Arab Journal of Administrative Sciences
1998, Vol. 5, No. 2, 519 - 532

MODELING SEVERITY OF AUTO ACCIDENTS IN KUWAIT

Hosny I. Hamdy
Mohamed A. Al-Mahmeed
Moawad H. El-Habshi
Kuwait University

This article sheds light on drivers' characteristics and their association with auto accident's severity in Kuwait. In particular, this study investigates the existence of significant association between the severity of an accident and a set of selected demographical predictors that might characterize drivers according to their anticipated risk; such as age, gender, nationality and place of accident. The analysis was conducted based on a collection of 345 police reports during the period 1994-1995. We extended our analysis to establish a model that quantifies the relationship between the probability of committing a severe or minor accident with a set of significant predictors through the logistic approach. Other interesting results are also included.

The market determines risk classification schemes and associated insurance prices that affect the societal allocation of the cost of risk. The importance of risk classification in automobile insurance has been debated for at least two decades. See for example (Spaht and Escolas, 1982; Austin, 1983; McCarthy and Turner, 1993; Harrington and Doerpinghous, 1993; Witt, 1979; Chang and Fairley, 1979; Sant, 1980 and Tryfos, 1980). Witt, (1979) found some supportive evidence for territorial rating in auto insurance pricing. However, his recommendation was that more research is needed in that direction, (Chang and Fairley 1979; Bennett, 1978 and Sant, 1980) contemplated the use of some traditional automobile rate making. They also considered the importance of a risk classification system in auto insurance risk assessment and pricing.

Submitted March 1997, accepted May 1997.

In many countries, government regulations have significantly restricted certain types of information regarding the characteristics of auto drivers involved in accident, including sex, marital status, nationality, and territory on discretionary underwriting. Such information can be used by the insurers to model anticipated risks of auto drivers, see for example (Shayer, 1978; Abraham 1985, 1986; Austin, 1983; Wortham, 1986 and Butler and William, 1988) for details.

Actuarial community generally argues that pricing and classification should be a cost-based process. That is, the price charged to a buyer should reflect the estimated expected value of claim. In addition to a loading cost that covers administrative costs in order to provide compensation for bearing the risk. Therefore, the objective of such price classification is to achieve the most homogeneous rate classes and predictive accuracy. (Spahr and Escolas, 1982) examined the effect of drivers' characteristics (age, sex and place of residence) on the frequency of accidents, the resulting physical injury, and the property damage. Their investigation indicated that these variables are reasonably good predictors to severity of accident.

The purpose of this paper is to examine the effect of drivers' characteristics on the severity of auto accidents in Kuwait. Specifically, we hypothesized the relationship between the severity of auto accidents in Kuwait and a set of selected predictors including age, sex, nationality, and place of accident. Therefore, we extend our investigation to establish a model that might characterize drivers according to their encountered risks. Moreover, we extend our investigation to establish a model that quantifies the relationship between significant predictors and the probability of committing a severe or a minor accident. Although the main objective of this paper is to help understand the characteristics of auto drivers' severity for the use in insurance business, such classification would help in other areas as well. For instance, it would shed light on traffic issues to find objective solutions. In particular, airbag legislative efforts would be an attempt to reduce the uncertainty or the likelihood of severe accidents. Moreover, it would undoubtedly serve to set up a merit/demerit system to penalize careless drivers... etc.

The Data

The data used in this analysis was collected from daily police reports of auto accidents in Kuwait. We gathered 345 reported cases of auto accidents in Kuwait over the period 1994-1995. Each report shows the time and the place of accident along with the driver's characteristics (age, sex, and nationality) as well as the severity of the accident as being severe or minor. However, only 337 cases were included in the analysis while the remaining 8 were excluded due to their insufficient or unsatisfactory information. All predictors are treated as indicator variables (dummy variables) except for age which is treated as a covariate.

Consider Y as a dichotomous random variable representing the degree of severity caused by an auto accident. The random variable Y assumes two values, 0 if the severity is minor (bodily injury), or 1 if the severity is major (resulting in death). The independent variables considered in this study are the age of the driver X_1, and gender of the driver, X_2, as being a male or female driver. The nationality of the driver, X_3, where we consider only two groups, Kuwaiti or non-Kuwait, and finally the place of accident, X_4. Furthermore, the place of accident is divided into 5 different categories according to the nature of the road as well as speed limits, whether the road is one lane, X_{41}, two lanes, X_{42}, a freeway, X_{43}. a high-way, X_{44}, or a ring road, X_{45}, (Kuwait has seven highways in the forms of rings that run across the country, with a maximum allowable driving speed of 120 Kilometers/hour).

First we investigate the degree of association between the severity, Y, and each of the independent variables to select the best subset of predictors that explain Y. The following Table (1) presents the results of the contingency between Y and the age groups of the drivers X_5. Note that Obs means observed frequency and Exp stands for expected frequency.

Table 1

Age Group	18-27		28-37		38-47		48-57		58 or more	
Y	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
0 (minor)	15	58.1	40	35.5	57	34.2	27	14.1	5	2.1
1 (severe)	121	77.9	43	47.5	23	45.8	6	18.9	0	2.9
% of severe accidents	**62.7**		**22.3**		**11.9**		**3.1**		**0.0**	

The analysis indicates that 62.7% of all severe accidents were committed by drivers in the age group 18-27 while 22.3% of the severe accidents were committed by drivers in the age goup 28-37. This percentage decreases in older age groups. It is also evident that 89% of total accidents committed by the age group 18-27 were severe. These observances point out an important characteristic of the risk regarding the drivers age group. To sum up the above findings, we would say that younger drivers are more likely to commit severe accidents regardless of nationality, sex, and place of accident.

Regarding the nationality of drivers, the following Table (2) provides details concerning the association between X_3 and the degree of severity Y.

Table 2

Driver's Nationality	Kuwaiti		Non Kuwaiti	
Y	Obs	Exp	Obs	Exp
0 (minor)	70	84.6	74	59.4
1 (severe)	128	113.4	64	79.6
% of severe accidents	**66.3**		**33.7**	

Table (2) shows that 58.8% of total accidents were committed by Kuwaiti drivers of which 64.7% were severe while 41.3% of accidents were committed by non-Kuwaiti drivers of which 46.8% were severe. In other words, Kuwaiti drivers are more likely to commit severe accidents than non-Kuwaiti drivers.

With respect to the association between the degree of severity Y and the gender of the driver X_2, a detailed table is omitted for brevity. Yet, analysis confirms that 70.9% of the accidents were committed by male drivers, of which 57.3% were severe. While for female drivers, out of the 29% of their accidents, 57.1% were severe.

Regarding the degree of severity of accidents, the condition of the road, and the speed limit, the analysis indicates significant association. between the severity of the accident, the nationality of the driver and the place of accident. Further details are given in the next section.

Further, if sex is controlled for male drivers Table (3) provides cross-tabulation results of the degree of severity and the age of both Kuwaiti and non-Kuwaiti drivers. A similar pattern holds when sex is controlled for female drivers as shown in Table (4).

Table 3

Male										
Nationality	**Kuwaiti**					**Non - Kuwaiti**				
Y Age Group	**18-**	**28-**	**38-**	**48-**	**58>**	**18-**	**28-**	**38-**	**48-57**	**58>**
(minor) Obs	7	16	13	9	2	4	10	29	9	3
Exp	25.7	9.9	7.2	3.4	0.7	14.0	12.4	20.5	6.5	1.6
(evere) Obs	68	13	8	1	0	22	13	9	3	0
Exp	49.3	19.1	13.8	6.6	1.3	12.0	10.6	17.5	5.5	1.4
% of severe accidents	**75.6**	**14.4**	**8.9**	**1.1**	**0.0**	**46.8**	**27.7**	**19.1**	**6.4**	**0.0**

Table 4

Male										
Nationality	**Kuwaiti**					**Non - Kuwaiti**				
Y Age Group	**18-**	**28-**	**38-**	**48-**	**58>**	**18-**	**28-**	**38-**	**48-57**	**58>**
(minor) Obs	4	8	7	4	-	0	6	8	5	-
Exp	10.6	6.4	4.1	1.9	-	3.6	7.2	5.1	3.1	-
(severe) Obs	24	9	4	1	-	7	8	2	1	-
Exp	17.4	10.6	6.9	3.1	-	3.4	6.8	4.9	2.9	-
% of severe accidents	**63.2**	**23.7**	**10.5**	**2.6**	**0**	**38.9**	**44.4**	**11.1**	**5.6**	**0.0**

With regard to the relationship between the degree of severity, nationality, and gender of the driver, Table (5) presents cross-tabulation results which illustrate the fact that if the severity is classified by gender and nationality, male driver are more likely to commit severe accidents than female drivers in general and among Kuwaiti drivers in particular. On the contrary, non-Kuwaiti males are more likely to commit minor accidents than Kuwaiti males.

Table 5

Y	0 (Minor)				1 (Severe)				
Driver Gender	Kuwaiti		Non-Kuwaiti		Kuwait		Non - Kuwaiti		P. Value
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	
Male	47	58.5	55	43.5	90	78.5	47	58.5	0.002
Female	23	26.1	19	15.9	38	34.9	18	21.1	0.186

So far we have identified some of the significant predictors that might explain the severity Y. Our goal in the following section is aimed at building a mathematical model that predicts the probability of committing severe or minor accidents given some significant characteristics of the driver. Although all characteristics as well as all possible interactions up to the fourth order have been tested against the model, we only report those terms which declare significance.

Logistic Regression Model

One of the advantages of logistic analysis is that we need only to know whether an event occurred and use a dichotomous value as our dependent variable. Since the severity Y is a dichotomous random variable, the normality assumption of the error term in the ordinary least square (OLS) method of estimation is no longer valid. Therefore, we utilize the logistic regression approach (Hair et al, 1992) to select a set of significant predictors, along with possible interactions to the fourth order, that can logically explain the behavior of Y. From this dichotomous value, the procedure predicts its estimate of the probability that the event will occur. If the predicted probability is greater than 0.5, then the prediction is severe accident given the set of predictors, otherwise it is minor. Thus, the predicted value in this case E(Y) is interpreted as the probability that the driver commits a severe or minor accident, given certain characteristics of the driver. The basic idea of the logistic analysis stems from the fact that the random variable Y is a Bernoulli trial, which describes the occurrence of two mutually exclusive events (either severe or minor accident).

Therefore, the expectation of the random varibale Y is the same as P (Y = 1). To this end, we write the logistic regression model as:

$$\frac{E(Y)}{(1-E(Y))} = e^{\beta_0 + \sum_i \beta_i X_i + \sum_{i \neq j} \beta_{ij} X_i X_j + \sum_{i \neq j \neq k} \beta_{ijk} X_i X_j X_k + \sum_{i \neq j \neq k \neq \ell} \beta_{ijk\ell} X_i X_j X_k X_\ell} \quad , (3.1)$$

where, $i = j = k = l = 1, 2, 3, 41, 42, 43, 44, 45$.

The procedure that calculates the regression coefficients is similar to the least squares criterion. Eventually, it makes comparisons of the probability of an event "severe accident" occuring with the probability "minor accident" occurring given the set of predictors. This odds ratio is expressed in logarithms, such that they need to be transformed back by taking the antilog, of the values so that their relative effect on the probabilities can be assessed more easily. However, computer programs do this procedure automatically. To this end we write the model in the form:

$$\frac{P(Y=1)}{P(Y=0)} = \frac{E(Y)}{(1-E(Y))} = e^{\beta_0 + \sum_i \beta_i X_i + \sum_{i \neq j} \beta_{ij} X_i X_j + \sum_{i \neq j \neq k} \beta_{ijk} X_i X_j X_k + \sum_{i \neq j \neq k \neq \ell} \beta_{ijk\ell} X_i X_j X_k X_\ell} \quad ,(3.2)$$

To obtain a linear regression model, we transform the above exponential model into the linear form as:

$$\ell n \left(\frac{E(Y)}{(1-E(Y))} \right) = \beta_0 + \sum_i \beta_i X_i + \sum_{i \neq j} \beta_{ij} X_i X_j + \sum_{i \neq j \neq k} \beta_{ijk} X_i X_j X_k + \sum_{i \neq j \neq k \neq \ell} \beta_{ijk\ell} X_i X_j X_k X_\ell \quad (3.3)$$

where, β_0 is the intercept (constant), β_i is the coefficient of X_i, β_{ij} is the coefficient of the 2-way interaction between X_j and X_j, β_{ijk} is the coefficients of the 3-way interaction between X_i , X_j and X_k, and $\beta_{ijk\ell}$ is the coefficient of the 4-way interaction between X_i, X_j, X_k and X_ℓ. Actually, the parameters β^s are measures of the relative changes in the odds ratio in (3.2). Furthermore, if the coefficient β is positive the antilog is greater than one, hence the odds ratio increases. This increase occurs when the predicted probability of the event "severe accident" occurring increases and the predicted probability of the event "minor accident" occurring is reduced. Likewise, if β is Zero, this indicates no change in the odds.

Wald's backward elimination procedure was performed to select the best subset of predictors that explain Y. The following Table (6) gives the estimates of the encountered parameters and the significance of each predictor. Other variables indicating insignificant relationship (P-$value > 0.1$) were dropped from the model.

Table 6

variables	β	S.E.	Wald	P-Value	R	e^{β}
X_{41}	3.6815	1.1074	11.052	0.0009	0.1403	39.705
X_{42}	6.4028	1.4775	18.7805	0	0.191	603.531
X_1 by X_{41}	-0.0883	0.0291	9.2023	0.0024	-0.1251	0.915
X_1 by X_3	-0.115	0.0258	19.8468	0	-0.197	0.891
X_2 by X_{42}	-2.3898	0.7968	8.9952	0.0027	-0.1233	0.092
X_2 by X_{44}	1.2608	0.5304	5.6514	0.0174	0.0891	3.528
X_1 by X_2 by X_{45}	0.041	0.0135	9.2537	0.0024	0.1256	1.042
X_1 by X_3 by X_{42}	-0.0548	0.0323	2.88	0.0897	-0.0437	0.947
X_1 by X_3 by X_{43}	0.0796	0.0297	7.1728	0.0074	0.106	1.083
X_1 by X_3 by X_{45}	-0.0445	0.0142	9.7613	0.0018	-0.1299	0.957
X_1 by X_2 by X_3 by X_{42}	0.0668	0.0342	3.8201	0.0506	0.0629	1.069
X_1 by X_2 by X_3 by X_{43}	-0.0573	0.0299	3.6735	0.0553	-0.0603	0.944
constant	**-2.3367**	**1.2531**	**3.4772**	**0.0622**		

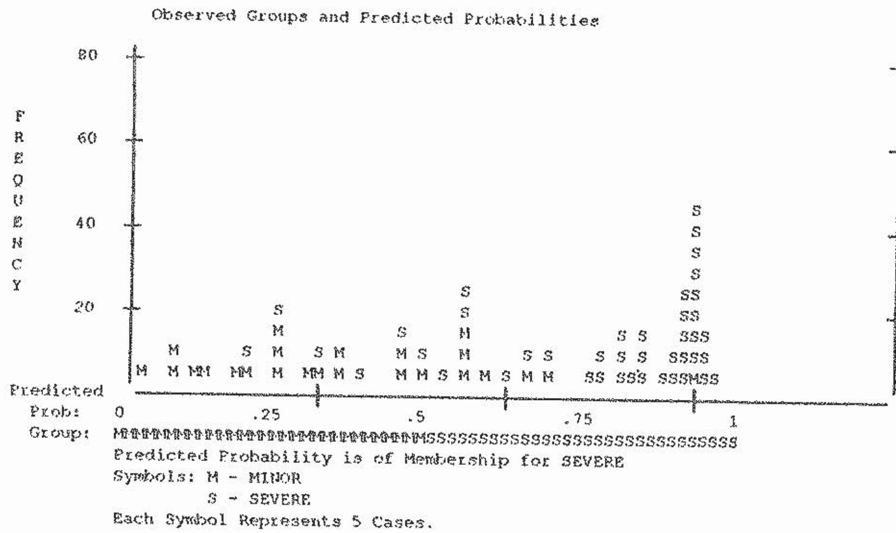
The result of logistic analysis shows that among the encountered characteristics of the drivers considered in this study and the nature of the road (being one lane, two lanes, freeway, highway or ring road) affect the severity of the accident. The interaction between the age of the driver travelling on a one lane or two lane road also determines the severity. It is also evident that the gender of a driver, travelling on a two lane road, can significantly indicate whether the accident results in death or bodily injury. Males are more likely to commit severe accidents than females when they drive on two lane roads or highways. Young Kuwaiti males driving on a two lane road or freeway are more likely to commit severe accidents. It would be also of interest to investigate the ability of the model to correctly classify accidents (severe or minor) based on their predicted probability. The following Table (7) summarizes the percentage of correct classifications based on the proposed model.

Table 7

Observed frequency	**Predicted Frequency**		**Percent Correct**
	Minor (M)	**Sever (S)**	
Minor(M)	103	41	71.53%
Severe(S)	30	163	84.46%
Overall Percentage of Correct Classification			**78.93%**

From Table (7), we see that the total percentage of correct classification is 78.93%. This means that among 144 minor accidents the model was able to correctly classify 103 cases. On the other hand, among 193 severe accidents the model was able to classify 163 cases correctly. The classification map Fig.(1) provides the predicted probabilities of drivers membership for severe (S) or minor (M) accidents.

Fig. 1



As we can see, the model overestimates the percentage of severe accidents by ≈ 7% relative bias. In contrast, it underestimates the percentage of minor accidents by ≈ 8% relative bias.

Hence, we give the fitted model as

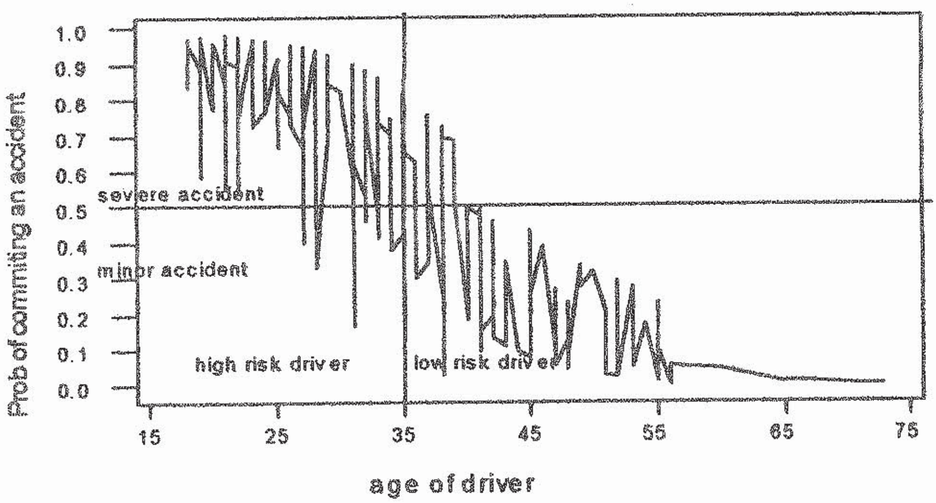
$$\hat{\pi} = -2.337 + 3.68X_{41} + 6.4X_{42} - .088X_1X_{41} - .115X_1X_{42} - 2.39X_2X_{42} + 1.26X_2X_{44} + 0.041X_1X_2X_{45} - .055X_1X_3X_{42} + 0.08X_1X_3X_{43} - .0445X_1X_3X_{45} + .0668X_1X_2X_3X_{42} - 0.057X_1X_2X_3X_4 \quad (3.4)$$

where, $\hat{\pi} = \ell n(\hat{p}/(1-\hat{p}))$ is the logarithm of the ratio between the probability $\hat{p}$ that the driver with certain charateristics will commit a severe accident and the probability (1-$\hat{p}$) that the driver with the same characteristics will commit a minor accident.

This study reveals several facts concerning classifications of drivers into severe risk or minor risk groups. According to their age, nationality, gender and place of accident. The logistic analysis shows that the interactions of these characteristics are the best predictors of the severity of an accident. Among the detected characteristics in this study, we find that the age of the driver plays a significant role in determining the severity of an accident. For instance, a male Kuwaiti driver belonging to the age group 18-27 has a high probability of commiting a severe accident on a one lane road or freeway. Accordingly, these highly risk drivers should be penalized in determining their premiums for auto insurance. Fig(2) presents the behavior of the predicted probability of committing an accident versus the driver's age. Apparently, young drivers ($\leq$ 35 years) are more likely to commit severe accidents than older drivers. The probability drops down considerably for older ages.

Fig 2

Classification of Predicted Risk by age of the driver



Conclusion

This study discloses several facts regarding classifications of drivers in Kuwait to high or low risk. We have found that age is a critical predictor, especially when interacting young drivers with gender (male). nationality (Kuwaiti), or even higher order interactions (young, male, Kuwaiti, place of accident). Further, the gender of the drive is a factor to be considered, in particular male drivers. Next, the nationality of the driver appears to play a role in evaluating the predicted probability, distinctly Kuwaiti drivers. Finally, road conditions are a significant factor, specially one lane, two lane, freeway and ring roads.

Although we have considered some major characteristics in this study to classify drivers according to their anticipated risk, other important factors were not available (not recorded) such as educational level, marital status, driving experience, alcoholic problems, psychological stability, driving record, occupation, vehicle specifications, vehicle age, limits of liability as well as territory. Therefore, we do not claim that we have reached a final identification model regarding drivers anticipated risk.

Rather, we say that further studies which include all related characteristics are required to better identify drivers according to their expected risk. While we have focused on human losses, property damage should also be addressed in future analysis.

As a closing statement, the natural question would be, who should bear the loss? According to our findings in this study, high risk drivers (reckless driver) should be penalized in the determination of their auto insurance premiums. In other words a fair pricing system should be imposed. It is also the responsibility of the authorities to protect other citizens by applying a merit/ demerit system and enforce safety requirements (Peltzman, 1975) such as airbags (Peterson, and Hoffer, 1996) and seat belt legislation (Adams, 1994). Finally, insurance companies also have a role to play regarding this issue.

References

Abraham, K. 1985. Efficiency and Fairness in Insurance Risk Classification. ***Virginia Law Review***, 71: 403-51

Abraham, K. 1986. Distributing Risk Insurance Legal Theory and Public Policy. ***Yale University Press New Haven, Conn.***

Adams, J.G. 1994. Seat Belt Legislation: The Evidence Revisited. ***Safety Science***, 18: 135-152.

Austin, R. 1983. The Insurance Classification Controversy. ***University of Pennsylvania Law review***, 131: 517-583.

Bennett, M.C. 1978. Models in Motor Insurance. ***Journal of Institute of Actuaries Students Society***, 22: 134 - 159.

Butler, P. and Williams, L. 1988. Sex-divided Mileage Accident and Ansurance Cost Data. ***Journal of Insurance Regulation***, b: 243 - 284.

Change, L. and Fairley, W. 1979. Princing Automobile Insurance Under Multivariate Classification of Risk: Additive Versus Multiplicative. ***The Journal of risk and Insurance***, 46: 75 - 93.

Harrington, S. and Doerpinghaus, H. 1993. The economics and politics of automobile insurance rate classification. ***The Journal of risk and Insurancee***, 60: 59 - 84.

Hair, j. F., Anderson, R.E., Tatham, R.L. and Black W.C. 1993. Multivariate data analysis with readings. ***Macmillan Publishing Company, New york*** 10022.

McCarty, D. and Turner, J. 1993. Risk Classification and six discrimination in pension plans. ***The journal of Risk and Insurance***, 60: 84 - 104.

Peltzman, S. 1975. The effect of Automobile safety Regulations. ***Journal of Political Economy***, 83: 677 - 725.

Peterson, S. and Hoffer, G.E. 1996. Auto insurance and Airbag: Comments. ***The Journal of Risk and Insurance***, 63: 515 - 523.

Sant, D. 1980. Estimating Expected Losses in auto Insurance. ***The Journal of Risk and Insurance***, 47: 132 - 151.

Shayers, N. 1978. Driver Classification in Automobile Insurance. ***Boston Division of Insurace.***

Sheppard, W. 1978. Pennsylvania Regulator Disagrees With Part of Massachusetts auto Rate Reform. ***Insurance Digest***, (December 16).

Hosny I. Hamdy (Ph.D. Oklahoma State University, 1983), Professor of Mathematics & Statistics, Kuwait University. His Area of Interest Includes, Sequential Analysis, Decision Theory, Application of Statistics in Business, Biomedical Research, Engineering, Statistics Reliability Theory, and Life Testing.

Mohamed A. Al-Mahmeed (Ph.D. Oklahoma State University, 1982), Associate Professor of Statistics, Kuwait University. His Area of Interest Includes, Time Series Analysis, Sequential Analysis, and Computational Statistics and Decision Theory.

Moawad H. El-Habashi (Ph.D. Shefield University, 1976), Professor of Insurance, Kuwait University. His Area of Interest Includes, Insurance Business, Investment, Protfolio Analysis and Risk Theory.

تصميم نموذج كمي لقياس وطأة الخسارة لحوادث السيارات بدولة الكويت

حسني إبراهيم حمدي
محمد عبدالهادي المحميد
معوض حسن الحبشي
جامعة الكويت

يهدف هذا البحث إلى دراسة العلاقة بين الصفات الديموغرافية للسائق ووطأة الخسارة (حجم الخسارة) في حالة وقوع حادث سيارة بدولة الكويت. وبالتحديد، فإن الدراسة تبحث في وجود علاقة بين وطأة الخسارة وبعض المتغيرات الديموغرافية للسائق مثل السن والجنس والجنسية ومكان وقوع الحادث. وقد اهتم البحث بدراسة هذه العلاقة من خلال تجميع بيانات خاصة بعدد من حوادث السيارات (٣٤٥) خلال الفترة ١٩٩٤ - ١٩٩٥ من واقع محاضر التحقيق الخاصة بهذه الحوادث في مخافر الشرطة بدولة الكويت. وفي هذا البحث تم استخدام التحليل اللوجستي والنماذج الخطية لبناء نموذج كمي لتقدير احتمال وقوع حادث ذي وطأة عالية (خسارة كبيرة) أو حادث ذي وطأة منخفضة (خسارة منخفضة). ويعرض البحث النتائج الخاصة بدرجة العلاقة بين المتغيرات الديموغرافية للسائق وحجم الخسارة.