

소개

- 강의형태 (2023학년도 1학기)
 - 대면강의
 - 녹화강의 제공 없음
 - 보강: 1+회 (5/22(월)) - 사전녹화 영상강의
 - 시험: 2회 (4/24(월), 6/7(수))
- 평가
 - 시험 (80%):
 - 중간시험 (4/24(월) 18:00-20:30) **원래 두달새는 두달!**
 - 기말시험 (6/07(수) 18:00-20:30)
 - 출석 (5%):
 - 2회(1주)를 초과한 결석은 감점 대상
 - 과제 (15%):
 - 3회 제출

1

소개

- 통계패키지 사용
 - SAS 설치 (3월말 이전)
 - 통계학과 사무실 (사회교육관 512호, 전화 02-710-9437)
 - 전산통계 미수강자를 위한 최소한의 사용법은 강의내용에 포함
- 공지사항
 - 강의계획서.pdf
 - 회귀분석정오표_2판1쇄.pdf
 - 회귀분석정오표_2판2쇄.pdf
 - 제1강의 강의자료 (강의자료는 수업전 공지사항 확인)

2

Introduction

■ 통계학 (Statistics)

주어진 문제에 대하여 합리적인 답을 줄 수 있도록
숫자로 표시되는 data를 수집·정리하여 해석하고
신뢰성있는 결론을 이끌어내는 방법을 연구하는 학문

3

Introduction (chatDPT)

ChatGPT

☀ Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →

⚡ Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests

⚠ Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

4

Introduction (Statistics)



What is statistics?



Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. It involves using mathematical and statistical methods and techniques to gather, analyze and interpret data, and draw meaningful conclusions from it.



Statistics is a crucial field of study as it helps us understand the world around us and make informed decisions based on evidence. It has a wide range of applications in various fields, including science, engineering, medicine, business, social sciences, and more.

There are two main branches of statistics: descriptive statistics and inferential statistics. Descriptive statistics involve methods for summarizing and visualizing data, such as computing measures of central tendency (e.g., mean, median, mode) and dispersion (e.g., range, standard deviation, variance), creating graphs and charts, and identifying patterns in data. Inferential statistics, on the other hand, involves making inferences or predictions about a larger population based on a smaller sample of data.

5

Introduction (Statistics)



통계학은 데이터의 수집, 분석, 해석, 제시 및 조직화에 대한 연구입니다. 수학적 및 통계적 방법과 기술을 사용하여 데이터를 수집, 분석 및 해석하고, 이로부터 의미 있는 결론을 도출하는 것을 포함합니다.



통계학은 우리 주변 세계를 이해하고, 증거에 기반하여 합리적인 결정을 내리는 데 매우 중요한 분야입니다. 이 분야는 과학, 공학, 의학, 비즈니스, 사회과학 등 다양한 분야에 널리 응용됩니다.

통계학에는 기술 통계학과 추론 통계학이라는 두 가지 주요 분야가 있습니다. 기술 통계학은 데이터의 요약과 시각화 방법을 다루며, 중심 경향성 측정 (예: 평균, 중앙값, 최빈값) 및 분산 측정 (예: 범위, 표준 편차, 분산) 계산, 그래프 및 차트 작성 및 데이터 내 패턴 식별 등의 방법을 포함합니다. 반면, 추론 통계학은 작은 샘플 데이터를 기반으로 큰 모집단에 대한 추론이나 예측을 수행합니다.

6

Introduction

예. 신체검사 자료

대상에 대한 특성
변수

관측값	번호	성별	연령	신장	몸무게	비만도	혈액형	입원건수
	1	남	28	174	64	정상	O	0
	2	남	29	172	73	과체중	A	0
	3	남	26	164	53	정상	A	1
	4	남	28	181	74	정상	B	0
	5	남	25	159	47	정상	AB	0
	6	여	26	167	55	정상	O	0
	7	남	27	174	78	비만	O	2
	8	남	25	175	56	저체중	B	1
	:	:	:	:	:	:	:	:

- 일변량자료 (univariate data)

↙ 다변량자료 (multivariate data)

↳ 회귀분석

7

Introduction

■ 자료의 형태

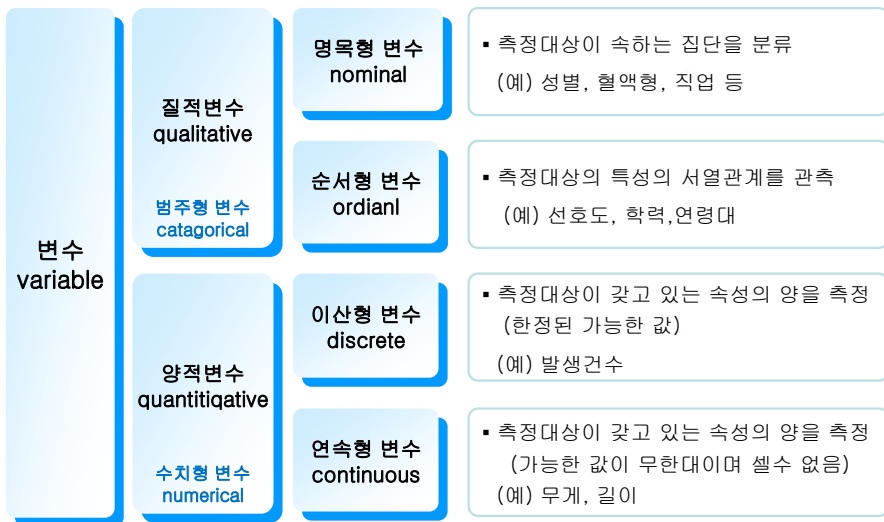


■ 양적자료는 구간형(interval)과 비율형(ratio) 자료로 분류 되기도 한다.

8

Introduction

■ 변수의 분류



9

Introduction

예. 신체검사 자료

변 수								
번호	성별	연령	신장	몸무게	비만도	혈액형	입원건수	
1	남	28	174	64	정상	O	0	
2	남	29	172	73	과체중	A	0	
3	남	26	164	53	정상	A	1	
4	남	28	181	74	정상	B	0	
5	남	25	159	47	정상	AB	0	
6	여	26	167	55	정상	O	0	
7	남	27	174	78	비만	O	2	
8	남	25	175	56	저체중	B	1	
:	:	:	:	:	:	:	:	

- 범주형 변수: 성별, 비만도, 혈액형

- 수치형 변수: 연령, 신장, 체중, 입원건수

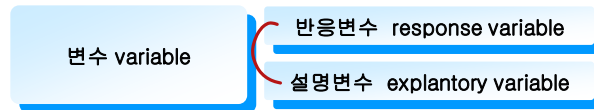
10

Introduction

■ 변수의 종류에 따른 분석방법

다변량자료의 분석

- 변수들간의 관계를 분석



- 반응변수는 종속변수 (dependent variable)
- 설명변수는 독립변수 (independent variable)

예) 연령에 따른 신장 : 반응변수(신장), 설명변수(연령)

설명변수	반응변수	분석방법
수치형 변수	수치형 변수	회귀분석
범주형 변수	수치형 변수	분산분석
범주형 변수	범주형 변수	범주형자료분석

11

Introduction

■ 회귀분석 (regression analysis)

- 변수들의 관계를 기술하고 형태를 파악하는 통계적방법
 - 설명변수와 반응변수의 관계

12

회귀 (回歸, regression)

NAVER

회귀 回歸 [회귀/회귀]

명사 한 바퀴 돌아 제자리로 돌아오거나 돌아감.

유의어 귀환 복고¹ 복귀

표준국어대사전

회귀 回歸 [회귀/회귀]

명사 수학 하나의 종속 변수와 두 개 이상의 독립 변수 사이에 나타나는 관계를 최소 제곱법으로 추정하는 방법.

우리말샘

회귀하다 回歸하다 [회귀하다/회귀하다]

동사 한 바퀴 돌아 제자리로 돌아오거나 돌아가다.

유의어 귀환하다 돌아오다 복귀하다

표준국어대사전

13

회귀 (回歸, regression)

DdM

회귀 回歸

본래의 자리로 돌아오거나 돌아감

회귀하다 回歸--

다시 돌아오거나 되돌아가다

회귀 분석 回歸分析

어떤 변수가 다른 변수에 의해 설명된다고 보고 그 함수 관계를 조사하는 통계적 해석 수법

회귀 본능 回歸本能

하천에서 태어난 물고기가 바다로 나가 자란 뒤에 알을 낳기 위하여 다시 자기가 태어난 하천으로 되돌아오는 것

14

Introduction (chatDPT)

ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

15

Introduction (Regression Analysis)

MW

What is regression analysis?



Regression analysis is a statistical method used to determine the relationship between a dependent variable and one or more independent variables. In other words, it helps to understand how changes in the independent variables affect the dependent variable. The goal of regression analysis is to find a mathematical equation that can predict the value of the dependent variable based on the values of the independent variables.

There are different types of regression analysis, including simple linear regression, multiple linear regression, logistic regression, and nonlinear regression. Simple linear regression is used when there is only one independent variable, while multiple linear regression is used when there are two or more independent variables. Logistic regression is used when the dependent variable is categorical, while nonlinear regression is used when there is a non-linear relationship between the independent and dependent variables.

Regression analysis is used in many fields, including economics, finance, psychology, and biology, to name a few. It is used to make predictions, identify trends, and explain the relationship between variables.

16

Introduction (Regression Analysis)



회귀 분석(Regression analysis)은 종속 변수와 하나 이상의 독립 변수 간의 관계를 분석하는 통계 기법입니다. 회귀 분석의 주요 목적은 이러한 변수들 간의 관계를 모델링하고, 이 모델을 사용하여 독립 변수의 값을 알고 있을 때 종속 변수의 값을 예측하는 것입니다.

회귀 분석은 경제학, 금융, 사회과학, 의학, 공학 등 여러 분야에서 활용됩니다. 예를 들어, 금융에서는 회귀 분석을 사용하여 두 종목의 수익률 간의 관계를 파악할 수 있습니다. 사회과학에서는 소득과 교육 수준 사이의 관계를 파악할 수 있습니다. 의학에서는 특정 위험 요소와 질병 발생 사이의 관계를 조사할 수 있습니다.

회귀 분석에는 선형 회귀 분석, 비선형 회귀 분석, 로지스틱 회귀 분석 등이 있습니다. 선형 회귀 분석은 변수 간의 관계가 선형적인 경우 사용되며, 최소 제곱법을 사용하여 최적의 회귀 계수를 추정합니다. 비선형 회귀 분석은 변수 간의 관계가 비선형적인 경우 사용되며, 모델을 이에 맞게 적합시킵니다. 로지스틱 회귀 분석은 종속 변수가 이항 변수(0 또는 1)인 경우 사용되며, 로지스틱 함수를 사용하여 종속 변수의 확률을 추정합니다.

총론적으로, 회귀 분석은 변수 간의 관계를 이해하는 데 유용한 강력한 도구이며, 다양한 분야에서 의사 결정 및 문제 해결에 유용하게 사용될 수 있습니다.

17

Introduction

■ 회귀분석 (regression analysis)

- 변수들의 관계를 기술하고 형태를 파악하는 통계적방법
 - 설명변수와 반응변수의 관계

18

1. 회귀분석의 소개

회귀분석 regression analysis

- (설명)변수와 (반응)변수 간의 함수관계를 규명하는 통계적인 분석방법

예)

설명

반응

- 국민소득이 증가하면 자동차 보유대수는 어느 정도 증가하는가?
- 국민소득이 증가하면 범죄 발생건수는 증가하는가 아니면 감소하는가?
- 자동차사고 발생건수가 증가하면 입원 환자수는 어느 정도 증가하는가?

- 설명변수: 국민소득, 자동차사고 발생건수
- 반응변수: 자동차 보유대수, 범죄 발생건수, 입원 환자수

2

2

1.1 회귀분석의 의미와 유래

회귀분석 regression analysis

- 변수들의 관계를 기술하고 형태를 파악하는 통계적인 기법
- (설명)변수에 기초하여 (반응)변수의 값을 묘사하거나 추정

회귀라는 이름은 분석과 분명히 거리
어색!

- 회귀분석의 「회귀」 용어의 기원

3

3

1.1 회귀분석의 의미와 유래

■ 회귀 용어의 기원

Francis Galton (1822-1911)



우생학자..

- 자녀의 키와 부모의 키의 관계에 대한 분석
: 흥미로운 결과

상당히

자녀의 키는 부모의 키가 크면(작으면)
대체로 크나(작으나)
부모의 키보다는 작으며(크며)
전체 자녀들의 평균 키에 근접하는 경향

「Regression towards mediocrity
in hereditary stature」

"평범한 쪽으로 회귀한다"
(평균에 근접한다)

4

4

1.1 회귀분석의 의미와 유래

■ 회귀 용어의 기원

Karl Pearson (1857 - 1936)



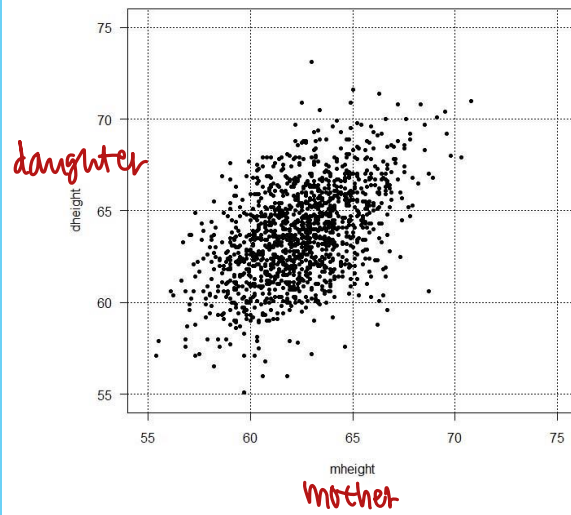
→ 생점도르파악

- 65세 이하 성인 여성과 18세 이상의 딸의 키
- 회귀분석의 이론을 수학적으로 정립

5

5

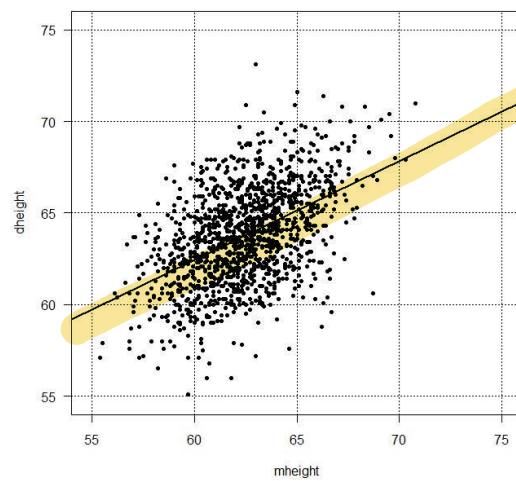
1.1 회귀분석의 의미와 유래



6

6

1.1 회귀분석의 의미와 유래



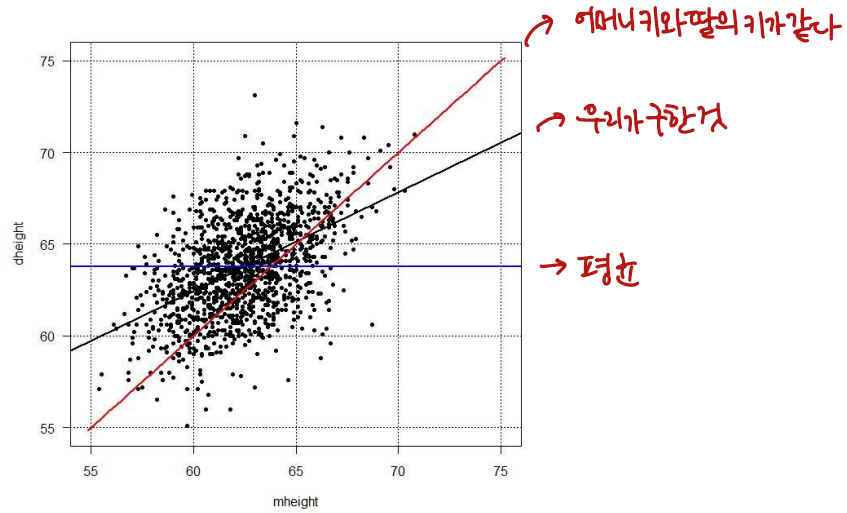
$$Y = 29.92 + 0.542X$$

대략 $y = 30 + 0.5x$

7

7

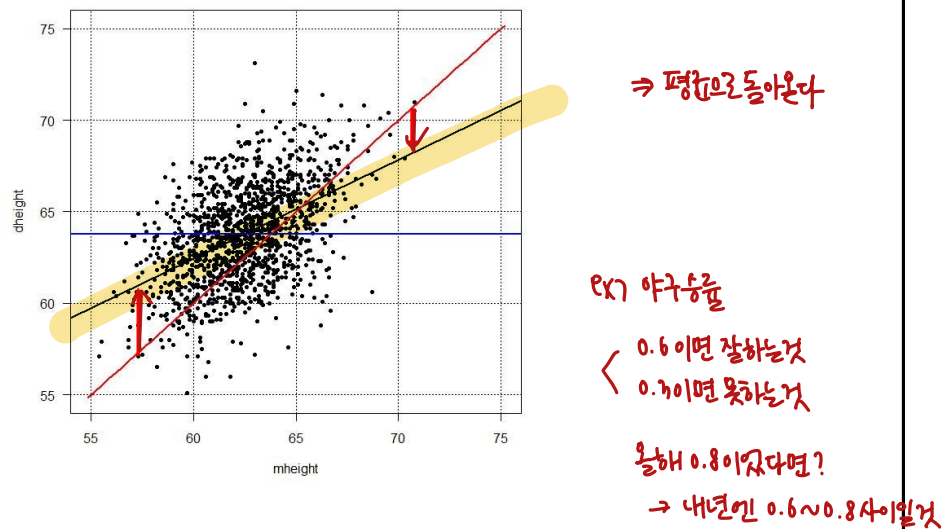
1.1 회귀분석의 의미와 유래



8

8

1.1 회귀분석의 의미와 유래



9

9

의미가 이상한 동어의 또 다른 예: 정규분포 normal distribution
↓
가우스분포

1.2 회귀분석의 활용의 예

예 1.1 포도수확량 자료

- 연간 포도 수확량 예측

X : 나무에 달린 열매개수의 평균 (수확기 전에 조사)

Y : 포도수확량(톤/에이커)

표 1.1 포도수확량 자료

연도	수확량(Y)	열매개수(X)	연도	수확량(Y)	열매개수(X)
1971	5.6	116.37	1978	4.8	125.24
1973	3.2	82.77	1979	4.9	116.15
1974	4.5	110.68	1980	4.7	117.36
1975	4.2	97.50	1981	4.1	93.31
1976	5.2	115.88	1982	4.4	107.46
1977	2.7	80.19	1983	5.4	122.30

자료원 : Casella와 Berger(2001)

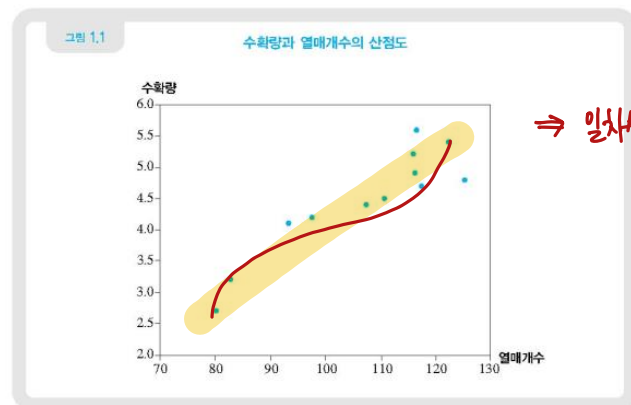
10

10

1.2 회귀분석의 활용의 예

- 산점도(scatter plot): 변수 간의 함수관계

▶ 수확량은 열매의 개수에 비례



11

11

1.2 회귀분석의 활용의 예

■ 관계식

$$Y = \beta_0 + \beta_1 X \quad (1.1)$$

- 변수간의 함수 관계: 자료를 반영하지 못함 → 실제로 일사석위에 있지 않음 (약간의 변동 有)
- 오차 ϵ 이 포함되면 올바른 관계를 나타냄

■ 회귀모형 (통계적 관계식)

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (1.2)$$

- X : 설명변수(explanatory variable) / 독립변수
- Y : 반응변수(response variable) / 종속변수

→ 우리가 알고자하는 것: 모집단의 특성

- 회귀계수(regression coefficient) / 모수(parameter)

- β_0 : X 가 0일 때의 Y 의 값, 절편(intercept)
- β_1 : X 가 한 단위 증가함에 따른 Y 의 변화량, 기울기(slope)

→ 표본을 이용해 추정 필요

12

12

1.2 회귀분석의 활용의 예

예 1.2 회사생산성

- 생산성 향상을 위한 재교육 훈련의 적정 기간의 결정

X : 훈련기간 → 독립변수

Y : 생산성 → 종속변수

- 훈련기간이 늘어나도 생산성이 계속 증가하지는 않음

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon \quad (1.3)$$

- β_2 곡선의 정도(curvature: 곡률)를 나타내는 모수
- X 와 Y 의 비선형의 관계를 기술

* 변수에 대해서는 비선형이지만 모수에 대해서는 선형

13

13

1.2 회귀분석의 활용의 예

- 선형모형 / 비선형모형

- ▶ 선형 여부: 설명변수를 기준 또는 모수를 기준

- 설명변수에 대한 선형

$\frac{\partial Y}{\partial X}$: 설명변수가 없어지면

- ✓ □ 모수에 대해 선형

$\frac{\partial Y}{\partial \beta_0}, \frac{\partial Y}{\partial \beta_1}, \frac{\partial Y}{\partial \beta_2}$: 모수가 없어지면

모형 (1.2) $Y = \beta_0 + \beta_1 X + \epsilon$ \Rightarrow 변수/모수에 대해 선형
 모형 (1.3) $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$ \Rightarrow 모수에 대해 선형
 (선형모형으로 간주)

14

14

1.2 회귀분석의 활용의 예

pass

- 자료수집 방법

- (예 1.1)과 (예 1.2)는 자료수집의 형태에서 차이

- 관찰에 의한 자료수집 (observational study)
 - 계획적이고 통제적인 실험 (controlled experiment에)

- ▶ 자료수집 형태에 의해 모형의 특성, 분석의 방법도 달라짐

15

15

1.2 회귀분석의 활용의 예

예 1.3 자동차 속도와 제동거리

- 제동거리에 속도의 영향 파악/적정 차간거리를 설정

X : 속도
 Y : 제동거리

- 제동거리(Y)는 속도(X)에 비례하므로 모형 (1.2) (?) $\rightarrow \times$

- 문제점

- 자동차의 속도 이외의 도로의 상태 등의 영향
- 속도와 도로 상태의 교호작용이 제동거리에 영향
- 속도와 제동거리의 선형관계는 일정한 속도 범위

- 제동거리와 여러 설명변수 사이의 선형관계

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \quad (1.4)$$

16

여러다른인들

자신이 아니냐면? $\rightarrow \times$
항을 아니냐는 요인이 있냐면?
:

16

1.2 회귀분석의 활용의 예

예 1.4 교통사고 자료

- 1980-2014년 자동차 등록대수, 교통사고 발생건수/사망자

4개 변수
 \rightarrow 2개씩 묶어 2개도 2개면

표 1.2 교통사고 자료

연도	자동차 등록대수 (천 대)	교통사고 발생건수 (천 건)	교통사고 사망자수 (명)	연도	자동차 등록대수 (천 대)	교통사고 발생건수 (천 건)	교통사고 사망자수 (명)
1980	528	120	5608	1998	10470	240	9057
1981	572	123	5804	1999	11164	276	9353
1982	647	141	6110	2000	12059	290	10236
1983	785	170	6834	2001	12914	261	8097
1984	948	134	7468	2002	13949	231	7222
1985	1113	147	7522	2003	14587	241	7212
1986	1309	154	7702	2004	14934	221	6563
1987	1611	176	7206	2005	15397	214	6376
1988	2035	225	11563	2006	15895	214	6327
1989	2660	256	12603	2007	16428	212	6166
1990	3395	255	12325	2008	16794	216	5870
1991	4248	266	13429	2009	17325	232	5838
1992	5231	257	11640	2010	17941	227	5505
1993	6274	261	10402	2011	18437	222	5229
1994	7404	266	10087	2012	18871	224	5392
1995	8469	249	10323	2013	19401	215	5092
1996	9553	265	12653	2014	20118	224	4762
1997	10413	246	11603				

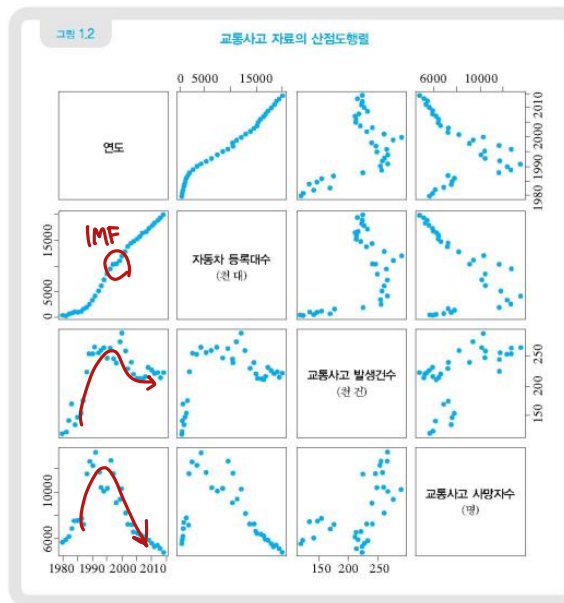
자료원: 통계청(2014)

17

17

1.2 회귀분석의 활용의 예

- 모든 변수들의 자료에 대한 산점도행렬 (scatterplot matrix)



18

18

1.2 회귀분석의 활용의 예

- 해가 지남에 따라
 - 자동차 등록대수: 계속적인 증가
 - 90년대 들어 급격히 증가
 - 2000년대 들어 증가 속도 둔화
 - 1998년에는 사회적인 현상으로 증가하지 않음
 - 교통사고 발생건수/교통사고로 인한 사망자의 수
 - 초반에는 증가
 - 90년대에 최고값을 기록하고 다시 감소
- 자동차 등록대수가 늘어남에 따라
 - 교통사고 발생건수와 사망자의 수
 - 증가하다가 다시 감소

19

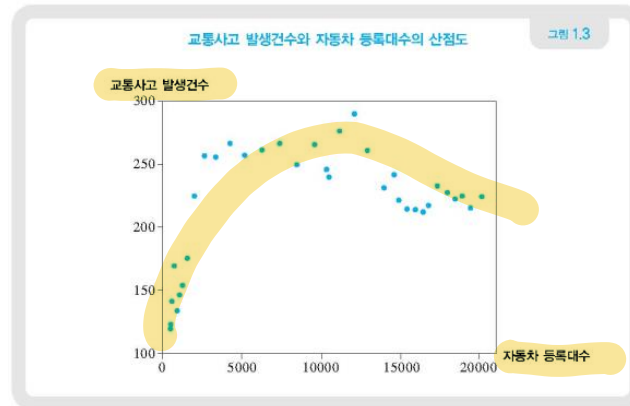
19

1.2 회귀분석의 활용의 예

- 교통사고 발생건수 Y 와 자동차 등록대수 X 의 관계
 - X 가 증가하면 Y 의 증가율이 떨어지거나 감소
 - 모형 (1.3)과 유사하고 절편 없는 모형

$$Y = \beta_1 X + \beta_2 X^2 + \epsilon$$

(1.5)



20

20

1.2 회귀분석의 활용의 예

예 1.5 편의점의 매출액

- 편의점 매출액 (Y)은 다음에 따라 결정된다는 마케팅 이론

- X_1 : 인근 지역의 인구
- X_2 : 소비자와 편의점의 실제 거리
- X_3 : 편의점의 매장 크기

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

(1.6)

- ▶ 편의점 위치 선정
- ▶ 매장 크기 결정

+ 근처에 다른 편의점이 있는지

⇒ 비용대비 매출극대화하기.

21

21

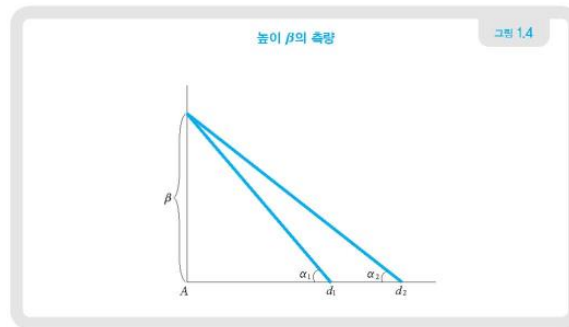
1.2 회귀분석의 활용의 예

예1.6 구조물의 높이 측정

- 직접 측정이 불가능한 구조물의 높이 측정
 - 구조물에서 거리가 d 인 지점에서 각도 α 측정

$$\tan \alpha = \beta / d \rightarrow \beta = d \cdot \tan \alpha$$

- 거리가 d_i 인 지점에서 α_i 또는 $\tan \alpha_i$ 측정, $i = 1, 2, \dots, n$



22

22

1.2 회귀분석의 활용의 예

- 오차가 없다면

$$\beta = (\tan \alpha_1)d_1 = (\tan \alpha_2)d_2 = \dots = (\tan \alpha_n)d_n \quad (1.7)$$

$$\tan \alpha_i = \beta / d_i, \quad i = 1, \dots, n$$

- y_i 를 $\tan \alpha_i$ 의 측정값이라 하고 측정오차(ϵ_i)를 감안하면

$$y_i = \tan \alpha_i + \epsilon_i = \beta \frac{1}{d_i} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (1.8)$$

오차누정

$$Y: \tan \alpha, \quad X: 1/d$$

- 반응변수를 y_i , 설명변수를 $x_i = 1/d_i$ 로 하는 모형

$$y_i = \beta x_i + \epsilon_i, \quad i = 1, 2, \dots, n \quad (1.9)$$

23

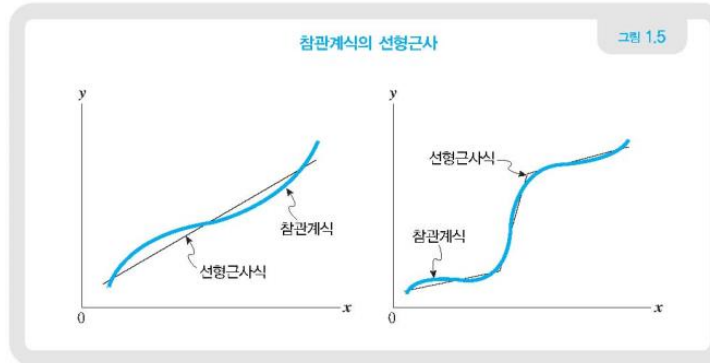
23

1.2 회귀분석의 활용의 예

■ 선형과 비선형모형

수익니까.

- 참의 관계식은 대부분 비선형관계이지만 **선형근사식** 이용
- 구조가 비선형이라도 이차함수 등으로 선형모형으로 표현
- 하나의 선형식으로 근사가 안되면 선형의 연결로 관계 근사



24

24

1.2 회귀분석의 활용의 예

예1.2 회사생산성

- 생산성 향상을 위한 재교육 훈련의 적정 기간의 결정

X : 훈련기간

Y : 생산성

- ▶ 훈련기간이 늘어나도 생산성이 계속 증가하지는 않음

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon \quad (1.3)$$

- β_2 : 곡선의 정도(curvature: 곡률)를 나타내는 모수
- X 와 Y 의 비선형의 관계를 기술
- 변수에 대해서는 비선형이지만 모수에 대해서는 선형

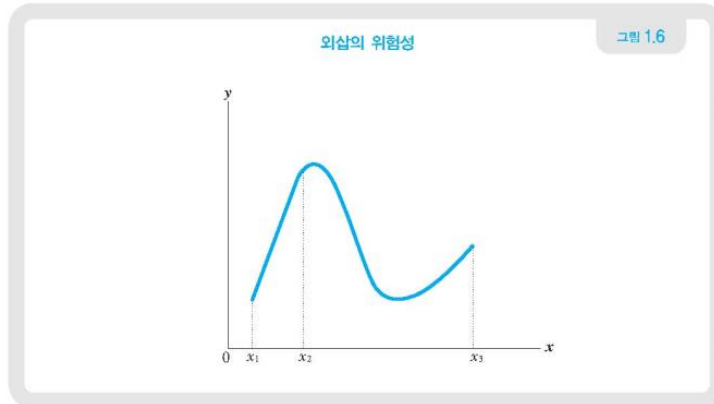
25

25

1.2 회귀분석의 활용의 예

- 외삽의 위험성

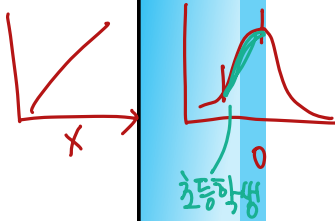
- 적절한 범위에서만 모형 (1.2) 등의 선형모형을 사용해야 함
- $x_1 \sim x_2$ 자료에 근거한 모형을 $x_2 \sim x_3$ 에 적용하는 것은 무리



26

26

ex) 귀성장



1.2 회귀분석의 활용의 예

- 비선형 모형

- 선형관계를 갖지 않은 모든 모형
- 이론적 설명이 가능(특수한 함수로 모형화)

$$Y = \alpha X_1^\beta X_2^\gamma \epsilon \quad (1.11)$$

$$Y = 1/(\beta_0 + \beta_1 X + \epsilon) \quad (1.12)$$

$$Y = \alpha e^{\beta X} + \epsilon \quad (1.13)$$

$$Y = \alpha + e^{\beta X} + \epsilon \quad (1.14)$$

- 모형 (1.11): 양변에 자연로그변환

$$\ln Y = \ln \alpha + \beta \ln X_1 + \gamma \ln X_2 + \ln \epsilon \quad (1.15)$$

- 모형 (1.12): Y의 역수를 새로운 Y로 치환

$$1/Y = \beta_0 + \beta_1 X + \epsilon \quad (1.16)$$

27

27

선형모형으로
변환가능

1.2 회귀분석의 활용의 예

■ 본질적/변환가능한 선형 (intrinsically/transformable linear)

▫ 모형 (1.15), (1.16)

■ 본질적 비선형 (intrinsically non-linear)

▫ 모형 (1.13), (1.14)

✕ 비선형모형 (non-linear model)

일반화선형모형 (generalized linear model)

▫ 교재 10장

28

28

1.2 회귀분석의 활용의 예

■ 모형개발

● 설명변수의 선택

- 관계식을 뒷받침해 줄 수 있는 적절한 변수
- 측정이 용이하고 측정오차가 적은 변수
- 외삽에 주의 (자료의 범위를 신중하게 선택)

● 설명변수의 선택의 과정

- 기존의 설명변수를 감안한 특정 설명변수의 설명력

너무 많으면 안됨. 최소한의 선택

● 설명변수와 반응변수의 변환이 더 좋은 결과를 보이는 경우가 있음

인식력, 제곱근, ...

⇒ 선형모형으로 변환 아 더 나은 결과

(회귀론에서 다룸)

29

29

1.3 회귀분석의 목적

■ 회귀분석의 기본적인 목적

- (1) 자료의 기술(data description)과 모수의 추정(estimation)
- (2) 적합한 모델을 통한 예측(prediction)
- (3) 통계적인 제어/통제(control)

예 1.1) - (2): 포도수확량을 미리 예측하여 수입 여부를 판단

예 1.2) - (1) (2) (3): 적절한 생산성에 필요한 교육기간 산정

예 1.3) - (1) (2): 도로에서의 안전거리를 설정

예 1.4), 1.5) - (1) (2): (예 1.3) 과 비슷한 논리

예 1.6) - (1): 모수의 추정으로 높이를 추정

- 예측 목적일 때 자료 범위의 타당성과 모수의 정확한 추정이 전제
(부실한 자료로 인한 모수의 추정과 예측은 정확성이 보장 안됨)

30