

ANALYSIS OF VARIANCE (ANOVA)



DR. SANGEETA MOHANTY

ANALYSIS OF VARIANCE IS A STATISTICAL METHOD FOR MAKING SIMULTANEOUS COMPARISONS BETWEEN TWO OR MORE MEANS OF A DEPENDENT VARIABLE AFFECTED BY NUMEROUS FACTORS OF AN INDEPENDENT VARIABLE. IT INVESTIGATES THE CAUSES OF DIFFERENCES OCCURRING IN THE MEAN VALUES OF TWO OR MORE POPULATIONS UNDER STUDY. THE NULL HYPOTHESIS FOR SUCH CASES IS 'ALL MEANS ARE EQUAL'. TYPES OF ANOVA

- I. ONE-WAY ANOVA**
- II. TWO-WAY ANOVA**
- III. FACTORIAL DESIGN**

ANALYSIS OF VARIANCE (ANOVA)

INTRODUCTION

ANALYSIS OF VARIANCE IS A STATISTICAL METHOD FOR MAKING SIMULTANEOUS COMPARISONS BETWEEN TWO OR MORE MEANS OF A DEPENDENT VARIABLE AFFECTED BY NUMEROUS FACTORS OF AN INDEPENDENT VARIABLE. IT INVESTIGATES THE CAUSES OF DIFFERENCES OCCURRING IN THE MEAN VALUES OF TWO OR MORE POPULATIONS UNDER STUDY. THE NULL HYPOTHESIS FOR SUCH CASES IS 'ALL MEANS ARE EQUAL'.

TYPES OF ANOVA

- I. ONE-WAY ANOVA
- II. TWO-WAY ANOVA
- III. FACTORIAL DESIGN

i) ONE WAY ANOVA

THE ONE-WAY ANALYSIS OF VARIANCE (ANOVA) IS USED TO DETERMINE WHETHER THERE IS ANY SIGNIFICANT DIFFERENCE BETWEEN THE MEANS OF A DEPENDENT VARIABLE FOR THREE OR MORE UNRELATED GROUPS OF AN INDEPENDENT VARIABLE. ONE-WAY ANOVA IS PARTICULARLY USED WHEN THERE IS ONLY ONE INDEPENDENT VARIABLE AND ONE DEPENDENT VARIABLE. THE DEPENDENT VARIABLE IS METRIC (INTERVAL OR RATIO SCALE) WHEREAS THE INDEPENDENT VARIABLE IS CATEGORICAL OR NOMINAL IN NATURE.

CASE ANALYSIS-1

PROBLEM

A MANAGER OF XYZ ELECTRIC BULBS COMPANY WANTS TO COMPARE THE LIFE SPAN OF THREE DIFFERENT BRANDS OF BULBS AVAILABLE IN THE MARKET. THE MANGER COLLECTED THE DATA OF LIFE SPAN OF THE BULBS OF BRANDS A, B AND C MEASURED IN HUNDREDS OF HOURS DURING THREE MONTHS PERIOD AS SHOWN BELOW.

TABLE-1: SAMPLE DATA

BRAND		
A	B	C
21	12	11
18	14	14
12	9	19

THE PURPOSE IS TEST WHETHER THE LIFETIMES OF FOUR BRANDS OF ELECTRIC BULBS ARE EQUAL OR NOT.

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : THE MEAN LIFETIMES FOR THREE BRANDS OF BULBS ARE EQUAL.

ALTERNATIVE HYPOTHESIS- H_1 : THE MEAN LIFE TIME OF AT LEAST TWO BRANDS OF BULBS DIFFER.

INPUT DATA

THE VARIABLE 'LIFE TIME' IS DEPENDENT VARIABLE AND THE VARIABLE 'BRAND' IS INDEPENDENT VARIABLE. THE INDEPENDENT VARIABLE IS CODED AS: 1 = BRAND A, 2 = BRAND B, 3 = BRAND C. THE FOLLOWING TABLE DEPICTS THE DEPENDENT VARIABLE ALONG WITH THE CODED INDEPENDENT VARIABLE AND IT IS TREATED AS THE INPUT DATA MATRIX FOR THE ANALYSIS.

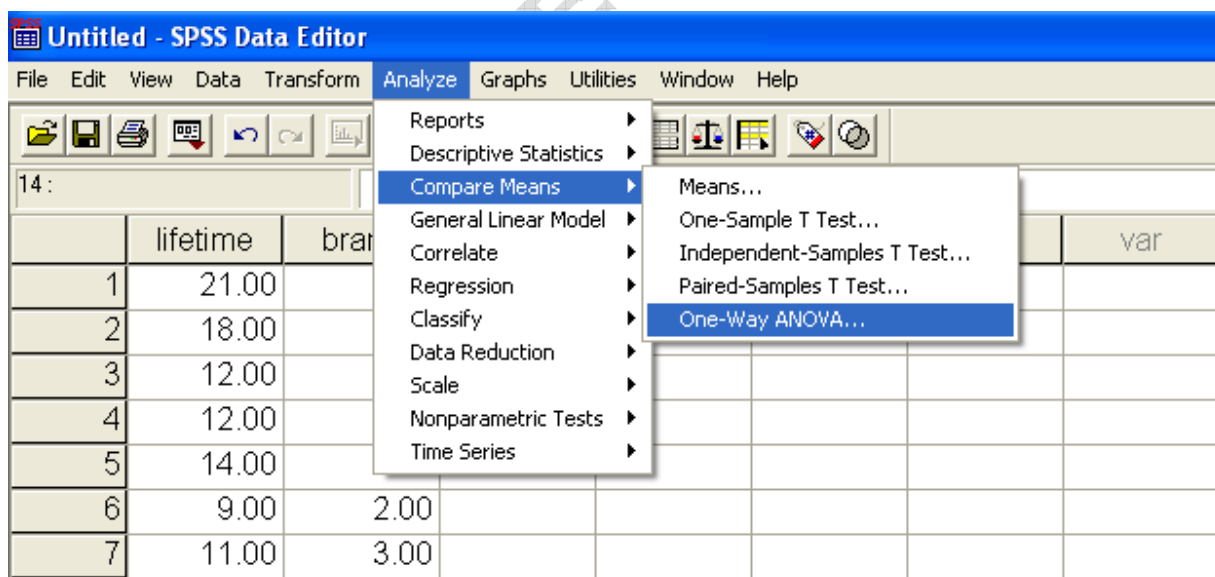
TABLE-2: INPUT DATA

SERIAL NUMBER	LIFE TIME (IN HUNDREDS OF HOURS)	BRAND
1	21	1
2	18	1
3	12	1
4	12	2
5	14	2
6	9	2
7	11	3
8	14	3
9	19	3

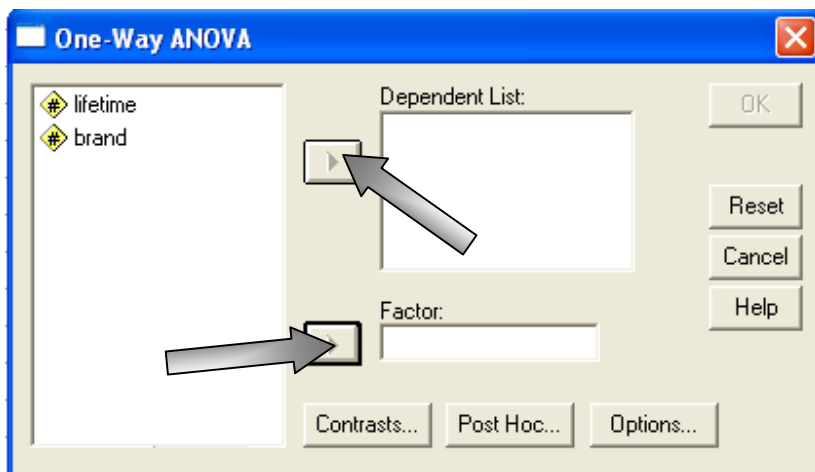
PERFORMING THE ANALYSIS WITH SPSS

FOR SPSS VERSION 11, CLICK ON **ANALYZE** → **COMPARE MEANS ONE-WAY ANOVA**

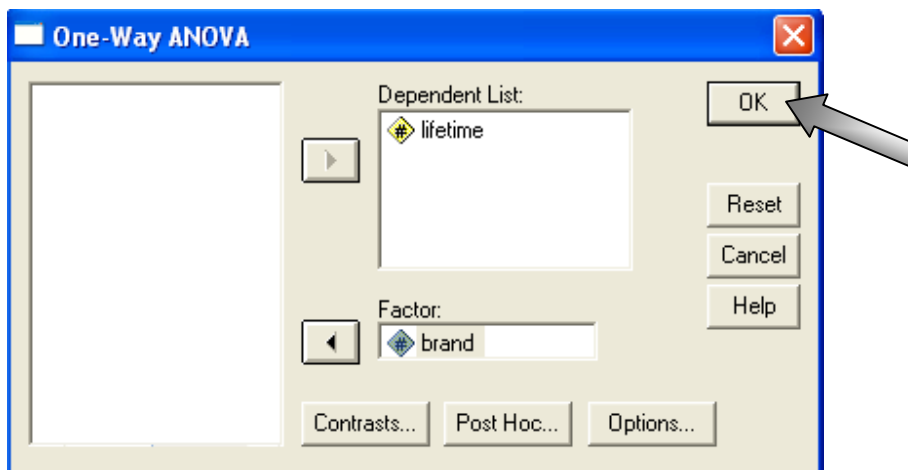
THIS WILL BRING UP THE SPSS SCREEN DIALOGUE BOX AS SHOWN BELOW.



AFTER CLICKING **ONE-WAY ANOVA**, THIS WILL BRING UP THE FOLLOWING SPSS SCREEN DIALOGUE BOX



SELECT THE DEPENDENT VARIABLE AND MOVE IT TO THE **DEPENDENT LIST** BOX. SIMILARLY SELECT THE CODED VARIABLE AND MOVE IT TO **FACTOR** BOX.



NOW, CLICK OK TO GET THE OUTPUT.

SPSS OUTPUT

THE SPSS OUTPUT OF THE ANALYSIS IS GIVEN IN THE FOLLOWING TABLE.

TABLE-3: ANOVA
LIFETIME

	SUM OF	DF	MEAN SQUARE	F	SIG.
BETWEEN	42.889	2	21.444	1.473	.302
WITHIN GROUPS	87.333	6	14.556		
TOTAL	130.222	8			

FROM THE OUTPUT, $F = 1.473$

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) ≤ 0.05

INTERPRETATION

THE P-VALUE IS 0.302 AND IT IS LESS THAN 0.05 (5% LEVEL OF SIGNIFICANCE), SO WE REJECT THE NULL HYPOTHESIS AND ACCEPT THE ALTERNATIVE HYPOTHESIS AT 5% LEVEL OF SIGNIFICANCE. IT CAN BE CONCLUDED THAT THE LIFE SPAN OF DIFFERENT BRANDS OF BULBS DIFFER SIGNIFICANTLY.

CASE ANALYSIS-2**PROBLEM**

A PEN MANUFACTURING UNIT IS INTERESTED TO COMPARE THE AVERAGE SALES OF ITS FOUR SALESMEN X, Y, Z AND W. THE FOLLOWING TABLE DEPICTS THE WEEKLY SALES BY THE SALESMEN IN HUNDREDS OF UNITS.

TABLE-1: SAMPLE DATA

SALESMEN			
X	Y	Z	W
15	22	14	10
18	8	9	11
9	13	8	18

THE PURPOSE IS TEST WHETHER THE LIFETIMES OF FOUR BRANDS OF ELECTRIC BULBS ARE EQUAL OR NOT.

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : THE AVERAGE SALES BY THREE SALESMEN ARE EQUAL.

ALTERNATIVE HYPOTHESIS- H_1 : THE AVERAGE SALES OF AT LEAST TWO SALESMEN DIFFER.

INPUT DATA

THE VARIABLE 'SALES' IS DEPENDENT VARIABLE AND THE VARIABLE 'SALESMEN' IS INDEPENDENT VARIABLE. THE INDEPENDENT VARIABLE IS CODED AS: 1 = SALESMAN X, 2 = SALESMAN Y, 3 = SALESMAN Z, 4 = SALESMAN W. THE FOLLOWING TABLE IS USED AS THE INPUT DATA FOR THE ANALYSIS.

TABLE-2: INPUT DATA

SERIAL NUMBER	SALES (IN HUNDREDS OF UNITS)	SALESMEN
1	15	1
2	18	1
3	9	1
4	22	2
5	8	2
6	13	2
7	14	3
8	9	3

9	8	3
10	10	4
11	11	4
12	18	4

SPSS OUTPUT

THE SPSS OUTPUT OF THE ANALYSIS IS GIVEN IN THE FOLLOWING TABLE.

TABLE-3: ANOVA

	SUM OF	DF	MEAN SQUARE	F	SIG.
BETWEEN	29.583	3	9.861	.392	.762
WITHIN GROUPS	201.333	8	25.167		
TOTAL	230.917	11			

FROM THE OUTPUT, $F = 0.392$

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) ≤ 0.05

INTERPRETATION

THE P-VALUE IS 0.762 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE), SO WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT THE AVERAGE SALES BY THE SALESMEN DO NOT DIFFER SIGNIFICANTLY.

SPSS COMMAND

1. CLICK ON ANALYZE AT THE SPSS MENU BAR (IN OLDER VERSIONS OF SPSS, CLICK ON STATISTICS INSTEAD OF ANALYZE).
2. CLICK ON COMPARE MEANS FOLLOWED BY ONE WAY ANOVA.
3. SELECT THE APPROPRIATE VARIABLE AND MOVE IT TO THE DEPENDENT LIST. SIMILARLY SELECT THE CODED VARIABLE AND MOVE IT TO FACTOR BOX.
4. CLICK OK TO GET THE OUTPUT.

ii) TWO-WAY ANOVA

A TWO-WAY ANOVA IS USED WHEN THERE IS THE PRESENCE OF AN ADDITIONAL VARIABLE AFFECTING THE RELATIONSHIP BETWEEN THE INDEPENDENT AND DEPENDENT VARIABLE. THE ADDITIONAL VARIABLE, SO COMES INTO VIEW IS TREATED AS SECOND INDEPENDENT VARIABLE. TWO-WAY ANOVA INVESTIGATES THE SOURCES OF VARIATIONS ARISING IN THE MEAN VALUES OF DEPENDENT VARIABLE DUE TO THE FIXED FACTORS. A "FACTOR" IS ANOTHER NAME FOR AN INDEPENDENT VARIABLE. THE FIXED FACTORS ARE TWO INDEPENDENT VARIABLES INFLUENCING THE DEPENDENT VARIABLE. THUS, A TWO WAY ANOVA USES TWO NULL HYPOTHESES FOR TWO INDEPENDENT VARIABLES WITH 'ALL MEANS EQUAL'.

CASE ANALYSIS-1

PROBLEM

AN EXPERIMENT IS CARRIED OUT TO TEST WHETHER THREE DIFFERENT TYPES OF DRUG A, B, C HAVE DIFFERENT EFFECTS ON MALES AND FEMALES. A SAMPLE OF 15 MALE AND 15 FEMALE RESPONDENTS (30 RESPONDENTS)

WERE SELECTED AND THE CURE PERCENTAGE AFTER USING THE DRUGS WAS RECORDED AS FOLLOWS.

TABLE-1: SAMPLE DATA

DRUG	CURE PERCENTAGE					
	A		B		C	
SERIAL NUMBER	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
1	55	90	75	30	70	55
2	58	80	85	43	80	65
3	65	70	95	60	90	75
4	90	45	90	50	40	85
5	75	65	75	80	45	95

THE PURPOSE IS TO TEST WHETHER THE LIFETIMES OF FOUR BRANDS OF ELECTRIC BULBS ARE EQUAL OR NOT.

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS-

H_{01} : THE AVERAGE PERCENTAGE CURED IS SAME FOR ALL THE DRUGS.

H_{02} : THE GENDER HAS NO EFFECT ON PERCENTAGE CURED.

ALTERNATIVE HYPOTHESIS-

H_{11} : THE AVERAGE PERCENTAGE CURED OF AT LEAST TWO DRUGS DIFFERS.

H_{12} : THE GENDER HAS EFFECT ON PERCENTAGE CURED.

INPUT DATA

THE VARIABLE 'PERCENTAGES CURED' IS DEPENDENT VARIABLE AND THE VARIABLE 'DRUG' IS INDEPENDENT VARIABLE. THE THIRD VARIABLE IS 'GENDER AND IT INFLUENCES THE RELATIONSHIP BETWEEN THE INDEPENDENT AND THE DEPENDENT VARIABLES. THE FIRST FIXED FACTOR IS THE INDEPENDENT VARIABLE 'DRUG' AND THE SECOND FIXED FACTOR IS 'GENDER'. THE CODES USED FOR THE VARIABLES ARE: 1 = MALE, 2 = FEMALE; 1 = DRUG A, 2 = DRUG B, 3 = DRUG C. THE FOLLOWING TABLE SHOWS THE DEPENDENT VARIABLE ALONG WITH THE CODED VARIABLES FOR FIXED FACTORS AND THE TABLE IS USED AS THE INPUT MATRIX FOR THE ANALYSIS.

TABLE-1: INPUT OUTPUT

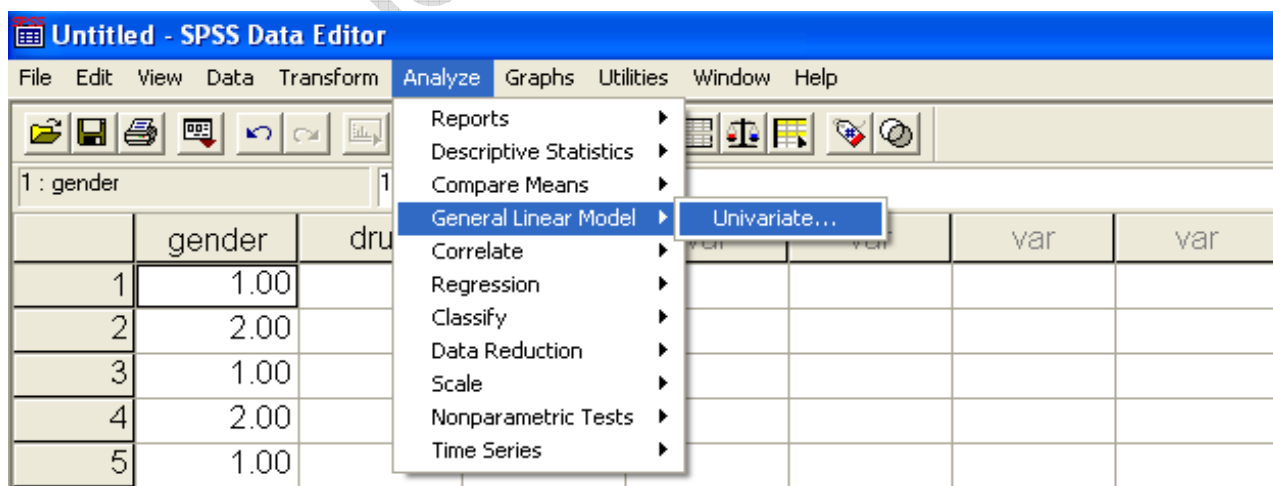
SERIAL NUMBER	GENDER	DRUGS	PERCENTAGE CURED
1	1	1	55
2	2	1	90
3	1	2	75
4	2	2	30
5	1	3	70
6	2	3	55
7	1	1	58

8	2	1	80
9	1	2	85
10	2	2	43
11	1	3	80
12	2	3	65
13	1	1	65
14	2	1	70
15	1	2	95
16	2	2	60
17	1	3	90
18	2	3	75
19	1	1	90
20	2	1	45
21	1	2	90
22	2	2	50
23	1	3	40
24	2	3	85
25	1	1	75
26	2	1	65
27	1	2	75
28	2	2	80
29	1	3	45
30	2	3	95

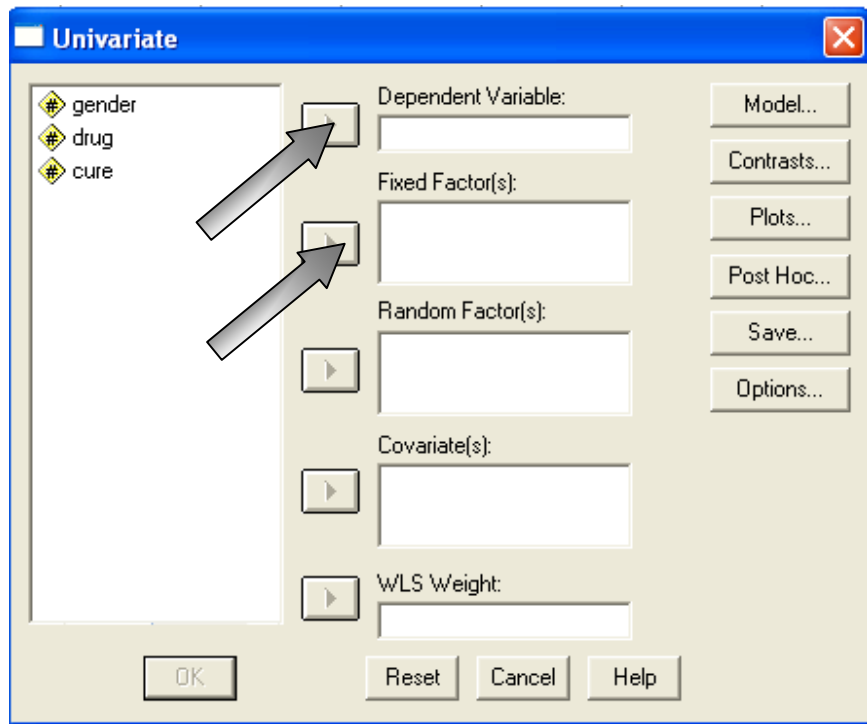
PERFORMING THE ANALYSIS WITH SPSS

FOR SPSS VERSION 11, CLICK ON **ANALYZE** → **GENERAL LINEAR MODEL** → **UNIVARIATE**

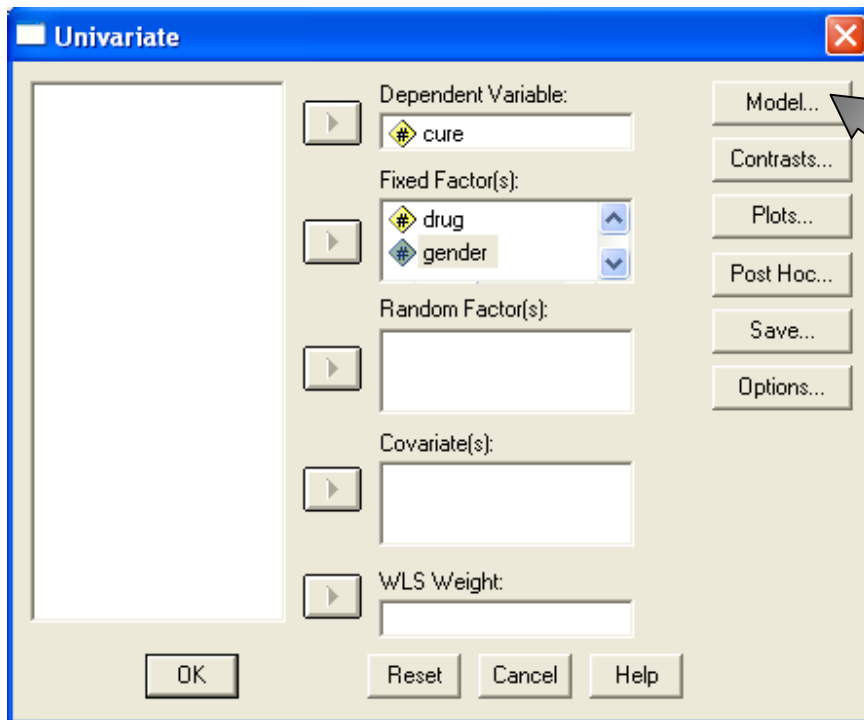
THIS WILL BRING UP THE SPSS SCREEN DIALOGUE BOX AS SHOWN BELOW.



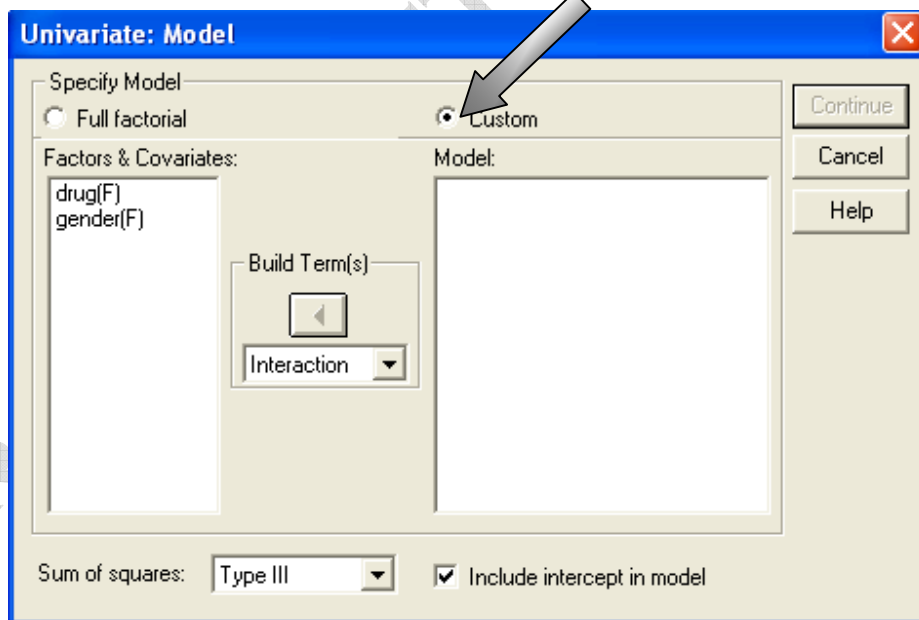
AFTER CLICKING **UNIVARIATE**, THIS WILL BRING UP THE FOLLOWING SPSS SCREEN DIALOGUE BOX.



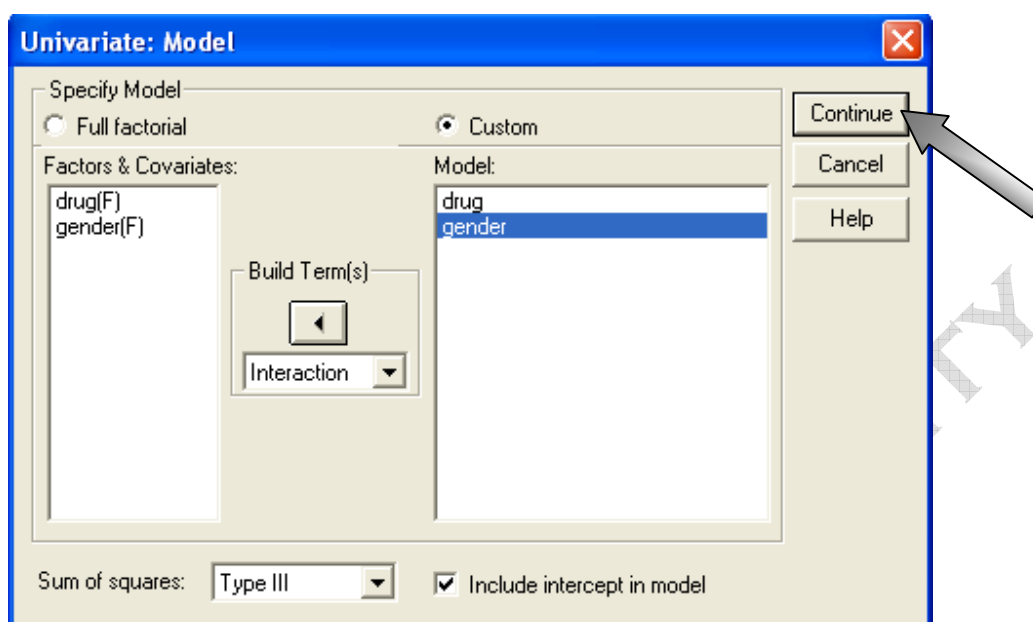
SELECT THE DEPENDENT VARIABLE “CURE” AND CLICK IT TO MOVE TO DEPENDENT VARIABLES BOX. SELECT THE FIRST FIXED FACTORS ‘DRUG’, THE SECOND FIXED FACTOR ‘GENDER’ AND MOVE THEM TO FIXED FACTOR(S) BOX.



THEN CLICK MODEL FOLLOWED BY CUSTOM. THIS WILL BRING UP THE FOLLOWING DIALOGUE BOX.



SELECT BOTH THE FACTORS 'DRUG' AND 'GENDER' AND MOVE THEM TO THE BOX CALLED MODEL.



THEN CLICK **CONTINUE** TO RETURN THE MAIN DIALOGUE BOX AND CLICK **OK** OF THE MAIN DIALOGUE BOX TO GET THE OUTPUT.

SPSS OUTPUT

THE SPSS OUTPUTS ARE ILLUSTRATED IN THE FOLLOWING TABLES.

TABLE-3: BETWEEN-SUBJECTS FACTORS

		N
GENDER	1.00	15
	2.00	15
DRUG	1.00	10
	2.00	10
	3.00	10

TABLE-3: TESTS OF BETWEEN-SUBJECTS EFFECTS

DEPENDENT VARIABLE: CURE

SOURCE	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F	SIG.
CORRECTED MODEL	347.933	3	115.978	.336	.800
INTERCEPT	143659.200	1	143659.200	415.900	.000
DRUG	14.600	2	7.300	.021	.979
GENDER	333.333	1	333.333	.965	.335
ERROR	8980.867	26	345.418		
TOTAL	152988.000	30			
CORRECTED TOTAL	9328.800	29			

A R SQUARED = .037 (ADJUSTED R SQUARED = -.074)

FROM THE OUTPUT, F = 0.021 (FOR DRUG) AND F = 0.965 (FOR GENDER)

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) ≤ 0.05

INTERPRETATION

THE P-VALUE CORRESPONDING TO THE HYPOTHESIS FOR DRUG IS 0.979 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT THE DRUGS HAVE SAME EFFECT TO BE USED AS ANTIDANDRUFF DRUG. BUT THE P-VALUE FOR GENDER IS 0.335 AND IT IS LESS THAN 0.05. SO WE REJECT THE NULL HYPOTHESIS AND CONCLUDE THAT THE PERCENTAGE CURED IS AFFECTED BY THE FACTOR GENDER.

CASE ANALYSIS-2

PROBLEM

THE PROBLEM RELATES TO THE POPULAR TELEVISION SHOWS VIEWED BY THE PEOPLE OF BALASORE TOWN. A RANDOM SAMPLE OF 20 RURAL AND 20 URBAN PEOPLE WAS SELECTED FROM BOTH URBAN AND RURAL AREA OF THE TOWN. THE RESPONDENTS WERE ASKED TO RATE THE TV SHOWS ON 1-10 SCALE. (1 = NOT LIKED AT ALL AND 10 = MOST LIKED SHOW)

TABLE-1: SAMPLE DATA

TV SHOWS	RATING							
	REALITY SHOWS		DAILY SERIALS		COMEDY SHOWS		NEWS	
	RURA L	URBA N	RURA L	URBA N	RURA L	URBA N	RURA L	URBA N
1	8	2	10	3	10	9	9	10
2	7	6	9	2	9	8	8	9
3	6	8	9	7	8	9	6	8
4	6	7	8	6	7	8	7	6
5	7	9	8	8	9	9	9	9

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS-

H₀₁: THE MEAN RATING OF TV SHOWS TYPE IS SAME FOR ALL FOUR SHOWS.

H₀₂: THE AREA HAS NO EFFECT ON MEAN RATINGS.

ALTERNATIVE HYPOTHESIS-

H₁₁: THE MEAN RATING OF AT LEAST TWO TV SHOWS DIFFERS.

H₁₂: THE AREA HAS EFFECT ON MEAN RATING.

INPUT DATA

THE VARIABLE 'RATING' IS DEPENDENT VARIABLE AND THE VARIABLE 'TV SHOW' IS INDEPENDENT VARIABLE. THE FIXED FACTORS ARE 'TV SHOWS' AND 'AREA'. THE CODES USED FOR THE VARIABLES ARE: 1 = RURAL, 2 = URBAN: 1 = REALITY SHOWS, 2 = DAILY SERIALS, 3 = COMEDY SHOWS AND 4 = NEWS AS SHOWN IN THE FOLLOWING TABLE AND IT IS USED AS THE INPUT DATA MATRIX FOR THE ANALYSIS.

TABLE-1: INPUT DATA

SERIAL NUMBER	AREA	TV SHOWS	RATING
1	1	1	8
2	1	1	7
3	1	1	6
4	1	1	6
5	1	1	7
6	2	1	2
7	2	1	6
8	2	1	8
9	2	1	7
10	2	1	9
11	1	2	10
12	1	2	9
13	1	2	9
14	1	2	8
15	1	2	8
16	2	2	3
17	2	2	2
18	2	2	7
19	2	2	6
20	2	2	8
21	1	3	10
22	1	3	9
23	1	3	8
24	1	3	7
25	1	3	9
26	2	3	9
27	2	3	8
28	2	3	9
29	2	3	8
30	2	3	9
31	1	4	9
32	1	4	8
33	1	4	6
34	1	4	7
35	1	4	9
36	2	4	10
37	2	4	9
38	2	4	8
39	2	4	6
40	2	4	9

SPSS OUTPUTS

THE SPSS OUTPUTS ARE GIVEN IN THE FOLLOWING TABLES.

TABLE-2: BETWEEN-SUBJECTS FACTORS

		N
TV SHOWS	1.00	10
	2.00	10
	3.00	10
	4.00	10
AREA	1.00	20
	2.00	20

TABLE-2: TESTS OF BETWEEN-SUBJECTS EFFECTS
DEPENDENT VARIABLE: RATINGS

SOURCE	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F	SIG.
CORRECTED MODEL	33.300	4	8.325	2.637	.050
INTERCEPT	2295.225	1	2295.225	727.159	.000
TV SHOWS	26.075	3	8.692	2.754	.057
AREA	7.225	1	7.225	2.289	.139
ERROR	110.475	35	3.156		
TOTAL	2439.000	40			
CORRECTED TOTAL	143.775	39			

A R SQUARED = .232 (ADJUSTED R SQUARED = .144)

FROM THE OUTPUT, $F = 2.754$ (FOR TV SHOWS) AND $F = 2.289$ (FOR AREA)
DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) ≤ 0.05

INTERPRETATION

THE P-VALUE CORRESPONDING TO THE HYPOTHESIS FOR TV SHOWS IS 0.057 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT ALL THE TV SHOWS HAVE EQUAL EFFECT ON VIEWERS. SIMILARLY THE P-VALUE FOR AREA IS 0.139 AND IT IS MORE THAN 0.05. SO WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT THE AREA HAS NO EFFECTS ON MEAN RATING.

SPSS COMMAND

1. CLICK ON ANALYZE AT THE SPSS MENU BAR (IN OLDER VERSIONS OF SPSS, CLICK ON STATISTICS INSTEAD OF ANALYZE).
2. CLICK ON GENERAL LINEAR MODEL FOLLOWED BY UNIVARIATE.
3. SELECT THE APPROPRIATE VARIABLE AND MOVE IT TO THE DEPENDENT LIST. SELECT THE FIRST FIXED FACTORS AND MOVE THEM TO FIXED FACTOR(S) BOX.
4. THEN CLICK MODEL FOLLOWED BY CUSTOM.
5. MOVE BOTH THE FACTORS TO THE BOX CALLED MODEL.
6. CLICKS CONTINUE TO RETURN THE MAIN DIALOGUE BOX.
7. CLICK OK TO GET THE OUTPUT FOR TWO-WAY ANOVA.

iii) FACTORIAL DESIGN

IF TWO OR MORE INDEPENDENT VARIABLES ARE TO BE TESTED THROUGH AN ANOVA, WE USE A FACTORIAL DESIGN. THE TEST IS SIMILAR TO TWO-WAY ANOVA. A TWO WAY ANOVA CAN ESTIMATE THE MAIN EFFECTS OF INDEPENDENT VARIABLE, BUT THE FACTORIAL DESIGN CAN ESTIMATE THE MAIN EFFECTS ALONG WITH THE INTERACTION EFFECT OF THE INDEPENDENT VARIABLES. FACTORIAL DESIGNS ARE DESCRIBED USING “A × B” NOTATION, “A” STANDS FOR THE NUMBER OF LEVELS OF ONE INDEPENDENT VARIABLE AND “B” STANDS FOR THE NUMBER OF LEVELS OF THE SECOND INDEPENDENT VARIABLE. THERE WOULD BE THREE NULL HYPOTHESES FOR TWO MAIN EFFECTS AND THE INTERACTION WITH ‘ALL MEANS EQUAL’.

TYPE OF THE VARIABLE

THE DEPENDENT VARIABLE SHOULD BE IN INTERVAL OR RATIO SCALE AND THE INDEPENDENT VARIABLES SHOULD BE IN NOMINAL SCALE.

CASE ANALYSIS-1

PROBLEM

A STUDY IS CONDUCTED TO FIND THE EFFECTS OF BRAND AND FLAVOUR ON SALES OF ICE-CREAMS. THE PURPOSE IS TO ESTIMATE:

- i) THE EFFECT OF THE INDEPENDENT VARIABLES ‘BRAND’ AND ‘FLAVOUR’ ON THE DEPENDENT VARIABLE ‘SALES’. (MAIN EFFECT)
- ii) THE COMBINED EFFECT OF ‘BRAND’ AND ‘FLAVOUR’ ON SALES. (INTERACTION EFFECT)

THE CODES USED FOR THE INDEPENDENT VARIABLES ARE AS FOLLOWS.

FLAVOUR: VANILLA (CODE =1), CHOCOLATE (CODE=2), STRAWBERRY (CODE=3), BUTTER SCOTCH (CODE=4).

BRAND: VADILAL (CODE =1), AMUL (CODE =2), KWALLITY (CODE =3), MILKO (CODE =4).

THERE ARE $4 \times 4 = 16$ LEVEL OF COMBINATIONS OF FLAVOR AND BRAND. THE DATA OF SALES HAVE BEEN COLLECTED ON THESE 16 COMBINATIONS FROM TWO RETAIL STORES DURING A MONTH PERIOD AS GIVEN BELOW.

TABLE-1: SAMPLE DATA

RETAIL STORE-1				RETAIL STORE-2			
SL. No.	FLAVOUR	BRAND	SALES IN KG	SL. No.	FLAVOUR	BRAND	SALES IN KG.
1	VANILLA	KWALLITY	12	17	VANILLA	KWALLITY	18
2	CHOCOLATE	KWALLITY	13	18	CHOCOLATE	KWALLITY	32
3	STRAWBERRY	KWALLITY	40	19	STRAWBERRY	KWALLITY	16
4	BUTTER SCOTCH	KWALLITY	18	20	BUTTER SCOTCH	KWALLITY	18
5	VANILLA	VADILAL	19	21	VANILLA	VADILAL	11
6	CHOCOLATE	VADILAL	17	22	CHOCOLATE	VADILAL	25
7	STRAWBERRY	VADILAL	16	23	STRAWBERRY	VADILAL	26
8	BUTTER SCOTCH	VADILAL	22	24	BUTTER SCOTCH	VADILAL	28

9	VANILLA	AMUL	35	25	VANILLA	AMUL	32
10	CHOCOLATE	AMUL	14	26	CHOCOLATE	AMUL	32
11	STRAWBERRY	AMUL	32	27	STRAWBERRY	AMUL	12
12	BUTTER SCOTCH	AMUL	19	28	BUTTER SCOTCH	AMUL	18
13	VANILLA	MILKO	12	29	VANILLA	MILKO	17
14	CHOCOLATE	MILKO	25	30	CHOCOLATE	MILKO	19
15	STRAWBERRY	MILKO	28	31	STRAWBERRY	MILKO	22
16	BUTTER SCOTCH	MILKO	29	32	BUTTER SCOTCH	MILKO	33

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS-

H_{01} : THE AVERAGE SALE REMAINS THE SAME FOR THREE TYPES OF FLAVOUR. (FIRST MAIN EFFECT)

H_{02} : THE AVERAGE SALE REMAINS THE SAME FOR ALL LEVELS OF BRAND. (SECOND MAIN EFFECT)

H_{03} : THE AVERAGE SALE REMAINS THE SAME FOR ALL COMBINATIONS OF FLAVOUR AND BRAND. (INTERACTION EFFECT)

ALTERNATIVE HYPOTHESIS-

H_{11} : THE AVERAGE SALE DIFFERS FOR AT LEAST TWO TYPES OF FLAVOUR.

H_{12} : THE AVERAGE SALE DIFFERS FOR AT LEAST TWO ALL LEVELS OF BRAND.

H_{13} : THE AVERAGE SALE DIFFERS FOR AT LEAST TWO COMBINATIONS OF FLAVOUR AND BRAND COMBINATION.

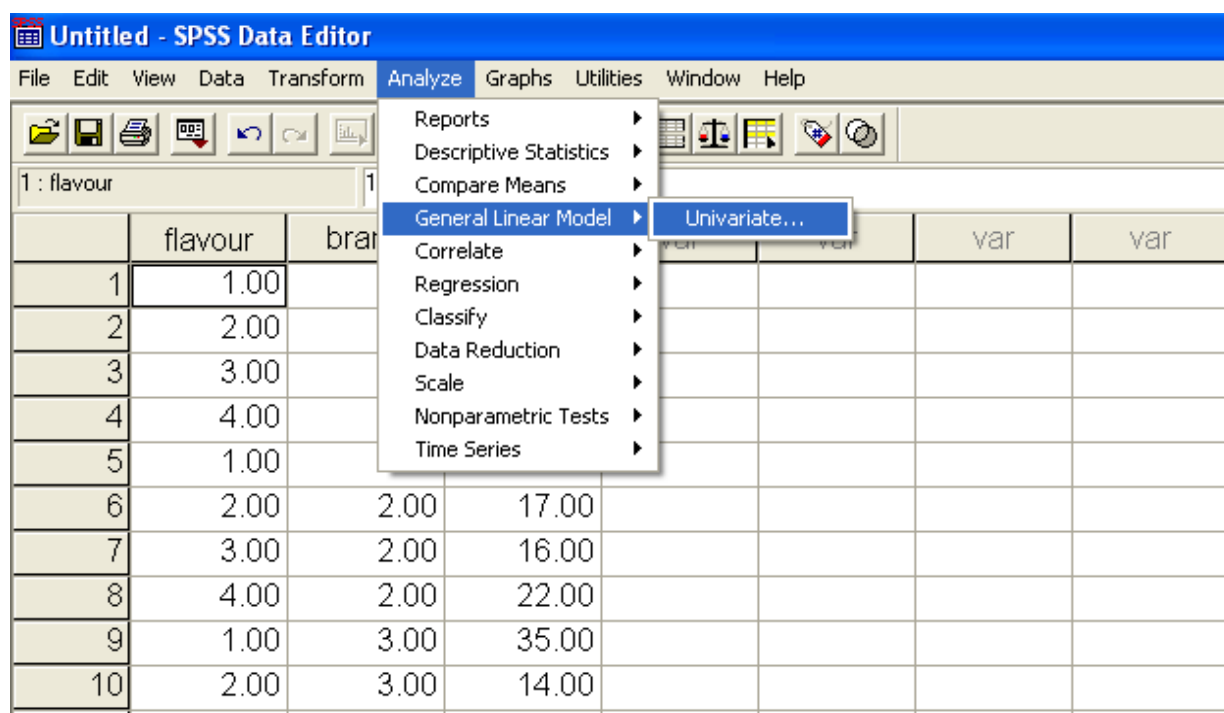
TABLE-2: INPUT DATA

SL. No.	FLAVOUR	BRAND	SALES
1	1	1	12
2	2	1	13
3	3	1	40
4	4	1	18
5	1	2	19
6	2	2	17
7	3	2	16
8	4	2	22
9	1	3	35
10	2	3	14
11	3	3	32
12	4	3	19
13	1	4	12
14	2	4	25
15	3	4	28
16	4	4	29
17	1	1	18

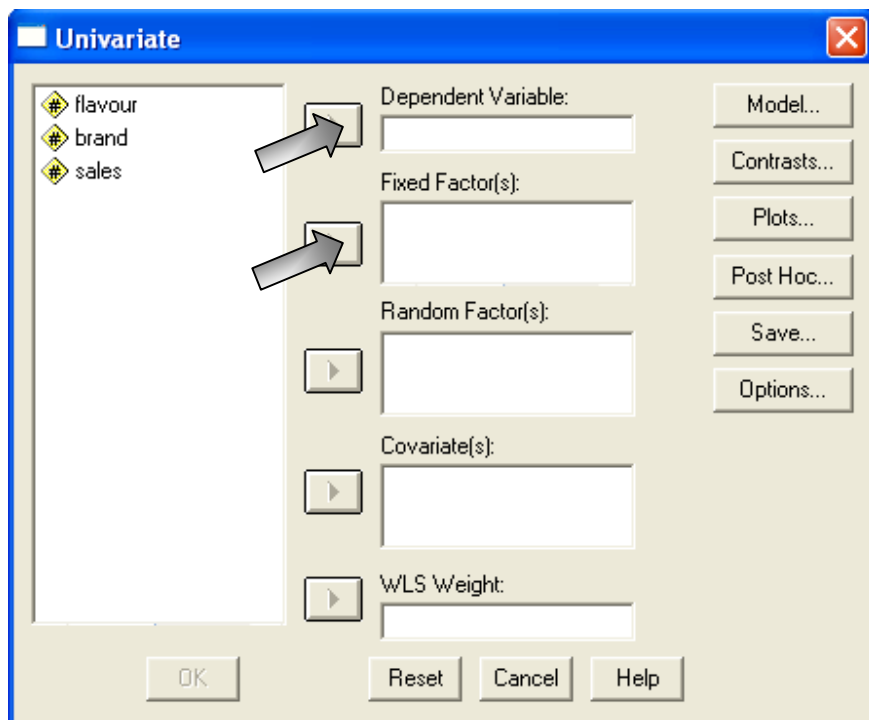
18	2	1	32
19	3	1	16
20	4	1	18
21	1	2	11
22	2	2	25
23	3	2	26
24	4	2	28
25	1	3	32
26	2	3	32
27	3	3	12
28	4	3	18
29	1	4	17
30	2	4	19
31	3	4	22
32	4	4	33

PERFORMING THE ANALYSIS WITH SPSS

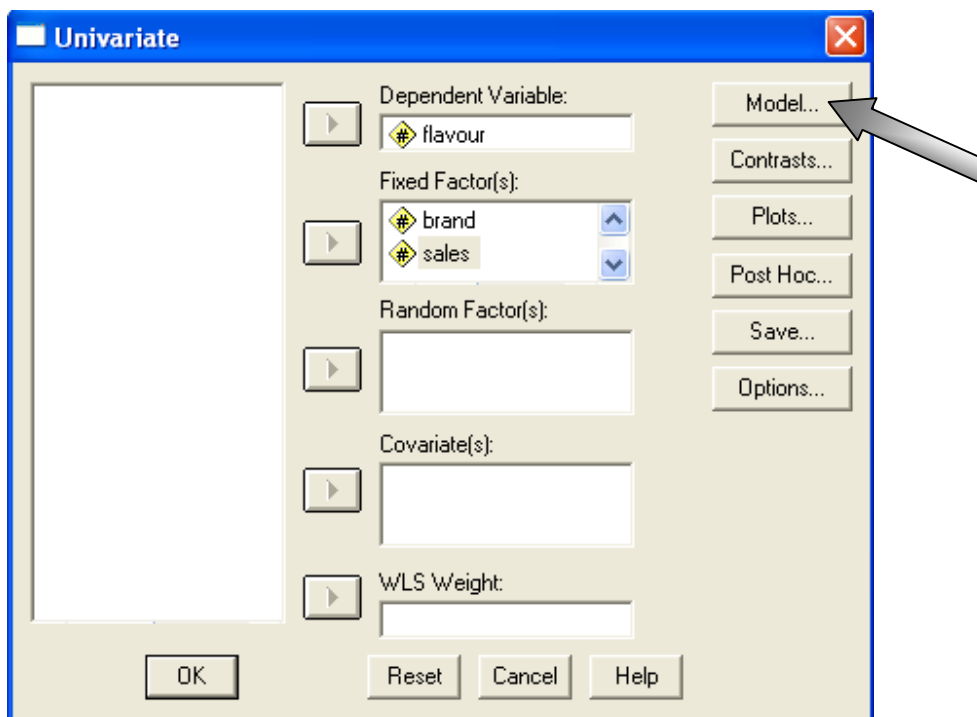
FOR SPSS VERSION 11, CLICK ON **ANALYZE** → **GENERAL LINEAR MODEL** → **UNIVARIATE**. THIS WILL BRING UP THE SPSS SCREEN DIALOGUE BOX AS SHOWN BELOW.



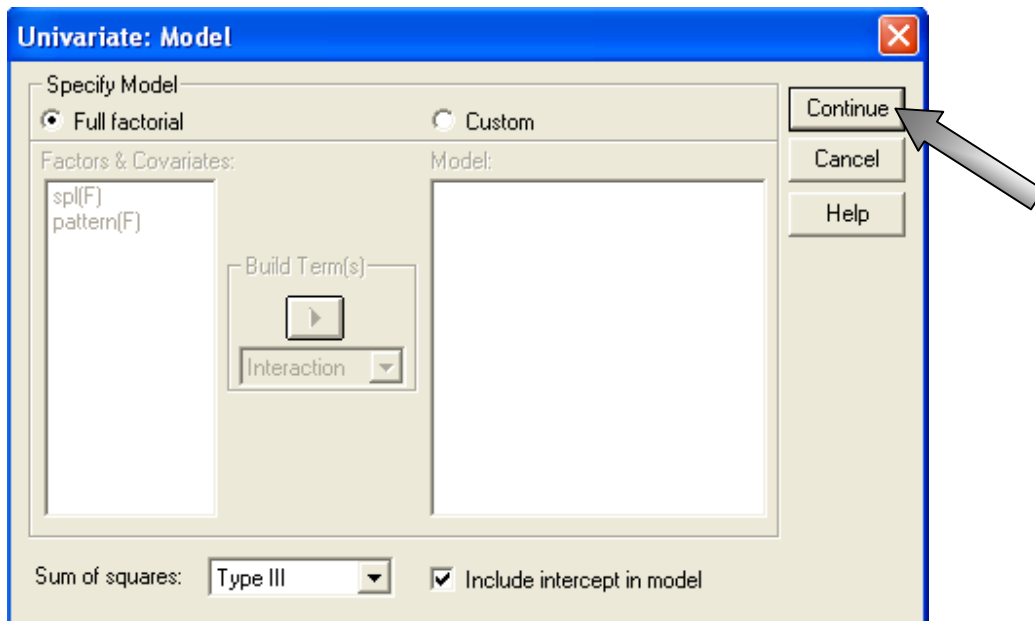
AFTER CLICKING **UNIVARIATE**, THIS WILL BRING UP THE FOLLOWING SPSS SCREEN DIALOGUE BOX.



SELECT THE DEPENDENT VARIABLE “FLAVOUR” AND CLICK IT TO MOVE TO DEPENDENT VARIABLES BOX. SELECT THE INDEPENDENT VARIABLES ‘BRAND’ AND ‘SALE’ AND MOVE THEM TO FIXED FACTOR(S) BOX.



THEN CLICK **MODEL** FOLLOWED BY **FULL FACTORIAL** AND THIS WILL BRING UP THE FOLLOWING DIALOGUE BOX.



THEN CLICK **CONTINUE** TO RETURN THE MAIN DIALOGUE BOX AND CLICK **OK** TO GET THE OUTPUT.

SPSS OUTPUT

THE SPSS OUTPUT OF THE ANALYSIS IS GIVEN IN THE FOLLOWING TABLES.

UNIVARIATE ANALYSIS OF VARIANCE

TABLE-2: BETWEEN-SUBJECTS FACTORS

		N
FLAVOUR	1.00	8
	2.00	8
	3.00	8
	4.00	8
BRAND	1.00	8
	2.00	8
	3.00	8
	4.00	8

TESTS OF BETWEEN-SUBJECTS EFFECTS

TABLE-3: DEPENDENT VARIABLE: SALES

SOURCE	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F	Sig.
CORRECTED MODEL	904.875	15	60.325	.926	.557
INTERCEPT	15753.125	1	15753.125	241.891	.000
FLAVOUR	91.125	3	30.375	.466	.710

BRAND	77.625	3	25.875	.397	.757
FLAVOUR * BRAND	736.125	9	81.792	1.256	.331
ERROR	1042.000	16	65.125		
TOTAL	17700.000	32			
CORRECTED TOTAL	1946.875	31			

A R SQUARED = .465 (ADJUSTED R SQUARED = -.037)

FROM THE OUTPUT, F = 0.466 (FLAVOUR), F = 0.397 (BRAND), F = 1.256 (FLAVOUR AND BRAND)

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) ≤ 0.05

INTERPRETATION

- i) THE P-VALUE FOR FLAVOR IS 0.710 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE), SO WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT FLAVOR HAS NO EFFECT ON SALES.**
- ii) THE P-VALUE FOR FLAVOR IS 0.757 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE) SO WE ACCEPT THE NULL HYPOTHESIS AT 5% LEVEL OF SIGNIFICANCE. IT CAN BE CONCLUDED THAT BRAND HAS NO EFFECT ON SALES.**
- iii) THE P-VALUE FOR FLAVOR AND BRAND IS 0.331 AND IT IS LESS THAN 0.05 (5% LEVEL OF SIGNIFICANCE) SO WE REJECT THE NULL HYPOTHESIS AT 5% LEVEL OF SIGNIFICANCE AND CONCLUDE THAT THE INTERACTION EFFECT OF FLAVOR AND BRAND IS INSIGNIFICANT.**

CASE ANALYSIS-2

PROBLEM

THE FOLLOWING DATA SET REFERS TO THE SALARY PACKAGE (IN LAKHS) OFFERED TO MBA GRADUATES WITH DIFFERENT SPECIALIZATION SELECTED FROM THREE DIFFERENT PATTERN OF STUDIES. THREE STUDENTS HAVE BEEN SELECTED FROM EACH PATTERN OF STUDIES WITH THE SAME SPECIALIZATION AS GIVEN BELOW.

TABLE-1: SAMPLE DATA

SPECIALIZATION	PATTERN OF STUDIES		
	REGULAR COURSE	DISTANCE COURSE	EXECUTIVE MBA
MARKETING	3	1.2	3
	4	3	2
	5	1	1
FINANCE	3	1	2
	4	2	1
	3	1	1
HUMAN	5	1.8	1

RESOURCE	4	3	2
	6	1	2

THE INTEREST IS TO TEST WHETHER THERE IS ANY EFFECT OF SPECIALIZATION (MAIN EFFECT DUE TO FIRST INDEPENDENT VARIABLE) ,THE PATTERN OF STUDIES (MAIN EFFECT DUE TO SECOND INDEPENDENT VARIABLE) AND THE COMBINED EFFECT OF PATTERN OF STUDIES AND THE SPECIALIZATION (INTERACTION EFFECT) ON THE SALARY PACKAGE (DEPENDENT VARIABLE) OR NOT.

CODE:

SPECIALIZATION	CODE	PATTERN OF STUDIES	CODE
MARKETING	1	REGULAR COURSE	1
FINANCE	2	DISTANCE COURSE	2
HUMAN RESOURCE	3	EXECUTIVE MBA	3

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS-

H₀₁: THE AVERAGE SALARY PACKAGE OFFERED REMAINS THE SAME FOR THREE TYPES OF SPECIALIZATION. (FIRST MAIN EFFECT)

H₀₂: THE AVERAGE SALARY PACKAGE OFFERED REMAINS THE SAME FOR THREE PATTERNS OF STUDIES. (SECOND MAIN EFFECT)

H₀₃: THE AVERAGE SALARY PACKAGE OFFERED REMAINS THE SAME FOR SPECIALIZATION AND PATTERN OF STUDIES. (INTERACTION EFFECT)

ALTERNATIVE HYPOTHESIS-

H₁₁: THE AVERAGE SALARY PACKAGE DIFFERS FOR AT LEAST TWO TYPES OF SPECIALIZATION.

H₁₂: THE AVERAGE SALARY PACKAGE DIFFERS FOR AT LEAST TWO LEVELS OF PATTERNS OF STUDIES.

H₁₃: THE AVERAGE SALARY PACKAGE DIFFERS FOR AT LEAST TWO COMBINATIONS OF SPECIALIZATION AND PATTERN OF STUDIES.

TABLE-2: INPUT DATA

SERIAL	SPECIALIZATIO	PATTERN OF	SALARY PACKAGE (IN
1	1	1	3
2	1	1	4
3	1	1	5
4	2	1	3
5	2	1	4
6	2	1	3
7	3	1	5
8	3	1	4
9	3	1	6
10	1	2	1.2
11	1	2	3
12	1	2	1
13	2	2	1
14	2	2	2
15	2	2	1

16	3	2	1.8
17	3	2	3
18	3	2	1
19	1	3	3
20	1	3	2
21	1	3	1
22	2	3	2
23	2	3	1
24	2	3	1
25	3	3	1
26	3	3	2
27	3	3	2

SPSS OUTPUT

THE SPSS OUTPUTS ARE ILLUSTRATED IN BELOW TABLES.

UNIVARIATE ANALYSIS OF VARIANCE

TABLE-2: BETWEEN-SUBJECTS FACTORS

		N
SPECIALIZATION	1.00	9
	2.00	9
	3.00	9
PATTERN OF STUDIES	1.00	9
	2.00	9
	3.00	9

TESTS OF BETWEEN-SUBJECTS EFFECTS

TABLE-3:DEPENDENT VARIABLE: SALARY PACKAGE

SOURCE	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F	SIG.
CORRECTED MODEL	41.301	8	5.163	7.083	.000
INTERCEPT	166.259	1	166.259	228.100	.000
SPECIALIZATION	3.505	2	1.753	2.404	.119
PATTERN OF STUDIES	35.852	2	17.926	24.593	.000
SPECIALIZATION * PATTERN OF STUDIES	1.944	4	.486	.667	.623
ERROR	13.120	18	.729		
TOTAL	220.680	27			
CORRECTED TOTAL	54.421	26			

A R SQUARED = .759 (ADJUSTED R SQUARED = .652)

FROM THE OUTPUT, F = 2.404 (SPECIALIZATION), F = 24.593 (PATTERN OF STUDIES), F = 0.667 (SPECIALIZATION * PATTERN OF STUDIES)

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE (SIG. VALUE) \leq 0.05

INTERPRETATION

- i) THE P-VALUE FOR SPECIALIZATION IS 0.119 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE), SO WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT FLAVOR HAS NO EFFECT ON SALARY PACKAGE.
- ii) THE P-VALUE FOR FLAVOR IS 0.000 AND IT IS LESS THAN 0.05 (5% LEVEL OF SIGNIFICANCE) SO WE REJECT THE NULL HYPOTHESIS AT 5% LEVEL OF SIGNIFICANCE. IT CAN BE CONCLUDED THAT SALARY PACKAGE IS AFFECTED BY THE PATTERN OF STUDIES.
- iii) THE P-VALUE FOR FLAVOR AND BRAND IS 0.623 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE) SO WE ACCEPT THE NULL HYPOTHESIS AT 5% LEVEL OF SIGNIFICANCE AND CONCLUDE THAT THE INTERACTION EFFECT OF SPECIALIZATION AND PATTERN OF STUDIES HAS NO EFFECT ON SALARY PACKAGE.

SPSS COMMAND

1. CLICK ON ANALYZE AT THE SPSS MENU BAR (IN OLDER VERSIONS OF SPSS, CLICK ON STATISTICS INSTEAD OF ANALYZE).
2. CLICK ON GENERAL LINEAR MODEL FOLLOWED BY UNIVARIATE.
3. SELECT THE APPROPRIATE VARIABLE AND MOVE IT TO THE DEPENDENT LIST. SELECT THE FIXED FACTORS AND MOVE THEM TO FIXED FACTOR(S) BOX.
4. THEN CLICK MODEL FOLLOWED BY FULL FACTORIAL.
5. CLICK CONTINUE TO RETURN THE MAIN DIALOGUE BOX.
6. FINALLY CLICK OK TO GET THE OUTPUT FOR FACTORIAL DESIGN.