Data, More Data, and Better Data

고려대학교 통계학과 이재창

한국조사연구학회
2006.12.1

초기단계에는 전수조사가 사회조사에서 주류를 이루고

1. B. Seebohm Rowntree(1901), "Poverty: A Study of Town Life"

2. A. N. Kiaer (1897), More penetrating, more detailed and increased scope of partial investigation by an approximate miniature of the population

3. Lucien March (1901, 1903), introduced the concepts of simple random sampling

4. Arthur L. Bowley (1906), an empirical verification to central limit theorem for simple random sampling

5. J. Neyman (1934), theory of point and interval estimation under randomization

6. Hansen and Hurwitz (1943), extented the idea of sampling with unequal inclusion probabilities for units in different strata

20세기 이전의 추론에 쓰인 Data

1. **Byproduct Data 의 예**

John Arbuthnot (1710) 은

"An argument for Divine Providence" 에서

1629-1710 사이의 London 의 아기들 의 세례기록으로 부터

H : "남녀 성비가 같다" 는 가설하에서 82년간 매년 남자아이가 더 많이 태어날 확률 즉 P-value 는 (1/2)의 82승 과 같다는 결론을 도출

2. Experimental Data의 예

Mendelian Genetics:

- 완두콩 실험 결과
- AA Aa aa
- 35 67 30
- Chi 자승 값 이 0.41 로 p-value 는 0.80 정도로 다항분포모형 (¼ , ½ , ¼)에 너무나 도 잘 적합하여 "Mendel 의 연구조수들이 매우 큰 도움을 주었다"고 들 이야기한다

Oskar Morgenstern 의 저서

"On the Accuracy of Economic Observations" (1963), Princeton U. Press

통계적 Data (또는 Observations) 가 사회과학과 자연과학에서 크게 다른 것을 지적하고 있다. 이러한 차이점이 현재에도 크게 달라 진 것이 없다.

자연과학에서는

- in the natural sciences the producer of the observations is usually also their user. If he does not exploit them fully himself, they are passed on to others who, in the tradition of the sciences, are precisely informed about the origin and the manner of obtaining these data. Also, the quality of the work of the observers is well-known, and this contributes to establishing a level of precisions of and confidence in the information. Even the most abstract theorists are exceedingly well informed about the precise nature, circumstances, and limitations of experiments and measurements.

사회과학에서는

It is not often feasible (because of the mass observations or for reasons of general lack of information) to be aware of the detailed nature of the data. Summarization of data is often performed by widely separated statistical workers who are likewise far removed from the later users. The tradition has simply not yet fully established itself for the users to insist upon being fully informed about all steps of the gathering and computing of statistics.

Data, Data, Data, data, ...

1. The stock of scientific knowledge is currently doubling every five to seven years, by the year 2020 it will double every 70-75 days.
2. Of all the scientists who ever lived, more than 90 percent are alive today.
3. Moore's law: Performance of technology will double every 18 months
4. The ICT revolution and the trend towards an increasingly integrated world economy are expected to continue in the next few decades. In an advanced society, an employed person now handles more information in a year than a person did in a lifetime at the beginning of 1900s

Rapid growth of supply and demand of information

1. New information sources, new ways of linking different data sources
2. New technology and statistical methods are accessible to all producers
3. Increased competition in the information market -- increased capabilities for competing organizations to produce 'statistics'
4. Diversification of user's data needs, more demanding users
5. Economies of scale in the information market ?

이런 문제를 해결하여 줄 수 있는 Meta-Data

- Meta-data is data *about* data.
- Meta-data is any information that might be needed to ensure a correct analysis or interpretation of a set of data or a statistical summary
- At the technical end it covers information about the location and content of data files, including the format of data records, the type, name and description of variables and the coding and labeling of classifications. At the opposite extreme are the abstract definitions of the concepts and terms underlying an investigation. In between are many diverse topics, such as sample design for surveys, and the correspondences between different versions of classifications.

Meta-data 의 중요성

To share data we need meta-data, whether simply transferring data to a colleague for further exploratory or confirmatory work, or co-ordinating members of a team collecting a set of data, or combining data from multiple sources, or publishing statistical summaries.

If we view data as a resource, as is the case with a data archive or a statistical production agency, then the importance of meta-data is clear. Secondary users of data must be able to discover the correct interpretation of all aspects of the data.

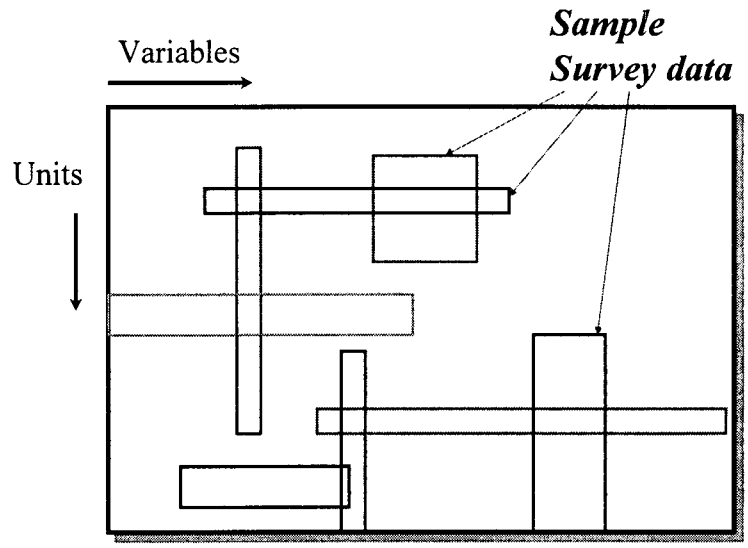
The availability of global networking through the Internet has introduced a wider issue. Potential users of information must be able to discover that information exists.

Data 의 홍수 속에서 "진주"를 찾을 수 있을까?—Better Data?

1. Data 의 결합으로 하나의 조사에서 는 구하지 못한 또 다른 변수들을 포함한 좋은 Data set 을 만들 수 없을까?
2. 다양한 조사에서 얻은 Data 들을 지속적으로 한의 file에 축적하여 더욱 유용한 data set 을 구 축할 수 없을까?
3. 또 다른 형태의 결합 (Data Matching)을 통해 paired experiment 가 현실적으로 불가능할때 Control Group 의 data 를 인위적으로 만들어 비교할 수 없을까?

1. Data 의 결합으로 하나의 조사에서 는 구하지 못한 또 다른 변수들을 포함한 좋은 Data set 을 만들 수 없을까?

   Combining Data from different Sources



<figure 2>

다양하고 지속적으로 생산되는 엄청난 Data에 따른 ombining Data 의 필요성

1. 시간적제약
2. 경제적제약
3. 긴 질문서: 응답질의 저하
4. 더 빈번한 조사: 응답자 부담, 참여율의 저하

통계적 결합

정확한 결합(linkage)

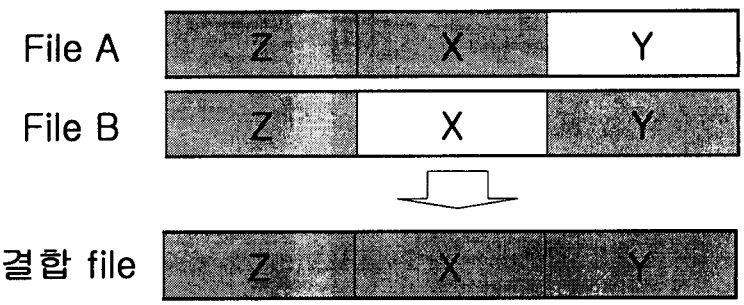
- 개인 식별 가능한 변수(주민등록번호)가 존재하는 자료의 결합 방법
- 두 자료에서 동일한 개체를 완전한게 결합하는 장점

- 개인식별 가능한 변수는 사생활 침해의 소지가 있으므로 자료에서 제거해야 함
- 서로 다른 모집단을 대상으로 한 자료는 동일한 개체가 없는 경우가 많음

통계적 결합

- 개인 식별 가능한 변수가 없거나 동일한 개체가 없는 자료의 결합 방법
- 공통으로 존재하는 변수를 통해 가장 유사한 특성을 갖는 개체를 결합
- 개체간의 근사성을 계산하는 방법에 따라 결합 결과의 차이 발생

통계적 결합



- 공통변수(Z)를 이용하여 File B의 Y를 File A에 추가
- File A를 수용파일(Receipt), File B를 제공파일(Donor)

통계적결합의 가정

1. 결합하는 두 개 이상의 Data Set 은 a set of common variables 를 내포한다
2. 각각의 data set 들은 unit 들이 다르다. (disjoint sets of units)

(* Record Linkage(또는 정확한 결합) 와의 비교: 위의 2 항의 위배, 각각의 data set 가 같은 units 를 가진다)

통계적결합 의 접근 방법

1. Micro 접근: 결과적으로 완전한 synthetic file 을 목적한다

2. Macro 접근: Source files 로 joint distribution function 이나 Correlation coeffcient 같은 분포의 특성을 추정하는 데 쓰인다.

통계적 결합

결측치 대입과 통계적 결합 비교

결측치 대입

- 관찰하지 못한 개체의 값을 다른 개체의 관측값을 이용해 추정하여 대입
- 자료의 크기는 변하지 않음
- 실제 관측값을 대입할 경우 통계적 결합과 유사함

통계적 결합

- 알고자 하는 변수가 없는 상태에서 그 변수가 존재하는 자료를 결합
- 변수 추가로 인해 자료의 크기는 증가
- 공통변수를 통해 개체간 근사성 계산
- 결측치 대입 방법의 적절한 활용 가능

최근접이웃(Nearest Neighbor) hot deck 방법

- **공통변수 Z 만을 이용해 가장 유사한 개체를 찾는 방법**
- **범주형 변수**
  - 같은 값을 갖는 개체를 그룹화
- **연속형 변수**
  - 그룹내에서 거리 측정 함수(Euclid, Mahalanobis)를 통해 개체간 근사성 계산
- 단점
  - 범주형 변수가 증가할수록 그룹수의 증가
  - 변수의 중요도가 거리 계산에 반영 안됨

예측 평균을 이용한 방법

결합 변수가 하나인 경우

- 공통변수와 결합변수를 통한 결합

$$y_i = \beta_0 + \beta_1 z_{1i} + \Lambda + \beta_p z_{pi} + e_i$$

- 제공파일에 의해 모수 추정
- 추정된 모수를 통해 제공파일과 수용파일의 예측 평균을 구함

$$\hat{\mu} = Z\hat{\beta}$$

- 예측평균이 가장 비슷한 개체를 찾아 Y변수의 값을 수용파일에 추가

수용파일

Unit no.	Z	X	예측평균
1			6.43
2			7.68
...			
n_A			5.90

제공파일의 Y_2 의 값을 수용파일 첫번째 개체에 추가

제공파일

Unit no.	Z	Y	예측평균
1			8.42
2			6.37
...			
n_B			6.98

예측평균을 이용한 방법의 장점

- **예측평균은 개체의 결합에만 이용**
  - 잘못된 모형 설정시 결과에 미치는 영향이 적음
- **공통변수의 정보를 예측평균이라는 하나의 상수로 변환**
  - 공통변수의 증가에 민감하지 않음
  - 공통변수의 중요도가 모수를 통해 적절히 반영
- **결합변수의 분포를 유지하기 위해 예측평균에 오차를 추가**
  - 정규분포에서 추출하거나 정규성 가정에 의존하지 않기 위해 잔차의 분포를 이용

결합 변수가 둘 이상인 경우

- 각각의 변수에 대한 선형 모형 설정(결합변수 q개)

$$y_{1i} = \beta_0 + \beta_1 z_{1i} + \Lambda + \beta_p z_{pi} + e_i$$
$$y_{2i} = \beta_0 + \beta_1 z_{1i} + \Lambda + \beta_p z_{pi} + e_i$$
$$y_{qi} = \beta_0 + \beta_1 z_{1i} + \Lambda + \beta_p z_{pi} + e_i$$

- 결합변수의 연관성을 유지해야 함
  - 예측평균벡터 $\{\hat{\mu}_1, \hat{\mu}_2, \Lambda, \hat{\mu}_q\}$를 통한 거리 계산
  - 마할라노비스 거리를 계산하여 결합

$$d(i,j) = [(\hat{\mu}_i - \hat{\mu}_j)' S_{\hat{\mu}}^{-1} (\hat{\mu}_i - \hat{\mu}_j)]^{1/2}$$

범주형 변수가 존재하는 경우

- 범주가 둘인 경우
  - 로지스틱 회귀분석을 통한 예측확률을 예측평균으로 이용

- 범주가 셋 이상인 경우
  - 예측확률을 구할 수 없음
  - 두개의 범주를 갖는 지시변수로 변환(K개의 범주 => K-1개의 지시변수)

$$I_1 = \begin{cases} 1, & \text{범주가 1인경우} \\ 0, & \text{그렇지 않은 경우} \end{cases} \quad \Lambda \quad I_{K-1} = \begin{cases} 1, & \text{범주가 } K-1\text{인경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

  - 결합변수가 증가하는 단점이 있지만 예측평균 방법의 이용 가능

순위합을 이용한 근사성 계산

$$R_i = Rank(|\mu_i^d - \mu_{ij}^r|)$$

μ_i^d 제공파일의 예측평균벡터, μ_{ij}^r 수용파일의 예측평균

순위합 = $\sum_{i=1}^{q} R_i$ q는 결합변수의 수

- **가장 낮은 순위합을 갖는 개체를 결합**
  - 가장 낮은 순위의 개체가 둘 이상인 경우 임의로 하나의 개체 선택
  - 결합변수들의 분포의 차이를 제거

사례

2004년 도시가계 자료

3847가구	공 통 변 수	결 합 변 수
범주형 변수	성별, 학력, 배우자 유무, 가구형태 (맞벌이,노인, 모자, 그외) 거처구분 (아파트, 단독, 연립다세대)	입주형태 (자가, 전세, 월세)
연속형 변수	가구원수, 취업인원수, 연령, 연간소득, 월세평가액, 사용면적(평), 월소득, 기타수입, 자동차소유대수, 세대수	월 가계지출

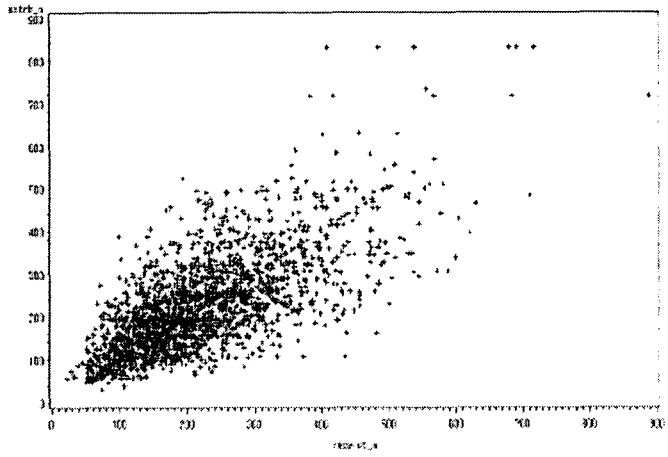
- 자료를 두개의 파일로 분리 후 하나의 파일에서 결합변수 제거
- 순위합 방법과 마할라노비스 거리를 이용한 방법의 결과 비교
  - 범주형 변수 : 명중률(hit rate)로 판단
  - 연속형 변수 : (참값 – 결합값) 분포에 의해 판단

- 마할라노비스(Mahalanobis)거리를 이용한 방법

		결합된 값			
		자가	전세	월세	정확결합
참값	자가	1141	77	30	91.4%
	전세	69	341	88	68.5%
	월세	22	86	69	40.0%
	계				80.7%

- 전체적으로 정확한 결합이 이루어짐
- 전세와 월세에서 공통변수의 차이가 거의 없어 결합 정확도가 낮음
- 자가 대 (전세, 월세)를 고려할 경우 결합의 정확도는 89.7%로 증가

- 마할라노비스(Mahalanobis)거리를 이용한 방법



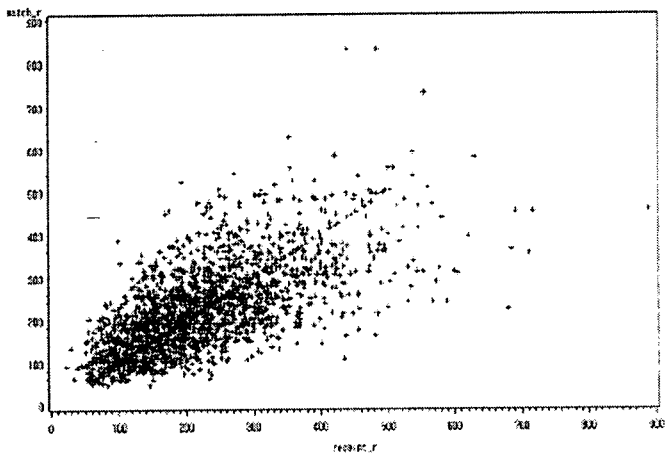
최소값	0
25백분위수	21
중위수	49
75백분위수	91
최대값	423

- 순위합을 이용한 방법

		결합된 값			
		자가	전세	월세	정확결합
참값	자가	1149	84	15	91.4%
	전세	69	347	82	69.7%
	월세	18	67	91	51.4%
	계				82.5%

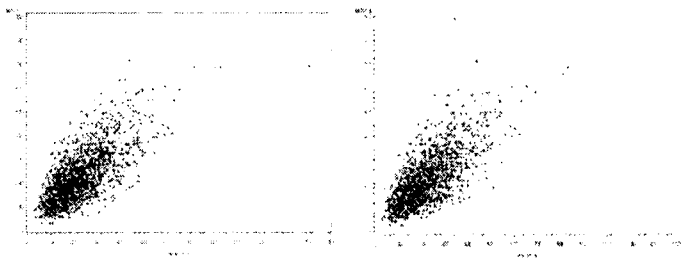
- 마할라노비스 거리를 이용한 방법보다 더 정확한 결합 결과
- 월세의 경우 결합 정확도가 10%이상 증가

- 순위합을 이용한 방법



최소값	0
25백분위수	20
중위수	49
75백분위수	90
최대값	456

- 연속형 변수만을 결합한 경우(월 가계지출, 기타수입)



	마할라노비스	순위합
최소값	0	0
25백분위수	22	20
중위수	47	49
75백분위수	85	87
최대값	808	521

2. 다양한 조사에서 얻은 Data 들을 지속적으로 한의 file에 축적하여 더욱 유용한 data set 을 구축할 수 없을까?

Data 를 담을 바구니가 필요—

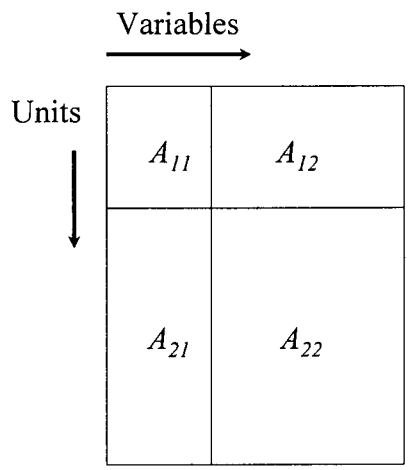
Census data 로 만든 cluster 를 활용가능

Combining Data from different **levels of Abstraction (for storage and enrichment)**

(Types of Combining)

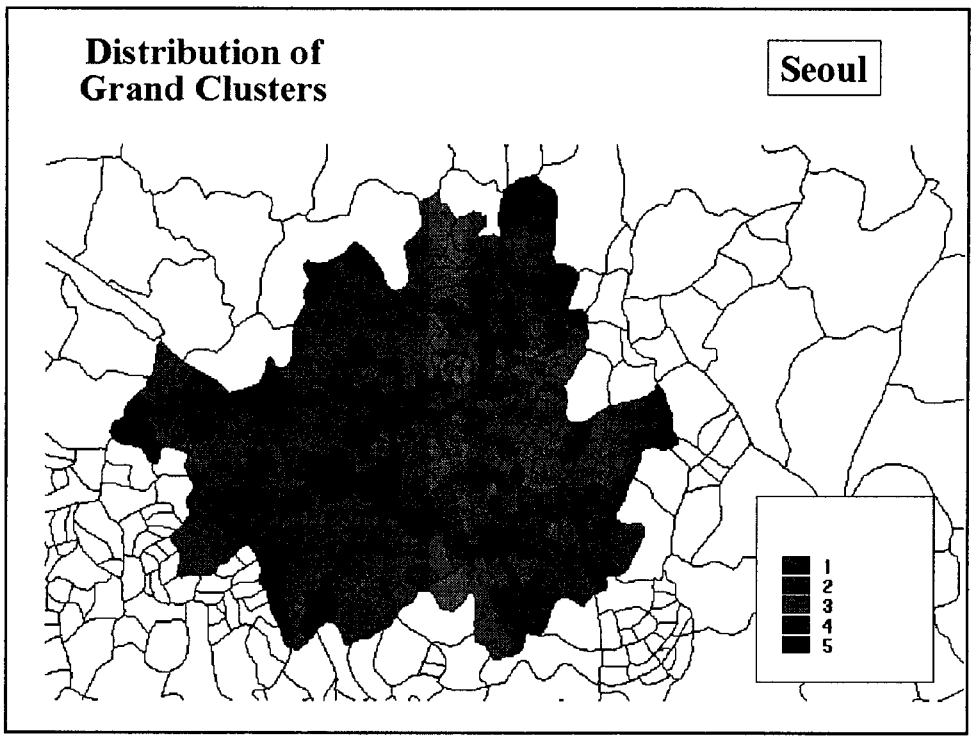
1. Record(unit) to record
2. Record to Group
   (higher level of abstraction)
3. Group to Group

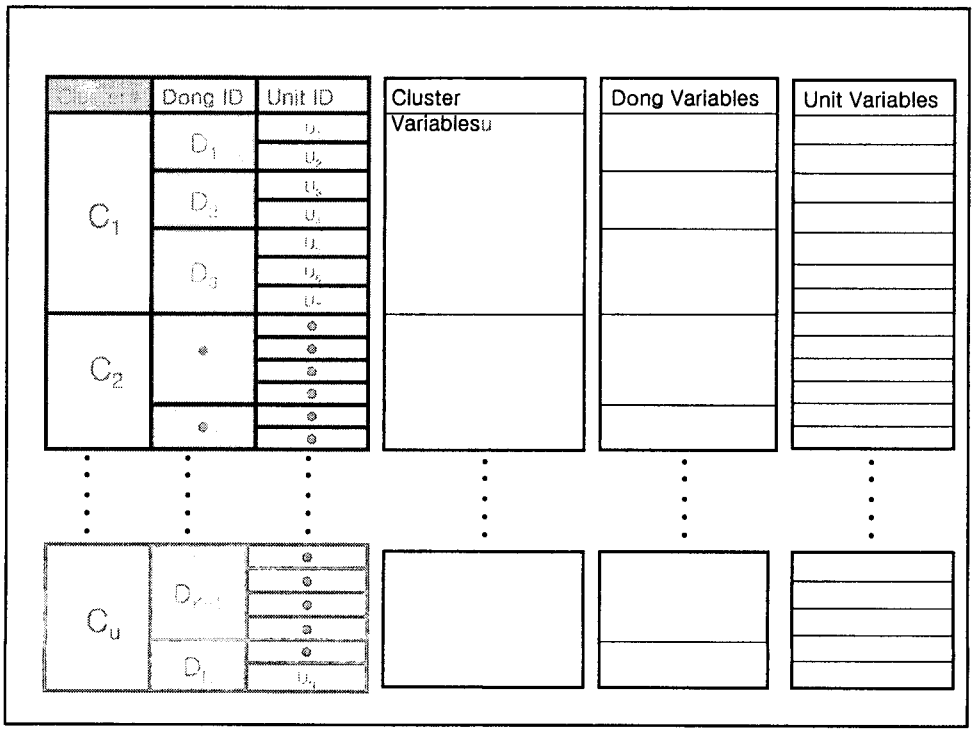
(Clustering is one possible grouping)



<figure 1>

- $A_{11} \cup A_{12}$ → ***Byproducts from business***
- $A_{11} \cup A_{21}$ → ***Census***
- A_{22} → ***No data available***


Distribution of
Grand Clusters
Seoul
1
2
3
4
5

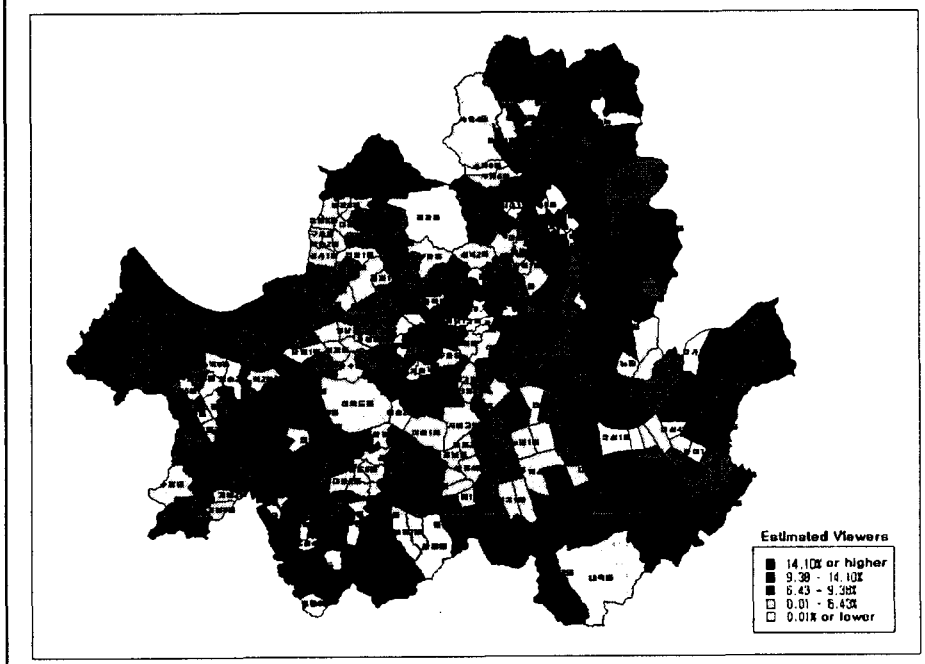

Dong ID
Unit ID
Cluster
Variablesu
Dong Variables
Unit Variables

Cluster #

Dong ID

Unit ID

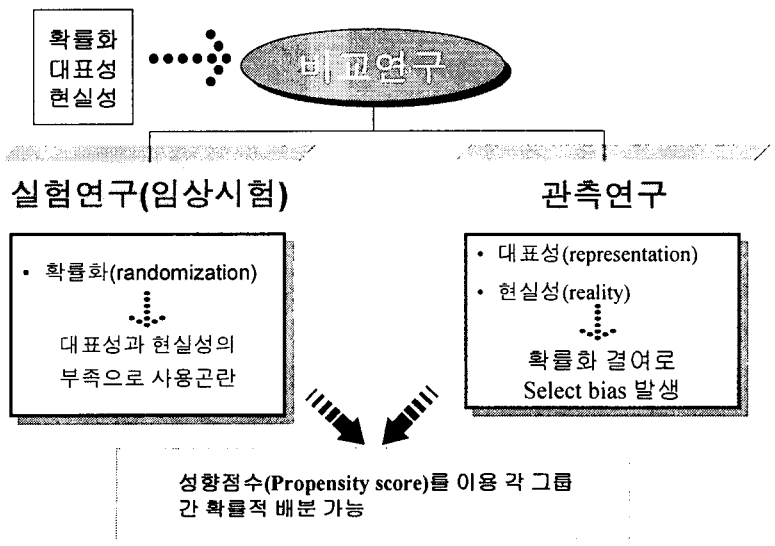
clu_id	dong_id	unit_id	inc_clu	age_clu	pop_clu	inc_dong	age_dong	status	class	pay_week	age
1	1101051	2	18362	48.3	785,493	12343	54.2	Don't Use	Professional	Weekly pay	Middle (25-35
1	1101051	3	18362	48.3	785,493	12343	54.2	Don't Use	Skilled Manual	Weekly pay	Young (< 25)
1	1101051	4	18362	48.3	785,493	12343	54.2	Use	Professional	Monthly salary	Middle (25-35
1	1101052	5	18362	48.3	785,493	4636	44.6	Use	Clerical	Monthly salary	Young (< 25)
1	1101052	6	18362	48.3	785,493	4636	44.6	Use	Management	Monthly salary	Young (< 25)
2	1101057	7	17635	42.7	12,034,475	23287	48.3	Use	Professional	Monthly salary	Old (> 35)
2	1101057	8	17635	42.7	12,034,475	23287	48.3	Don't Use	Professional	Monthly salary	Young (< 25)
2	1101058	9	17635	42.7	12,034,475	12438	38.1	Don't Use	Professional	Weekly pay	Young (< 25)
2	1101058	10	17635	42.7	12,034,475	12438	38.1	Don't Use	Clerical	Weekly pay	Young (< 25)
2	1101058	11	17635	42.7	12,034,475	12438	38.1	Don't Use	Unskilled	Weekly pay	Young (< 25)
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
26	1104844	46342	6598	58.1	380,948	9026	59.4	Don't Use	Skilled Manual	Weekly pay	Young (< 25)
26	1104844	46343	6598	58.1	380,948	14180	59.4	Don't Use	Professional	Monthly salary	Young (< 25)
26	1104844	46344	6598	58.1	380,948	11691	59.4	Don't Use	Clerical	Weekly pay	Young (< 25)
26	1104844	46345	6598	58.1	380,948	8266	59.4	Don't Use	Unskilled	Weekly pay	Young (< 25)



3. 또 다른 형태의 결합 (Data Matching)을 통해 paired experiment 가 현실적으로 불가능할때 Control Group 의 data 를 인위적으로 만들어 비교할 수 없을까?

사회적으로 사용 가능한 거대한 Data Base 를 사용하여 matching pair 를 찾아 볼 수 있다.
--건강보험자료, 국민연금자료 등.

관측연구에서의 성향점수(Propensity Score)



성향점수란?

성향점수

$$p(x_i) = \Pr(Z_i = 1 \mid X_i = x_i)$$

공변량의 균형달성

가정

가정**1.** $(Y_0, Y_1) \perp Z \mid X$ **[조건부 독립성의 가정]**

가정**2.** $0 < \Pr(Z = 1 \mid X) < 1$ **[공통영역의 가정]**

In 1983, Rosenbaum and Rubin

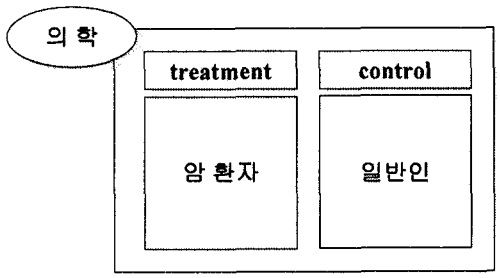
"Logistic Regression"

짝짓기(Matching)

하위집단 짝짓기 (Subclass Matching)-

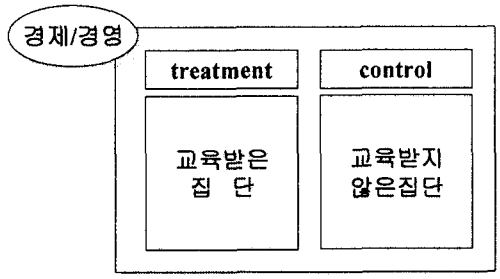
반경짝짓기 (Caliper Matching)	일대일 짝짓기 (One-to-one Matching)	층화 짝짓기 (Stratified Matching)
•성향점수를 이용 $\lvert p(x)_1 - p(x)_0 \rvert < \varepsilon$ 계산 •Random 추출 •균형의 점검	$\lvert p(x)_1 - p(x)_0 \rvert = 0$ • 균형의 점검	• 성향점수를 이용 10개 집단 층화 • Random추출 • 균형의 점검

비교연구에서의 성향점수 사용



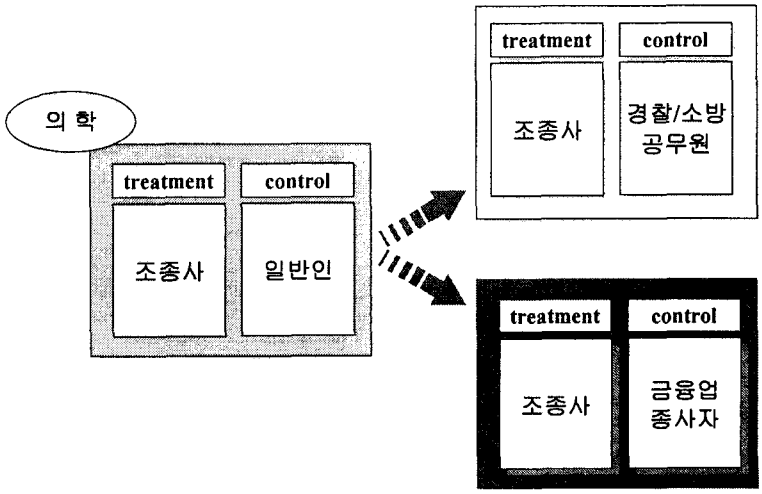
암환자와 일반인간의
특성차이(흡연등)를 비교한다.

※ 대조그룹 선정 시 본 논문 같은
대용량 데이터에 적용사례는 없음



교육을 받은 집단과 받지 않은
집단 간의 연봉차이를 비교한다.

실제자료에 적용



자료개요

Treatment Group
조종사 그룹 : 152명, 전투기 조종사(F-4,F-5,F-16)
2004년 정기 신검자료

Control Group
일반인 그룹 : 2004년 국민건강검진 자료 중
공교가입자 중 수검자 1,108,410명
직장가입자 중 수검자 3,7886,992명

공교가입자 (1,108,410명)	직장가입자 (3,7886,992명)
경찰/소방직 공무원 (98,109명)	금융업 종사자 (90,096명)

공변량 : 성별, 나이, 키, 몸무게(직업)
"비교변수 : 혈압(SBP)"

자료정리(직업군 선택)

	관측수	Age	빈도	Height (cm)	Weight (kg)	SBP (mmHg)	BBP (mmHg)	Blood Sugar (mg/dl)	Cholesterol (mg/dl)
조종사	152	40.13		173.24	72.80	120.48	79.01		
공교가입자	1,108,410	41.10		166.85	65.45	121.27	77.87	93.75	209.45
(male)	756,439	42.56		170.58	70.61	124.90	80.93	97.04	208.96
경찰/소방공무원	F	32.53	3448	162.04	55.66	112.57	71.18	85.97	176.24
	M	40.88	98,109	172.40	73.32	124.95	79.53	94.31	296.37
직장가입자	3,786,992	39.26		166.92	65.28	125.51	77.96	98.89	189.41
(male)	2,681,208	40.44		170.41	69.44	129.28	79.79	103.09	192.46
금융업	F	31.27	61,808	160.53	54.27	111.99	71.30	85.43	173.94
	M	39.33	90,096	171.37	71.43	123.49	78.56	94.11	193.50

* SBP : 수축기 혈압, BBP : 이완기 혈압, Blood Sugar : 공복시 혈당, Cholesterol : 콜레스테롤

❖ 조종사의 선택편의를 줄이기 위해 비행휴등으로 비행활동을 못하는 조종사의 혈압분포를 분석한 결과 자료의 조종사와 유사했으며 고혈압으로 비행을 중지하는 자가 매우 적어 편의를 발생시키지는 않는 것으로 보인다.

❖ 경찰/소방공무원은 그 직업의 특성(외근, 육체적 노동)및 선발 당시 조종사와 비슷한 환경으로 조종사와 직업상 유사 직종으로 간주 선택하였으며,

❖ 직장가입자는 비유사 직종으로 간주하여 선택하였다.

조종사 vs 경찰/소방공무원, 금융업

	Age	Height(cm)	Weight(kg)	SBP (mmHg)	BBP (mmHg)	Blood Sugar (mg/dl)	Cholesterol (mg/dl)
조종사	40.13	173.24	72.8	120.48	79.01		
경찰/소방직 공무원	40.88	172.4	73.32	124.95	79.53	94.31	296.37
Exact Matching	40.18	172.99	72.51	125.09	79.74	94.78	191.23
Caliper(ε=0.0000005)	39.51	173.47	73.65	124.81	79.21	89.34	199.05
Caliper(ε=0.000001)	40.84	173.15	72.3	122.77	78.68	91.93	189.68
Caliper(ε=0.00005)	40.02	172.93	71.99	125.37	79.17	90.73	193.82
Stratification	40.07	172.97	72.01	124.21	79.15	99.67	187.96

	Age	Height(cm)	Weight(kg)	SBP (mmHg)	BBP (mmHg)	Blood Sugar (mg/dl)	Cholesterol (mg/dl)
조종사	40.13	173.24	72.8	120.48	79.01		
금융업(male)	39.33	171.37	71.43	123.49	78.56	94.11	193.5
Exact Matching	40.02	173.11	72.64	122.55	77.89	94.07	201.53
Caliper(ε=0.00001)	39.64	171.82	72.41	122.76	78.02	93.07	189.24
Caliper(ε=0.00005)	39.93	173.52	73.14	121.64	77.67	94.37	204.73
Caliper(ε=0.0001)	40.03	173.27	72.43	122.37	76.55	93.76	190.67
Stratification	40.62	173.05	72.44	124.72	78.77	91.81	193.01

최종결과

	Matching Group	N	SBP	BBP	Mean difference	Std	T-TEST
							Pr >= |t|
조종사			120.48	79.01			
경찰/소방공무원	Male	152	124.95	79.53			
	Exact Matching	149	125.09	79.74	-4.804	16.477	0.0005
	Caliper(ε=0.00005)	152	125.37	79.17	-4.888	12.619	<.0001
	Stratification	152	124.21	79.15	-3.73	17.296	0.0087
금융업	Male	152	123.49	78.56			
	Exact Matching	148	122.55	77.89	-2.127	17.098	0.1241
	Caliper(ε=0.00005)	152	122.2	77.4	-1.724	16.165	0.1906
	Stratification	152	124.72	78.77	-2.908	15.984	0.0264

이러한 분석으로 몇 가지 추론이 가능하다

- 전투기 조종사의 혈압은 비슷한 신체적/직업적 환경에 놓여있는 경찰/소방 공무원그룹에 비해서 뿐 아니라 금융업종사자에 비해서도 낮음을 알 수 있었다.
- 또한 그 차이는 경찰/소방공무원과 더욱 크게 나타났다. 이는 조종사의 혈압 강하가 육체적 노동에 따른 혈압 상승과는 다소 차이가 나타나는 것을 반증하는 결과이다.
- 이는 육체적 노동에 속한 직업군인 경찰/소방공무원은 금융업종사자그룹보다 혈압이 높게 나타났으나, 그와 유사한 육체적 노동을 수반하는 전투조종사그룹은 매우 낮게 나타난 것으로 유추할 수 있다.
- 이상의 결과는 조종사들의 혈압이 비행시간(연령)에 따라 감소한다는 선행연구(고성경, 2004)와도 일치하는 것으로 나타났다. 다만, 비행시간과의 관계는 유의하지 않다. 이는 전투조종사의 표본수가 작아 그룹간 차이를 보기 어렵기 때문일 것이다.

- data 수집 방법은 1900년대 이후 계속해서 발전해옴.
- ICT 의 발전은 더 많은 data 를 생성, 축적하고 있음
- 연구를 위해 실험계획이나 조사설계로 얻은 data 가 아닌 byproduct data 를 더 좋은 data 로 만드는 기법의 개발이 요구되고 있음.

여러분 들의 귀중한 시간을 낭비하여
대단히 죄송합니다

그렇지만 다시 한번 감사 드립니다