



Analysis Some Risk Factors on Brain Cancer Disease by Using Factor Analysis

Roqia Abdulqadir Jamel

1. Technical Institute of Sulamany, Kurdistan, Iraq

. roqia_abdulkadr@yahoo.com.

Article Information

Submission date: 25 / 6 / 2020

Acceptance date: 10 / 7 / 2020

Publication date: 30 / 6 / 2020

Abstract

Brain cancer is the cancer of the central information-processing center of the body. Brain tumour can be malignant or benign, and can occur at any age. Only malignant tumors are cancerous. A primary brain tumor initially forms in the brain tissue. In addition, the aim of this study is to investigate the most important factors affecting on the brain cancer and identify the demographic of the disease of brain cancer, then factor analysis has been used to identify factors affecting brain cancer, The methods of this study was used descriptive statistics and factor analysis, then Factor Analysis (FA) is an exploratory technique applied to a set of observed variables that seeks to find underlying factors (subsets of variables) from which the observed variables were generated. Moreover, recorded data for brain cancer were taken from Hiwa Hospital in Sulaymaniyah Governorate during the period of (2018 to 2020). 150 patients and 8 variables were included in the study. After analysis study detected from factor which are explain 59.036% of the variance, and the result identified three factors such that, 1stFactor: This factor containing these variables (Age, Weight) by total variance 21.400%, 2ndFactor: This factor containing these variables (Gender, Smoking and Occupation) by total variance 20.665%, 3rdFactor: This factor containing the variable (Governorate, Blood type and Drinking) by total variance 16.971%, The papers limitation is including lack of variables taken for brain cancer and the number of samples was insufficient. Finally, the rate of 59.036% aforementioned variables influence on the brain cancer.

Keyword: Factor analysis, subsets of variables, brain cancer, underlying factors

Introduction

The fatality rate due to brain cancer is the highest in Asia ⁽¹⁾ Brain cancer develops in both the brain and spinal cord ⁽²⁾ The various symptoms of brain cancer include coordination issues, mood swings, frequent headaches, and difficulty in concentration, changes in speech, seizures and memory loss. Brain cancer is a kind of tumour which stays in the brain or central nervous system ⁽²⁾ Brain tumour are categorized into different types based on their origin, nature, rate of growth and progression stage ⁽³⁾. Brain tumour may be either benign or malignant. Benign brain tumour cells rarely invade surrounding healthy cells, have distinct borders and a slow progression rate (e.g., pituitary tumour, meningiomas, and astrocytomas). Malignant brain tumour cells (e.g., oligodendrogliomas, high-grade astrocytomas,) readily attack nearby cells in



the brain or spinal cord, have fuzzy borders and rapid progression rates. Based on their origin brain tumour can be further classified into two types: primary brain and secondary brain tumour. A primary tumour originates directly from the brain. If the tumour emerges in the brain due to cancer existing in some other body organ such as breasts, lungs, stomach etc., then it is known as a secondary brain tumour or metastasis. Further, grading of brain tumour is done according to the rate of growth of cancerous cells, i.e., from low to high grade. WHO categorizes brain tumour into four grades (I, II, III and IV) as per the rate of growth ^(2,4). (Brain tumour is also characterized by their progression stages (0, 1, 2, 3 and 4). Stage-0 refers to cancerous tumour cells, which are abnormal, but do not spread to neighbouring cells. Stages-1, 2 and 3 denote cells, which are cancerous and rapidly spreading. Finally in Stage 4 the cancer spreads throughout the body. Many lives could be saved if cancer were detected at an early stage by fast and cost-effective diagnosis techniques. However, it is very difficult to treat cancer at high stages where survival rates are low.

Brain cancer diagnosis may be either invasive or non-invasive. Biopsy is the invasive way where an incision is done to collect a tumour sample for examination. It is considered the gold standard for diagnosis of cancer where the pathologists observe various features of cells of the tumour sample under a microscope to confirm malignancy. The physical examination of the body and brain scanning using imaging modalities are considered as non-invasive approaches. The various imaging investigations such as magnetic resonance imaging (MRI) or computed tomography (CT), of brain are faster and safer techniques than biopsy. These imaging modalities help radiologists to locate brain disorders, observe disease progression and surgical planning ⁽⁵⁾ Brain scans or brain image reading to diagnose disorders is however subject to inter-reader accuracy and variability, which depends on the proficiency of the medical practitioner.

The cell is the main building block of the human body. It also defines the function of each organ within the body such as blood flow, oxygen flow, and waste materials management. Every cell has a central control system known as the nucleus, which contains 23 pairs of chromosomes consisting of millions of genes. The instructions for these genes are contained within deoxyribonucleic acid (DNA) ⁽⁶⁾. DNA is like a blueprint for genes and defines their behaviour. The protein of the gene is like a messenger that communicates between the cells or the genes themselves. Genes control the continuous process of the death of unwanted or unhealthy cells besides reproduction of healthy cells. The main cause of a cancer is uncontrolled growth of cells. A mutation alters this DNA sequence, which is the main cause of malfunctioning of the genes. There are many factors involved in DNA mutations such as lifestyle, environmental, and eating habits.

Worldwide, about 15 percent of cancers are caused by viruses ⁽⁷⁾. The viruses infect cells via altering DNA in the chromosomes which are responsible for converting proto-oncogenes into oncogenes. Only few cancers causing viruses have been identified i.e., DNA virus and retroviruses or Nocona viruses (an RNA virus). The four main DNA viruses responsible for human cancers are human papillomavirus, Hepatitis B, Epstein-Barr, and human herpes virus. The RNA viruses which cause cancer are type1 Human T lymphotropic and hepatitis C. Several

environmental factors also affect the cells. UV light, X-rays, viruses, tobacco products, pollution and many other daily use chemicals carry carcinogenic agents. Sunlight can also alter tumour suppressor genes in skin cells leading to skin cancer. Further, the carcinogenic compounds in smoke change the lung cells causing lung cancer ⁽⁸⁾.

Certain types of genes define the cancer. In various studies, connections are found between degree of mutation in genes and type of brain tumour. Tumour protein-53 (TP53) is involved in DNA repair and initiating apoptosis. Tp53 level is found to be quite abnormal in high-grade gliomas and mutations have been found in more than 80% of tumours ⁽⁹⁾. The retinoblastoma gene is a tumour suppression gene. Retinoblastoma mutation is found in approximately 75% of brain tumours and it is more relevant to glioblastoma ⁽⁹⁾. Brain tumour analysis using medical imaging is a challenging and complicated task, which can be categorized into pre-processing, classification and post-processing steps. There are many challenges associated with the aforementioned steps, which make this task more complicated. No ideal computer assisted tools available so far to conform, tumour malignancy and the degree of aggressiveness. Thus doctors depend on the biopsy test ⁽¹⁰⁾ only for all types of cancer.

Methods of Analysis

Factor Analysis:

1- Factor analysis procedures: are statistical methods used for examining the relationships within a group of observed variables, as measured through questions or items. It is important to note that factor analysis is not a singular statistical method, but rather a group of statistical analyses that share similar methodology and functionality. The theoretical and mathematical variations among the processes allow the analyses to accommodate breadth of purpose and theory in research and result in the widespread use of the tool across disciplines and applications; however, it is the flexibility of the statistical methods that fuel ongoing debate about the appropriate applications of these methods ⁽¹¹⁾. However, factor analysis method can effectively overcome these problems and make a important evaluation of the level of economic development between countries ⁽¹²⁾. Moreover, factor analysis attempts to represent a set of observed variables x_1, x_2, \dots, x_n in terms of a number of 'common' factors plus a factor which is unique to each variable. The common factors (sometimes called latent variables) are hypothetical variables which explain why a number of variables are correlated with each other- it is because they have one or more factors in common. A concrete physical example may help such as we measured the size of various parts of the body of a random sample of humans: for example, such things as height, leg, arm, finger, foot and toe lengths and head, chest, waist, arm and leg circumferences, the distance between eyes, etc. We'd expect that many of the measurements would be correlated, and we'd say that the explanation for these correlations is that there is a common underlying factor of body size. It is this kind of common factor that we are looking for with factor analysis, although in psychology the factor may be less tangible than body size. To carry the body measurement example further, it can't probably expected body size to explain all of the variability of the measurement for example, there might be a lankiness



factor, which would explain some of the variability of the circumference measures and limb lengths, and perhaps another factor for head size which would have some independence from body size (what factors emerge is very dependent on what variables measured) Even with a number of common factors such as by size, lankiness and head size, still it can't be expected to account for all of the variability in the measures (or explain all of the correlations), so the factor analysis model includes a unique factor for each variable which accounts for the variability of that variable which is not due to any of the common factors Smaller number of factors without losing too much information⁽¹³⁾.

1-(Amy S. Beavers, pp:1-13,,2013)• 2-(Anita Bai, pp:311-317,2015)

3-(Salfan, Aimad aldin, 1987). (Abbduljabarallbiati 1977)•

Factor analysis attempts to represent a set of observed variables x_1, x_2, \dots, x_n in terms of a number of 'common' factors plus a factor which is unique to each variable. The common factors (sometimes called latent variables) are hypothetical variables which explain why a number of variables are correlated with each other- it is because they have one or more factors in common. A concrete physical example may help. Say we measured the size of various parts of the body of a random sample of humans: for example, such things as height, leg, arm, finger, foot and toe lengths and head, chest, waist, arm and leg circumferences, the distance between eyes, etc. We'd expect that many of the measurements would be correlated, and we'd say that the explanation for these correlations is that there is a common underlying factor of body size. It is this kind of common factor that we are looking for with factor analysis, although in psychology the factor may be less tangible than body size. To carry the body measurement example further, we probably wouldn't expect body size to explain all of the variability of the measurement: for example, there might be a lankiness factor, which would explain some of the variability of the circumference measures and limb lengths, and perhaps another factor for head size which would have some independence from body size (what factors emerge is very dependent on what variables measured). Even with a number of common factors such as by size, lankiness and head size, we still wouldn't expect to account for all of the variability in the measures (or explain all of the correlations), so the factor analysis model includes a unique factor for each variable which accounts for the variability of that variable which is not due to any of the common factors. Smaller number of factors without losing too much information, we have achieved some economy of description, which is one of the goals of scientific investigation. It is also possible that factor analysis will allow us to test theories involving variables which are hard to measure directly. Finally, at a more prosaic level, factor analysis can help us establish that sets of questionnaire item (observed variables) are in fact all measuring the same underlying factor (perhaps with varying reliability) and so can be combined to formula a more reliable measure of that factor. There are a number of different varieties of factor analysis: the discussion here is limited to principal axis factor analysis and factor solution in which the common factors are uncorrelated with each other. It is also assumed that the observed variables are standardized (mean zero, standard deviation of one) and that the factor analysis is based on the correlation matrix of the observed variables.

2: The Factor Analysis Model

Models of factor analysis explain (p) of variables to sample size (n) based on linear function and composed by (p) of unique factor so that each variable ($m < p$), this linear model is formed as follows:

$$X_{px1} = \mu_{px1} + A_{pxm} F_{mx1} + U_{px1} \dots\dots\dots (11)$$

Where

X: represent vector of random variables for viewing

μ : represent vector of the mean variable

A: represent common factor matrix of variables

F: represent random common factor that have been tested of variables

U: represent vector of unique factor in variables

Based the main from of factor analysis in describing single value (i) and variable (j) denoted by the total of weighted value in factor. It can be formulated this model into

$$Z_{ij} = a_{j1} F_{1j} + a_{j2} F_{2j} + \dots + a_{jm} F_{mj} + U_{ji} \dots\dots\dots (12)$$

Where

Z_{ij} : Normative value of a single or variables

$F_{mi} \dots F_{1j}$: Normative value of single common factor

U_{ji} : Normative value of single (i) and unique factor of variables (j)

$a_{i1} \dots\dots\dots a_{jm}$: transferring weighted value in common factor

3: Basic Assumption of Factor Analysis

correlation Matrix or Covariance matrix is used in factor analysis, that in any fundamental primes that this analysis is based on the existence of correlation between the studied variables as factor analysis seeks to explain these correlations with the number of factors and dependent factor is used as the standard variables which have mean equal to zero and variance equal to one, this measurement units have different variables:

1. The mean vector of the variable is to be zero

$$E(X) = M = 0$$

2. The variance vector of the variable is to be one

$$Var(X) = 1$$

In this case, the form of factor analysis model as follows:

$$X = AF + U \dots\dots\dots (13)$$

1. Common factors (F) is independent among them that any

$$E(F) = mx_1 \cdot cov(F) = E(FF') = \varphi (m \times m)$$

2. The unique factors (U) is independent among them that any

$$E(U) = (p \times p)$$

Covariance matrix

$$Cov(U) = E(UU') = \Psi_{p \times p} = \begin{pmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & 0 & \dots & 0 \\ 0 & 0 & \dots & \Psi_p \end{pmatrix} \begin{pmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & 0 & \dots & 0 \\ 0 & 0 & \dots & \Psi_p \end{pmatrix}$$

3. Common factor and unique factor independent is which any

$$E(F' U') = \begin{bmatrix} E(FF') & E(FU') \\ E(FU') & E(UU') \end{bmatrix} = \begin{bmatrix} 0_{m \times m} & 0 \\ 0 & \Psi_{p \times p} \end{bmatrix} \begin{bmatrix} E(FF') & E(FU') \\ E(FU') & E(UU') \end{bmatrix} = \begin{bmatrix} 0_{m \times m} & 0 \\ 0 & \Psi_{p \times p} \end{bmatrix}$$

Where

φ : represent the coefficient of matrix co-factor (F)

Ψ : represent the unit of factor.

By using the main form (2.1) in factor analysis, derivative the covariance matrix for (X) variable as follows:

$$\begin{aligned} \underline{X} &= \underline{\mu} + \underline{A}\underline{F} + \underline{U} \\ (X - \mu)(X - \mu)' &= (\underline{A}\underline{F} + \underline{U})(\underline{A}\underline{F} + \underline{U})' \\ &= (\underline{A}\underline{F} + \underline{U})(\underline{A}\underline{F})' + \underline{U}\underline{U}' \\ &= \underline{A}\underline{F}(\underline{A}\underline{F})' + \underline{U}(\underline{A}\underline{F})' + \underline{A}\underline{F}\underline{U}' + \underline{U}\underline{U}' \\ \Sigma &= Cov(X) = E(X - \mu)(X - \mu)' = (X - \mu)' \\ &= \underline{A}E(\underline{F}\underline{F}')\underline{A}' + E(\underline{U}\underline{F}')\underline{A}' + \underline{A}E(\underline{F}\underline{U}) + E(\underline{U}\underline{U}') \\ &\text{In the previous hypotheses that} \\ &E(\underline{F}\underline{F}')\underline{A}' = E(\underline{U}\underline{F}') = E(\underline{F}\underline{U}) = E(\underline{U}\underline{U}') = \Psi \\ \Sigma &= \underline{A}\underline{A}' + \Psi' \end{aligned}$$

4: Variance of Factor Analysis

The quantities at the bottom of each factor are the sums of the squared lodgings for that factor, and how much of the total variance of the observed variables is accounted for by that factor

$$\begin{aligned} Var(X_i) &= 1 = a_{i1}^2 Var(f_1) + a_{i2}^2 Var(f_2) + \dots + a_{im}^2 Var(f_m) + Var(e_i) \\ &= a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 Var(e_i) \dots \dots \dots (14) \end{aligned}$$

The quantity $a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2$ is called the communality of x_i and is the part of the variance that is related to the common factors. $Var(e_i)$ is called the specificity of x_i that is unrelated to the common factors.

It can be shown that the coefficient between any two values x_i and x_j is given by

$$r_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + \dots + a_{im}a_{jm} \dots \dots \dots (15)$$

Hence two test scores can only be highly correlated if they have high lodgings on the same

factors. Also, since communality cannot be more than one, $-1 \leq a_{ij} \leq 1$ for every i and j .

5: How Many Factors

A factors solution with as many factors as variables would score highly on the amount of variance accounted for and the accurate reproduction of correlations, but would fail on economy of description, parsimony and explanatory power. The decision about the number of common factors to retain, or to use in rotation to simple structure must steer between the extremes of losing too much information about the original variables on one hand, and too many factors on the other. Various criteria have been suggested. The standard one (but not necessarily the best) is to keep all the factors which have eigenvalues greater than one in the original solution.

6: communalities

The amount of any common variable is a collection least transferring factors that variable is part of the total variance that determines which of the overlap between the variables and common factors derived from the analysis of the correlation matrix.

$$H_i = \sum_{k=1}^m a_{ki}^2 \dots\dots(1.6)$$

Where a_{ki} : matrix factor (F) weighting factor (k) and variables (i)

$J=1,2,3,\dots,p$: the number of variables.

$K=1,2,3,\dots,p$ the number of factor.

When we use a method of analyzing the main factor as one of the method of pickling factor analysis, we need to estimate the amount of the first common (h_j)

To replace element of the in the correlation matrix (R) unit and this is called correlation.

Since the communality values are part of the total variance, it is always a positive value and its value lies between zeros one.

$$0 \leq h_j \leq 1$$

There are three descriptions of communality values of variables

1. When the communality values close to one indicate a correlation between variable factors and extracted factor.
2. When the communality values are between zero and one, it indicates the presence of partial overlap between variable factors and extracted factors.
3. When the communality values close to zero, which is, not indicated a correlation between variable factors and extracted factors.

7: Factor Analysis Methods

We reach the last results in our study to the phenomenon using tow-step main factor analysis:

1. Initial solution
2. Final solution

There are many methods to get the first solution to estimate the matrix of factor

Principle Components Method

1. Principle Factor Method
2. Maximum Likelihood Method
3. The Centroid Method
4. Triangular Decomposition Method
5. Multiple Group Method
6. Uni-Factor Method
7. Two-Factor Method
8. Bi-Factor Method
9. Minimum Residuals Method
10. Alpha Method
11. Image Method
12. Rao Method
13. Un Weighted Least Square Method
14. Generalized Least Squares

7.1: Principle Factor Method

The third in the section format matrix can be written as follows

$$\begin{bmatrix} X_1 \\ \vdots \\ X_p \\ X_{p+1} \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} U_1 \\ \vdots \\ U_p \\ \vdots \\ U_n \end{bmatrix}$$

For estimation matrix factor we follow the following steps

1. Correlation Coefficient Matrix

Of the standard values (Stander value) for the variables, this is if the variables are different units of measurement, and if the variables have different units of measurement. If the variables have the same units of measurement, we use the Variance Covariance Matrix. (Variance Covariance Matrix)

$$R_{ij} = \frac{\sum_{k=1}^n X_{ik} X_{jk}}{n - 1}$$

Where R_{ij} : Variance Covariance Matrix

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}$$

2. Calculate the square of the multiple correlation coefficient for each variable with the rest

of the variables (R_{2rest}) first estimate of the common values to replace the single digits as elements of any country Replace (r) (R_{1rest}) and (r_{22}) in (R_{2rest}). Thus, at Reduced Correlation Matrix is symbolized (R_r).

3. The correlation matrix is extracted the sample values (Eigen Values) Positively under the characteristic equation as follows

$$Rr - \lambda I a = 0 \dots \dots \dots (18)$$

8: Extracting Factors and Rotation of Factors

The mathematical process used to obtain a factor solution .a correlation matrix is such that each successive factor, each of which is uncorrelated with the other factors, accounts for as much of the variance of the observed variables as possible. (The amount of variance accounted for by each factor is shown by a quantity called the eigenvalue, which is equal to the sum of the squared loadings for a given factor, as will be discussed below). This often means that all the variables have substantial lodgings on the first factor; i.e., that coefficient a_{11} , a_{21} , ..., a_{nm} are all greater than some arbitrary value each as .3 or .4. While this initial solution is consistent with the aim of accounting for as much and possible of the total variance of the observed variables with as few factors as possible, the initial pattern is often adjusted so that each individual variable has substantial loading on as few factors as possible (preferably only one). This adjustment is called rotation to simple structure, and seeks to provide a more interpret abler outcome. As will be seen in the example which we'll work through later, rotation to simple structure can be seen graphically as the moving or rotation of the axes (using the term 'axis' in the same way as it is used in 'x-axis' and 'y-axis') which represent the factors

9: Calculating Correlation from Factors

It was mentioned above that an aim of factor analysis is to 'explain' correlations among observed variables in terms of a relatively small number of factor. One way of gauging the success of a factor solution is to attempt to reproduce the original correlation matrix by using the loadings on the common factor and seeing how large a discrepancy there is between the original and reproduced correlations – the greater the discrepancy, the less successful the factor solution has been in preserving the information in the original correlation matrix. How are the correlations derived from the factor solution? When the factors are uncorrelated, the process is simple. The correlation between variables X_1 and X_2 is obtained by summing the product of the coefficients for the two variables through all common factor; for a three-factor solution, the quantity would be $(a_{11} * a_{21}) + (a_{12} * a_{22}) + (a_{13} * a_{23})$. This process will become clearer in the description of the hypothetical two-factor solution based on five observed variables in the next section.

10: kaiser – Meyer –Olkin and Bartlett's Test

10.1: Bartlett's test of It

It is used to test the hypothesis that the correlation matrix is an identity matrix (all diagonal terms are one and all off-diagonal terms are zero).you are looking for significance (less than 0.05) because you want the variables to be correlated .In other words, picture a correlation matrix: all items are perfectly correlated with themselves (one),and have some level of correlation with other items. If they are not correlated with the other items then they can't be part of the same factor!

10.2: Kaiser-Meyer-Olkin:(KMO)

Measure of sampling adequacy is used to compare the magnitudes of the observed correlation coefficients in relation to the magnitudes of the partial correlation coefficients. Large KMO values are good because correlation between pairs of variables (i.e., potential factor) can be explained by the other variables. If the KMO is below 0.5, factor analysis isn't do the data.

The formula for the KMO is (the sum of the observed correlation coefficients) + (the sum of the partial correlation coefficients between any pairs of variables). If the sum of the partial correlation coefficients between all pairs of variables is small when compared to the observed correlation coefficients, the KMO measure will be close to one. (Remember, a partial correlation is a measure of the strength of the relationship between any two variables when the other variables are held constant)

Aim of the study

The aim of this study is to investigate the most important factors affecting on the brain cancer and identify the demographic risk factors of the disease.

Research Summary

From the data we obtained, which was a sample size (150) patients and (8) variables under study, to determine the importance and role of all variables separately once, and the impact of each of them again. This paper shows that descriptive statistics for variables and using the factor analysis by statistical program (SPSS₂₂).

-Methodology

The study was conducted using factor analysis to investigate the most important factors affecting on the brain cancer recorded data for brain cancer were taken from Hiwa Hospital in Sulaymaniyah Governorate during the period of (2018 to 2020), which was a sample size (150) patients and (8) variables, The data entered the statistical program (SPSS₂₂) and the data for each of these variables were studied using descriptive statistics (frequency and percentage) of variables and the factor analysis (Principal Factor) to obtain the logical results according to the accuracy and validity of the data.

-Statistics analysis

Table (1): Study variables

X_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Study variables	Gender	Age	Governorate	Blood group	Smoking	Alcohol Drinking	Weight	Occupation

First: Descriptive statistics

Frequency distribution and relative percentage can be found from the variables information

Table (2): Gender

Gender	Number of Patients	Percent
Male	100	66.67%
Female	50	33.33%
Total	150	100%

Table (2) shows that the majority of gender was males, about 66.67% of the total. While the females is 33.33%



Table (3): Age

Age	Number of Patients	Percent
1 – 19	36	24%
20 – 39	50	33.33%
40 – 59	40	26.7%
More than 59	24	16%
Total	150	100.00

It can be seen in the table (3) that 33.33% of participant is between 20 to 39 years old, 26.7% of them were between 40 to 59 and 24 % of the participant were year between 1to 19 years old and also 16% of them were between more than 59 years old.

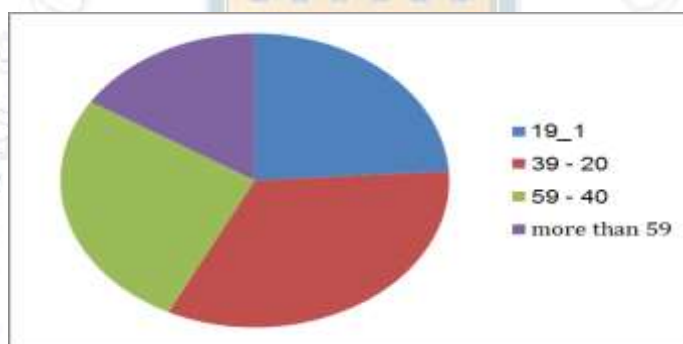


Table (4): Governorate

Governorate	Number of Patients	Percent
Sulaimani	65	43.34%
Karkuk	20	13.33%
Arbil	10	6.67%
Duhok	5	3.33%
Others	50	33.33%
Total	150	100.00

From the above table, the sample of the study in the Governorate was divided into five types. The highest percentage of the Governorate was (Sulaimani city) which was 65%, while the others was 50%, 13.33% of the total was from Karkuk city, while the rate of people living in (Arbil city) was 6.67%.

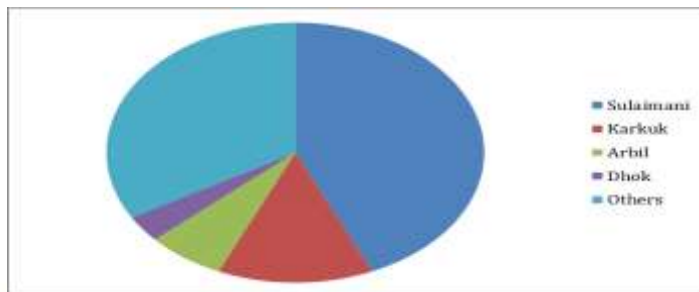


Table (5): Blood group

Blood group	Number of Patients	Percent
A	75	50%
B	25	16.67%
AB	30	20%
O	20	13.33%
Total	150	100.00

Table (5) shows the blood group of the sample of the study. The table shows that the blood group (A) of the study sample came first with 50% .Second place was the blood group (AB) by (20%). In the third place was the blood group (B) by (16.67%),



Table (6): Smoking

Smoking	Number of Patients	Percent
Yes	30	20%
No	120	80%
Total	150	100.00

Through the above table, we see that smoking people are less affected than those who are non-smokers, as the proportion of non-smokers is (80%) more than smokers, which constitutes (20%) .

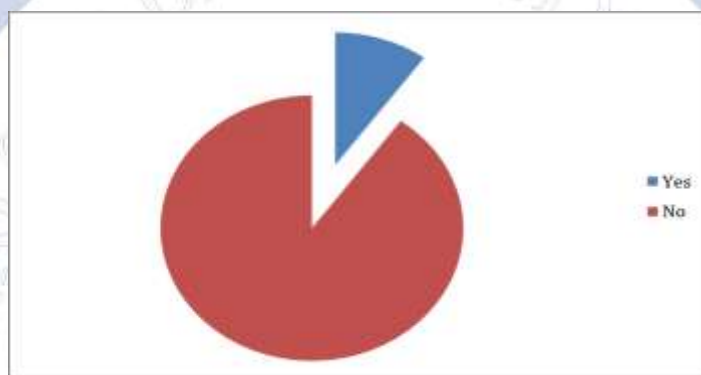


Table (7): Alcohol Drinking

Drinking	Number of Patients	Percent
Yes	15	10%
No	135	90%
Total	150	100.00

Through the above table, we see that the incidence is less in drinkers in comparison with those who are non- drinkers, as the proportion of non-smokers is (90%) but only (10%) in smokers.

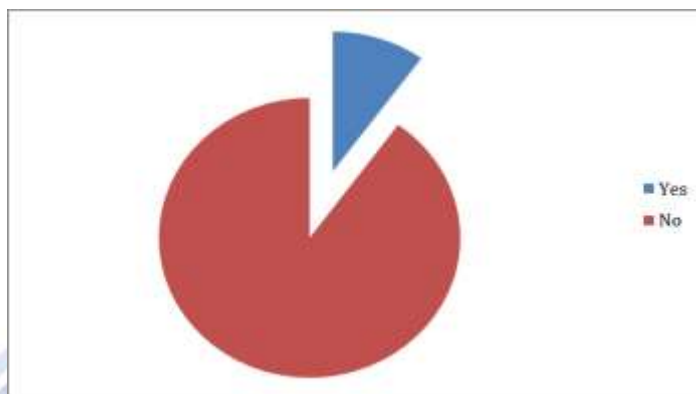


Table (8): Weight

Weight	Number of Patients	Percent
1-50	15	10%
51-100	105	70%
More than 100	30	20%
Total	150	100.00

It is obvious that, 70% weights range was between (51-100) kg, which was the highest rate among all types weight. While the rate of (1-50) Kg was 10%, which was the lowest rate.

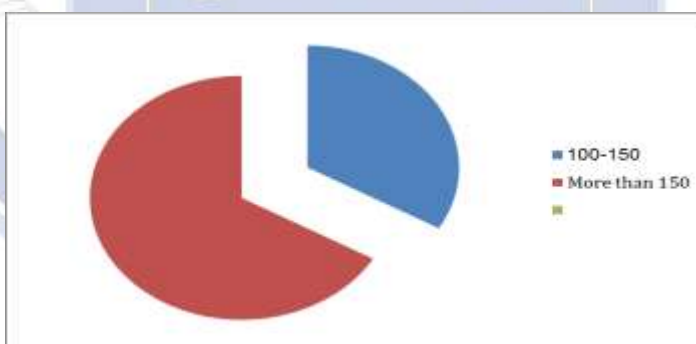
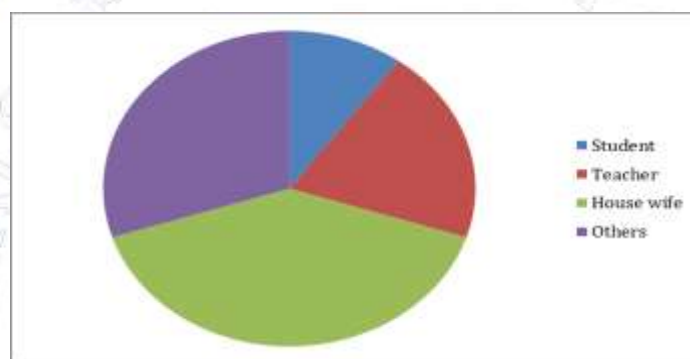


Table (9): Occupation

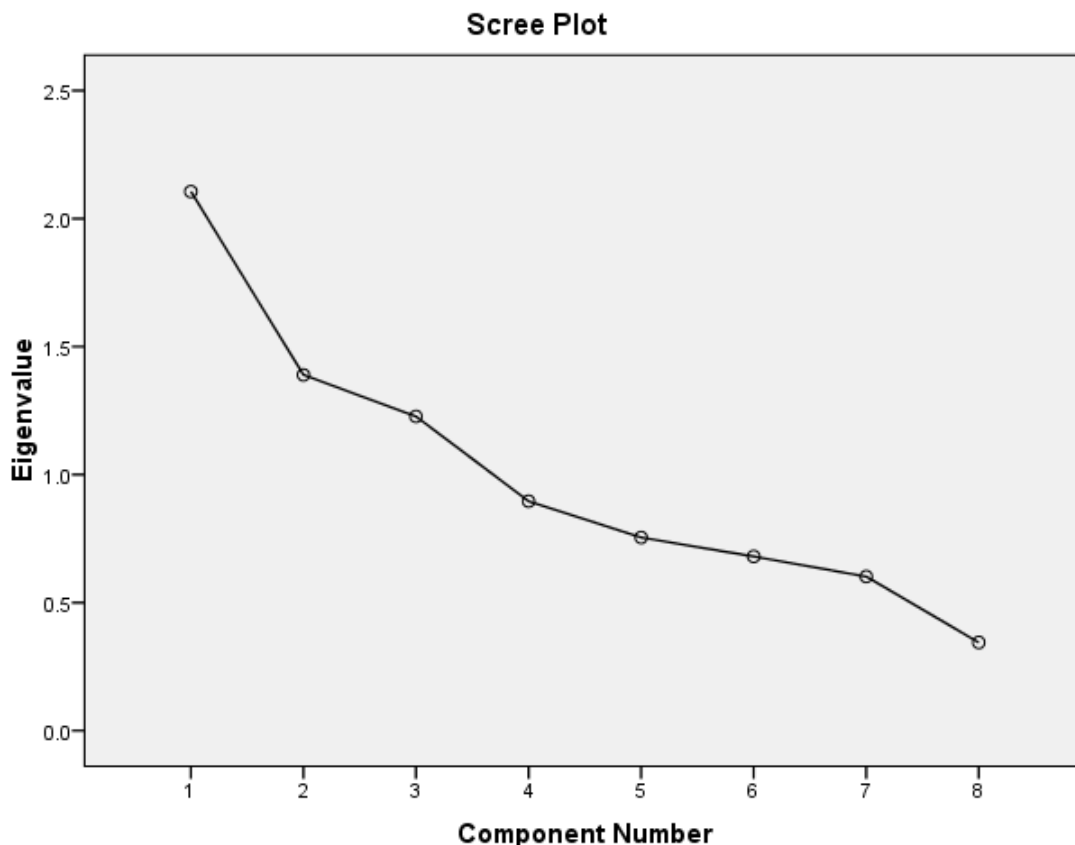
Occupation	Number of Patients	Percent
Student	15	10%
Teacher	30	20%
House wife	60	40%
Others	45	30%
Total	150	100.00

From the table, it appears that those who have a job title (House wife) are more frequent by (60) out of a total of (150), (40%).



Second: Factor Analysis

The analysis of the brain cancer after the data contained (8) variables were the results of the analysis of the variables that have an impact on the phenomenon studied and is the property of the analysis of the variables that affect the effects of the phenomenon of the variables have an impact on the phenomenon studied.



2.1-Kaiser-Meyer-Olkin and Bartlett's Test

(SPSS₂₂)Output one shows several very important parts of the output: the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. The KMO statistic varies between 0 and 1. A value of 0 indicates that the sum of partial correlations is large relative to the sum of correlations, indicating diffusion in the pattern of correlations (hence, factor analysis is likely to be inappropriate). A value close to 1 indicates that patterns of correlation are relatively compact and so factor analysis should yield distinct and reliable factors. Kaiser (1974) recommends accepting values greater than 0.5 as acceptance (values below this should lead you to either collect more data or rethink which variables to include). Furthermore, values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are great and values above 0.9 are superb. For these data the value is 0.609, which falls into the

range of being mediocre: so, we should be confident that factor analysis is appropriate for these data

Table (10): KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.609
Bartlett's Test of Sphericity	Approx. Chi-Square	155.935
	Df	28
	Sig.	0.000

Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. For factor analysis to work we need some relationships between variables and if the R-matrix were an identity matrix then all correction coefficients would be zero. Therefore, we want this test to be significant (is to study. have a significance value less than 0.5). Significant test tells us that the R-matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis. For these data, Bartlett's test

2.2-Factor Extraction

Table (11): Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.106	26.322	26.322	2.106	26.322	26.322	1.712	21.400	21.400
2	1.390	17.372	43.694	1.390	17.372	43.694	1.653	20.665	42.065
3	1.227	15.342	59.036	1.227	15.342	59.036	1.358	16.971	59.036
4	0.896	11.201	70.237						
5	0.755	9.432	79.668						
6	0.680	8.504	88.172						
7	0.602	7.523	95.695						
8	0.344	4.305	100.000						

Extraction Method: Principal Component Analysis.

Table (11) lists the eigenvalues associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction, SPSS has identified 8 linear components within the data set (we know that there should be as many eigenvectors as there are variables and so there will be as many factors as variables). The eigenvalues associated with each factor shows the variance explained by that particular linear component and table also displays the eigenvalue in terms of the percentage of variance explained (so factor 1 explain 26.322% of total variance). It should be clear that the first few factors explain relatively large amount of variance. SPSS then extracts all factors with eigenvalues greater than 1, which means that there are (3) factors. The eigenvalues associated with these factors are again displayed (and the percentage of variance explained) in the columns labeled Extraction Sums of Squared Loadings. The values in this part of the table are the same as the values before extraction; expect that the values for the discarded factors are ignored (hence, the table is blank after the three factors). In final part of the table (labeled Rotation Sums of Squared Loadings), the eigenvalues of the factors after rotation are displayed. Rotation has the effect of optimizing the factor structure and one consequence for these data is that the relative importance of the three factors is equalized. Before rotation 26.322%, compared to 17.372 and 15.342%. However, after extraction it accounts for only 21.400%, compared to 20.665 and 16.971%.

2.3-Component Matrix

Table (12): Component Matrix

	Component		
	1	2	3
Gender	0.456	0.168	0.575
Age	-0.734	0.306	0.376
Governorate	-0.068	0.713	-0.140
Bllood group	-0.091	0.493	-0.627
Somken	0.633	0.405	0.160
Drinking	0.286	0.493	-0.088
Weight	-0.701	0.313	0.458
Employer(job)	0.611	0.104	0.315

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Table (12) shows the component matrix before rotation. This matrix contains the loadings of each variable onto each factor. By default SPSS displays all loadings; however, we requested that all loadings less than 0.4 be suppressed in the output and so there are blank spaces for many of loadings. This matrix is not particularly important for interpretation. At this stage SPSS has extracted three factors. Factor analysis is an exploratory tool and so it should be used to guide the researcher to make various decisions: you should not leave the computer to make them. One important decision is the number of factors to extract. By Kaiser's criterion we should extract three factors and this is what SPSS has done. However, this criterion is accurate when there are

less than 30 variables and communalities after extraction are greater than 0.7 or when the sample size exceeds 150 and the average communality is greater than 0.6. The communalities are shown in table (12), and exceed 0.7. The average of the communalities can be found by adding them up and dividing by the number of association. So on both ground Kaiser's rule may not be accurate. However, you should consider the huge sample that we have, because the research into Kaiser's criterion gives recommendations for much smaller samples. We can also use the screen plot, which we asked SPSS to produce. The screen plot is shown below with a thunderbolt indicating the point of inflexion on the curve. This curve is difficult to interpret because the curve begins to tail off after six factors, but there is another drop after three factors before a stable plateau is reached.

2.4: Rotated Component Solution

Table (13): Rotated Component Matrix

	Component		
	1	2	3
Gender	0.065	0.745	-0.090
Age	0.864	-0.161	0.036
Governorate	0.234	0.073	0.687
Blloodgroup	-0.113	-0.332	0.722
Somken	-0.215	0.661	0.328
Drinking	-0.074	0.284	0.497
Weight	0.890	-0.081	0.006
Employer(job)	-0.222	0.658	-0.011

*Extraction Method: Principal Component Analysis. *Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Table (13) shows the rotated component matrix (also called the rotated factor matrix in factor analysis) which is a matrix of the factor loadings for each variable onto each factor. This matrix contains the same information as the component matrix in Table (12) except that it is calculated after rotation. There are several things to consider about the format of this matrix. First, factor loadings less than 0.4 have not been displayed because we asked for these loadings to be suppressed. If you did not select this option, or did not adjust the criterion value to 0.4, then your output will differ. Second, the variables are listed in the order of size of their factor loadings because we asked for the output to be sorted by size. If this option was not selected your output will look different. Finally, for all other parts of the output I suppressed the variable labels (for reasons of space) but for this matrix I have allowed the variable labels to print to aid interpretation. Compare this matrix with the un-rotated solution. Before rotation, most variables loaded highly onto the first factor and the remaining factors did not really get a look in. However, the rotation of the factor structure has clarified things considerably: there are three factors and variables load very highly onto only one factor (with the exception of one question). The suppression of loadings less than 0.4 and ordering variables by loading size also makes



interpretation considerably easier (because you do not have to scan the matrix to identify substantive loadings).

Conclusion

Study has some potential limitations. The relatively small number of variables precludes potentially important subgroup analyses. Our study suggests no possible role for occupational exposure in the etiology of brain cancer. Increased risk of brain cancer is associated with being male, and there seems to be a tendency in the study toward an elevated risk for those with blood group A.

The study, body weight, Age, smoking and alcohol drinking could not be confirmed as strong risk factors for patients with brain cancer.

The analysis of study identified from factors which are clarify 59.036% of the variance, and the result identified three factors: 1st Factor: This factor containing these variables (Age, Weight) by total variance 21.400%. ($F1 = 0.864X2 + 0.890X7$), 2ndFactor: This factor containing these variables (Gender, Smoking and Job) by total variance 20.665%. ($F2 = 0.745X1 + 0.661X5 + 0.658X8$), 3rd Factor: This factor containing these variables (Governorate, Blood type and Drinking) by total variance 16.971%, ($F3 = 0.687X3 + 0.722X4 + 0.497X6$). The number of sample size and variables wasn't sufficient in this study for recommendations for the investigators that they should take more variables on the brain cancer to examine in the future. Finally, the rate of 59.036% aforementioned variables influence on the brain cancer.

Conflict of Interests.

There are non-conflicts of interest .

Reference

1. International Agency for Research on Cancer. [(accessed on 1 November 2018)]; Available online: <https://gco.iarc.fr/>
2. Brain Tumor Basics. [(accessed on 1 November 2018)]; Available online: <https://www.thebraintumourcharity.org/>
3. American Cancer Society website. [(accessed on 1 November 2018)]; Available online: www.cancer.org/cancer.html.
4. WHO Statistics on Brain Cancer. [(accessed on 1 November 2018)]; Available online: <http://www.who.int/cancer/en/>
5. Mahaley M.S., Jr., Mettlin C., Natarajan N., Laws E.R., Jr., Peace B.B. National survey of patterns of care for brain-tumor patients. J. Neurosurg. 1989;71:826–836. doi: 10.3171/jns.1989.71.6.0826.
6. Griffiths A.J.F., Wessler S.R., Lewontin R.C., Gelbart W.M., Suzuki D.T., Miller J.H. An Introduction to Genetic Analysis. Macmillan; New York, NY, USA: 2005.
7. Liao J.B. Cancer issue: Viruses and human cancer. Yale J. Biol. Med. 2006;79:115–122.
8. Golemis E.A., Scheet P., Beck T.N., Scolnick E.M., Hunter D.J., Hawk E., Hopkins N. Molecular mechanisms of the preventable causes of cancer in the United States. Genes

- Dev. 2018;32:868–902. doi: 10.1101/gad.314849.118
9. Mabray M.C., Barajas R.F., Cha S. Modern brain tumor imaging. Brain tumor research and treatment. 2015;3:8–23. doi: 10.14791/btrt.2015.3.1.8. 10. Sharma H., Alekseychuk A., Leskovsky P., Hellwich O., Anand R.S., Zerbe N., Hufnagl P. Determining similarity in histological images using graph-theoretic description and matching methods for content-based image retrieval in medical diagnostics. Diagn. Pathol. 2012;7:134. doi: 10.1186/1746-1596-7-134
- 11-Aalten, P., de Vugt, M.E., Lousberg, R., Korten, E., Jaspers, N., Senden, B., Jolles, J. and Verhey, F.R., 2003. Behavioral problems in dementia: a factor analysis of the neuropsychiatric inventory. Dementia and geriatric cognitive disorders, 15(2), pp.99-105.
- 12-Badaruddoza and Brar, S.K., 2015. Factor analysis of traditional cardiovascular risk traits in Punjabi adolescents in India. Egyptian Journal of Basic and Applied Sciences, 2(1), pp.13-18.
- 13-Borůvka, L. Vacek, O. and Jehlička, J. 2005. Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. Geoderma, 128(3-4), pp.289-300.
- 14-Jackson, E.S., Rodgers, N.H. and Rodgers, D.B., 2019. An Exploratory Factor Analysis of Action Responses to Stuttering Anticipation. Journal of fluency disorders.
- 15-Munir, R.I.S. and Rahman, R.A., 2016. Determining dimensions of job satisfaction using factor analysis. Procedia Economics and Finance, 37, pp.488-496.
- 16-Pison, G., Rousseeuw, P.J., Filzmoser, P. and Croux, C., 2003. Robust factor analysis. Journal of Multivariate Analysis, 84(1), pp.145-172.
- 17-Ro, Y. and Ha, Y., 2019. A factor analysis of consumer expectations for autonomous cars. Journal of Computer Information Systems, 59(1), pp.52-60.
- 18-Walter, J., Chesnaux, R., Gaboury, D. and Cloutier, V., 2019. Subsampling of Regional-Scale Database for improving Multivariate Analysis Interpretation of Groundwater Chemical Evolution and Ion Sources. Geosciences, 9(3), p.139

الخلاصة

ان الهدف من هذه الدراسة هو التحقق من أهم العوامل التي تؤثر على سرطان الدماغ وتحديد التركيبة السكانية لمرض سرطان الدماغ، ثم استخدام تحليل العوامل لتحديد العوامل التي تؤثر على سرطان الدماغ، وكانت طرق هذه الدراسة استخدمت الإحصائيات الوصفية وتحليل العوامل، في تحليل العوامل تم استخدام (FA) هو تقنية استكشافية تطبق على مجموعة من المتغيرات المرصودة التي تسعى إلى العثور على العوامل الأساسية (مجموعات فرعية من المتغيرات) التي تم إنشاء المتغيرات الملحوظة منها. علاوة على ذلك، تم أخذ البيانات المسجلة لسرطان الدماغ من مستشفى هيوفا في محافظة السليمانية خلال الفترة (2018 إلى 2020). تم تضمين 150 مريضاً و 8 متغيرات في الدراسة. بعد أن كشفت دراسة التحليل من العامل الذي يفسر 59.036% من التباين، وحددت النتيجة ثلاثة عوامل مثل: العامل الأول: هذا العامل الذي يحتوي على هذه المتغيرات (العمر، الوزن) حسب التباين الكلي 21.400%، العامل الثاني: هذا العامل الذي يحتوي على هذه المتغيرات (النوع والتدخين والمهنة) حسب التباين الكلي 20.665%، العامل الثالث: هذا العامل الذي يحتوي على المتغير (المحافظة ونوع الدم والشرب) حسب التباين الكلي 16.971%، يشمل الحد من الأوراق نقص المتغيرات المأخوذة لسرطان المخ وعدد العينات لم يكن كافياً. وأخيراً، تؤثر نسبة 59.036% من المتغيرات المذكورة على سرطان الدماغ

الكلمات الدالة: تحليل العوامل، مجموعات فرعية من المتغيرات، سرطان الدماغ، العوامل الكامنة.