


## Inferential Statistics and Probability a Holistic Approach

---

### Chapter 12

### One Factor Analysis of Variance (ANOVA)

  
This Course Material by Maurice Geraghty is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Conditions for use are shown here: <https://creativecommons.org/licenses/by-sa/4.0/>

1

1

---

---

---


---

---

---

---

---



## ANOVA Definitions

- **Factor** – categorical variable that defines the populations.
- **Response** – variable that is being measured.
- **Levels** – the number of choices for the factor, represented by  $k$
- **Replicates** – the sample size for each level,  $n_1, n_2, \dots, n_k$ .
- If  $n_1 = n_2 = \dots = n_k$ , then the design is **balanced**.
  
- **H<sub>0</sub>**: There is no difference in the mean <response in context> due to the <factor in context>.
- **H<sub>a</sub>**: There is a difference in the mean <response in context> due to the <factor in context>.

2

2

---

---

---


---

---

---

---

---



## Underlying Assumptions for ANOVA

- The  $F$  distribution is also used for testing the equality of more than two means using a technique called analysis of variance (ANOVA). ANOVA requires the following conditions:
  - The populations being sampled are normally distributed.
  - The populations have equal standard deviations.
  - The samples are randomly selected and are independent.

3

3

---

---

---

---

---

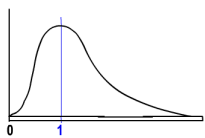
---

---

---

### Characteristics of F-Distribution

- There is a "family" of  $F$  Distributions.
- Each member of the family is determined by two parameters: the numerator degrees of freedom and the denominator degrees of freedom.
- $F$  cannot be negative, and it is a continuous distribution.
- The  $F$  distribution is positively skewed.
- Its values range from 0 to  $\infty$ .
  - As  $F \rightarrow \infty$  the curve approaches the X-axis.



4

---

---

---

---

---

---

---

---

4

### Analysis of Variance Procedure

- The Null Hypothesis:** the population means are the same.
- The Alternative Hypothesis:** at least one of the means is different.
- The Test Statistic:**  $F = (\text{between sample variance}) / (\text{within sample variance})$ .
- Decision rule:** For a given significance level  $\alpha$ , reject the null hypothesis if  $F(\text{computed})$  is greater than  $F(\text{table})$  with numerator and denominator degrees of freedom.

5

---

---

---

---

---

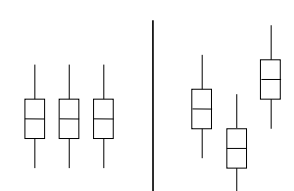
---

---

---

5

### ANOVA – Null Hypothesis



Ho is true - all means the same

Ho is false - not all means the same

6

---

---

---

---

---

---

---

---

6

### ANOVA NOTES

- If there are k populations being sampled (levels), then the  $df_{factor} = k-1$ 
  - If the sample size is n, then  $df_{error} = n-k$
- The test statistic is computed by:  $F = \frac{SS_F / (k-1)}{SS_E / (n-k)}$ .
- $SS_F$  represents the factor (between) sum of squares.
- $SS_E$  represents the error (within) sum of squares.
- Let  $T_c$  represent the column totals,  $n_c$  represent the number of observations in each column, and  $\sum X$  represent the sum of all the observations.
- These calculations are tedious, so technology is used to generate the **ANOVA table**.

7

7

---

---

---

---

---

---

---

---

---

---

### Formulas for ANOVA

$$SS_{Total} = \sum(X^2) - \frac{(\sum X)^2}{n}$$

$$SS_{Factor} = \sum\left(\frac{T_c^2}{n_c}\right) - \frac{(\sum X)^2}{n}$$

$$SS_{Error} = SS_{Total} - SS_{Factor}$$

8

8

---

---

---

---

---

---

---

---

---

---

### ANOVA Table

Source	SS	df	MS	F
Factor	$SS_{Factor}$	$k-1$	$SS_F / df_F$	$MS_F / MS_E$
Error	$SS_{Error}$	$n-k$	$SS_E / df_E$	
Total	$SS_{Total}$	$n-1$		

9

9

---

---

---

---

---

---

---

---

---

---

**EXAMPLE**

Party Pizza specializes in meals for students. Hsieh Li, President, recently developed a new tofu pizza.

- Before making it a part of the regular menu she decides to test it in several of her restaurants. She would like to know if there is a difference in the mean number of tofu pizzas sold per day at the Cupertino, San Jose, and Santa Clara pizzerias for sample of five days.
- At the .05 significance level can Hsieh Li conclude that there is a difference in the mean number of tofu pizzas sold per day at the three pizzerias?

10

---

---

---

---

---

---

---

---

10

**Example**

	Cupertino	San Jose	Santa Clara	Total
	13	10	18	
	12	12	16	
	14	13	17	
	12	11	17	
			17	
T	51	46	85	182
n	4	4	5	13
Means	12.75	11.5	17	14
$\Sigma^2$	653	534	1447	2634

11

---

---

---

---

---

---

---

---

11

**Example continued**

$$SS_{Total} = 2634 - \frac{182^2}{13} = 86$$

$$SS_{Factor} = 2624.25 - \frac{182^2}{13} = 76.25$$

$$SS_{Error} = 86 - 76.25 = 9.75$$

12

---

---

---

---


---

---

---

---

12

 **Example 4** *continued*

ANOVA TABLE

Source	SS	df	MS	F
Factor	76.25	2	38.125	39.10
Error	9.75	10	0.975	
Total	86.00	12		

13

13

---

---

---


---

---

---

---

---

 **EXAMPLE 4** *continued*

- **Design:**  $H_0: \mu_1 = \mu_2 = \mu_3$   
 $H_a$ : Not all the means are the same
- $\alpha = .05$
- Model: One Factor ANOVA
- $H_0$  is rejected if  $F > 4.10$
- **Data:** Test statistic:  $F = [76.25/2]/[9.75/10] = 39.1026$
- $H_0$  is rejected.
- **Conclusion:** There is a difference in the mean number of pizzas sold at each pizzeria.

14

14

---

---

---


---

---

---

---

---

 **One-way ANOVA: Cupertino, San Jose, Santa Clara**

Source	DF	SS	MS	F	P
Factor	2	76.250	38.125	39.10	0.000
Error	10	9.750	0.975		
Total	12	86.000			

S = 0.9874   R-Sq = 88.66%   R-Sq(adj) = 86.40%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
Cupertino	4	12.750	0.957
San Jose	4	11.500	1.291
Santa Clara	5	17.000	0.707

12.0   14.0   16.0   18.0

15

15

---

---

---

---

---

---

---

---

### Post Hoc Comparison Test

- Used for pairwise comparison
- Designed so the **overall** significance level is 5%.
- Use technology.
- Refer to **Tukey Test** Material in the textbook.

16

---

---

---

---

---

---

---

---

16

### Post Hoc Comparison Test

Grouping Information Using Tukey Method

	N	Mean	Grouping
Santa Clara	5	17.0000	A
Cupertino	4	12.7500	B
San Jose	4	11.5000	B

Means that do not share a letter are significantly different.

17

---

---

---

---

---

---

---

---

17

### Post Hoc Comparison Test

Individual Value Plot of Cupertino, San Jose, Santa Clara

Group	Mean
Cupertino	12.75
San Jose	11.50
Santa Clara	17.00

18

---

---

---

---

---

---

---

---

18

### Example – Oranges & Orchards

- Valencia oranges were tested for juiciness at 4 different orchards. Eight oranges were sampled from each orchard, and the total ml of juice per 20 gms of orange was calculated.
- Test for a difference in juiciness due to orchards using  $\alpha = .05$
- Perform all the pairwise comparisons using Tukey's Test and an overall risk level of 5%.

Orchard A:	Orchard B:	Orchard C:	Orchard D:
11,13,12,14, 9,13,11,9	10,9,8,10, 11,12,7,8	13,15,14,11, 12,10,16,11	9,7,11,9, 9,11,10,8

19

---

---

---

---

---

---

---

---

---

---

### Example - Definitions

- Factor: Orchard (A, B, C or D)
- Response: Juiciness of orange
- Levels:  $k = 4$
- Replicate:  $n_A = n_B = n_C = n_D = 8$
- Design: Balanced
- Sample size:  $n = 8 + 8 + 8 + 8 = 32$

20

---

---

---

---

---

---

---

---

---

---

### Example – Value Plot

Individual Value Plot of Juiciness vs Orchard

Orchard	Mean Juiciness
A	11.5
B	9.5
C	12.5
D	9.0

21

---

---

---

---

---


---

---

---

---

---



### Example – Stats & ANOVA Table

Level	N	Mean	StDev
Orchard A	8	11.500	1.852
Orchard B	8	9.375	1.685
Orchard C	8	12.750	2.121
Orchard D	8	9.250	1.389

Source	DF	SS	MS	F	P
Factor	3	69.59	23.20	7.31	0.001
Error	28	88.88	3.17		
Total	31	158.47			

22

---

---

---

---


---

---

---

---

22



### Example – Tukey Test Grouping

Grouping Information Using Tukey Method

	N	Mean	Grouping
Orchard C	8	12.750	A
Orchard A	8	11.500	A B
Orchard B	8	9.375	B
Orchard D	8	9.250	B

Means that do not share a letter are significantly different.

23

---

---

---

---

---

---

---

---

23