

Application of Analysis of variance and Chi- square to study diamond industry

Dr. Neelam Arora¹, Dr. Vinay Pandit²
Lala Lajpat Rai College

I. Introduction

1.1 Chi –Square

Chi -square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. For example, if, according to Mendel's laws, you expected 10 of 20 offspring from a cross to be male and the actual observed number was 8 males, then you might want to know about the "goodness to fit" between the observed and expected. Were the deviations (differences between observed and expected) the result of chance, or were they due to other factors. How much deviation can occur before you, the investigator, must conclude that something other than chance is at work, causing the observed to differ from the expected. The chi-square test is always testing what scientists call the **null hypothesis**, which states that there is no significant difference between the expected and observed result.

The formula for calculating chi-square is $\chi^2 = \sum \frac{(O-E)^2}{E}$

This value is then compared with the table value and tested at 1% or 5% LOS. If calculated value is greater than table value we reject the null hypothesis.

1.2 Analysis of Variance

ANOVA is a statistical method used to test differences between two or more means. It may seem odd that the technique is called "Analysis of Variance" rather than "Analysis of Means." As you will see, the name is appropriate because inferences about means are made by analyzing variance.

ANOVA is used to test general rather than specific differences among means. This can be seen best by example. In the case study "Smiles and Leniency," the effect of different types of smiles on the leniency shown to a person was investigated. Four different types of smiles (neutral, false, felt, miserable) were investigated.

1.3 Diamond Industry

The Indian diamond industry, similar to its origin, is based more in the villages, towns and cities of Gujarat, where most of the processing facilities are installed; the corporate operations of marketing and finance for all the diamond traders takes place from Mumbai, where all the major traders have their registered offices. Majority of the diamantaires procure the rough diamonds from the Diamond Trading Company (DTC, the marketing arm of the De Beers Group, which mines its diamonds in South Africa), which holds the maximum share of rough diamonds in the world. The DTC sells its rough diamonds through two channels: in the primary market to preferred clients called Sight holders, the world's leading diamantaires, carefully chosen for their diamond and marketing expertise; and also form a part of the DTC's Supplier of Choice program; the remainder of the rough diamonds are sold by the DTC in the secondary market worldwide. The other companies, besides DTC, supplying rough diamonds (but to a lesser extent) include Rio Tinto diamonds, Argyle, BHP Biliton and since recently, LevLeviev Diamonds. All the rough diamonds supplied by each of the companies mentioned follow the Kimberley Process Certification as a proof of its purity, identity and place of origin.

II. Research Methodology

2.1 Objective of Study:

- To study relationship between Income and buying behavior of customers with respect to Clarity of diamond.
- To study relationship between Demand of Diamond carat of Diamond.
- To study relationship between Demand of diamond and color of Diamond.
- To Study factors influencing your purchase decision.

2.2 Hypothesis:

The primary objective which is considered by the researcher for which the corresponding hypothesis, considered is as follows.

H₀: Clarity of diamond is independent on Income of customers.

H₁: Clarity of diamond is dependent on Income of customers.

H₀: Demand of Diamond is independent on carat of Diamond.

H₁: Demand of Diamond is dependent on carat of Diamond.

H₀: Demand of Diamond is independent on color of Diamond.

H₁: Demand of Diamond is dependent on color of Diamond.

2.3 Sources of Data

2.3.1 Primary data

Primary data collected by the researcher was through questionnaires. A structured questionnaire was built in correlation with objective of research and hypotheses.

2.3.2 Secondary Data

It is nothing but the backbone of research work. Secondary data is the one which has already been collected and analyzed by someone else. Usually this analyzed data is available in the published form. The articles which were based on the related topic were taken from Newspapers, Magazines, journals and websites which were published.

2.4 Research Design

The research design adopted for the study will be Quantitative Descriptive Cross-sectional design to cover the various facets of the study.

2.5 Sampling Design

A Sampling design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt to select units for the sample. It also indicate the number of units to be included in the sample also known as Sample size. Sampling design is determined before data are collected.

2.6 Type of Population

The researcher had made an attempt to clearly define the population under study. The population here considered by the researcher was Mumbai city.

2.7 Sampling Unit

The sampling unit was identified by the researcher before selection of a sample. Thus for the specific reason the sampling unit selected was south geographical region of Mumbai and the specific individuals belonging to this region.

2.8 Type of Sample

The researcher had deployed Non Probability Sampling method known as **Convenience Sampling**. The customers were identified through the convenient sample which were divided into four clusters namely housewife, students, service class, business class, and professional.

2.9 Size of the Sample

The total sample size decided by researcher was **794** across Mumbai city. All clusters namely housewife, students, service class, business class, and professional were considered for the same.

2.10 Research Area (Scope of Study)

For the intension to complete the research the researcher has made an attempt to collect data which encompasses central and western part of south Mumbai. This scope of study was taken due to logistical problem. All attempts were made to collect the data from different parts of south Mumbai. Also one more reason why this area was taken under consideration was that the area under research was heterogeneous in the sense, all class of people could be easily contacted. Further the scope of study was restricted to only one demographic factor income for the study. Also the scope of study was restricted to only one application of statistics namely chi square and ANOVA.

2.11 Limitations of Research

- 1) The researcher claims that limited time period and limited budget (cost) are indeed the limitations of research.
- 2) There is no control over respondent’s biasness. Thus, even though the researcher has made an attempt to collect authentic information from the respondents, it is observed that Respondent’s biasness while collecting primary information forms the prominent limitation of research.
- 4) However the study was restricted to Mumbai, Only selected region of Mumbai was considered for the research, thus neglecting the views of other respondents in those regions.
- 5) A limited sample size of 794 was considered by the researcher for this study.
- 6) The researcher has defined only one demographic factor which was considered for the research.
- 7) No other applied statistical tools was deployed other than chi square and ANOVA.

III. Data Analysis And Interpretation

To prove the hypotheses an attempt was made by the researcher to use chi square and ANOVA to arrive at the desire conclusion.

H0: Clarity of diamond is independent on Income of customers.

H1: Clarity of diamond is dependent on Income of customers.

Table 3.1

		Are you willing to spent more on diamonds if they are worth of clarity		Total
		Yes	No	
Clarity of diamond means more price	Highly Disagree	14 50.0%	14 50.0%	28 100.0%
	Disagree	43 59.7%	29 40.3%	72 100.0%
	Neither Agree Nor Disagree	95 57.9%	69 42.1%	164 100.0%
	Agree	231 49.3%	238 50.7%	469 100.0%
	Highly Agree	25 27.5%	66 72.5%	91 100.0%
Total		408 49.5%	416 50.5%	824 100.0%

Source: Survey

Chi-Square Tests

Table 3.2

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	25.346 ^a	4	.000
Likelihood Ratio	26.060	4	.000
Linear-by-Linear Association	13.982	1	.000
N of Valid Cases	824		

P value =0.000 < 0.05

Thus the Null hypothesis is rejected

Thus Clarity of diamond is dependent on Income of customers.

H0: Demand of Diamond is independent on carat of Diamond.

H1: Demand of Diamond is dependent on carat of Diamond.

Table 3.3

		Diamond market is inflated by carat		Total
		Yes	No	
Demand of Diamond	Very High	105	87	192
		54.7%	45.3%	100.0%
	High	120	98	218
		55.0%	45.0%	100.0%
	Moderate	64	70	134
		47.8%	52.2%	100.0%
Low	119	161	280	
	42.5%	57.5%	100.0%	
Total		408	416	824
		49.5%	50.5%	100.0%

Source: Survey

Chi-Square Tests
Table 3.4

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.400 ^a	3	.015
Likelihood Ratio	10.429	3	.015
Linear-by-Linear Association	9.317	1	.002
N of Valid Cases	824		

Interpretation:

P value= 0.015 < 0.05

Reject H0

Therefore Demand in Diamond market is influenced by carat.

Table 3.5

		Demand of diamond based on color		Total
		Yes	No	
Diamond and color has an association	Yes	226	205	431
		52.4%	47.6%	100.0%
	No	167	204	371
		45.0%	55.0%	100.0%
	May Be	15	7	22
		68.2%	31.8%	100.0%
Total		408	416	824
		49.5%	50.5%	100.0%

Source: Survey

Chi-Square Tests
Table 3.6

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.545 ^a	2	.023
Likelihood Ratio	7.619	2	.022
Linear-by-Linear Association	1.151	1	.283
N of Valid Cases	824		

Interpretation:

P value= 0.22 < 0.05

Reject H0

Thus Demand of Diamond is dependent on color of Diamond.

Table 3.3

Rank the following factors influencing your purchase decision		Sum of Squares	df	Mean Square	F	Sig.
Brand Name	Between Groups	5.699	4	1.425	.838	.501
	Within Groups	1391.830	819	1.699		
	Total	1397.529	823			
Price	Between Groups	36.565	4	9.141	4.526	.001
	Within Groups	1654.154	819	2.020		
	Total	1690.718	823			
Clarity	Between Groups	24.045	4	6.011	3.352	.010
	Within Groups	1468.934	819	1.794		
	Total	1492.979	823			
Value added services	Between Groups	16.250	4	4.062	2.201	.067
	Within Groups	1511.964	819	1.846		
	Total	1528.214	823			
Reference	Between Groups	13.847	4	3.462	1.835	.120
	Within Groups	1544.705	819	1.886		
	Total	1558.552	823			
Quick and efficient customer care service	Between Groups	10.056	4	2.514	1.457	.213
	Within Groups	1413.163	819	1.725		
	Total	1423.218	823			
Design	Between Groups	14.634	4	3.658	2.375	.05
	Within Groups	1261.691	819	1.541		
	Total	1276.325	823			
carat	Between Groups	61.617	4	15.404	7.863	.000
	Within Groups	1604.382	819	1.959		
	Total	1665.999	823			
Cut/ design	Between Groups	14.500	4	3.625	2.006	.032
	Within Groups	1479.858	819	1.807		
	Total	1494.358	823			
Brand ambassador	Between Groups	13.320	4	3.330	1.559	.183
	Within Groups	1749.593	819	2.136		
	Total	1762.913	823			
Value for Price	Between Groups	6.516	4	1.629	.919	.452
	Within Groups	1451.356	819	1.772		

	Total	1457.873	823			
Promotional Offers	Between Groups	14.529	4	3.632	2.205	.047
	Within Groups	1349.174	819	1.647		
	Total	1363.704	823			
Others please specify	Between Groups	.000	4	.000	.	.
	Within Groups	.000	819	.000		
	Total	.000	823			

P value for Price, Clarity, Design, carat, Cut/ design, Promotional Offers < 0.05
Thus these factors are significant for the purchase decision of the diamonds

IV. Conclusion

An attempt was made by the researcher to study the application of Analysis of variance and Chi- square to study diamond industry in Mumbai. The findings of the research reveal that the demand is strongly related to price of the diamond. Also price and color has a strong association. Here we add depth and context to our examination of the diamond market by taking a detailed look at the dynamics of consumer demand across the globe. Although the jewelry market is the main source of demand for diamonds, the overall industry is much larger: from the producers of rough diamonds to dealers, cutters and polishers to retail sales. In order to understand what drives growth of rough diamonds, one needs to examine factors behind the consumption of diamond jewelry. The research also presents a brief view of the consumer preferences, based on surveys of more than 800 diamond consumers around Mumbai. Also the research concluded that color and demand of diamond also has strong association. This leads to think that the color is one of the parameter which leads to increase or decrease demand of diamond. Also the rigorous attempt was made by the researcher to analyze the factors which are responsible for the buying diamonds. The prominent factors which lead to buying of diamonds were Price, Clarity, Design, carat, Cut/ design, Promotional Offers and other factors
We conclude with an update on the outlook for the diamond industry through 2020. The updated supply forecast is based on the latest developments of key diamond miners and the largest diamond mines worldwide. The 2020 demand outlook is based on our extensive market analysis and consumer research.

REFERENCES

- [1]. Gosall, Narinder Kaur Gosall, Guralp Singh (2012). *Doctor's Guide to Critical Appraisal*. (3. ed.). Knutsford: PasTest. pp. 129–130. ISBN 9781905635818.
- [2]. **Jump up**^ Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". *Philosophical Magazine Series 5* **50** (302): 157–175. doi:10.1080/14786440009463897. edit
- [3]. **Jump up**^ "1.3.6.7.4. Critical Values of the Chi-Square Distribution". Retrieved 14 October 2014.
- [4]. **Jump up**^ "Critical Values of the Chi-Squared Distribution". *NIST/SEMATECH e-Handbook of Statistical Methods*. National Institute of Standards and Technology.
- [5]. **Jump up**^ . See 'Discovering Statistics Using SPSS' by Andy Field for assumptions on Chi Square. ^[*citation needed*]
- [6]. Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification". *The Annals of Mathematical Statistics* **25** (3): 484. doi:10.1214/aoms/1177728717.
- [7]. Caliński, Tadeusz & Kageyama, Sanpei (2000). *Block designs: A Randomization approach, Volume I: Analysis*. Lecture Notes in Statistics **150**. New York: Springer-Verlag. ISBN 0-387-98578-6.
- [8]. Christensen, Ronald (2002). *Plane Answers to Complex Questions: The Theory of Linear Models* (Third ed.). New York: Springer. ISBN 0-387-95361-2.
- [9]. Cox, David R. & Reid, Nancy M. (2000). *The theory of design of experiments*. (Chapman & Hall/CRC). ISBN 978-1-58488-195-7
- [10]. Fisher, Ronald (1918). "Studies in Crop Variation. I. An examination of the yield of dressed grain from Broadbalk". *Journal of Agricultural Science* **11**: 107–135. doi:10.1017/S002185960003750.
- [11]. Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, II. Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification". *The Annals of Mathematical Statistics* **25** (3): 484. doi:10.1214/aoms/1177728717.
- [12]. Caliński, Tadeusz & Kageyama, Sanpei (2000). *Block designs: A Randomization approach, Volume I: Analysis*. Lecture Notes in Statistics **150**. New York: Springer-Verlag. ISBN 0-387-98578-6.
- [13]. Christensen, Ronald (2002). *Plane Answers to Complex Questions: The Theory of Linear Models* (Third ed.). New York: Springer. ISBN 0-387-95361-2.
- [14]. Cox, David R. & Reid, Nancy M. (2000). *The theory of design of experiments*. (Chapman & Hall/CRC). ISBN 978-1-58488-195-7