BINUS UNIVERSITY

GREATER JAKARTA • BANDUNG • MALANG

People
Innovation
Excellence

STAT6171001
Basic Statistics

Two-Way ANOVA
Session 12

Raymond Bahana
rbahana@binus.edu

Session Learning Outcomes

Upon completion of this session, students are expected to be able to

- LO 1. Explain basic statistics concept
- LO 2. Analyze a problem by using the basic concept of descriptive and inferential statistics
- LO 3. Design a descriptive and inferential statistics solution to meet a given set of computing requirements in the context of computer science
- LO4. Produce descriptive and inferential statistics solutions

Topic

- Analysis of Variance

What Is a Two-Way ANOVA?

- A two-way analysis of variance is used to determine whether there is a statistically significant difference between the means of three or more independent groups that have been split on two variables (or factors).
- The two-way ANOVA can be applied in different fields, such as economics, finance, social science, and medicine.

When to Use a Two-Way ANOVA

- Use a two-way ANOVA when you'd like to know how two factors affect a response variable and whether or not there is an interaction effect between the two factors on the response variable

Example

- A botanist wants to explore how sunlight exposure and watering frequency affect plant growth.
- She plants 40 seeds and lets them grow for 2 months under different conditions for sunlight exposure and watering frequency.
- After 2 months, she records the height of each plant.

- In this case, we have the following variables:
  - Response variable (dependent variable): plant growth
  - Factors (independent variables): sunlight exposure, watering frequency

Example

- And we would like to answer the following questions:
  - Does sunlight exposure affect plant growth?
  - Does watering frequency affect plant growth?
  - Is there an interaction effect between sunlight exposure and watering frequency? (e.g., the effect that sunlight exposure has on the plants is dependent on watering frequency)
- We would use a two-way ANOVA for this analysis because we have two factors.
- If instead we wanted to know how only watering frequency affected plant growth, we would use a one-way ANOVA since we would only be working with one factor.

Two-Way ANOVA Assumptions

- The two-way ANOVA test has got two independent variables hence the term two-way.
- That there as several assumptions that relate to the two-way analysis of variance, as follows:
    - The sample population must be approximately normally or normally distributed
    - It is mandatory for the samples to be independent
    - It is mandatory that the populations variances be equal
    - The sample size of the groups must be the same

A Two-Way ANOVA Hypotheses

- The two-way ANOVA has several sets to hypotheses.
- The null hypotheses are as follows:
  - The first factors population means are equal. It is the same as the one-way ANOVA when it comes to the row factor
  - The second factors population means are equal. It is the same as the one-way ANOVA when it comes to column factor
  - Between the two factors, no interaction exists. It is the same as administering a test for independence using contingency tables.

Example

- The results are shown (in inches):
- In the table, we see that there were five plants grown under each combination of conditions.

	Sunlight Exposure			
Watering Frequency	None	Low	Medium	High
Daily	4.8	5	6.4	6.3
	4.4	5.2	6.2	6.4
	3.2	5.6	4.7	5.6
	3.9	4.3	5.5	4.8
	4.4	4.8	5.8	5.8
Weekly	4.4	4.9	5.8	6
	4.2	5.3	6.2	4.9
	3.8	5.7	6.3	4.6
	3.7	5.4	6.5	5.6
	3.9	4.8	5.5	5.5

Example

- For example, there were five plants grown with daily watering and no sunlight and their heights after two months were 4.8 inches, 4.4 inches, 3.2 inches, 3.9 inches, and 4.4 inches:

	Sunlight Exposure			
Watering Frequency	None	Low	Medium	High
Daily	4.8	5	6.4	6.3
	4.4	5.2	6.2	6.4
	3.2	5.6	4.7	5.6
	3.9	4.3	5.5	4.8
	4.4	4.8	5.8	5.8
Weekly	4.4	4.9	5.8	6
	4.2	5.3	6.2	4.9
	3.8	5.7	6.3	4.6
	3.7	5.4	6.5	5.6
	3.9	4.8	5.5	5.5

Example: Step by Step

- **Step 1**: **Calculate Sum of Squares for First Factor (Watering Frequency)**
- First, calculate the grand mean height of all 40 plants:
  Grand mean = (4.8 + 5 + 6.4 + 6.3 + … + 3.9 + 4.8 + 5.5 + 5.5) / 40 = 5.1525

- Next, we will calculate the mean height of all plants watered daily:
  Mean of Daily = (4.8 + 5 + 6.4 + 6.3 + … + 4.4 + 4.8 + 5.8 + 5.8) /20 = 5.155

- Next, we will calculate the mean height of all plants watered weekly:
  Mean of Weekly = (4.4 + 4.9 + 5.8 + 6 + … + 3.9 + 4.8 + 5.5 + 5.5) /20 = 5.15

Example: Step by Step

- Next, we will calculate the sum of squares for the factor “watering frequency” by using the following formula:

$$\Sigma n(X_j - X..)^2$$

- where:

  n: the sample size of group j
  Σ: a greek symbol that means “sum”
  X_j: the mean of group j
  X..: the grand mean

- Calculate the sum of squares for the factor “watering frequency”: $20*(5.155-5.1525)^2 + 20*(5.15-5.1525)^2 = 0.00025$

Example: Step by Step

- **Step 2: Calculate Sum of Squares for Second Factor (Sunlight Exposure)**
- First, we will calculate the grand mean height of all 40 plants:
  Grand mean = (4.8 + 5 + 6.4 + 6.3 + … + 3.9 + 4.8 + 5.5 + 5.5) / 40 = 5.1525

- Next, we will calculate the mean height of all plants with no sunlight exposure:
  Mean of No Sunlight = (4.8+4.4+3.2+3.9+4.4+4.4+4.2+3.8+3.7+3.9) / 10 = 4.07
  Mean of Low Sunlight = 5.1
  Mean of Medium Sunlight = 5.89
  Mean of High Sunlight = 5.55

Example: Step by Step

- Next, we will calculate the sum of squares for the factor “sunlight exposure”.
- Calculate the sum of squares for the factor “sunlight exposure”:
  $10*(4.07-5.1525)^2 + 10*(5.1-5.1525)^2 + 10*(5.89-5.1525)^2 + 10*(5.55-5.1525)^2 = 18.76475$

Example: Step by Step

- **Step 3: Calculate Sum of Squares Within (Error)**
- Next, calculate the sum of squares within by taking the sum of squared differences between each combination of factors and the individual plant heights.
- For example, the mean height of all plants watered daily with no sunlight exposure is 4.14.
- We can then calculate the sum of squared differences for each of these individual plants as:

  SS for daily watering and no sunlight: $(4.8-4.14)^2 + (4.4-4.14)^2 + (3.2-4.14)^2 + (3.9-4.14)^2 + (4.4-4.14)^2 = 1.512$

Example: Step by Step

- Repeat this process for each combination of factors:
  SS for daily watering and low sunlight: 0.928
  SS for daily watering and medium sunlight: 1.788
  SS for daily watering and high sunlight: 1.648
  SS for weekly watering and no sunlight: 0.34
  SS for weekly watering and low sunlight: 0.548
  SS for weekly watering and medium sunlight: 0.652
  SS for weekly watering and high sunlight: 1.268
- Sums of squares within = 1.512 + .928 + 1.788 + 1.648 + .34 + .548 + .652 + 1.268 = 8.684

Example: Step by Step

- **Step 4: Calculate Total Sum of Squares**
- Next, calculate the total sum of squares by taking the sum of the differences between each individual plant height and the grand mean:

  Total Sum of Squares

  $= (4.8 – 5.1525)^2 + (5 – 5.1525)^2 + … + (5.5 – 5.1525)^2$

  $= 28.45975$

Example: Step by Step

- **Step 5: Calculate Sum of Squares Interaction**
- Next, calculate the sum of squares interaction by using the following formula:

  SS Interaction = SS Total – SS Factor 1 – SS Factor 2 – SS Within

  SS Interaction = 28.45975 – .00025 – 18.76475 – 8.684

  SS Interaction = 1.01075

Example: Step by Step

- **Step 6: Fill in ANOVA Table**

  df Watering Frequency: j-1 = 2-1 = 1

  df Sunlight Exposure: k-1 = 4-1 = 3

  df Interaction: (j-1)*(k-1) = 1*3 = 3

  df Within: n – (j*k) = 40 – (2*4) = 32

  df total: n-1 = 40-1 = 39

  MS: SS/ df

  F Watering Frequency: MS Watering Frequency / MS Within

  F Sunlight Exposure: MS Sunlight Exposure / MS Within

Example: Step by Step

F Interaction: MS Interaction / MS Within

p-value Watering Frequency: The p-value that corresponds to F value of 0.000921 with df numerator = 1 and df denominator = 32

p-value Sunlight Exposure: The p-value that corresponds to F value of 23.04898 with df numerator = 3 and df denominator = 32

p-value Interaction: The p-value that corresponds to F value of 1.241517 with df numerator = 3 and df denominator = 32

- **Note #1**: n = total observations, j = number of levels for watering frequency, k = number of levels for sunlight exposure.
- **Note #2**: The p-values that correspond to the F-value were calculated using the F Distribution Calculator.

Example

- Performs a two-way ANOVA in Excel and ends up with the following output:

G	H	I	J	K	L	M
SUMMARY	None	Low	Medium	High	Total	
Daily						
Count	5	5	5	5	20	
Sum	20.7	24.9	28.6	28.9	103.1	
Average	4.14	4.98	5.72	5.78	5.155	
Variance	0.378	0.232	0.447	0.412	0.775237	
Weekly						
Count	5	5	5	5	20	
Sum	20	26.1	30.3	26.6	103	
Average	4	5.22	6.06	5.32	5.15	
Variance	0.085	0.137	0.163	0.317	0.722632	
Total						
Count	10	10	10	10		
Sum	40.7	51	58.9	55.5		
Average	4.07	5.1	5.89	5.55		
Variance	0.211222	0.18	0.303222	0.382778		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample (Watering)	0.00025	1	0.00025	0.000921	0.975975	4.149097
Columns (Sunlight)	18.76475	3	6.254917	23.04898	3.9E-08	2.90112
Interaction	1.01075	3	0.336917	1.241517	0.310898	2.90112
Within	8.684	32	0.271375			
Total	28.45975	39				

Example: Step by Step

- **Step 7: Interpret the results**
- Observe the following from the ANOVA table:
    - The p-value for the interaction between watering frequency and sunlight exposure was 0.311 → This is not statistically significant at $\alpha = 0.05$.
    - The p-value for watering frequency was 0.975 → This is not statistically significant at $\alpha = 0.05$.
    - The p-value for sunlight exposure was < 0.000 → This is statistically significant at $\alpha = 0.05$.

Example

- These results indicate that sunlight exposure is the only factor that has a statistically significant effect on plant height.
- And because there is no interaction effect, the effect of sunlight exposure is consistent across each level of watering frequency.
- That is, whether a plant is watered daily or weekly has no impact on how sunlight exposure affects a plant.

A Two-Way vs. One-Way ANOVA

- Analysis of variance exists in two types: one-way and two-way which is also known as unidirectional and bidirectional, respectively.
- The two analyses of variance refer to independent variables number in the analysis of variance test.

- One-Way ANOVA
  - It’s a type of analysis of variance used to evaluate the sole factors impact on a single response variable.
  - It also tests the samples to find out if they are equal.
  - In addition, it can be used to determine whether, between the means of three or more unrelated groups, there are significant differences in the statistics.

A Two-Way vs. One-Way ANOVA

- Two-Way ANOVA
  - There are two independent variables in this type of analysis of variance.
  - With the two-way variables test, a company can easily compare the productivity of its workers, based on two independent variables such as skills and salary.
  - It can utilize the test to observe how those two factors interact with each other.
  - It also tests the two factors effect simultaneously.

Two-way ANOVA table

Source of variation	df	Sums of squares	Mean square	F
Factor A	$k - 1$	SSA	$MSA = \frac{SSA}{k-1}$	$F_A = \frac{MSA}{MSE}$
Factor B	$l - 1$	SSB	$MSB = \frac{SSB}{l-1}$	$F_B = \frac{MSB}{MSE}$
Interaction AB	$(k - 1)(l - 1)$	SSAB	$MSAB = \frac{SSAB}{(k-1)(l-1)}$	$F_{AB} = \frac{MSAB}{MSE}$
Error	$kl(m - 1)$	SSE	$MSE = \frac{SSE}{kl(m-1)}$	
Total	$klm - 1$	SSTo		

2nd Example

- A reputed marketing agency in New Zealand has three different training programs for its salesmen.
- The three programs are Method – A, B, C.
- To assess the success of the programs, 4 salesmen from each of the programs were sent to the field.

2nd Example

- Their performances in terms of sales are given in the following table.

Salesmen	Methods		
	A	B	C
1	4	6	2
2	6	10	6
3	5	7	4
4	7	5	4

- Test whether there is significant difference among methods and among salesmen.

2nd Example

- **Step 1 : Hypotheses**
  - Null Hypotheses: H_{01} : $\mu_{M1} = \mu_{M2} = \mu_{M3}$ (for treatments)
  - That is, there is no significant difference among the three programs in their mean sales.
  - H_{02} : $\mu_{S1} = \mu_{S2} = \mu_{S3} = \mu_{S4}$ (for block / salesman)

- Alternative Hypotheses:
  - H_{11} : At least one average is different from the other, among the three programs.
  - H_{12} : At least one average is different from the other, among the four salesmen.

2^{nd} Example

- **Step 2 : Data**

Salesmen	Methods		
	A	B	C
1	4	6	2
2	6	10	6
3	5	7	4
4	7	5	4

People
Innovation
Excellence

GREATER JAKARTA • BANDUNG • MALANG

2^{nd} Example

- **Step 3 : Level of significance** α = 5%

- **Step 4 : Test Statistic**

$$F_{0t}\,(\text{treatment}) = \frac{MST}{MSE}$$

$$F_{0b}\,(\text{block}) = \frac{MSB}{MSE}$$

2^{nd} Example

- **Step 5 : Calculation of the Test Statistic**

	Methods			Total $x_{i.}$	$x_{i.}^2$
	A	B	C		
1	4	6	2	12	144
2	6	10	6	22	484
3	5	7	4	16	256
4	7	5	4	16	256
x_{i}	22	28	16	66	1140
$x_{i.}^2$	484	784	256	1524	

Squares

16	36	4
36	100	36
25	49	16
49	25	16
		$\sum\sum x_{ij}^2 = 408$

2^{nd} Example

Correction Factor: $CF = \frac{G^2}{n} = \frac{(66)^2}{12} = \frac{4356}{12} = 363$

Total Sum of Squares: $TSS = \sum\sum x_{ij}^2 - C.F$

$= 408 - 363 = 45$

Sum of Squares due to Treatments: $SST = \frac{\sum_{i=1}^{k} x_{.j}^2}{k} - C.F$

$= \frac{1140}{3} - 363$

$= 380 - 363 = 17$

2nd Example

Sum of Squares due to Blocks:

$$SSB = \frac{\sum_{i=1}^{k} x_{\cdot j}^2}{k} - C.F$$

$$= \frac{1524}{4} - 363$$

$$= 381 - 363$$

$$= 18$$

Sum of Squares due to Error:

$$SSE = TSS - SST - SSB$$

$$= 45 - 17 - 18 = 10$$

Mean sum of squares:

$$MST = \frac{SST}{k-1} = \frac{17}{2} = 8.5$$

$$MSB = \frac{SSB}{m-1} = \frac{18}{3} = 6$$

$$MSE = \frac{SSE}{(k-1)(m-1)} = \frac{10}{6} = 1.67$$

2nd Example

- ANOVA Table (two-way)

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	F-ratio
Between treatments (Programs)	17	3	5.67	$F_{ot} = \frac{5.67}{1.67} = 3.40$
Between blocks (Salesmen)	18	2	9	$F_{ob} = \frac{9}{1.67} = 5.39$
Error	10	6	1.67	
Total		11		

2nd Example

- **Step 6 : Critical values**
  - $f_{(3,\ 6),0.05} = 4.7571$ (for treatments)
  - $f_{(2,\ 6),0.05} = 5.1456$ (for blocks)

2nd Example

- **Step 7 : Decision**
- (i) Calculated $F_{0t} = 3.40 < f_{(3, 6),0.05} = 4.7571$, the null hypothesis is not rejected, and we conclude that there does not exist significant difference in the mean sales among the three programs.
- (ii) Calculate $F_{0b} = 5.39 > f_{(2, 6),0.05} = 5.1456$, the null hypothesis is rejected and conclude that there is significant difference in the mean sales among the four salesmen.

References

- https://www.statology.org/two-way-anova/
- https://thebusinessprofessor.com/en_US/research-analysis-decision-science/two-way-anova-definition
- https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-6-two-way-analysis-of-variance/
- https://www.brainkart.com/article/Two-Way-ANOVA_39242/

Thank you