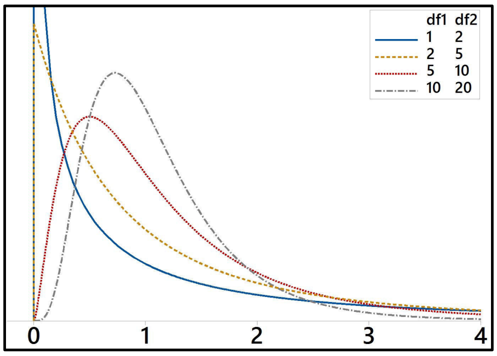
ONE-FACTOR ANOVA AND MULTIPLE COMPARISONS

- F-Distribution
- Sources of Variation
- One-Factor ANOVA Test
- Multiple Comparisons

University of Pittsburgh

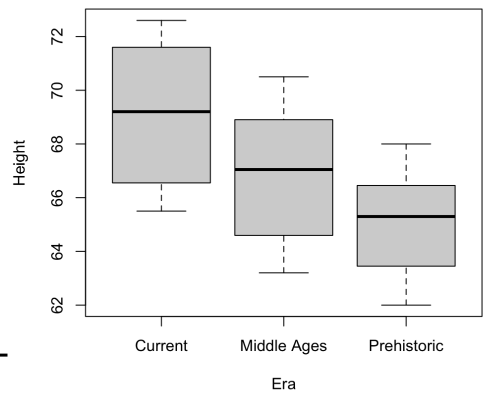
F-DISTRIBUTION

- **F-distribution:** continuous probability distribution that has the following properties:
  - Unimodal and right-skewed
  - Always non-negative
  - Two parameters for degrees of freedom
    - One for numerator and one for denominator
  - Used to compare the ratio of two sources of variability
  - **Key Fact:** Large F-statistics are evidence against the null hypothesis.



MOTIVATION: ONE-FACTOR ANOVA

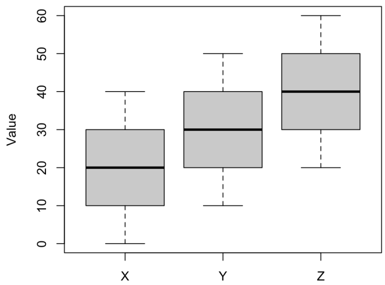
- **Scenario:** Archaeologists want to analyze the skeletal remains of 8 males from each of three time periods (current era, Middle Ages, and Prehistoric Era)
- **Question:** Are the mean heights of the skeletal remains equal across all three eras?
- **Answer:** From the side-by-side boxplots:
  - Means appear to be _______________
  - But there is a great amount of _________ within each group
  - Sample size is ____________
- **Takeaway:** Need an inferential technique that compares ______________________



One-Factor Analysis of Variance (ANOVA)

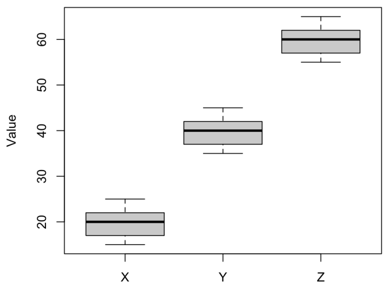
- **One-factor analysis of variance (ANOVA):** statistical technique used to compare the means of three or more populations
  - Uses two sources of variability to compare means
  - **Between group variation:** measures that amount of variability between the sample means of individual groups
    - "How different are the sample means from one another?"
  - **Within group variation:** measures the amount of variability that exists within the samples
    - "How different are the individual observations from one another within each group?"

Comparing Types of Variation



Small between group variation
- Means are ________________ (20, 30, and 40)

Large within group variation
- Observations within groups are ___________
- Range for each sample is about ___



Large between group variation
- Means are ___________ (20, 40, and 60)

Small within group variation
- Observations within groups are ________________
- Range for each sample is about ___

One-Factor ANOVA: Hypotheses and Conditions

Step	Description
Used for	Comparing the means of three or more populations
Hypotheses	$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ H_A: At least two means are significantly different
Conditions	1. Sampling distributions of all sample means must be approximately normal 2. The spreads of all samples must be approximately equal *$\frac{s_L^2}{s_S^2} < 2$ where s_L represents the largest sample standard deviation and s_S is the smallest sample standard deviation*

Example: One-Factor ANOVA

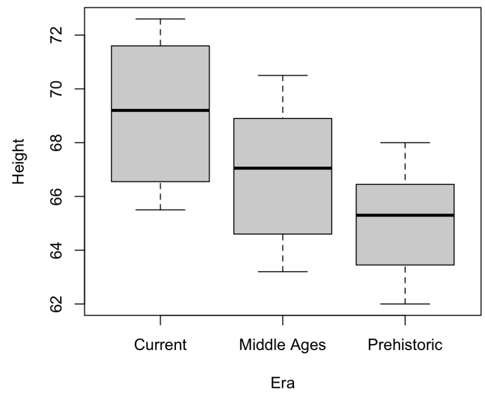
- **Scenario:** Archaeologists want to analyze the skeletal remains of 8 males from each of three time periods (current era, Middle Ages, and Prehistoric Era)
- **Question:** Are the mean heights of the skeletal remains equal across all three eras?
- **Hypotheses:**
  - H_0: ______________________
  - H_A: At least _____________________ are __________________________

Example: One-Factor ANOVA (Cont.)

- **Question:** Are the mean heights of the skeletal remains equal across all three eras?

Statistic	Current	Middle Ages	Prehistoric
Mean	69.10	66.85	65.05
Std. Dev.	2.842	2.644	2.056

- **Conditions:**
  - **Normality:** _____________
    - Sample size of 8 is _______, but boxplots are all _____________________
  - **Equal Spread:** _____________
    - **Largest SD:** ___________
    - **Smallest SD:** ___________
    - **Ratio:** _________ = ______ < 2



Grand Mean

- **Grand mean:** the mean of all observations, disregarding the group from which the observations were sampled
  - Used in the calculation of the between group variation because it helps us understand how different the sample means are.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

where

- $\bar{x}_i$ is the mean of the observations from group i
- n_i is the number of observations sampled from group i

Sources of Variation

- **Between Group Variation:** How different are the sample means?

$$SST = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

Sample Size   Sample Mean

- **Within Group Variation:** How different are the observations within each group?

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2$$

Sample Size   Sample Variance

- **Total Variation:** $SSY = SST + SSE$

Example: Sources of Variation

Stat	Current	Mid. Ages	Prehistoric
Mean	69.10	66.85	65.05
SD	2.842	2.644	2.056
n	8	8	8

- **Grand mean:**

$$\bar{\bar{x}} = \frac{8(69.1) + 8(66.85) + 8(65.05)}{8 + 8 + 8} = \_\_\_\_$$

- **Between group variation:**

$$SST = 8(69.1 - 67)^2 + 8(66.85 - 67)^2 + 8(65.05 - 67)^2$$
$$= \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$
$$= \_\_\_\_\_\_$$

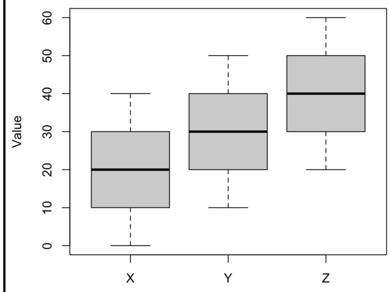
- **Within group variation:**

$$SSE = (8 - 1)(2.842)^2 + (8 - 1)(2.644)^2 + (8 - 1)(2.056)^2$$
$$= \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$
$$= \_\_\_\_\_\_\_$$

One-Factor ANOVA: Test Statistic and Conclusion

Step	Description
Mean Squared Treatment	$MST = \frac{SST}{k-1}$ ("Average between group error per group")
Mean Squared Error	$MSE = \frac{SSE}{n-k}$ ("Average within group error per observation")
Test Statistic	$F_{k-1,n-k} = \frac{MST}{MSE}$ ("Ratio of average between group variation to average within group variation")
Conclusion	**If H_0 is not rejected:** None of the means are significantly different **If H_0 is rejected:** At least two means are significantly different

Comparing Sources of Variation with F-Statistic



Small between group variation

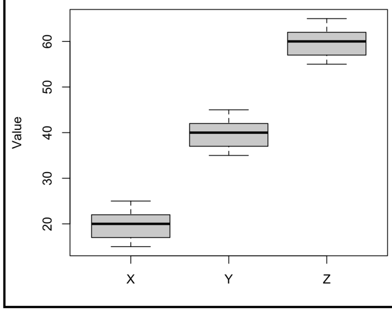
- Means are close together → __________

Large within group variation

- Observations within groups are far apart → __________

F-statistic $\left(\frac{MST}{MSE}\right)$ **will be** _________

- Evidence supporting ____________________



Large between group variation

- Means are far apart → __________

Small within group variation

- Observations within groups are close together → __________

F-statistic $\left(\frac{MST}{MSE}\right)$ **will be** _________

- Evidence supporting _______________________________

ANOVA Table and Code

- **ANOVA table:** summary of the sums of squares, degrees of freedom, mean squared terms and test statistic from an ANOVA

Source	DF	Sums of Squares	Mean Squares	Test Statistic
Between Group	$k-1$	SST	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
Within Group	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	SSY		

- To run a one-factor ANOVA in R and print the ANOVA table:

```
model = aov(height ~ era, data = heights)
summary(model)
```


Example: Test Statistic and ANOVA Table

- **ANOVA Table:**

Source	DF	Sums of Squares	Mean Squares	Test Statistic	P-Value
Between Group	2	65.88	32.94	5.122	0.0154
Within Group	21	135.06	6.43		
Total	23	200.94			

- **Test Statistic:** $F =$ _______
- **P-Value:** $p = 0.0154$
- **Conclusion:** ___________ and conclude that the average heights of the skeletal remains for ____ _____________.

Drawback of ANOVA

- When we reject H_0, the only conclusion that can be drawn at that point is that at least two means are not equal.
- **Problem:** ________________ of rejecting H_0
  - Don't know how many pairs of means differ or which pairs differ
    - $\boxed{\mu_1 \neq \mu_2}, \mu_1 = \mu_3, \mu_2 = \mu_3$
    - $\mu_1 = \mu_2, \boxed{\mu_1 \neq \mu_3}, \mu_2 = \mu_3$
    - $\mu_1 = \mu_2, \mu_1 = \mu_3, \boxed{\mu_2 \neq \mu_3}$

    ______ pair of means not equal

    - $\boxed{\mu_1 \neq \mu_2}, \boxed{\mu_1 \neq \mu_3}, \mu_2 = \mu_3$
    - $\boxed{\mu_1 \neq \mu_2}, \mu_1 = \mu_3, \boxed{\mu_2 \neq \mu_3}$
    - $\mu_1 = \mu_2, \boxed{\mu_1 \neq \mu_3}, \boxed{\mu_2 \neq \mu_3}$

    ______ pairs of means not equal

    - $\mu_1 \neq \mu_2 \neq \mu_3$

    ________ are equal

Number of comparisons increases exponentially as the number of groups increases.

- **Solution:** ____________________ will tell us which scenario is true.

Multiple Comparisons

- **Multiple comparisons:** procedure used to determine exactly which pairs of means are significantly different
  - Extension of ANOVA
  - Perform a hypothesis test interval for each pair of means, but…
  - …make adjustment to the level of significance based on how many comparisons need to be made
    - Want the overall Type I error rate to be 5% in total
    - Level of significance for each individual test must be smaller than 5%
  - Many different techniques
    - Fisher's Least Significant Difference Method
    - Tukey's Method of Multiple Comparisons
    - Bonferroni Adjustment Method

Bonferroni Adjustment Method

- **Bonferroni adjustment method:** a method of multiple comparisons where a test statistic is calculated for each pairwise comparison being done, but where the level of significance is adjusted according to how many comparisons are being made
  - **Hypotheses:** $H_0: \mu_i = \mu_j$ vs. $H_A: \mu_i \neq \mu_j$ when comparing groups i **and** j
  - **Test Statistic:**

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \text{ which has } n - k \text{ df}$$

  - **Decision:** If the overall level of significance is α and the number of comparisons being made is k, then each multiple comparisons test is made using a level of significance of $\frac{\alpha}{k}$.

Example: Preparing for Multiple Comparisons

- **Scenario:** Archaeologists want to analyze the skeletal remains of 8 males from each of three time periods (Prehistoric, Middle Ages, and current era)
- **Question:** How many comparisons do we need to make?
- **Answer:** ___
  - ________ vs. ___________
  - ________ vs. ___________
  - ___________ vs. ___________
- **Question:** If we want to use an overall 5% level of significance, what should the level of significance be for each individual test?
- **Answer:** _________________

Obtaining P-Values Using R

- After running the one-way ANOVA test, multiple comparisons p-values can be obtained:

Quantitative variable

Categorical variable

```
pairwise.t.test(heights$height, heights$era,
    p.adjust = "none")
```

Tells R not to use a different method of multiple comparisons

- R will output a table of p-values that contains the result of every comparison:

```
            Current Middle Ages
Middle Ages 0.0905  -
Prehistoric 0.0044  0.1704
```


Example: Multiple Comparisons

- **Task:** Test if the average heights of the skeletal remains from the current era and Middle Ages are significantly different.

Stat	Current	Mid. Ages	Prehistoric
Mean	69.10	66.85	65.05
SD	2.842	2.644	2.056
n	8	8	8

```
          Df Sum Sq Mean Sq F value Pr(>F)
era        2  65.88   32.94   5.122 0.0154 *
Residuals 21 135.06    6.43
```

```
            Current Middle Ages
Middle Ages 0.0905  -
Prehistoric 0.0044  0.1704
```

- **Hypotheses:** H_0: ___________ vs. H_A: ___________
- **Test Statistic:** $t =$ _____________ $=$ _________
- **P-Value:** $p =$ _________

(because _______ > _______)

- **Conclusion:** ______________ and conclude that the mean heights for the current era and Middle Ages are _________________________.

EXAMPLE: MULTIPLE COMPARISONS

- **Task:** Test if the average heights of the skeletal remains from the current era and Prehistoric Era are significantly different.

Stat	Current	Mid. Ages	Prehistoric
Mean	69.10	66.85	65.05
SD	2.842	2.644	2.056
n	8	8	8

```
          Df Sum Sq Mean Sq F value Pr(>F)
era        2  65.88   32.94   5.122 0.0154 *
Residuals 21 135.06    6.43
```

```
            Current Middle Ages
Middle Ages 0.0905  -
Prehistoric 0.0044  0.1704
```

- **Hypotheses:** H_0: __________ vs. H_A: __________
- **Test Statistic:** $t =$ ____________ $=$ ________
- **P-Value:** $p =$ ________ (because ______ < ______)
- **Conclusion:** ________ and conclude that the mean heights for the current era and Prehistoric Era are ____________________.

EXAMPLE: MULTIPLE COMPARISONS

- **Task:** Test if the average heights of the skeletal remains from the Middle Ages and Prehistoric Era are significantly different.

Stat	Current	Mid. Ages	Prehistoric
Mean	69.10	66.85	65.05
SD	2.842	2.644	2.056
n	8	8	8

```
          Df Sum Sq Mean Sq F value Pr(>F)
era        2  65.88   32.94   5.122 0.0154 *
Residuals 21 135.06    6.43
```

```
            Current Middle Ages
Middle Ages 0.0905  -
Prehistoric 0.0044  0.1704
```

- **Hypotheses:** H_0: __________ vs. H_A: __________
- **Test Statistic:** $t =$ ____________ $=$ ________
- **P-Value:** $p =$ ________ (because ______ > ______)
- **Conclusion:** _____________ and conclude that the mean heights for the Middle Ages and Prehistoric Era are ______________________.

EXAMPLE: DIFFERENT SAMPLE SIZES

- **Question:** What would happen to the multiple comparisons tests if the observations were distributed differently?
- **Answer:** ______________ would be different in each comparison

Original Data

Stat.	Current	Mid. Ages	Prehistoric
SD	2.842	2.644	2.056
n	8	8	8
MSE	6.43		

All SE are $\sqrt{6.43\left(\frac{1}{8}+\frac{1}{8}\right)} =$ ____ regardless of the comparison being made.

Different Sample Sizes

Stat	Current	Mid. Ages	Prehistoric
SD	2.842	2.644	2.056
n	10	8	6
MSE	$[9(2.842)^2 + 7(2.644)^2 + 5(2.056)^2]/21 = 6.80$		

Current vs. Mid. Ages: $\sqrt{6.80\left(\frac{1}{10}+\frac{1}{8}\right)} =$ ____

Current vs. Prehistoric: $\sqrt{6.80\left(\frac{1}{10}+\frac{1}{6}\right)} =$ ____

Mid. Ages vs. Prehistoric: $\sqrt{6.80\left(\frac{1}{8}+\frac{1}{6}\right)} =$ ____

EXAMPLE: FAILING TO REJECT ANOVA

- **Scenario:** Suppose the heights for the current era were 1" shorter, resulting in summary statistics and test results shown below.

Stat	Current	Mid. Ages	Prehistoric
Mean	68.10	66.85	65.05
SD	2.842	2.644	2.056
n	8	8	8

```
          Df Sum Sq Mean Sq F value Pr(>F)
era        2  37.61  18.807   2.924 0.0758 .
Residuals 21 135.06   6.431
```

```
            Current Middle Ages
Middle Ages 0.335   -
Prehistoric 0.025   0.170
```

- **Question:** What conclusions could we draw?
- **Answer:**
  - **ANOVA:** ______________ and conclude _____________________ are significantly different
  - **Multiple Comparisons:** Would ________________ to perform the tests
    - ANOVA already told us that _______________________
    - All p-values are _______________