Analytical Study using Association Rules and Mean of Confidence to Analyze The Poverty Factors in Baghdad Capital of Iraq

Ali Sami*, Dr. Ali Mohammed Sahan*, Ghazi Johnny*

*IT Dept., Technical College of Management, Middle Technical University, Baghdad, Iraq

Abstract:

Poverty is a negative phenomenon that seriously threatened the progress and growth of communities. Studying poverty using scientific research leads to identify the most important influential factors of this phenomenon. In this paper, the poverty in Iraq has been studied through a sample of families in Baghdad city which is the largest and most important city in Iraq. The relational databases as well as the factors related with poverty phenomenon including gender age, the number of family members, educational level, skilled, corruption, disability and disease, wars and disasters have been considered. The association rules and mean of confidence have been used to analyze the main factors for the spread of poverty phenomenon in order to provide clear diagnosis of the most influential factors in the proliferation of this phenomenon, which leads improve appropriate solutions to reduce the impact of these factors.

Keywords:– Association rules, mean of confidence, poverty factors, relational databases.

دراسة تحليلية باستخدام القواعد الارتباطية ومتوسط الموثوقية لتحليل عوامل الفقر في بغداد عاصمة العراق

المستخلص

الفقر هو ظاهرة سلبية تهدد بجدية تقدم ونموالمجتمعات، ان دراسة الفقر باستخدام البحث العلمي يؤدي الى تحديد اهم العوامل المؤثرة في هذه الظاهرة. في هذا البحث تم دراسة الفقر في العراق من خلال دراسة عينة من العوائل في مدينة بغداد كونها اكبر واهم مدن العراق. قواعد البيانات العلائقية بالإضافة الى العوامل المرتبطة بظاهرة الفقر تتضمن الجنس، العمر، عدد أفراد العائلة، المستوى التعليمي، اصحاب المهارات، الفساد، الاعاقة والمرض، الحروب والكوارث. قواعد الارتباط ومتوسط الموثوقية استخدمت في تحليل العوامل الرئيسية لتفشي من اجل توفير تشخيص واضح أكثر العوامل تأثيرا في تفشي هذه الظاهرة مما يقود الى وضع حلول مناسبة لتقليل تأثير هذه العوامل.

الكلمات المفتاحية: العوامل المرتبطة، معدا الثقة، عوامل الفقر، البيانات العلائقية.

1. Introduction

Poverty is a serious, complicated, social, economic, political, historical phenomenon that families and individuals face in different societies, no matter how rich or developed these societies are. So, societies of these countries work hard in order to limit this phenomenon and its effects on family and society level. Contemporary human development has contributed in a way or another in complicating this phenomenon especially in out growth societies whose most people suffer from lack of basic needs (food, residence, health ,education and goods that are needed at home; fridge, oven…etc.) besides other poverty data like; feeling insecure, being afraid of future and lack of self-development opportunities ,which is the destiny of the majority of people in underdeveloped countries ,in which poverty is no longer a new thing .The new thing is realizing this situation and working through every possible mean to reduce or finish it . Because poverty in underdeveloped countries doesn't prove that there are no factors or potential forces which lead to the development and reducing poverty. It is the lack of way and means through which all these factors and forces can become capable of creating tangible growth and development in reality. Poverty in the Arab World is an extremely complicated, economic and social phenomenon as a result of interacting different factors, which don't belong to a certain geographical area, but it is part of rural and urban societies as well. Within the same society, poverty is more common in the countryside more than in a city and among the illiterate more than the literate. Despite a lot of similar things in the Arab World, there are also some exceptions. Oil-producing countries with low population suffer less poverty than non-producing oil countries which depend basically on agriculture. Sudan has the highest poverty level %46.55 then comes Yemen %45.2, Palestine %38, Lebanon %28 and Iraq %19.9 which means that 6 million Iraqi out of 33 million are living below the line of poorness in a country whose annual budget goes beyond 100 billion dollars among an obvious class disparity in incomes as a result of increasing administrative and financial corruption after the year 2003 within the absence of social injustice. Bagdad, the capital of Iraq, in which we intend to examine poverty, its population is 7.665 million represents %21 of the total population in Iraq with % 50.7 male and % 49.3 female. Poverty in Baghdad outruns %13 of its population due to bad security conditions over the last ten years, in addition to the increasing numbers of widows, orphans, people with no family provider and large numbers of unemployed graduates [1,2,3, 4, 5 ,6].Data Mining is an automated exploratory process of useful data among enormous data sources. Data mining techniques are spread to scour hugedatabase in order to find out useful patterns, which could be unknown without it.

Data Mining is a supplementary part of (KDD) knowledge discovery in data base that is the whole process, through which raw data is transformed into useful information [7, 8,9].

2. Association Rules Mining:

Association rule mining, one of the very powerful and well-researched methods of data mining, was first introduced in [Agrawal et al. 1993]. The main goal of it's to find the important and exciting correlations, recurrent model, associations or informal Structures among group of items in the transaction databases or other data repositories. The official interpretation of Apriori Algorithm that is applied to detect the association rules, the support of a group of item I, sup(I), is represented as the number of agreement in the database

holding I. min_sup (minimal support) is a threshold defined by the user. A group of item is frequent if its support isn't less than the min_sup. A group of item with k items in nameda k- group of item.

Let D be a collection of transactions and I = {i1, i2, …,im}, A group of item is a subgroup of I. Given X and Y are group of sets, an association rule is of the form

$X \Rightarrow Y$, where $X \subset I, Y \subset I$, and $X \cap Y = \Phi$,

where the $sup(X \cup Y) > $ min_sup, and the confidence of $X \Rightarrow Y$ is not below a predefined threshold, min_conf, the confidence of $X \Rightarrow Y$ is. $\dfrac{sup(X \cup Y)}{sup(X)}$

After finding out all frequent group of sets, the algorithm of generating association rules uses the subgroups of a frequent group of set as antecedents to create the rules.

The form $X \Rightarrow Y$ with confidence 60% where X = Mobile phone and Y= Mobile applications, for example,most of the customers who buy a Mobile phone also bought the Mobile applications .
And the form of $X \Rightarrow Y$ is not equal to the form $Y \Rightarrow X$ as a result of

unless X=Y which is called perfect rule.

This must be taken into consideration when analyzing the results of association rules [7,8,9].
Also, we used the Mean of the items confidence (MOC) which can be defined as follows:

MOC $(A) = \{conf\,(A \to B) + conf\,(A \to C) + conf\,(A \to D)\}/3.$

The formula gives the mean of the occurrence of item A among the items B, C and D which means the influence of item A on the others [10, 11].

3. Study sample:

Data has been collected incorporation with Ministry of Planning and the Central Statistical Organization; a random sample of poor families has been chosen.
The following factors are signified in order to be analyzed using Association Rule Mining.

1- A represents Gender; female 1, male 0. Because the rates are almost equal and, families provided with partnerless-woman are poorer.

2-B represents Age >= 20years; people, who are 20 or older face more difficulties and suffering due to bad conditions the country has been through.

3- C represents Family members >= 4; the more members a family has, the poorer it is.

4- D represents Educational level; represents a lack of job opportunities for the literate.

5-E represents Posses skills; represents a lack of job opportunities for Posses skill.

6- F represents existence of Corruption; corruption and injustice relating to job opportunities

7- G represents Disability and disease; it refers to the effects on individuals as a result of wars the country has been through.

8-H represents War and disaster; so many families have been impoverished directly or indirectly because of wars and disasters the country has been through.

4. Experimental analysis:
After analyzing collected data, a lot of relations that show the cause of poverty in the capital Bagdad. More important relations were chosen with high confidence. Table 1 presented a part of these relations and show clearly the spread of poverty phenomenon in a country which is considered one of the richest countries in the world.

$$\dfrac{Sup\ (X\ U\ Y)}{Sup(X)}\ NOT\ EQUAL\ \dfrac{Sup\ (Y\ U\ X)}{Sup(Y)}$$

Table (1) some of the relations extracted from given data

Code	Symble	Equ	Relat	Conf
9)	D.H./H.	64/ 82	H-->D	78%
9)	D.H./D.	64/ 80	D-->H	80%
10)	C.H./H.	61/ 82	H-->C	74%
10)	C.H./C.	61/ 76	C-->H	80%
11)	B.H./H.	80/ 82	H-->B	98%
11)	B.H./B.	80/ 96	B-->H	83%
12)	B.F./F.	81/ 83	F-->B	98%
12)	B.F./B.	81/ 96	B-->F	84%
13)	E.H./H.	33/ 82	H-->E	40%
13)	E.H./E.	33/ 44	E-->H	75%
14)	F.G./G.	18/ 21	G-->F	86%
14)	F.G./F.	18/ 83	F-->G	22%
15)	E.G./G.	5/ 21	G-->E	24%
15)	E.G./E.	5/ 44	E-->G	11%

No Row = 3290

4.1 Association rules results:

Regarding Gender 50%, Age>= 20years 96%, Family members >= 4 76%, Educational level 80%, Possess kills 44%, Corruption 83%, Disability and disease 21%, War and disaster 82%. Figure 1shows that the highest percentage is for people who are >= 20 years and suffering from corruption, wars, and disasters.

1- people with disability or disease are 20 years old or above.

2- All women who have families with 4 members or more, and become poor because of wars are 20 years old or above.

3- All women in a member-families >= 4 and suffer from corruption are 20>= years old.

4- 80% of the literate become poor because of wars and disasters.

5- 80% of member-families >= 4 become poor because of wars and disasters.

6- 98% of people suffering because of wars and disasters, are 20 years old or above.

7- 98% of individuals suffering from corruption are 20 years old or above.

8- 86% of the disabled suffer from corruption.

9- 88% of highly-educated people suffer from corruption.

10- 77% of craft or skill-workers suffer from corruption.

11- 84% of women suffer from corruption.

12- 84% of skill workers are providers of member-families >= 4.

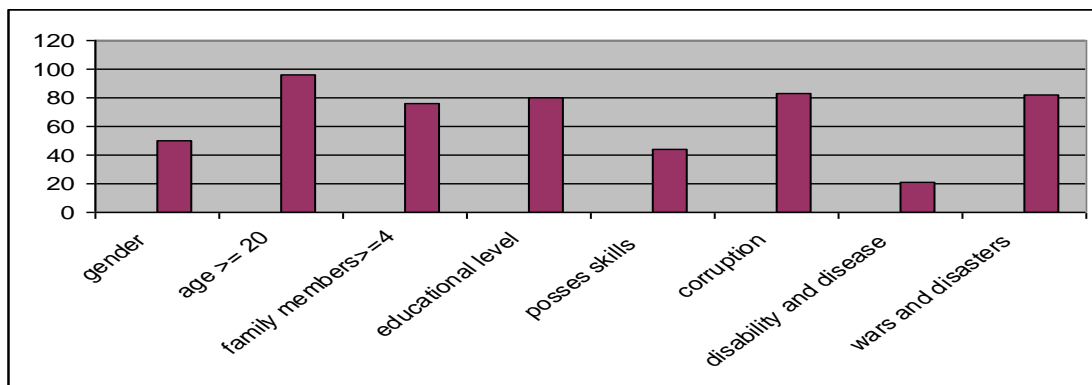13- 86% of educated people, who become poor because of wars, suffer from corruption.

Figure (1) factors percentage from the givendata

4.2. Mean of confidence and the effects of factors:

1. The occurrence of gender is female with other factors.

Regarding Age >= 20years 96%, Family members >= 4 56%, Educational level 78%, Posses skills 38%, Corruption 84 %, Disability and disease 20%, War and disaster 86%. Figure 2 shows that the most effective factors that occur with gender; female, being widows are 20>= years, corruption, wars, and disasters.
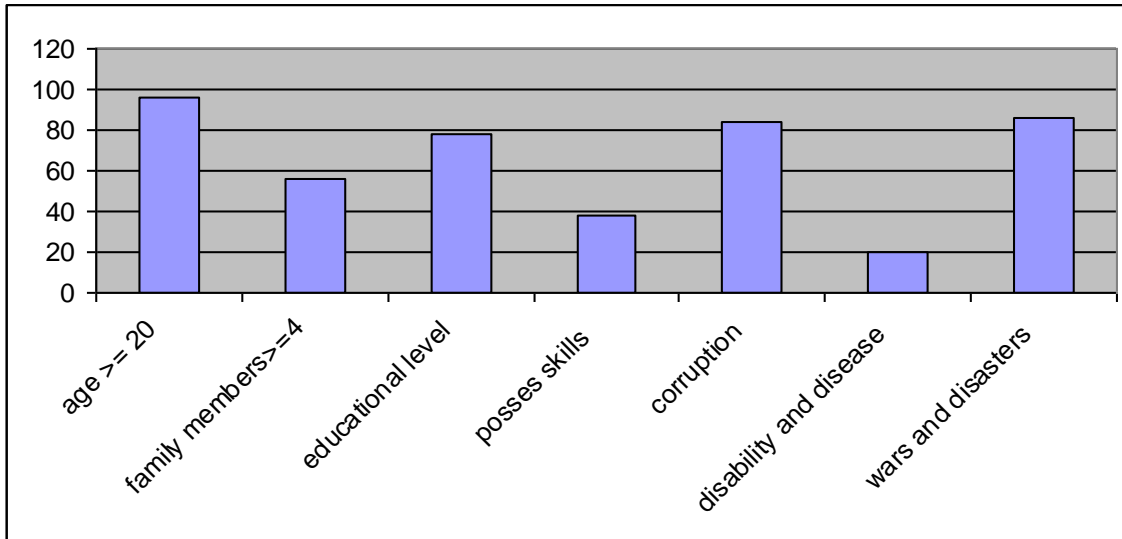

Figure (2) shows the Occurrence of gender with the other factors

2. The occurrence of age >= 20 years with other factors.

Regarding Gender 50%, Family members >= 4 77%, Educational level 79%, Posses skills 44%, Corruption 84 %, Disability and disease 22%, War and disaster 83%. Figure 3 shows that the most effective factors that occur with age are corruption, warctivs, and disasters.
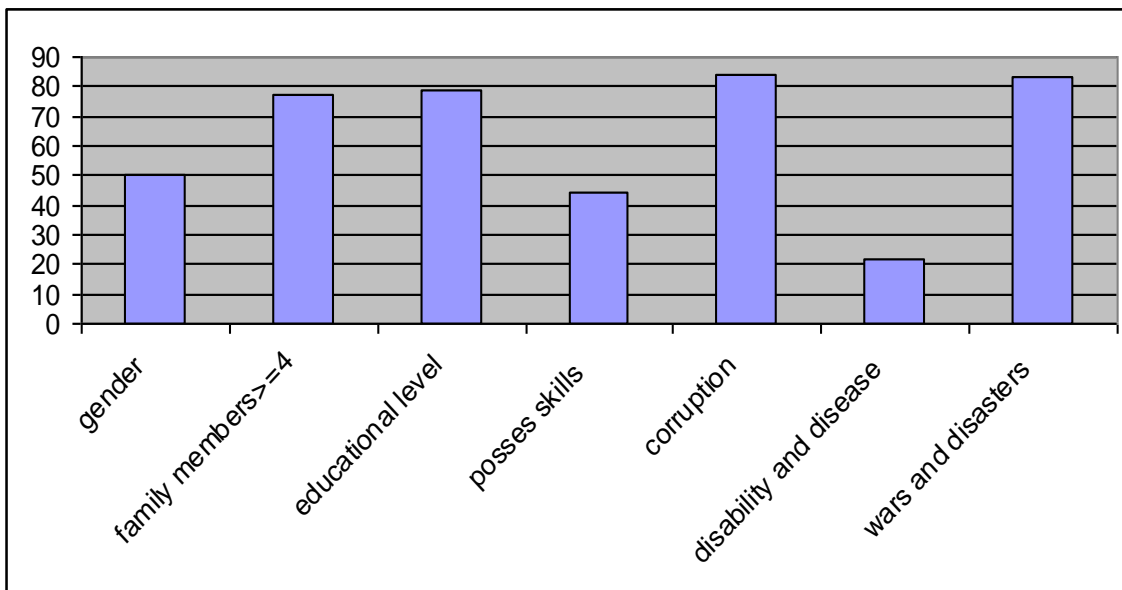

Figure (3) shows the Occurrence of age >= 20 years with the other factors

3. The occurrence of family members >= 4 with the other factors.
   Regarding Gender 37%, Age >= 20years 97%,
   Educational level 84%, Posses skills 49%, Corruption 83 %, Disability and disease 26%, War and disaster 80%. Figure 4 shows that the most effective factors, that occur with the family of members are age, educational level, and corruption.
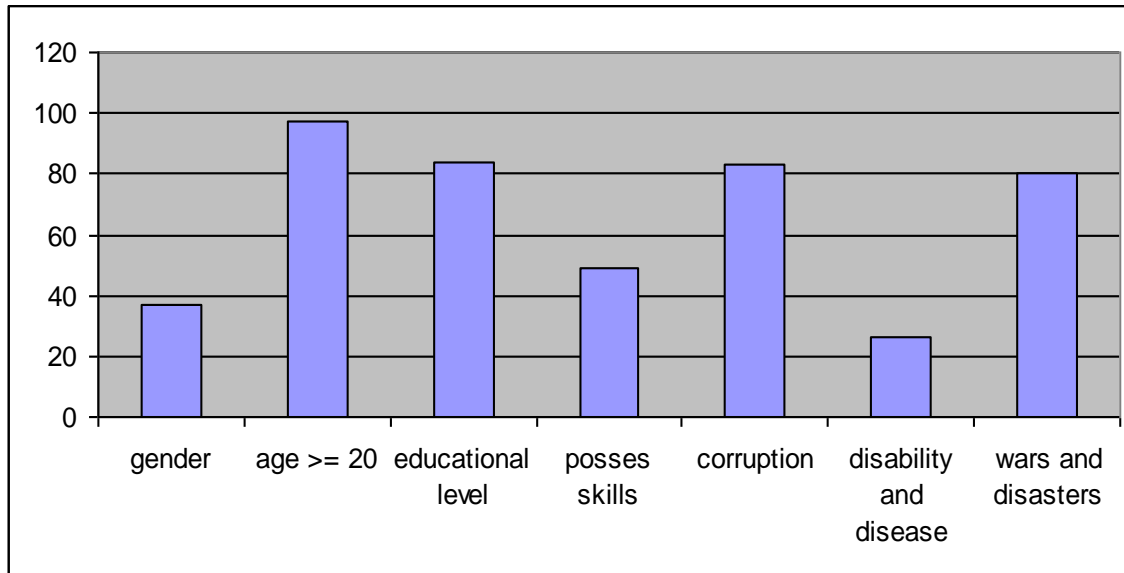

Figure (4) shows the Occurrence of family members >= 4with the other factors

4. The occurrenceof Educational levels with the other factors.
   Regarding Gender 49%, Age >= 20years 95%, Family members >= 4 80%, Posses skills 48%, Corruption 80 %, Disability and disease 19%, War and disaster 80%. Figure 5 shows that the most effective factors that occur with educational level are age, and equally the number of family members, corruption, wars, and disasters.
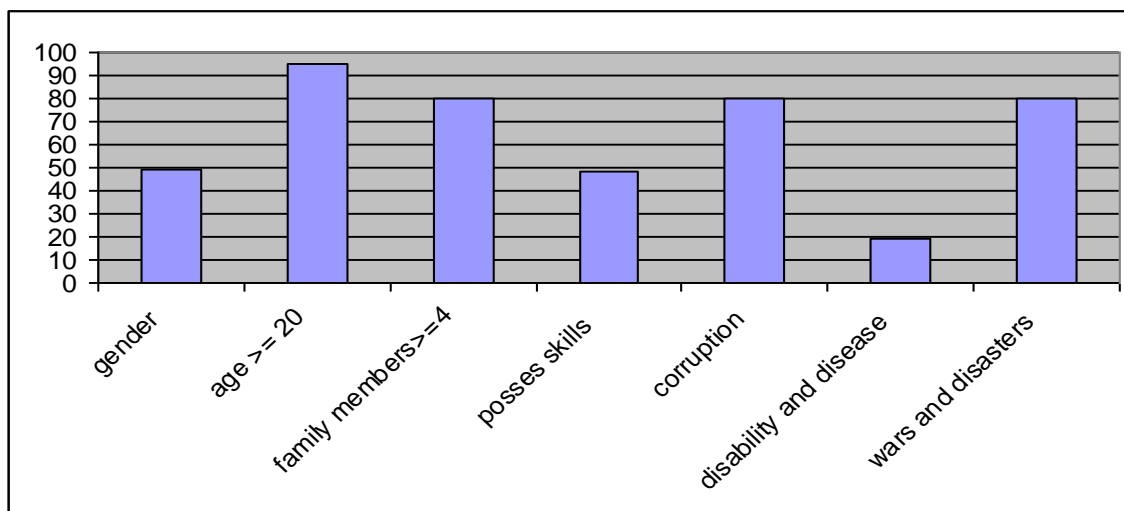

Figure (5) shows the Occurrence of Educational level with the other factors

5. The occurrence of posses skills with the other factors.
Regarding Gender 43%, Age >= 20 years 95%, Family members >= 4 84%, Educational level 86%, Corruption 77 %, Disability and disease 11%, War and disaster 75%. Figure 6 shows that the most effective factors that occur with Posses skills are age, educational level and the family members.
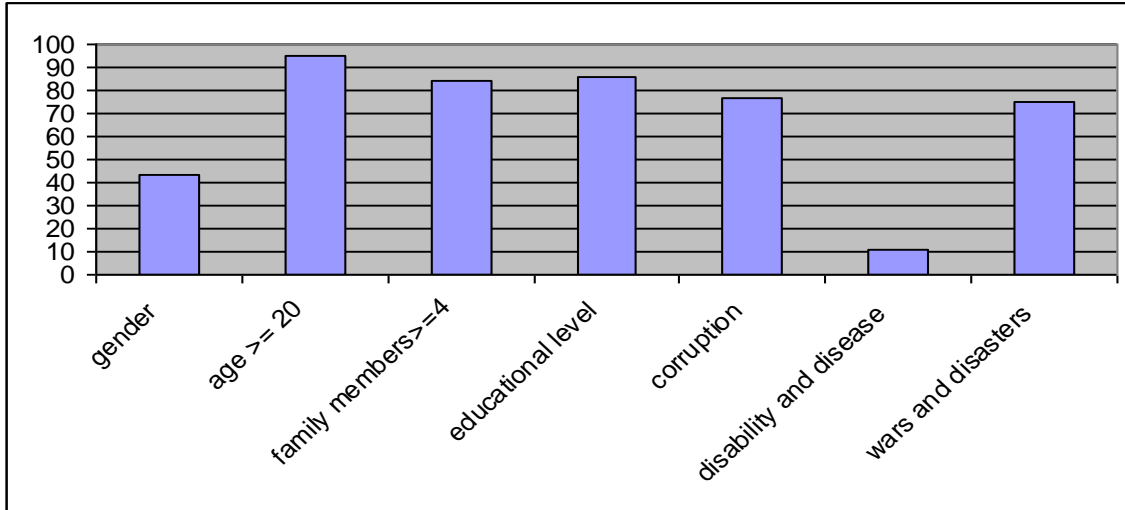

Figure (6) shows the Occurrence of posses skills with the other factors

6. The occurrence of corruption with the other factors.
Regarding Gender 51%, Age >= 20 years 98%, Family members >= 4 76%, Educational level 77%, Posses skills 41%, Disability and disease 22%, War and disaster 87%. Figure 7 shows that the most effective factors that occur with corruption are age, educational level, wars, and disasters.
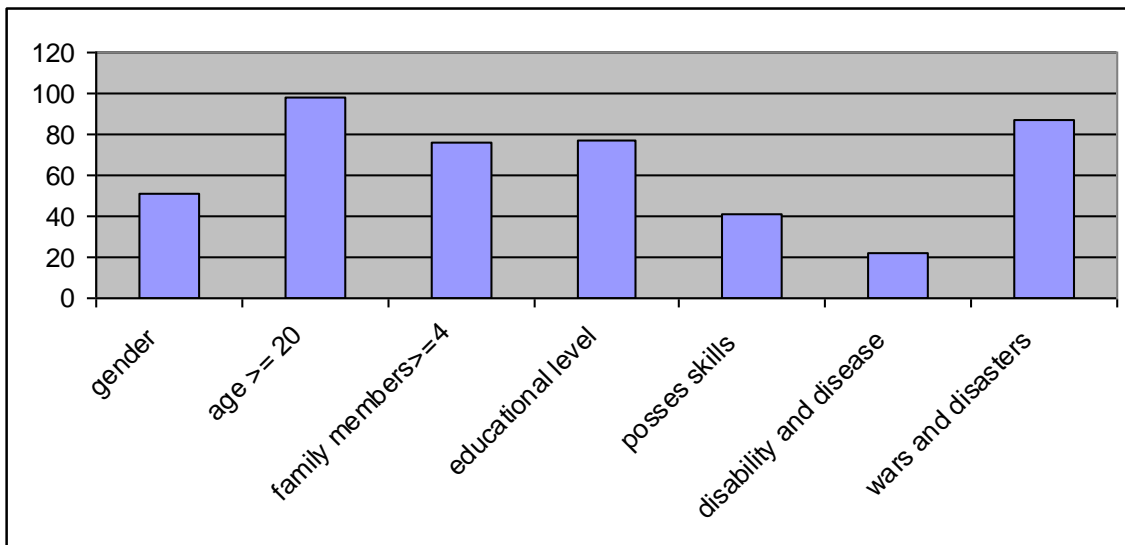

Figure (7) shows the Occurrence of corruption with the other factors

7. The occurrence of Disability and disease with the other factors.
Regarding Gender 48%, Age >= 20years 100%, Family members >= 4 95%, Educational level 71%, Posses skills 24%, Corruption 86 %, War and disaster 86%. Figure (8) shows that the most effective factors that occur with disability and disease are age, family members and equally corruption, wars, and disasters.
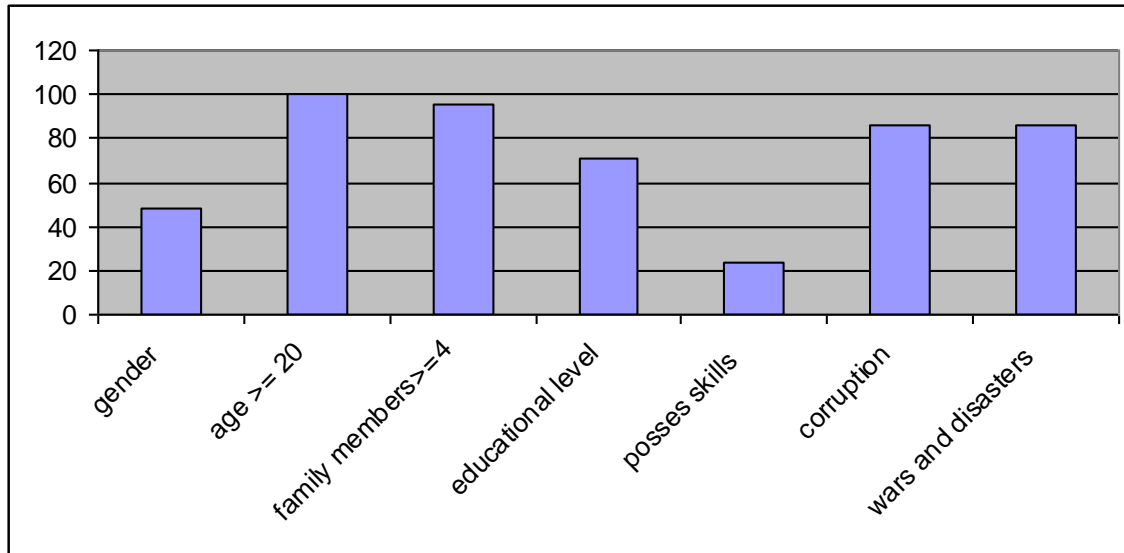

Figure (8) shows the Occurrence of Disability and disease with the other factors

8. The occurrence of War and disaster with the other factors.
Regarding Gender 50%, Age >= 20years 96%, Family members >= 4 76%, Educational level 80%, Posses skills 44%, Corruption 83 %, Disability and disease 21%. Figure 9 shows that the most effective factors that occur with war and disaster are age, corruption.
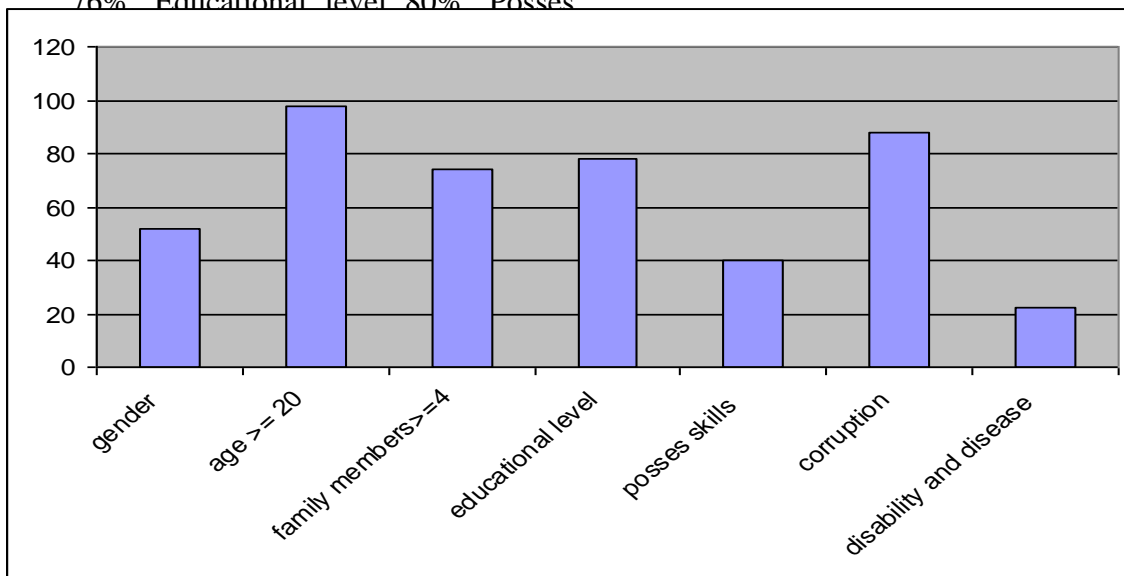

Figure (9) shows the Occurrence of the War and disasters with another factor

9. Mean of confidence (occurrence of the other factors with each single factor), Regarding Gender 65%, Age >= 20years 63%, Family members >= 4 65%, Educational level 64%, Posses skills 67%, Corruption 65 %, Disability and disease 73%, War and disaster 65%. Figure 10 shows that the average of means of occurrence for each factor on the other factors focuses mainly on disability - disease factor and Posses skill factor.
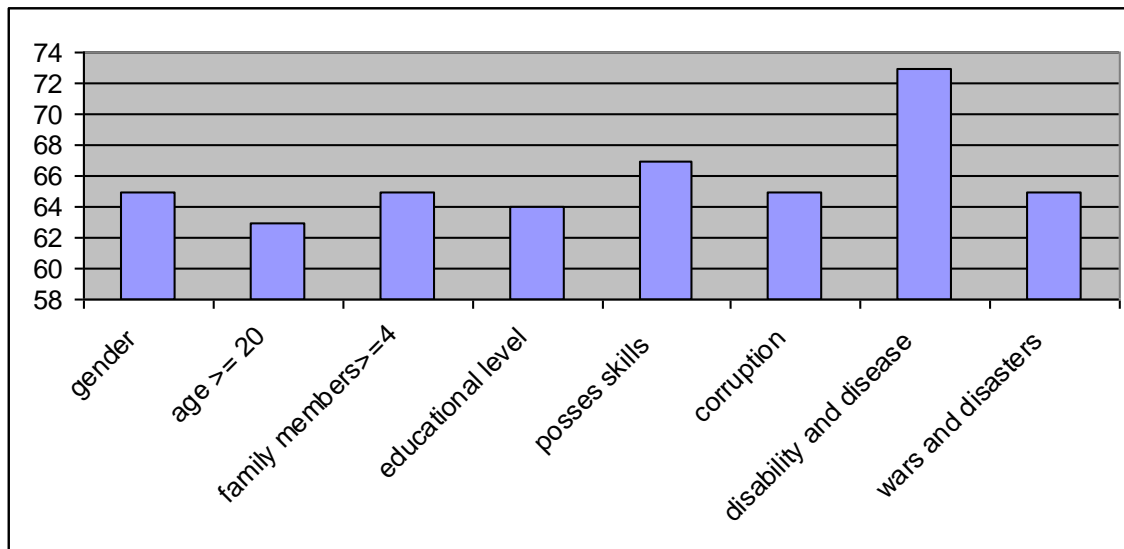

Figure (10) shows the Mean of confidence

5. Conclusions and Recommendations:

5.1 Conclusions:

1- The percentage of women, who are providers for poor families, are very close to a male percentage, because most of these women are widows, who lost their husbands in wars.

2- All people, who are at their 20s now, are orphans who lost their fathers in wars especially the Gulf War in 1991 and American Invasion in 2003. people, who are above the 20s, have participated in, suffered from the wars or widows.

3- Families that have 4 or more members: this is natural in the third world countries; people there don't follow birth control policy.

4- Educational level

The 70s have witnessed a great concern of education with all its sections elementary, general, professional and in military schools. Because being employed by the government was worthy; Iraqi dinar equaled 3.33 dollar at that time.

Though many generations of the educated and Posses skill have emerged, most of which died during the wars. people who managed to survive, suffered from the consequences of wars, and for them being employed by the government is no longer worthy ;1 dollar=3000 Iraqi dinar. The collapse of national currency has affected all aspects of life.

5- As a consequence of wars, economic collapse and lack of opportunities, a corrupted class emerged, they have practiced every type of corruption in order to keep their positions in the

government. This was so obvious after invading Bagdad; the country's infrastructure was destroyed, unplanned construction started. Many people made advantage of these circumstances; they caused poverty and unemployment for so many families because of their being corrupted besides destroying the private sector massively.Shutting down many important factories helped spreading poverty and unemployment.

6- One of the consequences of wars is the emergence of a lot of disabled and sick people due to neglecting public health issues in the country.

7- Wars are the main reason behind destroying any country. Disasters are also caused by wars; destroying Iraq and making its people poor is a consequence of the wars Iraq has been through.
All the country's incomes are spent to lengthen the war, the thing that weakens every aspect of life in this country. This is the dictatorship when it rules it causes destruction and when it's gone there would be destruction and misery as well.

5.2 Recommendations:

1-A country of integral institutions, non-insulating decision making, well choosing for competences owners within a system that neglects all inherited means, which caused spread of corruption, all these are the means needed to build a country free of corruption.

2-Establishing a new institution for widows to check people who are under the line of poverty and to provide job opportunities.

3-Establishing an institution for orphans who lost their fathers as a consequence of the wars the country has been through.

4-Rehabilitating every governmental companies that have been shut down, and rerun all the factories especially military industry companies, and make them civil companies.

5-Encouraging human investment through training and improving individuals to make them more efficient and productive, so they could get more wages or incomes.

6-Fighting against illiteracy and spreading cultural consciousness in the society.

7-Developing administrative work in health, education, family and social services fields.

8-Developing countryside and showing interest in agriculture which has been neglected for decades, encouraging rural industries using middle technological means to provide as job opportunities as possible.

9-Fighting finical and administrative corruption.

10-Take advantage of the civilized world experiences such as Japanese, Chinese and Indian experience. Homelands are built by loyal people, but people who are corrupted with sick, individual dreams destroy homelands.

References:
[1] B.M. HUSSEIN, Measuring the poverty in the Economic and Social Commission for Western Asia, a series of anti-poverty studies / United Nations / New York/1996.

[2] H. M. ALRUFAAI, the problem of poverty in the Islamic world / causes and solutions, ALNUKASH House for printing, publishing and distribution/Lebanon 2006.

[3] S. AZIZ. ALATBI, Statistic in Geography, AKRAM Press, Baghdad 2012.

[4] S. ALTAEER, Poverty and corruption in the Arabian world/DAR ALSAQI/London 2009.

[5] A. A. ALBALDAWI, Statistical methods for Economic Sciences and Business Administration with the use of SPSS program/First Edition /WAEL Publishing house/Jordan /Oman2009.

[6] A. ALFARIS, Poverty and the distribution of income in the Arabian land / Center for Arab Unity Studies/Bruit 2001.

[7] Y.S. Koh and N. Rountree, Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection, Information Science Reference (an imprint of IGI Global), New York - 2010.

[8] A. gkoulals and V. S. Verykios, Association Rule Hiding for Data Mining, Springer Science, Business Media, LLC, New York, USA ,2010.

[9] Zhang C. and S. Zhanger , Association Rule Mining Models and Algorithms, Springer-Verlag Berlin Heidelberg ,2002.

[10] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", In Proc. of Int. Conf. on Very Large Data Bases (VLDB'94), Santiago, Chile, September 1994, pp. 487-499.

[11] G. Johnny and H.S. ABBAS "Analytical Study using Association Rules and Mean of Confidence to Assess the Acceptance of (Industrial Schools/Computer) students at computer systems Dept\ Institute of Administration Rusafa" 2nd conference, Technical College of Management – Baghdad 28-29 /11/2012.