DATA ANALYSIS REPORT

FACTORS INFLUENCING POCKET EXPENSES OF COLLEGE STUDENTS

Date: 11th June, 2004

TABLE OF CONTENTS

Executive Summary

- 1. Introduction
- 2. Methodology
- 3. Descriptive Analysis
- 4. Inferential Analysis
- 5. Conclusion
- **6. Excel Functions Used**
- 7. Appendix: The Questionnaire

INTRODUCTION

The area under discussion in the following report is the relationship between the factors affecting pocket expenses of college students. It envelops a range of processes and techniques, which were employed to collect data regarding the above-mentioned theme, as well as a detailed analysis of the same. Suitable diagrams and graphs have been included in the report so as to make it interesting and simple for the reader. Concurrently it puts in numerous tests like f-distribution test, chi square distribution test, test for goodness of fit, z-test for correlation and other miscellaneous tests to clarify the subject matter. In addition it enlightens the Excel functions being used for the analysis as well as a brief summary and conclusion of the whole report.

METHODOLOGY

The key word in statistics is 'data'. Data refers is all the information collected in any form for analysis. Data may be expressly collected for a specific purpose. Such data are known as primary data. The collection of facts and figures relating to the population in the census provides primary data. Often, however data collected for some purpose, frequently for administrative reasons, may be used. Such data are known as secondary data. The following methods are generally adopted for collecting the data.

- Postal questionnaire
- Questionnaires to be filled in by enumerators
- Telephonic interview
- Observation Reports
- Results of experiments

Some other modern techniques used for collecting data are:

- E-mail
- Having questionnaires put up on chat rooms

Collection of Data

We followed a planned strategy for the collection of data. As our topic was concerned with college students we targeted the universities, cafeterias, call centres, cinemas and other rendezvous places such as Priya's, DT malls etc. As we were a group of 3 we divided these places according to locations convenient for each. As it was time for admissions we knew the target would be the Delhi University and there is where we got most of our data. We stood near the admission office and as students came to collect admission forms we asked them to spare a minute explaining them the purpose of the questionnaires. To get a wide array of people we also went to colleges of fashion designing (NIFD), interior decoration, engineering colleges and other colleges of vocational studies.

We also e-mailed the questionnaires to our friends staying or studying abroad in order to see how their expenditure differs from ours. The questionnaires were mailed to U.S.A, U.K and Australia. We converted their currency in rupees to have the data in the same currency in order to make interpretation easier. We also posted the questionnaires on the yahoo chat room notice board for others to fill in n give their opinion. As we kept the name as optional most of them filled in as they knew their identity would be kept secret.

DESCRIPTIVE ANALYSIS

Statistics is a subject, which can be (and is) applied to every aspect of our lives. The aim of statistical methods is simple: to present information in a clear, concise and accurate manner. The difficulty in analysing many phenomena, be they economic, social or otherwise, is that there is simply too much information for the mind to assimilate. The task of descriptive methods is therefore to summarize all this information and draw out the main features, without distorting the picture. Descriptive statistic uses graphs and statistical formulae in order to interpret the information and get desired results.

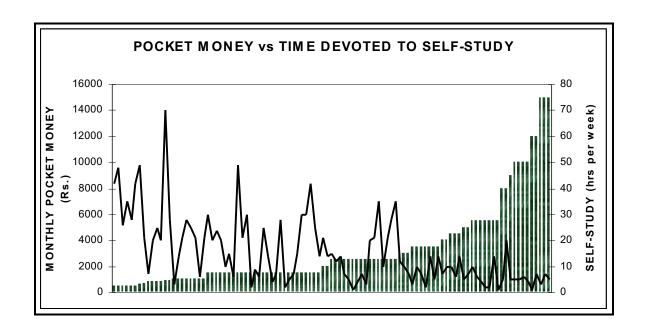
Our research considered various quantitative as well as qualitative factors that influence the pocket expenses of college students. The qualitative data was made quantifiable with the help of dummy variables. Also, factors like the highest level of education of parents were converted into years of education. All data was clearly represented in a spreadsheet format in Microsoft Excel, where the serial number of the entry corresponded to the survey form number on the questionnaire. Here functions such as 'sort in ascending order' were used. With the help of this it was evident that the sample space used was appropriate to comment on the population. [As the monthly pocket money, 'Y' ranged from Rs.500 to Rs.15000.]

This section on descriptive analysis involves the inferences drawn out when different variables are linked either individually with the 'Y' factor or with each other with the help of histograms, pie charts, and other data analysis tools.

The observations were as follows:

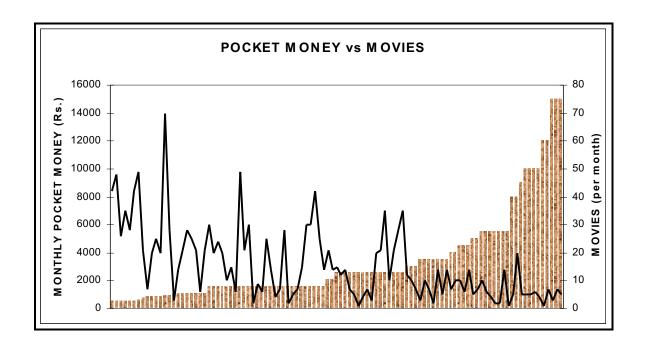
1. Relation of time devoted to self-study with the monthly pocket money:

An inverse relation is evident, when plotting monthly pocket money against the time devoted to self-study (hrs per week). It was observed that on an average, college students receiving pocket money around Rs.500 per month utilize about 30-50 hrs per week for self-study. On the other hand, as the pocket money rises, study time falls to as low as 5hrs per week. This analysis gives us an insight into student psychology. It is certain that students get diverted with too much money in hand, and education takes a backseat.



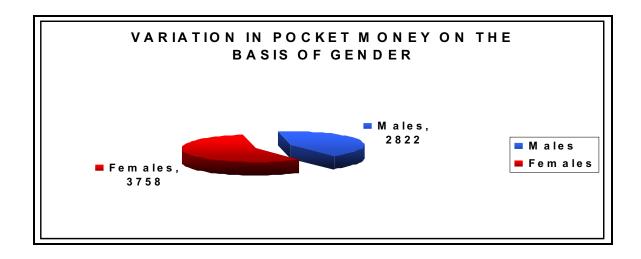
2. Relation between the number of movies seen in a month with the pocket money received:

A clearly incremental graph is obtained when the number of movies seen (per month) is considered alongside the monthly pocket money. With monthly pocket money of Rs.500 an average student watches about 2 movies per month, while one receiving Rs.15000 can afford to watch more than 10. From this we can infer that movies being a common expenditure of college students, the more money is available, the more movies are seen per month.



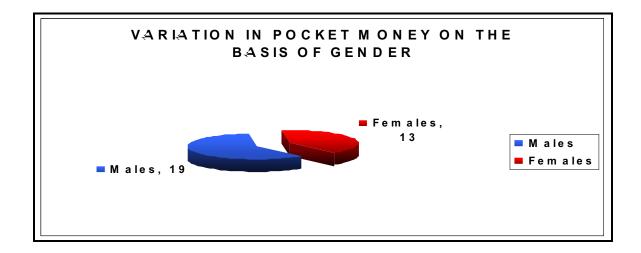
3. Variation of average pocket money on the basis of gender:

52 females and 51 males provided the 103 data entries. This uniformity provides us the opportunity to obtain a value for their respective average monthly pocket money. It is noticeable that an average female receives a higher pocket money than an average male. While the average pocket money of males is close to Rs.3000, that of an average female is more than Rs.3500.



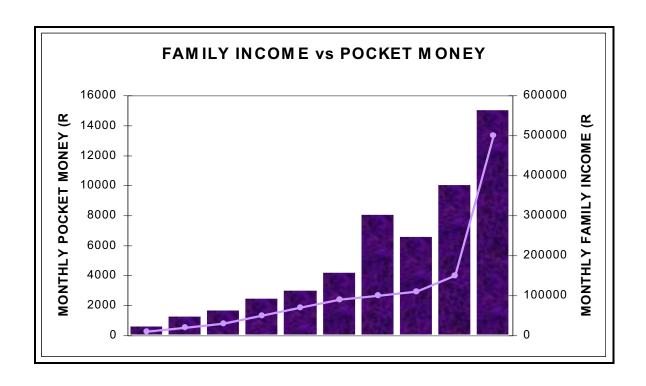
4. Variation in time devoted to self-study on the basis of gender:

It is a realised trend that females usually devote more time to self-study than males. However, the results of our analysis do not conform to it. As per our analysis males append about 19hrs per week on self-study, while females approximately 13hrs. As per our analysis, it was observed that females receive a higher pocket money than males and also that with an increase in pocket money the time devoted to self-study decreases. This could be a possible cause for the obtained result and deviation from the trend.



5. Relation of Monthly Family Income with Monthly Pocket Money:

Distribution is always with respect to a whole. Similarly, the apparent fact of pocket money given to children being dependent on the income of the family is verified by the graph below. It is clear from the graph that pocket money follows a direct relation with the family income.



INFERENCIAL ANALYSIS

Inferential statistics basically relates the sample characteristics to population characteristics. We take random samples from the populations because we cannot take the entire populations for the purpose of our study. Samples are simply the means to an end and not an end in themselves. Statistical inference is concerned with making inferences about the population parameters. Typical population parameters are the mean, standard deviation, the area under the curve between two values of a variable etc.

Inferences about Regression as a whole: Overall Significance

Using F-test

The F test is used to determine whether there is a significant relationship between the dependent variable and the set of all the independent variables. It helps us check whether the value of R-square really indicated that the independent variables explains Y or has happened just by chance.

"Is the regression as a whole significant?"

We see how all the X_i's taken together significantly explain the variability explained in Y. Our hypotheses is

 H_0 : $B_1 = B_2 = B_3 \dots = B_k$ Null hypotheses: Y does not depend on X_i 's.

 $H_i \!\!: At \ least \ one \ B_i \!\!\neq\!\! 0$ Alternate Hypotheses: Y depend upon at least one X_i 's

In discussing the Variation in Y we look at three different terms, we denote these by:

Three different Sums of Squares,

SST = Total Sum of Squares (i.e. explained part) = $\sum (Y - \overline{y})^2$

SSR = Regression Sum of Squares (i.e. the unemployed part) $\sum (Y-\overline{y})^2$

SSE = Error Sum of Squares (i.e. the unemployed part) $\sum (Y - \nabla)^2$

These are related by Equation

"Decomposing the total Variation in Y"

SST = SSR + SSE

Which says that the total variation in Y can be broken down into two parts:

Explained Part & Unexplained part.

Each of these squares has associated (n-1) degrees of freedom (n observation 1) degrees of freedom because sample mean is fixed. SSR has k Degrees of Freedom because there are 'k' independent variables being used to explain Y.

Finally, SSE has (n-K-1) degrees of freedom because we used our 'n' observation to estimate

(K+1) constants, a, b1, b2.....b_k.

If the null hypothesis is true the ratio below is equal to.

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

From the summary output: We see that SSR = 727000000 (with K=7 degrees of freedom)

SSE = 378000000(with n-k-1=102-7-1= 95 degrees of freedom)

So,
$$F = \frac{727000000/7}{3780000000/95}$$

= 26.14

the entries in the "MS" column are just the sums of squares divided by degrees of freedom for 7 numerator degrees of freedom & 95-denominator degree of freedom. The table tells us that 2.12 is the upper limit of acceptance region. For a significance level of

$$\alpha = 0.05$$
.

Our calculated F value of 26.14 is far above 2.13 so we see that regression as a whole is highly significant. We can reach the same conclusion by noting that the output tells us that "significance F" = 1.31e-19

Inferences about an Individual Factor: Individual Significance

The z-test

The T Test is used to determine whether each of the individual independent variables is significant. A separate t test is conducted for each of the independent variables. The purpose of the t-test is to see whether we can use the sample data to conclude that

The regression plane is derived from a sample and not from the entire population. As a result, we cannot expect the true regression equation:

$$Y = A + B_1X_1 + B_2X_2 + \dots + B_kX_k$$

(The one for the entire population)

To be exactly the same, as the equation estimated from the sample observations,

$$Y = a + b_1X_1 + b_2X_2 + + b_kX_k$$

Even so we can use the value of b_1 , one of the factors, we calculate from the sample, to test hypotheses value of B_1 , one of the factors of the regression plane for the entire population.

We test the following Hypothesis about B₁.

$$H_0: B_1 = 0$$

Null Hypotheses

$$H_1: B_1 \neq 0$$

Alternate Hypotheses

If H_0 is rejected, we will conclude that $B_1 \neq 0$ and that the two variables have a statistically significant relationship. However, if H_0 cannot be rejected [We reject H_0 if $(t < -t_{\alpha/2})$ or if $(t > t_{\alpha/2})$], we would not have sufficient evidence to conclude that there is a relation between the variables.

Age:

Here,
test statistic,
$$t = 2.891083$$

 $t_{0.025} = 1.50$

With 2.891083>1.50 ($t_{0.025}$), we reject H_0 and conclude at the 0.05 significance that B_1 is not equal to zero.

The statistical evidence shows that there is a significant relationship between age and monthly expenses.

Number of working parents:

Here,

test statistic,
$$t = 2.155667$$

 $t_{0.025} = 1.50$

With 2.155667>1.50 ($t_{0.025}$), we reject H_0 and conclude at the 0.05 significance that B_1 is not equal to zero.

The statistical evidence shows that there is significant relationship between Number of working parents with monthly expense of students.

Monthly income:

Here,

test statistic, t = 2.613987 $t_{0.025} = 1.50$

With 2.613987<>.50 ($t_{0.025}$), we reject H_0 and conclude at the 0.05 significance that B_1 is not equal to zero.

The statistical evidence shows that there is significant relationship between family monthly income and monthly expenses of students,

Number of siblings:

Here.

test statistic, t = -0.902742 $t_{0.025} = 1.50$

With -0.902742 < 1.50 ($t_{0.025}$), we accept H_0 and conclude at the 0.05 significance that B_1 is equal to zero.

The statistical evidence shows that there is no significant relationship between number of siblings and monthly expenses of students,

Number of movies seen in a month:

Here,

test statistic, t = 3.142832 $t_{0.025} = 1.50$

With 3.142832>1.50 ($t_{0.025}$), we reject H_0 and conclude at the 0.05 significance that B_1 is not equal to zero.

The statistical evidence shows that there is significant relationship between Number of movies seen in a month and monthly expenses of students.

Visits to restaurants per month:

Here.

test statistic, t = 5.421227

$$t_{0.025} = 3.96$$

With 5.421227>1.50 ($t_{0.025}$), we reject H_0 and conclude at the 0.05 significance that B_1 is not equal to zero.

The statistical evidence shows that there is significant relationship between Number of restaurant visits per month and monthly expenses of students.

Time devoted to self study:

Here.

test statistics, t = -1.747169 $t_{0.025} = 3.96$

With -1.747169 < 1.50 (t_{0.025}), we accept H₀ and conclude at the 0.05 significance that B₁ is equal to zero.

The statistical evidence shows that there is no significant relationship between time devoted to self study and monthly expenses of students.

Also, we use the P-value to test whether the factors are significant explanatory variables or not. The entries in the column are probability values for the two-tailed test of the hypotheses.

 $H_0: B_3 = 0$

 $H_1: B_3 \neq 0$

We only compare the probability value with α the significant level of the test, to determine where QF is a significant variable of Y.

Testing the significance of an explanatory variable is always a two-tailed test. The independent variable X_i is a significant explanatory variable if " b_i " is significantly different from zero, that is, if to is a large positive or a large negative number.

Testing Hypothesis for independence of two categories though χ^2

CHI square test

CHI square is used as goodness of fit test, it is most of the used as a test of independence to determine if the paused observations obtained on two or more nominal variables are independent of each other or not. It is some times necessary to deal with the idea of two

variables being related in the sense that the value of one variable depends upon the value of other corresponding variable.

Then the null hypothesis for independence can be tested by: -

$$\chi^2 = \frac{\sum (f_o - f_e)^2}{f_e}$$

Where:

 F_o = Observed frequency

F_e = Expected frequency = row total*column total/100

 \sum = summation overall cells

The degrees of freedom in a contingency table are given below;

$$df = (r-1)(k-1)$$

Where:

R = number of rows

K = number of columns

Relationship Between Family Income and Fathers Education.

 $H_0 = 0$: Null Hypothesis: no relationship between income and education.

 H_1 = Alternate Hypothesis: relationship between income and education.

income	1	2	Total
10000-30000	25	6	31
30001-50000	22	6	28
50001-70000	10	4	14
70001-90000	6	6	12
90001-110000	8	8	16
>110000	2	1	3
Total	73	31	104

fo	f _e	f _o -f _e	$x = f_0 - f_e $ - 0.5	x ²	x²/fe
25	22.63	2.37	1.87	3.5	0.1546
6	9.61	-3.61	3.11	9.6721	1.0064
22	20.44	1.56	1.06	1.1236	0.0549

6	8.68	-2.68	2.18	4.7524	0.5475
10	10.22	-0.22	0.28	0.0784	0.0076
4	4.34	-0.34	0.16	0.0256	0.0059
6	8.76	-2.76	2.26	5.1076	0.583
6	3.72	2.28	1.78	3.1684	0.8517
				10.112	
8	11.68	-3.68	3.18	4	0.8658
				12.531	
8	4.96	-4.04	3.54	6	2.5265
2	2.19	19	0.31	0.0961	0.0438
1	0.93	.07	.43	0.1849	0.1988
					6.8465

Solving the problem we get the value of χ^2 as 6.8465

Looking at the value of χ^2 from the table for 95% confidence (or α = 0.5) & at (6-1)(2-1) = 5 degrees of freedom, we get:

$$\chi^2 = 11.0705$$

Since our calculated value of χ^2 is less than the critical value of χ^2 i.e. 11.0705; we accept the null hypothesis and reject Alternate hypothesis, concluding that there is no relationship between family income and number of working members in the family.

Relationship between amount of time devoted to self study and number of movies seen in a month

 H_0 = 0: Null Hypothesis: no relationship between time spent on self study and number of movies seen in a month.

 H_1 = Alternate Hypothesis: relationship between time spent on self study and number of movies seen in a month

				7 or	
Study Movies	1-2	3-4	5-6	more	Total
<15	27	13	13	12	65
15-25	10	4	4	2	19
25-35	2	2	4	5	13
35-45	0	1	2	0	3
>45	3	1	0	0	4
Total	42	21	22	19	104

f _o	f _e	f _o -f _e	$x = f_0 - f_e - 0.5$	X ²	x² / fe
				0.0	,
27	27.3	-0.3	0.2	4	0.0014
				0.0	
13	13.65	-0.65	0.15	25	0.0019
				0.6	
13	14.3	-1.3	0.8	4	0.0447
				0.0	
12	12.35	-0.35	0.15	25	0.002
				2.3	
40	7.00	0.00	4.50	10	0.0005
10	7.98	2.02	1.52	4	0.2895
				0.2	
4	2.00	0.01	0.40	40	0.0602
4	3.99	0.01	0.49	0.1	0.0602
				0.1	
4	4.18	-0.18	0.32	4	0.0245
	4.10	0.10	0.02	1.2	0.0240
				32	
2	3.61	-1.61	1.11	1	0.3413
				3.9	
				60	
2	5.46	-3.46	1.99	1	0.7253
				0.0	
				52	
2	2.73	-0.73	0.23	9	0.0194
				0.4	
_				09	
4	2.86	1.14	0.64	6	0.1432
				4.1	
_	0.47	0.50	0.00	20	4 000 4
5	2.47	2.53	2.03	9	1.6684
				0.5	
0	1.26	-1.26	0.76	77 6	0.0458
	7.20	-1.20	0.76	0.0	0.0438
				16	
1	0.63	0.37	0.13	9	0.0268
-	3.30	J.J.	0.20	0.7	0.0200
				05	
2	0.66	1.34	0.84	6	1.0691
				0.0	
				04	
0	0.57	-0.57	0.07	9	0.0086

				0.6 72	
3	1.68	1.32	0.82	4	0.4002
1	0.84	0.16	0.34	0.11 56	0.1376
0	0.88	-0.12	0.38	0.14 44	0.1641
0	0.76	-0.76	0.26	0.06 76	0.0089
TOTAL					5.1829

Solving the problem we get the value of χ^2 as 5.1829

Looking at the value of χ^2 from the table for 95% confidence (or α = 0.5) & at (5-1)(4-1) =12 degrees of freedom, we get:

$$\chi^2 = 21.0261$$

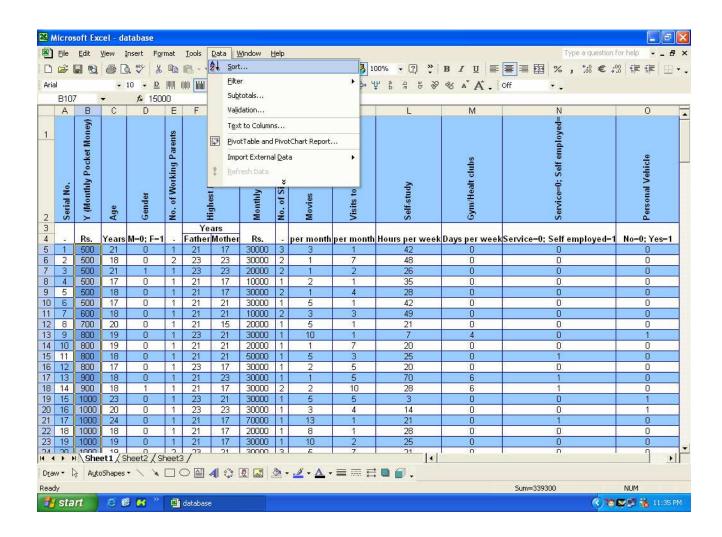
Since our calculated value of χ^2 is less than the critical value of χ^2 i.e. 21.0261; we accept the null hypothesis and reject Alternate hypothesis, concluding that there is no relationship between number of movies seen in a month and time spent on self study.

EXCEL FUNCTIONS USED

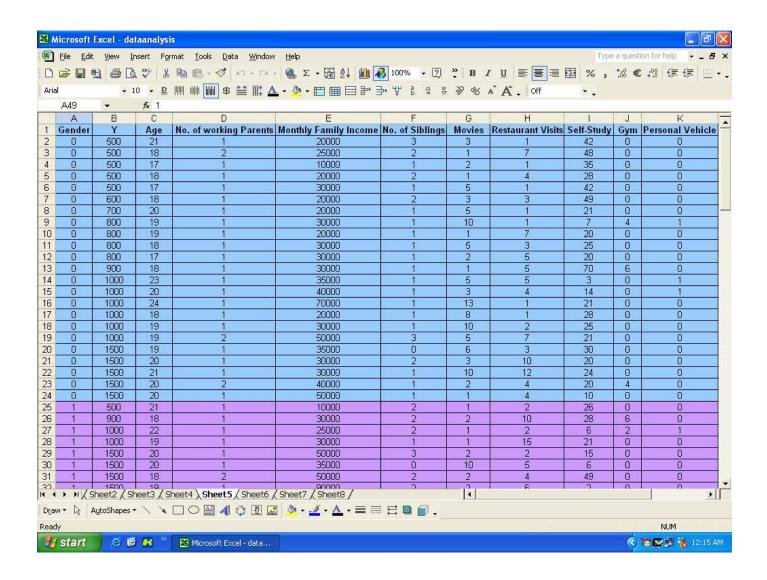
Alongside the essential functions of regression, t-test, f-test, chi square test, mean, certain other functions were also used to make interpretations and calculations easier. An example of this is the sort function.

Use of 'Sort' Function:

The Y factor, namely the monthly pocket money was arranged in ascending order. This Excel function provided the convenience of analysing whether the sample space had been appropriately selected, so as to represent the population. The data collected was significant as it included a broad range that varied between Rs. 500 to Rs. 15000 per month. Also was identifiable a gap in the incremental trend of the data collected. The data does not include any entries with pocket money ranging from Rs. 5500 to Rs. 8000. A method to rectify this is by entering all useful data simultaneously as it is obtained. However, in this research the impact of this drawback is reduced as the number of questionnaires considered was many. i.e. 103



This function was also used to segregate the data on a gender basis. Dummy variables were used to depict this variable. Females were represented by 1, and the males with a 0 as for binary coding. These values were then sorted to obtain a data segregated on the basis of gender. This made further analysis on this basis easier.



CONCLUSION

From a sample space of around 100 students we made inferences about the various factors influencing the pocket expenses of students. Subsequent to the analysis, it was notices that certain factors have a direct relation with monthly pocket money while others demonstrate an inverse relation.

Variables with direct relation:

Age:

As this questionnaire was targeted at the college students the age group selected was 17 – 24. It was observed that in this age group your pocket expenses increase with age as a rise was observed in the expenditure on education and entertainment.

Working parents:

It was observed that both parents working, leading to greater family income had a direct impact on the pocket money of students. Also these students were observed to be more outgoing and spent major part of their pocket expense on entertainment.

• Entertainment:

It is however obvious that greater the pocket money greater would be the money spent on entertainment. As in the questionnaire we divided entertainment in various sub heads such as movies, eating out, money spent on gym and health club etc, we observed the major portion of money spent was entertainment.

Variables with Indirect Relation:

Number of siblings:

From the observed data we concluded that more the number of siblings lesser is the pocket many i.e the number of siblings and the monthly pocket money of students has an inverse relation. It is thus seen the more the number of siblings, the money gets distributed hence lesser money for each student.

Time devoted to self-study:

Our analysis showed that students who spend more time on self study spend less on entertainment and hence have quite reasonable monthly expenses. Thus an inverse relation is seen between both variables.

APPENDIX

The Questionnaire:

SURVEY FORM NO.:

FACTORS INFLUENCING POCKET EXPENSES OF COLLEGE STUDENTS

All the information provide be kept confidential.		•		-		
Name:						
Age:yrs						
Gender: M	F					
Contact (email/phone no.)						
1. Who is/are the working		•				
2. What is the highest leve	l of education of	your pare	nts?			
Father Secondary School		ligh chool		Graduate	□ Postgr	aduate
Mother □ Secondary : School		ligh chool		Graduate	□ Postgr	aduate
3. What is the main source	e of your family i	ncome?				
□ Service □ Se	lf- Employed Ea	rnings				
4. What is your family's a	verage monthly i	ncome? (F	rom all	sources)		
\Box < Rs.20,000		s.20,000 s.40,000	-	- □ Rs	40,000 – Rs.6	60,000

Rs.80,000		Rs.80,000 Rs.1,00,000		> Rs.1,00,000
5. How many siblings d	lo you have?			
□ 0	□ 1	□ 2	□ 3	□ 4
6. What is your averag	e monthly pock	et money?		
 □ Rs.500 - Rs.1 □ Rs.3,000 - Rs □ If other, plea 	,000 □ s.4,000 □ se specify	Rs.1,000 – Rs.2,000 Rs.4,000 – Rs.5,000		Rs.2,000 – Rs.3,000 Rs.5,000 – Rs.6000
7. Do you own a person	nal vehicle?			
□ Yes □	No			
8. How many movies d	o you watch in a	n month? (per n	nonth)	
9. How many times do month)	you visit to a re	staurant/eating joints	in a mont	th? (visits per
10. Do you visit a fitnes	ss centre (gym/h	ealth club)?		
□ Yes □	No			
If yes, how many da	ys in a week	(days per week)		
11. How much time do	you devote to se	elf study? (hour	rs per wee	k)
			Sig	gnature:
			Da	nte:
		-THANKYOU		