

공식통계자료의 표현방법

홍종선¹⁾ 임한승²⁾

요약

공공기관에서 발간하는 공식통계자료들을 살펴보면 대부분 관찰값으로 총 빈도수나 또는 전체를 기준으로 하여 그 빈도수가 차지하는 퍼센트 그리고 지수 등으로 나타나 있다. 이러한 자료는 단순히 공무원들에게 행정용으로 활용되고는 있으나 일반인들이 자료를 이해하고 나아가 활용하기는 어렵다. 이런 자료들이 일반인을 위한 자료가 되기 위해서는 국민 한사람(또는 기본 단위)당 그 발생 확률을 구하여 제시하고 나아가 개개인의 여러 복잡한 현실 상황을 고려해도 그 확률 계산이 용이하도록 기초적인 자료를 제공하는 것이 바람직하다고 사료된다. 즉, 육하(六何)원칙을 근거로한 현상에 대하여 확률을 구하고 활용할 수 있는 방안을 제시한다. 이 논문에서는 경찰청에서 발표된 교통사고에 대한 통계자료와 대검찰청에서 발표된 범죄사건 통계자료를 통계학의 기본인 확률의 개념을 도입하여 보다 이해가 쉽고, 나아가 교통사고와 범죄 피해를 최소한으로 줄일 수 있는 자료로 변환하여 설명하고자 한다.

1. 서론

현재 우리나라에서 집계되어 발간되는 공식통계자료(official statistical data)는 많은 종류가 있다. 대표적인 범죄(crime)와 교통사고(car accidents)에 관한 공식통계자료를 살펴보고자 한다. 우선 공식범죄자료로는 대검찰청에서 1963년부터 분기별로 발간하다가 1995년부터 매년 발간하는 「범죄분석」, 경찰청에서 매년 발간하는 「범죄분석」, 법무연수원에서 매년 발간하는 「범죄백서」, 문화체육부에서 매년 발간하는 「청소년백서」 등이 있으며, 또한 교통사고와 관련된 공식통계자료는 경찰청에서 매년 발간하는 「교통사고통계」, 도로교통안전협회에서 발간하는 「교통사고통계분석」, 교통안전공단에서 발간한 「교통판례집」, 건설교통부에서 매년 발간하는 「건설교통통계연보」와 「교통안전연차보고서」 등이 있다.

이러한 공식통계자료는 기본적으로 '얼마나 많은 건수(범죄나 교통사고 등)가 발생하는가'를 중심으로 제시되고 있는데 이는 수사기관이나 행정기관의 독자적인 목적을 우선하여 작성되고 있다. 따라서 사회과학적 연구를 위한 통계자료나 혹은 일반인들이 쉽게 이해할 수 있는 통계로는 한계가 있다. 이런 한계를 극복하기 위해 인구 10만명당 범죄발생건수를 계산하는 범죄율(crime rate)을 사용한다. 범죄율은 인구 대비로 범죄발생 건수를 비교한다는 점에서 유용한 자료이지만 중요한 범죄와 가벼운 범죄가 동등한 범죄로 취급된다는 문제점이 지적된다. 이러한 문제점을 해결하기 위하여 중요범죄(index crime)의 중요도를 구분하여 범죄율 조사를 수행하곤 한다. 이 외에도 범죄시계(crime clock)라는 것이 있다. 이는 미국범죄통계(UCR)에 나오는 것으로 매시간의 범죄발생 현황을 표시한 것을 말

1) (110-745) 서울. 종로구 명륜동3가, 성균관대학교 경제학부 통계학전공, 교수, cshong@stat.skku.ac.kr.

2) (110-745) 서울. 종로구 명륜동3가, 성균관대학교 경제학부 통계학전공, 강사, mylims@dragon.skku.ac.kr.

한다. 예를 들어 몇 초마다 몇 건의 살인사건이 발생한다던가, 몇 분마다 몇 건의 강간사건이 발생한다고 하는 것이다. 이러한 범죄시계는 일반인들에게는 쉽게 이해할 수 있는 통계자료로써 범죄예방홍보 등에 많이 이용되고 있으나 인구성장률을 반영하지 않고 있으며, 시간을 고정적인 비교단위로 사용하는 문제점이 있기 때문에 통계적 가치는 없다고 할 수 있다 (박상기, 손동권, 이순래(1998)와 Sellin, Wolfgang(1964) 참조).

본 논문에서는 공공기관에서 발표하는 많은 공식통계자료를 국민이 피부로 느낄 수 있는 유용한 정보로 변환시켜서 사회과학적 연구를 위한 통계로 활용할 수 있는 두 가지 방법을 제안한다. 특히 여기에서는 1998년에 경찰청에서 발표된 교통사고에 대한 통계자료와 대검찰청에서 발표하는 범죄사건 통계자료를 통계학의 기본인 확률(probability)의 개념을 도입하여 논의하고자 한다. 확률은 통계학 개념의 강좌 초반부에서 공부한다. 이때 확률과 우리들의 일상생활과는 매우 밀접한 관계를 갖고 있으며, 모든 부분에서의 모든 의사결정에 중요한 역할을 한다고 배우고 익힌다. 그러나 현실 세계에서는 많이 활용하지 못하고 있는 실정이다. 이런 한계를 극복하기 위해 많은 종류의 공식자료들을 이해하기 쉽고 활용하기에 좋은 자료로 변환하는 방법으로 확률과 조건부 확률(conditional probability)을 사용하는 것을 제안하고자 한다.

예를 들어 모든 자동차 사고 가운데 승용차의 사고는 다른 종류보다 제일 많은 60% 이상을 차지한다(표 2.2 참조). 그러나 가장 많은 사고 점유율을 나타냄으로 가장 위험한 차종으로 간주한다는 것은 위험한 발상이다. 왜냐하면 승용차는 등록된 모든 차량종류에서 70% 이상을 점유하고 있기 때문이다. 그러므로 차종별 등록대수와 비교하여 사고율을 구하여야 한다. 그리고 "40대의 남자가 서울에서 강도를 당하는데 그 피해 액수가 100만원 이하일 경우"와 같이 현실 생활에서 발생하는 갖가지 경우의 확률을 계산할 수 있다면, 일반인들에게 자신이 어떤 범죄의 피해자가 될 확률 등을 알려줌으로서 범죄를 피하고 예방할 수 있는 행동지침을 제공해주는 유용한 정보가 될 것이다.

본 논문의 구성은 다음과 같다. 2절에서는 빈도수로 나타난 자료를 확률로 표현했을 때 결과 해석에 대하여 큰 변화가 있는 몇 가지 경우를 예를 들어 설명했으며, 3절에서는 주변 확률과 조건부 확률로 자료를 표현하고 조건부 독립(conditional independence)이라는 개념을 도입하여 육하(六何)원칙에 준한 현실 상황에 대한 확률을 구할 수 있는 방법을 예제와 함께 제공하였다.

2. 확률로 표현된 자료

확률의 개념을 도입하여 자료를 다시 표현해보면, 원자료(raw data)에서 발견 못한 현상을 발견하는 경우가 많은 것은 자명한 사실이다. 그럼에도 불구하고 공식통계자료에서는 잘 활용되지 않는 실정이다. 여기에서는 공식자료 중 교통사고와 범죄통계자료에 대하여 확률로 변환했을 때에 추가적인 정보와 인지할 수 있는 변화의 정도가 큰 경우의 몇 가지 예를 설명하고자 한다.

매년 경찰청에서 발간하는 「교통사고통계」를 살펴보면 표 2.1의 '사고수'와 같은 종류의 빈도수(frequency) 또는 지수(index) 등의 자료를 제공한다. 이런 종류의 숫자들로부터

는 약간의 변동만 있다는 결과 외의 다른 분석결과를 유도하기 어렵다. 그러나 매년 꾸준히 증가하고 있는 우리나라 총 인구수(통계청이 발표한 추계인구)나 등록된 자동차의 수(건설교통부 자료)를 사고건수와 비교하여 교통사고를 당할 확률을 구하여 표 2.1에 첨가하여 정리하였다.

표 2.1: 연도별 교통사고와 확률(백인율)

		1991	1992	1993	1994	1995	1996	1997
사고수	총	265964	257194	260921	266107	248865	265052	246452
	증감	4.2	-3.3	1.4	2.0	-6.5	6.5	-7.1
발생확률		0.614	0.588	0.590	0.596	0.552	0.582	0.536
자동차사고확률		6.261	4.916	4.159	3.594	2.939	2.775	2.367

자료: 1) 「교통사고통계」, 경찰청, 1998., 2) 「건설교통통계연보」, 건설교통부, 1998.

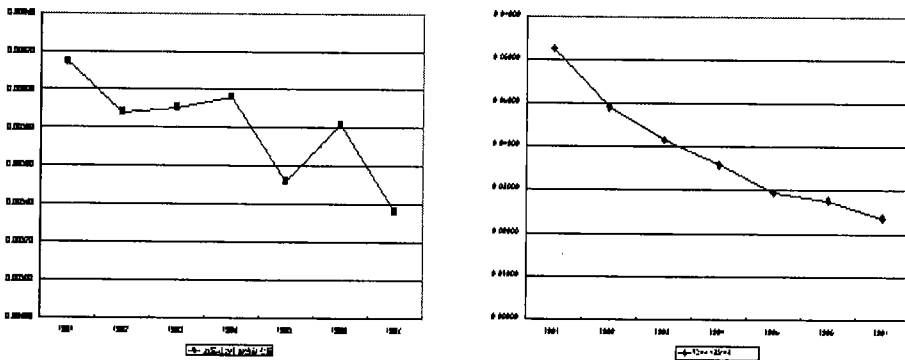


그림 2.1: 교통사고를 당할 확률과 자동차가 사고 날 확률

경찰청에서 발표된 공식통계자료는 일반인들에게는 큰 도움을 주지 못한다. 통계자료가 우리의 생활과 밀접한 관계를 갖기 위해서는 우리들이 생활하면서 느낄 수 있는 인지된 숫자를 제공하여야 하는데 표 2.1의 확률들을 살펴보면 대한민국 국민의 한 사람으로 교통사고를 당할 확률과 우리가 소유하고 있는 자동차가 사고 날 확률들을 파악할 수 있으며 이 값들은 매년 감소하고 있는 추세라는 사실을 발견할 수 있다(그림 2.1참조). 여기에서 우리는 1997년도에 한 사람이 교통사고를 당할 확률은 0.00536으로 십만명 중 536명이 교통사고를 당한다고 해석할 수 있으며 이는 $1/0.00536 = 187$ 명당 1명이 교통사고를 당한다는 것을 의미한다. 그리고 자동차는 42대당 한대 꼴로 사고가 발생한다는 것을 인지할 수 있다. 이러한 숫자의 개념은 해당 관청에서 발표하는 내용과는 전혀 다른 내용이며 일상생활에서 피부로 느낄 수 있는 유용한 정보로 활용될 수 있다.

이런 유사한 결과를 나타내는 몇 가지 사례를 분석하여보자. 우선 자동차 종별로 구분된 교통사고 발생 건수 자료와 분석내용은 표 2.2와 그림 2.2에 나타나 있다. 이 표를 살펴

보면 승용차 사고가 제일 많고 그 다음으로 화물차임을 알 수 있다. 그러나 등록된 자동차별 대수와 비교한 사고 확률을 살펴보면 오히려 제일 사고율이 낮은 차종이 승용차와 화물차임을 파악할 수 있다. 그리고 특수차의 사고율이 제일 높고 다음으로 승합차의 사고율순으로 높음을 알 수 있다. 1997년도에 승용차는 50대당 한대 꼴로 특수차는 26대당 한대 꼴로 사고가 발생한다.

표 2.2: 차종별 교통사고와 확률(백인율)

	승용차	승합차	화물차	특수차	이륜차
사고수	152909	23842	46370	1605	10790
등록대수	7586474	719127	2072256	35570	2540150
사고확률	2.0	3.3	2.2	3.9	0.43

자료: 1) 「교통사고통계」, 경찰청, 1998., 2) 「건설교통통계연보」, 건설교통부, 1998.

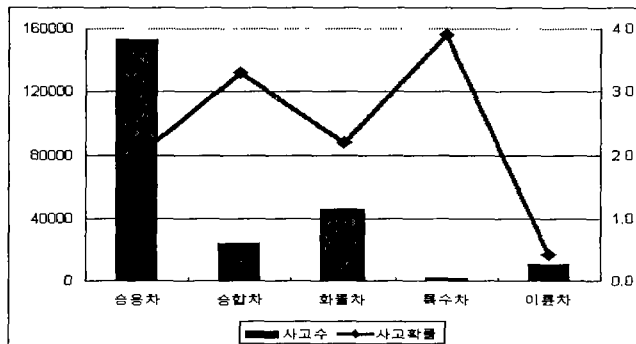


그림 2.2: 차종별 교통사고 건수와 백인율

표 2.3의 제일 상단부에는 1997년도에 발생한 교통사고 건수가 도로종류별로 나타나 있는데 이러한 교통사고의 공식통계자료는 우리들에게 친숙한 자료이다. 그러나 이런 자료에서 도로별 총길이가 서로 큰 차이를 나타낸다. 예를 들어 일반국도에서 발생한 사고수는 고속도로에서 발생한 사고수의 약 9.6배 정도이나 고속도로의 총 길이는 일반국도의 15%정도이다. 그러므로 도로종류별 사고 건수는 도로 길이를 기준으로 비교 분석하여야 한다. 즉, 도로의 총길이에 따른 도로 1km당 발생하는 교통사고 건수를 조사하여 표 2.3의 중단부에 나타나 있다. 여기에서 알 수 있듯이 지방도로에서의 사고발생 건수가 1.4로 제일 적으며, 다음으로는 시군도로에서 2.2를 나타낸다. 역시 일반국도에서의 발생 건수가 약 5.5건으로 제일 많으며, 특별시도와 고속도로에서의 발생건수가 3.2와 3.8로 비슷하게 나타났다.

또한 자동차가 각 도로종류에서 일년동안에 교통사고를 당할 확률을 구하여 보자. 우리나라 자동차가 매년 15,000km를 평균적으로 주행한다고 가정할 때, 발생할 평균 교통사고 건수를 계산하여 보자. 표 2.3에 나타나 있는 사고 건수에 일년동안에 운행하는 평균거리인 15,000km를 곱하여 주고 해당연도의 차량 등록대수를 나누어주면 매년 15,000km 운행할 때

표 2.3: 도로 종류별 교통사고와 확률(백인율)

	일반국도	지방도	특별시도	시군도	고속도로
사고수	68898	24766	55452	78341	7169
총 길이	12459	17089	17243	36288	1889
1Km당 사고수	5.5	1.4	3.2	2.2	3.8
15,000Km 운행시 사고확률	0.792	0.202	0.460	0.317	0.547

자료 : 1) 「교통사고통계」, 경찰청, 1998., 2) 「건설교통통계연보」, 건설교통부, 1998.

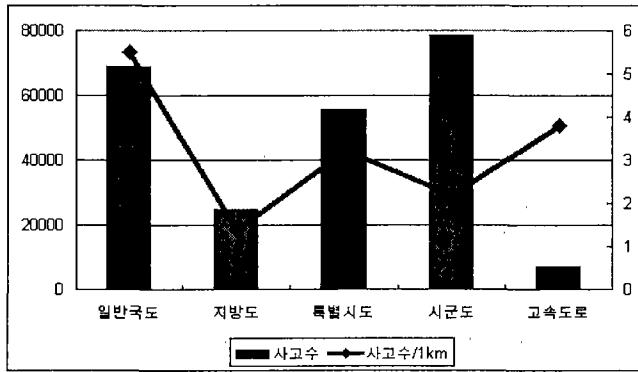


그림 2.3: 도로 종류별 교통사고 건수와 1Km당 사고수

발생할 사고확률이 되는데, 이 결과는 표 2.3의 하단부에 나타나 있다. 대체적으로 자동차가 15,000km 운행시 일반국도와 고속도로, 특별시도에서의 위험율은 다른 도로보다 높게 나타남을 알 수 있다. 일반국도를 15,000km 운행한다고 가정할 때 126대당 한대 꼴로 사고가 발생하며, 고속도로에서는 182대당 한대이고, 특별시도로에서는 217대당 한대 꼴로 교통사고가 발생한다.

지역별 교통사고에 대한 1997년도 자료를 살펴보면, 6대 도시 중에서는 서울이 그리고 6대 도시를 제외한 지역에서는 경기도가 교통사고가 가장 많이 발생하는 것을 알 수 있으며 또한 일반적으로 이 두 지역이 교통 문화가 제일 심각하다고 생각하기가 쉬울 것이다. 그러나 이 두 지역이 시도별 인구나 등록된 자동차수가 제일 많다는 사실을 감안하여 인구수와 자동차 대수를 기준으로 비교하여 살펴보면 표 2.4와 같이 또 다른 사실을 발견할 수 있다. 즉, 6대 도시 중에서는 부산, 서울과 광주가 사고율이 낮게 나타났으며 인천과 대구의 사고율이 가장 높게 나타났다. 또한 도별로는 경기도와 경상남도의 사고율이 낮게 나타났으며, 강원도와 충청북도의 사고율이 높게 나타났음을 알 수 있다. 만약 이 자료에서 모든 사고가 해당되는 지역에 거주하는 사람과 등록된 자동차가 발생한 사고라고 가정한다면 우리는 다음과 같은 재미있는 결론을 내릴 수 있다: "1997년도에는 부산 시민 284명중 한명 꼴로 교통사고가 발생하고 강원도민은 110명중 한명 꼴로 교통사고가 난다. 또한 광주

시에 등록된 자동차 56대중에 한대 꼴로 사고가 발생하고 충청남도에 등록된 자동차 19대 중 한대 꼴로 사고가 발생한다”고 유추할 수 있다.

표 2.4: 지역별 교통사고와 확률(백인율, *, 1인당, **, 1대당)

	6대 도시						기타 지역								
	서울	부산	대구	인천	광주	대전	경기	강원	충북	충남	전북	전남	경북	경남	제주
사고수	40863	13904	14195	14468	5081	5823	42425	13547	11579	21649	11712	17313	21432	20219	3146
사고확률*	0.379	0.352	0.628	0.592	0.393	0.47	0.499	0.907	0.81	0.803	0.629	0.555	0.757	0.505	0.6
사고확률**	1.817	1.931	2.184	2.184	1.782	1.795	2.098	3.764	3.474	5.37	2.869	4.432	3.272	2.938	2.229

자료: 1) 「교통사고통계」, 경찰청, 1998., 2) 「건설교통통계연보」, 건설교통부, 1998.

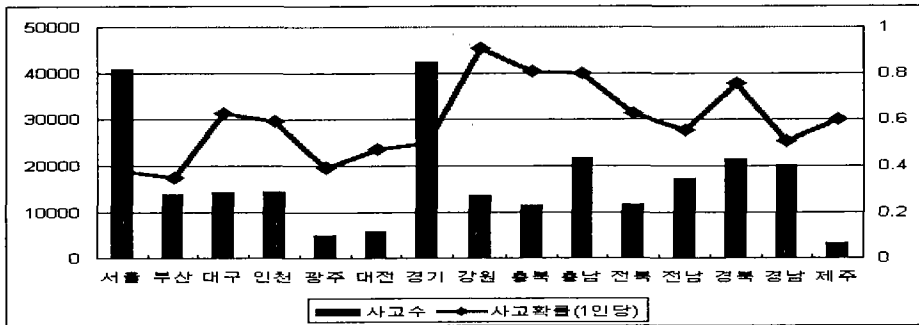


그림 2.4: 지역별 교통사고 건수와 사고확률(백인율)

마지막으로 지역별 범죄에 대한 자료인 표 2.5의 왼쪽표를 분석해 보자. 도시별 인구수를 기준으로 확률을 계산하여 표 2.5의 오른쪽에 나타내었다. 범죄의 발생확률은 값이 너무 작기 때문에 계산된 발생확률에 일만을 곱하여 구한 만인율(萬人率)로 표현하였다. 표 2.5의 사건수를 살펴보면, 우선 서울이 제일 포악한 도시라는 느낌을 받을 수 있다. 그러나 인구수를 고려한 표 2.5의 확률값을 살펴보면 오히려 서울이 제일 치안이 보장된 안전한 도시라고 판단할 수 있다. 그리고 표 2.5에서 죄목별로 자세히 살펴보면, 절도, 강도, 강간, 폭행, 절도, 공갈이 발생할 가능성이 많은 지역은 광주인데 특히 절도, 강도, 폭행은 다른 도시보다 월등히 광주가 높은 발생확률을 갖고 있다는 것이 눈에 띈다. 또한 상해와 간통이 발생할 가능성이 많은 지역은 대구이며 살인이 발생할 가능성이 많은 지역은 부산이라는 사실을 쉽게 파악할 수 있다.

이와 같이 전체적으로 발생 빈도수를 제공하는 공식통계자료들에 확률의 개념을 도입하여 살펴보면, 자료의 의미가 완전히 다른 결과가 유추되기도 하는데 이렇게 가공된 자료는 일반인들이 피부로 느낄 수 있는 확률로 활용될 수 있는 정보로 제공된다.

표 2.5: 도시별 범죄 사건과 도시별 범죄 발생확률(만인율)

사건수	서울	부산	대구	인천	광주	대전
절도	12511	8283	4790	6633	5717	4462
살인	139	78	43	38	22	20
강도	937	327	224	384	283	119
강간	1115	385	258	353	207	156
폭행	1805	172	202	186	483	51
상해	5642	1868	1315	1284	658	497
공갈	245	154	102	70	84	62
간통	968	435	330	318	171	163

사건수	서울	부산	대구	인천	광주	대전
절도	12.04	21.43	19.15	26.95	43.10	33.73
살인	0.13	0.20	0.17	0.15	0.17	0.15
강도	0.90	0.85	0.90	1.56	2.13	0.90
강간	1.07	1.00	1.03	1.43	1.56	1.18
폭행	1.74	0.45	0.81	0.76	3.64	0.39
상해	5.43	4.83	5.26	5.22	4.96	3.76
공갈	0.24	0.40	0.41	0.28	0.63	0.47
간통	0.93	1.13	1.32	1.29	1.29	1.23

자료 : 1) 「범죄분석」, 대검찰청, 1998., 2) 「장래인구추계」, 통계청, 1998.

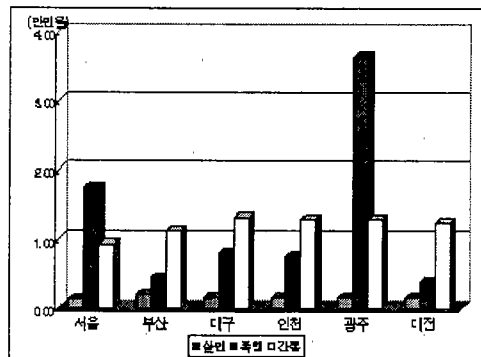
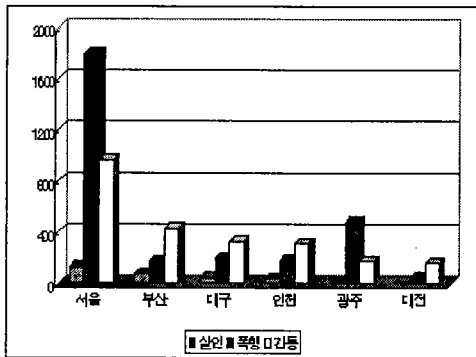


그림 2.5: 도시별 범죄 사건과 도시별 범죄 발생확률(만인율)

3. 조건부 확률로 표현된 자료

2절 마지막에서 도시별 범죄 발생확률은 각 도시별의 인구를 기준으로 하였다. 그러므로 모든 사람은 자기가 살고있는 도시에서 발생하는 범죄에만 연루된다고 가정해야한다. 따라서 이런 문제점을 개선하기 위해서는 전체인구를 고려하여야 한다. 그러나 이 값은 기준이 되는 분모의 수가 커짐으로 인하여 표 2.5(B)에서 나타난 값(만인율) 보다도 더욱 작기 때문에 이에 대한 보완책으로 주변확률(marginal probability)과 조건부 확률(conditional probability)을 이용하여 표현하고자 한다(1절에서도 언급하였듯이 일반적인 범죄분석에서는 인구 10만명당 범죄발생 건수를 계산하는 범죄율(crime rate)이 있는데 본 논문은 범죄 분석만을 다루는 것이 아니고 일반적인 공식통계자료를 대상으로 하기 때문에 적절한 만인율을 사용하였다). 예를 들어 서울에서 절도가 발생할 확률은 다음과 같이 간단한 원리를 응용하여 구할 수 있다.

$$P(\text{서울} \cap \text{절도}) = P(\text{서울}|\text{절도})P(\text{절도}). \tag{3.1}$$

표 3.1: 도시별 조건부 범죄 발생확률(백인율)

조건부 확률	서울	부산	대구	인천	광주	대전	기타지역	주변확률(만인율)
절도	15.00	9.93	5.74	7.95	6.86	5.35	49.16	18.13
살인	17.62	9.89	5.45	4.82	2.79	2.53	56.91	0.17
강도	21.04	7.34	5.03	8.62	6.35	2.67	48.94	0.94
강간	19.66	6.79	4.55	6.23	3.65	2.75	56.37	1.25
폭행	33.69	3.21	3.77	3.47	9.02	0.95	45.88	1.22
상해	24.39	8.08	5.69	5.55	2.84	2.15	51.30	5.21
공갈	17.86	11.22	7.43	5.10	6.12	4.52	47.74	0.3
간통	16.67	7.49	5.68	5.48	2.94	2.81	58.94	1.28

자료: 1) 「범죄분석」, 대검찰청, 1998., 2) 「장래인구추계」, 1998, 통계청.

즉, (3-1)식의 좌변의 확률은 주변확률과 조건부 확률의 곱이므로 각각 구분하여 표 3.1에 나타내었다. 표 3.1에서는 주요도시를 중심으로 조건부 확률의 백분율을 나타내고 있으며, 죄목별 발생확률을 나타내는 주변확률은 값이 작기 때문에 만인율로 나타내고 있으며 표의 제일 오른쪽에 위치하고 있다. 그리고 나머지 칸에서는 어떤 범죄발생을 조건으로 어떤 도시에서 발생할 확률인 조건부 확률을 백분율로 나타내고 있다. 예를 들어 우선 죄목별 발생확률로부터 절도가 제일 많이 발생하고 있는데 만명당 약 18명 정도로 피해를 당한다고 말할 수 있다. 그리고 절도로 인한 피해 중 15%가 서울에서 발생한다고 할 수 있다. 그러므로 서울에서 절도를 당할 확률은

$$\begin{aligned}
 P(\text{서울} \cap \text{절도}) &= P(\text{절도})P(\text{서울}|\text{절도}) \\
 &= 18.13/10000 \times 15.00/100 \\
 &= 2.72 \times 10^{-4},
 \end{aligned}$$

으로 일만명 중 약 3명이 이런 경우에 속한다. 이 절에서는 조건부 확률에 대한 저자의 의견을 집중적으로 피력하기 위하여 범죄사건 자료에 대하여만 계속 예를 들어 설명하고자 한다. 다음으로는 피해자의 성별에 따라 연령별의 범죄 발생확률을 계산하였다. 표 3.2와 표 3.3에서 죄목별 조건부확률을 살펴보면 대부분 30대가 범죄로부터 피해자가 될 확률이 크지만 공갈로 인한 피해자 중에는 13세 15세까지의 남자중학생인 경우가 월등히 높으며 강도, 강간의 피해자 중에는 20대 여자인 경우가 가장 많음을 알 수 있다.

여기에서 표 3.2와 표 3.3를 표 3.1과 연결하여 도시 변수(A)와 성별로 구분되어 있는 연령변수(B)를 동시에 고려하여 보자. 연령 변수와 도시 변수는 통계적 독립(stochastic independence)이라고 가정할 수 있으며, 나아가 각 죄목이 조건일 때도 독립이라고 가정하자. 이런 조건부 독립(conditional independence)의 가정 하에서 우리는 다음이 만족됨을 알 수 있다.

$$P(A \cap B | \text{강도}) = P(A | \text{강도})P(B | \text{강도}).$$

예를 들어 부산(A)에서 30대의 남자(B)가 강도를 당할 확률을 구하여 보자.

$$\begin{aligned}
 P(A \cap B \cap \text{강도}) &= P(\text{강도})P(A|\text{강도})P(B|\text{강도}) \\
 &= 0.94/10000 \times 7.34/100 \times 30.11/100 \\
 &= 2.08 \times 10^{-6}.
 \end{aligned}$$

그러므로 30대의 남자 일백만명 중에 2명 꼴로 부산에서 강도를 당한다고 판단된다.

표 3.2: 피해자 남자 연령별 조건부 범죄 발생확률(백인율)

조건부확률	-6세	7-15세	16-20세	21-30세	31-40세	41-50세	51-60세	60세-
절도	0.07	0.73	4.75	23.36	39.00	21.47	7.81	2.81
살인	3.17	1.73	4.90	21.61	33.72	20.46	6.92	7.49
강도	0.00	3.62	11.28	23.99	30.11	19.30	7.24	4.45
강간	1.49	14.89	14.64	26.55	29.53	9.68	2.23	0.99
폭행	0.19	3.36	6.12	15.50	32.26	22.55	11.95	8.07
상해	0.20	2.23	4.77	16.35	34.15	24.12	11.75	6.43
공갈	0.38	37.67	18.02	5.76	14.89	13.02	6.01	4.26
간통	0.09	0.00	0.04	7.36	48.17	35.96	7.49	0.89

표 3.3: 피해자 여자 연령별 조건부 범죄 발생확률(백인율)

조건부확률	-6세	7-15세	16-20세	21-30세	31-40세	41-50세	51-60세	60세-
절도	0.04	0.39	4.01	26.90	35.70	22.31	7.58	3.07
살인	6.02	3.39	5.26	19.55	34.96	15.79	7.14	7.89
강도	0.13	2.75	9.25	31.06	27.88	17.94	6.00	5.00
강간	3.04	16.92	23.45	25.15	18.29	9.39	2.40	1.37
폭행	0.43	1.83	4.43	15.65	30.52	23.83	12.96	10.35
상해	0.18	0.93	3.08	14.00	30.92	26.42	14.99	9.48
공갈	0.00	12.69	8.63	19.80	25.89	21.83	8.63	2.54
간통	0.00	0.00	0.25	16.27	52.52	25.54	4.92	0.50

자료: 1) 「범죄분석」, 대검찰청, 1998., 2) 「장래인구추계」, 1998, 통계청.

다음으로는 3절에서 언급한 변수들과 조건부 독립을 가정하는데 무리가 없는 시간을 나타내는 변수를 추가하여 시간대별로 발생하는 조건부 확률을 구하여 표 3.4에 정리하였다.

그리고 조건부 독립의 가정 하에 범죄자의 연령을 고려한 조건부 확률을 표 3.5에 나타내었다. 표 3.4으로부터 주로 밤에 발생하는 범죄들은 절도, 살인, 강도, 강간이며, 폭행, 상해, 공갈은 주로 오후(넓게 표현하면 낮)에 많이 발생함을 알 수 있다. 그리고 표 3.5에서는 절도, 강도, 공갈의 범죄자는 주로 20세 미만의 청소년이라는 사실과 살인, 강간은 주로 20대와 30대가 범행을 저지르고, 상해와 간통은 30대와 40대의 범죄자가 많다는 사실을 얻을 수 있다. 특히 폭행죄로 검거된 자들은 주로 30대 초반이다. 여기에서도 표 3.1부터 표 3.5를 이용하여 다음과 같은 가상의 경우가 발생할 확률을 구하여 보자. "30대의 여자가 오후시간에 광주에서 40대의 범인으로부터 폭행을 당할 확률"은 다음과 같다.

$$\begin{aligned}
 & P(\text{폭행})P(\text{광주|폭행})P(30\text{대여자|폭행}) \times P(\text{오후|폭행})P(40\text{대범인|폭행}) \\
 & = 1.22/10000 \times 9.02/100 \times 30.52/100 \times 39.52/100 \times 10.06/100 \\
 & = 1.34 \times 10^{-7}.
 \end{aligned}$$

이 값은 30대의 여자 일천만명 중 1.34명 비율로 오후 시간에 광주에서 40대의 범인으로부터 폭행을 당한다는 것으로 파악된다. 여기까지 우리가 고려한 변수는 범죄 발생 지역, 피

표 3.4: 시간별*조건부 범죄 발생확률(백인율)

조건부확률	새벽	아침	낮	오전	오후	저녁	밤
절도	5.50	3.83	31.28	10.08	21.21	7.90	34.48
살인	9.76	3.80	24.46	7.60	16.86	6.46	39.80
강도	10.04	1.66	19.74	6.22	13.52	3.97	45.55
강간	8.92	2.45	18.84	5.29	13.54	4.90	39.93
폭행	2.86	6.12	55.18	15.66	39.52	9.11	6.72
상해	2.22	6.93	60.11	17.96	42.15	8.73	3.19
공갈	1.09	3.06	50.22	14.07	36.15	5.47	5.54

*: 새벽: 04:00-06:59 아침: 07:00-08:59 오전: 09:00-11:59

오후: 12:00-17:59 밤: 18:00-19:59 저녁: 20:00-03:59 낮: 09:00-17:59

자료: 1) 「범죄분석」, 대검찰청, 1998., 2) 「장래인구추계」, 1998, 통계청.

해자의 성별과 연령, 시간, 그리고 범죄자의 연령이었다. 이 변수들은 상호 독립적이라고 할 수 있으며, 또한 죄목별로도 조건부 독립이라고 가정해도 일반적으로 무리가 없었다. 나아가 이런 조건부 독립성을 만족하는 다음과 같은 여러 변수들을 고려한 범죄 자료에 대하여 분석이 가능하다.

- 누 가 : 피해자와 범죄자의 성별과 연령, 범죄자의 학력, 직업 또는 전과 여부, 재판자의 재판기간 및 종류, 공범여부나 공범관계 등의 변수
 언 제 : 범죄 시간(월별, 요일별, 공휴일 등의 특수사정, 눈 비 등의 일기상황, 시간대별, 재판 기간 등)의 변수

표 3.5: 범죄자 연령별 조건부 범죄 발생확률(백인율)

	-15세	16-20세	21-25세	26-30세	31-35세	36-40세	41-50세	51-60세	61-70세	70세-
절도	29.41	34.23	8.59	7.56	5.97	5.52	5.68	2.29	0.64	0.11
살인	0.74	9.71	11.79	15.72	17.81	16.34	18.92	6.63	2.21	0.12
강도	9.55	49.56	18.72	10.19	5.74	3.26	2.21	0.64	0.11	0.02
강간	2.16	9.74	15.94	20.54	16.06	13.51	14.12	5.65	1.91	0.36
폭행	0.96	1.93	2.57	4.76	63.91	8.22	10.60	5.03	1.70	0.33
상해	1.06	4.09	6.42	11.51	15.59	19.85	25.42	11.88	3.39	0.80
공갈	28.60	24.78	2.50	4.41	6.18	10.74	13.90	6.91	1.84	0.15
간통	0.10	0.67	4.75	12.97	23.29	28.21	24.68	4.82	0.43	0.08

자료: 1) 「범죄분석」, 대검찰청, 1998., 2) 「장래인구추계」, 1998, 통계청.

어디서 : 변행 장소(시도별, 장소별) 등의 변수

무엇을 : 각종 범죄의 죄명별, 동종 및 이종 범죄별 등의 변수

어떻게 : 재산과 신체의 피해정도, 신체 상해정도, 피해자와 범죄자의 관계 및 공범들과의 관계, 공범수단 변행 도구, 변행후 은신처, 그리고 침입적도의 침입구와 방법 및 도구, 변행도구와 입수 방법, 범죄와 판결의 관계 등의 변수

왜 : 변행 동기, 건거 장소 등의 변수

이보다 더욱 다양한 변수들을 고려한 원자료들은 대검찰청에서 발간하는 범죄분석 등의 자료집에 수록되어 있다. 이런 공개된 자료를 앞에서 언급한 방법을 이용하여 조건부 확률로 변환하여 분석한다면, 육하(六何)원칙보다도 더욱 복잡한 우리들의 현실 세계의 현상에서 어떤 범죄가 발생할 확률을 구할 수 있겠다.

그리고 2절에서 주로 다룬 교통사고에 대한 분석에서도 조건부 독립성을 만족하는 다음과 같은 여러 변수들을 고려하여 복잡하고 다양한 상황에서 교통사고 당할 확률, 부상 당할 확률, 또는 사망할 확률을 구한 후 분석할 수 있다.

누 가 : 연령별, 성별 등의 변수

언 제 : 시간별(시간대, 주야, 윤일, 그리고 월별), 일기상태 등의 변수

어디서 : 지역별, 도로종류별 등의 변수

무엇을 : 교통사고, 부상, 또는 사망 그리고 음주운전 및 어린이 교통사고

어떻게 : 차종별, 사고 유형별, 법규위반별 등의 변수

왜 : 사고시의 심리상태 등의 변수

이러한 확률을 일상생활에 적용한다면, 범죄 피해와 교통사고 피해뿐만 아니라 모든 위험으로부터 피해를 최소한으로 줄일 수 있는 과학적 사고를 할 수 있다.

4. 결론

본 논문에서는 공공기관에서 발표하는 많은 공식통계자료(예를 들어 1998년에 발표한 교통사고와 범죄에 관한 통계자료)가 주로 행정용으로만 사용되는 것에서 탈피하여 일반 국민이 쉽게 인지하고 피부로 느낄 수 있는 정보를 얻기 위한 두 가지 방법을 제안하였다.

첫번째 방법은 기본 단위(개인 또는 자동차)가 발생할 확률로 계산되어 실생활에서 느낄 수 있는 값으로 변환시키는 방법이고, 두번째는 여러 상황에서 발생할 확률을 구할 수 있는 조건부 확률로 나타내는 방법이다. 특히 첫번째 방법으로 구한 확률값이 매우 작은 경우에는 그 의미를 파악하기 어렵기 때문에 어떤 사상(사고 또는 사건)을 조건으로 하는 조건부 확률이 유용하다.

확률의 개념은 통계학개론 수준의 강좌 처음부분에 배우는 기본적인 내용이다. 그리고 확률은 학문에서만뿐만 아니라 일상생활에서도 자주 애용되는 단어이다. 그럼에도 불구하고 일상생활에서 확률값을 갖고 의사 결정을 하는 경우는 거의 없다. 그 이유로는 확률의 개념을 도입하기위해 자료를 가공하고자할 때 계산할 수 있는 정보를 얻기가 힘들다는 것이다. 본 논문에서는 경찰청에서 발표하는 교통사고자료와 대검찰청에서 발표하는 범죄자료에 통계청에서 제공하는 인구에 관한 자료와 건설교통부에서 제공하는 자동차등록대수 자료 등과 병행하여 논문에서 제안한 두가지 방법을 사용하여 분석하였다. 그리고 그 결과들이 현실세계에서 의사결정 하는데 중요한 역할을 한다는 것을 설명하였다.

결론적으로 말하여 공공기관에서 공식통계자료를 발표할 때, 원자료 뿐만 아니라 확률로 나타나는 정보 그리고 현실 상황에 활용할 수 있는 조건부 확률도 함께 제공해주길 바란다.

참고문헌

- [1] 건설교통부 (1996, 1997, 1998). <건설교통통계연보>.
- [2] 건설교통부 (1996, 1997, 1998). <교통안전연차보고서>.
- [3] 경찰청 (1996, 1997, 1998). <교통사고통계>.
- [4] 경찰청 (1996, 1997, 1998). <범죄분석>.
- [5] 교통안전공단 (1997). <교통판례집>.
- [6] 도로교통안전협회 (1997, 1998). <교통사고통계 분석>.
- [7] 대검찰청 (1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998). <범죄분석>.
- [8] 문화체육부 (1996, 1997, 1998). <청소년 백서>.
- [9] 박상기, 손동권, 이순래 (1998). <형사정책>, 한국형사정책연구원.

[10] 법무연수원 (1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998). <범죄백서>.

[11] 통계청 (1996, 1997, 1998). <장래인구추계>.

[12] Sellin, T. and Wolfgang, M. (1964). *The Measurement of Delinquency*, New York.

[1999년 2월 접수, 1999년 9월 최종수정]

A Representation Method for Official Statistical Data

Chong Sun, Hong¹⁾ Han Seung, Lim²⁾

ABSTRACT

Most of official statistical data issued by our government offices are consisted with frequencies, percentages or some indices as observed values. These data are not used by public people but by government officers for the administration only. In order for these data to be useful to the public, the probabilities which are occurring to per person (or the basic unit) had better to be transformed from the data. And by obtaining the conditional probabilities as well as the probabilities, we can calculate and obtain the probability which a complex situation combined many events will happen in the real world. The values of these probabilities might play very important role in decision making process in our everyday life. By using two concepts of the probability and the conditional probability in this paper, we demonstrate and explore two official data of 1997' car-accident and crime to be minimize our losses from recent panic.

1) Professor, Department of Statistics, SungKyunKwan University, Seoul, 110-745, Korea, cshong@stat.skku.ac.kr.

2) Lecturer, Department of Statistics, SungKyunKwan University, Seoul, 110-745, Korea, mylims@dragon.skku.ac.kr.