3+ groups we use a new test called ***ANOVA*** and a new statistic called ***F***.

ANOVA

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

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 mean is different than others.

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.

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.
    - Especially important when the sample sizes are small.
    - Can check with qq-plots

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.
    - Especially important when the sample sizes are small.
    - Can check with qq-plots
3. The variability across the groups should be about equal.
    - Especially important when the sample sizes differ between groups.
    - Can check with boxplots

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):

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

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):

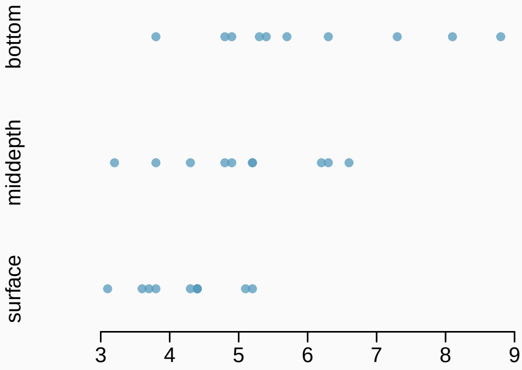
$$F = \frac{\text{variability btw groups}}{\text{variability w/in groups}}$$

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

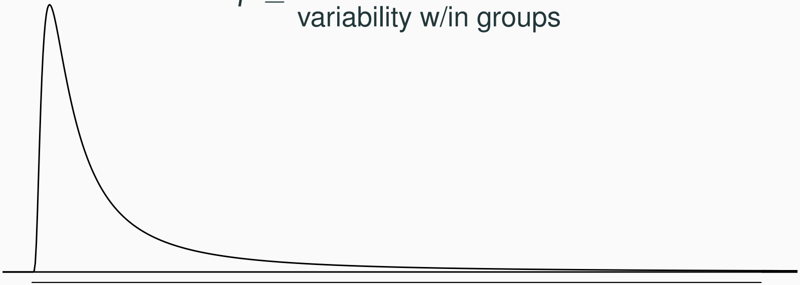
Does there appear to be a lot of variability within groups? How about between groups?

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



F distribution and p-value

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



- 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 - Deconstructed

		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		
	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: $df_E = df_T - df_G$

ANOVA Output - Deconstructed

		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		
	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 - Deconstructed 2

		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		
	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.

ANOVA Output - Deconstructed 2

		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		
	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

ANOVA Output - Deconstructed 2

		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		
	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

$$SSG = \left(10 \times (6.04 - 5.1)^2\right)$$

ANOVA Output - Deconstructed 2

		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		
	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

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

ANOVA Output - Deconstructed 2

		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		
	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

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

ANOVA Output - Deconstructed 2

		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		
	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

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

ANOVA Output - Deconstructed 3

		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		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

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

where x_i represent each observation in the dataset.

ANOVA Output - Deconstructed 3

		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		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

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

where x_i represent each observation in the dataset.

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

ANOVA Output - Deconstructed 3

		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		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

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

where x_i represent each observation in the dataset.

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

ANOVA Output - Deconstructed 3

		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		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

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

where x_i represent each observation in the dataset.

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

ANOVA Output - Deconstructed 3

		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		
	Total	29	54.29			

Sum of squares total, SST

Measures the variability between groups

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

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 - Deconstructed 4

		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		
	Total	29	54.29			

Sum of squares error, SSE

Measures the variability within groups:

$$SSE = SST - SSG$$

ANOVA Output - Deconstructed 4

		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 - Deconstructed 5

		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		
	Total	29	54.29			

Mean square error

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

ANOVA Output - Deconstructed 5

		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		
	Total	29	54.29			

Mean square error

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

$$MSG = 16.96/2 = 8.48$$

ANOVA Output - Deconstructed 5

		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		
	Total	29	54.29			

Mean square error

Mean square error 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 - Deconstructed 6

		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		
	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.

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

ANOVA Output - Deconstructed 6

		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		
	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 - Deconstructed 7

		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		
	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.

ANOVA Output - Deconstructed 7

		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		
	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 - in context

What is the conclusion of the hypothesis test?

The data provide convincing evidence that the average aldrin concentration

A) is different for all groups.
B) on the surface is lower than the other levels.
C) is different for at least one group.
D) is the same for all groups.

Conclusion - in context

What is the conclusion of the hypothesis test?

The data provide convincing evidence that the average aldrin concentration

A) is different for all groups.

B) on the surface is lower than the other levels.

C) is different for at least one group.

D) is the same for all groups.