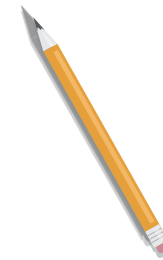


10강

역학연구의 통계적 방법

정보통계학과
이태림 교수



목차

1. 역학연구의 기본개념

역학의 정의와 내용

2. 역학연구의 설계와 측정

폐쇄집단 연구

단면적 연구

사례 대조군 연구

위험률 발생률 유병률

특이도 민감도

3. 타당성 연구

교란과 치우침 통제

짜짓기 자료 분석

학습하기

1

1. 역학연구의 기본개념



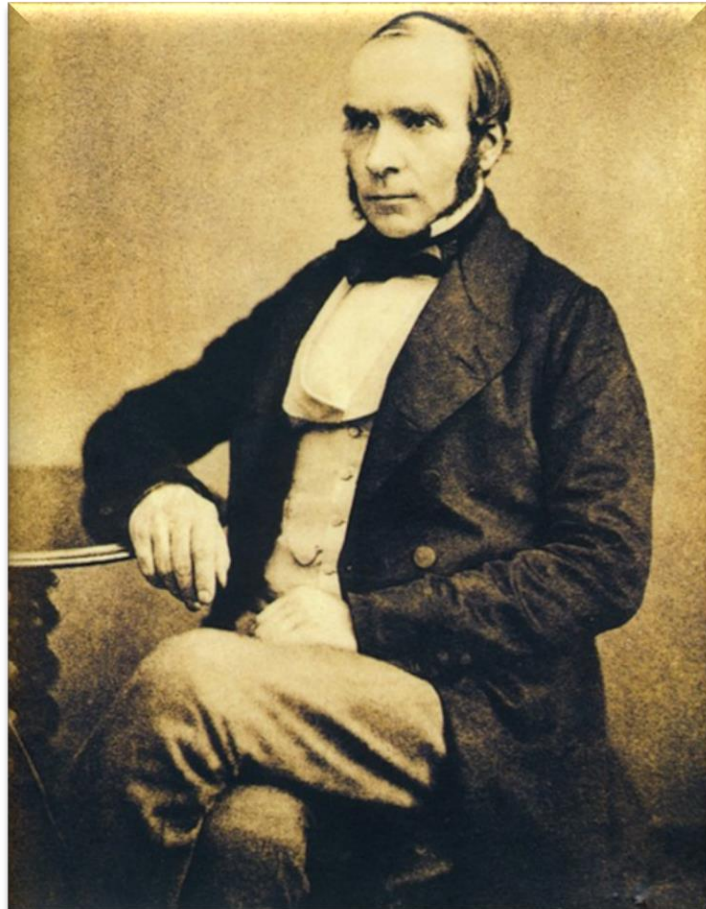
역학(Epidemiology)

인구집단에서 발생하는 질병의 분포상태를 관찰하여 원인을 찾고 질병 발생을 예측하고 효과적인 예방법을 개발하는 연구분야

- 우리나라 국민건강조사
: 1962년 보건사회부 건강조사

- 1954 British Doctors Study by Doll & Hill
“Smoking & lung Cancer”

콜레라 유행의 원인 연구



John Snow

1813년 3월 15일 ~ 1858년 6월 16일

빅토리아 여왕 시대 영국의 의사.

1854년 [런던](#) 소호에서 창궐한 [콜레라](#)가 오염된 물을 통해서 퍼졌다는 것을 연구를 통해 밝혀 내어 수많은 목숨을 구한 역학자

그가 한 방법들을 기초적인 [빅 데이터 프로세싱](#)과 첨단 기법을 동원해 발전시킨 것일 정도로 역학의 성립에 기여

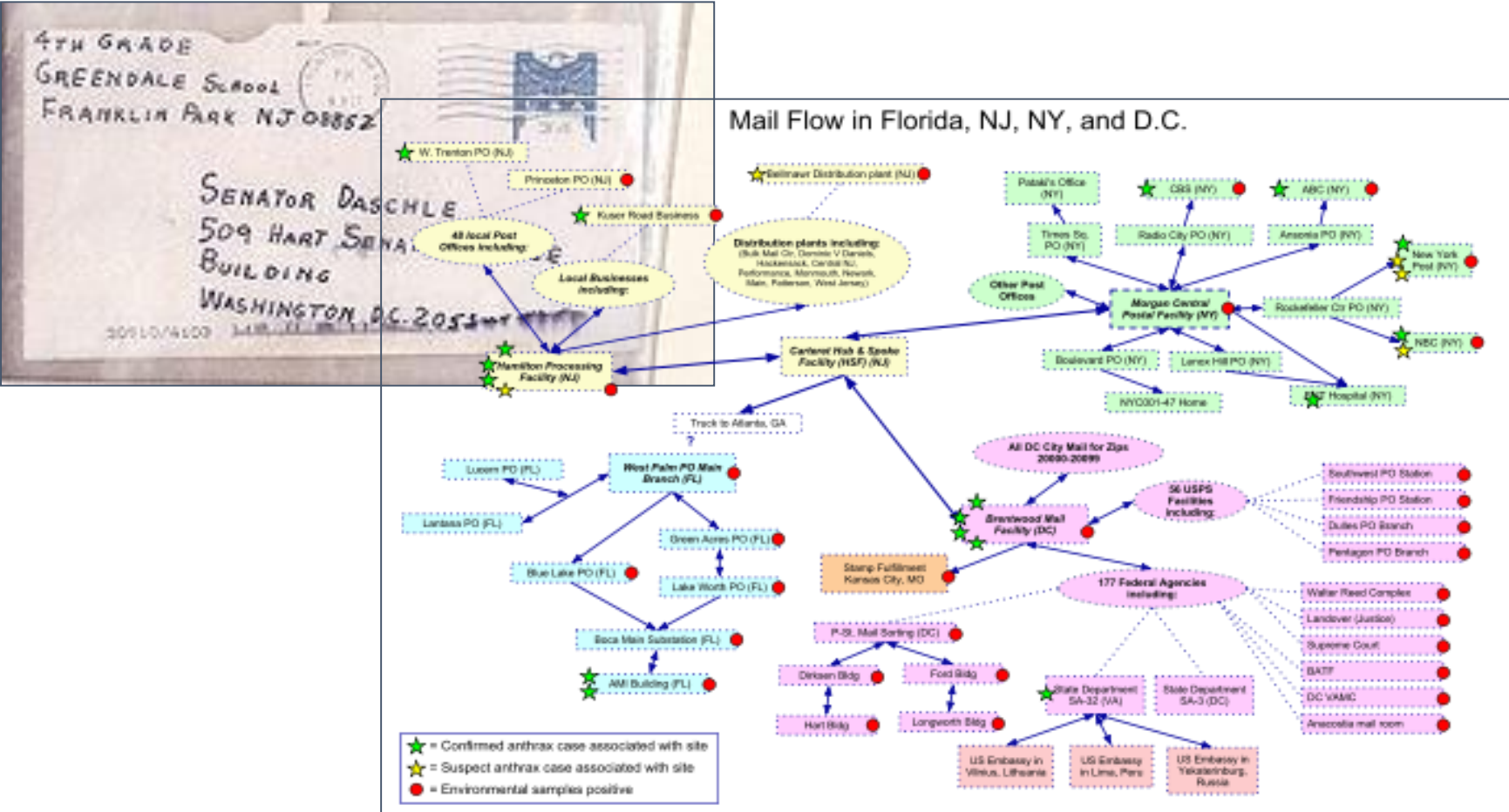


역학(Epidemiology)

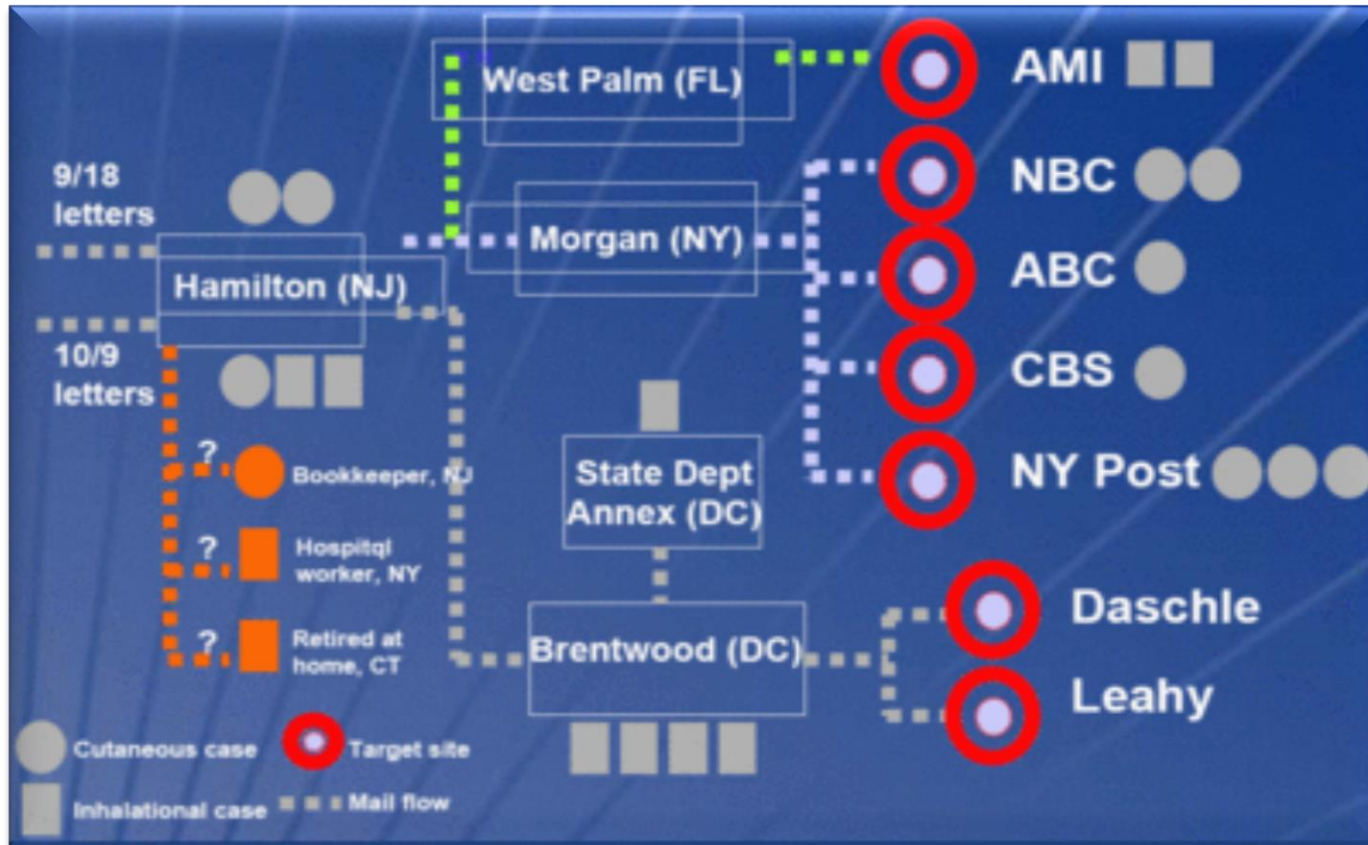




Globalization PHI: Internet & Bioterrorism



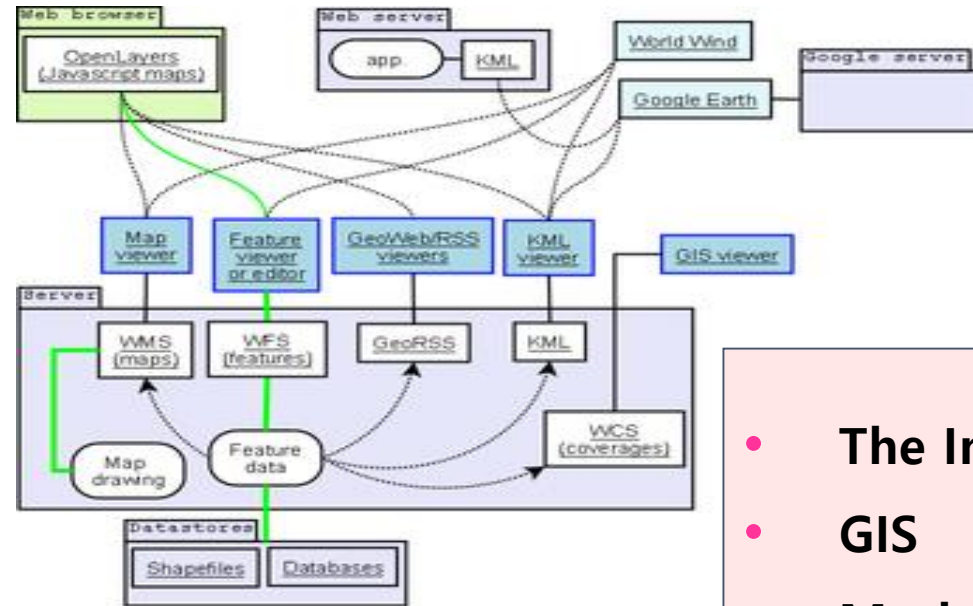
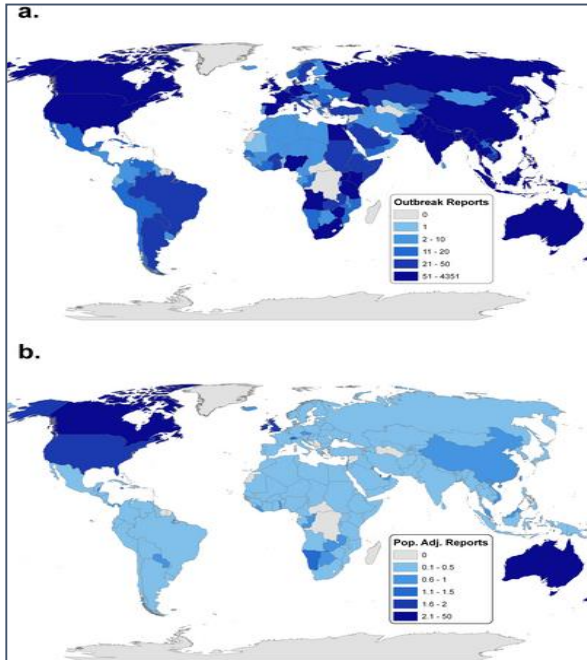
Globalization PHI: Internet & Bioterrorism



Seven letters are believed to have been mailed in the U.S., resulting in 22 infections; five people died.

HealthMap

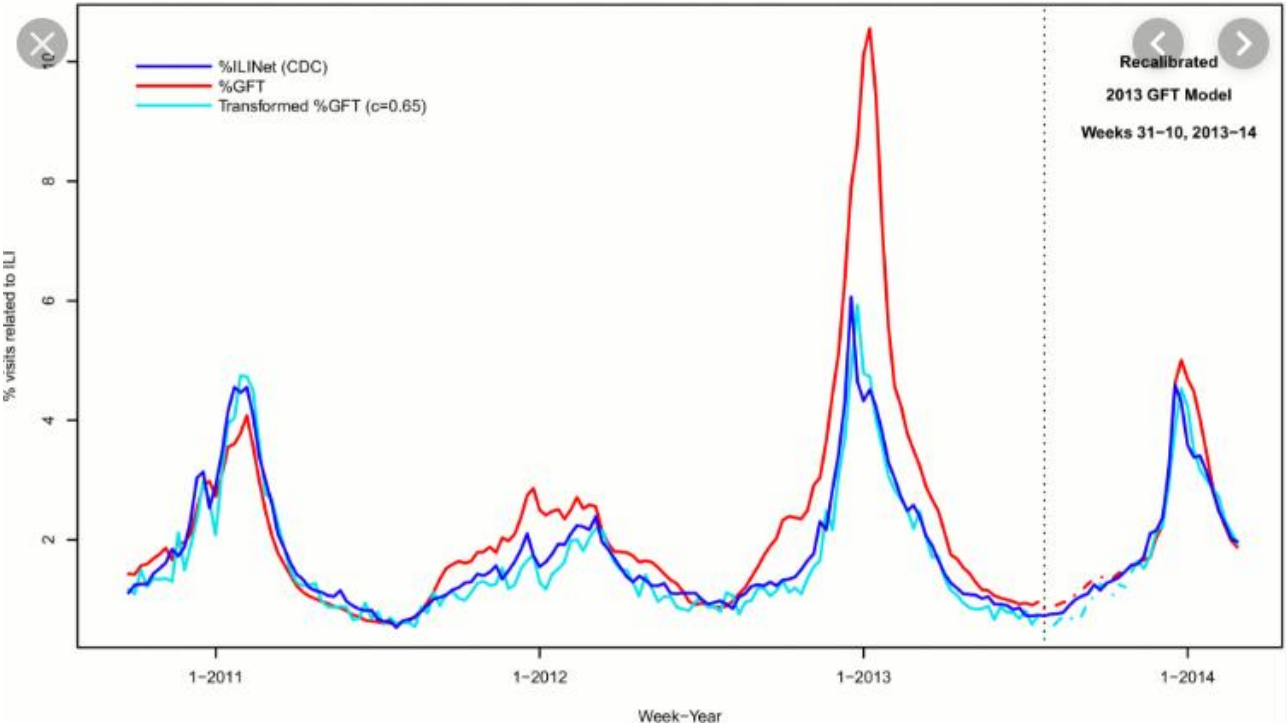
1. 역학연구의 기본개념



- The Internet
- GIS
- Mashup

Geographic coverage of English language news reports of **infectious disease outbreaks** collected through Google News from October 1, 2006 through July 18, 2007. (A) displays **counts of disease outbreak reports**; (B) displays **population-adjusted outbreak reporting** as number of reports per million inhabitants.

Google Flue Trend



질병분포의 측정

1. 역학연구의 기본개념

- 위험률 - 질병이 발생할 확률

- 발생률(incidence rate)

- 특정기간 동안 발견된 질병 환자수를 전체 집단의 수로 나눈 것

- 유병률(prevalence rate)

- 조사당시 질병환자수를 조사대상자수로 나눈 것

- ① 시점유병률(point prevalence)

- 특정시점에 앓고 있는 사람을 대상으로 한 유병률

- ② 기간유병률 (period prevalence)

- 일정기간 동안 관찰하여 얻는 유병률

- * 폐쇄집단(cohort)

- 일정기간 관찰 대상이 되는 인구집단으로 같은 특성을 가진 집단을 의미

타당성(validity)

어떤 측정치나 측정방법이 측정하고자 목적하는 것을 성취하는 정도

예) 결핵환자 발견을 위한 X-선 촬영 방법,
미생물 배양법

	유	무	계
양성	a	b	a+b
음성	c	d	c+d

타당성(validity)

	유	무	계
양성	a	b	a+b
음성	c	d	c+d

① 민감도(sensitivity) = $\frac{\text{검사 양성자수}}{\text{총 환자수}} = \frac{a}{a+c}$

② 특이도(specificity) = $\frac{\text{검사 음성자수}}{\text{총 비환자수}} = \frac{d}{b+d}$

③ 예측도(predictability)

양성 예측도 = $\frac{\text{검사 음성자수}}{\text{총 검사 양성자수}}$

음성 예측도 = $\frac{\text{검사 음성자수}}{\text{총 검사음성자수}}$

예) 당뇨병 환자의 경우 혈당치 측정
결핵 환자 - BCG 접종
자궁암 환자 - 세포 검사

1. 역학연구의 기본개념

치우침(Bias)

- 선택성 편견(selection bias)
: 연구 대상자를 선정할 때 발생
- 정보성 편견(information bias)
: 부정확한 측정이나 부정확한 정보수집으로 발생
- 외연인자에 의한 편견(confounding bias)
: 연구대상 요인효과가 외적변수효과
[extraneous factor]와 혼합되게 발생

학습하기

1

2

2. 역학연구의 설계와 측정



폐쇄집단 연구(cohort study)

- * 폐쇄집단(cohort)
 - 일정기간 관찰 대상이 되는 인구집단으로 같은 특성을 가진 집단을 의미

- $\frac{a}{a+b}$ 와 $\frac{c}{c+d}$ 를 비교한다.
- 질병원인에 대한 가설을 검증하는 역학적 연구에 있어 매우 중요한 방법 중 하나이다.
- 단점 : 장기간 추적하는 동안 누락이 많이 발생하고 장기간 관찰을 위해 표본의 규모가 커야한다.

폐쇄집단 연구(cohort study)

- 희귀하게 발생하는 질병이나 잠복기간이 긴 질병에 유용하다.
- 적합한 대조군의 선정이 어렵고 confounding 제어가 쉽지 않다.

질병발생의 측정

- ① 비율
 - 위험률 비(risk odds)
- ② 평균율 (average rate)의 추정
- ③ 위험률(risk rate)의 추정

사례-대조군 연구(case-control study)

2. 역학연구의 설계와 측정

- $\frac{a}{a+b}$ 와 $\frac{c}{c+d}$ 의 차이가 유의한가 검토
- 교차비(odds ratio)로 상대위험도를 추정하여
요인과 질병 발생과의 연관성을 규명한다.

$$OR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{ad}{bc}$$

- 지역사회 표본인구에 대한 조사이므로 대상
질환의 유병률을 얻을 수 있고 차후 cohort
연구의 대상 집단으로 활용하면 유용하다.

- 짧은 시간내 수행이 가능하고 표본수가 적어도
되므로 시간, 경비노력이 절약된다.



학습하기

2

3

3. 질병발생의 측정



질병발생의 측정

3. 질병발생의 측정

- ① 비율
 - 위험률 비(risk odds)
- ② 평균율 (average rate)의 추정
- ③ 위험률(risk rate)의 추정

- 위험비
 - 의심되는 요인에 폭로된 집단에서의 특정질환 발생 비율과 이 요인에 폭로되지 않은 집단에서의 이 질환 발생 비율과의 대비

- 폭로군과 비폭로군의 특정질환 발생 비율이 동일할때는 1에 가까워지면 이들간 발생비율의 차이가 크면 클수록 통계적 연관성이 강하다.

질병빈도 및 사망률 측정

- ① 상병빈도의 측정 - 현재 특정 질환을 앓고 있는
사람수를 추정
- ② 사망정도의 측정
 - 사망률(mortality rate)
: 인구집단에서의 특정 질환에 의한 사망률
 - 치명률(fatality rate)
: 특정 질환을 가진 사람 중에서의 사망률
 - 사망률(death rate)
: 전체 집단에서의 모든 질병에 의한 사망률

질병빈도 및 사망률 측정

3. 질병발생의 측정

③ 인과성 정도의 측정

– 상대위험도(relative risk)

$$= \frac{\text{위험 노출집단의 특정 질병 발생률}}{\text{위험에 노출되지 않은 집단의 특정 질병 발생률}}$$

= 오즈비 (ods ratio)

$$OR = \frac{a/b}{c/d} = \frac{ad}{bc}$$

– 귀속위험도 (attributable risk)

$$AR = \frac{b(r-1)}{b(r-1) + 1} \times 100$$

[단, b는 전체인구중에서
위험에 노출된 비율, r은 상대위험도]

연령정정사망률

(1) 직접법 (Direct Method)

필요자료 : 해당인구의 연령별 사망률, 표준인구와 그 연령구성

- 연령정정사망률

= $\frac{\text{지역A의 연령별 사망률이 표준인구에 적용될때의 총사망자수}}{\text{표준인구의 수}}$

$$= \frac{\sum_x P_S(x) \times M_A(x)}{\sum_x P_S(x)}$$

- A지역 연령정정사망률

$$\begin{aligned} &= \frac{\sum_x P_S(x) \times M_A(x)}{\sum_x P_S(x)} \times 1000 = \\ &= \frac{(15000 \times 0.005) + (30000 \times 0.01) + (15000 \times 0.015)}{15000 + 30000 + 15000} \times 1000 \\ &= 10 \end{aligned}$$

연령정정사망률

(1) 직접법 (Direct Method)

- 필요자료 : 해당인구의 연령별 사망률, 표준인구와 그 연령구성

- **연령정정사망률**

= $\frac{\text{지역A의 연령별 사망률이 표준인구에 적용될때의 총사망자수}}{\text{표준인구의 수}}$

$$= \frac{\sum_x P_S [x] \times M_A [x]}{\sum_x P_S [x]}$$

- B지역 연령정정사망률

$$= \frac{(15000 \times 0.005) + (30000 \times 0.01) + (15000 \times 0.015)}{15000 + 30000 + 15000} \times 1000$$

$$= 10$$

[2] 간접법 Indirect Method]

- 해당 지역인구의 보통사망률과 연령별 인구통계는 있으나 연령별 사망 통계 자료는 없고 표준인구의 사망률과 연령별 사망률 자료가 있을때

D_A : A 지역 인구의 보통사망률

$P_A(x)$: A지역 인구의 연령별 사망률

D_S : 표준인구의 보통사망률

$M_S(x)$: 표준인구의 연령별 사망률

$$D_A' = \frac{\text{표준인구의 연령별 사망률이 A지역 연령별인구에 적용될때의 총 사망자수}}{\text{A지역 총 인구}}$$

$$= \frac{\sum_x M_S(x) P_A(x)}{\sum_x P_A(x)} \times 1000$$

여기서 D_A' 는 표준인구의 사망률이 A지역 인구에 적용되었다고 가정하였을때 기대되는 이론적인 보통사망률

연령정정사망률

[2] 간접법 Indirect Method]

- 해당 지역인구의 보통사망률과 연령별 인구통계는 있으나 연령별 사망 통계 자료는 없고 표준인구의 사망률과 연령별 사망률 자료가 있을때

D_A : A 지역 인구의 보통사망률

$P_A(x)$: A지역 인구의 연령별 사망률

D_S : 표준인구의 보통사망률

$M_S(x)$: 표준인구의 연령별 사망률

$$\begin{aligned} D_{A'} &= \frac{\text{표준인구의 연령별 사망률이 A지역 연령별인구에 적용될때의 총 사망자수}}{\text{A지역 총 인구}} \\ &= \frac{\sum_x M_S(x) P_A(x)}{\sum P_A(x)} \times 1000 \end{aligned}$$

[2] 간접법 Indirect Method]

- 해당 지역인구의 보통사망률과 연령별 인구통계는 있으나 연령별 사망 통계 자료는 없고 표준인구의 사망률과 연령별 사망률 자료가 있을때

D_A : A 지역 인구의 보통사망률

$P_A(x)$: A 지역 인구의 연령별 사망률

D_S : 표준인구의 보통사망률

$M_S(x)$: 표준인구의 연령별 사망률

표준인구의 연령별 사망률이 A 지역 연령별인구에
적용될때의 총 사망자수

$$D_{A'} = \frac{\text{표준인구의 연령별 사망률이 A 지역 연령별인구에
적용될때의 총 사망자수}}{\text{A 지역 총 인구}}$$

=

- 정정지수 : $\frac{D_S}{D_{A'}} = C$

D_S : 표준인구의 연령별 인구에 이 사망률이
적용되었을때의 값

$D_{A'}$: 지역 A 인구의 연령별 인구에 그 사망률이
적용되었을때의 값

- A 지역 연령정정사망률 = $D_{A'} \times C$

연령정정사망률

[2] 간접법 (Indirect Method)

P.276 <표6.12>

연령	연령별 사망률
< 21	1.0
21~60	2.0
60 <	3.0

<표6.13>

연령	인구
< 21	10000
21~60	30000
60 <	10000

1. D_A'

$$= \frac{(10000 \times 0.0001) + (30000 \times 0.0002) + (10000 \times 0.0003)}{10000 + 30000 + 10000} \times 1000 = 0.2$$

2. $C = \frac{D_S}{D_A'} = \frac{0.22}{0.2} = 1.1$

3. $D_A \times C = 0.35 \times 1.1 = 0.385$

학습하기

3

4

4. 타당성 연구



주요내용

교란과 치우침



자료수집 이전의 통제기법



자료수집 이후의 통제기법



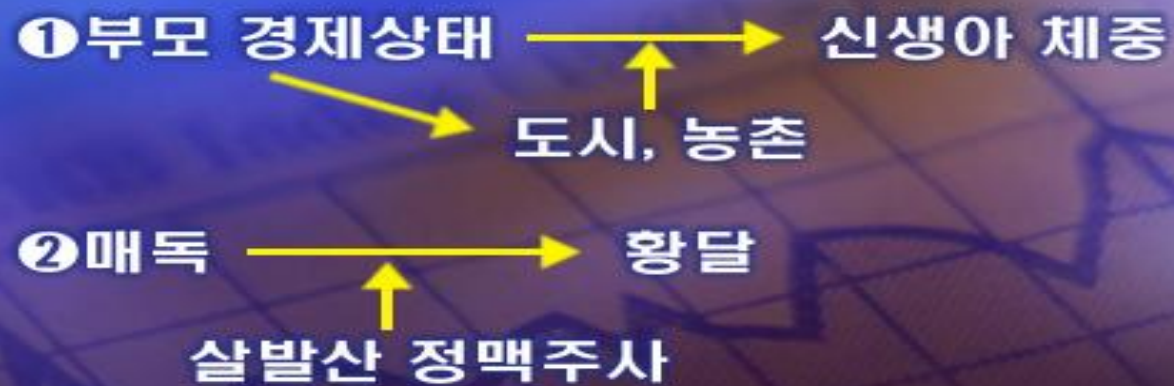
짜짓기 자료의 분석방법



교란에 의한 편견(confounding bias)

- 교란인자(confounding factor)
 - 제3의 변수로 원인과 결과 간의 관계를 체계적으로 교란시키는 인자

- 예



교란인자 통제방법

연구설계시의 통제방법

- 확률적 처리 배치
- 짝짓기(matching) – 교란인자수를 토대로 하는 방법
 - 개체수를 토대로 하는 방법

결과분석 단계에서의 통제

- 표준화법
- 총화법
- 다변량기법

교란인자를 토대로 한 통제방법

- 단일변수 짝짓기의 가정
 - ① 폐쇄집단 연구 - 결과, 사례대조군 연구 - 원인
⇒ 측정결과가 이항적으로 나타난다
 - ② 두 집단간의 차이는 교란인자와의 상호작용 없이 측정된다
 - ③ 두 집단의 표본수는 동일하게 한다



교란인자를 토대로 한 통제방법

- 단일변수 짝짓기의 가정

- ① 폐쇄집단 연구 - 결과, 사례대조군 연구 - 원인
⇒ 측정결과가 이항적으로 나타난다
- ② 두 집단간의 차이는 교란인자와의 상호작용 없이 측정된다
- ③ 두 집단의 표본수는 동일하게 한다
- ④ 조사집단을 구성하는 개체는 이미 선정되어서 고정되어 있다

교란인자를 토대로 한 통제방법



- 반경짜짓기
- 최근접짜짓기
- 총화짜짓기
- 빈도짜짓기
- 평균짜짓기

반경 짝짓기(caliper matching)

- 교란인자의 수치가 일정한 거리내에 위치하는 개체들을 사례군과 대조군에서 각각 뽑아 1:1로 짝짓는 방법

반경 짝짓기(caliper matching)

- 교란인자의 수치가 일정한 거리내에 위치하는 개체들을 사례군과 대조군에서 각각 뽑아 1:1로 짝짓는 방법
- 예) 흡연이 혈압에 미치는 영향조사
- 교란인자 : 연령

흡연군	비흡연군
37	36, 38
38	38, 40
40	40, 42
45	43
50	49



흡연군	비흡연군
37	36
38	38
40	40
45	43
50	49

최근접짝짓기(nearest available matching)

사례군과 대조군의 개체를 교란인자 수치에 따라 일정하게 배열하고 사례군 개체의 일정한 순서에 따라 비교군으로부터 가장 가까운 교란인자 수치를 갖는 개체선택

흡연군	비흡연군
40	30
50	32
45	46
	49
	55



흡연군	비흡연군
40	46
50	49
45	55

층화짝짓기(stratified matching)

- 교란인자가 범주형일때 일정 형식으로 층화하여 두 집단을 분류한 후 동일한 교란인자의 층내에서 짝을 이루는 방법

층화짝짓기(stratified matching)

- 교란인자가 범주형일때 일정 형식으로 층화하여 두 집단을 분류한 후 동일한 교란인자의 층내에서 짝을 이루는 방법
- 단점
 - 대조군의 도수가 사례군보다 작으면 도수에 대한 짝을 이룰수 없다
 - 교란인자가 연속변수일때 구분해야 한다

평균짜짓기 (mean matching)

교란인자의 평균치를 두 집단에 동일하게 하는 방법.
사례군의 평균연령을 구하여 대조군 중에서 평균연령과
같거나 비슷한 사람을 우선 뽑는다.

평균짜짓기 (mean matching)

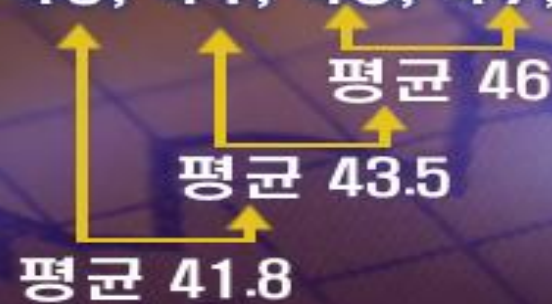
교란인자의 평균치를 두 집단에 동일하게 하는 방법.
사례군의 평균연령을 구하여 대조군 중에서 평균연령과
같거나 비슷한 사람을 우선 뽑는다.

- 예] 흡연자 40, 45, 50 → 평균연령 : 45세
비흡연자 32, 35, 40, 41, 45, 47, 55

평균 짝짓기 (mean matching)

- 교란인자의 평균치를 두 집단에 동일하게 하는 방법. 사례군의 평균연령을 구하여 대조군 중에서 평균연령과 같거나 비슷한 사람을 우선 뽑는다.

- 예) 흡연자 40, 45, 50 → 평균연령 : 45세
비흡연자 32, 35, 40, 41, 45, 47, 55



개체수를 토대로 하는 방법

- 사례군에서 하나를 뽑고 대조군에서 하나보다 더 많은 개수의 대조군을 1:R 로 짝짓기한다.



개체수를 토대로 하는 방법

- 사례군에서 하나를 뽑고 대조군에서 하나보다 더 많은 개수의 대조군을 1:R 로 짝짓기한다.

대조군의 수에 따른 효율

R	효율
1	1.000
2	1.333
3	1.500
4	1.600
5	1.667
6	1.714
+ ∞	2.000



적정한 R은 효율성과 함께 경제성도 고려해야 한다.

자료수집 이후의 통제기법

- 표준화법
- 총화법
- 다변량법
- 로그선형 분석법
- 로지스틱 회귀법
- 중회귀분석법
- 공분산분석

짜짓기 자료의 분석방법

단일짜 분석

4. 타당성 연구

짜짓기 자료의 분석방법

단일짜 분석

- 이분적 결과(dichotomous outcome)
- 다항결과(polychotomous outcome)

짜짓기 자료의 분석방법

단일짜 분석

- 이분적 결과(dichotomous outcome)
- 다항결과(polychotomous outcome)
- 다항폭로수준(polychotomous exposure levels)

2) 복수짜 분석 ; 대조군으로 M개를 이용하여 짝을 지은 경우

짜짓기 자료의 분석방법

단일짜 분석

- 이분적 결과(dichotomous outcome)
- 다항결과(polychotomous outcome)
- 다항폭로수준(polychotomous exposure levels)

① 이분적 결과 <표 8.6>

사례군	대조군		
	요인 존재	비존재	계
요인 존재	a	b	a + b
비존재	c	d	c + d
계	a + c	b + d	n

Q & A



다음시간에는

▶ 10강 역학연구의 통계적 방법

11강 임상시험1

12강 임상시험2