QUESTION

Describe a practical problem in any field of interest for which a regression model will be appropriate.Do a research to obtain real data relating to the problem described.

PROBLEM STATEMENT

There has been an upsurge of accidents on the Accra Kumasi Highway in recent times.In spite of this increased occurrence of accidents on the roads little have been done to find out the main causes of the life threatening situation.Although there have been massive reconstruction of the roads at several parts of the highway and in spite of the numerous road safety campaigns that have been undertaken,very little changes have occurred.In order to curb the occurrence of accidents on our roads,there is therefore the need to conduct research to find out the causes of the accidents and to find out the extent to which they do so.

METHODOLOGY

A survey was conducted at the Asafo Accra-Kumasi Lorry Park.In the survey, fifty commercial drivers were selected at random and they were asked questions on their past experiences in the last twelve months.The drivers answered questions on the number of accidents they had encountered, the number of times they maintained their vehicles ,the average speed at which they drove their vehicles and their ages. The number of accidents the drivers encountered was regressed on the number of times they maintained their vehicles,the average speed at which they drove their vehicles and their ages using the statistical package,SPSS.

AIMS

- To know if the number of accidents encountered by vehicles on the Accra Kumasi highway are related to number of times the drivers maintained their vehicles, the average speed at which they drove their vehicles and their ages.
- To find out the strength of the relationship that exist between the number of accidents encountered by the commercial vehicles, and the number of times they maintained their vehicles, the average speed at which they drove their vehicles and their ages if there exists any at all.
- To make predictions on the number of accidents the drivers could encounter given the number of times they maintained their vehicles, the average speed at which they drove their vehicles and their ages.
- To advice drivers and the highway authorities on ways to which they can help minimize accidents on the Accra Kumasi Highway.

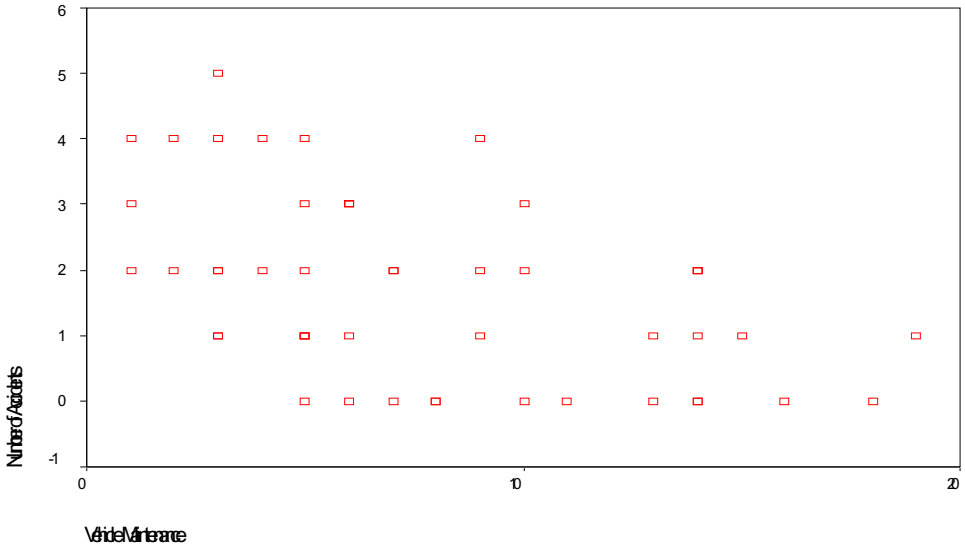
Data Values

Let Y: Number of accidents
X1: Vehicle Maintenance
X2: Average speed of driver
X3: Age of driver

Y	X1	X2	X3
2	14	85	35
1	13	75	56
2	7	95	48
4	5	100	29
2	7	70	43
0	18	75	57
3	6	110	37
2	14	100	40
4	4	100	25
1	14	80	30
1	19	70	34
3	10	95	32
2	10	85	36
4	3	70	27
1	15	90	45
2	14	75	60
3	5	100	29
4	9	85	41
1	3	75	26
0	16	70	58
2	9	75	46
1	5	70	53
0	14	70	48
0	8	60	52
2	4	75	36
0	7	65	50
3	6	90	34

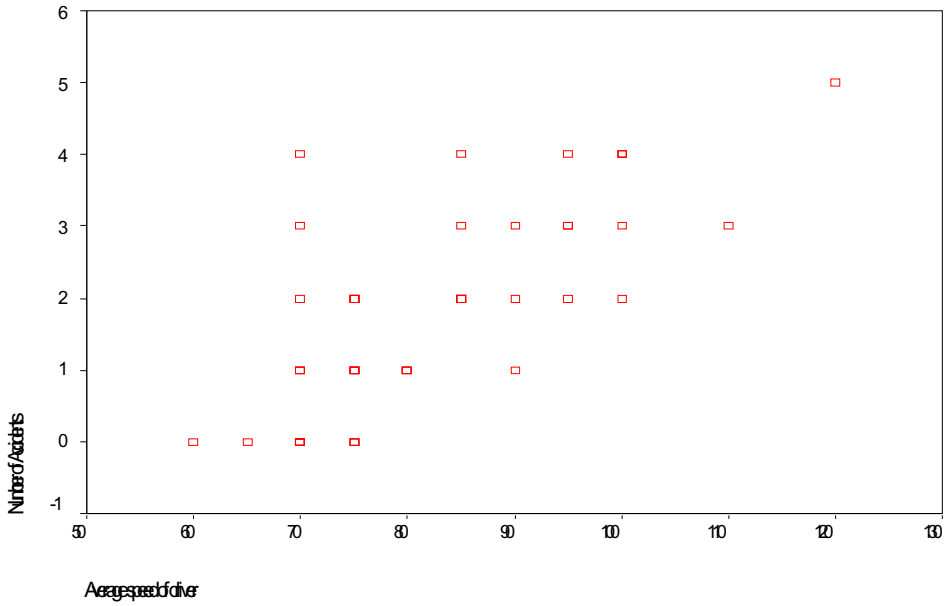
5	3	120	35
2	1	85	41
3	6	70	29
3	6	85	34
0	13	70	53
2	5	75	34
4	2	95	35
4	1	100	53
2	2	85	36
1	5	70	34
1	5	75	27
0	6	70	56
1	3	80	44
0	8	70	53
0	11	75	37
3	1	95	38
2	3	90	43
2	3	85	25
1	6	80	36
0	5	75	54
0	14	70	45
0	10	70	38
1	9	75	57

Number of Accidents vrs Vehicle Maintenance



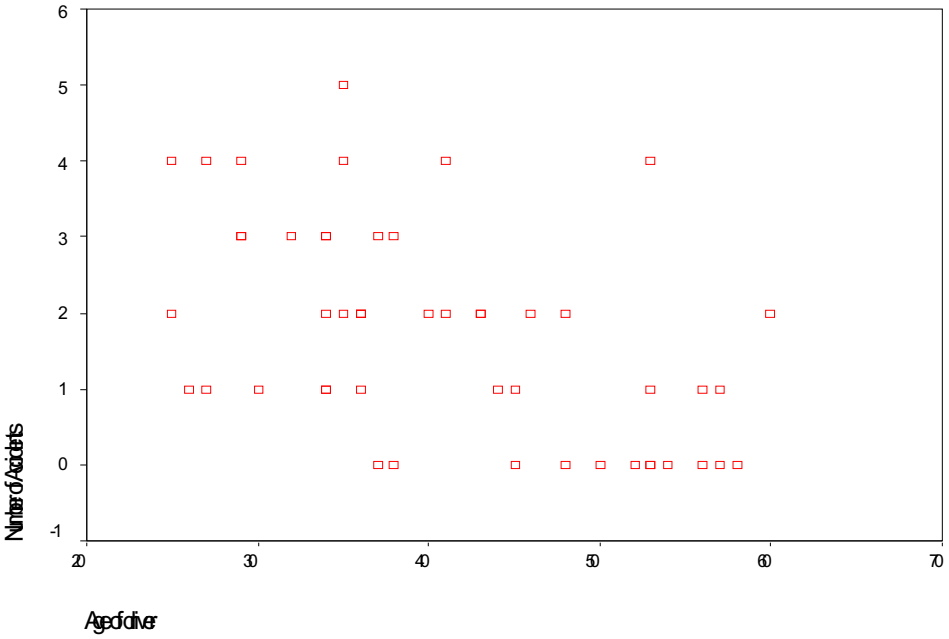
It is observed that there is a negative linear relationship between the number of accidents and the age of driver.

Number Accidents vrs Average speed of driver



It is observed that there is a positive linear relationship between the number of accidents and the average speed of driver.

Number Accidents vrs Age of Driver



It is observed that there is a negative linear relationship between the number of accidents and the age of driver.

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age of driver, Average speed of driver, Vehicle Maintenance[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Number of Accidents

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.812[a]	.659	.637	.842

a. Predictors: (Constant), Age of driver, Average speed of driver, Vehicle Maintenance

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	62.995	3	20.998	29.616	.000 [a]
	Residual	32.625	46	.709		
	Total	95.620	49			

a. Predictors: (Constant), Age of driver, Average speed of driver, Vehicle Maintenance

b. Dependent Variable: Number of Accidents

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.745	1.151		-1.516	.136
	Vehicle Maintenance	-6.51E-02	.028	-.222	-2.346	.023
	Average speed of driver	6.45E-02	.010	.589	6.245	.000
	Age of driver	-3.11E-02	.013	-.225	-2.345	.023

a. Dependent Variable: Number of Accidents

Correlations

Correlations

		Number of Accidents	Vehicle Maintenance	Average speed of driver	Age of driver
Number of Accidents	Pearson Correlation	1	-.489 **	.739 **	-.513 **
	Sig. (2-tailed)	.	.000	.000	.000
	N	50	50	50	50
Vehicle Maintenance	Pearson Correlation	-.489 **	1	-.316 *	.360 *
	Sig. (2-tailed)	.000	.	.025	.010
	N	50	50	50	50
Average speed of driver	Pearson Correlation	.739 **	-.316 *	1	-.354 *
	Sig. (2-tailed)	.000	.025	.	.012
	N	50	50	50	50
Age of driver	Pearson Correlation	-.513 **	.360 *	-.354 *	1
	Sig. (2-tailed)	.000	.010	.012	.
	N	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

DESCRIPTION OF MODEL

The regression model is of the form

$$\boldsymbol{Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3}$$

with *Y* being the number of accidents encountered by the drivers and the dependent variable. X_1, X_2, X_3 represents the number of times they maintained their vehicles, the average speed at which they drove their vehicles and their ages respectively;they represent the independent variables.

$Y = -1.745 - 0.0651X_1 + 0.06454X_2 - 0.0311X_3$ $R^2 = 0.695$
(1.51) (0.028) (0.010) (0.013) S.e

The estimated regression model above depicts the relationship between the predictor variables and the response variable and their standard errors.

The value of R (0.812) depicts a strong linear relationship between all the predictor variables and the response variable.
An R^2 value of 0.659 means that the regression model fits well to the set of data points and that 65.9 percent of the variation of the number of accidents the drivers had is

explained by the multiple regression model. The Adjusted R^2 means that 63 percent of the regression model can be predicted with the estimated regression model when the number of variables is taken into account.

.

TESTS OF SIGNIFICANCE

The test of the overall significance and viability of the estimated model is as follows;

$H_0 : B_j = 0$

$H_1 : B_j \neq 0$

for at least one, j = 1,2,3

From the anova table, the p-value is given to be 0.00, since it is less than $\propto = 0.05$, we reject the null hypothesis and conclude that the number of accidents a driver encountered is at least dependent on one of the following; the number of times they maintained their vehicles, the average speed at which they drove their vehicles and their ages.

To find out the strength of each predictor variable with the response variable, t-tests are carried out as follows;

$H_0 : B_k = 0$

$H_1 : B_k \neq 0$

To test whether vehicle maintenance is related to the number of accidents encountered by the drivers, take k = 1.

$H_0 : B_1 = 0$

$H_1 : B_1 \neq 0$

Since a p-value of 0.023 is less than 0.05, we reject H_0 and conclude that the number of accidents encountered by the drivers depends on the number of times they maintained their vehicles.

To test whether the average speed of drivers is related to the number of accidents encountered by the drivers, take k = 2.

$H_0 : B_2 = 0$

$H_1 : B_2 \neq 0$

Since a p-value of 0.00 is less than 0.05, we reject H_0 and conclude that the number of accidents encountered by the drivers depends on the average speed they drove their vehicles.

To test whether age of driver is related to the number of accidents encountered by the drivers,take k = 3.
$H_0 : B_3 = 0$

$H_1 : B_3 \neq 0$

Since a p-value of 0.023 is less than 0.05,we reject H_0 and conclude that there is evidence from the data that the number of accidents encountered by the drivers depend on the age of the driver.

CORRELATION ANALYSIS

The correlations table depicts the strength of the relationship that exist between the number of accidents encountered by the commercial vehicles as the dependent variable,and the number of times they maintained their vehicles, the average speed at which they drove their vehicles and their ages.

It is easily observed that the number of times the drivers maintained their vehicles is negatively related to the number of times they had accidents but the Pearson Correlation Coefficient of 0.489 in absolute value shows that the relationship is a weak one.This suggests that drivers who maintained their cars more often experienced less number of accidents.

It is also observed that the average speed at which the commercial drivers drove on the Accra Kumasi Highway is positively related to the number of accidents they encountered.Thus,the faster they drove,the more they risked getting an accident.The Pearson Correlation Coefficient of 0.739 reveals that the relationship between the average speed at which the commercial drivers drove and the number of accidents they encountered is very strong.

The age of the drivers is observed to be negatively related to the number of times they had accidents and the Pearson Correlation Coefficient of 0.513 in absolute value shows that this relationship is a good one.

Nevertheless, the correlation between each variable and itself is a perfect one with a Correlation Coefficient of 1.

PREDICTION

To predict the number of accidents a driver of 25 years of age who maintained his vehicle 4 times and drove at an average speed of 100 km/h the data is put into the regression model as follows;

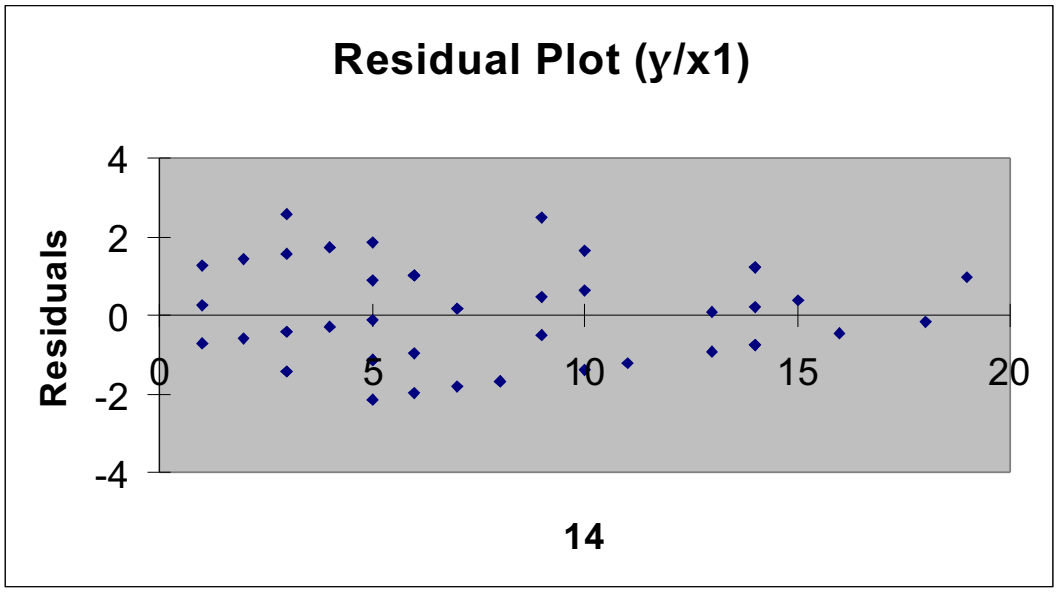
$$Y= -1.745-0.0651X_1+0.06454X_2-0.0311X_3$$

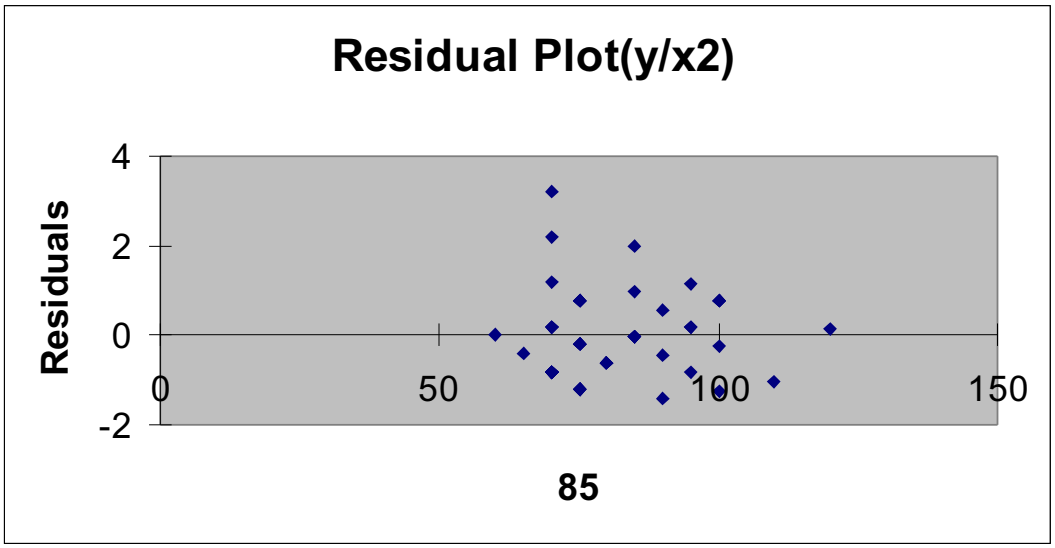
$$Y = -1.745-0.06514(4)+0.06454(100)-0.0311(25)$$

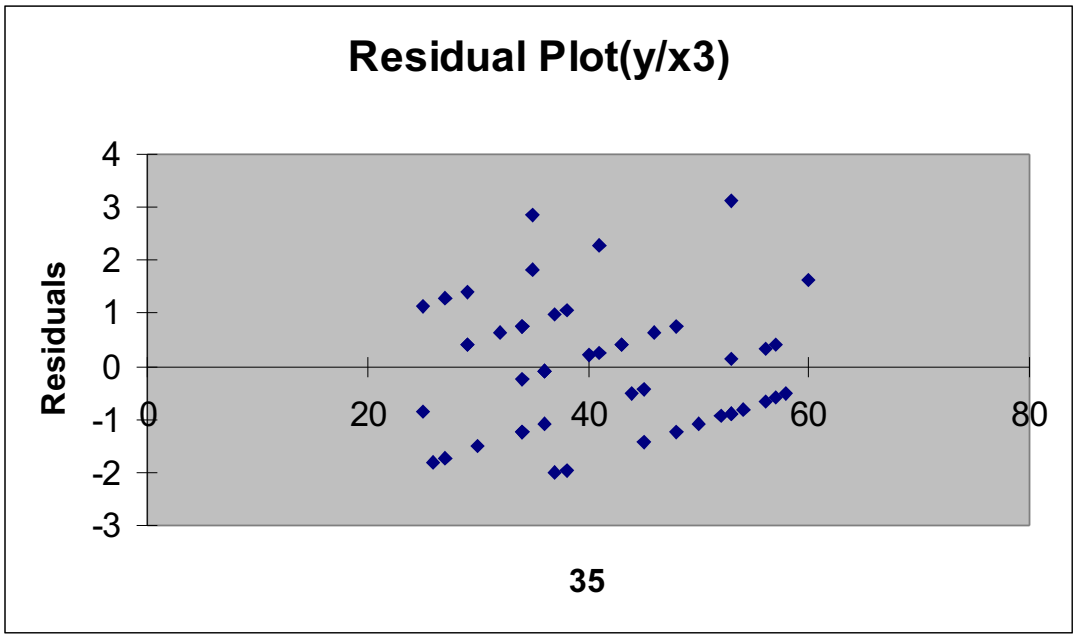
$$=3.67$$

Hence a driver of 25 years of age who maintained his vehicle 4 times and drove at an average speed of 100 km/h may encounter 3.67 accidents which is 0.33 away from the observed value of 4.Hence,with a residual error of just 0.33,the regression model can be said to be adequate with vehicle maintenance and average speed of driver being most influencial in the model.

RESIDUAL PLOTS







.

CONCLUSION

The number of accidents encountered by vehicles on the Accra Kumasi highway are related to number of times the drivers maintained their vehicles, the average speed at which they drove their vehicles and their ages.

From the model it is observed that the more a driver did vehicle maintenance,the less he got accidents.Younger drivers got more.The faster a driver drove,the more he encountered accidents.

ROAD SAFETY PROPOSALS

In order to minimize road accidents;

- Drivers who ply the highway must be forced to maintain their vehicles more often by way of conduction of regular checkups at the exit points of the lorry station,so that vehicles that are not in good condition are stopped from embarking on the journey.
- There should be more speed limit indications and road safety signs at various points on the highway.An uncompromising law enforcement agency should be

posted at various segments of the highway to make sure defaulters are made to face the law.

- Young drivers must be educated on the dangers they impose on themselves whenever they drive at high speeds.

REFERENCES

Econometric Models and Forecasts (Irwin & Mc-Graw-Hill); Robert S. Pindyck & Danile.

Elementary Statistics (Mc Graw-Hill); Allan G. Bluman