

交通事故の因子について分析

平成20年1月21日

早稲田大学商学部

＝目次＝

1. はじめに
2. 仮説やモデル式
3. 回帰式全体の分析・考察
4. 個々の説明変数に対する分析・考察
5. サンプルに対する分析・考察
6. 今回使用したデータについて
7. TSPのアウトプット

1. はじめに

最近、飲酒運転による交通事故に対するマスメディアや世間の注目が高まりつつある。結果、飲酒運転に対する国民意識は大いに喚起されたが、いまだに一部の者によって悲劇的に引き起こされる事故は後を絶たない。

そこで、交通事故はなぜ引き起こされるのかということへの原点に振り返ってみる。交通事故が多い場所としてよく言われるのは大都市圏である。しかし、報道では飲酒運転は電車などの公共交通機関が発達していない地方が多い。6章に掲載している統計を見ても、地方ではなく大都市における事故が多そうだ。

今回は、大都市における事故が多さに着目し、交通事故が人口密度によって説明できるのではないかと考える。しかし、当たり前だが、車が多いところでも交通事故は多いはずなので、これも説明変数に加えて分析する。

2. 仮説やモデル式

仮説：「交通事故は人口密度や車の保有台数によって説明できる。」

この仮説に基づいて、モデル式を次のようにした。

交通事故数を被説明変数 A 、説明変数を自動車の保有台数を M 、人口密度を P とすれば、

$$A = \alpha + \beta M + \gamma P \quad (\alpha, \beta, \gamma \text{ は実数の定数})$$

これに対し、以下のデーターを適用した。

平成14年度の都道府県別統計 () は出展元

- ・ A : 交通事故件数 (警察庁)
- ・ M : 自動車保有台数 (財団法人自動車検査協力会)
- ・ P : 1 km²当たりの人口密度 (総務省統計局)

TSPで回帰分析した結果、以下のような結果が得られた。

標本数: 47

DW値= 1.81915 [<.360]

F 値= 656.936 [.000]

自由度調整済み決定係数= 0.966123

Coefficient	誤差	t-statistic	P-value
$\alpha = -4541.68$	907.524	-5.00447	[.000]
$\beta = 0.010929$.621735E-03	17.5782	[.000]
$\gamma = 4.76593$.447141	10.6587	[.000]

3. 回帰式全体の分析・考察

回帰式全体に関する基本統計量は以下になる。

標本数: 47 説明変数: 2
DW値= 1.81915 [$<.360$]
F 値= 656.936 [$.000$]
自由度調整済み決定係数= 0.966123

自由度調整済み決定係数は1に限りなく近く、回帰式の当てはまりは極めてよい。DW検定を行おう。サンプル数47、被説明変数の数2より、表から $L = 1.43 \sim 1.46$ 、 $U = 1.62 \sim 1.63$ なる。DW値 $1.82 > U$ より、残差に正の系列相関がないと判断された。さらに、DW値 $< 4 - U$ でもあるので、負の系列相関も存在しない。よって、DW検定の結果は有意である。F検定も行おう。サンプル数47、被説明変数の数2より、分子自由度 $2-1=1$ 、分母自由度 $47-2=45$ であるから、F検定の分布表より1%の有意水準で臨界値は7~8である。回帰分析の結果のF値は約657であり、明らかに有意である。

さらにホワイト検定も実施しよう。今回の残差は以下のようになった。

1	-8296.64174	13	-2344.281087	25	-718.1227472	37	2863.741812
2	641.3893904	14	1177.245726	26	-1973.189126	38	716.7068355
3	-1990.68509	15	-2315.123255	27	-5161.771044	39	405.67952
4	-2973.051184	16	-154.8930014	28	6213.779699	40	11635.34254
5	-1159.232485	17	-4095.406036	29	-3166.269055	41	5067.887902
6	1640.449317	18	9358.808261	30	711.7313928	42	-1299.275937
7	823.7206908	19	389.8258726	31	-474.9759031	43	529.9964284
8	735.743282	20	224.070384	32	-930.450329	44	-296.6175083
9	242.9055978	21	-761.8947515	33	5999.590498	45	-295.5796998
10	-336.1574008	22	-3627.865684	34	1282.626758	46	863.4473363
11	882.365069	23	-3708.690614	35	-1084.519377	47	-4795.076044
12	1725.630987	24	-3054.068857	36	881.1526554		

これを被説明変数とし、従来の説明変数で回帰分析を実施した（このアウトプットはあとで掲載している）。結果、F値は「.933996E-14 [1.00]」はほぼ0と出力された。よって、帰無仮説は採択され、分散不均一とみなされた。

4. 個々の説明変数に対する分析・考察

回帰式全体に関する基本統計量は以下になる。

Coefficient	誤差	t-statistic	P-value
$\alpha = -4541.68$	907.524	-5.00447	[.000]
$\beta = 0.010929$.621735E-03	17.5782	[.000]
$\gamma = 4.76593$.447141	10.6587	[.000]

まずは係数の値を見よう。 β 、 γ ともに正であり、交通事故は人口密度や車の保有台数によって説明できるとする仮説が実証された。今回用いたデータのうち、事故件数と自動車保有台数は指数化された値ではないので、0.01 という結果は、自動車保有台数 100 台に対して、年間で、新たに 1 件の交通事故が引き起こされるということの意味している。

次に t 検定も行おう。P 値をみれば明らかだが、t 分布の t 値から分布の裾までを積分区間とした面積は、いずれの係数においてもきわめて 0 に近い。よって、有意水準を 1 % として設定しても、t 検定は有意であると結論づけられた。

5. サンプルに対する分析・考察

第3章で取り上げた残差を標準化したものは以下である。

1	北海道	-2.322307519	17	長野	-1.146342403	33	岡山	1.67934142
2	青森	0.179530881	18	静岡	2.619617849	34	広島	0.35901921
3	岩手	-0.557211351	19	富山	0.109115903	35	山口	-0.303567104
4	宮城	-0.832184796	20	石川	0.062719393	36	徳島	0.246642859
5	秋田	-0.324480001	21	福井	-0.213261457	37	香川	0.801588082
6	山形	0.459177087	22	岐阜	-1.015473492	38	愛媛	0.200612938
7	福島	0.230567115	23	愛知	-1.038097144	39	高知	0.113553487
8	東京	0.205941417	24	三重	-0.8548624	40	福岡	3.256841059
9	茨城	0.067991546	25	滋賀	-0.201009265	41	佐賀	1.418549161
10	栃木	-0.094093596	26	京都	-0.552314067	42	長崎	-0.363679471
11	群馬	0.246982225	27	大阪	-1.444827929	43	熊本	0.148350951
12	埼玉	0.483020232	28	兵庫	1.739294978	44	大分	-0.083026011
13	千葉	-0.656186173	29	奈良	-0.886268284	45	宮崎	-0.082735519
14	神奈川	0.329522075	30	和歌山	0.199220265	46	鹿児島	0.241686975
15	新潟	-0.648024624	31	鳥取	-0.132950192	47	沖縄	-1.342186574
16	山梨	-0.043355998	32	島根	-0.260441738			

これを見れば、全国のデータを使った回帰式に対して、それぞれの都道府県がどれほど乖離しているか判明する。とくに、北海道・静岡・福岡は著しい。福岡の理由は不明だが、静岡は東名自動車道の通過区間によって、モデル式より正の方向に事故件数が乖離し、北海道はあまりにも幅が広く交通量も少ない道路事情によって、モデル式より負の方向に乖離したのではないかと推察される。

6. 今回使用したデータについて

平成14年度の都道府県別統計（ ）は出展元

- ・ 交通事故件数（警察庁）
- ・ 自動車保有台数（財団法人自動車検査協力会）
- ・ 1k㎡当たりの人口密度（総務省統計局）

	都道府県名	事故件数	自動車保有台数	人口密度
1	北海道	28674	3685295	259.3
2	青森	9126	991262	460.1
3	岩手	5928	973974	381
4	宮城	12864	1534362	757.4
5	秋田	5082	822973	375.3
6	山形	9134	911395	435.3
7	福島	15434	1532703	503.8
8	東京	88512	4654863	8696.1
9	茨城	24699	2325140	752.5
10	栃木	15450	1562455	682.3
11	群馬	18936	1681535	885
12	埼玉	51617	3794603	2719.7
13	千葉	37788	3341524	1711
14	神奈川	67660	3938094	5871.9
15	新潟	14923	1752262	551.7
16	山梨	7534	710748	936.4
17	長野	14543	1830151	666.9
18	静岡	41082	2714518	1384.4
19	富山	8082	855217	605.8
20	石川	8998	845613	854.8
21	福井	5261	627254	778.3
22	岐阜	13976	1596991	984.5
23	愛知	55552	4791092	2400.5
24	三重	12048	1395899	920.7
25	滋賀	9766	917103	1049.7
26	京都	19174	1351594	2290.7

27	大阪	63802	3799757	6709.7
28	兵庫	43064	2905856	2021.4
29	奈良	9250	812185	1695.7
30	和歌山	8797	731579	971.8
31	鳥取	3075	447207	672.3
32	島根	3258	534714	605.6
33	岡山	21344	1434339	883.4
34	広島	22344	1785842	1276.9
35	山口	9905	1041231	871.1
36	徳島	6793	605537	804.8
37	香川	11272	735142	1031.4
38	愛媛	11198	985309	892.7
39	高知	5362	565683	695.7
40	福岡	49348	3065008	1837.4
41	佐賀	10443	622187	654
42	長崎	8420	896833	935.7
43	熊本	12964	1257783	677.6
44	大分	7778	853536	689.9
45	宮崎	7818	880257	636.8
46	鹿児島	12884	1275640	549.9
47	沖縄	5759	881509	1146

7. TSPの出力結果

```

-----
|           this copy licensed           |
|           for use by:                   |
| TSP/GiveWin User      #45AGT1103 |
-----

```

TSP Version 4.5
(11/07/03) TSP/GiveWin 4MB
Copyright (C) 2003 TSP International
ALL RIGHTS RESERVED
01/22/08 12:30PM

In case of questions or problems, see your local TSP consultant or send a description of the problem and the

associated TSP output to:

TSP International
P.O. Box 61015
Palo Alto, CA 94306
USA

PROGRAM

COMMAND *****

```
| 1 freq n;  
| 2 smpl 1 47;  
| 3 load;  
| 4 regopt(pvprint) t fst;  
| 5 olsq a c m p;  
| 6 end;
```

EXECUTION

Current sample: 1 to 47

```
| 3 load a m p;  
| 3 28674 3685295 259.3  
| 3 9126 991262 460.1  
| 3 5928 973974 381  
| 3 12864 1534362 757.4  
| 3 5082 822973 375.3  
| 3 9134 911395 435.3  
| 3 15434 1532703 503.8  
| 3 88512 4654863 8696.1  
| 3 24699 2325140 752.5  
| 3 15450 1562455 682.3  
| 3 18936 1681535 885  
| 3 51617 3794603 2719.7  
| 3 37788 3341524 1711  
| 3 67660 3938094 5871.9  
| 3 14923 1752262 551.7  
| 3 7534 710748 936.4  
| 3 14543 1830151 666.9  
| 3 41082 2714518 1384.4
```


| 3 8082 855217 605.8
| 3 8998 845613 854.8
| 3 5261 627254 778.3
| 3 13976 1596991 984.5
| 3 55552 4791092 2400.5
| 3 12048 1395899 920.7
| 3 9766 917103 1049.7
| 3 19174 1351594 2290.7
| 3 63802 3799757 6709.7
| 3 43064 2905856 2021.4
| 3 9250 812185 1695.7
| 3 8797 731579 971.8
| 3 3075 447207 672.3
| 3 3258 534714 605.6
| 3 21344 1434339 883.4
| 3 22344 1785842 1276.9
| 3 9905 1041231 871.1
| 3 6793 605537 804.8
| 3 11272 735142 1031.4
| 3 11198 985309 892.7
| 3 5362 565683 695.7
| 3 49348 3065008 1837.4
| 3 10443 622187 654
| 3 8420 896833 935.7
| 3 12964 1257783 677.6
| 3 7778 853536 689.9
| 3 7818 880257 636.8
| 3 12884 1275640 549.9
| 3 5759 881509 1146

Equation 1

=====

Method of estimation = Ordinary Least Squares

TSP Version 4.5
(11/07/03) TSP/GiveWin 4MB
Copyright (C) 2003 TSP International
ALL RIGHTS RESERVED
01/22/08 1:25 PM

In case of questions or problems, see your local TSP
consultant or send a description of the problem and the
associated TSP output to:

TSP International
P.O. Box 61015
Palo Alto, CA 94306
USA

PROGRAM

COMMAND *****

```
| 1 freq n;  
| 2 smpl 1 47;  
| 3 load;  
| 4 regopt(pvprint) t fst;  
| 5 olsq u c m p;  
| 6 end;
```

EXECUTION

Current sample: 1 to 47

```
| 3 load u m p;  
| 3 -8296.64174 3685295 259.3  
| 3 641.3893904 991262 460.1  
| 3 -1990.68509 973974 381  
| 3 -2973.051184 1534362 757.4  
| 3 -1159.232485 822973 375.3  
| 3 1640.449317 911395 435.3  
| 3 823.7206908 1532703 503.8  
| 3 735.743282 4654863 8696.1  
| 3 242.9055978 2325140 752.5  
| 3 -336.1574008 1562455 682.3  
| 3 882.365069 1681535 885
```

| 3 1725.630987 3794603 2719.7
| 3 -2344.281087 3341524 1711
| 3 1177.245726 3938094 5871.9
| 3 -2315.123255 1752262 551.7
| 3 -154.8930014 710748 936.4
| 3 -4095.406036 1830151 666.9
| 3 9358.808261 2714518 1384.4
| 3 389.8258726 855217 605.8
| 3 224.070384 845613 854.8
| 3 -761.8947515 627254 778.3
| 3 -3627.865684 1596991 984.5
| 3 -3708.690614 4791092 2400.5
| 3 -3054.068857 1395899 920.7
| 3 -718.1227472 917103 1049.7
| 3 -1973.189126 1351594 2290.7
| 3 -5161.771044 3799757 6709.7
| 3 6213.779699 2905856 2021.4
| 3 -3166.269055 812185 1695.7
| 3 711.7313928 731579 971.8
| 3 -474.9759031 447207 672.3
| 3 -930.450329 534714 605.6
| 3 5999.590498 1434339 883.4
| 3 1282.626758 1785842 1276.9
| 3 -1084.519377 1041231 871.1
| 3 881.1526554 605537 804.8
| 3 2863.741812 735142 1031.4
| 3 716.7068355 985309 892.7
| 3 405.67952 565683 695.7
| 3 11635.34254 3065008 1837.4
| 3 5067.887902 622187 654
| 3 -1299.275937 896833 935.7
| 3 529.9964284 1257783 677.6
| 3 -296.6175083 853536 689.9
| 3 -295.5796998 880257 636.8
| 3 863.4473363 1275640 549.9
| 3 -4795.076044 881509 1146

Equation 1

=====

Method of estimation = Ordinary Least Squares

Dependent variable: U

Current sample: 1 to 47

Number of observations: 47

Mean of dep. var. = .194793E-05 LM het. test = 5.49225 [.019]
Std. dev. of dep. var. = 3494.06 Durbin-Watson = 1.81915 [<.360]
Sum of squared residuals = .561588E+09 Jarque-Bera test = 19.1985 [.000]
Variance of residuals = .127634E+08 Ramsey's RESET2 = 5.84706 [.020]
Std. error of regression = 3572.59 F (zero slopes) = .933996E-14 [1.00]
R-squared = .107767E-15 Schwarz B.I.C. = 455.424
Adjusted R-squared = -.045455 Log likelihood = -449.649

	Estimated	Standard		
Variable	Coefficient	Error	t-statistic	P-value
C	-.423816E-04	907.524	-.467003E-07	[1.00]
M	.425494E-10	.621735E-03	.684365E-07	[1.00]
P	-.187565E-07	.447141	-.419476E-07	[1.00]

END OF OUTPUT.