

## 주요내용

- 통계란?
- 통계해석
  - 숫자로 이루어진 통계 해석
  - 그림으로 이루어진 통계 해석
  - 엉뚱한 연관
  - 문항설계
- 관찰연구와 실험연구
- 연관성 분석
  - 연속형  $\Rightarrow$  상관분석
  - 이산형  $\Rightarrow$  카이제곱 검정
- 통계의 사용 보기

## 이 세상의 세 가지 거짓말-Benjamin Disraeli

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

## 문화인의 기본 소양-H. G. Wells

통계적 사고가 읽고 쓰기만큼 시민의 기본 소양이 되는 날이 올 것이다.

## 통계(Statistics)의 어원

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

# 생활 속의 확률

## 기대값

조폐공사나 한국은행을 털지 않는 도둑

## 확률 1

광화문 네거리를 무단횡단해 볼까?

## 확률 2

눈/비율 확률 50%?

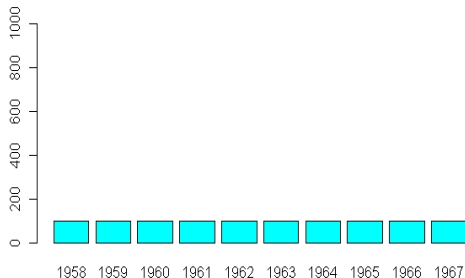
## 확률 3

로또 1등 확률  ${}_{45}C_6 = 8,145,060$

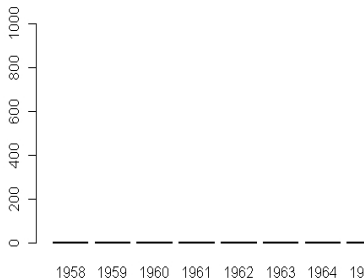
## 광고

- 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

N of cars in 1,000



N of cars in 1,000

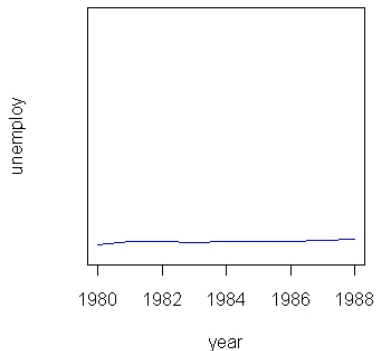
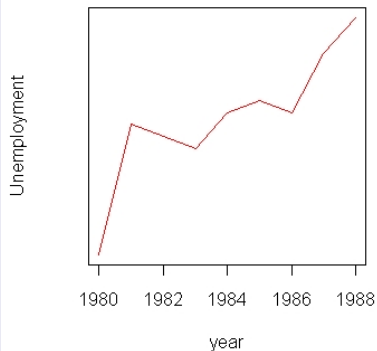


## 평균?, 중앙값?

- 달동네 주민의 월평균 소득 10,000,000
- 실제: 10억 1명, 0이 99명
- 프로야구 선수의 평균 연봉이 6790여만 원
- 프로야구 선수의 56%가 연봉 3000만 원 미만.
- 1억 원이 넘는 선수는 10% 남짓
- <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=001&oid=022&aid=0000078894&>

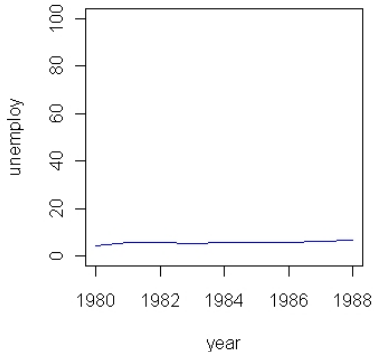
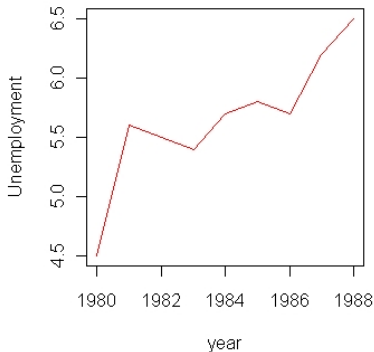
# 그림에 대한 검토

## 실업율에 대한 두 개의 그래프

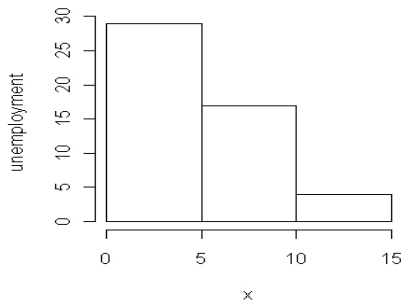
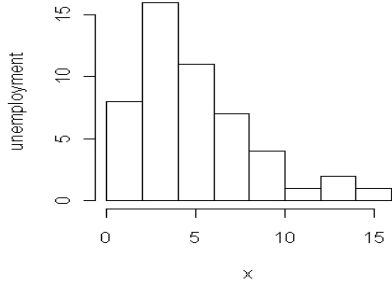


# 그림에 대한 검토

## 실업율에 대한 두 개의 그래프



## 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	

## 기동 수

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

등

# 정의

## 실업율

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?

## 근로자 가구 소득조사

2010년 1/4분기 월평균

경상소득	4,012,580
근로소득	3,582,857
사업소득	165,323
재산소득	6,335
이전소득	248,065
비경상소득	161,615
소득합계	4,174,197

자료 : 통계청 사회통계국 사회복지통계과  
용어정의

# 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)

## 2006년 13대 수출품목(산자부) 자료백업

자동차	32,922	자동차부품	10,230	반도체	37,360
무선통신기기	27,018	일반기계	23,920	석유화학	24,099
선박류	22,123	철강제품	19,429	석유제품	20,404
가전	14,553	컴퓨터	12,576	섬유류	13,232
액정디바이스	12,212	단위 백만USD			

# 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

<http://news.mk.co.kr/newsRead.php?year=2011&no=544607>

## 교수들 나눠먹기식 자리할당...10명에 3명이 보직교수

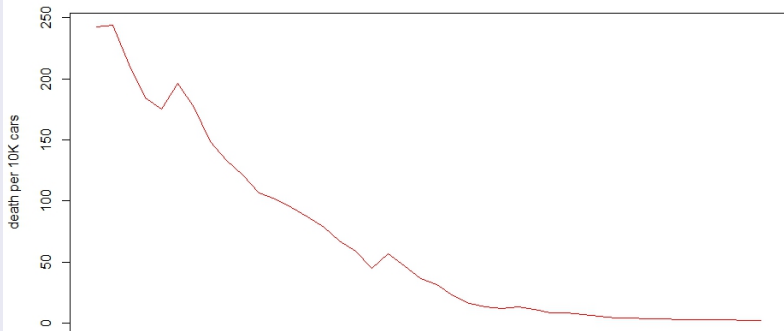
대학의 보직교수가 전임교수 10명당 3명꼴이나 될 정도로 많아 방만한 학교운영과 예산낭비를 초래하고 있다는 지적을 받고 있다. 8일 교육부에 따르면 전국 26개 국공립대학의 보직교수는 전임교수 1만1천1명 가운데 29.5%인 3천2백54명인 것으로 나타났다.

사립대도 지난해 4월 현재 전임교수 2만7천6백63명 가운데 29.2%인 8천74명으로 집계됐다. 사립대는 학교에 따라 보직교수 비율이 6.6%에서 100%까지 큰 차이가 있는 것으로 나타났다.

서울대의 보직교수는 전임교수 1천4백42명 중 3백68명으로 25.5%였고 △부산대 23.2% △서울시립대 34.9% △경북대 37.6% △공주대 39.3% 등이며 부경대는 무려 43.9%나 됐다.

국공립대의 경우 보직교수는 처장의 경우 많게는 월 2백만원, 학장보의 경우 월 50만원 정도의 수당을 받고 있고 직책에 따라 강의시간이 줄고 연금산정에서 유리해지는 등의 혜택을 받고 있다. (이진녕기자) 1998/05/09(토) 06:39 <http://jupiter.hallym.ac.kr/>

## 자동차 1만대당 사망자수와 카파라치



## 차량 1만대당 교통사고 사망자수

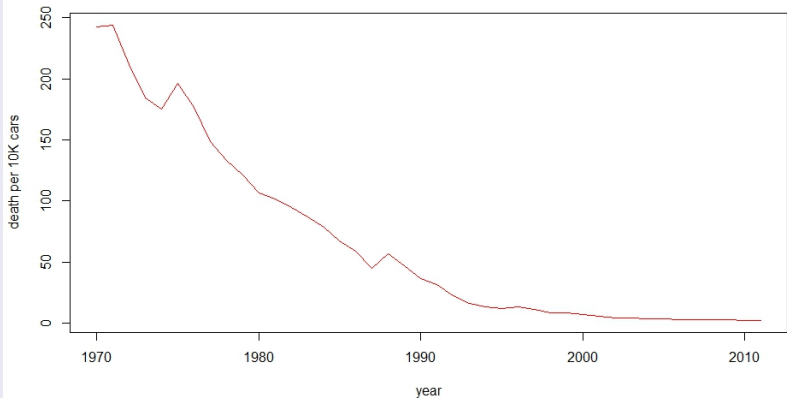
연도	70	71	72	73	74	75	76	77	78	79
사망자	243	244	211	184	176	195	176	149	133	122
연도	80	81	82	83	84	85	86	87	88	89
사망자	106	102	94	87	79	68	59	45	57	47
연도	90	91	92	93	94	95	96	97	98	99
사망자	36	32	22	17	14	12	13	11	9	8
연도	00	01	02	03	04	05	06	07	08	09
사망자	8	6	4.6	4.4	3.9	3.4	3.2	3.1	2.9	2.8
연도	10	11								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

카파라치 01.3.-02.12

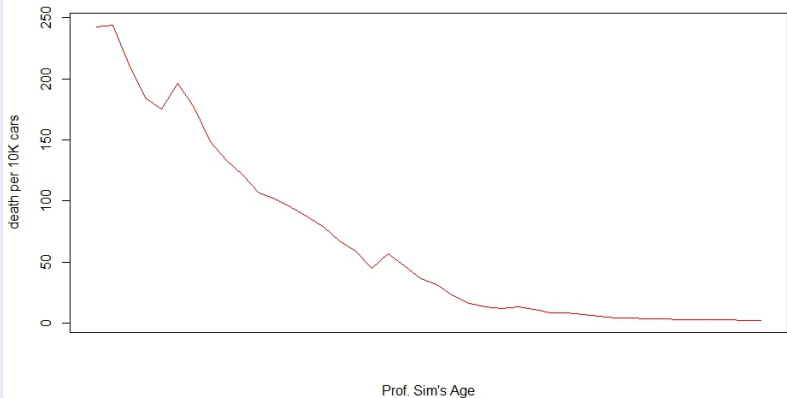
자료:도로교통공단 (<http://www.koroad.or.kr/>)



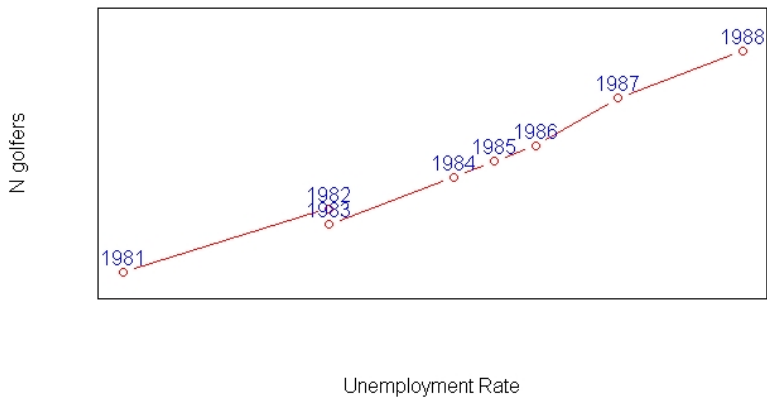
## 자동차 1만대당 사망자수와 카파라치



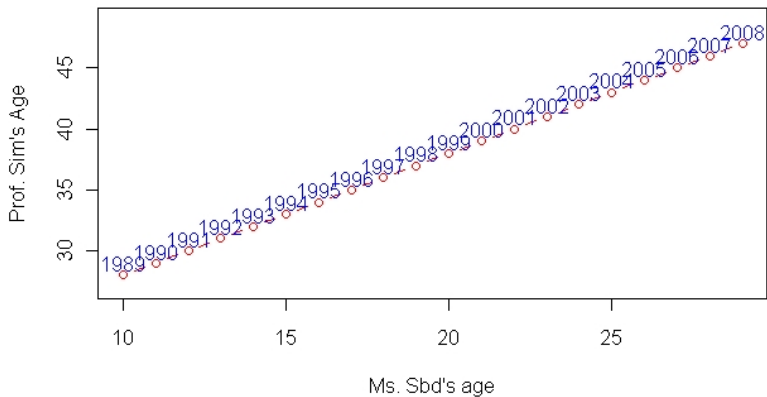
## 자동차 1만대당 사망자수와 카파라치



## 골프인구와 실업률



## 두 사람의 나이



## 담배와 건강보험료

(단위: 억원)

구분		2006년	2007년	2010년	2011년
수입	계	232,631	260,498	343,822	389,151
	보험료등	188,106	217,287	284,577	329,221
	국고지원금 등	35,994	70,253	48,614	50,362
	담배부담금	9,664	9,676	10,631	9,568
지출	계	228,178	258,885	-	-
	보험급여비	215,880	245,601	337,493	358,302
	관리운영비	7,785	7,196	-	-
	기타 지출	4,513	6,080	-	-

(자료: 건강보험통계 <http://www.nhic.or.kr>)

# 연령대별 1인당 월의료비

## 2010년 연령대별 진료비-자료:국민건강보험 보도자료

연령대	금액(억원)	인구비중	진료비비중	월진료비(원)
9세 이하	33,374	10.1	7.6	60,646
10대	20,180	13.3	4.6	27,805
20대	24,810	13.2	5.7	34,359
30대	40,819	16.4	9.4	45,776
40대	56,628	17.0	13.0	61,162
50대	80,364	14.1	18.4	104,305
60대	80,543	8.6	18.5	171,332
70세+	99,565	7.2	22.8	252,831

<http://www.nhic.or.kr/>

# 65세 이상인구와 보험료

2011년 65세이상 노인 진료비 15조 3,768억원 (33.3%)

구 분	2006	2007	2010	2011
적용인구(천명)	47,410	47,820	48,160	48,614
노인인구(천명)	4,073	4,387	4,979	5,184
점유율 (%)	8.6	9.2	10.2	10.5
총진료비(억원)	284,103	323,892	436,283	462,379
노인진료비(억원)	73,504	91,189	141,350	153,768
점유율 (%)	25.9	28.2	32.4	33.3
1인당 월평균진료비(원)	150,400	173,217	236,588	247,166

자료: 국민건강보험공단

# 사과와 오렌지 비교하기

## ① (2005년 행자부)

개인별로는 총인구 (주민등록인구 4천 878만 명) 의 27.3%에 해당하는 1천 334만 명이 토지를 소유하고 있고 세대별로는 주민등록총세대수 (1천 785만 세대) 의 59.9%에 해당하는 1천 70만 세대가 토지를 보유하고 있는 것으로 집계됐다.



- ② 부동산관련 세금과 자동차 관련세금 - 미국과 우리나라 비교.
- ③ 국민부담율 = 조세부담율 + 사회보장기여금  
한국 25.3%, 미국 25.6%, 일본 25.3% 과 유럽국가들(2003년 기준, 자료:OECD)

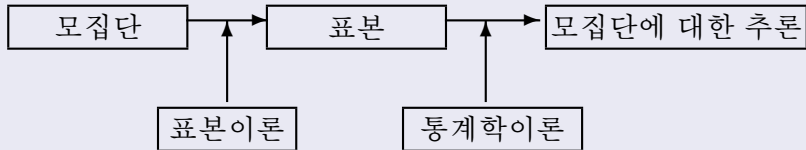


# 각국의 Gini 계수

나라	기준년	계수	나라	기준년	계수
Australia	1994	35.2	Bangladesh	2000	33.4
Belgium	2000	33.0	Brazil	2004	57.0
Canada	2000	32.6	Belgium	2000	33.0
Chile	2003	54.9	China	2004	46.9
Hong Kong-China	1996	43.4	France	1995	32.7
Germany	2000	28.3	India	2004	36.8
Italy	2000	36.0	Japan	1993	24.9
Korea, Rep.	1998	31.6	Malaysia	1997	49.2
Mexico	2004	46.1	New Zealand	1997	36.2
Norway	2000	25.8	Philippines	2003	44.5
Russian Fed	2002	39.9	Singapore	1998	42.5
Spain	2000	34.7	Sweden	2000	25.0
Thailand	2002	42.0	United States	2000c	40.8

자료: 세계은행(World Bank; <http://www.worldbank.org/>)

## 통계적 자료분석에 대한 흐름도



# 대표성과 표본

## 모집단(population)

관심의 대상 전체

## 표본(sample)

모집단의 일부

- 간을 맞추기 위해 국을 다 마실 필요는 없다
- 하지만 국을 잘 저은 후에
- 모집단(술 전체)에서 정확한 표본(국자) ⇒ 대표성 확보
- 휘젓지 않은 국 ⇒ 특정지역에서만 여론조사

## 확률표본과 비확률표본

- 확률표본(probability sampling)
  - 단순임의추출(simple random sampling)
  - 계통추출(systemtic sampling)
  - 집락추출(cluster sampling)
  - 층화추출(stratified sampling)
- 비확률표본(nonprobability sampling)
  - 누적표본추출(snowball sampling) - 피라미드식
  - 편의추출(convenience sampling) - 지나가는 사람 100명에게 물었습니다.
  - 할당표본추출(quota sampling)

## 문항설계

- 우리나라의 발전을 저해하는 요소는 부패와 대충대충하는 적당주의이라고 생각한다. (Yes, No)
- 어떤 스캔들에 휘말린 것으로 보도된 공직자가
  - ① 그 스캔들이 사실이면 물러나야 한다.
  - ② 무조건 물러나야 한다.
  - ③ 물러날 필요가 없다.
- 우리나라에 믿을 수 있는 정치인도 있다. (Yes, No)  
우리나라에 믿을 수 있는 정치인이 있다. (Yes, No)
- ‘소련의 기자가 워싱턴에서 자유롭게 취재활동을 할 수 있도록 허락해야 하는가?’ (Yes, No)  
‘미국의 기자가 모스크바에서 자유롭게 취재활동을 벌일 수 있어야 한다고 생각하는가?’ (Yes, No)

## 문항설계-Conti.

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

다음 중 잠자기 좋은 시간은?

- ① 오전 9시 ② 오후 9시

## 문항설계-Conti.

- '우리 군인'과 '미군' (미국 월남전 여론조사)
- OR의 시간 예측 서비스에 만족하십니까?
- 범행의 동기 및 원인, 피고인의 태양
- 불만족의 원인 아닌 것은?

## 관찰연구와 인과관계

- 초음파 검사 여부와 신생의 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	75	68	71

(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.

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.

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)

<http://www.donga.com/fbin/output?f=total&n=200806010008&top20=>

## 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)