

계량 사회과학 입문

정오표

i! 에이콘 출판사

[p.32 : 1.3절 1행]

이 책의 남은 부분의

→

이 책을 활용하기 위한

[p.60 : 마지막행]

실험 연구 설계로 수집됐다. 이는 처치변수 또는 연구자가 관심 있는 인과변수가 결과변수에 미치는 인과효과를 연구하고자 조작됐다. 이 예제에서 처치변수는 가상 구직자의 인종이며, 이는 이력서의 적힌 이름으로 유추된다. 결과변수는 전화 회신(콜백) 여부로 후보자가 정해진다. 연구자가 관심 있는 것은 다른 이름의 이력서가 회신율의 차이를 가져오는지에 있다.

→

실험 연구 설계로부터 수집됐다. 이력서 예제의 가상 구직자의 인종이 처치 변수가 되며, 인종은 이력서의 이름으로 유추한다. 지원한 회사로부터 전화 회신(콜백)을 받는지 여부가 결과 변수가 된다. 연구자는 이름에 따라 회신율이 달라지는지 여부에 관심이 있다.

[p.64 : 마지막행]

있으며, 이는

→

있다. 두 논리값은

[p.69 : 4행]

그런 다음 명령문은 call 변수의 값이 1인 부분 선택된 관측값의 비율과 같은 mean() 함수를 사용해 이 부분 선택된 벡터의 샘플 평균을 계산한다.

→

그런 다음 명령문은 mean() 함수를 사용해서 부분 선택된 벡터의 표본평균을 계산한다. mean() 함수는 call 변수의 값이 1인 부분 선택된 관측값의 비율과 같다.

[p.74 : 5행]

sort() 함수를 사용해 프레젠테이션을 쉽게 하고자 결과를 오름차순으로 정렬할 수 있다.

→

프레젠테이션을 쉽게 하고자 sort() 함수를 사용해 결과를 오름차순으로 정렬할 수 있다.

[p.75 : 아래에서 10행]

이력서를 고용주에게 발송하지

→

동일한 이력서를 같은 고용주에게

[p.75 : 아래에서 8행]

연구자들이 최저임금을 올리면 실업률이 증가하는지를 알아내는 데 관심이 있는 또 다른 사례를 생각해 보자.

→

또 다른 사례로 최저임금을 올리면 실업률이 증가하는지의 여부에 관심 있는 연구를 생각해 보자.

[p.76 : 7행]

처치변수(treatment variable) T가 결과변수 Y의 변화를 일으키는지 연구자들은 두 가지 잠재적 결과, 즉 각각 Y(1) 및 Y(0)로 표현된 처치의 존재와 부재로 실현된 Y의 잠재적 가치를 살펴봐야 한다.

→

처치변수(treatment variable) T가 결과변수 Y의 변화를 일으키는지 여부를 알기 위해서는 두 가지 잠재적 결과 값을 확인해야 한다. 예를 들어 각각 Y(1) 및 Y(0)로 표현된 처치의 존재와 부재로 실현된 Y의 잠재적 가치들이 있다.

[p.79 : 7행]

평균 결과 차이 평균의

→

평균 결과 차이 혹은 평균의

[p.81 : 3행]

특히 미시간 주 예비 선거기간 중에 그들은

→

특히 연구자들은 미시간 주 예비 선거기간 중에

[p.87 : 2행]

인상됐는지 확인한다.

→

인상됐는지 여부를 먼저 확인한다.

[p.88 : 7행]

이 이론을 테스트하고자 뉴저지 법이 시행된 후 뉴저지와 펜실베이니아 식당들 사이에서 이 변수의 표본평균을 간단히 비교해 정규직 직원의 비율을 주요 결과변수로 조사한다.

→

이 이론을 테스트하고자 정규직 직원의 비율을 주요한 결과변수로 조사한다. 그러기 위해 뉴저지에서 법률이 시행된 후 뉴저지와 펜실베이니아 식당들 사이에서 이 변수의 표본평균을 간단히 비교한다.

[p.88 : 아래에서 7행]

처치군과 대조군은

→

처치그룹과 통제그룹은

[p.90 : 6행]

예를 들어, 버거킹이 다른 패스트푸드 체인과 다른 고용 정책을 갖고 있다면 두 주의 이러한 차이는 최저임금 인상과 고용 사이의 관계를 교란시킬 수 있다.

→

두 주의 이러한 차이는 버거킹이 다른 패스트푸드 체인과 다른 고용 정책을 갖고 있는 경우에는 최저임금 인상과 고용 사이의 관계를 교란시킬 수 있다.

[p.90 : 8행]

가능성을 처리하려고

→

가능성에 대처하기 위해

[p.90 : 9행]

이러한 분석을 통해서 통계적 제어로

→

이러한 분석으로 통계적 제어를 통해서

[p.91 : 3행]

다른 교란요인이 될 가능성이 있는 것은 점포의 위치다.

→

점포의 위치는 또 다른 잠재 교란요인이다.

[p.93 : 2행]

평행의 경향(parallel trend)을 따른다는

→

평행 추세(parallel trend)를 따른다는

[p.94 : 2행]

병렬 시간추세 가정에서

→

평행 시간추세 가정에서

[p.99 : 식(2,3)]

entry

→

투입값

[p.110 : 5행]

그러나 전쟁 기간의 민간인 경험과 여론을 측정하는 것은 설문 조사자와 응답자에게 잠재적 위협을 가하는 혹독한 치안 조건 때문에 어려운 과제다.

→

그러나 설문 조사자와 응답자에게 잠재적 위협을 가하는 혹독한 치안 조건 때문에 전쟁 기간의 민간인 경험과 여론을 측정하는 것은 어려운 과제다.

[p.113 : 4행]

이제 이 변수에 대한

→

이제 월 소득에 대한

[p.113 : 6행, 8행, 10행]

결측 데이터

→

결측값

[p.115 : 8행]

여러 범주가 있는(2.2.5절 참고) 요인변수의 분포를 요약하기 위해서 table() 및 prop.table() 함수를 사용해 위에서 생성한 총수 또는 비율이 있는 간단한 표만으로도 충분하다.

→

여러 범주가 있는(2.2.5절 참고) 요인변수의 분포를 요약하기 위해서는 **위에서** 생성한 총수 또는 비율이 있는 간단한 표만으로도 충분하다.

[p.115 : 10행]

하지만 막대그래프(bar plot)를 사용해 분포를 시각화할 수 있다.

→

이를 위해 table() 및 prop.table() 함수를 사용한다. 하지만 막대그래프(bar plot)를 사용해 분포를 시각화할 수 있다.

[p.120 : 마지막행]

따라서 교육의 중앙 연도가

→

교육 연수의 중간값이

[p.124 : 12행]

장치를 닫는 방식으로 명령으로

→

장치를 닫는 방식으로

[p.129 : 1행]

다단계 클러스터 샘플링 방법은 먼저 큰 단위를 샘플링한 다음에 무작위로 작은 단위를 각각의 선택된 큰 단위에서 샘플링하는 여러 단계로 진행된다.

→

다단계 클러스터 샘플링 방법은 먼저 큰 단위를 샘플링한 다음에 **각각의 선택된 큰 단위에서 무작위로 작은 단위를 샘플링하는 여러 단계로 진행된다.**

[p.135 : 2행]

사회과학자들은 이 문제를 극복하고자

→

사회과학자들은 **잠재적 편향**을 극복하고자

[p.139 : 3행]

대신에 경제적 차원은 당파적 차이의 지배적인 설명으로 보이며,

→

대신에 경제적 차원은 당파적 차이를 지배적으로 **설명하는 것으로** 보이며,

[p.140 : 아래에서 6행]

줄거리는 두 정당의 이념적 중심이 시간이 지남에 따라

→

플롯은 두 정당의 이념적 중심이 시간이 지남에 따라

[p.141 : 4행]

소득 불평등을 측정하고자 그래픽으로 가장 잘 이해되는 지니계수(Gini Index)를 사용한다.

→

그래픽으로 가장 잘 이해되는 지니계수(Gini Index)를 사용해 소득 불평등을 측정하고자 한다.

[p.145 : 15행]

연관성의 강점과 약점을 비교할 수 있다.

→

연관성의 **강함**과 **약함**을 비교할 수 있다.

[p.146 : 4행]

상관관계가 긍정적이고 상당히 높은 것으로 나타났으며,

→

상관관계가 **양(+)**이고 상당히 높은 것으로 나타났으며,

[p.153 : 아래에서 4행]

적용하기 전에 입력을

→

적용하기 전에 **입력값**을

[p.175 : 아래에서 3행]

루프 내에서 중첩돼 있음을

→

루프 내에서 중첩(**nested**)돼 있음을

[p.177 : 6행]

데이터의 부분 선택을 만든다.

→

데이터를 부분 **선택한다.**

[p.180 : 아래에서 2행]

투표 예측과 오바마의 실제 마진(차이)을 나타낸다.

→

여론조사 예측과 오바마의 실제 마진(차이)을 나타낸다.

[p.183 : 아래에서 6행]

각 후보의 평균 지지를 계산하고,

→

각 후보의 평균 지지~~를~~을 계산하고,

[p.184 : 3행]

또한 루프 내에서 데이터 pollsUS08의 부분 선택을 설정해 결과 데이터에 지난 7일 동안 수행된 설문조사와 그 날 자체만 포함되도록 한다.

→

또한 **생성된 데이터가 선거일과 지난 7일 이내에 행해진 여론조사만을 포함하도록 루프 내에서 pollsUS08의 부분 선택을 정의한다.**

[p.184 : 7행]

부분 선택 데이터의 이 변수는

→

데이터의 부분 선택을 정의하는 이 변수는

[p.186 : 1행]

사전선거 투표가 상당히 정확하다는 것을

→

사전선거 **설문조사**가 상당히 정확하다는 것을

[p.192 : 1행]

선형회귀모형을 맞추고자

→

선형회귀모형을 **적합하기 위해**

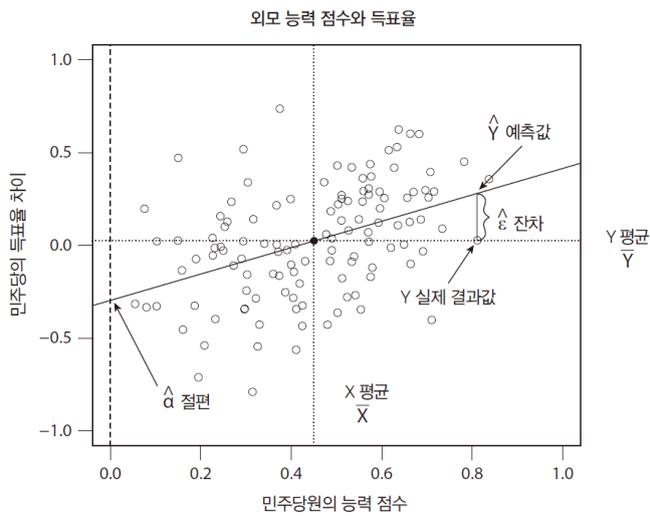
[p.193 : 아래에서 3행]

lm() 함수에서 출력 객체를 산점도에 회귀선을 추가하는 것은 abline() 함수를 사용하면 간단하다.

→

lm() 함수에서 **출력된 객체를 입력값으로 하는 abline() 함수를 사용하면 간단하게 산점도에 회귀선을 추가하는 것이 가능하다.**

[p.194 : 그림]



‘의’ 삭제

[p.195 : 아래에서 5행]

p.190의 그림에서 이것이 실제 사실임을

→

p.194의 그림에서 이것이 실제 사실임을

[p.197 : 아래에서 10행]

그는 이것을 "부모의 키의 평균치가 전체 평균보다~

→

그는 **평균으로의 회귀**를 "부모의 키의 평균치가 전체 평균보다

[p.198 : 아래에서 2행]

기울기로 알 수 있듯이 부모가 더 큰 자녀의 경우도 평균적으로 더 큰 경향이 있다.

→

기울기로 알 수 있듯이 **부모의 키가 큰 자녀의** 경우도 평균적으로 키가 큰 경향이 있다.

[p.202 : 5행]

현재 예에서는 두 데이터프레임의 동일한 행에 전미 각 주가 반드시 동일한 행에 있어야 한다.

→

현재 예에서는 **전미 각 주가 반드시 두 데이터프레임의** 동일한 행에 있어야 한다.

[p.206 : 7행]

46% 정도만 2012년에 전회를 상회하는

→

46% 정도만 2012년에 **지난번 대선을** 상회하는

[p.208 : 4행]

출력값을 입력으로 받는다

→

출력값을 입력값으로 한다

[p.209 : 3행]

얻은 결정계수는 이전 선거 결과를 사용해 같은 정당 후보자의 득표를 예측했기 때문에 상대적으로 낮게 나타난다.

→

이전 선거 결과를 사용해 같은 정당 후보자의 득표를 예측했기 때문에 결정계수는 상대적으로 낮게 나타난다.

[p.209 : 5행]

대응하는 출력 객체가 주 레벨의 회귀분석의 출력을 나타내는 fit1인 회귀분석의 R² 값을 계산할 수 있다.

→

회귀분석의 R² 값을 주 레벨의 회귀분석 결과를 나타내는 출력 객체인 fit1을 사용해서 계산할 수 있다.

[p.211 : 아래에서 1행]

팜비치 없이 결정계수가 0.51에서 0.85로 크게 증가하는 것을 볼 수 있다.

→

팜비치를 뺀 결정계수는 0.51에서 0.85로 크게 증가하는 것을 볼 수 있다.

[p.214 : 8행]

의장으로 선출된 비율을 계산하는 것으로 확인해 보자.

→

의장으로 선출된 비율을 계산해 보자.

[p.219 : 9행]

fitted() 함수와 마찬가지로 이 함수는 lm() 함수에서 출력을 가져와 예측값을 계산한다.

→

이 함수는 fitted() 함수와 마찬가지로 lm() 함수에서 출력값을 입력값으로 취해서 예측값을 계산한다.

[p.220 : 아래에서 9행]

결과적으로 각 처치조건(Civic Duty, Hawthorne 또는 Neighbors) 통제그룹과 비교한 평균처치효과(ATE)는 각 처치그룹의 계수에서 절편이 없는 모형의 기준그룹으로 통제그룹의 계수를 뺀 값으로 계산할 수 있다.

→

결과적으로 각 처치그룹(Civic Duty, Hawthorne, Neighbors)의 통제그룹과 비교한 평균처치효과(ATE)는 각 처치그룹의 계수에서 절편이 없는 모형의 기준그룹인 통제그룹의 계수를 뺀 값으로 계산할 수 있다.

[p.222 : 3행]

않은 모두를 얻을 수 있다(7.3절 참고).

→

않은 R^2 모두를 얻을 수 있다(7.3절 참고).

[p.223 : 3행]

그들은 그러한 처치에 가장 큰 영향을

→

그들은 **사회적 압력** 처치에 가장 큰 영향을

[p.229 : 2행]

함수를 사용해 다양한 시나리오에서

→

함수를 사용하는 것이다. 즉, 다양한 시나리오에서

[p.247 : 아래에서 2행]

스태밍(stemming)이라고

→

어간화/스태밍(stemming)이라고

[p.255 : 5행]

출력을 입력으로 사용한다.

→

출력값을 **입력값**으로 사용한다.

[p.259 : 1행]

부분 선택해 이러한 단어만 포함시키도록 한다.

→

부분 선택해 **앞서 언급한 10개의** **단어들**만 포함시키도록 한다.

[p.259 : 3행]

그런 다음 각 저자의 전체 문서에서 해밀턴과 매디슨의

→

그런 다음 **각각 해밀턴과 매디슨의 전체 문서에서**

[p.260 : 5행]

저자를 예측하고자 예비분석을 기초로 선택한 4개 단어의 용어빈도를 사용한다

→

예비분석을 기초로 선택한 4개 단어의 용어빈도를 **사용해서 저자를 예측하고자** 한다.

[p.261 : 아래에서 7행]

피팅하고자 사용했던 같은 데이터를 예측 정확도를 평가하는 데 다시 사용한다.

→

피팅하고자 사용했던 것과 같은 데이터가 예측 정확도를 평가하는 데 다시 사용된다.

[p.267 : 아래에서 3행]

지정해 인접행렬의 모든 대각선 요소에 대해 값이 0이 된다.

→

지정해서 인접행렬의 대각선상의 모든 요소의 값이 0이 되게 한다.

[p.268 : 아래에서 8행]

차수 대신, 주어진 노드에서 직접 연결되지 않은 노드를 포함한 그래프 내의 다른 모든 노드까지 에지의 합을 계산할 수 있다.

→

차수 대신에 직접적으로 연결되지 않은 노드를 포함하는 주어진 노드에서 그래프 내의 다른 모든 노드까지 에지의 총 합을 계산할 수 있다.

[p.268 : 아래에서 2행]

두 노드 간의 거리는 관심의 두 노드 사이에 연결된 노드의 최단 시퀀스인 최단 경로의 에지 수다.

→

두 노드 사이의 거리는 둘을 잇는 최단 에지 수이며, 두 개의 관심 노드 간의 최단 경로의 노드 연결 수이다.

[p.270 : 2행]

특히 한 쌍의 노드 사이의 커뮤니케이션이 최단 경로를 통해서 발생한다고 가정한다면 이러한 최단 경로에 있는

→

구체적으로 한 쌍의 노드 사이 커뮤니케이션이 최단 경로를 통해서 발생한다고 가정한다면 최단 경로에 있는

[p.270 : 6행]

토르나부오니(Tornabuon) 가문

→

토르나부온(Tornabuon) 가문

[p.270 : 아래에서 3행]

지금까지 메디치 가족이 가장 높은 매개성의 값을 갖고 있는 것을 발견했다. 실제로 모든 주어진 노드는 105의 다른 노드와 고유하게 쌍을 이룰 수 있기 때문에 메디치 가문은 45% 이상의 가능한 모든 다른 노드 쌍의 최단 경로에 있다.

→

지금까지 메디치 가문이 가장 높은 매개성의 값을 갖고 있는 것을 알 수 있다. 실제로 어떠한 주어진 노드는 다른 105개의 노드들과 고유하게 쌍을 이룰 수 있기 때문에 메디치 가문은 가능한 모든 다른 노드 쌍의 45% 이상의 최단 경로를 매개하고 있다.

[p.273 : 아래에서 8행]

유방향 네트워크에 대해 앞서 논의한 세 가지의 중심성 척도들을 정의할 수 있다.

→

앞서 논의한 세 가지의 중심성 척도들은 유방향 네트워크에 대해 정의할 수 있다.

[p.275 : 1행]

유방향 네트워크 데이터를 정의할 수 있다.

→

유방향 네트워크 데이터에서도 똑같이 정의할 수 있다.

[p.277 : 7행]

중심성 척도의 마지막 대안으로

→

마지막으로, 중심성 척도의 하나인

[p.277 : 7행]

페이지랭크는 구글의 공동 창업자인 세르게이 브린과 래리 페이지에 의해 검색엔진의 결과인 웹사이트 순위를 최적화하고자 개발됐다.

→

페이지랭크는 구글의 공동 창업자인 세르게이 브린과 래리 페이지에 의해 개발된 것으로, 검색엔진의 결과인 웹사이트 순위를 최적화한다.

[p.277 : 아래에서 6행]

또한 만약 노드가 더 많은 수의 내향에지를 갖는 또 다른 노드로부터 내향에지를 가질 때는, 적은 수의 내향에지를 가진 노드로부터 내향에지를 갖는 경우보다 큰 페이지랭크 값을 얻는다.

→

만약 노드가 더 많은 내향에지를 가지는 또 다른 노드로부터 내향에지를 가질 때는, 적은 내향에지밖에 못 갖는 또 다른 노드로부터 내향에지를 가질 경우보다 큰 페이지랭크 값을 얻는다.

[p.281 : 3행]

페이지랭크 값 차이의 최대 절대값으로 0.001을 사용한다.

→

페이지랭크 값 차이의 절대값의 최대값으로 0.001을 사용한다.

[p.281 : 아래에서 5행]

가장 많은 내향에지가 있기 때문에

→

가장 많은 내향에지를 가지고 있기 때문에

[p.282 : 12행, 18행]

워터

→

수도

[p.282 : 아래에서 8행]

그 당시 조사 결과 위에서 언급한 펌프 우물의 물을 마시는 습관을 가진 사람들을 제외하고는 런던의 이 지역에서 콜레라의 창궐은 없었다.

→

그 당시 조사 결과, 런던의 이 지역에서 위에서 언급한 펌프 우물의 물을 마시는 습관을 가진 사람들을 제외하고는 콜레라의 창궐 혹은 증가는 없었다.

[p.283 : 1행]

존 스노는 겹치는 지역은

→

존 스노는 수도 공급이 중복하는 지역은

[p.283 : 3행]

두 그룹의 고객들은 다른 측면에서는 유사하다고 가정할 때 콜레라 발병 비율의 차이는 회사 선택에 따른 결과다.

→

두 그룹의 고객들이 다른 모든 측면에서는 유사하다고 가정할 때 콜레라 감염율의 차이는 회사 선택에 따른 결과다.

[p.284 : 3행]

콜레라 발병률과 관련된 변수들을

→

콜레라 감염률과 관련된 변수들을

[p.285 : 3행]

모든 사회계층의 사람이 자신의 선택 없이 대부분은 두 그룹으로 나뉘었다.

→

모든 사회계층의 사람이 사람들이 선택의 여지 없이 대부분 두 그룹으로 나뉘었다.

[p.287 : 5행]

함수를 사용해 이러한 도시를 지도에 추가하고

→

함수를 사용해 인구수가 가장 많은 7개 도시를 지도에 추가하고

[p.291 : 1행]

이 그림에서 plot() 함수에서 산만함을 피하고자

→

이 그림에서는 산만함을 피하고자 plot() 함수에서

[p.292 : 8행]

우승했기 때문에 파란색으로 채색한다.

→

승리했기 때문에 파란색으로 채색한다.

[p.295 : 1행]

미국의 다국적 물류 소매 체인인

→

미국의 다국적 **소매 물류** 체인인

[p.295 : 3행]

그 후 수십 년 동안 미국 내뿐 아니라

→

그 후 수십 년 동안 미국 **뿐만** 아니라

[p.299 : 아래에서 2행]

국내 남은 지역으로의 확장 속도이며,

→

미국 내 미진출 지역으로의 확장 속도이며,

[p.309 : 3행]

노이즈는 실제 모든 데이터에서 발생하는, 관련 없는

→

노이즈는 실제 모든 데이터**셋**에서 발생하는 **패턴과는** 관련 없는

[p.309 : 아래에서 3행]

확률은 수많은 유용한 정리가 도출된 세 가지 간단한 공리 집합을 기반으로 한다.

→

확률은 세 가지 단순한 공리의 집합에 기반 한다. 거기에서 셀 수 없이 많고 유용한 정리가 도출된다.

[p.310 : 11행]

동전 던지기는 이러한 조건에는 초기 동전의 각도 및 속도,

→

동전 던지기의 **경우에는** 이러한 조건에는 **동전의 초기** 각도 및 속도,

[p.312 : 11행]

1. 실험: 관심 있는 확률적 사건을 생성하는 작업 또는 일련의 작업.

→

1. 실험: **관심 확률 사건을 생성하는 하나의 행위 또는 행위의 집합.**

[p.317 : 3행]

n개의 고유한 대상 중 k개의 대상을 배열할 때의

→

n개의 고유한 대상 중 k개의 **개체**를 배열할 때의

[p.321 : 7행]

여러 번 인터뷰하지 않기 때문에 대체하지 않은 표본추출의 예다.

→

여러 번 인터뷰하지 않기 때문에 **비복원추출**의 예다.

[p.321 : 아래에서 6행]

시뮬레이션 추정값은 분석 해답인 0.507에

→

시뮬레이션 추정값은 **분석해인** 0.507에

[p.325 : 9행]

6.3.1 조건부 확률, 주변확률, 결합확률

→

6.2.1 조건부 확률, 주변확률, 결합확률

[p.344 : 6행]

조건부 확률 $P(\text{인종} | \text{성})$ 이 가장 큰 인종 범주가 자기신고 인종과 동일한 경우 인종이 올바르게 분류된 것으로 간주된다.

→

여기서는 인종의 조건부 확률 $P(\text{인종} | \text{성})$ 이 최대가 되는 인종 범주와 자기신고한 인종이 일치할 때 인종이 올바르게 분류된 것으로 간주된다.

[p.347 : 11행]

특정 선거구의 거주자 중 각 인종 범주의 비율

→

특정 선거구의 거주자 중 각 **인종** 범주의 비율

[p.352 : 1행]

그러니 일단 그들의 거주지를 고려하면 그 사람들이 흑인이라는 예측확률은

→

그러나 일단 거주지를 고려하면 **성이 'White'인** 사람이 흑인이라는 예측확률은

[p.355 : 7행]

값은 잘 일어나지 않기 때문이다.

→

어떠한 값도 나타나지 않기 때문이다.

[p.356 : 아래에서 2행]

불가산 무한개의 값을 취하는 것으로

→

불가산 **무한대**의 값을 취하는 것으로

[p.358 : 아래에서 5행]

여기서 소개한 두 분포에는 편리한 연결이 있다.

→

여기서 **도입한** 두 분포에는 **유용한** 편리한 연결을 공유한다.

[p.370 : 7행]

식 (6.30)에 주어진 확률모델에서 2012년 오바마의 득표에 대해 도출된 정규분포에서 2008년의 득표율에 상당하는 면적을 계산할 수 있다.

→

식 (6.30)의 **확률모델에서** 2012년 오바마의 득표에 대해 도출된 정규분포하에서 2008년의 득표율에 상당하는 면적을 계산할 수 있다.

[p.385 : 아래에서 3행]

정리는 표본평균의 z 점수가 표본크기가 증가함에 따라

→

중심극한정리는 표본평균의 z 점수가 표본크기가 증가함에 따라

[p.397 : 4행]

무작위 대조시험에서 평균 결과가 처치그룹과 통제그룹 간에 다를 경우

→

무작위 대조시험에서 **처치그룹과 통제그룹 간의 평균 결과가** 다를 경우

[p.399 : 9행 #일괄변경#]

표본분포

→

표집분포

[p.400 : 아래에서 1행]

왜냐하면 그것이 모집단 평균에

→

왜냐하면 **표본평균**이 모집단 평균에

[p.402 : 3행]

그러나 각 단위에 대해 하나의 잠재적 결과밖에 관찰할 수 없어서 어느 단위의 처치효과도 관찰할 수 없으므로 SATE는 미지의 것이다.

→

그러나 각 단위에 대해 하나의 잠재적 결과밖에 관찰할 수 **없기 때문에**, 어떠한 단위에 관해서도 **처치효과는 관찰할 수 없다**. **그렇기 때문에** SATE는 미지의 것이다.

[p.402 : 12행]

그것이

→

단위 i 가

[p.418 : 2행]

90%의 횡수를 포함해야만 한다.

→

90%의 횡수로 포함해야만 한다.

[p.424 : 아래에서 6행]

표준규모 학급 등에 무작위로 배정된다. 소규모 학급이

→

표준규모 학급 등으로 무작위 배정된다. 이 실험에서는 소규모 학급이

[p.433 : 아래에서 7행]

올바르게 분류했을 가능성과 관련이 있다.

→

올바르게 분류했는지에 대한 가능성과 관련이 있다.

[p.438 : 아래에서 6행]

이 함수의 주 입력은

→

이 함수의 주 입력값은

[p.440 : 7행]

보다 작거나 같다면 귀무가설을 기각하고

→

유의수준 α 보다 작거나 같다면 귀무가설을 기각하고

[p.445 : 5행]

전자에는 효과의 크기에 대한 정보도 포함돼 있어서

→

왜냐하면 신뢰구간은 효과의 크기에 대한 정보도 포함돼 있어서

[p.450 : 9행, 11행]

구직자

→

지원자

[p.450 : 아래에서 5행]

콜백을 받을 확률이 흑인처럼 들리는 이름의 이력서와 백인 이름처럼 들리는 이력서 간에 동일하다는 귀무가설을 검정한다.

→

콜백을 받을 확률이 흑인 이름처럼 들리는 이력서와 백인처럼 들리는 이름의 이력서 간에 동일하다는 귀무가설을 검정한다.

[p.453 : 5행]

경험적 결과가 통계적으로 유의한지를 결정하는 기준으로 간주된다.

→

경험적/실증적 결과가 통계적으로 유의한지를 결정하는 기준으로 간주된다.

[p.454 : 2행]

바로 아래에 있는 경험적 결과들을

→

바로 아래에 있는 경험적/실증적 결과들을

[p.466 : 아래에서 9행]

2장에서 하나의 전략으로 처치받은

→

하나의 전략으로 2장에서는

[p.480 : 10행]

선형회귀모형은 일단 계수를 추정해 모형의 예측변수 값이 주어진 결과변수를 예측하고자 모형을 사용할 수 있다.

→

선형회귀모형의 경우, 한 번 계수들을 추정하면 주어진 모형의 예측변수 값으로 결과변수를 예측하기 위해 모형을 사용하는 것이 가능하다.

[p.481 : 7행]

$\hat{\alpha}$ 과 $\hat{\beta}$ 은 서로 독립되지 않을 수 있으므로 위에서 도입한 일반적인 공식을 사용해 예측변수가 특정값 x 와 같을 때 예측값 \hat{Y} 의 아래의 분산을 얻는다.

→

$\hat{\alpha}$ 과 $\hat{\beta}$ 은 서로 독립되지 않을 수 있으므로 위에서 도입한 일반적인 공식을 사용하면 예측변수 X 가 특정값 x 와 같을 때 아래와 같은 예측값 \hat{Y} 의 분산을 얻는다.

[p.482 : 4행]

이 연구에서 영국에서 선거에 당선하는 것이 후보자의 자산에 미치는 평균효과를 추정했다.

→

이 연구에서는 영국에서 선거 당선 **이** 후보자의 자산에 미치는 평균효과를 추정했다.

[p.486 : 아래에서 6행]

실제로 선거에서 당선의 평균효과는

→

실제로 당선의 평균효과는

[p.498 : 아래에서 9행]

대부분 대표 샘플과는 거리가 멀다.

→

대부분 대표 표본과는 거리가 멀다.

[p.499 : 아래에서 10행]

연구자들은 지역 조사기관과 함께 지역 지도자들과 무장세력들과의 협상을 통해서만 농촌 마을에 접근할 수 있었다.

→

연구자들은 지역 **지도자들과 무장세력들과의 협상을 통해서만 지역 조사기관과 함께** 농촌 마을에 접근할 수 있었다.