SCSI2143: PROBABILITY & STATISTICAL DATA ANALYSIS

CHAPTER 8

Analysis of Variance (ANOVA)

Outline

- Introduction
- One-Way ANOVA with Equal Sample Sizes.
- One-Way ANOVA with Unequal Sample Sizes.
- Two-Way ANOVA.

Not covered in this syllabus

Introduction

- ANOVA is a method of testing the equality of three or more population means by analyzing sample variances.
- The purpose of ANOVA is to test for significant differences between Means.
- Elementary concepts provide a brief introduction to the basics of statistical significance testing.

One-Way ANOVA with Equal Sample Sizes

- We assume that the populations have normal distribution and same variance (or standard deviation), and the samples are random and independent of each other.

ANOVA Notation:

n = size of each sample

k = number of populations or treatments being compared

$s^2_{\overline{X}}$ = variance sample means

s^2_P = pooled variance obtained by calculating the mean of the sample variances

One-Way ANOVA with Equal Sample Sizes

- Test statistic: $$F = \frac{\text{variance between sample}}{\text{variance within sample}} = \frac{nS_{\bar{X}}^2}{S_P^2}$$

- Variance between sample:
  - Also called variation due to treatment
  - An estimate of the common population variance σ^2 that is based on the variability among the sample **means.**
- Variance within sample:
  - Also called variation due to error.
  - An estimate of the common population variance σ^2 based on the sample **variances.**

One-Way ANOVA with Equal Sample Sizes

- The critical value of *F*:
  - numerator degrees of freedom = $k - 1$
  - denominator degrees of freedom = $k(n - 1)$
  - k = number of population or treatments being compared.
  - n = sample size

Example 1

Table below lists the head injury to car crash test dummies for four different types of cars. Use a 0.05 significance level to test the null hypothesis that the different types of car have the same mean.

No. of head injury			
Subcompact Cars	Compact Cars	Midsize Cars	Full-Size Cars
681	643	469	384
428	655	727	656
917	442	525	602
898	514	454	687
420	525	259	360

Example 1

- Step 1: Define hypothesis.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1 : at least one mean is different.

- Step 2: For each category, find $n, \bar{x}$ and s

Example 1

- Category 1: (Subcompact Cars)

$$n = 5$$

$$\bar{x} = \frac{681+428+917+898+420}{5} = 668.8$$

$$s = \sqrt{\frac{(681-668.8)^2+(428-668.8)^2+(917-668.8)^2+(898-668.8)^2+(420-668.8)^2}{5-1}}$$

$$= 242.0$$

- Category 2: (Compact Cars)

$$n = 5$$

$$\bar{x} = \frac{643+655+442+514+525}{5} = 555.8$$

$$s = \sqrt{\frac{(643-555.8)^2+(655-555.8)^2+(442-555.8)^2+(514-555.8)^2+(525-555.8)^2}{5-1}}$$

$$= 91.0$$

Example 1

- Category 3: (Midsize Cars)

$$n = 5$$

$$\bar{x} = \frac{469+727+525+454+259}{5} = 486.8$$

$$s = \sqrt{\frac{(469-486.8)^2+(727-486.8)^2+(525-486.8)^2+(454-486.8)^2+(259-486.8)^2}{5-1}}$$

$$= 167.7$$

- Category 4: (Fullsize Cars)

$$n = 5$$

$$\bar{x} = \frac{384+656+602+687+360}{5} = 537.8$$

$$s = \sqrt{\frac{(384-537.8)^2+(656-537.8)^2+(602-537.8)^2+(687-537.8)^2+(360-537.8)^2}{5-1}}$$

$$= 154.6$$

Example 1

- Step 3: Find variance between samples

Step 3a: Find mean between samples

$$\bar{\bar{x}} = \frac{668.8 + 555.8 + 486.8 + 537.8}{k = 4} = 562.3$$

Step 3b: Find standard deviation between samples

$$s_{\bar{x}} = \sqrt{\frac{(668.8 - 562.3)^2 + (555.8 - 562.3)^2 + (486.8 - 562.3)^2 + (537.8 - 562.3)^2}{4 - 1}}$$

$$= 76.779$$

Step 3c: Find variance between samples

$$ns_{\bar{x}}^2 = 5(76.779)^2 = 29475.1$$

Example 1

- Step 4: Find variance within samples

$$s_p^2 = \frac{(242.0)^2 + (91.0)^2 + (167.7)^2 + (154.6)^2}{k = 4}$$

$$= 29717.4$$

- Step 5: Calculate test statistic, *F*

$$F = \frac{\text{variance between samples}}{\text{variance within samples}} = \frac{ns_{\bar{x}}^2}{s_p^2}$$

$$= \frac{29475.1}{29717.4}$$

$$= 0.992$$

Example 1

- Step 6: Calculate numerator and denominator degree of freedom
  - Numerator = $k - 1 = 4 - 1 = 3$
  - Denominator = $k(n - 1) = 4(5 - 1) = 16$

- Step 7: Find critical value of F with $\alpha = 0.05$ from F-distribution table
  - F critical value = 3.2389

Example 1

- Step 8: Test the claim and state the conclusion.
  - Since $F_{test\ statistic} < F_{critical\ value}$ (0.992 < 3.2389), we fail to reject the null hypothesis.
  - There is sufficient evidence to claim that the different types of cars have the same mean for head injury.

Example 2

Table below lists the chest deceleration to car crash test dummies for four different types of cars. Use a 0.05 significance level to test the null hypothesis that the different types of car have the same mean.

Subcompact Cars	Compact Cars	Midsize Cars	Full-Size Cars
n = 5	n = 5	n = 5	n = 5
$\bar{x} = 50.4$	$\bar{x} = 53.0$	$\bar{x} = 48.8$	$\bar{x} = 46.0$
s = 6.69	s = 4.64	s = 3.35	s = 7.11

Example 2

- Step 1: State the hypothesis.
  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
  H_1 : at least one mean is different.
- Step 2: (Done – refer table).
- Step 3: Find variance between samples

Step 3a: Find mean between samples

$$\bar{\bar{x}} = \frac{50.4 + 53.0 + 48.8 + 46.0}{4} = 49.55$$

Step 3b: Find standard deviation between samples

$$s_{\bar{x}} = \sqrt{\frac{(50.4-49.55)^2 + (53.0-49.55)^2 + (48.8-49.55)^2 + (46.0-49.55)^2}{4-1}}$$

$$= 2.93$$

Step 3c: Find variance between samples

$$ns_{\bar{x}}^2 = 5(2.93)^2 = 42.92$$

Example 2

- Step 4: Find variance within samples

$$s_p^2 = \frac{(6.69)^2 + (4.64)^2 + (3.35)^2 + (7.11)^2}{4}$$
$$= 32.02$$

- Step 5: Calculate test statistic, *F*

$$F = \frac{\text{variance between samples}}{\text{variance within samples}} = \frac{ns_{\bar{x}}^2}{s_p^2}$$
$$= \frac{42.92}{32.02}$$
$$= 1.34$$

Example 2

- Step 6: Calculate numerator and denominator degree of freedom
  - Numerator = $k - 1 = 4 - 1 = 3$
  - Denominator = $k(n - 1) = 4(5 - 1) = 16$

- Step 7: Find critical value of *F* with $\alpha = 0.05$ from *F*-distribution table
  - *F* critical value = 3.2389

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Example 2

- Step 8: Test the claim and state the conclusion.
  - Since $F_{test\ statistic} < F_{critical\ value}$ (1.34 < 3.2389), we fail to reject the null hypothesis.
  - There is sufficient evidence to claim that the different types of cars have the same mean for chest deceleration.

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Exercise

Table 1 lists the body temperatures of 5 randomly selected subjects from each of 3 different age groups. Informal examination of the 3 sample means (97.940, 98.580, 97.800) seems to suggest that the 3 samples come from populations with means that are not significantly different. Test the claim that the 3 age-group populations have the same mean body temperature. Use α= 0.05.

Exercise

Table 1: Body Temperature (ºF) Categorized by Age

18-20	21-29	30 and older
98.0	99.6	98.6
98.4	98.2	98.6
97.7	99.0	97.0
98.5	98.2	97.5
97.1	97.9	97.3

n_1=5	n_2=5	n_3=5
Mean: x_1=97.940	x_2=98.580	x_3=97.800
Std. Dev: s_1=0.568	s_2=0.701	s_3=0.752