Chapter 9 – ANOVA

one-way analysis of variance (ANOVA) tests measure significant effects of one factor only, two-way analysis of variance (ANOVA) tests (also called two-factor analysis of variance) measure the effects of two factors simultaneously. For example, an experiment might be defined by two parameters, such as treatment and time point. One-way ANOVA tests would be able to assess only the treatment effect or the time effect. Two-way ANOVA on the other hand would not only be able to assess both time and treatment in the same test, but also whether there is an interaction between the parameters.

One way ANOVA

Investigates how much of variations in grouped data comes from differences between the groups and how much is just random observational error.
performs a hypothesis test:

H_0 : $\mu_1 = \mu_2 = ... = \mu_k$

H_1 : At least one is different i.e. $\mu_i \neq \mu_j$ **for i≠j**

Two way ANOVA : Allows us to analyze both row effect and column effect,i.e. differences between rows and differences between columns.
splits the variations among the observations into row effect , column effect and random error.
Performs a hypothesis test :

H_0 : $\alpha_i = 0$ **for all i (testing no row effect at all)**

H_1 : At least one is none zero

H_0 : $\beta_j = 0$ **for all j (testing no column effect at all)**

H_1 : At least one is none zero

Interaction in 2-way ANOVA :

Additive structure $\mu_{ij} = \mu + \alpha_i + \beta_j$ **allows us to talk clearly about differences in row/column differences**
2-way ANOVA allows us to analyze the interaction between the two variables (row variables & column variables)so we can study how combinations of these variables influence behavior.
Interaction describes how the effect of one independent variable is influenced/effected by the value of the other independent variable.
Equivalently , the effect of one independent variable varies with the value of the other independent variable.

Construction of 2- way ANOVA Model

- We need a continuous r.v.
- 2 characteristics described as r rows , c columns such that the r.v. Xij appears in the cell corresponding to the ith row and jth column i = 1,…,r and j= 1,…., c

-Define E[Xij]= μ_{ij}

-Key Model assumptions : Xij are normal r.v.'s with mean μ_{ij} and constant variance σ^2 : Xij ~ N(μ_{ij}, σ^2)

-Additional assumption: cell population means μ_{ij} have additive structure: $\mu_{ij} = \mu + \alpha_i + \beta_j$

μ : overall mean , α_i : row effect , β_j : column effect

Steps in building the ANOVA Table : 2 – way Anova

1. State the usual 2-way ANOVA assumptions.
2. Calculate row means : $\bar{x}_{1.}$,….
3. Calculate overall mean x..
4. Calculate Sample row means variance : S_R^2
5. Calculate Mean sum squares between rows (mean SS) = c S_R^2 , (c = No. of columns)
6. Calculate sum squares between rows(SS) = **(r-1) c** S_R^2 , (r= No. of rows)
7. Calculate column means $\bar{x}_{.1}$, ...
8. Calculate Sample column means variance S_C^2
9. Calculate Mean sum squares between columns (mean SS) = **r** S_C^2
10. Calculate Sum squares between columns (SS) = **(c-1) r** S_C^2
11. Calculate the sample variance of all observations : S_T
12. Calculate total sum squares : **(rc - 1)** S_T^2
13. Calculate Residual Sum Squares = Total SS – SS between rows – SS between columns
14. Calculate Mean Residual SS = **S²** = $\dfrac{ResidualSS}{(r-1)(c-1)}$
15. Fill the ANOVA Table :

Source	D.F.	Sum squresSS	MSE	F-value
row	r - 1	Row SS	Mean row SS	$F_R = c S_R^2 / S^2$
column	c - 1	Column SS	Mean Column SS	$F_C = r S_C^2 / S^2$
Error	(r-1)(c-1)	Residual SS	Mean Residual SS	
Total	rc – 1			

Hypothesis testing :

- Testing row effect : $H_0 : \alpha_i = 0 \quad \forall i$

H_1 : non zero row effect

Under H_0 : $F_R = c S_R^2 / S^2 \sim F_{r-1,(r-1)(c-1)}$

Criterion: Reject H_0 if $F_R \geq F_{\alpha,r-1,(r-1)(c-1)}$

- Testing column effect: $H_0 : \beta_j = 0 \quad \forall j$

H1 : non zero column effect

Under H_0 : $F_C = \dfrac{rS_C^2}{S^2} \sim F_{c-1,(r-1)(c-1)}$

The criterion : Reject H_0 if $F_C \geq F_{\alpha,c-1,(r-1)(c-1)}$

Simultaneous C.I.

(SCI) used for all possible pairs of α_i's

The confidence interval in case of difference of population means α_i - α_j is given by:

$$\overline{X}_i - \overline{X}_j \pm S\sqrt{(r-1)F_{\alpha,r-1,(r-1)(c-1)}(2/c)}$$

S^2: Mean Residual SS

One Way – ANOVA One characteristic

Collection of k independent samples of constant variance population σ^2

Steps :

1. Mean within group SS : $S^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + \ldots}{n-k}$
2. The within group Sum of squares : $(n-k)S^2$, n =No. of observations , k= No. of samples
3. The variance of all observations : S_T^2
4. Total Sum of Squares **:** $(n-1)\ S_T^2$
5. Between groups Sum of Squares = $(k-1)\ S_B^2$ **=** Total Sum of Squares- within group SS
6. ANOVA Table

Source	D.F.	Sum squresSS	MSE	F-value
Between Groups	k -1	Between Groups SS	Mean Between Groups SS	$F = S_B^2 / S^2$
Within Groups	n – k	Within Groups SS	Mean Within Groups SS	
Total	n – 1	**total**		

Hypothesis testing :

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots\ldots = \mu_k$, H_1 : Not all the μ_i are equal

The test statistic : $F = \dfrac{S_B^2}{S^2}$, Reject H_0 if $\dfrac{S_B^2}{S^2} > F_{\alpha,k-1,n-k}$

Confidence Intervals :

Two -way ANOVA ***(Example 63 S.G , p107)*** , r = 4 , c = 3

Potato			
Location	A	B	C
1	18	13	12
2	20	23	21
3	14	12	9
4	11	17	10

$\bar{x}_{.1} = 15.75$ $\bar{x}_{.2} = 16.25$, Mean Residual SS = S² =7.03

1. Single interval :

For $\alpha_i - \alpha_j$: $\bar{X}_i - \bar{X}_j \pm t_{\frac{\alpha}{2},(r-1)(c-1)} S\sqrt{2/c}$

For $\beta_i - \beta_j$: $\bar{X}_i - \bar{X}_j \pm t_{\frac{\alpha}{2},(r-1)(c-1)} S\sqrt{2/r}$

e.g. a 95% C.I. for $\beta_1 - \beta_2$:

$15.75 - 16.25 \pm t_{0.025,6}\sqrt{7.03}\sqrt{2/4}$ = -0.5 $\pm$ 4.69 (Table 10 : $t_{0.025,6}$ = 2.447)

2. Simultaneous Intervals :

For $\alpha_i - \alpha_j$: $\bar{X}_i - \bar{X}_j \pm S\sqrt{(r-1)F_{\alpha,r-1,(r-1)(c-1)}(2/c)}$

For $\beta_i - \beta_j$: $\bar{X}_i - \bar{X}_j \pm S\sqrt{(c-1)F_{\alpha,c-1,(r-1)(c-1)}(2/r)}$

e.g. a 95% simultaneous C.I. for $\beta_1 - \beta_2$:

$15.75 - 16.25 \pm \sqrt{7.03}\sqrt{2F_{0.05,2,6}(2/4)}$ = - 0.5 $\pm$ 6.01(Table 12b : $F_{0.05,2,6}$ = 5.143)

One -way ANOVA : ***(Example 61 S.G , p100)*** , n = 19 , k = 4

Cereals			
1	2	3	4
9.3	13.4	12.5	14.0
10.8	12.2	14.7	15.6
8.4	12.4	12.9	14.1
9.7	12.8	11.8	
9.5	12.2		
7.9			
9.5			

$\bar{X}_1$ = 9.3 $\bar{X}_2$ = 12.6 , S : Mean within group SS : S² = 0.8330

1. Single interval : for $\mu_i - \mu_j$

$\bar{X}_i - \bar{X}_j \pm t_{\alpha,n-k} S\sqrt{(1/n_i + 1/n_j)}$

e.g. a 95% C.I. for $\mu_1 - \mu_2$

$\bar{X}_1 - \bar{X}_2 \pm t_{0.05,15}\sqrt{0.8330}\sqrt{(1/7 + 1/5)}$, (Table 10 : $t_{0.05,15}$ = 2.13)

9.3 - 12.6 $\pm$ (2.13)(0.913)(0.5855) = - 3.3 $\pm$ 1.1386

2. Simultaneous Interval :

$\bar{X}_i - \bar{X}_j \pm S\sqrt{(k-1)F_{\alpha,k-1,n-k}(1/n_i + 1/n_j)}$

e.g. a 95% simultaneous C.I. for $\mu_1 - \mu_2$:

9.3 - 12.6 $\pm$ (0.913) $\sqrt{F_{0.05,3,15}(1/7 + 1/5)}$ = -3.3 $\pm$ 1.68 (Table 12b : $F_{0.05,3,15}$ = 3.287)