Advanced Statistics
DS2003 (BDS-4A)
Lecture 15

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

12 April, 2022

Previous Lecture

- More about the t-distribution
  - How it differs from the normal distribution (common rule of thumb)
- The paired t-test
  - Main assumptions
- Example: Instructor giving two exams to students, are they equally difficult?

Lecture For Today

- Difference in two means (t or z tests)
- Difference in more than two means (ANOVA)

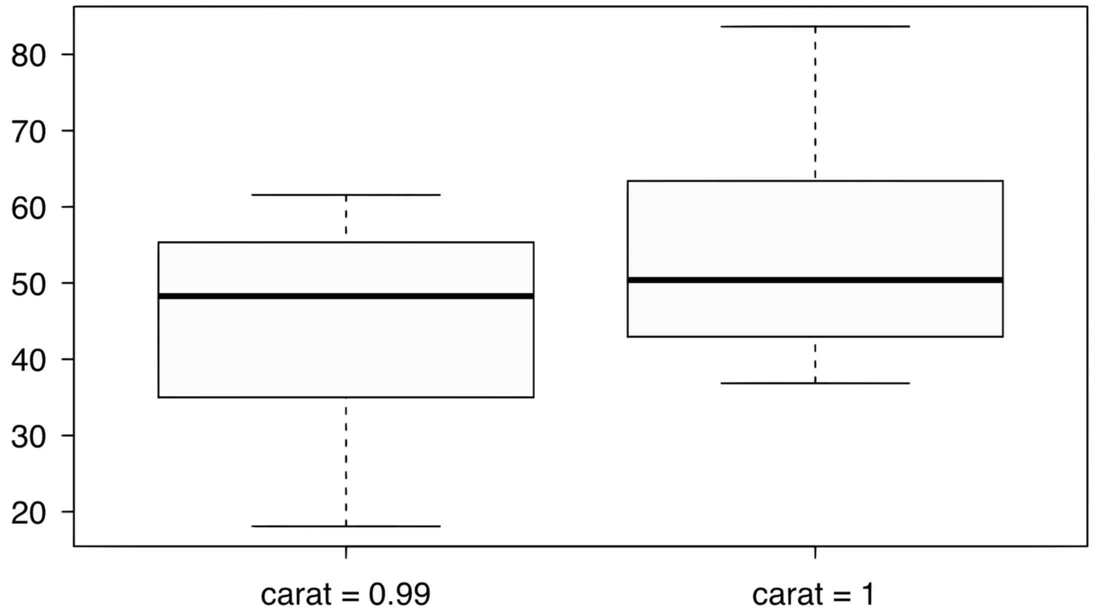
Difference in two means

Slides developed by Mine Çetinkaya-Rundel of OpenIntro
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.
The slides may be copied, edited, and/or shared via the CC BY-SA license
Some images may be included under fair use guidelines (educational purposes)

Diamonds

- Weights of diamonds are measured in carats
- 1 carat = 100 points, 0.99 carats = 99 points, etc.
- The difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but does the price of a 1 carat diamond tend to be higher than the price of a 0.99 diamond?
- We are going to test to see if there is a difference between the average prices of 0.99 and 1 carat diamonds
- In order to be able to compare equivalent units, we divide the prices of 0.99 carat diamonds by 99 and 1 carat diamonds by 100, and compare the average point prices

Data



	0.99 carat	*1 carat*
	pt99	pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

Note: These data are a random sample from the diamonds data set in ggplot2 R package.

Parameter and point estimate

- *Parameter of interest*: Average difference between the point prices of *all* 0.99 carat and 1 carat diamonds

$$\mu_{pt99} - \mu_{pt100}$$

Parameter and point estimate

- *Parameter of interest*: Average difference between the point prices of *all* 0.99 carat and 1 carat diamonds

$$\mu_{pt99} - \mu_{pt100}$$

- *Point estimate*: Average difference between the point prices of *sampled* 0.99 carat and 1 carat diamonds

$$\bar{x}_{pt99} - \bar{x}_{pt100}$$

Hypotheses

Which of the following is the correct set of hypotheses for testing if the average point price of 1 carat diamonds (pt100) is higher than the average point price of 0.99 carat diamonds (pt99)?

A. $H_0 : \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A : \boldsymbol{\mu}_{\text{pt99}} \neq \boldsymbol{\mu}_{\text{pt100}}$

B. $H_0 : \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A : \boldsymbol{\mu}_{\text{pt99}} > \boldsymbol{\mu}_{\text{pt100}}$

C. $H_0 : \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A : \boldsymbol{\mu}_{\text{pt99}} < \boldsymbol{\mu}_{\text{pt100}}$

D. $H_0 : \bar{x}_{pt99} = \bar{x}_{pt100}$

$\bar{x}_{pt99} < \bar{x}_{pt100}$

Hypotheses

Which of the following is the correct set of hypotheses for testing if the average point price of 1 carat diamonds (pt100) is higher than the average point price of 0.99 carat diamonds (pt99)?

A. $H_0: \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A: \boldsymbol{\mu}_{\text{pt99}} \neq \boldsymbol{\mu}_{\text{pt100}}$

B. $H_0: \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A: \boldsymbol{\mu}_{\text{pt99}} > \boldsymbol{\mu}_{\text{pt100}}$

C. $H_0: \boldsymbol{\mu}_{\text{pt99}} = \boldsymbol{\mu}_{\text{pt100}}$

$H_A: \boldsymbol{\mu}_{\text{pt99}} < \boldsymbol{\mu}_{\text{pt100}}$

D. $H_0: \bar{x}_{pt99} = \bar{x}_{pt100}$

$\bar{x}_{pt99} < \bar{x}_{pt100}$

Conditions

Which of the following does not need to be satisfied in order to conduct this hypothesis test using theoretical methods?

A. Point price of one 0.99 carat diamond in the sample should be independent of another, and the point price of one 1 carat diamond should independent of another as well
B. Point prices of 0.99 carat and 1 carat diamonds in the sample should be independent.
C. Distributions of point prices of 0.99 and 1 carat diamonds should not be extremely skewed
D. Both sample sizes should be at least 30

Conditions

Which of the following does not need to be satisfied in order to conduct this hypothesis test using theoretical methods?

A. Point price of one 0.99 carat diamond in the sample should be independent of another, and the point price of one 1 carat diamond should independent of another as well
B. Point prices of 0.99 carat and 1 carat diamonds in the sample should be independent.
C. Distributions of point prices of 0.99 and 1 carat diamonds should not be extremely skewed
D. *Both sample sizes should be at least 30*

Test statistics

Test statistic for inference on the difference of two small sample means

The test statistic for inference on the difference of two means where σ_1 and σ_2 are unknown is the *T* statistic.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

where

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad \text{and} \qquad df = min(n_1 - 1, n_2 - 1)$$

Note: The calculation of the *df* is actually much more complicated. For simplicity we'll use the above formula to <u>estimate</u> the true df when conducting the analysis by hand

Test statistics (cont.)

	0.99 carat	*1 carat*
	pt99	pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

In context...

Test statistics (cont.)

	0.99 carat	*1 carat*
	pt99	pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

In context...

$$T = \frac{\text{point estimate} - \text{null value}}{SE}$$

Test statistics (cont.)

	0.99 carat	*1 carat*
	pt99	pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

In context...

$$T = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{(44.50 - 53.43) - 0}{\sqrt{\frac{13.32^2}{23} + \frac{12.22^2}{30}}}$$

Test statistics (cont.)

	0.99 carat pt99	*1 carat* pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

In context...

$$\begin{aligned} T &= \frac{\text{point estimate} - \text{null value}}{SE} \\ &= \frac{(44.50 - 53.43) - 0}{\sqrt{\frac{13.32^2}{23} + \frac{12.22^2}{30}}} \\ &= \frac{-8.93}{3.56} \end{aligned}$$

Test statistics (cont.)

	0.99 carat pt99	*1 carat* pt100
$\bar{x}$	44.50	53.43
s	13.32	12.22
n	23	30

In context...

$$
\begin{aligned}
T &= \frac{\text{point estimate} - \text{null value}}{SE} \\
&= \frac{(44.50 - 53.43) - 0}{\sqrt{\frac{13.32^2}{23} + \frac{12.22^2}{30}}} \\
&= \frac{-8.93}{3.56} \\
&= -2.508
\end{aligned}
$$

Test statistics (cont.)

Which of the following is the correct *df* for this hypothesis test?

A. 22
B. 23
C. 30
D. 29
E. 52

Test statistics (cont.)

Which of the following is the correct *df* for this hypothesis test?

A. *22*
B. 23
C. 30
D. 29
E. 52

$$df = \min(n_{pt99} - 1, n_{pt100} - 1)$$
$$= \min(23 - 1, 30 - 1)$$
$$= \min(22, 29)$$

p-value

Which of the following is the correct p-value for this hypothesis test?

$$T = -2.508 \qquad df = 22$$

A. between 0.005 and 0.01
B. between 0.01 and 0.025
C. between 0.02 and 0.05
D. between 0.01 and 0.02

p-value

Which of the following is the correct p-value for this hypothesis test?

$$T = -2.508 \qquad\qquad df = 22$$

A. between 0.005 and 0.01
B. *between 0.01 and 0.025*
C. between 0.02 and 0.05
D. between 0.01 and 0.02

```
> pt(q = -2.508, df = 22)
[1] 0.0100071
```

Synthesis

What is the conclusion of the hypothesis test? How (if at all) would this conclusion change your behavior if you went diamond shopping?

Synthesis

What is the conclusion of the hypothesis test? How (if at all) would this conclusion change your behavior if you went diamond shopping?

- p-value is small so reject H_0. The data provide convincing evidence to suggest that the point price of 0.99 carat diamonds is lower than the point price of 1 carat diamonds
- Maybe buy a 0.99 carat diamond? It looks like a 1 carat, but is significantly cheaper

Comparing means with ANOVA

Slides developed by Mine Çetinkaya-Rundel of OpenIntro
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.
The slides may be copied, edited, and/or shared via the CC BY-SA license
Some images may be included under fair use guidelines (educational purposes)

- 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)

- 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 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

- 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 sediment, they are more likely to be found in higher concentrations near the bottom

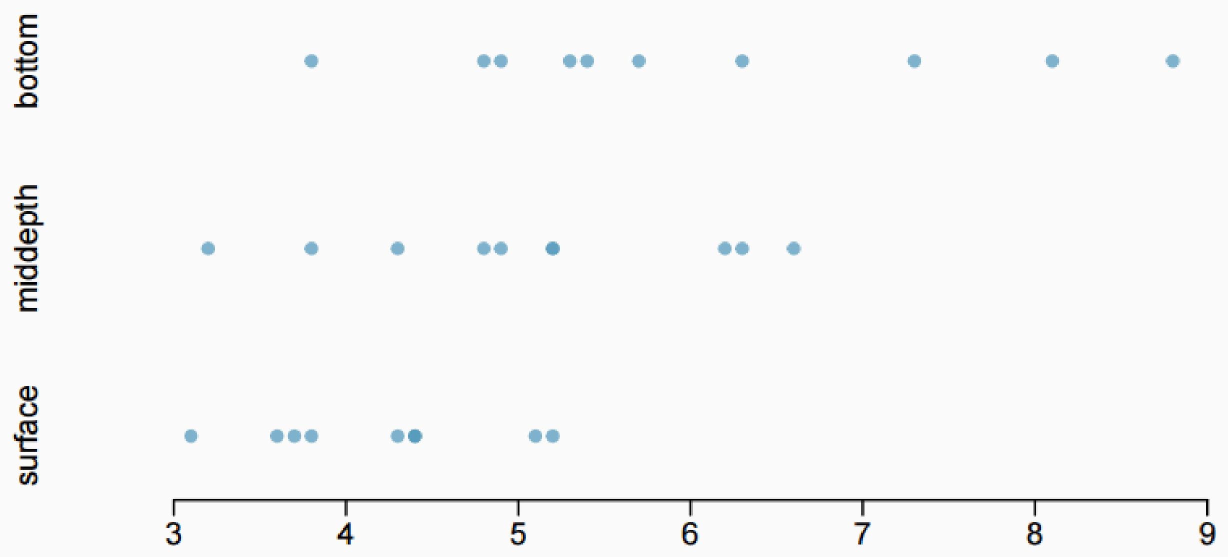
Data

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	5.1 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

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,

$$\mu_1 = \mu_2 = \ldots = \mu_k,$$

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

How do we check for normality?

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

How do we check for normality?

3. The variability across the groups should be about equal
   - Especially important when the sample sizes differ between groups

How can we check this condition?

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

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

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)

$$F = \frac{variability\ bet.groups}{variability\ within\ groups}$$

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)

$$F = \frac{variability\ bet.\ groups}{variability\ within\ groups}$$

- 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

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

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

Hypotheses

A. $H_0 : \mu_B = \mu_M = \mu_S$

$H_A : \mu_B \neq \mu_M \neq \mu_S$

B. $H_0 : \mu_B \neq \mu_M \neq \mu_S$

$H_A : \mu_B = \mu_M = \mu_S$

C. $H_0 : \mu_B = \mu_M = \mu_S$

H_A : At least one mean is different

D. $H_0 : \mu_B = \mu_M = \mu_S = 0$

H_A : At least one mean is different

E. $H_0 : \mu_B = \mu_M = \mu_S$

$H_A : \mu_B > \mu_M > \mu_S$

Hypotheses

A. $H_0 : \mu_B = \mu_M = \mu_S$

$H_A : \mu_B \neq \mu_M \neq \mu_S$

B. $H_0 : \mu_B \neq \mu_M \neq \mu_S$

$H_A : \mu_B = \mu_M = \mu_S$

C. $H_0 : \mu_B = \mu_M = \mu_S$

H_A : *At least one mean is different*

D. $H_0 : \mu_B = \mu_M = \mu_S = 0$

H_A : At least one mean is different

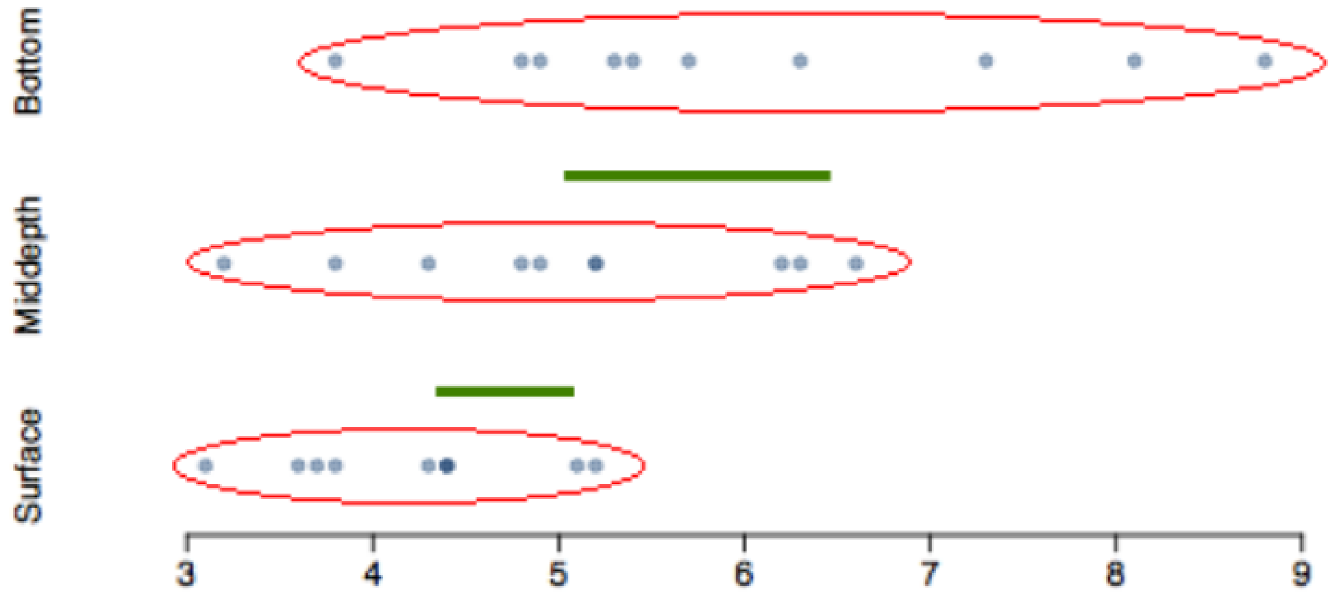
E. $H_0 : \mu_B = \mu_M = \mu_S$

$H_A : \mu_B > \mu_M > \mu_S$

Test statistic

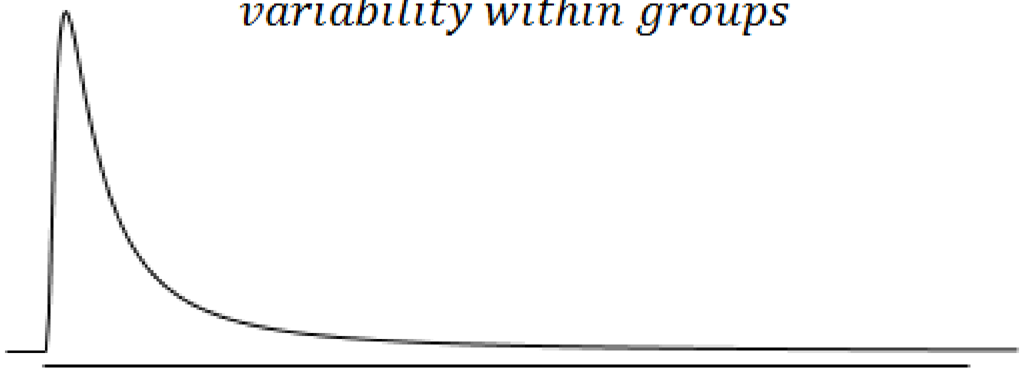
Does there appear to be a lot of variability within groups? How about between groups?

$$F = \frac{variability\ bet.\ groups}{variability\ within\ groups}$$



F distribution and p-value

$$F = \frac{\textit{variability bet. groups}}{\textit{variability within groups}}$$



- 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

		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	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	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	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)
(*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

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

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

		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, $\bar{x}_i$ ıe average for each group, is tł $\bar{x}$ overall (grand) mean

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

		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, $\bar{x}_i$ ıe average for each group, is tł $\bar{x}$ overall (grand) mean

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

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

		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, $\bar{x}_i$ ıe average for each group, is th $\bar{x}$ overall (grand) mean

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

$$SSG = (10 \times (6.04 - 5.1)^2)$$
$$+ (10 \times (5.05 - 5.1)^2)$$

		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, $\bar{x}_i$ ıe average for each group, is th $\bar{x}$ overall (grand) mean

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

$$SSG = (10 \times (6.04 - 5.1)^2) + (10 \times (5.05 - 5.1)^2) + (10 \times (4.2 - 5.1)^2)$$

		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, $\bar{x}_i$ ıe average for each group, is tł $\bar{x}$ overall (grand) mean

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

$$SSG = (10 \times (6.04 - 5.1)^2) + (10 \times (5.05 - 5.1)^2) + (10 \times (4.2 - 5.1)^2)$$

		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, $\bar{x}_i$ ıe average for each group, is tl $\bar{x}$ 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)
(*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

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

where x_i represent each observation in the dataset

		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

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

		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

$$SST = (3.8 - 5.1)^2 + (4.8 - 5.1)^2 + (4.9 - 5.1)^2 + \ldots + (5.2 - 5.1)^2$$
$$= (-1.3)^2 + (-0.3)^2 + (-0.2)^2 + \ldots + (0.1)^2$$

		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

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

		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

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

		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:

$$SSE = SST - SSG$$

		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

		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

$$MSG = 16.96/2 = 8.48$$

		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)
(*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

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

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

		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			

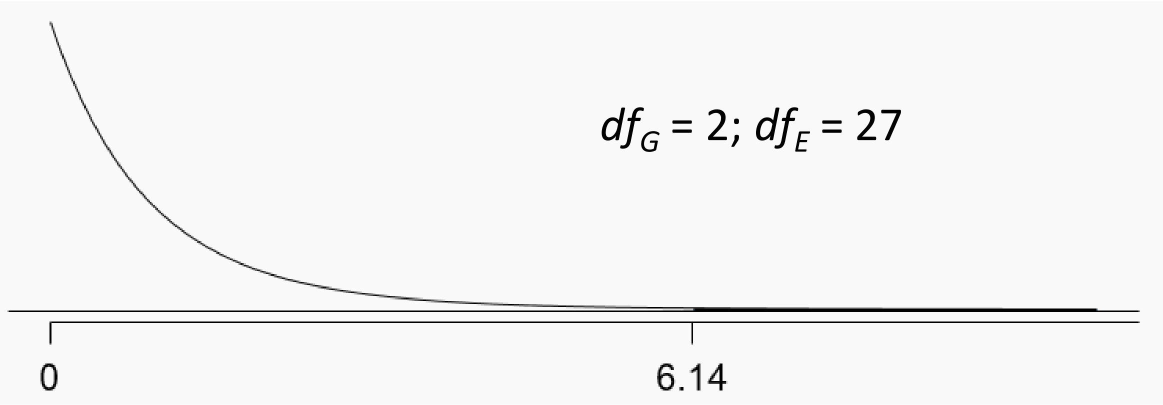
p-value

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.

		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 - 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

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)

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)

Practice: 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 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

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

Sources

- openintro.org/os (Chapter 7, Section 7.3 and Section 7.5)
- http://www2.stat.duke.edu/~cr173/Sta102_Fa15/Lec/Lec15.pdf

Helpful Links: