FSA 327 FIRE RELATED HUMAN BEHAVIOR

WESTERN OREGON UNIVERSITY

CORRELATION AND REGRESSION AN EXAMPLE WITH POPULATION AND EMS CALLS FOR THE CITY OF BRAMPTON

Len Papp, CET, GIFireE, CFEI
Fire Prevention Officer
Brampton Fire & Emergency Services
Phone: (416) 722-8164
e-mail: len.papp@city.brampton.on.ca

PROFESSOR: Howard A. Reeher

Issued: July 7, 2004

Revised: September 14, 2005

Revised: October 12, 2005

Table of Contents

Introduction

Correlation is the degree of relationship between two variables. For example, everyone would agree that there is a high correlation between smoking and lung cancer. If we compared the number of smokers with the incidence of lung cancer we would find that as the number of smokers increases, the incidence of lung cancer also increases. The relationship between smoking and lung cancer is called a positive correlation. A negative correlation would show the opposite effect.

The hypothesis of this paper will be to investigate through statistical analysis utilizing correlations if a positive correlation exists between population and EMS calls. If a positive relationship exists, regression analysis will be used to predict future EMS calls based on population.

Part 1

POPULATION AND EMS CALLS 1999 TO 2003

Exhibit 1: Population & EMS calls Brampton Fire & Emergency Services

Year	Population	EMS calls
1999	299,000	4700
2000	311,000	5226
2001	325,000	5596
2002	352,983	6360
2003	371,939	6490
2004	386,539	

Population Mean (average) = 331 984.4

EMS Mean (average) = 5674.4

Calculating the Variance, Standard Deviation and Average (1999-2003)

The data from exhibit 1 will be used to calculate the following.

Exhibit 2:
Averages and Standard Deviations for Selected Years(1999 – 2003)

	Average	**Standard Deviation**
Population (in thousands)	331 984.4	30 059.98
EMS calls	5674.4	756.95

To calculate the Variance and Standard Deviation for population and EMS calls the following calculations are performed:

$$\text{Sample Variance} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\text{Sample Standard Deviation} = s = \sqrt{\frac{\sum (x_i - x)^2}{n-1}}$$

Calculating the Variance and Standard Deviation (1999-2003)

Population:

			Squared
299 000 – 331 984.4	=	-32 984.4	1.0879 09
311 000 – 331 984.4	=	-20 984.4	4.4034 08
325 000 – 331 984.4	=	- 6 984.4	48 781 843
352 983 – 331 984.4	=	20 998.6	4.4094 08
371,939 – 331 984.4	=	39 954.6	1.5963 09
Total			3.6144 09 / 4
		Variance =	9.036 08

The square root of the variance is the sample standard deviation.

Therefore, the standard deviation for population is 30 059.98 since this is the square root of 9.036 08.

EMS calls:

			Squared
4700 – 5674.4	=	-974.4	949 455.36
5226 – 5674.4	=	-448.4	201 062.56
5596 – 5674.4	=	-78.4	6 146.56
6360 – 5674.4	=	685.6	470 047.36
6490 – 5674.4	=	815.6	665 203.36
Total			229 191.2 / 4
		Variance =	572 978.8

Therefore, the standard deviation for population is 756.95 since this is the square root of 572 978.8.

Part 2

Calculating the Correlation

To calculate the correlation the following steps will be performed:

1. Convert the values of both variables to standard units.
2. Take the product of the standard unit for each pair.
3. Sum the resulting values and divide by the number of points minus 1.

Standard Units = $\frac{xi - \bar{x}}{s}$

$\bar{x}$ = arithmetic mean (average)

s = sample standard deviation

xi = independent variable

Population Standard Units:

$\frac{299\,000 - 331\,984.4}{30\,059.98} = -1.09$

$\frac{311\,000 - 331\,984.4}{30\,059.98} = -0.70$

$\frac{325\,000 - 331\,984.4}{30\,059.98} = -0.23$

$\frac{352\,983 - 331\,984.4}{30\,059.98} = 0.69$

$\frac{371\,939 - 331\,984.4}{30\,059.98} = 1.32$

Calculating the Correlation

EMS Standard Units:

$\frac{4700 - 5674.4}{756.95} = -1.28$

$\frac{5226 - 5674.4}{756.95} = -0.59$

$\frac{5596 - 5674.4}{756.95} = -0.10$

$\frac{6360 - 5674.4}{756.95} = 0.90$

$\frac{6490 - 5674.4}{756.95} = 1.07$

Exhibit 3: Correlation Calculation

Population	EMS calls	Population Standard Units	EMS Standard Units	Product
299 000	4700	-1.09	-1.28	1.39
311 000	5226	-0.70	-0.59	0.41
325 000	5596	-0.23	-0.10	0.02
352 983	6360	0.69	0.90	0.62
371 939	6490	1.32	1.07	1.41
Total				3.85
Correlation (Total divided by 4) = **.96**				

Calculating the Correlation

Therefore, a correlation of **.96** exists between EMS calls and population.

Based on this statistical information we can conclude that this high correlation indicates a strong association between the two variables.

Calculating the Regression Line

The regression line has the general form:

$$y = mx + b$$

where y is the dependent variable (EMS calls) and x is the independent variable (population), m is the slope (regression coefficient, and b is the intercept (or constant).

I will now calculate the slope and intercept based on these variables.

In exhibit 4 I have placed a dot at the pair formed by the two averages (331 984.4 population and 5674.4 EMS calls). We want the regression line to pass through this point. I get a second dot by moving one standard deviation for the population to the right and upward by one standard deviation for EMS calls **times** the correlation. The regression line is then formed graphically by drawing a straight line through these two points.

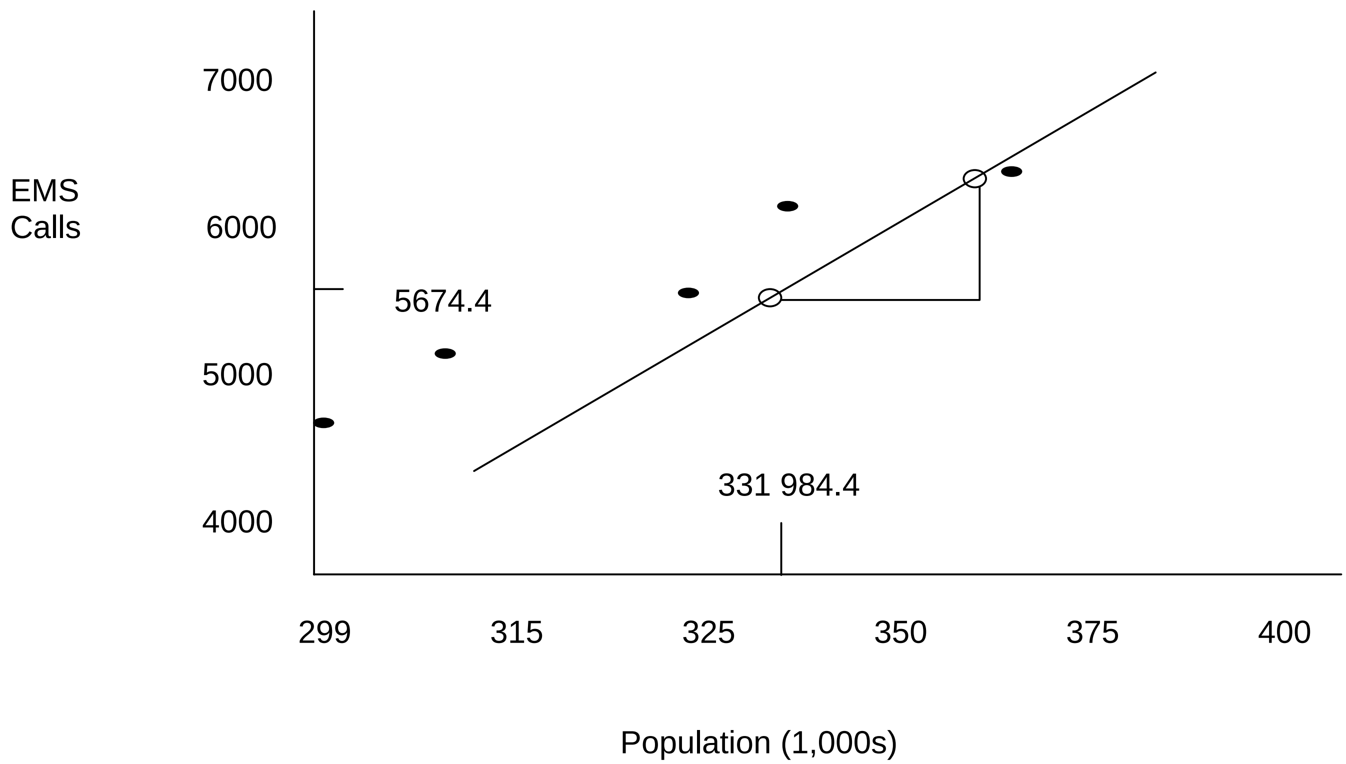
Algebraically, the slope of the straight line is:

$$\text{Slope} = \frac{\text{r x S.D. of EMS calls}}{\text{S.D. of population}} = \frac{.96 \times 756.95}{30\ 059.98}$$

Slope = **0.02**

Calculating the Regression Line

Exhibit 4
Scatter Diagram of Population & EMS calls = ●
Graphical Formulation of Regression Line = ○



Calculating the intercept takes advantage of the fact that the regression line goes through the point created by the two averages. Therefore, we determine the intercept, b, by:

EMS calls average = m x Population average + b

5674.4 = 0.02 x 331 984.4 + b

-965.28 = b

The final result is the regression line:

EMS calls = 0.02 x Population -965.28

Calculating the Regression Line

To see how to use this regression, we look at our 2000 population of 311 000 for the City of Brampton. The number of EMS calls estimated by the regression line is then:

EMS calls = (0.02 x 311 000) – 965.28

= 5254

Based on the statistics from our communication department Brampton actually experienced 5226 EMS calls, a difference of 28.72 calls. Therefore, the regression line was off by 28.72 EMS calls or **0.5 %.**

Checking our statistics, for our population average of 331 984.4 we have the following from the regression line:

EMS calls = (0.02 x Population) – 965.28

= (0.02 x 331 984.4) – 965.28

= 5674.4

The result is 5674.4, which is the average number of EMS calls for the years 1999 to 2003.

The regression line can also be used to estimate the number of EMS calls in the future. The following chart is an estimate of the projected population growth for the City of Brampton. How many EMS calls can be expected with this projected population?

Part 3

Estimated EMS calls using Regression Line

We can estimate the future EMS calls with the following calculation:

EMS calls = (0.02 x 500 000) –965.28

= 9034.72

Therefore, the City of Brampton can estimate running 9000 EMS calls with a population of 500 000.

Projected population	Projected EMS calls
386 539 (2004 estimate)	6765.5
450 000	8034.72
500 000	9034.72
550 000	10 034.72
600 000	11 034.72
650 000	12 034.72

Summary/Recommendations

Correlation and regression analysis are two very powerful tools for the fire service administrator for analyzing the relationship between two continuous variables. The example utilized in this paper between population and EMS calls showed a positive relationship as the correlation between these two variables were .96. Correlations close to –1 or +1 indicate a strong relationship between two variables.

Regression analysis provides a way to quantify the relationship of two variables. We can apply the regression line to estimate the value of one variable given the other variable. With the regression of population on EMS calls I have shown that we can make estimates on the number of EMS calls that the City of Brampton Fire and Emergency Services can expect given it's projected population.

It is recommended to fire administration that correlation and regression analysis is utilized to determine future fire safety inspections and fires based on the number of buildings within the Municipality. This will assist the fire service administrator in determining the amount of staff required in the future to complete these duties.

List of Works Cited

Federal Emergency Management Agency, United States Fire Administration. Fire Data Analysis Handbook, 1999

Federal Emergency Management Agency, United States Fire Administration. Fire Related Human Behavior Course Guide, 1999

Schefler, William C. Statistics: Concepts and Applications. Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc., 1988