

CHI-SQUARE TEST



DR. SANGEETA MOHANTY

WHEN OUR DATA CONSISTS OF ONLY THE FREQUENCIES OF VARIOUS EVENTS, THE MOST COMMONLY USED STATISTICS IS THE CHI SQUARE (χ^2). CHI-SQUARE TEST TABULATES A VARIABLE INTO CATEGORIES AND COMPUTES A CHI-SQUARE STATISTIC BASED ON THE DIFFERENCES BETWEEN OBSERVED AND EXPECTED FREQUENCIES.

DR. SANG

CHI-SQUARE TEST

INTRODUCTION

WHEN OUR DATA CONSISTS OF ONLY THE FREQUENCIES OF VARIOUS EVENTS, THE MOST COMMONLY USED STATISTICS IS THE CHI SQUARE (X^2). CHI-SQUARE TEST TABULATES A VARIABLE INTO CATEGORIES AND COMPUTES A CHI-SQUARE STATISTIC BASED ON THE DIFFERENCES BETWEEN OBSERVED AND EXPECTED FREQUENCIES.

APPLICATION OF CHI-SQUARE

- I. TEST FOR GOODNESS OF FIT
- II. CHI-SQUARE TEST FOR CROSS TABS
- III. CHI-SQUARE TEST FOR PROPORTIONS

I) TEST FOR GOODNESS OF FIT

A GOODNESS FIT OF CHI-SQUARE COMPARES THE OBSERVED FREQUENCIES WITH THE EXPECTED OR PREDICTED FREQUENCIES. IT TESTS HOW WELL THE OBSERVED DATA SUPPORTS THE ASSUMPTION ABOUT THE POPULATION.

CASE ANALYSIS-1

PROBLEM

THE MANAGER OF A RETAIL STORE HAS TO TAKE THE DECISION REGARDING THE QUANTITY OF THE DIFFERENT TYPES OF CHOCOLATE BARS TO BE STOCKED TO SATISFY THE CUSTOMERS DEMAND. THE MANAGER CLAIMS THAT THE DEMAND FOR FOUR TYPES OF CHOCOLATE; CADBURY, MUNCH, KIT-KAT, FIVE STAR IS ALMOST THE SAME IN A MONTH PERIOD. TO TEST THE VALIDITY OF THE CLAIM; A RANDOM SAMPLE OF 50 CUSTOMERS WAS SELECTED AND IT PRODUCED THE FOLLOWING DATA.

TABLE-1: SAMPLE DATA

TYPE OF CHOCOLATE BAR	CADBURY	MUNCH	KIT-KAT	FIVE STAR
NUMBER PREFERRING	11	22	9	8

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : THE DEMAND FOR EACH CHOCOLATE IS THE SAME

ALTERNATIVE HYPOTHESIS- H_1 : THE DEMAND VARIES FOR EACH CHOCOLATE TYPE

INPUT DATA

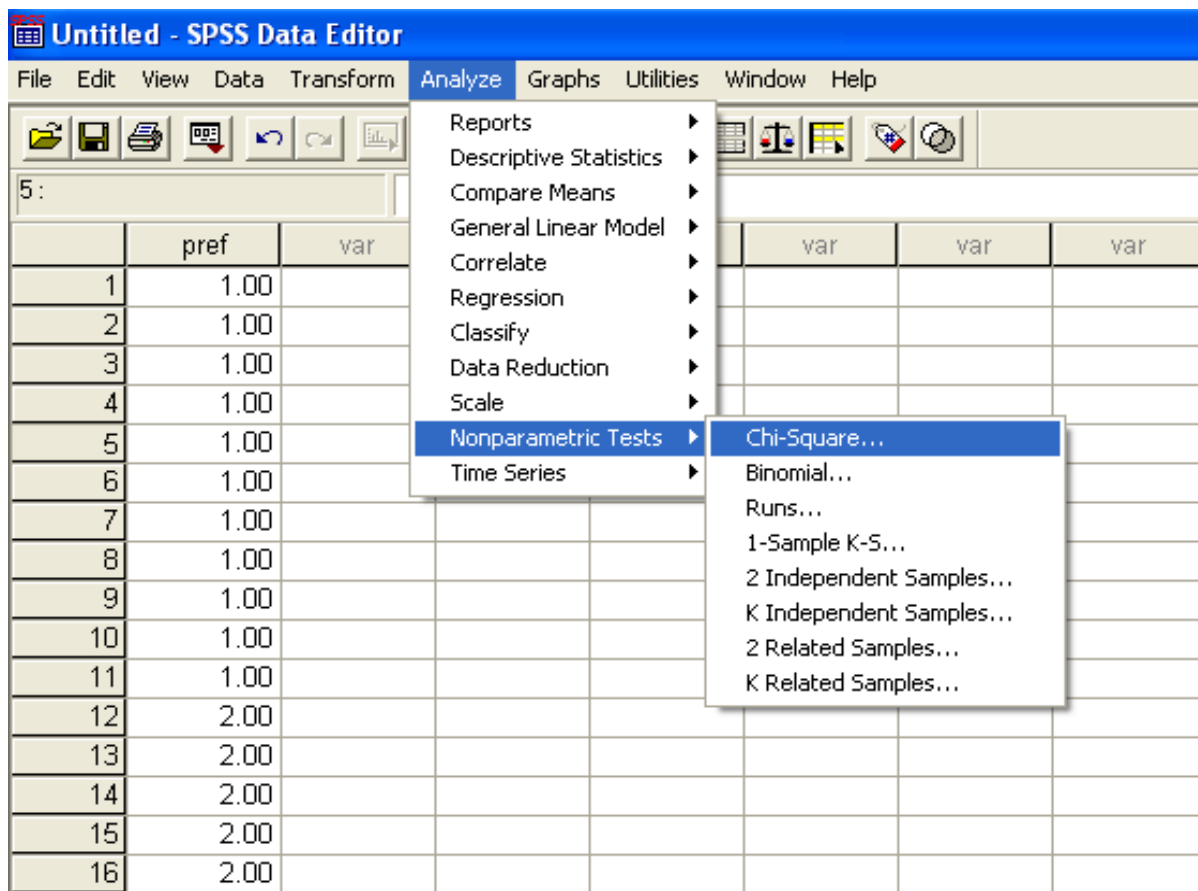
THE DATA ARE ENTERED INTO CODED VARIABLE (CODES: 1 = CADBURY, 2 = MUNCH, 3 = KIT-KAT, 4 = FIVE STAR). SO, THERE ARE 11 TIMES 1; 22 TIMES 2; 9 TIMES 3; 8 TIMES 4 IN THE DATA SHEET AS SHOWN BELOW.

TABLE-2: INPUT DATA

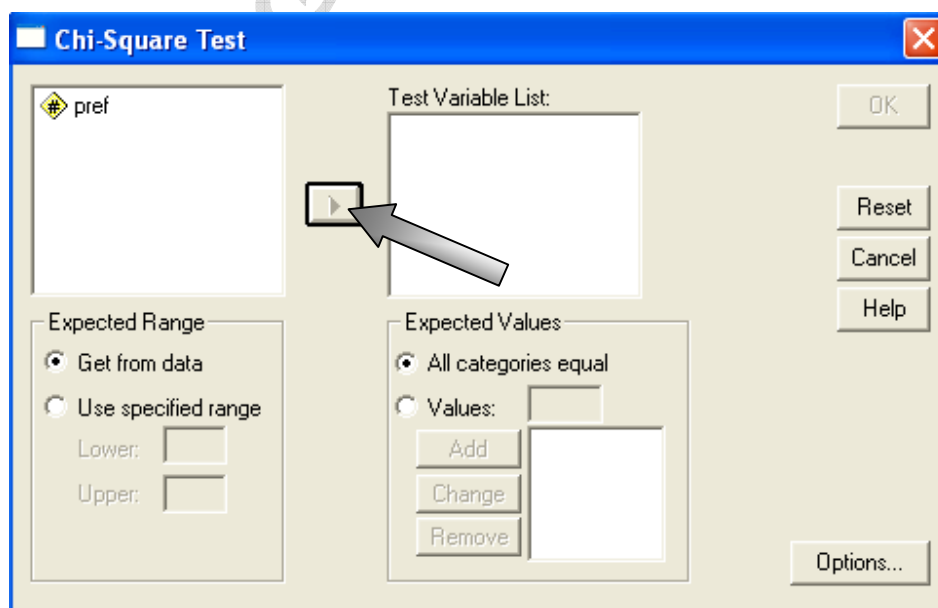
RESPONDENT NUMBER	PREFERENCE OF CHOCOLATE	RESPONDENT NUMBER	PREFERENCE OF CHOCOLATE
1	1	26	2
2	1	27	2
3	1	28	2
4	1	29	2
5	1	30	2
6	1	31	2
7	1	32	2
8	1	33	2
9	1	34	3
10	1	35	3
11	1	36	3
12	2	37	3
13	2	38	3
14	2	39	3
15	2	40	3
16	2	41	3
17	2	42	3
18	2	43	4
19	2	44	4
20	2	45	4
21	2	46	4
22	2	47	4
23	2	48	4
24	2	49	4
25	2	50	4

PERFORMING THE ANALYSIS WITH SPSS

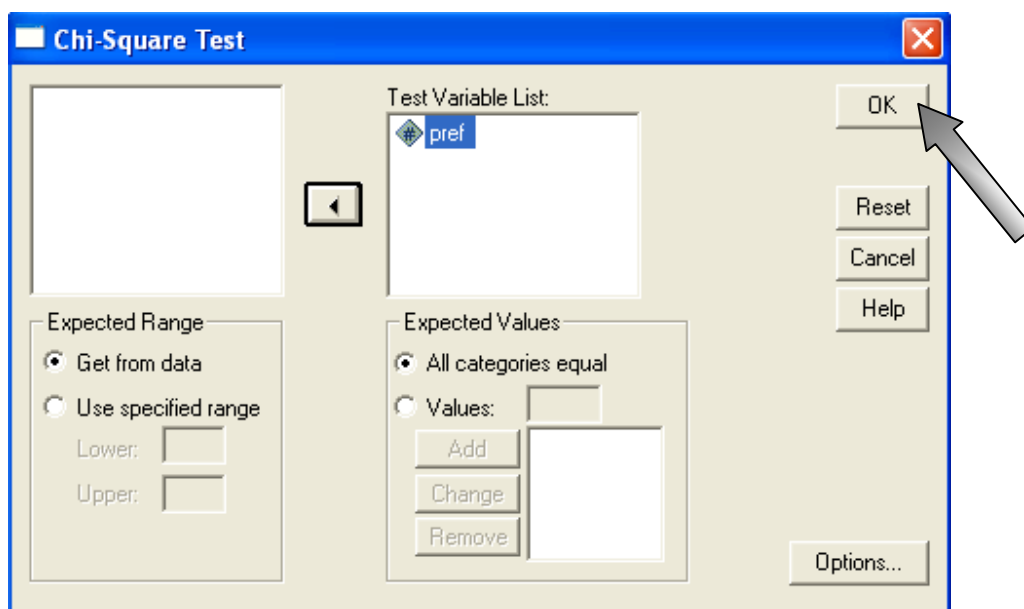
FOR SPSS VERSION 11, CLICK ON ANALYZE → NON PARAMETRIC TESTS → CHI-SQUARE. THIS WILL BRING UP THE SPSS SCREEN DIALOGUE BOX AS SHOWN BELOW-SQUARE.



AFTER CLICKING CHI-SQUARE, THIS WILL BRING UP THE FOLLOWING SPSS DIALOGUE BOX.



SELECT THE VARIABLE AND MOVE IT TO THE TEST VARIABLE LIST BOX AND CLICK OK.



SPSS OUTPUT

THE SPSS OUTPUTS OF THE ANALYSIS ARE DEPICTED IN TABLE-2 AND TABLE-3.

CHI-SQUARE TEST FREQUENCIES

TABLE-3: PREF

	OBSERVED N	EXPECTED N	RESIDUAL
1.00	11	12.5	-1.5
2.00	22	12.5	9.5
3.00	9	12.5	-3.5
4.00	8	12.5	-4.5
TOTAL	50		

TABLE-4: TEST STATISTICS

	PREF
CHI-SQUARE	10.000
DF	3
ASYMP. SIG.	.019

0 CELLS (.0%) HAVE EXPECTED FREQUENCIES LESS THAN 5. THE MINIMUM EXPECTED CELL FREQUENCY IS 12.5.

FROM THE OUTPUT CHI-SQUARE = 10.000

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE \leq 0.05

INTERPRETATION

THE P-VALUE IS 0.019 AND IT IS LESS THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE REJECT THE NULL HYPOTHESIS AND CONCLUDE THAT THE AVERAGE DEMAND VARIES FOR EACH CATEGORY OF CHOCOLATE.

CASE ANALYSIS-2

PROBLEM

A LARGE HOTEL SCHEDULES THE SHIFT DUTIES OF THE STAFFS ASSUMING THAT THE DWELLERS CHECK-OUT THE HOTEL AT A CONSTANT RATE THROUGH THE WEEK PERIOD. BECAUSE OF STAFF SHORTAGES, THE HOTEL ADMINISTRATION WANTS TO OBSERVE WHETHER THE NUMBER OF CHECK-OUTS VARIES IN A WEEK PERIOD OR NOT. THE CHECK-OUT DATA RECORDS WERE COLLECTED FOR A WEEK PERIOD AS SHOWN BELOW.

TABLE-1: SAMPLE DATA

DAYS	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
	Y	Y	DAY	AY	AY	DAY	Y
NUMBER OF CHECK-OUTS	10	9	7	8	15	3	4

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : CHECK-OUT RATE IS UNIFORMLY DISTRIBUTED THROUGHOUT THE WEEK.

ALTERNATIVE HYPOTHESIS- H_1 : CHECK-OUT RATE IS NOT UNIFORM IN A WEEK PERIOD.

CODED VARIABLE

THE CODES USED FOR THE WEEK DAYS ARE (CODES: 1 = MONDAY, 2 = TUESDAY, 3 = WEDNESDAY, 4 = THURSDAY, 5 = FRIDAY, 6 = SATURDAY, 7 = SUNDAY). THE FOLLOWING DATA SHEET IS USED AS INPUT DATA.

TABLE-2: INPUT DATA

SERIAL NUMBER	NUMBER OF CHECK-	SERIAL	NUMBER OF
1	1	29	4
2	1	30	4
3	1	31	4
4	1	32	4
5	1	33	4
6	1	34	4
7	1	35	5

8	1	36	5
9	1	37	5
10	1	38	5
11	2	39	5
12	2	40	5
13	2	41	5
14	2	42	5
15	2	43	5
16	2	44	5
17	2	45	5
18	2	46	5
19	2	47	5
20	3	48	5
21	3	49	5
22	3	50	6
23	3	51	6
24	3	52	6
25	3	53	7
26	3	54	7
27	4	55	7
28	4	56	7

SPSS OUTPUT

**CHI-SQUARE TEST
FREQUENCIES**

TABLE-3: CHECKOUT

	OBSERVED N	EXPECTED N	RESIDUAL
1.00	10	8.0	2.0
2.00	9	8.0	1.0
3.00	7	8.0	-1.0
4.00	8	8.0	.0
5.00	15	8.0	7.0
6.00	3	8.0	-5.0
7.00	4	8.0	-4.0
TOTAL	56		

TABLE-2: TEST STATISTICS

	CHECKOUT
CHI-SQUARE	12.000
DF	6
ASYMP. SIG.	.062

A 0 CELLS (.0%) HAVE EXPECTED FREQUENCIES LESS THAN 5. THE MINIMUM EXPECTED CELL FREQUENCY IS 8.0.

INTERPRETATION

THE P-VALUE IS 0.062 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT THE CHECK-OUT RATE IS UNIFORMLY DISTRIBUTED IN A WEEK PERIOD.

SPSS COMMANDS

1. CLICK ON ANALYZE AT THE SPSS MENU BAR (IN OLDER VERSIONS OF SPSS, CLICK ON STATISTICS INSTEAD OF ANALYZE).
2. CLICK ON NONPARAMETRIC FOLLOWED BY CHI-SQUARE
3. SELECT THE VARIABLES AND MOVE THEM TO TEST VARIABLE LIST BOX
4. NOW CLICK ON THE BUTTON OK.

ii) CHI-SQUARE TEST FOR CROSS TABS

CHI-SQUARE TEST FOR CROSS TABS IS USED TO TEST THE ASSOCIATION OF TWO VARIABLES UNDER STUDY. THE TEST FORMS A BIVARIATE CROSS TABULATION OF THE VARIABLES WITH THE FREQUENCY FOR EACH CELL.

CASE ANALYSIS-1

WE ARE ATTEMPTING TO FIND OUT IF THERE IS ANY SIGNIFICANT RELATIONSHIP BETWEEN THE INCOME AND THE NUMBER OF TIMES OF USING A CREDIT CARD. THE FOLLOWING DATA SET HAS BEEN PREPARED FROM 40 SELECTED PEOPLE OF THE DIFFERENT INCOME GROUPS (< 3 LAKHS, 3-5 LAKHS, 5-7 LAKHS, AND > 7 LAKHS) WHO ARE USING THE CREDIT CARD IN DAILY LIFE.

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : CREDIT CARD USAGE IS INDEPENDENT OF INCOME.

ALTERNATIVE HYPOTHESIS- H_1 : CREDIT CARD USAGE PATTERN IS ASSOCIATED WITH INCOME.

CODED VARIABLE

THE CODES USED FOR INCOME ARE: (CODES: < 3 LAKHS = 1, 3-5 LAKHS= 2, 5-7 LAKHS= 3, > 7 LAKHS = 4) AND THE CODES FOR CREDIT CARD USAGE PATTERN ARE: (CODES: < 2 TIMES = 1, 2-5 TIMES = 3, > 7 TIMES = 3). THE FOLLOWING TABLE IS TREATED AS THE INPUT DATA.

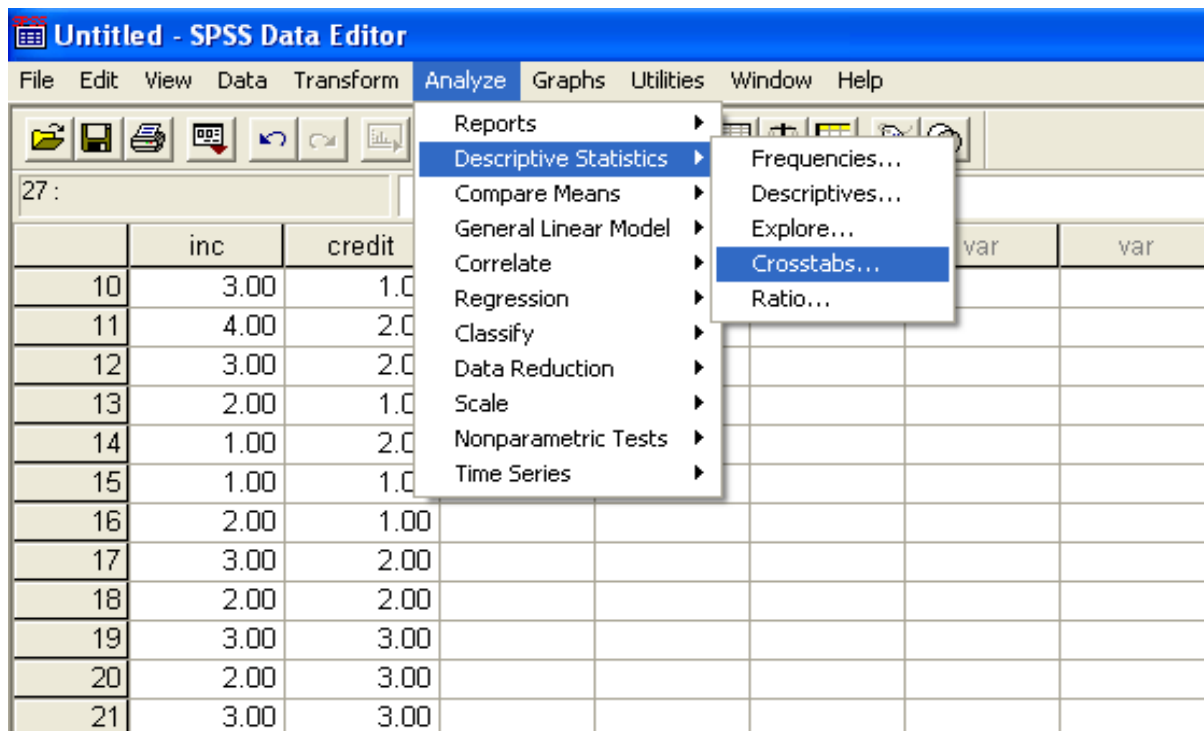
TABLE-1: INPUT OUT

SERIAL NUMBER	INCOME PER ANNUM	CODES FOR INCOME	USE OF CREDIT CARD IN MONTH	CODES FOR CREDIT CARD USAGE
1	< 3 LAKHS	1	< 2 TIMES	1
2	3-5 LAKHS	2	< 2 TIMES	1

3	< 3 LAKHS	1	< 2 TIMES	1
4	3-5 LAKHS	2	< 2 TIMES	1
5	3-5 LAKHS	2	< 2 TIMES	1
6	3-5 LAKHS	2	< 2 TIMES	1
7	3-5 LAKHS	2	2-5 TIMES	2
8	3-5 LAKHS	2	2-5 TIMES	2
9	3-5 LAKHS	2	< 2 TIMES	1
10	5-7 LAKHS	3	< 2 TIMES	1
11	> 7 LAKHS	4	2-5 TIMES	2
12	5-7 LAKHS	3	2-5 TIMES	2
13	3-5 LAKHS	2	< 2 TIMES	1
14	< 3 LAKHS	1	2-5 TIMES	2
15	< 3 LAKHS	1	< 2 TIMES	1
16	3-5 LAKHS	2	< 2 TIMES	1
17	5-7 LAKHS	3	2-5 TIMES	2
18	3-5 LAKHS	2	2-5 TIMES	2
19	5-7 LAKHS	3	> 7 TIMES	3
20	3-5 LAKHS	2	> 7 TIMES	3
21	5-7 LAKHS	3	> 7 TIMES	3
22	5-7 LAKHS	3	> 7 TIMES	3
23	5-7 LAKHS	3	2-5 TIMES	2
24	5-7 LAKHS	3	2-5 TIMES	2
25	3-5 LAKHS	2	2-5 TIMES	2
26	3-5 LAKHS	2	2-5 TIMES	2
27	5-7 LAKHS	3	2-5 TIMES	2
28	3-5 LAKHS	2	< 2 TIMES	1
29	5-7 LAKHS	3	2-5 TIMES	2
30	> 7 LAKHS	4	< 2 TIMES	1
31	3-5 LAKHS	2	2-5 TIMES	2
32	5-7 LAKHS	3	< 2 TIMES	1
33	3-5 LAKHS	2	2-5 TIMES	2
34	5-7 LAKHS	3	2-5 TIMES	2
35	> 7 LAKHS	4	> 7 TIMES	3
36	> 7 LAKHS	4	> 7 TIMES	3
37	> 7 LAKHS	4	> 7 TIMES	3
38	5-7 LAKHS	3	> 7 TIMES	3
39	> 7 LAKHS	4	> 7 TIMES	3
40	> 7 LAKHS	4	< 2 TIMES	1

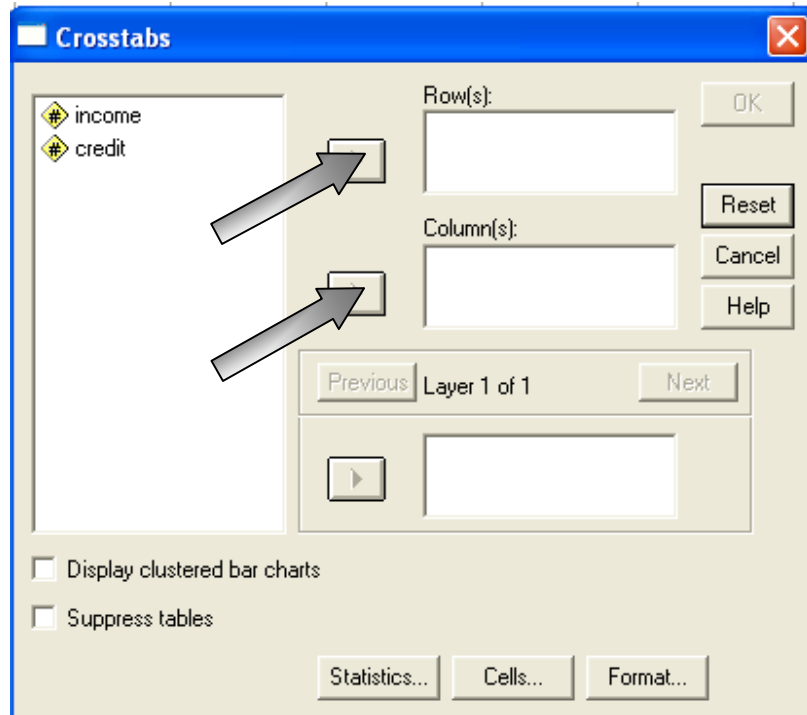
PERFORMING THE ANALYSIS WITH SPSS

FOR SPSS VERSION 11, CLICK ON **ANALYZE** → **DESCRIPTIVE STATISTICS** → **CROSSTABS**. THIS WILL BRING UP THE SPSS SCREEN DIALOGUE BOX AS SHOWN BELOW.

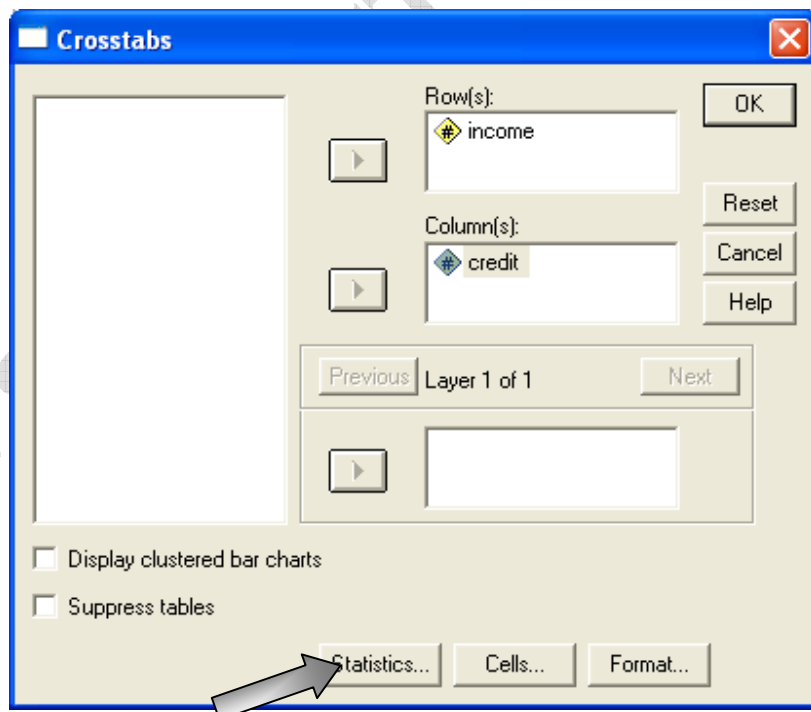


AFTER CLICKING **CROSSTABS**, THIS WILL BRING UP THE FOLLOWING SPSS SCREEN DIALOGUE BOX

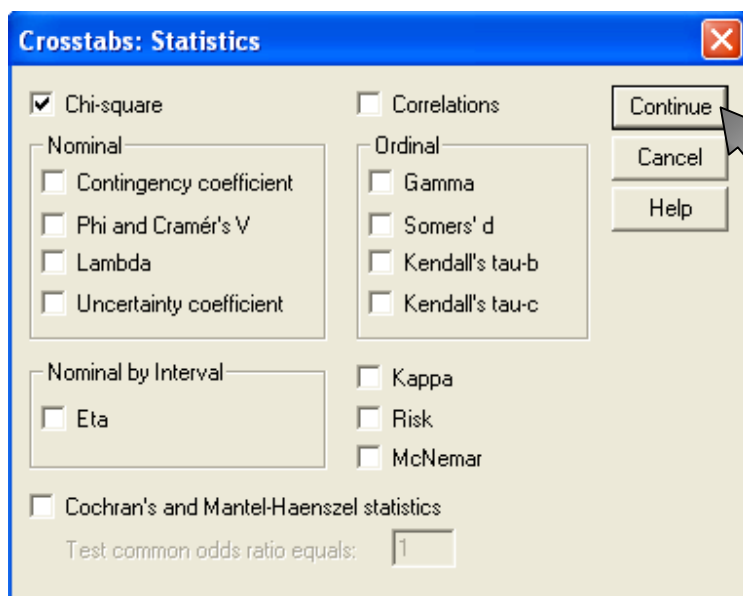
DR. SANGHVI



SELECT THE ROW AND COLUMN VARIABLES AND MOVE THEM TO ROW(S) AND COLUMN (S) BOX.



SELECT STATISTICS, CLICK CHI-SQUARE AND THEN CLICK CONTINUE TO RETURN TO THE CROSS TABS DIALOGUE BOX.



CLICK OK OF CROSS TABS DIALOGUE BOX TO GET THE OUTPUT.

SPSS OUTPUT

THE SPSS OUTPUTS OF THE ANALYSIS ARE DEPICTED IN TABLE-2 TO TABLE-4

TABLE-1: CASE PROCESSING SUMMARY

	CASES		MISSING	TOTAL	
	VALID				
	N	PERCENT	N	PERCENT	PERCENT
INCOME *	40	100.0%	0	.0%	100.0%

TABLE-2: COUNT INCOME * CREDIT CROSS TABULATION

		CREDIT			TOTAL
		1.00	2.00	3.00	
INCOME	1.00	3	1		4
	2.00	8	7	1	16
	3.00	2	7	4	13
	4.00	2	1	4	7
TOTAL		15	16	9	40

TABLE-3: CHI-SQUARE TESTS

	VALUE	DF	ASYMP. SIG. (2-
PEARSON CHI-	12.979	6	.043
LIKELIHOOD RATIO	13.925	6	.030
LINEAR-BY-LINEAR	8.236	1	.004
N OF VALID CASES	40		

A 9 CELLS (75.0%) HAVE EXPECTED COUNT LESS THAN 5. THE MINIMUM EXPECTED COUNT IS .90.

FROM THE OUTPUT CHI-SQUARE = 12.979

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE \leq 0.05

INTERPRETATION

THE P-VALUE IS 0.043 AND IT IS LESS THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE REJECT THE NULL HYPOTHESIS AND CONCLUDE THAT THE CREDIT CARD USAGE PATTERN IS ASSOCIATED WITH THE INCOME.

CASE ANALYSIS-2

A MANAGEMENT EDUCATIONAL INSTITUTE IS INTERESTED TO TEST WHETHER THERE EXISTS ANY RELATIONSHIP BETWEEN THE EDUCATIONAL BACK GROUND AND THE CHANCE OF GETTING JOBS WITH GOOD SALARY PACKAGE AFTER THE DEGREE.

THE HYPOTHESES FOR THE ANALYSIS ARE:

NULL HYPOTHESIS- H_0 : THE CHANCE OF GETTING JOBS WITH GOOD PACKAGE IS INDEPENDENT OF THE EDUCATIONAL BACK GROUND.

ALTERNATIVE HYPOTHESIS- H_1 : THE CHANCE OF GETTING JOBS WITH GOOD PACKAGE IS ASSOCIATED WITH THE EDUCATIONAL BACK GROUND.

CODED VARIABLE

CODES FOR EDUCATIONAL BACKGROUND ARE: COMMERCE = 1, ENGINEERING = 2, ARTS = 3, SCIENCE =4

SALARY PACKAGE CODES ARE: LESS THAN 4 LAKHS PER ANNUM =1, 4-8 LAKHS =2, MORE THAN 8 LAKHS PER ANNUM = 3. THE FOLLOWING TABLE IS USED AS THE INPUT DATA FOR THE ANALYSIS.

TABLE-1: INPUT OUT

STUDENT NUMBER	EDUCATIONAL BACKGROUND	CODES FOR	SALARY PACKAGE	CODES FOR
1	COMMERCE	1	LESS THAN 4 LAKHS	1
2	COMMERCE	1	MORE THAN 8 LAKHS	3
3	COMMERCE	1	4-8 LAKHS	2
4	COMMERCE	1	MORE THAN 8 LAKHS	3
5	COMMERCE	1	MORE THAN 8 LAKHS	3
6	COMMERCE	1	MORE THAN 8 LAKHS	3
7	ENGINEERING	2	MORE THAN 8 LAKHS	3
8	ENGINEERING	2	LESS THAN 4 LAKHS	1
9	ENGINEERING	2	4-8 LAKHS	2
10	ENGINEERING	2	4-8 LAKHS	2

11	ENGINEERING	2	4-8 LAKHS	2
12	ARTS	3	MORE THAN 8 LAKHS	3
13	ARTS	3	LESS THAN 4 LAKHS	1
14	ARTS	3	LESS THAN 4 LAKHS	1
15	ARTS	3	LESS THAN 4 LAKHS	1
16	ARTS	3	MORE THAN 8 LAKHS	3
17	ARTS	3	MORE THAN 8 LAKHS	3
18	ARTS	3	4-8 LAKHS	2
19	ARTS	3	MORE THAN 8 LAKHS	3
20	SCIENCE	4	MORE THAN 8 LAKHS	3
21	SCIENCE	4	MORE THAN 8 LAKHS	3
22	SCIENCE	4	MORE THAN 8 LAKHS	3
23	SCIENCE	4	4-8 LAKHS	2
24	SCIENCE	4	LESS THAN 4 LAKHS	1
25	SCIENCE	4	MORE THAN 8 LAKHS	3

SPSS OUTPUT

TABLE -2: CASE PROCESSING SUMMARY

	CASES		MISSING		TOTAL	
	VALID		N	PERCENT	N	PERCENT
	N	PERCENT	N	PERCENT	N	PERCENT
EDN*SALARY	25	62.5%	15	37.5%	40	100.0%

TABLE -3: EDN*SALARY CROSS TABULATION COUNT

		SALARY			TOTAL
		1.00	2.00	3.00	
EDN	1.00	1	1	4	6
	2.00	1	3	1	5
	3.00	3	1	4	8
	4.00	1	1	4	6
TOTAL		6	6	13	25

TABLE -3: CHI-SQUARE TESTS

	VALUE	DF	ASYMP. SIG. (2-
PEARSON CHI-	5.807	6	.445
LIKELIHOOD RATIO	5.340	6	.501
LINEAR-BY-LINEAR	.000	1	.986
N OF VALID CASES	25		

A 12 CELLS (100.0%) HAVE EXPECTED COUNT LESS THAN 5. THE MINIMUM EXPECTED COUNT IS 1.20.

FROM THE OUTPUT CHI-SQUARE = 5.807

DECISION

REJECT THE NULL HYPOTHESIS IF P-VALUE \leq 0.05

INTERPRETATION

THE P-VALUE IS 0.445 AND IT IS MORE THAN 0.05 (5% LEVEL OF SIGNIFICANCE). THEREFORE WE ACCEPT THE NULL HYPOTHESIS AND CONCLUDE THAT THE CHANCE OF GETTING JOBS WITH GOOD SALARY PACKAGE IS INDEPENDENT OF THE EDUCATIONAL BACK GROUND.

SPSS COMMANDS

1. CLICK ON ANALYZE AT THE SPSS MENU BAR (IN OLDER VERSIONS OF SPSS, CLICK ON STATISTICS INSTEAD OF ANALYZE).
2. CLICK ON DESCRIPTIVE FOLLOWED BY CROSS TABS.
3. SELECT THE ROW AND COLUMN VARIABLES AND MOVE THEM TO ROW(S) AND COLUMNS(S) BOX
4. SELECT STATISTICS, CLICK CHI-SQUARE AND THEN CONTINUE TO RETURN THE MAIN DIALOGUE BOX.
5. NOW CLICK ON THE BUTTON OK.

DR. SANGEETA MOHANTY