Death on the Track, the Econometrics Behind Formula 1

Jane Lee
Junior Sophister

Econometric techniques need not be constrained to economic theory and models. Jane Lee examines the incidence of death in Formula 1 racing since its beginging in 1950, with various factors concerning the circuits that have been used

Introduction

In 1950 the first Formula One Grand Prix was held. Since its creation, the World Championship has undergone numerous regulation changes. For reasons of safety, the governing body has sought to channel the ingenuity of the engineers, whose prime objective is always to improve the performance of their machine. In just over twenty pages, the championships regulations clearly set out the limits in which the engineers can work. These are very important, considering the lengths to which Formula One teams will go to gain that all important extra one hundredth of a second.

It is the area of safety that my project is concerned. My hypothesis is that the deaths of drivers during the history of Formula One can be explained by the characteristics of the circuits on which the deaths occurred.

The Dependent Variable

My Y variable concerns the number of deaths that have occurred on a particular track from a sample of 17 tracks that have been used. This variable has been chosen for a number of reasons, mainly due to data collection problems. Originally time-series data concerning the number of deaths per year was to be used. This didn't vary much from year to year, and in some cases the number was zero. This would make it very difficult to find any correlation with other independent variables. The next approach was to consider including serious accidents for each year. Two problems emerged. Firstly, although the data was available, the sheer volume would have prompted a tedious task as well as the potential problem in inconsistency. Secondly, there was the problem of defining what was meant by a "serious accident." Analyzing the data in a cross-sectional format seemed to be the option that would avoid these problems and would also allow for a larger number of possible X variables.

The Data

The data was taken from two books and the Internet. The Deaths have been recorded since the start of Formula 1. As can be seen from the diagram below, quite a large number of these have occurred on the old Nurburing. In fact all of the numerical data that was observed on this circuit outweighed the others by a significant amount. A test for stability could be carried by running two separate regressions using the same data. This would show whether this observation is influencing the calculated R-squared values.

It must also be noted that cars and tracks change every season and that this factor would have to be taken account of. Perhaps then, the model is not as sound as initially believed, as the need for constancy is necessary in such a model.

This is the list of tracks used, where they are and the number of deaths that occurred on them:

TRACK	DEATHS
Old Nurburing (Germany)	6
New Nurburing (Germany)	1
Magny-Cours (France)	1
Spa Francorchamps (Belgium)	3
Monza (Italy)	4
Monte Carlo (Monaco)	2
Kyalami (South Africa)	2
Gilles Villeneuve (Canada)	2
Imola (San Marino)	3
Interlagos (Brazil)	1
Silverstone (Britain)	1
Catalunya (Spain)	1
Hungaroring (Hungary)	1
Suzuka (Japan)	1
Adelaide (Austrailia)	1
Estoril (Portugal)	1
Oscar Alfredo Galvez (Argentina)	1

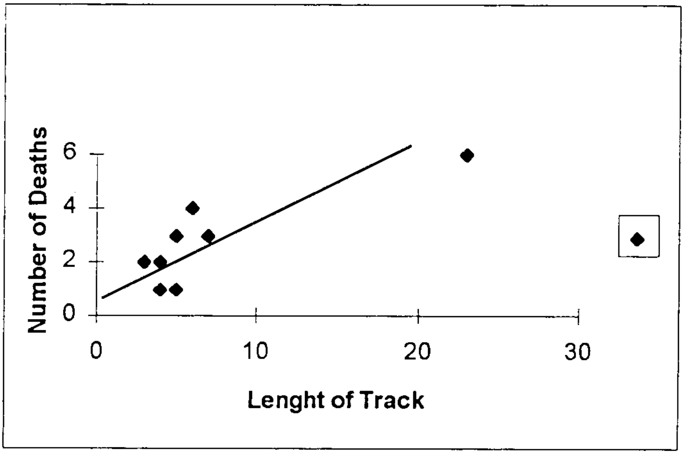
The Independent Variables

Much of Formula One is concerned with the design, speed and safety of the car itself. With top speeds of 340km per hour, and using up to 100 litres of fuel per 100km at top speed, the strict safety standards are well justified. Some of the commentary during a race will be concerned with the circuit that is being used, often with safety in mind. Some circuits are known to be more dangerous than others. My independent variables are concerned with different aspects of the track that could be potentially dangerous, and hence may explain the differences in deaths across different circuits.

Variable One (X_1)

The first variable in this test is the length of each track in kilometers. Any driver must devote a good proportion of his time to learning the layout of the track and the different maneuvers needed to negotiate its shape. A corner or turn can often be misjudged by a matter of a one hundredth of a second, resulting in a crash that can be fatal for both driver and spectators. A longer track is hypothesized to be more difficult to learn, increasing the likelihood of error.

Y regressed on X_1



As can be seen from this scatter diagram there is a positive relationship between the two variables. (The difference in the values for the Old Nurburing are reflected in the point that is plotted on it own in the top right hand corner of the graph). This relationship can be seen more clearly by examining the regression that has been done:

Multiple R	.80401	Adjusted R Square	.62286
R Square	.64643	Standard Error	.86530

Variable	B	SE B	Beta	T	Sig T
Length	0.254	0.0485	0.804	5.237	0.0001
(Constant)	0.421	0.3491		1.206	0.2464

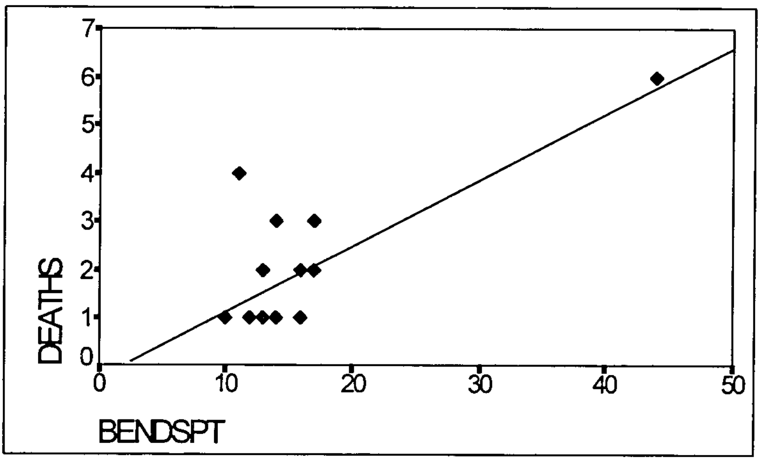
The R squared value of 0.646 tells that over 64% of the changes in Y can be explained by the changes in X_1. The regression line can be plotted by substituting our results into the following line: $Yi = a + \beta Xi$

The line can then be estimated as follows : $Y_1 = 0.421 + 0.804\ (X_1)$
The t-value is significant for a probability of less then 0.001 which indicates that the model is significant.

Variable Two (X_2)

The second variable examines the number of bends and corners that are incorporated in a circuit's design. The reasons for this variable are similar to the reasons for my first variable. The relationship is again believed to be positive. The greater the number of corners and bends, the greater the chance of driver error. Sharp corners often mean that bottlenecks are created. These are potentially dangerous as the cumulation of cars slowing down and then accelerating could easily hit one another and cause a serious accident.

Y regressed on X_2



Multiple R	.74229	Adjusted R Square	.52105
R Square	.55099	Standard Error	.97512

Variable	B	SE B	Beta	T	Sig T
Length	0.137	0.032	0.742	4.29	0.0006
(Constant)	-0.262	0.553		-o.475	0.6414

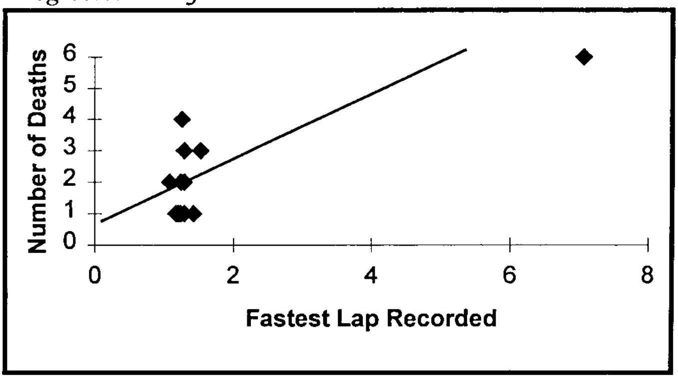
By looking at the regression that was performed it is clear that there is a positive relationship between these two variables. The R squared illustrates that about

55% of the changes in Y can be explained by the changes in X_2. This result is also significant with a probability consistent with X_1.

Our line can be estimated as follows: $Y_2 = -0.26289 + 0.742285(X_2)$

Variable Three (X_3)
The third and final variable that has been used in this analysis concerns the speed of the race. The figures used are the fastest ever lap times that have been recorded on each circuit during qualifying. Fast and slow circuits are quite a common description in Formula One. The speed can depend on a number of factors such as the angle between corners, how close they are together and the number of straight stretches in the track. This suggests a negative relationship between the number of deaths and the shortest time needed to complete a lap. The quicker the lap, the greater the potential seriousness of any accident.

Y regressed on X_3



Multiple R	.76513	Adjusted R Square	.55778
R Square	.58542	Standard Error	.93698

Variable	B	SE B	Beta	T	Sig T
FastestLap	0.764	0.166	0.765	4.602	0.0003
(Constant)	0.659	0.349		1.886	0.0788

R2 is high, but the hypothesis must be rejected as the relationship has proved positive. This indicates that the changes in Deaths are not explained by a quicker lap time.
Our line can be estimated as follows: $Y_3 = .659448 + .76513(X_3)$

Multiple Regression on all Three Variables:
The following is the multiple regression on all three variable simultaneously.

Multiple R	.85623	Adjusted R Square	.67155
R Square	.73313	Standard Error	.80751

Variable	B	SE B	Beta	T	Sig T
Fastest lap	-3.119	1.522	-3.122	-2.049	0.0612
Lengthot	0.959	0.358	3.034	2.679	0.0189
Bendspt	0.170	0.116	0.918	1.474	0.1643
(Constant)	-1.299	1.048		-1.283	0.2376

The estimated R square is quite a strong result when multiple regression has been performed on the three variables. The line can be estimated as follows:

$$\mathbf{Y = -1.298565 + (-)3.122281(X_1) + 3.034716(X_2) + .918981(X_3)}$$

Problems with the Variables and Data:
There are many problems with the variables due to the nature of the data and experiment. It is common practice for circuits to have their layout redesigned from year to year. A particular bend may be taken out and replaced by another. Data describing these changes was unavailable, and would have made this project impossible to estimate. The layouts that have been used are those that were used in the 1996 Championship. In some cases tracks are completely redesigned. Some data was available, (i.e. The old and the new Nurburing.) but much of it was not. This could seriously affect the consistency of the test, probably changing some of the properties of the model.

The data came from books and Internet pages (see bibliography). There were small problems with both of these sources. Information had slight variations, and some of the sources were more up to date than others. This meant that the information had to be clarified on another website to check its accuracy.

Another likely problem is the possible relationship between the X Variables. This may result in multicollinearity. In the multiple regression we can see that X3, the fastest lap, has become negative. This result could rest in the theory that the fastest lap is likely to be a function of the length and the bends on the track. However, this will not affect multiple regression

Conclusion:
From the regression analysis that has been conducted it is easy to see that there is a significant relationship between the dependent variable and the independent

variables. The estimates suggest that my hypothesis should be accepted. But the story does not end there. The problems with the variables and the nature of the data are too large to ignore, and must be addressed in full to gain any worthwhile conclusion to this model.

Bibliography:

Tremayne, D. (1996) "Formula One" Parragon
William's Renault (1994), "Formula 1 Motor Racing" Dorling Kindersley.

Internet Sources

http://www.u-net.com/enigma/formula1/circuits/countrygphistory.html
http://Student-www.eng.hawaii.edu/carina/formula 1
grandprixdeaths.html