# The Using of Statistical Analyses Methods to Study Some Affecting Factors Increasing the rate of Cancer Incidence

### Wafaa Muhie Eldeen Taha

Department of Mathematics, College of Education, University of Tikrit, Tilrit, Iraq

## Abstract:

This research was designed to study the benefit of using the statistical and standard methods in the medical field and the practical application of these methods in the analyses of the factors that leads to increase incidence of cancer.

The correlation analysis was used to analyze some variables like time periods, age group, and the incidence of cancer. On the other hand, Regression analysis was used to find the quantitative relationship between factors that affect it and also, the ANOVA and Duncan's tests. Finally we used the forecasting method to determine the expected future incidence of the disease till 2010.

#### Introduction:

Cancer is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. Most types of cancer cells from a mass called a tumor and named after the part of the body where the tumor originates. [1]

Cancer is caused by both external factors (tobacco, chemicals, radiation, poor nutrition and infectious organisms) and internal factors (inherited mutations, hormones, immune conditions, and mutation that occur from metabolism). These causal factors may act together or in sequence to initial or promote carcinogenesis. [2]

Ten or more years often pass between exposure to external factors and detectable cancer. Cancer is treated by surgery, radiation, chemotherapy, hormones and immunotherapy. The statistical studies of cancer provide usually health professional, researchers and policy-maker with detailed information regarding incidence and mortality of the most common types cancer by age, sex, time period and province or territory [5].

In this research, the statistical and standard measures including correlation analysis, regression analysis, analysis of variance, Duncan's test and forecasting with application where used of analysis the factors that affect the increasing incidence of cancer like age, time period, and which types of cancer that are more prevalence in studied period in Salahaddin in province. 1188 male patients reviewed in Tikrit teaching hospital from January 1995 to December 2005. the patients age range from 1 month to 100 years. They were then allocated to this age group:

- 1. 1 month 1 year
- 2. 1 year 20 year
- 3. 20 year 40 year
- 4. 40 year 60 year
- 5. 60 year 80 year
- 6. 80 year 100 year

The patients were diseased by about 18 type of cancers and the diagnosis confirmed by the specialist doctors in the hospital. The types include:

No.	Type of cancer
1	leukemia
2	lung cancer
3	bladder cancer
4	prostate cancer
5	bone cancer
6	brain cancer
7	stomach cancer
8	pancreatic cancer
9	bowel cancer
No.	Type of cancer
10	kidney cancer
10 11	kidney cancer liver cancer
10 11 12	kidney cancer liver cancer urinary tract cancer
10 11 12 13	kidney cancer liver cancer urinary tract cancer laryngeal cancer
10 11 12 13 14	kidney cancer liver cancer urinary tract cancer laryngeal cancer thyroid cancer
10 11 12 13 14 15	kidney cancer liver cancer urinary tract cancer laryngeal cancer thyroid cancer colonic cancer
10 11 12 13 14 15 16	kidney cancer liver cancer urinary tract cancer laryngeal cancer thyroid cancer colonic cancer lymphoma cancer
10 11 12 13 14 15 16 17	kidney cancer liver cancer urinary tract cancer laryngeal cancer thyroid cancer colonic cancer lymphoma cancer intestinal cancer

## 1. Theoretical Part

## 1.1. Correlation analysis:

The Pearson's simple correlation coefficient for the time, age group, type of cancer, and the incidence of cancer were calculated. The significance of correlation was determined and the results were verified in Table (1) with signal (\*) or (\*\*) referring to the significance of correlation coefficient. The former refers to (0.05) and latter to (0.01) significance level. [3].

#### 1.2. Regression analysis:

This type of statistical used to evaluate the quantitative relationship between the incidence of cancer as a dependent factor and other factors like time, and age group. Because the relation between the independent variables and dependent variables were not determined previously, the estimation of four models to this relation was introduced (linear model, normal complete logarithmic model, normal semi-logarithmic lift side model and normal logarithmic right side model). The appropriate model was selected among the estimated model according to standard and statistical measures. They effect of each factors on the incidence of disease was estimated and the "F test" was used to estimated the total significance as shown in table (2). [4].

#### 1.3. Analysis of variance (ANOVA):

This type of statistical or standard analyses used to determine the significant effects of the cancer incidence according to the age groups, type of cancer, and time. The ANOVA test enhances the regression analysis, but does not substitute it. This is because the regression analysis estimate the relationship between variables in a quantitative manner, where as the ANOVA searches for the presence of significant differences in the dependent variable values as a result of differences in the independent variables levels. the results were shown in t-able (3). [6]

## 1.4. Duncan's Test

Because of the usage of analysis of variance, the needed statistical analysis was performed to confirm which of the variable group were different from the others. For this reason we used the Duncan's test which could complete the ANOVA and classify the variables as similar and different groups significantly. [7].

## 1.5. Forecasting

Today the ability for forecasting became easy with the use of different forecasting methods. In this research the regression equation for the incidence of the cancer by time period was used and this equation used for estimating the average incidence of the disease from 2006 to 2010.

$$\mathbf{M} = \boldsymbol{\alpha} + \boldsymbol{\beta} \mathbf{A} + \boldsymbol{\gamma} \mathbf{T}$$

Where

M = the incidence of cancer T = time

A = age group

## 2. Application Part

## 2.1. Description of data

In this research use the spss statistical program to find the results and graph the figures. First age group (1 month - 1 years) show that the leukemia is more increased than the other types of cancer through the study period and the highest increasing is in years 1995, 1998, 2002, 2005 as shown in Fig.(1).

Second age group (1 years -20 years) show increasing the types leukemia, brain and bone in comparison with other cancer types as shown in Fig.(2).

Third age group (20 years -40 years) reveal that leukemia, lung, intestinal and urinary are increased in this age group as shown in Fig.(3).

Fourth age group (40 years - 60 years) show the increasing of most type of cancer but the lung and leukemia types reveal the higher rate of incidence especially in 1995 as shown in Fig.(4).

Fifth age group (60 years - 80 years) reveal an increasing in a wide type of cancer and the highest types are leukemia, bladder, liver, lung and prostatic as shown in Fig.(5).

Sixth age group (80 years -100 years) show increasing of all the types of cancer as shown in Fig.(6).



Fig. (1): First Age Group



Fig. (3): Third Age Group





Fig. (6): Sixth Age Group

#### 2.2. Results of correlation coefficient

The reveal to Table(1) that there is significant correlation between the incidence of cancer and age group with significance level of (0.01).

Table (1)					
Deerson		Age	incidence		
correlation	Age		0.253**		
	Incidence				

Time = Time period (1995 - 2005)
----------------------------------

Incidence = the incidence of cancer during the last period. Age = age group.

#### 2.3. Result of regression analysis:

Table (2) reveal that the best model was estimated to show the relation between cancer incidence rate as dependent variable(M) and between age ranges(A), time period(T) as independent variables which affect the dependent variable (first model) mean the linear model  $M = \alpha + \beta A + \gamma T$  and the estimated function as: 
$$\begin{split} M &= -0.0753 + 0.25A + 0.008721T\\ SE &= 0.141 \quad 0.028 \quad 0.015\\ T &= -0.407 \quad 9.014^* \quad 0.581\\ R &= 0.254\\ R^2 &= 0.064 \end{split}$$

```
F = 40.794*
```

The estimated model obey to standard tests like t-test fro estimated parameters at significance level 0.05 as in Table (2).

F-test was used for estimating the total significance model. The analysis of variance for regression produced that F value is significant under total level 0.001. This value supports the model efficacy for analyzing and forecasting as in Table (2).

The equation of estimated forecasting produced that they incidence of cancer increased by about 0.25 with increase of the age group one step, and also increased by 0.008721 with increase in time one years.

	(1)	n ·	CC <sup>1</sup> 1		1.	1		C	•	
1 9 NIA	1710	Rogroccion	coatticiant	Actimatod	according	r to tha	tuno	of roor	acción /	annation
I ADIC	\ <i>4</i> 1.	RUEIUSSIUII	COULICICIU	Usumateu	according	2 10 100		ULICEL	CSSIOII 1	Judation.
	× .				···· · · · · · · · · · · · · · · · · ·			0		

		0	71	0		
Madal		Coff.		Tests		
Model	α	β	γ	R	$\mathbf{R}^2$	F
$M = \alpha + \beta A + \gamma T$	-0.0573	0.250	0.008721	0.254	0.064	40.794*
T test	-0.407	9.014*	0.581	0.030	0.001	0.0194
$\ln M = \alpha + \beta A + \gamma T$	0.588	0.01134	0.002759			
T test	5.291*	0.556	0.266	0.094	0.007	1.517
$\ln M = \alpha + \beta A + \gamma \ln T$	0.725	0.01218	-0.0758	0.084		
T test	6.147*	0.599	-1.648			
$M = \alpha + \beta A + \gamma \ln T$	0.05185	0.250	-0.0357	0.254	54 0.064	40.794*
T test	0.342	9.014*	-0.535			

(\*) Significance range 0.05.

#### 2.4. Results of ANOVA:

The variance of the cancer incidence data according to the age groups, years, and cancer type were analyzed. The results are shown in the Table (3).

<b>T</b>	( <b>n</b> )		<b>TT</b> 1	1	•	c	•
Table (	- 1	۱.	The	analy	VSIS	ot.	variance
I UDIC (	· •	, •	1110	unui	y 010	O1	variance

Model	Ss	Df	MS	F	Sig.
Age	230.596	5	46.119	21.107	0.000
Time	48.052	10	4.805	2.199	0.016
Error	2521.516	1154			
Total	4289.090	1188			

From the above table we concluded that when performing the ANOVA test for incidence of cancer in male patients according to the type, age group, and years, there were significant differences between the mean incidence and the age group, years, and type of cancer with significance level below 0.05.

#### 2.5. Duncan's Test Results:

The results of Duncan's test were allocated according to the type of cancer, the age group and the years, as in table (4), (5), and (6).

Table (4): Duncan's test results for the incidence by age

group.					
Incidence	Sig				
0.3283	а				
0.3939	ab				
0.6566	b				
1.0621	с				
1.3131	cd				
1.4747	d				
	Incidence           0.3283           0.3939           0.6566           1.0621           1.3131           1.4747				

 Table (5): Duncan's test results for the incidence by time

	(Years).	
Time	Incidence	Sig.
1997	0.6111	а
1998	0.6574	а
2002	0.7157	а
1999	0.7315	а
1996	0.8145	ab
2001	0.8148	ab
2003	0.8889	ab
2004	0.8889	ab
2000	1.000	ab
2005	1.1944	b
1995	1.2685	b

 Table (6): Duncan's test results for the incidence by the

type of cancer					
Kind	Incidence	Sig.			
0	0.000	а			
18	0.1667	ab			
16	0.3182	ab			
14	0.3485	ab			
9	0.4848	abc			
7	0.5000	abc			
8	0.5333	abcd			
13	0.5455	abcd			
11	0.5606	abcd			
15	0.6061	abcd			
12	0.6971	abcd			
17	0.7424	abcd			
10	0.7576	bcd			
5	0.9091	cd			
4	1.1364	cde			
3	1.1561	cde			
6	1.2576	de			
2	1.6818	e			
1	3.3333	f			

When Duncan's test performed according to the age groups, a significant differences were noticed in the incidence of cancer in the age group 4 as compared with other age group except 5.

Also we noticed this result between age group 6 and all other age group except 5, and also between age group 3 with other age groups except 1.

It is clear that the incidence of the disease in the advanced age groups (4, 5, 6) was more than that seen with younger age groups (1, 2, 3) and this difference was significant.

When Duncan's test performed to estimate the significant differences in the cancer incidence according to the time (years of the study), we found that the earliest 1 year (1995) and the last 1 year (2005) were similar to each other but differ from other years. The is a vague condition because the incidence of the disease started high at 1995 then decline throughout the years of the study then increases again at 2005. among the other years from 1996 to 2004 there were no significant differences in the incidence of cancer.

It is clear from the above results that the cancer type 1 was significantly different from the other types and that it showed a higher incidence than others.

Type 2 diseases were also significantly different from the other types except type 6, 3, and 4. There were no other significant differences among other types of cancer to be mentioned.

#### 2.6. Forecasting:

The incidence of cancer forecasting from 2006 to 2010 performed by using the following forecasting equation:

M = -0.0573 + 0.025A + 0.008721T

For age group and according to the Table (7). Fig.(7) Show an increasing in the incidence of cancer rate

especially in the oldest age groups.



Figure (7): Estimating incidence for different age's group.

#### **Conclusion:**

In this research the statistical, standard method and their clinical applications were used in the analysis of the factors leading to increase the incidence of cancer diseases. There are significant relationships at value (P < 0.01) between the incidence of the disease and the age groups by using correlation analysis. The use of regression analysis relationship demonstrate a significant relationship according to F-Test.

The application of ANOVA analysis demonstrate that there were significant differences (P < 0.05) in the incidence of disease in male patient according to other listed factors (age group, time periodic and type of cancer).

The enhancement of ANOVA analysis results by the use of Duncan's test found that the incidence of the disease among the advanced age group (from 40 years – 100 years) was significantly higher then the incidence among the younger age group from (1 month – 40 years).

Duncan's test found also that the incidence of all types of cancer was higher in 1995 and in 2005 and decline in the period between the years.

According to the Duncan's test results the type Leukemia was significantly different from other types of the cancer and showed higher incidence than others.

The estimation of the incidence of cancer till 2010 by the use of regression equation demonstrates that there is an increase in the incidence with the increase of years.

#### References

- 1. American Cancer Society, 2005, web sit www.cancer.org
- 2. American Cancer Society, 2006, web sit www.cancer.org
- Campbell R. McConnel and Harish C.gupta, (1977), Economics Principle, Problems, and Policies, Tata, McGraw- Hill Publishing Company, New Delhi.
- 4. George W. Sendecor and William G. Cochran, (1978), statistical Methods, AMES, IOWA, U.S.A.
- 5. Heart Disease and Stroke Statistics 2006 Update, American Heart Association

6. Johnston J., (1984), Econometric Methods, Third Edition, New York. 7. IMF. International Financial Statistics yearbook, 1998. Washington, DC: International Monetary Found.

# استخدام طرق التحليل الاحصائي لدراسة بعض العوامل المؤثرة في زيادة معدل الاصابة في السرطان

# وفاء محي الدين طه

## قسم الرياضيات، كلية التربية، جامعة تكريت، تكريت، جمهورية العراق

## الملخص:

تم في هذا البحث دراسة الفائدة المستنبطة من استخدام الادوات الاحصائية والقياسية في العلوم الطبية مع التطبيق العملي في تحليل العوامل المؤثرة في زيادة نسبة الاصابة بامراض السرطان اذ تم استخدام اسلوب تحليل الارتباط بين متغيرات الزمن والفئات العمرية ثم معدل الاصابة.

استخدم اسلوب تحليل الانحدار لايجاد العلاقة الكمية بين معدل الاصابة بمرض السرطان كعامل معتمد وبين عدة عوامل مؤثرة عليه. ثم استخدمت طريقة تحليل التباين واختبار دنكن لتعزيز اسلوب تحليل التباين تم التنبؤ بالقيم المستقبلية لمعدل الاصابة لغاية سنة ٢٠١٠ .