American Journal of Applied Sciences 6 (8): 1580-1585, 2009
ISSN 1546-9239

On Regression Modeling of Human Immunodeficiency Virus Patients

Hadeel Salim AL-Kutubi
Institute for Mathematical Research, University Putra Malaysia,
43400, UPM, Serdang, Selangor, Malaysia

Abstract: Problem statement: The main propose of this study was to evaluate the HIV patients for the period 1990-2008 depend on three variables age, gender and ethnicity. **Approach:** The data was analyzed using regression and correlation methods to get the mathematical model that explain the relationship and the effect between the age, gender and ethnicity. SSPS program V. 17.0 was used throughout this study to analyze the data and to generate the various Tables. **Results:** Using SPSS program to obtain regression models for each year in the period 1990-2008 depend on three variables age group, gender and ethnicity. Also obtained the relationship between all three variables in HIV patients using correlation methods. **Conclusion:** The age effect on gender and ethnicity in three years 1991, 2001 and 2002 are stronger than other years. In regression models, there exist significance effect between age and gender in two models, but there is no significance effect between age and ethnicity in all models. In correlation, there is no significance relationship between age and gender, age and ethnicity, ethnicity and genders in all years from 1990-2008.

Key words: linear regression, correlation coefficient, HIV patients, SPSS program

INTRODUCTION

In many problems there are two or more variables that are related and it is of interest to model and explore this relationship.

Suppose that there is a single dependent variable or response y that depend on k independent for example $x_1 \ldots, x_n$. The relationship between these variables is characterized by a mathematical model called a regression model[2].

In regression analysis, the age effect on gender and ethnicity with mathematical model that explain the significance relationship between all variable was presented. Also, the relationship between all variables (age group, gender and ethnicity) in HIV patients was presented in correlation analysis.

This study consist of 1434 cases from 1990 to 2008 taken from one hospital in Malaysia. The SSPS program V. 17.0 was used throughout this study to analyze the data and to generate the various tables.

MATERIALS AND METHODS

Linear regression: The statistical procedure for finding this best fitting line is called the method of least squares and the line is called the regression line[2]. The formal derivation of this procedure, which requires differential calculus, is presented in advanced statistical texts.

First, it is necessary to introduce some useful notation:

(X_i, Y_i) = ith pair of observations (1)

$$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{n} = \sum xy \quad (2)$$

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = \sum y^2 \quad (3)$$

$$\sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n} = \sum x^2 \quad (4)$$

The sample regression line is written $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ where the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are:

$$\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} \text{ and } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

The values $\hat{\beta}_0$ and $\hat{\beta}_1$ are calculated from a sample of observations from the entire population of interest and are estimates of the "true" population values" β_0 and β_1. As was the case with $\bar{Y}$ and s, the values $\hat{\beta}_0$ and $\hat{\beta}_1$ are subject to sampling variation and therefore may

vary from sample to sample. The value $\hat{Y}$ obtained for a given X is the predicted mean of the population of all possible Y values that could occur at the given value X. Just as there is a sample standard deviation associated with each $\bar{Y}$, there is a standard deviation associated with the regression line and $\hat{Y}$ This quantity, denoted by s_{yx} to signify regression, is called the standard error of the estimate it is given by:

$$s_{y.x} = \sqrt{SSE / (n-2)}$$

where, n is the number of pairs of observations and Sum of Squares for Error (SSE) is defined as:

$$SSE = \sum (Y - \hat{Y})^2$$

The Standard Error (SE) for $\hat{Y}$ at a given X value would be:

$$SE(\hat{Y}) = s_{y.x}\sqrt{\frac{1}{n} + (X - \bar{X})^2 / \sum x^2}$$

And

$$SSE = \sum (Y - \hat{Y})^2$$

Correlation coefficient: The most widely used measure of this degree of association between Y and X is provided by r, the coefficient of correlation. The formula for r[1] is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

The values of r lie in the interval $-1 \leq r \leq +1$ with a "large" value of r (either positive or negative) indicating a strong relationship between X and Y. A negative value of r indicates that high X values are associated with low Y values, or, low X values associated with high Y values. A positive r, on the other hand, indicates that high values of X are associated with high values of Y and low values of x are associated with low values of Y. A further explanation of r may be seen by comparing it with $\hat{\beta}_1$, the slope of the regression line. In the formulas for r and $\hat{\beta}_1$, numerators are identical (the denominators for both will always be positive); therefore, r and $\hat{\beta}_1$ will have the same sign. When the slope of the line is negative, the correlation is also negative thus indicating a negative, or inverse relationship between Y and X. Similarly, a positive slope and a positive correlation indicate direct relationship between variables. Further, if an exact positive relationship exists between Y and X (i.e., all points lie exactly on the regression line), then the value of r is +1. An exact negative relationship will yield an r of -1.

When $\hat{\beta}_1 = 0$, r = 0 and hence no linear relationship between Y and X is indicated. As was the case with $\hat{\beta}_1$, the value r is the sample estimate of a true population correlation value denoted by ρ and is subject to sampling variation. It is of interest therefore to test the hypothesis that the true population correlation equals zero. A value of ρ = 0 indicates that there is no linear association between the variables under study. The test statistic for testing $H_0 : \rho = 0$ is:

$$t = r\sqrt{\frac{n-2}{1-r^2}},\ \text{n-2 degrees of freedom}$$

RESULTS

Using SPSS program, we get (Table 1-20) explains regression models for each year in the period 1990-2008 depend on three variables age, gender and ethnicity.

Regression analysis:

Correlation coefficient: Also using SPSS program, we found positive and negative relationship between all three variables age, gender and ethnicity in HIV patients in (Table 21-39) using correlation methods.

Table 1: Regression models summary

Model	R	R^2	Adjusted R^2	Std. error
1990	0.172	0.030	-0.456	8.42897
1991	0.397	0.158	0.052	8.42101
1992	0.187	0.035	-0.079	15.89379
1993	0.257	0.066	-0.006	7.76293
1994	0.259	0.067	0.018	7.30564
1995	0.055	0.003	-0.048	12.25415
1996	0.034	0.001	-0.031	9.52153
1997	0.072	0.005	-0.014	11.35318
1998	0.193	0.037	0.017	10.66039
1999	0.283	0.080	0.063	8.65190
2000	0.207	0.043	0.024	12.18368
2001	0.449	0.201	0.179	10.77842
2002	0.477	0.227	0.131	9.48309
2003	0.160	0.026	-0.020	11.41613
2004	0.140	0.020	-0.024	13.70288
2005	0.205	0.042	0.018	11.86708
2006	0.093	0.009	-0.007	11.13336
2007	0.159	0.025	0.012	10.35178
2008	0.156	0.024	0.015	11.05048

Table 2: Regression summary for dependent variable (age in 1990)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	53.048	4.113		12.898	0.000
Gender	0.619	9.379	0.033	0.066	0.951
Ethnicity	1.429	4.505	0.161	0.317	0.767

$\hat{y}_1 = 53.048 + 0.619x_1 + 1.429x_2$

Table 3: Regression summary for dependent variable (age in 1991)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	43.284	2.779		15.575	0.000
Gender	6.868	4.838	0.333	1.420	0.175
Ethnicity	1.479	2.164	0.160	0.683	0.504

$\hat{y}_2 = 43.284 + 6.868x_1 + 1.479x_2$

Table 4: Regression summary for dependent variable (age in 1992)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	44.892	6.913		6.494	0.000
Gender	2.108	17.332	0.031	0.122	0.905
Ethnicity	3.586	4.650	0.195	0.771	0.451

$\hat{y}_3 = 44.892 + 2.108x_1 + 3.586x_2$

Table 5: Regression summary for dependent variable (age in 1993)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	48.460	2.372		20.433	0.000
Gender	11.380	8.409	0.273	1.353	0.188
Ethnicity	0.960	2.372	0.082	0.405	0.689

$\hat{y}_4 = 48.46 + 11.38x_1 + 0.96x_2$

Table 6: Regression summary for dependent variable (age in 1994)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	45.176	1.744		25.910	0.000
Gender	2.176	7.511	0.460	0.290	0.774
Ethnicity	2.189	1.412	0.247	1.550	0.129

$\hat{y}_5 = 45.176 + 2.176x_1 + 2.189x_2$

Table 7: Regression summary for dependent variable (age in 1995)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	47.502	3.345		14.201	0.000
Gender	1.804	8.428	0.039	0.214	0.832
Ethnicity	0.988	2.935	0.062	0.337	0.738

$\hat{y}_6 = 47.52 + 1.804x_1 + 0.988x_2$

Table 8: Regression summary for dependent variable (age in 1996)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	50.800	2.110		24.077	0.000
Gender	0.671	4.597	0.019	0.146	0.884
Ethnicity	0.293	1.641	0.023	0.179	0.859

$\hat{y}_7 = 50.8 + 0.671x_1 + 0.293x_2$

Table 9: Regression summary for dependent variable (age in 1997)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	48.399	1.933		25.032	0.000
Gender	1.132	3.478	0.032	0.325	0.746
Ethnicity	1.031	1.560	0.650	0.661	0.510

$\hat{y}_8 = 48.399 + 1.132x_1 + 1.031x_2$

Table 10: Regression summary for dependent variable (age in 1998)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	46.824	1.650	28.381	0.000	
Gender	4.435	3.834	0.119	1.157	0.250
Ethnicity	1.678	1.364	0.126	1.231	0.221

$\hat{y}_9 = 46.824 + 4.435x_1 + 1.678x_2$

Table 11: Regression summary for dependent variable (Age in 1999)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	44.564	1.306		34.117	0.000
Gender	7.905	2.555	0.292	3.094	0.002
Ethnicity	1.08	1.041	0.098	1.038	0.302

$\hat{y}_{10} = 44.564 + 7.905x_1 + 1.08x_2$

Table 12: Regression summary for dependent variable (age in 2000)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	43.247	1.981		21.826	0.000
Gender	6.112	2.916	0.206	2.096	0.039
Ethnicity	1.219	1.657	0.072	0.736	0.464

$\hat{y}_{11} = 43.247 + 6.112x_1 + 1.219x_2$

Table 13: Regression summary for dependent variable (age in 2001)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	44.040	1.969		22.365	0.000
Gender	12.546	2.951	0.477	4.251	0.000
Ethnicity	2.730	1.577	0.194	1.732	0.088

$\hat{y}_{12} = 44.04 + 12.546x_1 + 2.73x_2$

Table 14: Regression summary for dependent variable (age in 2002)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	44.740	2.999		14.919	0.000
Gender	13.628	6.319	0.502	2.157	0.047
Ethnicity	1.416	2.938	0.112	0.482	0.636

$\hat{y}_{13} = 44.74 + 13.628x_1 + 1.416x_2$

Table 15: Regression summary for dependent variable (Age in 2003)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	48.715	2.962		16.448	0.000
Gender	0.685	4.686	0.022	0.146	0.884
Ethnicity	2.458	2.330	0.159	1.055	0.297

$\hat{y}_{14} = 48.715 + 0.685x_1 + 2.458x_2$

Table 16: Regression summary for dependent variable (age in 2004)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	44.608	3.505		12.727	0.000
Gender	3.942	5.073	0.115	0.777	0.441
Ethnicity	1.624	2.805	0.086	0.579	0.566

$\hat{y}_{15} = 44.608 + 3.942x_1 + 1.624x_2$

Table 17: Regression summary for dependent variable (age in 2005)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	40.373	2.257		17.891	0.000
Gender	5.089	2.912	0.194	1.748	0.084
Ethnicity	1.460	1.540	0.105	0.948	0.346

$\hat{y}_{16} = 40.373 + 5.089x_1 + 1.460x_2$

Table 18: Regression summary for dependent variable (Age in 2006)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	42.212	1.806		23.378	0.000
Gender	2.272	2.255	0.090	1.008	0.316
Ethnicity	0.374	1.224	0.027	0.306	0.760

$\hat{y}_{17} = 42.212 + 2.272x_1 + 0.374x_2$

Table 19: Regression summary for dependent variable (age in 2007)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	42.210	1.673		25.232	0.000
Gender	2.781	0.905	0.118	1.460	0.147
Ethnicity	1.630	1.270	0.103	1.283	0.201

$\hat{y}_{18} = 42.210 + 2.781x_1 + 1.630x_2$

Table 20: Regression summary for dependent variable (age in 2008)

	Un standardized coefficients		Standardized coefficients		
Model	B	Std. error	Beta	t	Sig.
Constant	43.9936	1.479		29.698	0.000
Gender	1.8790	0.903	0.141	2.082	0.039
Ethnicity	0.99	1.108	0.061	0.902	0.368

$\hat{y}_{19} = 43.9936 + 1.879x_1 + 0.99x_2$

Table 21: Correlation for 1990

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.072	0.169
Gender	0.072	1.000	0.240
Ethnicity	0.169	0.240	1.000
Sig. (1-tailed)			
Age	1.000	0.439	0.359
Gender	0.439	1.000	0.302
Ethnicity	0.359	0.302	1.000

Table 22: Correlation for 1991

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.365	0.227
Gender	0.365	1.000	0.201
Ethnicity	0.227	0.201	1.000
Sig. (1-tailed)			
Age	1.000	0.062	0.175
Gender	0.062	1.000	0.204
Ethnicity	0.175	0.204	1.000

Table 23: Correlation for 1992

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.035	0.185
Gender	0.035	1.000	0.339
Ethnicity	0.185	0.339	1.000
Sig. (1-tailed)			
Age	1.000	0.441	0.218
Gender	0.441	1.000	0.072
Ethnicity	0.218	0.072	1.000

Table 24: Correlation for 1993

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.245	0.012
Gender	0.245	1.000	0.342
Ethnicity	0.012	0.342	1.000
Sig. (1-tailed)			
Age	1.000	0.100	0.476
Gender	0.100	1.000	0.034
Ethnicity	0.476	0.034	1.000

Table 25: Correlation for 1994

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.089	0.255
Gender	0.089	1.000	0.174
Ethnicity	0.255	0.174	1.000
Sig. (1-tailed)			
Age	1.000	0.290	0.054
Gender	0.290	1.000	0.138
Ethnicity	0.054	0.138	1.000

Table 26: Correlation for 1995

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.009	0.043
Gender	0.009	1.000	0.491
Ethnicity	0.043	0.491	1.000
Sig. (1-tailed)			
Age	1.000	0.478	0.395
Gender	0.478	1.000	0.000
Ethnicity	0.395	0.000	1.000

Table 27: Correlation for 1996

	Age	Gender	Ethnicity
Correlation			
Age	1.0000	0.025	0.028
Gender	0.0025	1.000	0.268
Ethnicity	0.0280	0.268	1.000
Sig. (1-tailed)			
Age	1.0000	0.420	0.410
Gender	0.4200	1.000	0.015
Ethnicity	0.4100	0.015	1.000

Table 28: Correlation for 1997

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.032	0.064
Gender	0.032	1.000	0.005
Ethnicity	0.064	0.005	1.000
Sig. (1-tailed)			
Age	1.000	0.374	0.255
Gender	0.374	1.000	0.481
Ethnicity	0.255	0.481	1.000

Table 29: Correlation for 1998

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.148	0.154
Gender	0.148	1.000	0.237
Ethnicity	0.154	0.237	1.000
Sig. (1-tailed)			
Age	1.000	0.070	0.063
Gender	0.070	1.000	0.009
Ethnicity	0.063	0.009	1.000

Table 30: Correlation for 1999

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.266	0.023
Gender	0.266	1.000	0.257
Ethnicity	0.023	0.257	1.000
Sig. (1-tailed)			
Age	1.000	0.002	0.405
Gender	0.002	1.000	0.003
Ethnicity	0.405	0.003	1.000

Table 31: Correlation for 2000

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.194	0.038
Gender	0.194	1.000	0.168
Ethnicity	0.038	0.168	1.000
Sig. (1-tailed)			
Age	1.000	0.024	0.351
Gender	0.024	1.000	0.043
Ethnicity	0.351	0.043	1.000

Table 32: Correlation for 2001

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.410	0.030
Gender	0.410	1.000	0.343
Ethnicity	0.030	0.343	1.000
Sig. (1-tailed)			
Age	1.000	0.000	0.398
Gender	0.000	1.000	0.001
Ethnicity	0.398	0.001	1.000

Table 33: Correlation for 2002

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.465	0.053
Gender	0.465	1.000	0.329
Ethnicity	0.053	0.329	1.000
Sig. (1-tailed)			
Age	1.000	0.022	0.414
Gender	0.022	1.000	0.084
Ethnicity	0.414	0.084	1.000

Table 34: Correlation for 2003

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.022	0.159
Gender	0.022	1.000	0.000
Ethnicity	0.159	0.000	1.000
Sig. (1-tailed)			
Age	1.000	0.442	0.146
Gender	0.442	1.000	0.500
Ethnicity	0.146	0.500	1.000

Table 35: Correlation for 2004

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.111	0.080
Gender	0.111	1.000	0.047
Ethnicity	0.080	0.047	1.000
Sig. (1-tailed)			
Age	1.000	0.227	0.294
Gender	0.227	1.000	0.375
Ethnicity	0.294	0.375	1.000

Table 36: Correlation for 2005

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.177	0.074
Gender	0.177	1.000	0.163
Ethnicity	0.074	0.163	1.000
Sig. (1-tailed)			
Age	1.000	0.055	0.254
Gender	0.055	1.000	0.071
Ethnicity	0.254	0.071	1.000

Table 37: Correlation for 2006

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.089	0.024
Gender	0.089	1.000	0.040
Ethnicity	0.024	0.040	1.000
Sig. (1-tailed)			
Age	1.000	0.160	0.396
Gender	0.160	1.000	0.327
Ethnicity	0.396	0.327	1.000

Table 38: Correlation for 2007

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.121	0.107
Gender	0.121	1.000	0.031
Ethnicity	0.107	0.031	1.000
Sig. (1-tailed)			
Age	1.000	0.068	0.094
Gender	0.068	1.000	0.354
Ethnicity	0.094	0.354	1.000

Table 39: Correlation for 2008

	Age	Gender	Ethnicity
Correlation			
Age	1.000	0.143	0.065
Gender	0.143	1.000	0.030
Ethnicity	0.065	0.030	1.000
Sig. (1-tailed)			
Age	1.000	0.018	0.170
Gender	0.018	1.000	0.331
Ethnicity	0.170	0.331	1.000

DISCUSSION

In Table 1, we found the age effect on gender and ethnicity in three years 1991, 2001 and 2002 are stronger than other years.

In regression models $\hat{y}_n$ in Table 2-20, there exist significance effect between age and gender in the models $\hat{y}_{10}, \hat{y}_{13}$ and $\hat{y}_{19}$. And there is no significance effect in the models $\hat{y}_1, \hat{y}_2$, $\hat{y}_3, \hat{y}_4$, $\hat{y}_5, \hat{y}_6$, $\hat{y}_7, \hat{y}_8$, $\hat{y}_9, \hat{y}_{11}$, $\hat{y}_{12}, \hat{y}_{14}, \hat{y}_{15}, \hat{y}_{16}, \hat{y}_{17}, \hat{y}_{18}$ and $\hat{y}_{19}$.

But there is no significance effect between age and ethnicity in all models, $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4, \hat{y}_5, \hat{y}_6, \hat{y}_7, \hat{y}_8, \hat{y}_9, \hat{y}_{10}$, $\hat{y}_{11}, \hat{y}_{12}, \hat{y}_{13}, \hat{y}_{14}, \hat{y}_{15}, \hat{y}_{16}, \hat{y}_{17}, \hat{y}_{18}$ and $\hat{y}_{19}$..

Correlation coefficient shows there is no significance relationship between age and gender in all years (1990-2008). Also there is no significance relationship between age and ethnicity, ethnicity and genders.

CONCLUSION

The age effect on gender and ethnicity in three years 1991, 2001 and 2002 are stronger than other years. In regression models $\hat{y}_n$, there exist significance effect between age and gender in the models $\hat{y}_{10}, \hat{y}_{13}$, but there is no significance effect between age and ethnicity in all models. In correlation, there is no significance relationship between age and gender, age and ethnicity, ethnicity and genders in all years from 1990-2008.

REFERENCES

1. Martin, B., 2000. An Introduction to Medical Statistics. 3rd Edn., Oxford University Press, Inc., New York, ISBN: 0192632698, pp: 405.
2. Montgomery, D.C., 2001. Design and Analysis of Experiments. 5th Edn., John Wiley and Sons, Inc., New York, ISBN: 10: 0471316490, pp: 672.