Lesson 20:
Chapter 12

Caleb Moxley

BSC Mathematics

16 November 15

An **analysis of variance (ANOVA)** test compares the means of multiple populations/subpopulations.

§12.1 Inference for One-Way ANOVA

An **analysis of variance (ANOVA)** test compares the means of multiple populations/subpopulations. We've compared two means before, but a one-way ANOVA compares multiple means. It follows the

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

§12.1 Inference for One-Way ANOVA

An **analysis of variance (ANOVA)** test compares the means of multiple populations/subpopulations. We've compared two means before, but a one-way ANOVA compares multiple means. It follows the

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

This model arises from the assumption that all the means are equal (our hull hypothesis). And thus, the sample means should follow

$$x_{ij} = \mu_i + \epsilon_{ij},$$

where $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$ and ϵ_{ij} all have $N(0, \sigma)$ distribution.

§12.1 Inference for One-Way ANOVA

The estimate for σ is s_p which is given by

$$\sqrt{\frac{(n_1 - 1)s_1^2 + \cdots + (n_i - 1)s_i^2}{(n_1 - 1) + \cdots + (n_i - 1)}}.$$

§12.1 Inference for One-Way ANOVA

The estimate for σ is s_p which is given by

$$\sqrt{\frac{(n_1-1)s_1^2+\cdots+(n_i-1)s_i^2}{(n_1-1)+\cdots+(n_i-1)}}.$$

We always pool the standard deviation for a one-way ANOVA.

§12.1 Inference for One-Way ANOVA

Definition (hypothesis for one-way ANOVA)

The **null and alternative hypothesis** for a one-way ANOVA are

$$H_0 : \mu_1 = \cdots = \mu_i$$

$$H_a : \text{ not all of the } \mu_i \text{ are equal.}$$

§12.1 Inference for One-Way ANOVA

Definition (hypothesis for one-way ANOVA)

The **null and alternative hypothesis** for a one-way ANOVA are

$$H_0 : \mu_1 = \cdots = \mu_i$$

$$H_a : \text{ not all of the } \mu_i \text{ are equal.}$$

The conclusion of a one-way ANOVA can often be reached by finding the P-value from Minitab.

§12.1 Inference for One-Way ANOVA

Definition (hypothesis for one-way ANOVA)

The **null and alternative hypothesis** for a one-way ANOVA are

$$H_0 : \mu_1 = \cdots = \mu_i$$

$$H_a : \text{ not all of the } \mu_i \text{ are equal.}$$

The conclusion of a one-way ANOVA can often be reached by finding the P-value from Minitab. We will also discuss how the computations might be done by hand.

Definition (sum of squares, degrees of freedom, and mean squares)

Sum of squares represent variation in the data. They are calculated by summing square deviations. There are three sources of variation in a one-way ANOVA.

$$\text{SST} = \text{SSG} + \text{SSE}$$

The **degrees of freedom** are associated with each sum of squares.

$$\text{DFT} = \text{DFG} + \text{DFE}$$

The **mean squares** are $\frac{\text{sum of squares}}{\text{degrees of freedom}}$.

§12.1 Inference for One-Way ANOVA

Use the table below to calculate the values discussed in the previous slide.

§12.1 Inference for One-Way ANOVA

Use the table below to calculate the values discussed in the previous slide.

Source	DF	SS	MS	F
Groups	$I-1$	$\sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$	$\frac{\text{SSG}}{\text{DFG}}$	MSG/MSE
Error	$N-I$	$\sum_{\text{groups}} (n_i - 1)s_i^2$	$\frac{\text{SSE}}{\text{DFE}}$	
Total	$N-1$	$\sum_{\text{observations}} (x_{ij} - \bar{x})^2$		

§12.1 Inference for One-Way ANOVA

Definition (one-way ANOVA F test)

To test the null hypothesis that three or more population means are the same versus the alternative that at least one of them if different from the others, use the F statistic

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

The P-value is the probability that a random variable having the $F(I-1, N-I)$ distribution is greater than or equal to the calculated value of the F statistic.

Definition (one-way ANOVA F test)

To test the null hypothesis that three or more population means are the same versus the alternative that at least one of them if different from the others, use the F statistic

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

The P-value is the probability that a random variable having the $F(I-1, N-I)$ distribution is greater than or equal to the calculated value of the F statistic.

This is a right-tailed F test.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

- The smallest s_i must be larger than half of the largest s_i so that we can assume pooled standard deviations safely.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

- The smallest s_i must be larger than half of the largest s_i so that we can assume pooled standard deviations safely.
- Each population should be roughly normal.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

- The smallest s_i must be larger than half of the largest s_i so that we can assume pooled standard deviations safely.
- Each population should be roughly normal.
- The samples must be independent.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

- The smallest s_i must be larger than half of the largest s_i so that we can assume pooled standard deviations safely.
- Each population should be roughly normal.
- The samples must be independent.
- There must be only a single factor separating populations.

§12.1 Inference for One-Way ANOVA

In order to perform a one-way ANOVA test, the following requirements must be mets.

- The smallest s_i must be larger than half of the largest s_i so that we can assume pooled standard deviations safely.
- Each population should be roughly normal.
- The samples must be independent.
- There must be only a single factor separating populations.
- The samples must be SRSs.

§12.1 Inference for One-Way ANOVA

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

§12.1 Inference for One-Way ANOVA

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

18-25	25-50	over 50
650	725	700
625	670	660
645	770	750
600	590	700
590	700	760
500		725
595		

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Thus,

$$s_p = \sqrt{\frac{6(50.450353)^2 + 4(67.305275)^2 + 5(36.934627)^2}{6+4+5}} = 51.777.$$

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Thus,

$$s_p = \sqrt{\frac{6(50.450353)^2 + 4(67.305275)^2 + 5(36.934627)^2}{6+4+5}} = 51.777.$$

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Also, $\bar{x}_{\text{total}} = 664.16667$, so we have

$$\text{SSG} = 7(600.71429 - 664.16667)^2 + 5(691 - 664.16667)^2 +$$

$$6(715.83333 - 664.1667)^2 = 47800.214.$$

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Also, we have

$$SSE = 6(50.450353)^2 + 4(67.305275)^2 + 5(36.934627)^2 = 40212.262.$$

Example (mean credit score)

We want to see if the mean credit scores are the same across three age groups. We collect the following data from three SRSs.

	18-25	25-50	over 50
$\bar{x}$	600.71429	691	715.83333
s	50.450353	67.305275	36.934627
n	7	5	6

Thus, we must have SST $= 40212.262 + 47800.214 = 88012.5$.

§12.1 Inference for One-Way ANOVA

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	$I - 1$	$\sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$	$\frac{\text{SSG}}{\text{DFG}}$	MSG/MSE
Error	$N - I$	$\sum_{\text{groups}} (n_i - 1)s_i^2$	$\frac{\text{SSE}}{\text{DFE}}$	
Total	$N - 1$	$\sum_{\text{observations}} (x_{ij} - \bar{x})^2$		

§12.1 Inference for One-Way ANOVA

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	$\frac{SSG}{DFG}$	MSG/MSE
Error	$N - I$	$\sum_{\text{groups}} (n_i - 1)s_i^2$	$\frac{SSE}{DFE}$	
Total	$N - 1$	$\sum_{\text{observations}} (x_{ij} - \bar{x})^2$		

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	$\frac{SSG}{DFG}$	MSG/MSE
Error	15	40212.262	$\frac{SSE}{DFE}$	
Total	$N-1$	$\sum_{\text{observations}} (x_{ij} - \bar{x})^2$		

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	$\frac{SSG}{DFG}$	MSG/MSE
Error	15	40212.262	$\frac{SSE}{DFE}$	
Total	17	88012.5		

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	23900.107	8.915
Error	15	40212.262	2680.817	
Total	17	88012.5		

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	23900.107	8.915
Error	15	40212.262	2680.817	
Total	17	88012.5		

Thus, the P-value for this test is $P(F > 8.915) = 0.003$.

Example (mean credit score)

Let's fill in our table:

Source	DF	SS	MS	F
Groups	2	47800.214	23900.107	8.915
Error	15	40212.262	2680.817	
Total	17	88012.5		

Thus, the P-value for this test is $P(F > 8.915) = 0.003$. So if we conducted the one-way ANOVA at a 5% significance level, we would reject the null hypothesis. We would not support the claim that the mean credit scores are the same across all three of these age groups.

Note: The full one-way ANOVA test can be conducted in Minitab! You should make sure that you know how the one-way ANOVA table could be filled in given just a few SS and DF values, though!

Note: The full one-way ANOVA test can be conducted in Minitab! You should make sure that you know how the one-way ANOVA table could be filled in given just a few SS and DF values, though! Let's do one more test using Minitab.

Example (labor costs)

A company would like to see if its labor costs are the same over all four seasons. It takes a SRS of labor costs in each season over several years and gets the following data.

Spring	Summer	Fall	Winter
300000	295000	259000	260000
450000	430000	460000	475000
375000	380000	385000	380000
257000	277000	259000	300000
280000	285000	275000	290000
300000	299000	301000	298000
440000	445000	440500	444000

Example (labor costs)

A company would like to see if its labor costs are the same over all four seasons. It takes a SRS of labor costs in each season over several years and gets the following data. We get the following table.

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Columns	3	3.3774107e8	1.1258036e8	0.017332332	0.9968
Error	24	1.558895e11	6.4953958e9		
Total	27	1.5622724e11			

§12.1 Inference for One-Way ANOVA

Example (labor costs)

A company would like to see if its labor costs are the same over all four seasons. It takes a SRS of labor costs in each season over several years and gets the following data. We get the following table.

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Columns	3	3.3774107e8	1.1258036e8	0.017332332	0.9968
Error	24	1.558895e11	6.4953958e9		
Total	27	1.5622724e11			

Thus, we fail to reject the null hypothesis and support the claim that the labor costs are the same across all four seasons.