

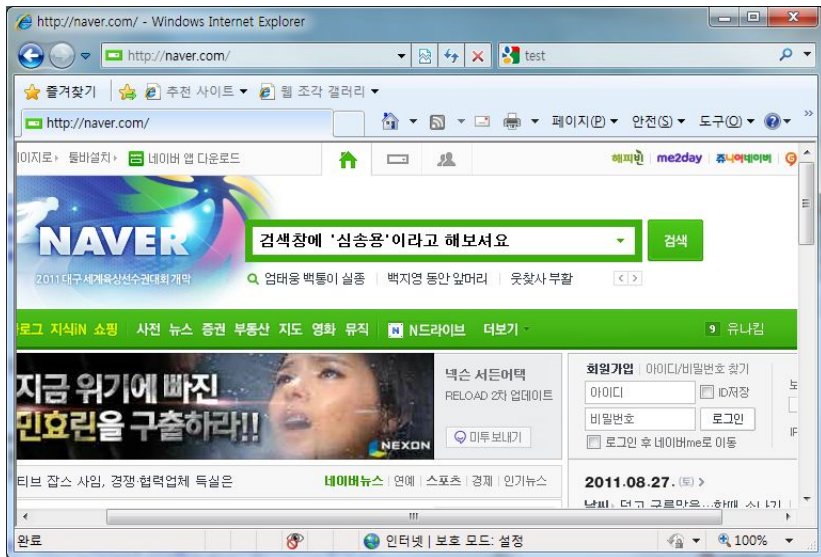
빅데이터와 통계해석

MS by L^AT_EX

심송용(<http://jupiter.hallym.ac.kr>)

한림대학교 데이터사이언스학부

강사는 누구인가?



거짓말

이 세상의 세 가지 종류의 거짓말
(three kinds of lies)

거짓말

이 세상의 세 가지 종류의 거짓말
(three kinds of lies)

- 1 거짓말 (lies)

거짓말

이 세상의 세 가지 종류의 거짓말
(three kinds of lies)

- ① 거짓말 (lies)
- ② 새빨간 거짓말(damned lies), 그리고

거짓말

이 세상의 세 가지 종류의 거짓말
(three kinds of lies)

- ① 거짓말 (lies)
- ② 새빨간 거짓말(damned lies), 그리고
- ③ 통계 (statistics)

거짓말

이 세상의 세 가지 종류의 거짓말
(three kinds of lies)

- ① 거짓말 (lies)
- ② 새빨간 거짓말(damned lies), 그리고
- ③ 통계 (statistics)

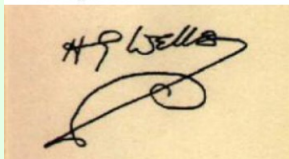
Prime Minister UK

20 Feb 1874 – 21 Apr 1880



Benjamin Disraeli
(1804-1881)

문화인의 기본 소양 – H. G. Wells



“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”

언젠가는 통계적 사고가 읽고 쓰기만큼 중요한 문화인의 능력이 될 것이다.

Herbert G. Wells (1866 - 1946)

세종 때의 여론조사

- 1430년 3월 새로운 세금제도인 공법제에 대해 전국의 관리와 백성을 대상으로 찬반을 물음.

세종 때의 여론조사

- 1430년 3월 새로운 세금제도인 공법제에 대해 전국의 관리와 백성을 대상으로 찬반을 물음.
- 경기, 경상, 전라 등 토지가 좋은 곳은 찬성이 많고 평안, 함길, 강원은 반대가 많음. 관리는 찬반이 비슷

세종 때의 여론조사

- 1430년 3월 새로운 세금제도인 공법제에 대해 전국의 관리와 백성을 대상으로 찬반을 물음.
- 경기, 경상, 전라 등 토지가 좋은 곳은 찬성이 많고 평안, 함길, 강원은 반대가 많음. 관리는 찬반이 비슷
- 여론이 갈리자 시행을 미루고 보완을 거쳐 1438년 경상, 전라도만 시범실시, 3년후 충청도 실시

세종 때의 여론조사

- 1430년 3월 새로운 세금제도인 공법제에 대해 전국의 관리와 백성을 대상으로 찬반을 물음.
- 경기, 경상, 전라 등 토지가 좋은 곳은 찬성이 많고 평안, 함길, 강원은 반대가 많음. 관리는 찬반이 비슷
- 여론이 갈리자 시행을 미루고 보완을 거쳐 1438년 경상, 전라도만 시범실시, 3년후 충청도 실시
- 1444년 연분 9등급(풍년의 정도), 전분 6등급(토지 등급)

세종 때의 여론조사

- 1430년 3월 새로운 세금제도인 공법제에 대해 전국의 관리와 백성을 대상으로 찬반을 물음.
- 경기, 경상, 전라 등 토지가 좋은 곳은 찬성이 많고 평안, 함길, 강원은 반대가 많음. 관리는 찬반이 비슷
- 여론이 갈리자 시행을 미루고 보완을 거쳐 1438년 경상, 전라도만 시범실시, 3년후 충청도 실시
- 1444년 연분 9등급(풍년의 정도), 전분 6등급(토지 등급)

기미독립선언문

... 統計數字上(통계숫자상) 虛飾(허식)의 下(하)에서 利害相反(이해상반)한 兩(양) 民族間(민족간)에 永遠(영원)히 和同(화동)할 수 없는 怨溝(원구)를 去益深造(거익심조)하는 ...

사전 등재

Statistics: Encyclopedia Britannica (1797)

Statistics = State + Arithmetic(Number)

사전 등재

Statistics: Encyclopedia Britannica (1797)

Statistics = State + Arithmetic(Number)

통계(Statistics)의 어원

- 라틴어의 statisticum collegium $\Rightarrow$ council of state(국가위원회)

사전 등재

Statistics: Encyclopedia Britannica (1797)

Statistics = State + Arithmetic(Number)

통계(Statistics)의 어원

- 라틴어의 statisticum collegium $\Rightarrow$ council of state(국가위원회)
- 이탈리아어의 statista $\Rightarrow$ statesman 또는 politician

사전 등재

Statistics: Encyclopedia Britannica (1797)

Statistics = State + Arithmetic(Number)

통계(Statistics)의 어원

- 라틴어의 statisticum collegium $\Rightarrow$ council of state(국가위원회)
- 이탈리아어의 statista $\Rightarrow$ statesman 또는 politician
- 독일어인 Statistik $\Rightarrow$ science of state, political arithmetic

통계란

사전 등재

Statistics: Encyclopedia Brittanica (1797)

Statistics = State + Arithmetic(Number)

통계(Statistics)의 어원

- 라틴어의 statisticum collegium $\Rightarrow$ council of state(국가위원회)
- 이탈리아어의 statista $\Rightarrow$ statesman 또는 politician
- 독일어인 Statistik $\Rightarrow$ science of state, political arithmetic

십만양병설

10만의 병력을 유지하기 위해 필요한 인구, 경제력은?

기댓값

조폐공사나 한국은행을 털지 않는 도둑

기댓값

조폐공사나 한국은행을 털지 않는 도둑

기댓값과 분산: 약속시간 맞추기

지방 출장 시간: 지하철 $\Rightarrow$ 고속버스 $\Rightarrow$ 택시

생활 속의 확률

기댓값

조폐공사나 한국은행을 털지 않는 도둑

기댓값과 분산: 약속시간 맞추기

지방 출장 시간: 지하철 $\Rightarrow$ 고속버스 $\Rightarrow$ 택시

확률 1

광화문 네거리를 무단횡단해 볼까?

생활 속의 확률

기댓값

조폐공사나 한국은행을 털지 않는 도둑

기댓값과 분산: 약속시간 맞추기

지방 출장 시간: 지하철 $\Rightarrow$ 고속버스 $\Rightarrow$ 택시

확률 1

광화문 네거리를 무단횡단해 볼까?

확률 2

눈/비올 확률 50%?

확률3: 로또 명당

- 부산동구 '부일카서비스'는 26건의 1등 당첨자를 배출

확률3: 로또 명당

- 부산동구 '부일카서비스'는 26건의 1등 당첨자를 배출
- 서울 노원구 '스파' 판매점 23건의 1등 당첨자를 배출

확률3: 로또 명당

- 부산동구 '부일카서비스'는 26건의 1등 당첨자를 배출
- 서울 노원구 '스파' 판매점 23건의 1등 당첨자를 배출
- 현재까지 최고액 당첨자는 춘천경찰서 직원('03년, 407억2천만원)

확률3: 로또 명당

- 부산동구 '부일카서비스'는 26건의 1등 당첨자를 배출
- 서울 노원구 '스파' 판매점 23건의 1등 당첨자를 배출
- 현재까지 최고액 당첨자는 춘천경찰서 직원('03년, 407억2천만원)



확률3: 로또 명당

- 부산동구 '부일카서비스'는 26건의 1등 당첨자를 배출
- 서울 노원구 '스파' 판매점 23건의 1등 당첨자를 배출
- 현재까지 최고액 당첨자는 춘천경찰서 직원('03년, 407억2천만원)



[관련기사](#)

생활 속의 확률(속)

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{{}_{45}C_6} = \frac{1}{8,145,060}$$

생활 속의 확률(속)

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{{}_{45}C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{45C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.
- 돼지꿈??, 조상꿈??, 땡꿈?? 을 꾀다.

로또 1등 당첨자 가장 많이 꾀 꿈 “조상꿈”

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{45C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.
- 돼지꿈??, 조상꿈??, 땡꿈?? 을 꾀다.
로또 1등 당첨자 가장 많이 꾀 꿈 “조상꿈”
- 자주 나온 번호는 피해야 한다?

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{{}_{45}C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.
- 돼지꿈??, 조상꿈??, 땡꿈?? 을 꾀다.
로또 1등 당첨자 가장 많이 꾀 꿈 “조상꿈”
- 자주 나온 번호는 피해야 한다?
- 자주 나온 번호를 골라야 한다?

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{{}_{45}C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.
- 돼지꿈??, 조상꿈??, 땡꿈?? 을 꾀다.
로또 1등 당첨자 가장 많이 꾀 꿈 “조상꿈”
- 자주 나온 번호는 피해야 한다?
- 자주 나온 번호를 골라야 한다?
- 함께 나온 번호??

생활 속의 확률(속)

당첨확률 높이기 전략

- 로또 1등 확률:

$$\frac{1}{{}_{45}C_6} = \frac{1}{8,145,060}$$

- 로또명당을 찾아간다.
- 돼지꿈??, 조상꿈??, 땡꿈?? 을 끈다.
로또 1등 당첨자 가장 많이 끈 꿈 "조상꿈"
- 자주 나온 번호는 피해야 한다?
- 자주 나온 번호를 골라야 한다?
- 함께 나온 번호??

급기한 : 2019/09/23
970 91491 20894 09584 51672 37548 33585
017MPBDFHBJA33 141100800/0000006784

수	등	08	12	23	26	31	38
수	등	03	07	14	17	37	42
수	등	08	15	21	31	33	38
수	등	02	09	28	33	37	45
수	등	01	04	06	22	26	43

₩5,000
금 액
970 91491 20894 09584 51672 37548 33585

이번주 로또1등 당첨 예상번호 "7, 13, 28, 39.. R

2019년 811회, 841회 1등 당첨... 놀라운 당첨 확률을 보여주고 있는 "월드로또"는 무료로 행운의 당첨번호를 지급하고 있다. 월드로또는 매주 빅데이터 기반 회차별, 당첨 패턴 등으로 번호를 분석하는데, 룰렛을 돌리기만 하면 "1등당첨 5 조합"을 3개월간 무료로 받아볼 수 있다...

광고

- 지난 11년간 등록된 볼보 자동차의 95% 이상이 아직도 사용되고 있습니다. Over 95% of all Volvos registered here in last 11 years are still on the road.

광고

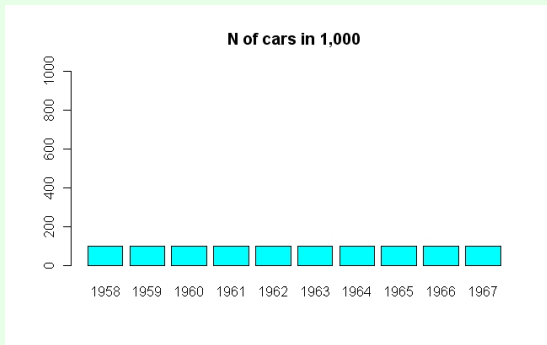
- 지난 11년간 등록된 볼보 자동차의 95% 이상이 아직도 사용되고 있습니다. Over 95% of all Volvos registered here in last 11 years are still on the road. (1968년 볼보 광고)

광고

- 지난 11년간 등록된 볼보 자동차의 95% 이상이 아직도 사용되고 있습니다. Over 95% of all Volvos registered here in last 11 years are still on the road. (1968년 볼보 광고)
- 실제 생산 대수

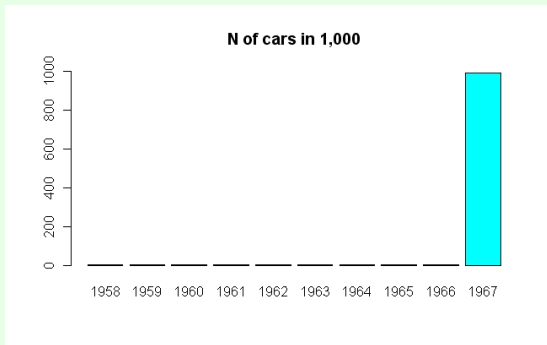
광고

- 지난 11년간 등록된 볼보 자동차의 95% 이상이 아직도 사용되고 있습니다. Over 95% of all Volvos registered here in last 11 years are still on the road. (1968년 볼보 광고)
- 실제 생산 대수



광고

- 지난 11년간 등록된 볼보 자동차의 95% 이상이 아직도 사용되고 있습니다. Over 95% of all Volvos registered here in last 11 years are still on the road. (1968년 볼보 광고)
- 실제 생산 대수



What You Guest

Years ago	Cars Prod.
10	100,000
9	100,000
8	100,000
7	100,000
6	100,000
5	100,000
4	100,000
3	100,000
2	100,000
1	100,000
합	1,000,000
In use	950,000

What You Guest

Years ago	Cars Prod.
10	100,000
9	100,000
8	100,000
7	100,000
6	100,000
5	100,000
4	100,000
3	100,000
2	100,000
1	100,000
합	1,000,000
In use	950,000

Reality Might Be

Years ago	Cars Prod.
10	1,000
9	1,000
8	1,000
7	1,000
6	1,000
5	1,000
4	1,000
3	1,000
2	1,000
1	991,000
합	1,000,000
In use	950,000

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)
 - 신인 선수 연봉은 2400만원으로 고정(관련 기사)

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)
 - 신인 선수 연봉은 2400만원으로 고정(관련 기사)
 - 전체 선수는 기존선수 471명과 신인선수 68명, 외국인선수 14명의 자료.

평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)
 - 신인 선수 연봉은 2400만원으로 고정(관련 기사)
 - 전체 선수는 기존선수 471명과 신인선수 68명, 외국인선수 14명의 자료.
 - 2014년 처음으로 1억을 넘음(관련 기사)

평균?, 중앙값?

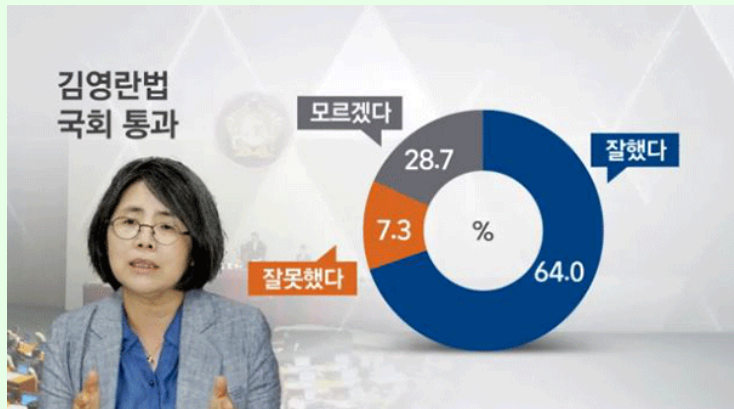
- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)
 - 신인 선수 연봉은 2400만원으로 고정(관련 기사)
 - 전체 선수는 기존선수 471명과 신인선수 68명, 외국인선수 14명의 자료.
 - 2014년 처음으로 1억을 넘음(관련 기사)
- 평균 수심과 당나라 군사

평균?, 중앙값?

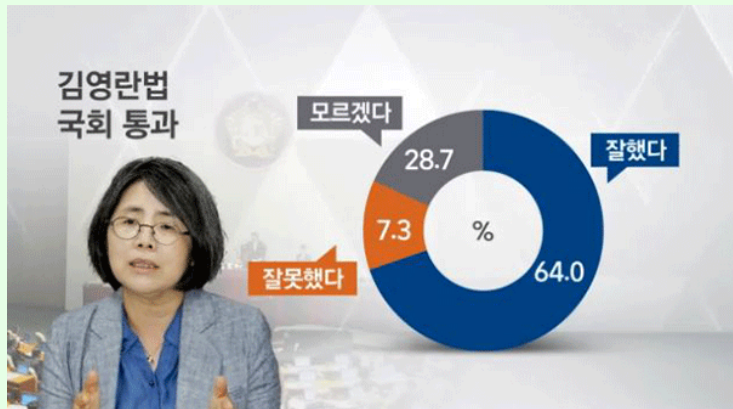
- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 2013년 외국인/신인 선수 제외한 프로야구 선수의 평균 연봉이 9496만 원 (관련 기사)
 - 신인 선수 연봉은 2400만원으로 고정(관련 기사)
 - 전체 선수는 기존선수 471명과 신인선수 68명, 외국인선수 14명의 자료.
 - 2014년 처음으로 1억을 넘음(관련 기사)
- 평균 수심과 당나라 군사



김영란법에 대한 찬반 비율

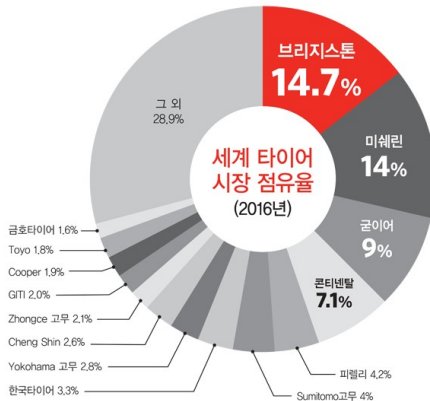


김영란법에 대한 찬반 비율



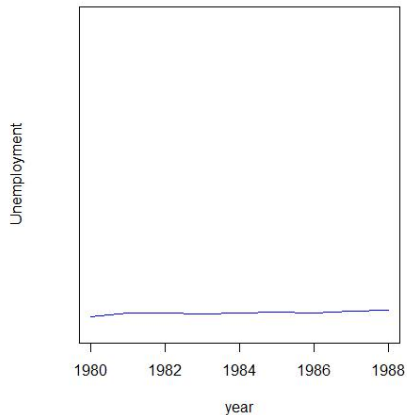
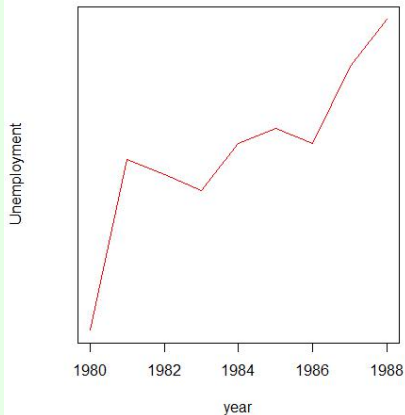
[관련기사](#)

타이어 시장 점유율(2016년)



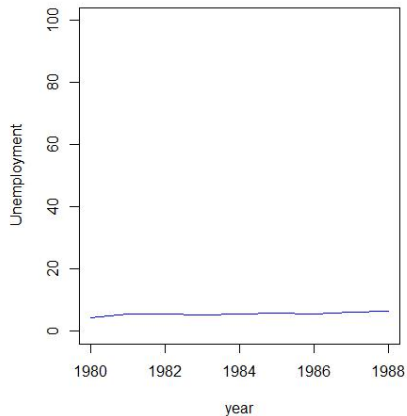
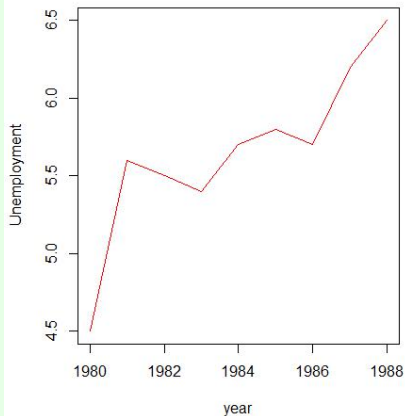
그림에 대한 검토

실업에 대한 두 개의 그래프



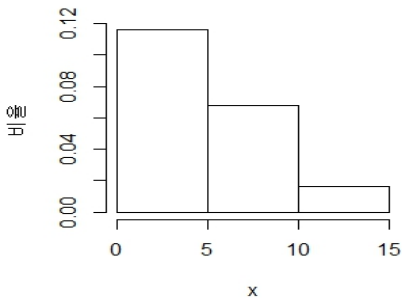
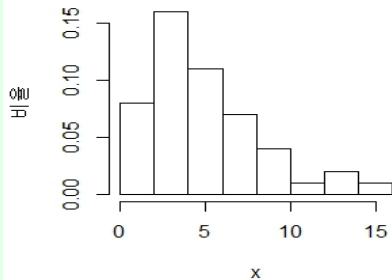
그림에 대한 검토

실업률에 대한 두 개의 그래프



그림에 대한 검토

두 개의 히스토그램



그림에 대한 검토

원래자료

8	2	10	1	1	6	6	3	4	4	3	4	5	3	14	5	3
4	7	9	5	1	10	6	2	4	6	8	3	4	13	3	5	7
7	6	6	7	3	8	4	4	5	1	1	12	2	9	3	15	

그림에 대한 검토

원래자료

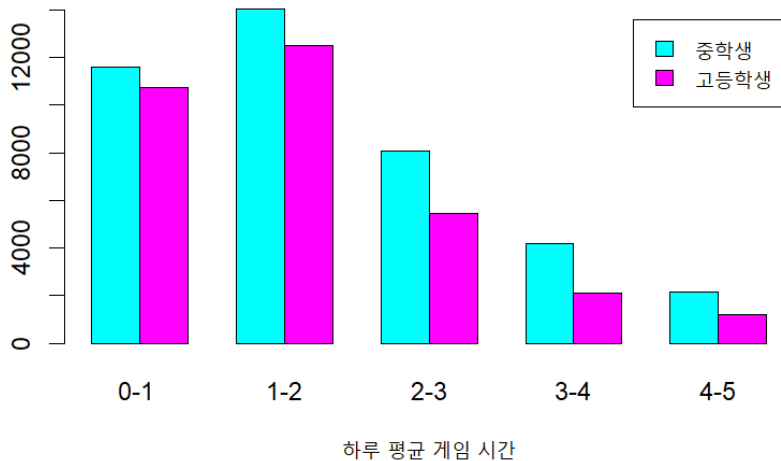
8	2	10	1	1	6	6	3	4	4	3	4	5	3	14	5	3
4	7	9	5	1	10	6	2	4	6	8	3	4	13	3	5	7
7	6	6	7	3	8	4	4	5	1	1	12	2	9	3	15	

기동 수

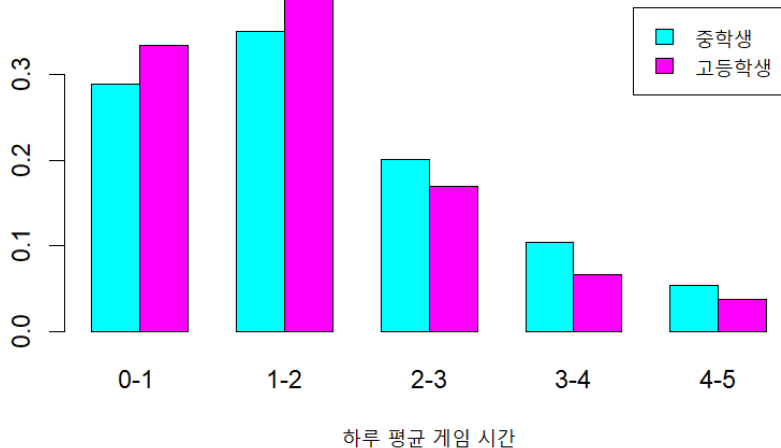
$$1 + 3.3 \log(n), \quad n > 15$$

등

중고등학생의 게임시간 비교



중고등학생의 게임시간 비교



비교그래프와 빈도의 종류

실제 조사결과 빈도(단위: 명)

	0 ^{이상} -1 ^{미만}	1-2	2-3	3-4	4-5	합계
중학생	11,580	14,040	8,060	4,170	2,150	40,000
고등학생	10,720	12,520	5,440	2,120	1,200	32,000

위 자료로 추정한 평균 게임시간

중학생 : 1.78시간, 고등학생 1.58시간

비교그래프와 빈도의 종류

실제 조사결과 빈도(단위: 명)

	0 ^{이상} -1 ^{미만}	1-2	2-3	3-4	4-5	합계
중학생	11,580	14,040	8,060	4,170	2,150	40,000
고등학생	10,720	12,520	5,440	2,120	1,200	32,000

위 자료로 추정한 평균 게임시간

중학생 : 1.78시간, 고등학생 1.58시간

상대빈도로 변환(단위: %)

	0 ^{이상} -1 ^{미만}	1-2	2-3	3-4	4-5	합계
중학생	28.95	35.10	20.15	10.42	5.38	100.00
고등학생	33.50	39.12	17.00	6.62	3.75	100.00

실업

조사기관	실업자 수
The National Industrial Conference Board	9,177,000
Government Committee on Economic Security	10,913,000
The American Federation of Labor	10,077,000
National Research League	14,173,000
Labor Research Association	17,029,000

Cohen (1938) The Misuse of Statistics, *Journ. Amer. Statist. Assoc.* **38**

용어정의

실업

조사기관	실업자 수
The National Industrial Conference Board	9,177,000
Government Committee on Economic Security	10,913,000
The American Federation of Labor	10,077,000
National Research League	14,173,000
Labor Research Association	17,029,000

Cohen (1938) The Misuse of Statistics, *Journ. Amer. Statist. Assoc.* **38**

용어정의

독서율

무엇이 독서인가?

정의? 잘못된 연관?

한국 1인당 석유소비량

지난 2006년 한국의 1인당 석유소비량은 16.18배럴로 세계 5위 수준이었으며 비산유국 가운데에서는 2위, 아시아권에서는 가장 많았다.(파이낸셜 뉴스, 연합뉴스, YTN 등)

정의? 잘못된 연관?

한국 1인당 석유소비량

지난 2006년 한국의 1인당 석유소비량은 16.18배럴로 세계 5위 수준이었으며 비산유국 가운데에서는 2위, 아시아권에서는 가장 많았다.(파이낸셜 뉴스, 연합뉴스, YTN 등)

2018년 13대 수출품목(산자부) 2005년 2006년 2011년 2014년

자동차	40,899	자동차부품	23,122	반도체	126,713
무선통신기기	14,615	일반기계	53,567	석유화학	50,062
선박류	21,261	철강제품	33,996	석유제품	46,762
가 전	7,218	컴퓨터	10,762	섬유류	14,086
평판디스플레이	24,681	단위 백만USD			

2011년 석유소비량

2011년 소비량 알아보기

정의? 잘못된 연관?

한국 1인당 석유소비량

품목별 수출실적 (단위=억달러)

• 2010년(1~7월)

1	반도체	287
2	선박	278
3	화공품	269
4	기계류	243
5	철강	205
6	정보통신기기	204
7	자동차	179
8	석유제품	170

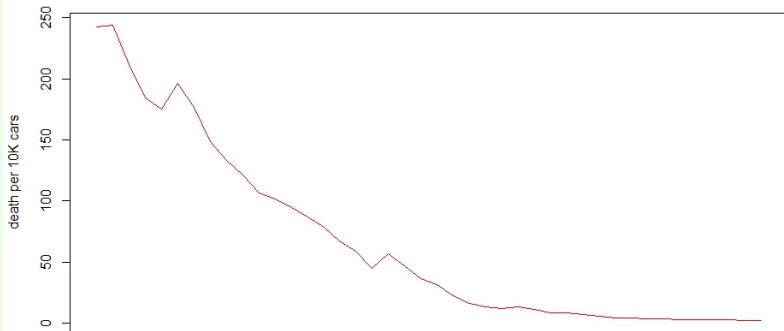
• 2011년(1~7월)

1	선박	361
2	화공품	349
3	기계류	314
4	석유제품	295
5	반도체	292
6	철강	279
7	자동차	231
8	정보통신기기	224

※ 자료=관세청

기사 원문 석유류 수출 통계

자동차 1만대당 사망자수와 카파라치



차량 1만대당 교통사고 사망자수

연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122

엉뚱한 연관 1

차량 1만대당 교통사고 사망자수

연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122
연도	80	81	82	83	84	85	86	87	88	89
사망자	106	102	94	87	79	68	59	45	57	47

차량 1만대당 교통사고 사망자수

연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122
연도	80	81	82	83	84	85	86	87	88	89
사망자	106	102	94	87	79	68	59	45	57	47
연도	90	91	92	93	94	95	96	97	98	99
사망자	36	32	22	17	14	12	13	11	9	8

엉뚱한 연관 1

차량 1만대당 교통사고 사망자수

연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122
연도	80	81	82	83	84	85	86	87	88	89
사망자	106	102	94	87	79	68	59	45	57	47
연도	90	91	92	93	94	95	96	97	98	99
사망자	36	32	22	17	14	12	13	11	9	8
연도	00	01	02	03	04	05	06	07	08	09
사망자	8	6	4.6	4.4	3.9	3.4	3.2	3.1	2.9	2.8

엉뚱한 연관 1

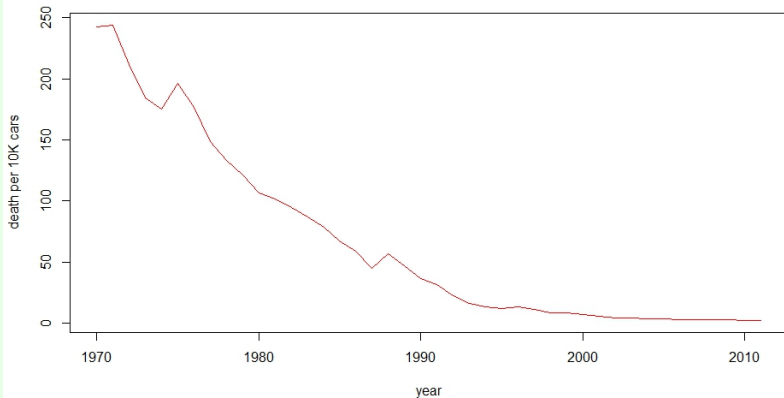
차량 1만대당 교통사고 사망자수

연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122
연도	80	81	82	83	84	85	86	87	88	89
사망자	106	102	94	87	79	68	59	45	57	47
연도	90	91	92	93	94	95	96	97	98	99
사망자	36	32	22	17	14	12	13	11	9	8
연도	00	01	02	03	04	05	06	07	08	09
사망자	8	6	4.6	4.4	3.9	3.4	3.2	3.1	2.9	2.8
연도	10	11	12	13	14	15				
사망자	2.6	2.4	2.4	2.2	2.0	1.96				

카파라치 01.3.-02.12

자료:도로교통공단 (<http://www.koroad.or.kr/>)

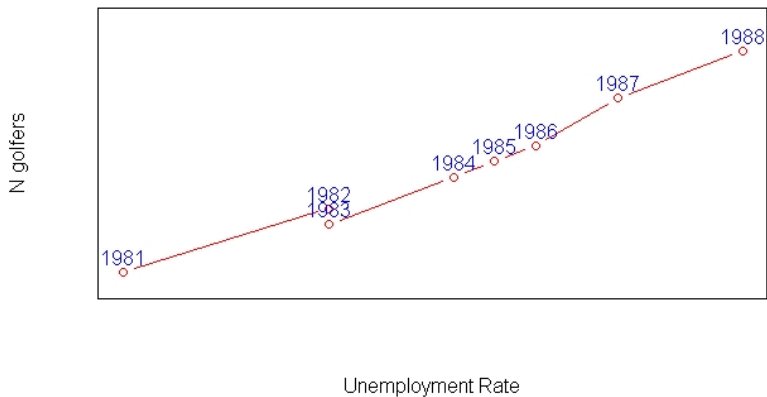
자동차 1만대당 사망자수와 카파라치



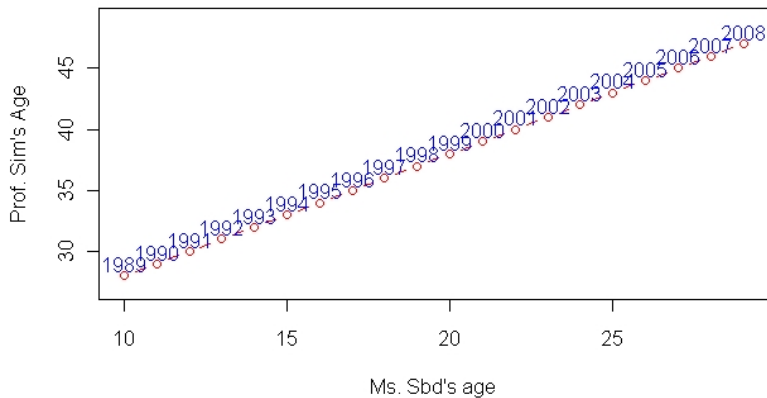
자동차 1만대당 사망자수와 카파라치



골프인구와 실업률



두 사람의 나이



쌀을 먹으면 안되는 100가지 이유

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.
- ③ 교통사고 가해자의 80% 이상이 사고를 내기 10시간 이내에 쌀을 섭취

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.
- ③ 교통사고 가해자의 80% 이상이 사고를 내기 10시간 이내에 쌀을 섭취
- ④ 교도소 수감자를 조사했더니 범죄 24시간 이내에 쌀을 먹은 수감자가 99.8%

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.
- ③ 교통사고 가해자의 80% 이상이 사고를 내기 10시간 이내에 쌀을 섭취
- ④ 교도소 수감자를 조사했더니 범죄 24시간 이내에 쌀을 먹은 수감자가 99.8%
- ⑤ 매일 한 번 이상 쌀을 먹은 경우 100년 이내에 사망할 가능성 99.9% 이상

쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.
- ③ 교통사고 가해자의 80% 이상이 사고를 내기 10시간 이내에 쌀을 섭취
- ④ 교도소 수감자를 조사했더니 범죄 24시간 이내에 쌀을 먹은 수감자가 99.8%
- ⑤ 매일 한 번 이상 쌀을 먹은 경우 100년 이내에 사망할 가능성 99.9% 이상
- ⋮
- ⋮

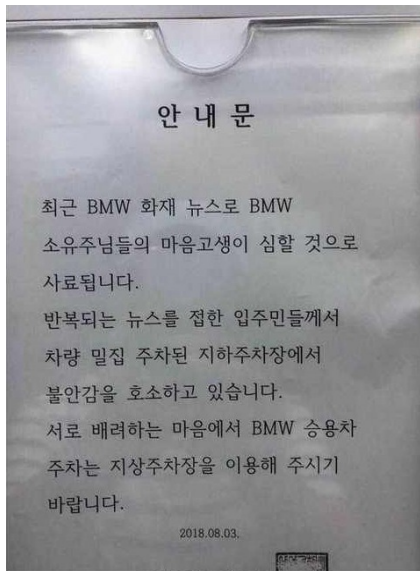
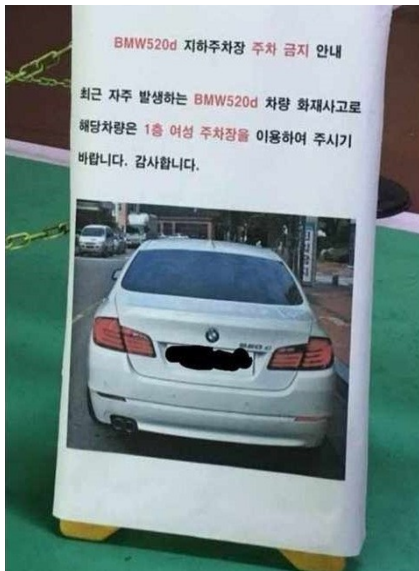
쌀을 먹으면 안되는 100가지 이유

- ① 학교 폭력 가해 학생들의 85% 이상이 하루에 한 번 이상 쌀을 섭취
- ② 미스 코리아 입상자의 경우 다른 사람에 비해 쌀의 섭취량이 적음.
- ③ 교통사고 가해자의 80% 이상이 사고를 내기 10시간 이내에 쌀을 섭취
- ④ 교도소 수감자를 조사했더니 범죄 24시간 이내에 쌀을 먹은 수감자가 99.8%
- ⑤ 매일 한 번 이상 쌀을 먹은 경우 100년 이내에 사망할 가능성 99.9% 이상
- ⋮
- ⋮
- ⑩ 지금 이런 말하는 이 사람도 24시간 이내에 쌀을 섭취하였음.

서로다른 모집단 - 불타는 BMW 520d



서로다른 모집단 - 불타는 BMW 520d



서로다른 모집단 – 불타는 BMW 520d



서로다른 모집단 - 불타는 BMW 520d

주요 자동차 브랜드별, 차종별 화재 발생 현황

	등록대수(대)	화재발생(건)	연평균 화재(건)	1만대당 화재발생(건)
현대 스타렉스	206,523	728	121.33	35.25
현대 포터	1,267,215	2,842	473.67	22.43
기아 봉고	714,260	1,572	262.00	22.01
현대 쏘나타	1,273,904	1,720	286.67	13.50
기아 쏘렌토	226,885	205	34.17	9.04
BMW	253,469	200	33.33	7.89
벤츠	225,460	167	27.83	7.41
기아 스포티지	475,569	226	37.67	4.75
기아 모닝	808,208	224	37.33	2.77
폭스바겐	158,254	42	7.00	2.65
쉐보레 스파크	377,167	43	7.17	1.14

등록대수: 한국자동차공업협회 2015년 12월 기준

화재발생: 소방방재청 화재 발생 출동 기록 기준 2010년 1월~2015년 12월

기사

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함
- 조사기간도 확인

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함
- 조사기간도 확인
- 연식 및 주행거리

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함
- 조사기간도 확인
- 연식 및 주행거리
- 적재함화재 – 담배꽂초

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함
- 조사기간도 확인
- 연식 및 주행거리
- 적재함화재 – 담배꽂초
- 소유주-회사/개인

서로다른 모집단 – 불타는 BMW 520d

- 화재 출동통계로 교통사고, 방화 등이 모두 포함
- 조사기간도 확인
- 연식 및 주행거리
- 적재함화재 – 담배꽂초
- 소유주-회사/개인
- 수입차/국산차 수리비

올 상반기 차량 화재사고 발생 빈도

(단위:만대당 건)



※()안은2017년화재사고발생빈도

자료:민경욱 자유한국당 의원, 소방청

[기사](#)

서로 다른 모집단

교통사고율 남성이 여성보다 3배 이상 높아

남성이 여성보다 3배 이상 교통사고를 더 내는 것으로 나타났다. 난폭운전 등 남성의 잘못된 운전 습관이 가장 큰 이유다.

서로 다른 모집단

교통사고율 남성이 여성보다 3배 이상 높아

남성이 여성보다 3배 이상 교통사고를 더 내는 것으로 나타났다. 난폭운전 등 남성의 잘못된 운전 습관이 가장 큰 이유다.

교통안전공단이 13일 내놓은 "남녀 교통사고 특성 비교"에 따르면 운전면허소지자 100명 당 발생 건수가 **남성은 1.13건인 데 비해 여성은 0.34건으로** 나타나 남성이 여성에 비해 교통사고 발생 빈도가 3.3배나 높았다.

서로 다른 모집단

교통사고율 남성이 여성보다 3배 이상 높아

남성이 여성보다 3배 이상 교통사고를 더 내는 것으로 나타났다. 난폭운전 등 남성의 잘못된 운전 습관이 가장 큰 이유다.

교통안전공단이 13일 내놓은 "남녀 교통사고 특성 비교"에 따르면 운전면허소지자 100명 당 발생 건수가 **남성은 1.13건인 데 비해 여성은 0.34건으로** 나타나 남성이 여성에 비해 교통사고 발생 빈도가 3.3배나 높았다.

교통안전공단 관계자는 "여성운전자가 대체로 남성보다 운행거리가 짧은 탓도 있지만 여성이 순발력 등 운전기능이 남성보다 떨어짐에도 불구하고 급가속, 급출발 등 난폭운전을 안하고 조심운전을 하기 때문인 것으로 풀이된다"고 말했다.

[원문 기사 1](#) [원문 기사 2](#)

서로 다른 모집단

자동차 사고율 '남성 보다 여성, 기혼 보다 미혼'

운전자 가운데 기혼보다는 미혼이 자동차 사고를 더 많이 내는 것으로 나타났다. 또 통계적으로 남성보다는 여성이, 대도시보다는 중소도시 거주자가 더 많이 사고를 내는 것으로 집계됐다.

서로 다른 모집단

자동차 사고율 '남성 보다 여성, 기혼 보다 미혼'

운전자 가운데 기혼보다는 미혼이 자동차 사고를 더 많이 내는 것으로 나타났다. 또 통계적으로 남성보다는 여성이, 대도시보다는 중소도시 거주자가 더 많이 사고를 내는 것으로 집계됐다.

보험연구원이 자동차 운전자 574명을 대상으로 조사한 결과에 따르면 최근 1년 새 운전 중 사고가 발생했다고 응답한 사람은 전체 응답자의 12.7%였으며, 평균 사고 횟수는 1년에 0.15회였다.

서로 다른 모집단

자동차 사고율 '남성 보다 여성, 기혼 보다 미혼'

운전자 가운데 기혼보다는 미혼이 자동차 사고를 더 많이 내는 것으로 나타났다. 또 통계적으로 남성보다는 여성이, 대도시보다는 중소도시 거주자가 더 많이 사고를 내는 것으로 집계됐다.

보험연구원이 자동차 운전자 574명을 대상으로 조사한 결과에 따르면 최근 1년 새 운전 중 사고가 발생했다고 응답한 사람은 전체 응답자의 12.7%였으며, 평균 사고 횟수는 1년에 0.15회였다.

성별로 보면 **여성 운전자의 사고 발생횟수가 0.18회로 남성 운전자(0.13회)보다 더 많았다.** 결혼 여부로 따지면 기혼(0.14회)보다 미혼 운전자(0.19회)의 사고 발생률이 더 높았다.

서로 다른 모집단

자동차 사고율 '남성 보다 여성, 기혼 보다 미혼'

운전자 가운데 기혼보다는 미혼이 자동차 사고를 더 많이 내는 것으로 나타났다. 또 통계적으로 남성보다는 여성이, 대도시보다는 중소도시 거주자가 더 많이 사고를 내는 것으로 집계됐다.

보험연구원이 자동차 운전자 574명을 대상으로 조사한 결과에 따르면 최근 1년 새 운전 중 사고가 발생했다고 응답한 사람은 전체 응답자의 12.7%였으며, 평균 사고 횟수는 1년에 0.15회였다.

성별로 보면 **여성 운전자의 사고 발생횟수가 0.18회로 남성 운전자(0.13회)보다 더 많았다.** 결혼 여부로 따지면 기혼(0.14회)보다 미혼 운전자(0.19회)의 사고 발생률이 더 높았다.

직업별로 보면 **주부의 사고 발생횟수가 0.22회로 가장 높았으며,** 다음은 자영업(0.15회)이었다.

원문 기사 김여사 김사장

사과와 오렌지 비교

현대車, 국내 급발진 1위 ... 불명예
2위는 기아, 3위는 르노삼성

사과와 오렌지 비교

현대車, 국내 급발진 1위 ... 불명예
2위는 기아, 3위는 르노삼성

현대자동차가 국내 급발진 발생 1위를 하는 불명예를 안았다.
19일 강동원 의원에 따르면 2010년부터 올해 7월 말까지 국내 급발진
차량으로 신고된 차량 총 482대 가운데 신고차량 43.6%(210건)가
현대자동차 제조 차량으로 드러났다. 기아차는 16.2%(78건)으로 2위를
기록했으며 르노삼성 차량이 14.7%(71건)로 3위를 차지했다.

사과와 오렌지 비교

현대車, 국내 급발진 1위 ... 불명예 2위는 기아, 3위는 르노삼성

현대자동차가 국내 급발진 발생 1위를 하는 불명예를 안았다. 19일 강동원 의원에 따르면 2010년부터 올해 7월 말까지 국내 급발진 차량으로 신고된 차량 총 482대 가운데 신고차량 43.6%(210건)가 현대자동차 제조 차량으로 드러났다. 기아차는 16.2%(78건)으로 2위를 기록했으며 르노삼성 차량이 14.7%(71건)로 3위를 차지했다. 급발진 신고 차량 가운데 에어백 미전개 차량도 43.4%(209대)로 나타났다. 현대차가 88대와 기아차는 33대, 르노삼성이 33대, 쌍용 56.7% 17대로 나타났다. 수입차로는 벤츠 4대, 도요타 5대도 포함됐다. 신고된 국내 급발진 차량을 분석해본 결과 2006-2010년 제작된 차량이 전체의 40.5%(195대), 2011-2014년 제작된 차량이 전체의 41.7%(201대)를 나타냈다.

[원문기사](#) [유사한 기사](#)

한국 노인빈곤율 OECD 1위

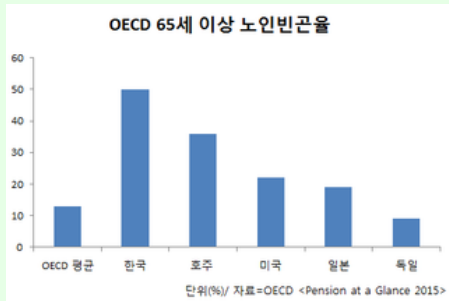
한국 노인빈곤율 OECD 1위

OECD조사에 따르면 2014년 기준 한국의 노인 빈곤율은 49.6%로 OECD 34개 국가 중 가장 높았다. 소득분배의 불평등 정도를 나타내는 지니 계수는 2014년 기준으로 0.302를 기록했는데, 이는 OECD 국가 중 17번째로 중간 정도 수준이다.

다른 시대의 비교

한국 노인빈곤율 OECD 1위

OECD조사에 따르면 2014년 기준 한국의 노인 빈곤율은 49.6%로 OECD 34개 국가 중 가장 높았다. 소득분배의 불평등 정도를 나타내는 지니 계수는 2014년 기준으로 0.302를 기록했는데, 이는 OECD 국가 중 17 번째로 중간 정도 수준이다.



[원문기사](#)
[유사한 기사](#)
[OECD report](#)

간단한 산수문제

- 아빠한테 500원 빌리고 엄마한테 500원 빌려 총 1,000원으로 970원짜리 과자를 샀다.

간단한 산수문제

- 아빠한테 500원 빌리고 엄마한테 500원 빌려 총 1,000원으로 970원짜리 과자를 샀다.
- 거스름돈 30원으로 아빠 10원, 엄마 10원을 각각 갚고 나머지 10원을 내가 가졌다.

간단한 산수문제

- 아빠한테 500원 빌리고 엄마한테 500원 빌려 총 1,000원으로 970원짜리 과자를 샀다.
- 거스름돈 30원으로 아빠 10원, 엄마 10원을 각각 갚고 나머지 10원을 내가 가졌다.
- 그러면 엄마아빠한테 각각 490원씩 빚진건데
 $490 + 490 = 980$ 원에서
내가 가진돈 10원을 포함하면 990원이다.

간단한 산수문제

- 아빠한테 500원 빌리고 엄마한테 500원 빌려 총 1,000원으로 970원짜리 과자를 샀다.
- 거스름돈 30원으로 아빠 10원, 엄마 10원을 각각 갚고 나머지 10원을 내가 가졌다.
- 그러면 엄마아빠한테 각각 490원씩 빚진건데
 $490 + 490 = 980$ 원에서
내가 가진돈 10원을 포함하면 990원이다.
- 나머지 10원은 어디로 갚을까요?

간단한 산수문제-2

- 어느 보석가게에 손님이 70만원짜리 목걸이를 사고 100만원 짜리 수표를 내밀었다.

간단한 산수문제-2

- 어느 보석가게에 손님이 70만원짜리 목걸이를 사고 100만원 짜리 수표를 내밀었다.
- 주인은 거스름돈이 없어 옆집에서 수표를 100만원의 현금으로 바꾸어 손님에게 30만원의 거스름돈을 지불했다.

간단한 산수문제-2

- 어느 보석가게에 손님이 70만원짜리 목걸이를 사고 100만원 짜리 수표를 내밀었다.
- 주인은 거스름돈이 없어 옆집에서 수표를 100만원의 현금으로 바꾸어 손님에게 30만원의 거스름돈을 지불했다.
- 다음 날 옆집에서 어제의 수표가 부도수표라며 환불을 요구하여 100만원을 돌려주었다.

간단한 산수문제-2

- 어느 보석가게에 손님이 70만원짜리 목걸이를 사고 100만원 짜리 수표를 내밀었다.
- 주인은 거스름돈이 없어 옆집에서 수표를 100만원의 현금으로 바꾸어 손님에게 30만원의 거스름돈을 지불했다.
- 다음 날 옆집에서 어제의 수표가 부도수표라며 환불을 요구하여 100만원을 돌려주었다.
- 보석가게의 주인의 손해는 얼마인가?

① 100만원 ② 130만원 ③ 170만원 ④ 200만원

문항설계-Conti.

- 우리나라 사람의 95% 는 잠자기 가장 좋은 시간으로 밤 9시라고 답하였다.

문항설계-Conti.

- 우리나라 사람의 95% 는 잠자기 가장 좋은 시간으로 밤 9시라고 답하였다.

다음 중 잠자기 좋은 시간은?

① 오전 9시 ② 오후 9시

문항설계-Conti.

- 우리나라 사람의 95% 는 잠자기 가장 좋은 시간으로 밤 9시라고 답하였다.

다음 중 잠자기 좋은 시간은?

① 오전 9시 ② 오후 9시

- 우리나라에 믿을 수 있는 정치인도 있다. (Yes, No)
우리나라에 믿을 수 있는 정치인이 있다. (Yes, No)

관찰연구와 인과관계

- 초음파 검사 여부와 신생아의 체중에 대한 관찰연구

검사유무	저체중여부	
	저체중	정상+
유	70	30
무	20	80

(가상자료)

관찰연구와 인과관계

- 초음파 검사 여부와 신생아의 체중에 대한 관찰연구

검사유무	저체중여부	
	저체중	정상+
유	70	30
무	20	80

(가상자료)

- Once started, never stop (smoking)!!

Smoking	Never	Quit	Continued
Life span	83	68	75

(가상자료)

골프 치면 5년 더 장수

골퍼들이 골프를 하지 않는 사람들보다 평균 5년 정도 더 장수를 누린다는 연구 결과가 나왔다고 영국 데일리 메일 신문이 지난 31일 보도했다.

골프 치면 5년 더 장수

골퍼들이 골프를 하지 않는 사람들보다 평균 5년 정도 더 장수를 누린다는 연구 결과가 나왔다고 영국 데일리 메일 신문이 지난 31일 보도했다.

스웨덴 카롤린스카 연구소의 연구진은 30만명의 골퍼들을 대상으로 실시한 연구결과 골프를 하는 사람이 그렇지 않은 사람보다 일정 시점에 사망할 가능성이 40% 더 낮은 것으로 나타났다고 발표했다. 이것은 평균 5년 더 오래 사는 것과 맞먹는다고 연구진은 말했다.

골프 치면 5년 더 장수

골퍼들이 골프를 하지 않는 사람들보다 평균 5년 정도 더 장수를 누린다는 연구 결과가 나왔다고 영국 데일리 메일 신문이 지난 31일 보도했다.

스웨덴 카롤린스카 연구소의 연구진은 30만명의 골퍼들을 대상으로 실시한 연구결과 골프를 하는 사람이 그렇지 않은 사람보다 일정 시점에 사망할 가능성이 40% 더 낮은 것으로 나타났다고 발표했다. 이것은 평균 5년 더 오래 사는 것과 맞먹는다고 연구진은 말했다.

이 연구에서 핸디캡으로 따져 최고의 플레이어들이 가장 건강한 사람인 것으로 나타났다. 최고의 골퍼는 같은 연령의 골프를 하지 않는 사람에 비해 일정 시점에서 사망할 확률이 47% 더 낮았다. 연구진은 골프가 육체적으로 격렬한 운동은 아니지만, 18홀 한 라운드가 보통 4마일(약 6.4km) 이상 걷기를 수반하는 건강에 좋은 운동이라고 말했다.

골프 치면 5년 더 장수

이 연구를 이끈 카롤린스카 연구소의 안더스 아봄 교수는 "한 라운드의 골프는 필드에서 6-7km의 거리를 빠른 속도로 걸으며 4-5시간 있다는 것을 뜻한다"며 걷기는 이미 건강에 좋은 것으로 잘 알려져 있다고 말했다.

골프 치면 5년 더 장수

이 연구를 이끈 카롤린스카 연구소의 안더스 아봄 교수는 "한 라운드의 골프는 필드에서 6-7km의 거리를 빠른 속도로 걸으며 4-5시간 있다는 것을 뜻한다"며 걷기는 이미 건강에 좋은 것으로 잘 알려져 있다고 말했다.

아봄 교수는 "낮은 핸디캡을 유지한다는 것은 골프를 많이 친다는 말이 되고, 결국 골프라는 운동 자체가 건강에 좋다는 것을 말한다"고 설명했다. 이와 함께 다른 사람들과 어울리는 골프의 사교적 성격도 수명을 늘리는 데 긍정적인 요인으로 작용할 수 있다고 연구진은 말했다. (런던=연합뉴스)

[원문기사](#)

실험연구-Controlled Experiment

- 처리 수준의 결정자

실험연구-Contolled Experiment

- 처리 수준의 결정자
 - ① 맹검(blind test)

실험연구-Contolled Experiment

- 처리 수준의 결정자
 - 1 맹검(blind test)
 - 2 이중 맹검(double blind test)

실험연구-Contolled Experiment

- 처리 수준의 결정자
 - 1 맹검(blind test)
 - 2 이중 맹검(double blind test)
 - 3 ANOVA

실험연구-Contolled Experiment

- 처리 수준의 결정자
 - ① 맹검(blind test)
 - ② 이중 맹검(double blind test)
 - ③ ANOVA
- 실험연구 $\Rightarrow$ 인과관계(cause & effect)

실험연구-Contolled Experiment

- 처리 수준의 결정자
 - ① 맹검(blind test)
 - ② 이중 맹검(double blind test)
 - ③ ANOVA
- 실험연구 ⇒ 인과관계(cause & effect)
- 관찰연구 ⇒ 연관 (association)

신용카드 사기

Credit Card Fraud Detection System(FDS)

- 장기 미사용 신용카드
- 전자제품 등 현금화가 쉬운 제품 구입
- 짧은 시간에 많은 결제
- ⋮

⇒ 높은 확률의 불법 사용

⇒ 승인거부

신용카드 사기

Credit Card Fraud Detection System(FDS)

- 장기 미사용 신용카드
- 전자제품 등 현금화가 쉬운 제품 구입
- 짧은 시간에 많은 결제
- ⋮

⇒ 높은 확률의 불법 사용

⇒ 승인거부

신용카드 사기

- 보이스 피싱에도 응용 가능
- 금융정보분석원이 수집하는 정보

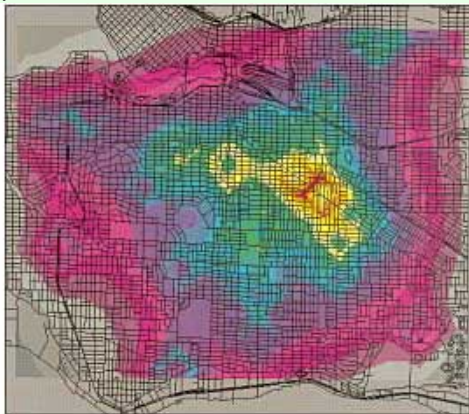
지리적 프로파일링

Geographical profile: 캐나다 밴쿠버의 연쇄 무장강도 사례

Figure 6.1

A geoprofile of
a series of armed
robberies in
Vancouver, British
Columbia, Canada.

Source: Det. Insp.
D. Kim Rossmo,
Vancouver,
British Columbia,
Police Department.
Reproduced by
permission.



지리적 프로파일링

범죄 발생 예측에도 사용: 미국 LA 의 발생확률

GIS & Investigations: Forecasting Crime

Figure 6.2

A map showing probability boundaries in the Los Angeles, California, Motorcycle Bandit case.

Source: Geggler, 1998, Reproduced by permission.



⇒ 높은 확률에 더 많은 순찰 배치

- 경주시의 관광객수 추정(통계청)

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용
- 대학입시와 판별분석 $\Rightarrow$ 신입사원 채용

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용
- 대학입시와 판별분석 $\Rightarrow$ 신입사원 채용
- 신체치수 사이의 관계(허리둘레와 목둘레)

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용
- 대학입시와 판별분석 $\Rightarrow$ 신입사원 채용
- 신체치수 사이의 관계(허리둘레와 목둘레)
- 무응답에 대한 판별분석

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용
- 대학입시와 판별분석 $\Rightarrow$ 신입사원 채용
- 신체치수 사이의 관계(허리둘레와 목둘레)
- 무응답에 대한 판별분석
- 군집분석

통계와 빅데이터 사용

- 경주시의 관광객수 추정(통계청)
- 장바구니 분석
- 보험료 산정(국민/공무원/사학/군인 등 공적연금은 성차별을 하는가?)
- 파생금융상품(ELS, ELT 등)의 설계
- 교통모형과 시뮬레이션
- Web log의 활용
- 대학입시와 판별분석 $\Rightarrow$ 신입사원 채용
- 신체치수 사이의 관계(허리둘레와 목둘레)
- 무응답에 대한 판별분석
- 군집분석
- 비정형자료의 분석(SNS text mining)

감사합니다!!