On Multiple Regression Modeling of Breast Cancer For the period 2005-2012 Using SPSS Program

Hadeel Salim AL-Kutubi

Informatics Center for Research & Rehabilitation, Kufa University, Iraq

Abstract:

The main propose of this study is to evaluate the Breast Cancer in AL-Najaf city in Iraq for the period 2005-2012 for three variables that is age, weight and Tumor level. The data is analyzed using multiple regression analysis to get the mathematical model, the age, weight effect on tumor level and to get the positive or negative relationship between variables. SPSS program V. 17.0 was used throughout this study to analyzed the data and to generate the various Tables .

From the results, we find the age and weight effect on tumor level is very weak. So there is no significant effect between age , weight and tumors level. Also , from all regression lines, there is a negative relationship between Y (Tumor level) and X_1 (Age) in 2005,2006, 2007, 2008, 2009, 2010. But there is a positive relationship between Y (Tumor level) and X_1 (Age) in 2011 and 2012. Also there is a negative relationship between Y (Tumor level) and X_2 (Weight) in 2005, 2011, 2012. And there is a positive relationship between Y (Tumor level) and X_2 (Weight) in 2006, 2007, 2008, 2009, and 2010.

Key words: Multiple Regression , Breast Cancer (2005-2012) , SPSS program.

1.Introduction

The statistical technique of extending simple linear regression so that it considers two or more independent variables is called multiple linear regression. With the development of electronic computers, some times as many as fifteen independent variables are considered. It is quite obvious that multiple linear regression is a very useful technique and it is very frequently used in business and economic problems. We shall mainly consider linear regression which is of the form:[7]

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

In this study, we evaluate the Breast Cancer in Al-Najaf City in Iraq for the period 2005-2012 and for three variables that is age, weight and

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Tumor. The data is analyzed using multiple regression. SPSS program V. 17.0 was used throughout this study to analyzed the data and to generate the various Tables .

2.Materials and Methods

In this part of study we present the theoretical method of multiple regression. We depend on Douglas Montgomery [4] to explain the multiple regression method .

The test for significance of regression is a test to determine if there is a linear relationship between the response variable y and a subset of the variables $X_1, X_2, \dots, X_k$. The appropriate hypotheses are

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$H_1: \beta_j \neq 0$ for at least one j

Rejection of H_0 implies that at least one of the variables $X_1, X_2, \dots, X_k$

contributes significantly to the model. The test procedure involves an analysis of variance partitioning of the total sum of square SS_T into a sum of squares due to the model (or to regression) and a sum of square due to residual (or error), say $SS_T = SS_R + SS_E$. Now if the null hypothesis

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is true, then $\frac{SS_R}{\sigma^2}$ is distributed as ${\chi_k}^2$

Where the number of degrees of freedom for χ^2 are equal to the number of variables in the model. The test procedure for $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is to compute

$$F_0 = \frac{MS_R}{MS_E}$$

Alternatively, we could use the p value approach to hypothesis testing And, thus reject H_0 if the p value for the statistic F_0 is less than α . the test is usually summarized in an analysis of variance table such as:

Table A. Analysis of variance for regression model

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Source of variation	Sum of squares	Degrees of freedom	Mean square	F_0
Regression	SS_R	K	MS_R	$\frac{MS_R}{MS_E}$
Error or residual	SS_E	n-k-1	MS_E	
Total	SS_T	n-1		

3.Results and Discussion

In this study , we find regression models for each year in the period 2005-2012, by using SPSS program with depend on three variables age, weight and tumor. Also find the positive and negative relationship between these variables.

Table 1. Regression Model Summary for Age, Weight and Tumor in 2005

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.710[a]	0.503	0.338	1.05914

a. Predictors: (constant), Weight, Age

In this Table, we can see the R^2=50%. This means that, the age and weight effect on tumor level is 50 %. This effect is not strong.

Table 2. Analysis of Variance for Age, Weight and Tumor in 2005

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	6.825	2	3.412	3.042	0.122[a]
Residual	6.731	6	1.122		
Total	13.556	8			

a. Predictors: (constant), Weight, Age b. Dependent variable: Tumor

Table 3. Regression Coefficients for Age, weight and Tumor in 2005

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	7.068	1.712		4.129	0.006
Age	-0.031	0.050	-0.271	0.614	0.562
Weight	-0.042	0.038	-0.482	1.091	0.317

a. Dependent variable : Tumor

The regression line is $\hat{Y} = 7.068 - 0.031\,X_1 - 0.042X_2$. That is means, there is negative relationship between Y (Tumor) , X_1 (Age) and X_2 (Weight).

Table 4. Regression Model Summary for Age, Weight and Tumor in 2006

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.222[a]	0.049	0.040	1.04130

a. Predictors: (constant), Weight, Age

In this Table, we can see the R^2=4%. This means that, the age and weight effect on tumor level is 4 %. This effect is very week.

Table 5. Analysis of Variance for Age, Weight and Tumor in 2006

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	11.999	2	6.000	5.533	0.005[a]
Residual	230.959	213	1.084		
Total	242.958	215			

a:Predictors: (constant), Weight, Age b. Dependent variable: Tumor

Table 6. Regression Coefficients for Age, weight and Tumor in 2006

Model	Unstandardized Coefficients	Standardized Coefficients	T	Sig.

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

	B	Std. Error	Beta		
Constant	3.755	0.453		8.285	0.000
Age	-0.020	0.006	-0.224	3.326	0.001
Weight	0.003	0.006	0.033	0.496	0.620

a. Dependent variable : Tumor

The regression line is $\hat{Y} = 3.755 - 0.020\ X_1 + 0.003X_2$. That is means, there is a negative relationship between Y (Tumor) and X_1 (Age). And a positive relationship between Y (Tumor) and X_2 (Weight).

Table 7. Regression Model Summary for Age, Weight and Tumor in 2007

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.219[a]	0.048	0.039	0.81361

a. Predictors: (constant), Weight, Age

In this Table, we can see the R^2=4.8%. This means that, the age and weight effect on tumor level is 4.8 %. This effect is very week.

Table 8. Analysis of Variance for Age, Weight and Tumor in 2007

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	6.967	2	3.484	5.263	0.006[a]
Residual	137.687	208	0.662		
Total	144.654	210			

a.Predictors: (constant), Weight, Age b.Dependent variable: Tumor

Table 9. Regression Coefficients for Age, weight and Tumor in 2007

Model	Unstandardized Coefficients	Standardized Coefficients	T	Sig.

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

	B	Std. Error	Beta		
Constant	3.913	0.278		14.061	0.000
Age	-0.017	0.005	-0.231	3.183	0.002
Weight	0.002	0.004	0.040	0.551	0.582

a.Dependent variable : Tumor

The regression line is $\hat{Y} = 3.913 - 0.017\ X_1 + 0.002X_2$. That is means, there is negative relationship between Y (Tumor) and X_1 (Age). And a positive relationship between Y (Tumor) and X_2 (Weight).

Table 10. Regression Model Summary for Age, Weight and Tumor in 2008

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.271[a]	0.073	0.066	0.96993

a.Predictors: (constant), Weight, Age

In this Table, we can see the R^2=7.3%. This means that, the age and weight effect on tumor level is 7.3 %. This effect is very week.

Table 11. Analysis of Variance for Age, Weight and Tumor in 2008

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	18.724	2	9.362	9.952	0.000[a]
Residual	236.130	251	0.941		
Total	254.854	253			

a.Predictors: (constant), Weight, Age b.Dependent variable: Tumor

Table 12. Regression Coefficients for Age, weight and Tumor in 2008

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	3.048	0.307		9.943	0.000
Age	-0.021	0.005	-0.241	3.776	0.000
Weight	0.014	0.004	0.218	3.417	0.001

a.Dependent variable : Tumor

The regression line is $\hat{Y} = 3.048 - 0.021X_1 + 0.014X_2$. That is means, there is a negative relationship between Y (Tumor) and X_1 (Age). And a positive relationship between Y (Tumor) and X_2 (Weight).

Table 13. Regression Model Summary for Age, Weight and Tumor in 2009

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.110[a]	0.012	0.003	0.93812

a.Predictors: (constant), Age

In this Table, we can see the R^2=1.2%. This means that, the age and weight effect on tumor level is 1.2 %. This effect is very week.

Table 14. Analysis of Variance for Age, Weight and Tumor in 2009

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.249	2	1.124	1.2777	0.281[a]
Residual	182.175	207	0.880		
Total	184.424	209			

a.Predictors: (constant), Weight, Age b.Dependent variable: Tumor

Table 15. Regression Coefficients for Age, weight and Tumor in 2009

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	3.355	0.283		11.850	0.000
Age	-0.009	0.006	-0.119	1.565	0.119
Weight	0.001	0.004	0.027	0.352	0.725

a.Dependent variable : Tumor

The regression line is $\hat{Y} = 3.355 - 0.009X_1 + 0.001X_2$. That is means, there is a negative relationship between Y (Tumor) and X_1 (Age). And a positive relationship between Y (Tumor) and X_2 (Weight).

Table 16. Regression Model Summary for Age, Weight and Tumor in

Created using
nitroPDF professional
download the free trial online at nitropdf.com/professional

2010

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.090[a]	0.008	0.002	1.02008

a.Predictors: (constant),Weight, Age

In this Table, the age and weight effect on tumor level is very week.

Table 17. Analysis of Variance for Age, Weight and Tumor in 2010

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.896	2	1.448	1.391	0.250[a]
Residual	356.911	343	1.041		
Total	359.806	345			

a.Predictors: (constant), weight, Age b.Dependent variable: Tumor

Table 18. Regression Coefficients for Age, weight and Tumor in 2010

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	3.762	0.320		11.747	0.000
Age	-0.007	0.004	-0.090	1.666	0.097
Weight	0.000	0.004	-0.006	0.103	0.918

a.Dependent variable : Tumor

The regression line is $\hat{Y} = 3.762 - 0.007X_1 + 0.000X_2$. That is means, there is negative relationship between Y (Tumor) and X_1 (Age).

Table 19. Regression Model Summary for Age, Weight and Tumor in 2011

nitroPDF professional
download the free trial online at nitropdf.com/professional

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.197[a]	0.039	0.031	0.98994

a.Predictors: (constant), Weight, Age

In this Table, we can see the R^2=3.9%. This means that, the age effect on tumor level is 3.9 %. This effect is very week.

Table 20. Analysis of Variance for Age, Weight and Tumor in 2011

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	9.839	2	4.919	5.020	0.007[a]
Residual	244.015	249	0.980		
Total	253.853	251			

a.Predictors: (constant), Weight, Ageb.Dependent variable: Tumor

Table 21. Regression Coefficients for Age, weight and Tumor in 2011

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	3.757	0.495		7.589	0.000
Age	0.011	0.004	0.163	2.621	0.009
Weight	-0.014	0.007	-1.912	1.912	0.057

a.Dependent variable : Tumo

The regression line is $\hat{Y} = 3.757 + 0.011X_1 - 0.014X_2$. That is means, there is positive relationship between Y (Tumor) and X_1 (Age). And a negative relationship between Y (Tumor) and X_2 (Weight).

Table 22. Regression Model Summary for Age, Weight and Tumor in 2012

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.172[a]	0.030	-0.024	1.15790

a.Predictors: (constant), Weight, Age

In this Table, we can see the R^2=3%. This means that, the age and weight effect on tumor level is 3 %. This effect is very week.

Table 23. Analysis of Variance for Age, Weight and Tumor in 2012

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.477	2	0.738	0.551	0.581[a]
Residual	48.267	36	1.341		
Total	49.744	38			

a.Predictors: (constant), Age, weightb.Dependent variable: Tumor

Table 24. Regression Coefficients for Age, weight and Tumor in 2012

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	4.627	1.767		2.618	0.013
Age	0.003	0.012	0.042	0.250	0.804
Weight	-0.024	0.024	-0.161	0.969	0.339

a.Dependent variable : Tumor

The regression line is $\hat{Y} = 4.627 + 0.003X_1 - 0.024X_2$. That is means, there is a positive relationship between Y (Tumor) and X_1 (Age). And a negative relationship between Y (Tumor) and X_2 (weight).

4.Conclusion

Regression Models for Age, weight and Tumors in the years 2005, 2006, 2007, 2008, 2009,2010, 2011 and 2012 , shows, the values of R^2 present the age and weight effect on tumor level is very week. So there is no significant effect between age , weight and tumors level.

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

Also , from all regression lines, there is a negative relationship between Y (Tumor level) and X_1 (Age) in 2005,2006, 2007, 2008, 2009, 2010. But there is a positive relationship between Y (Tumor level) and X_1 (Age) in 2011 and 2012. Also there is a negative relationship between Y (Tumor level) and X_2 (Weight) in 2005, 2011, 2012. And there is a positive relationship between Y (Tumor level) and X_2 (Weight) in 2006, 2007, 2008, 2009, and 2010.

5.References

1. Alkutubi H.S, 2005. Evaluation of cancer disease for the period (1995-2005) in Tikrit teaching hospital, Tikrit journal of pharmaceutical science,vol.1,no.2.
2. Bland Martin, 2000. An introduction to medical statistics, 3th ed. Oxford University Press.
3. Lindgren B.W. 1968. Statistical theory, 2nd ed., New York, Macmillan companies.
4. Montgomery D.C., 2001. Design and analysis of experiments.5th edition.John Wiley & Sons, Inc.
5. Mood A.M. and Graybill F.A. 1974. Introduction to the theory of statistics, 3th ed. MCGraw- Hill kogakusha, Lth,.
6. Taylor J. K. and Cihon C., 2004. Statistical Techniques for data analysis. 2nd ed..Chapman & Hall / CRC.
7. Yamane T. 1973. Statistic , An introductory analysis. 3th edition. Harper international edition. New York.

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

حول موديل الانحدار المتعدد لسرطان الثدي للفترة 2005-2012

باستخدام برنامج إحصائي

هديل سليم الكتبي

مركز البحث والتأهيل ألمعلوماتي، جامعة الكوفة، العراق

الملخص

الهدف الرئيسي من هذه الدراسة هو تقييم أعداد الإصابات **بسرطان الثدي للفترة** 2005-2012 في مدينة الصدر الطبية بمحافظة النجف بالاعتماد على ثلاث متغيرات العمر، الوزن ومستوى الورم **(مستوى السرطان بالجسم)**. تم تحليل البيانات بالاعتماد على تحليل الانحدار المتعدد للحصول على الموديل الرياضي للانحدار ومعرفة تأثير العمر والوزن **على مستوى السرطان (المستوى الأول ، الثاني، الثالث أو الرابع) بالجسم** . أيضا إيجاد العلاقة الموجبة والسالبة بين المتغيرات الثلاث. للحصول على هذه الأهداف تم استخدام برنامج إحصائي لتحليل البيانات **وللخروج بالنتائج المطلوبة**.

استنتجنا من التحليل الإحصائي إن تأثير العمر والوزن على مستوى السرطان بالجسم ضعيف جدا وأيضا لا توجد علاقة معنوية بين العمر والوزن ومستوى السرطان بالجسم. **خطوط** الانحدار للسنين المختلفة أعطتنا علاقة سالبة بين مستوى السرطان بالجسم وعمر الشخص المصاب وللسنين التالية: 2005، 2006، 2007، 2008، 2009 **و** 2010. لكن هناك علاقة موجبة بين مستوى السرطان بالجسم وعمر الشخص المصاب للسنين التالية 2011 **و** 2012. أيضا هناك علاقة سالبة بين مستوى السرطان ووزن الشخص المصاب للسنين التالية: 2005،

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional

2011و 2012 . فضلا عن وجود علاقة موجبة بين مستوى السرطان بالجسم ووزن الشخص المصاب للسنين التالية : 2006، 2007، 2008، 2009 و 2010.

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional