Chapter 13: Analysis of Variance

DSCC 462

Computational Introduction to Statistics

Anson Kahng

Fall 2022

Some Midterm Comments

Some Midterm Comments

- Trends:

Some Midterm Comments

- Trends:
  - Some people did **well**

Some Midterm Comments

- Trends:
    - Some people did **well**
    - Some people did **poorly**

Some Midterm Comments

- Trends:
  - Some people did **well**
  - Some people did **poorly**
  - Some people **cheated…**

Some Midterm Comments

- Trends:
  - Some people did **well**
  - Some people did **poorly**
  - Some people **cheated…**
- Actions:

Some Midterm Comments

- Trends:
  - Some people did **well**
  - Some people did **poorly**
  - Some people **cheated…**
- Actions:
  - There will almost surely be a **curve** (yet to be determined)

Some Midterm Comments

- Trends:
  - Some people did **well**
  - Some people did **poorly**
  - Some people **cheated…**
- Actions:
  - There will almost surely be a **curve** (yet to be determined)
  - If you cheated, prepare to hear from me next week

Some Assignment Comments

Some Assignment Comments

- Update to homework schedule:

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)
- Project:

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)
- Project:
  - Think about project groups (3-4 people per group)

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)
- Project:
  - Think about project groups (3-4 people per group)
  - Project description will be released next Tuesday, November 15

Some Assignment Comments

- Update to homework schedule:
  - 6 total assignments, not 7 (= you get 100% on the 7th assignment)
  - HW 5 released: next Thursday, November 17
  - HW 5 due: Thursday, December 6 (2 weeks to complete it)
- Project:
  - Think about project groups (3-4 people per group)
  - Project description will be released next Tuesday, November 15
  - Groups by next Thursday, November 18

Plan for Today

Plan for Today

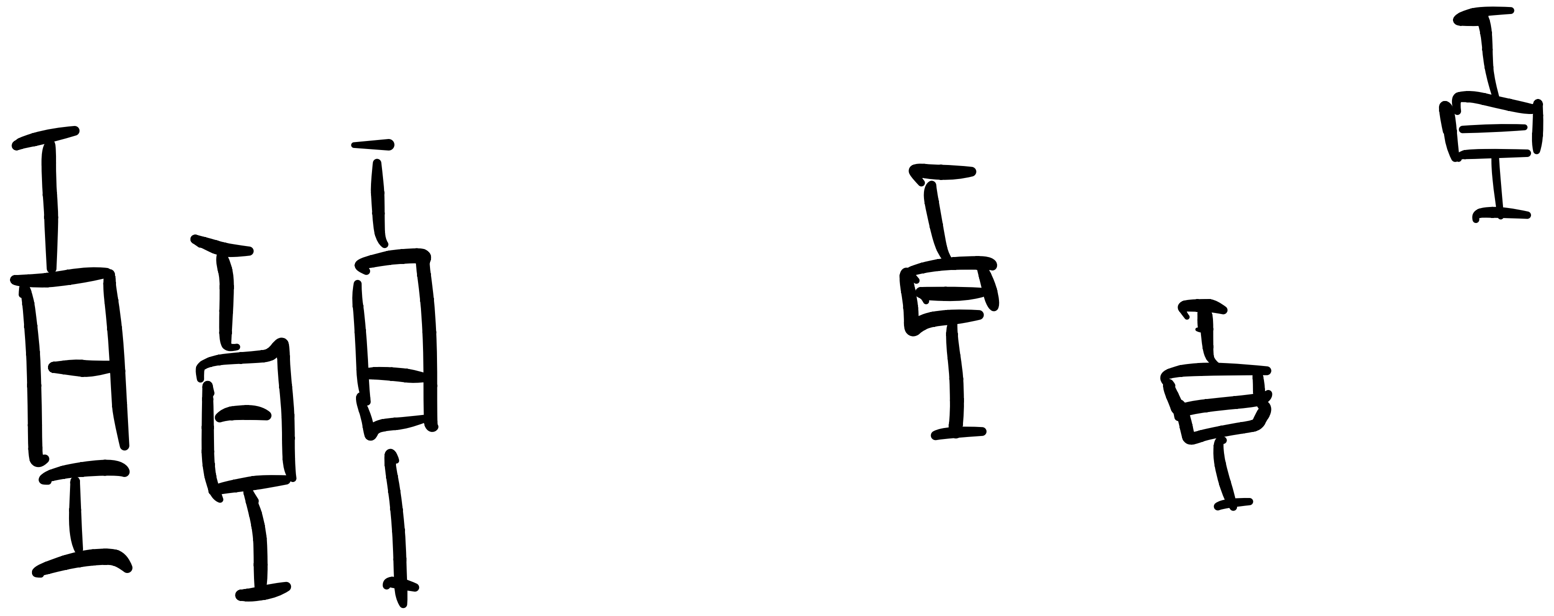
- How do we compare sample means for more than two groups?

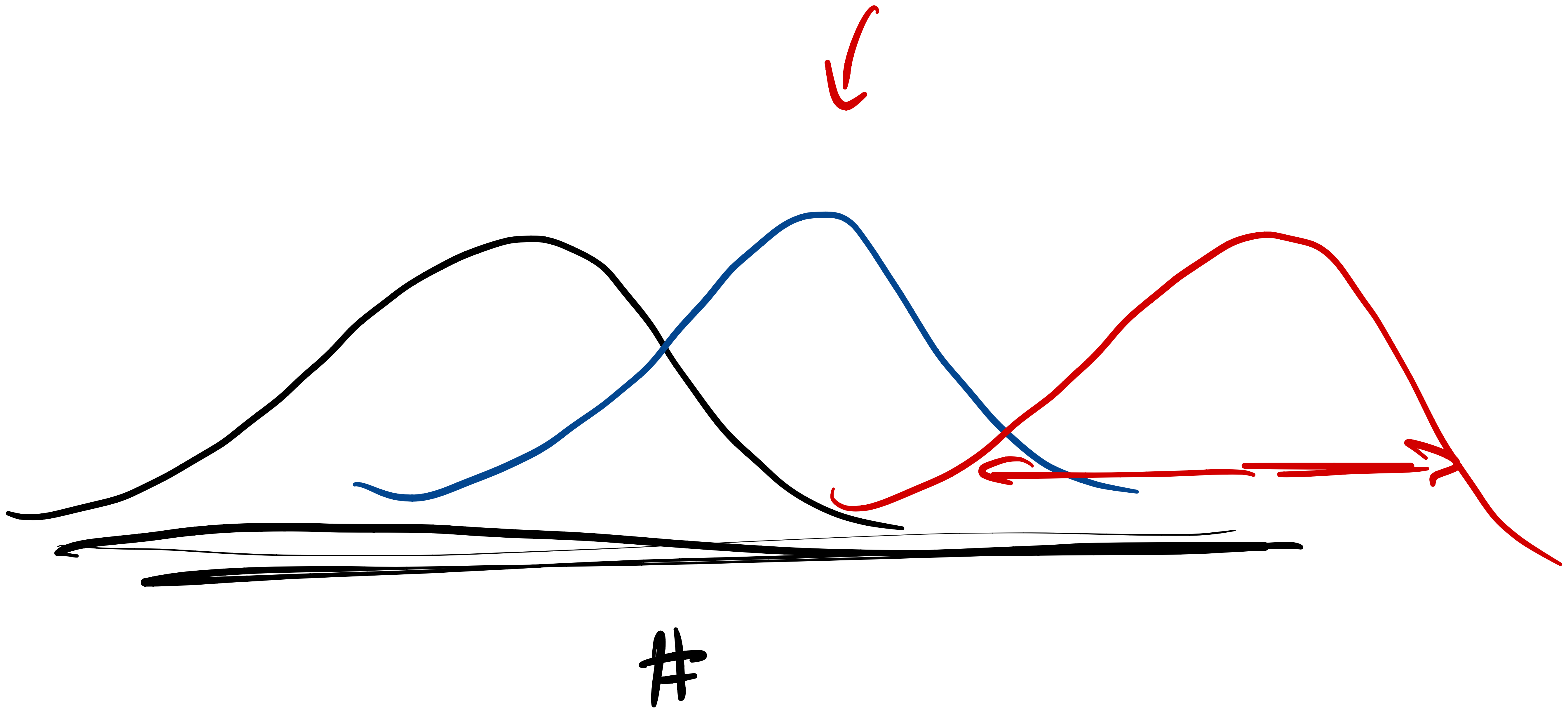
Plan for Today

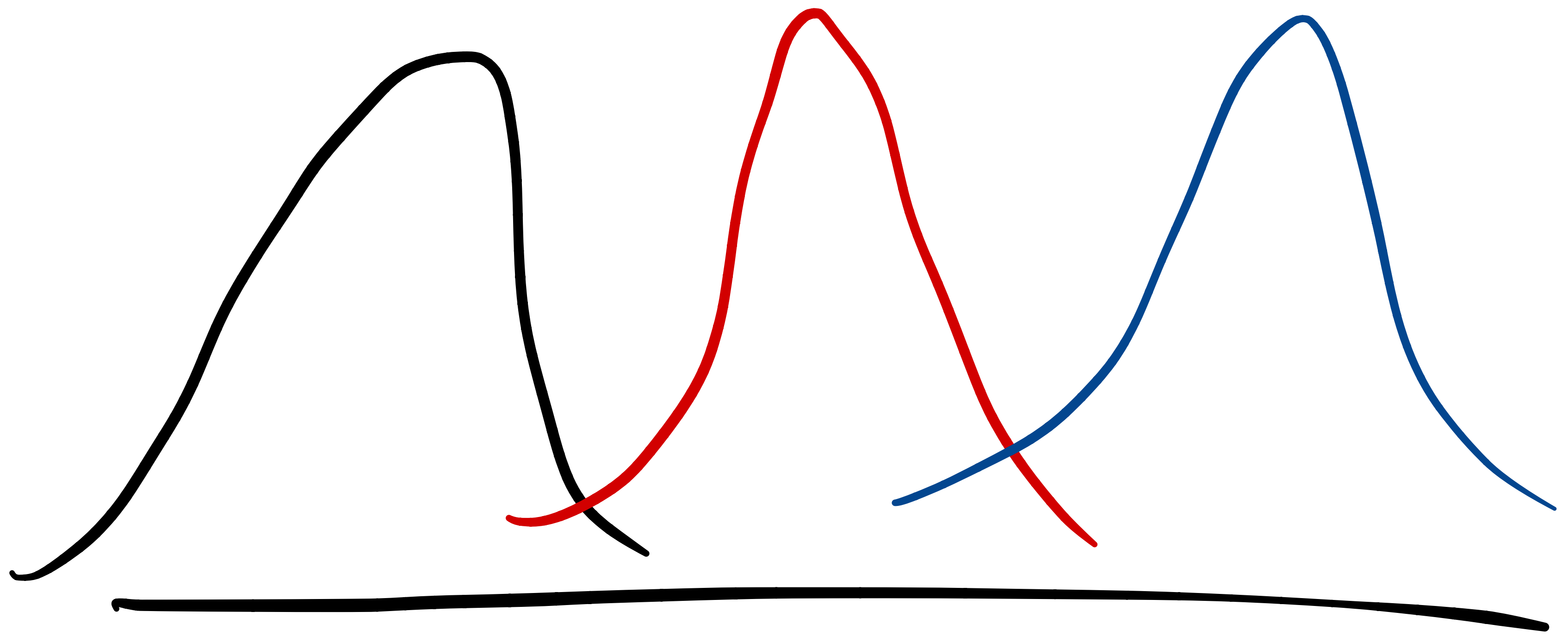
- How do we compare sample means for more than two groups?
  - One-way ANOVA

Plan for Today

- How do we compare sample means for more than two groups?
  - One-way ANOVA
- Motivating illustration:







Analysis of Variance: Motivation

Analysis of Variance: Motivation

- We previously discussed the use of t-tests for comparing sample means of two groups

Analysis of Variance: Motivation

- We previously discussed the use of t-tests for comparing sample means of two groups
- What if we want to compare sample means for more than two groups?

Analysis of Variance: Motivation

- We previously discussed the use of t-tests for comparing sample means of two groups
- What if we want to compare sample means for more than two groups?
  - Use analysis of variance (ANOVA)

One-Way ANOVA: Motivating Example

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old
- These three groups have means $\mu_1, \mu_2,$ and μ_3, respectively

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old
- These three groups have means $\mu_1, \mu_2,$ and μ_3, respectively
- We want to test the null hypothesis that the population means are identical

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old
- These three groups have means $\mu_1, \mu_2,$ and μ_3, respectively
- We want to test the null hypothesis that the population means are identical
  - $H_0 : \mu_1 = \mu_2 = \mu_3$

One-Way ANOVA: Motivating Example

- Suppose we are interested in the average weight of adult Americans, but we are looking at three different age groups
  - Group 1: 18-30 years old
  - Group 2: 31-50 years old
  - Group 3: Over 50 years old
- These three groups have means $\mu_1, \mu_2,$ and μ_3, respectively
- We want to test the null hypothesis that the population means are identical
  - $H_0 : \mu_1 = \mu_2 = \mu_3$
  - H_1 : At least one of the population means differs from one of the others

One-Way ANOVA

One-Way ANOVA

- In general, we are interested in comparing k different populations

One-Way ANOVA

- In general, we are interested in comparing k different populations
- Assume the k populations are independent and normally distributed

One-Way ANOVA

- In general, we are interested in comparing k different populations
- Assume the k populations are independent and normally distributed
- Draw a sample of size n_i from group i that has population mean μ_i and population variance σ_i^2

One-Way ANOVA

- In general, we are interested in comparing k different populations
- Assume the k populations are independent and normally distributed
- Draw a sample of size n_i from group i that has population mean μ_i and population variance σ_i^2
- For this sample, we get sample mean $\bar{x}_i$ and sample variance s_i^2

One-Way ANOVA

- In general, we are interested in comparing k different populations
- Assume the k populations are independent and normally distributed
- Draw a sample of size n_i from group i that has population mean μ_i and population variance σ_i^2
- For this sample, we get sample mean $\bar{x}_i$ and sample variance s_i^2
- Note: The number of observations in each sample does not need to be the same

One-Way ANOVA

One-Way ANOVA

- Consider our weight example over the three age groups, and let’s say we have the following data:

One-Way ANOVA

- Consider our weight example over the three age groups, and let's say we have the following data:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$

One-Way ANOVA

- Consider our weight example over the three age groups, and let's say we have the following data:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
  - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$

One-Way ANOVA

- Consider our weight example over the three age groups, and let's say we have the following data:
  - Group 1: $n_1 = 26, \bar{x}_1 = 151, s_1 = 8.9$
  - Group 2: $n_2 = 31, \bar{x}_2 = 174, s_2 = 11.4$
  - Group 3: $n_3 = 44, \bar{x}_3 = 162, s_3 = 9.9$

One-Way ANOVA

- Consider our weight example over the three age groups, and let's say we have the following data:
  - Group 1: $n_1 = 26$, $\bar{x}_1 = 151$, $s_1 = 8.9$
  - Group 2: $n_2 = 31$, $\bar{x}_2 = 174$, $s_2 = 11.4$
  - Group 3: $n_3 = 44$, $\bar{x}_3 = 162$, $s_3 = 9.9$

H_0

- Question: Are the average weights for the three age groups the same?

One-Way ANOVA

One-Way ANOVA

- We could answer this question by doing $\binom{3}{2} = 3$ different paired t-tests

One-Way ANOVA

- We could answer this question by doing $\binom{3}{2} = 3$ different paired t-tests
  - Group 1 vs. Group 2, Group 1 vs. Group 3, and Group 2 vs. Group 3

One-Way ANOVA

- We could answer this question by doing $\binom{3}{2} = 3$ different paired t-tests
  - Group 1 vs. Group 2, Group 1 vs. Group 3, and Group 2 vs. Group 3
- This may seem feasible for three groups, but as the number of groups increases, it quickly becomes infeasible

One-Way ANOVA

- We could answer this question by doing $\binom{3}{2} = 3$ different paired t-tests
    - Group 1 vs. Group 2, Group 1 vs. Group 3, and Group 2 vs. Group 3
- This may seem feasible for three groups, but as the number of groups increases, it quickly becomes infeasible
    - For instance, if we have $k = 10$ groups, we need to do $\binom{10}{2} = 45$ paired t-tests

One-Way ANOVA

One-Way ANOVA

- There is a more important problem, though: if we perform all possible two-sample t-tests, we are likely to reach an incorrect conclusion!

One-Way ANOVA

- There is a more important problem, though: if we perform all possible two-sample t-tests, we are likely to reach an incorrect conclusion!
- Suppose we do the three hypothesis tests for $k = 3$ groups, each at a significance level of $\alpha = 0.05$

One-Way ANOVA

- There is a more important problem, though: if we perform all possible two-sample t-tests, we are likely to reach an incorrect conclusion!
- Suppose we do the three hypothesis tests for $k = 3$ groups, each at a significance level of $\alpha = 0.05$
- What's the probability of a type I error (rejecting the null hypothesis given that the null hypothesis is true)?

One-Way ANOVA

	A	R
T_1	.95	.05
T_2	.95	.05
T_3	.95	.05

- There is a more important problem, though: if we perform all possible two-sample t-tests, we are likely to reach an incorrect conclusion!

- Suppose we do the three hypothesis tests for $k = 3$ groups, each at a significance level of $\underline{\alpha = 0.05}$

- What's the probability of a type I error (rejecting the null hypothesis given that the null hypothesis is true)?

$$\mathrm{Pr}(\overbrace{\text{Reject } H_0 \text{ given } H_0 \text{ is true}}^{R}) = 1 - \mathrm{Pr}(\text{Fail to reject in all three tests})$$

$$Pr(R) = Pr(R_1 \vee R_2 \vee R_3)$$

$$= 1 - (1 - 0.05)^3$$

$$= 1 - 0.857 = \underline{0.143} \quad > \quad 0.05$$

One-Way ANOVA

One-Way ANOVA

- The probability of rejecting the null hypothesis in at least one of these tests is higher than the α we use for each pairwise test

One-Way ANOVA

- The probability of rejecting the null hypothesis in at least one of these tests is higher than the α we use for each pairwise test
- In other words, if we know that the null hypothesis is true, then this type I error rate (0.143) is much higher than the desired standard (0.05)

One-Way ANOVA

- The probability of rejecting the null hypothesis in at least one of these tests is higher than the α we use for each pairwise test
- In other words, if we know that the null hypothesis is true, then this type I error rate (0.143) is much higher than the desired standard (0.05)
- With one-way ANOVA, we are able to keep the desired significance level α, unlike if we perform multiple t-tests

One-Way ANOVA

One-Way ANOVA

- Key idea: One-way ANOVA is dependent on estimates of spread or dispersion

One-Way ANOVA

- Key idea: One-way ANOVA is dependent on estimates of spread or dispersion
- “One-way” indicates that there is a single factor or characteristic that distinguishes the populations from each other

One-Way ANOVA

- Key idea: One-way ANOVA is dependent on estimates of spread or dispersion
- "One-way" indicates that there is a single factor or characteristic that distinguishes the populations from each other
    - In our example, age is the distinguishing factor

One-Way ANOVA: Assumptions

One-Way ANOVA: Assumptions

- Three assumptions must hold:

One-Way ANOVA: Assumptions

- Three assumptions must hold:
  - Normality: Each group follows a normal distribution

One-Way ANOVA: Assumptions

- Three assumptions must hold:
  - Normality: Each group follows a normal distribution
  - Equal variances: Population variances for each group are equal

One-Way ANOVA: Assumptions

- Three assumptions must hold:
  - Normality: Each group follows a normal distribution
  - Equal variances: Population variances for each group are equal
  - Independence: Observations are not correlated

Sources of Variation

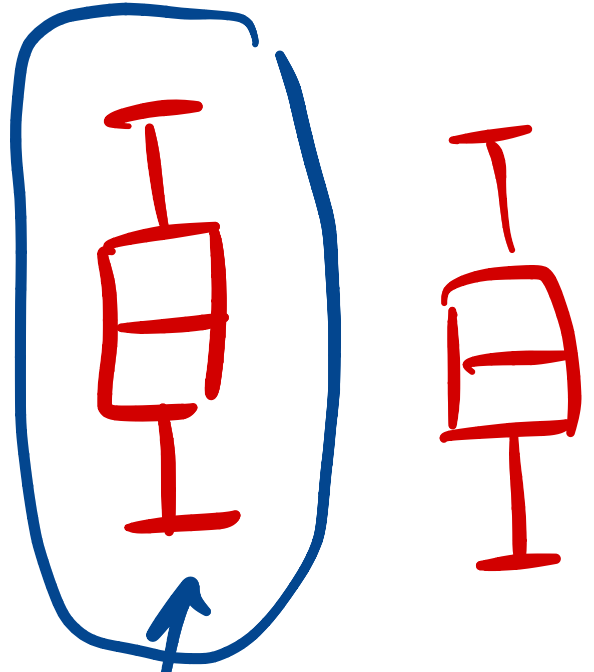
Sources of Variation

- We assume that there is a common variance σ^2 for the populations

Sources of Variation

- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:

Sources of Variation



- We assume that there is a common variance σ^2 for the populations

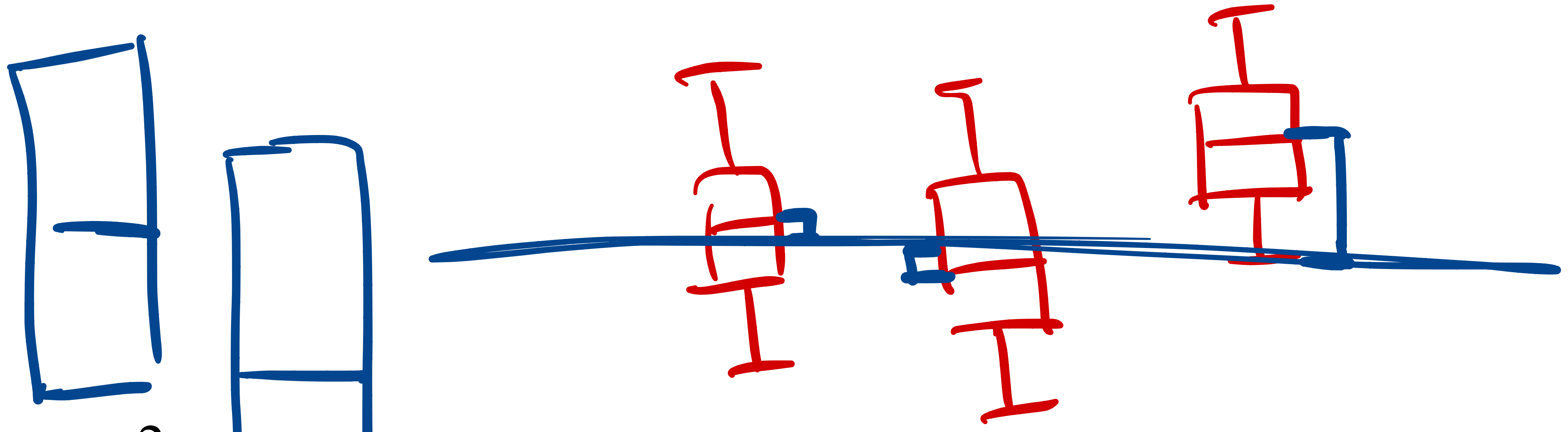
- There are two sources of variation in this ANOVA setup:

  - **Within-group variability** (s_w^2): Variation of the individual values around their group population means

Sources of Variation

- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:
  - **Within-group variability** (s_w^2): Variation of the individual values around their population means
    - All groups are assumed to have the same variability

Sources of Variation



- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:
  - **Within-group variability** (s_w^2): Variation of the individual values around their population means
    - All groups are assumed to have the same variability
  - **Between-group variability** (s_b^2): Variation of the population means around the grand (overall) mean

Sources of Variation

- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:
  - **Within-group variability** (s_w^2): Variation of the individual values around their population means
    - All groups are assumed to have the same variability
  - **Between-group variability** (s_b^2): Variation of the population means around the grand (overall) mean
- When the variability *within* the k populations is small relative to the variability among their respective means, this suggests that the population means are indeed different

Sources of Variation

- We assume that there is a common variance σ^2 for the populations
- There are two sources of variation in this ANOVA setup:
  - **Within-group variability** (s_w^2): Variation of the individual values around their population means
    - All groups are assumed to have the same variability
  - **Between-group variability** (s_b^2): Variation of the population means around the grand (overall) mean
- When the variability *within* the k populations is small relative to the variability among their respective means, this suggests that the population means are indeed different
  - Intuition: tightly clustered and separated from each other

Sources of Variation

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k} \leftarrow \frac{SSW}{df}$$

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \ldots + (n_k-1)s_k^2}{n_1+n_2+\ldots+n_k-k}$$
$$= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \ldots + (n_k-1)s_k^2}{n-k}$$

- **Between groups**: Variability of the population means around the grand (overall) mean

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k}$$

- **Between groups**: Variability of the population means around the grand (overall) mean

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$
$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k}$$

- **Between groups**: Variability of the population means around the grand (overall) mean

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

- The *grand mean*, $\bar{x}$, is defined as

Sources of Variation

- **Within groups**: Variability of the individuals around their population means

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k}$$

- **Between groups**: Variability of the population means around the grand (overall) mean

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

- The *grand mean*, $\bar{x}$, is defined as

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n}.$$

ANOVA Testing Procedures

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?
  - I.e., is $s_b^2 > s_w^2$? (Recall that s_w^2 measures within-group variability, s_b^2 measures between-group variability)

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?
  - I.e., is $s_b^2 > s_w^2$? (Recall that s_w^2 measures within-group variability, s_b^2 measures between-group variability)
- If the answer to this question is “yes”, this provides evidence that the population means are different

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?
    - I.e., is $s_b^2 > s_w^2$? (Recall that s_w^2 measures within-group variability, s_b^2 measures between-group variability)
- If the answer to this question is "yes", this provides evidence that the population means are different
- How can we compare s_w^2 and s_b^2?

ANOVA Testing Procedures

- Recall that we are interested in the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Our proxy question: Do sample means vary around the grand mean *more* than the individual observations around the sample means?
  - I.e., is $s_b^2 > s_w^2$? (Recall that s_w^2 measures within-group variability, s_b^2 measures between-group variability)
- If the answer to this question is "yes", this provides evidence that the population means are different
- How can we compare s_w^2 and s_b^2?
  - Thinking back to variances, s_b^2/s_w^2 follows an F distribution

ANOVA Testing Procedures

ANOVA Testing Procedures

- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2 / s_w^2$

ANOVA Testing Procedures

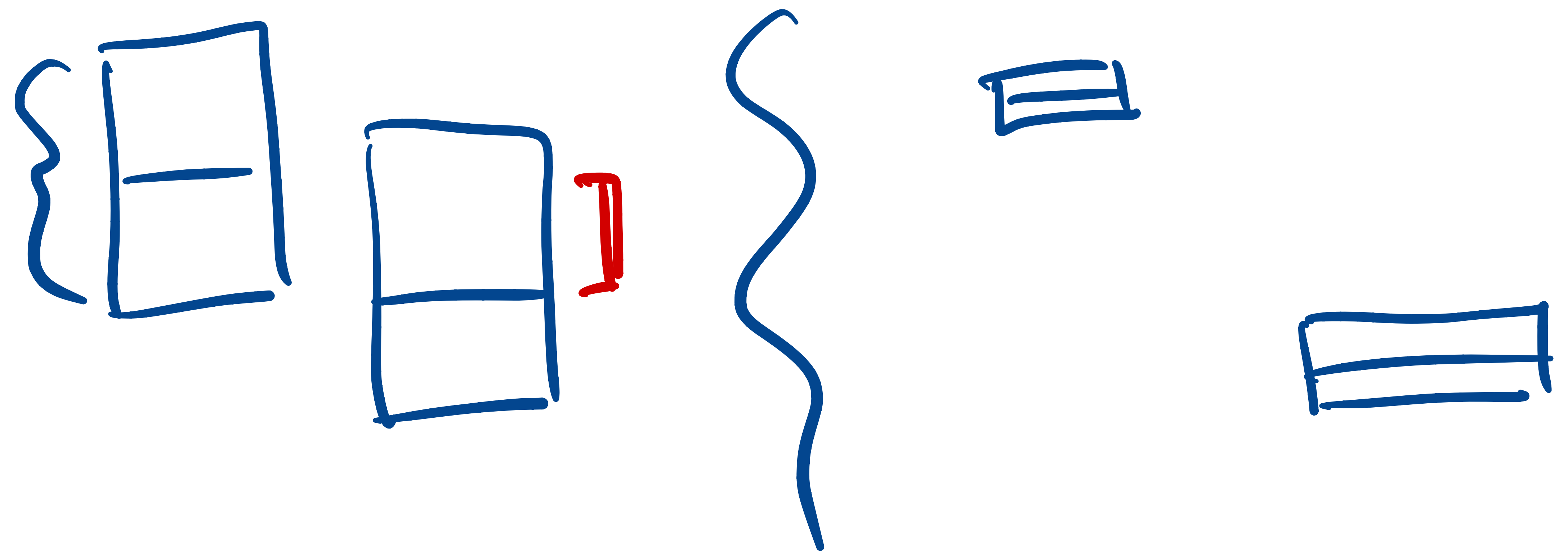
- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2 / s_w^2$
- Under the null hypothesis H_0, both s_b^2 and s_w^2 estimate the true σ^2

ANOVA Testing Procedures

- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2/s_w^2$
- Under the null hypothesis H_0, both s_b^2 and s_w^2 estimate the true σ^2
- Thus, we expect F to be close to 1

ANOVA Testing Procedures

- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2 / s_w^2$
- Under the null hypothesis H_0, both s_b^2 and s_w^2 estimate the true σ^2
- Thus, we expect F to be close to 1
- If a difference exists between population means, $s_b^2 > s_w^2$ and thus F will be larger than 1



ANOVA Testing Procedures

- To test the null hypothesis with a certain significance level α, se use the following test statistic: $F = s_b^2/s_w^2$
- Under the null hypothesis H_0, both s_b^2 and s_w^2 estimate the true σ^2
- Thus, we expect F to be close to 1
- If a difference exists between population means, $s_b^2 > s_w^2$ and thus F will be larger than 1
  - If $s_b^2 \leq s_w^2$, then there is no difference between population means

ANOVA Testing Procedures

ANOVA Testing Procedures

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = \overbrace{n_1 + n_2 + \cdots + n_k}^{n} - k$$

- Under H_0, $F = s_b^2 / s_w^2$ has an F distribution with $k - 1$ degrees of freedom in the numerator and $n - k$ degrees of freedom in the denominator

ANOVA Testing Procedures

- Under H_0, $F = s_b^2/s_w^2$ has an F distribution with $k-1$ degrees of freedom in the numerator and $n-k$ degrees of freedom in the denominator

- If $k=2$, this F-test reduces to a two-sample t-test

t^2

ANOVA Testing Procedures

ANOVA Testing Procedures

- Once F is calculated, we can calculate a p-value, p, based on the F distribution with degrees of freedom df1 and df2

ANOVA Testing Procedures

- Once F is calculated, we can calculate a p-value, p, based on the F distribution with degrees of freedom df1 and df2
    - Reject H_0 if $p \leq \alpha$

ANOVA Testing Procedures

- Once F is calculated, we can calculate a p-value, p, based on the F distribution with degrees of freedom df1 and df2
  - Reject H_0 if $p \leq \alpha$
- To get p-values in R (recall that we are only interested in the upper tail probability): `1-pf(F,df1,df2)`

ANOVA Testing Procedures

- Once F is calculated, we can calculate a p-value, p, based on the F distribution with degrees of freedom df1 and df2
  - Reject H_0 if $p \leq \alpha$
- To get p-values in R (recall that we are only interested in the upper tail probability): `1-pf(F,df1,df2)`
- Or, we can compare our test statistic to the critical value that cuts of the upper $\alpha \cdot 100\,\%$ of the F distribution with degrees of freedom df1 and df2

ANOVA Example

ANOVA Example

- Consider our weight example over the three age groups

ANOVA Example

- Consider our weight example over the three age groups
- We are interested in comparing the mean weight for three age groups: 18-30 years old, 31-50 years old, and 51+ years old

ANOVA Example

- Consider our weight example over the three age groups
- We are interested in comparing the mean weight for three age groups: 18-30 years old, 31-50 years old, and 51+ years old
- At the $\alpha = 0.05$ significance level, we want to test $H_0 : \mu_1 = \mu_2 = \mu_3$ against H_1 : at least one of the age groups has an average weight that is different from at least one of the other age groups

ANOVA Example

ANOVA Example

- Recall our sample summary statistics:

ANOVA Example

- Recall our sample summary statistics:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$

ANOVA Example

- Recall our sample summary statistics:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
  - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$

ANOVA Example

- Recall our sample summary statistics:
    - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
    - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$
    - Group 3: $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$

ANOVA Example

- Recall our sample summary statistics:
  - Group 1: $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$
  - Group 2: $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$
  - Group 3: $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$
- Using this information, we can calculate s_w^2, s_b^2, and our F-statistic

ANOVA Example

ANOVA Example

- Given $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$, $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$, and $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$

ANOVA Example

- Given $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$, $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$, and $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$
- $s_w^2 =$

ANOVA Example

- Given $n_1 = 26,\ \bar{x}_1 = 151,\ s_1 = 8.9$, $n_2 = 31,\ \bar{x}_2 = 174,\ s_2 = 11.4$, and $n_3 = 44,\ \bar{x}_3 = 162,\ s_3 = 9.9$
- $s_w^2 =$
- $\bar{x} =$

ANOVA Example

- Given $n_1 = 26$, $\bar{x}_1 = 151$, $s_1 = 8.9$, $n_2 = 31$, $\bar{x}_2 = 174$, $s_2 = 11.4$, and $n_3 = 44$, $\bar{x}_3 = 162$, $s_3 = 9.9$

"within"

- $s_w^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{n_1+n_2+n_3-3} = \dfrac{(26-1)\times 8.9^2 + (31-1)\times 11.4^2 + (44-1)\times 9.9^2}{26+31+44-3}$

  $= \boxed{103.}$

- $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1+n_2+n_3} = \underline{162.85}$

- $s_b^2 = \dfrac{n_1(\bar{x}_1-\bar{x})^2 + \cdots + n_3(\bar{x}_3-\bar{x})^2}{k-1 = 3-1 = 2} = \boxed{7521}$

$\dfrac{s_b^2}{s_w^2} \approx 70$

ANOVA Example

ANOVA Example

- Therefore, our test statistic is $F = s_b^2 / s_w^2 =$

ANOVA Example

- Therefore, our test statistic is $F = s_b^2 / s_w^2 =$
- For an F distribution with $k - 1 =$ and $n - k =$ degrees of freedom, we get a p-value of $p =$

ANOVA Example

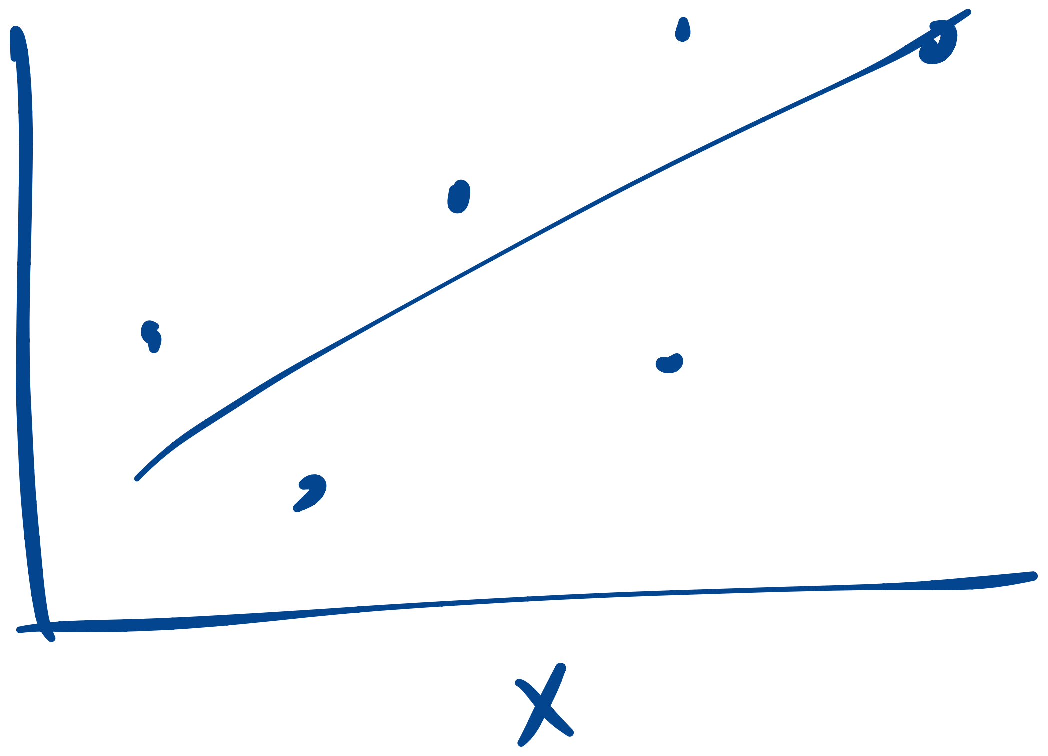
- Therefore, our test statistic is $F = s_b^2/s_w^2 =$ 73
- For an F distribution with $k - 1 =$ 2 and $n - k =$ 98 degrees of freedom, we get a p-value of $p =$ 1- pf(73, 2, 98) = 0
- Conclusion: $p <<< \alpha$

ANOVA Table

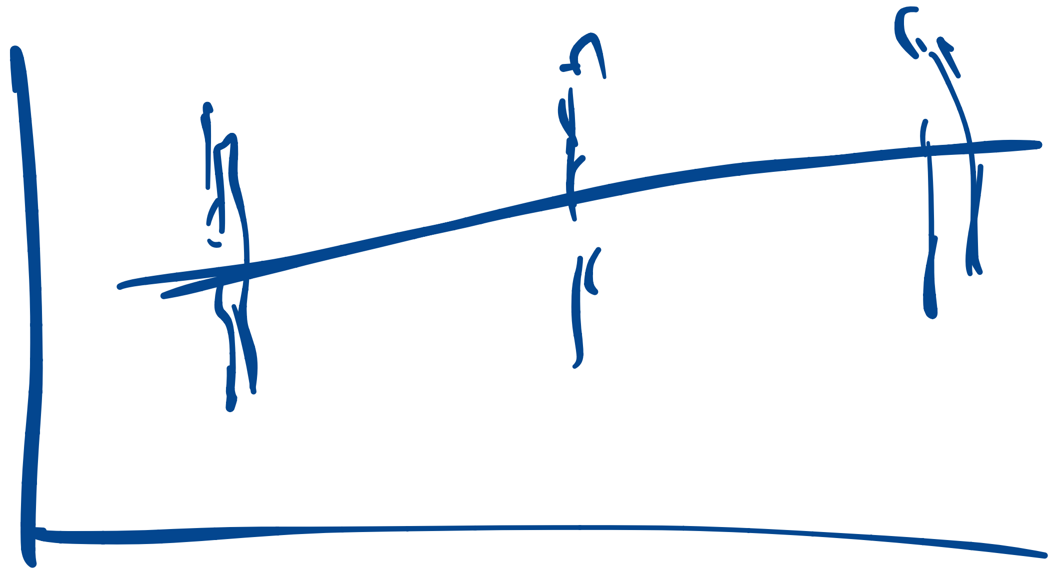
Source of Variation	Sum of Squares (SS)	df	Mean Squares (MS)	F	P-value
Between (treatment)	$SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$	$k-1$	$s_b^2 = \frac{SSB}{k-1}$	$\frac{s_b^2}{s_w^2}$	p^{***}
Within (error)	$SSE = \sum_{i=1}^{k} (n_i - 1)s_i^2$	$n-k$	$s_w^2 = \frac{SSE}{n-k}$		
Total	$\boxed{SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}$	$n-1$			

$1 - pf(F, k-1, n-k)$

ANOVA Table for Example



Source of Variation	Sum of Squares (SS)	df	Mean Squares (MS)	F	P-value
explain Between (treatment)	15041.754	$k-1$ 2	$s_b^2 = 7500$	≈ 75	0
unexplained Within (error)	10093.51	$n-k$ 98	$s_w^2 = 100$		
Total					



ANOVA

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Once H_0 is rejected, we can conclude that the population means are not all equal

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Once H_0 is rejected, we can conclude that the population means are not all equal
- However, we don't know exactly which means differ!

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Once H_0 is rejected, we can conclude that the population means are not all equal
- However, we don't know exactly which means differ!
- We need to conduct additional tests to find where the differences are

ANOVA

- Our null hypothesis is $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
- Once H_0 is rejected, we can conclude that the population means are not all equal
- However, we don't know exactly which means differ!

"post hoc"

- We need to conduct additional tests to find where the differences are
- In this case, we are performing *multiple comparisons*

Multiple Comparisons

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually
- Recall that in performing the $\binom{k}{2}$ possible two-sample t-tests, we increase our overall probability of committing a type I error

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually
- Recall that in performing the $\binom{k}{2}$ possible two-sample t-tests, we increase our overall probability of committing a type I error
- We correct for this by being more conservative in the individual comparisons

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually
- Recall that in performing the $\binom{k}{2}$ possible two-sample t-tests, we increase our overall probability of committing a type I error
- We correct for this by being more conservative in the individual comparisons
- Make it more difficult to reject each individual comparison so that the overall significance level remains at α

Multiple Comparisons

- Typically, we will be interested in comparing each pair of means individually
- Recall that in performing the $\binom{k}{2}$ possible two-sample t-tests, we increase our overall probability of committing a type I error
- We correct for this by being more conservative in the individual comparisons
- Make it more difficult to reject each individual comparison so that the overall significance level remains at α
    - We call this the *familywise type I error*, or α_{FWE}

Multiple Comparisons

Multiple Comparisons

- Intuition: If we are performing $\binom{k}{2}$ tests, then we can separate α evenly between these tests

Multiple Comparisons

- Intuition: If we are performing $\binom{k}{2}$ tests, then we can separate α evenly between these tests
    - $\alpha* = \frac{\alpha}{\binom{k}{2}}$ is the significance level for an individual comparison

Multiple Comparisons

- Intuition: If we are performing $\binom{k}{2}$ tests, then we can separate α evenly between these tests
    - $\alpha^* = \dfrac{\alpha}{\binom{k}{2}}$ is the significance level for an individual comparison
- This is called the *Bonferroni correction*

Multiple Comparisons

- Intuition: If we are performing $\binom{k}{2}$ tests, then we can separate α evenly between these tests
- $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$ is the significance level for an individual comparison
- This is called the *Bonferroni correction*
- For instance, if we want $\alpha = 0.05$ significance for $k = 5$ populations, then each pairwise test should have significance level $\alpha* = 0.05/10 = \underline{0.005}$

Multiple Comparisons

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}}$

"within"

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$
  - This is the test statistic for a two-sample t-test

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha^* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$
    - This is the test statistic for a two-sample t-test
    - But we estimate σ^2 based on all populations, not just populations i and j

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha^* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$
    - This is the test statistic for a two-sample t-test
    - But we estimate σ^2 based on all populations, not just populations i and j
    - Under null hypothesis H_0, t_{ij} has a t distribution with df = $n - k$

<u>not</u> $n_i + n_j - 2$

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\alpha* = \dfrac{\alpha}{\binom{k}{2}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$
  - This is the test statistic for a two-sample t-test
  - But we estimate σ^2 based on all populations, not just populations i and j
  - Under null hypothesis H_0, t_{ij} has a t distribution with df = $n - k$
- Calculate the p-value, p, based on a t distribution with $n - k$ degrees of freedom

Multiple Comparisons

- Consider the null hypothesis $H_0 : \mu_i = \mu_j$ that compares populations i and j
- Suppose that we want to test this hypothesis with a significance level of $\boxed{\alpha^* = \frac{\alpha}{\binom{k}{2}}}$
- Calculate test statistic $t_{ij} = \dfrac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$
  - This is the test statistic for a two-sample t-test
  - But we estimate σ^2 based on all populations, not just populations i and j
  - Under null hypothesis H_0, t_{ij} has a t distribution with $\boxed{\text{df} = n - k}$
- Calculate the p-value, p, based on a t distribution with $n - k$ degrees of freedom
- Reject H_0 if $p \leq \alpha^*$ $= \alpha / \binom{k}{2}$

Multiple Comparisons: Example

Multiple Comparisons: Example

- Return to the weight by age group example

Multiple Comparisons: Example

- Return to the weight by age group example
- We found that the population means were not all identical

Multiple Comparisons: Example

- Return to the weight by age group example
- We found that the population means were not all identical
- Now, we must compare each pair of age groups to see where the differences are

Multiple Comparisons: Example

- Return to the weight by age group example
- We found that the population means were not all identical
- Now, we must compare each pair of age groups to see where the differences are
- Total of $\binom{k}{2} = \binom{3}{2} = 3$ comparisons

Multiple Comparisons: Example

$\bar{x}_1 = 151$

$\bar{x}_2 = 174$

$\bar{x}_3 = 162$

- Return to the weight by age group example
- We found that the population means were not all identical
- Now, we must compare each pair of age groups to see where the differences are
- Total of $\binom{k}{2} = \binom{3}{2} = 3$ comparisons
- Overall desired significance $\alpha = 0.05$

$s_w^2 = 100$

$\alpha^* = \frac{\alpha}{\binom{k}{2}} = 0.0167$

Multiple Comparisons: Example

Multiple Comparisons: Example

- $t_{12} =$

Multiple Comparisons: Example

- $t_{12} =$

- $t_{13} =$

Multiple Comparisons: Example

- $t_{12} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_w^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{151 - 174}{\sqrt{100\left(\frac{1}{26} + \frac{1}{31}\right)}} = \underline{-8.52}$

- $t_{13} = \underline{-4.38}$

- $t_{23} = \underline{5.04}$

$df = n-k = 98$

Multiple Comparisons: Example

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$ -8.5, $t_{13} =$ -4.4, and $t_{23} =$ 5

$$\alpha^* = \frac{\alpha}{\binom{4}{2}} = \boxed{0.0167}$$

$$df = 101 - 3 = \underline{98}.$$

$$p_{12} = 2 \times pt(-8.5, df = 98) = 0.000000000000000195$$

$$p_{13}, \quad p_{23} < 1 \times 10^{-4}$$

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =
- Calculating p-values

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =
- Calculating p-values
  - $p_{12} =$

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =
- Calculating p-values
  - $p_{12} =$
  - $p_{13} =$

Multiple Comparisons: Example

- From previous slide, we have $t_{12} =$, $t_{13} =$, and $t_{23} =$
- Calculating other parameters: $\alpha^* =$, df =
- Calculating p-values
  - $p_{12} =$
  - $p_{13} =$
  - $p_{23} =$

Multiple Comparisons: Example

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?
  - Group 1 vs. Group 2:

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?
  - Group 1 vs. Group 2:
  - Group 1 vs. Group 3:

Multiple Comparisons: Example

- Given these p-values, what conclusions can we draw?
  - Group 1 vs. Group 2:
  - Group 1 vs. Group 3:
  - Group 2 vs. Group 3:

Reject H_0.

Different.

Multiple Comparisons: Example

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case
- Some populations may be the same whereas others are different

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case
- Some populations may be the same whereas others are different
- Conclusions from ANOVA and multiple comparisons may contradict each other!

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case
- Some populations may be the same whereas others are different
- Conclusions from ANOVA and multiple comparisons may contradict each other!
    - Significant ANOVA and no significant pairwise comparisons: Overly conservative pairwise comparisons test

Multiple Comparisons: Example

- In this case, all three comparisons were found to be significant
- This does not always have to be the case
- Some populations may be the same whereas others are different
- Conclusions from ANOVA and multiple comparisons may contradict each other!
  - Significant ANOVA and no significant pairwise comparisons: Overly conservative pairwise comparisons test
  - Non-significant ANOVA but significant pairwise comparisons: Generally consider pairwise comparisons result valid

Multiple Comparisons: Other Methods

Multiple Comparisons: Other Methods

- Other testing procedures than the Bonferroni procedure exist

Multiple Comparisons: Other Methods

- Other testing procedures than the Bonferroni procedure exist
  - Often called *post-hoc* analysis methods

Multiple Comparisons: Other Methods

- Other testing procedures than the Bonferroni procedure exist
  - Often called *post-hoc* analysis methods
- Bonferroni is traditionally one of the most conservative measures and can suffer from lack of power

Multiple Comparisons: Other Methods

- Other testing procedures than the Bonferroni procedure exist
  - Often called *post-hoc* analysis methods
- Bonferroni is traditionally one of the most conservative measures and can suffer from lack of power
- Some others: Tukey, Newman-Keuls, Scheffee, Dunnett, etc.

ANOVA in R

ANOVA in R

- First, create an ANOVA object using the `aov()` function

ANOVA in R

- First, create an ANOVA object using the `aov()` function
  - Let Y be the continuous variable (e.g., weight) and X be the grouping variable (e.g., age ranges)

ANOVA in R

- First, create an ANOVA object using the `aov()` function
  - Let Y be the continuous variable (e.g., weight) and X be the grouping variable (e.g., age ranges)
  - `model1=aov(Y~X)`

ANOVA in R

- First, create an ANOVA object using the `aov()` function
  - Let Y be the continuous variable (e.g., weight) and X be the grouping variable (e.g., age ranges)
  - `model1=aov(Y~X)`
- Summarize using the anova() function

ANOVA in R

- First, create an ANOVA object using the `aov()` function
  - Let Y be the continuous variable (e.g., weight) and X be the grouping variable (e.g., age ranges)
  - `model1=aov(Y~X)`
- Summarize using the anova() function
  - `anova(model1)`