

통계와 윤리

박은식

전남대학교 통계학과

연구 정직성 및 올바른 자료/결과 처리

- 연구의 정직성과 통계

연구의 정직성과 통계의 올바른 이용이 어떤 관련이 있는지 살펴본다.

- 올바른 자료/결과 처리

- 연구가 진행되는 주요 단계별로 나누어서, 각 단계에서 주의해야 할 통계의 올바른 이용법을 살펴본다.
- 통계가 연구의 과학성과 윤리적인 측면에 미치는 영향의 중요성을 공유하고 이해한다.

연구의 정직성과 통계 I

- 연구의 정직성과 통계와의 관련성
 - 통계의 올바른 이용은 연구의 과학성을 높이고 연구윤리를 지키기 위한 중요한 역할을 담당한다.
 - 통계의 올바른 이용에 관한 이해를 돕기 위하여 통계의 잘못된 이용에 대하여 살펴보자. 통계의 잘못된 이용은 다양한 원인이 있는데 통계 이론과 방법에 관한 지식의 부족, 통계적 응용과정에서 발생하는 순수한 실수, 심각한 부주의, 고의적 부정행위 등을 들 수 있다.
 - 특히 고의적 부정행위는 학술지 논문등재를 위한 통계적 기준을 만족하기 위하여 인위적으로 원자료(raw data)를 조작할 필요성에 기인하는 경우가 있다.

연구의 정직성과 통계 II

- 연구의 정직성과 통계와의 관련성
 - 통계의 잘못된 이용은 연구의 정직성, 객관성, 정확성, 공개 등 몇 가지 윤리적 의무를 위반하는 결과에 이를 수 있다. 통계의 잘못된 이용이 고의적으로 속인 결과라면 위법행위 (misconduct) 로 분류할 수 있다.
 - 존재하지 않는 자료 (data)나 결과를 허위로 만들어 낸다면 위조(fabrication)에 해당하며, 연구 자료나 결과를 인위적으로 조작, 변형, 삭제하는 행위는 변조 (falsification)에 해당한다.
 - 고의적인 속임수를 동반하지 않은 통계의 잘못된 이용은 허용되는 연구 관행을 벗어난 심각한 일탈 행위로 간주할 수 있다.

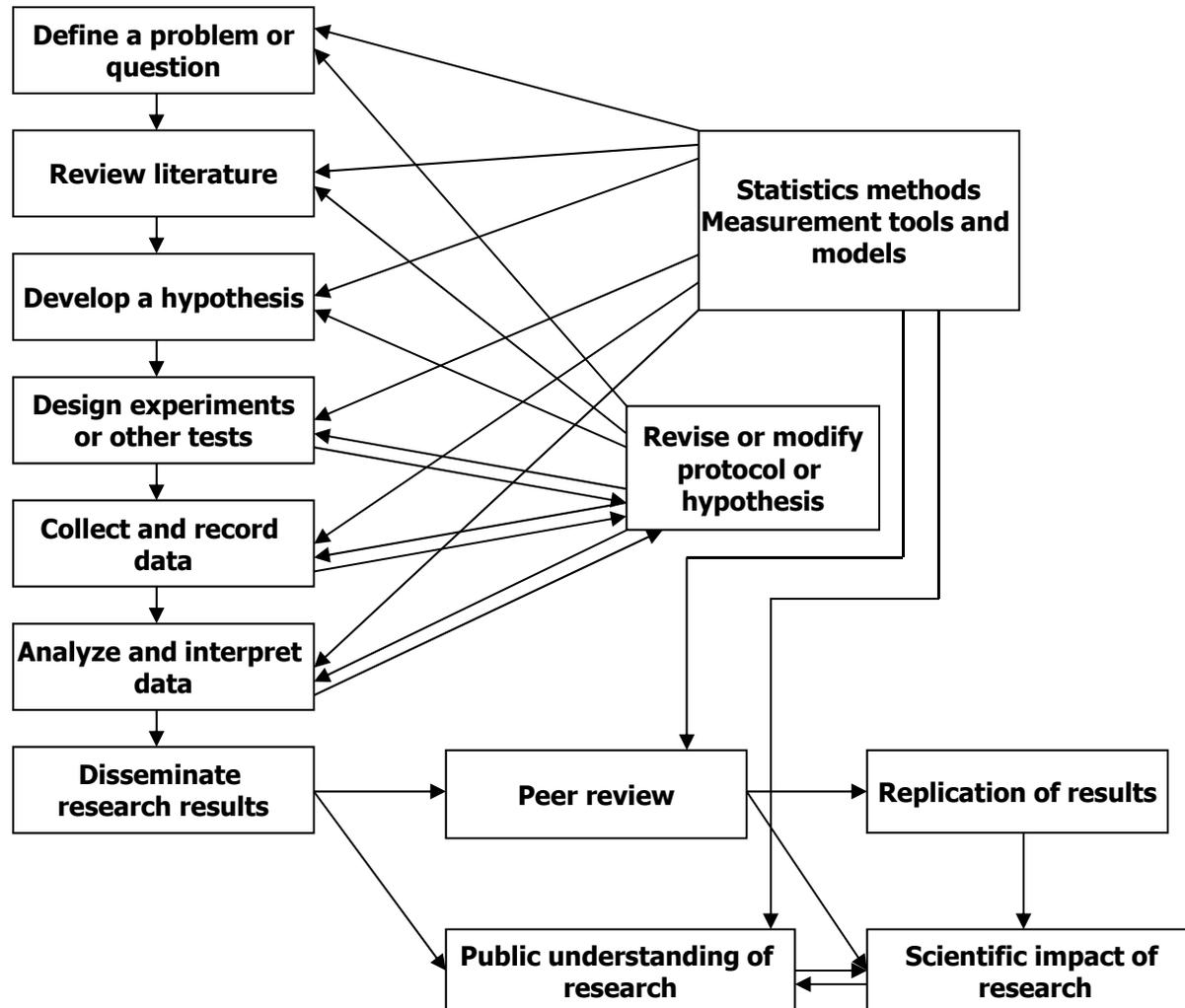
연구의 정직성과 통계 III

■ 연구의 정직성과 통계와의 관련성

- 통계의 잘못된 이용은 절대적인 것이 아니며, 연구 목적과 디자인에 적합하지 않은 통계 방법을 사용한 것을 말한다. 즉, 연구의 목적에 부합하도록 설계하여 자료를 얻고, 연구 디자인에 합당한 통계적 분석 방법을 선택하여 적용하여야 한다.
- 따라서 매우 다양한 연구 디자인과 분석 방법이 있지만, 본 강의에서는 널리 쓰이는 연구 디자인과 분석 방법을 대상으로 기본적인 통계의 올바른 이용법을 살펴보고, 통계의 이용이 연구 윤리와 어떤 연관이 있는지를 함께 살펴보고자 한다.
- 통계의 올바른 이용법은 연구가 진행되는 주요 단계별로 나누어 살펴보고자 하며, 개별 단계에서 주의해야 할 사항들을 완전하게 다루기보다는, 통계가 연구의 과학성과 윤리적인 측면에 미치는 영향의 중요성을 공유하는데 중점을 두고자 한다.

The Role of Statistics in the Scientific Research

Gardenier JS, Resnik DB (2002). The misuse of statistics: concepts, tools, and a research agenda. *Accountability in Research* 9, 65-74.



올바른 자료/결과 처리

연구의 정직성을 확립하고 과학성을 높이기 위하여 널리 쓰이는 통계적 방법의 올바른 이용법을 살펴보고자 한다

- 연구 설계의 중요성
- 표본크기 산정과 연구윤리와의 관련성
- 연구의 객관성 확보에 중요한 비교 그룹 배정 원칙
- 자료의 수집 및 관리와 연구결과의 신뢰도와의 연관성
- 이상치의 바람직한 자료/결과 처리 방법
- 표준편차와 표준오차
- 심프슨의 역설 (Simpson's Paradox)
- P-value와 구간추정의 바람직한 결과 표기법
- 통계적 방법과 분석 결과의 작성시 주의사항
- 분석결과의 해석

Design Issues I

- 연구의 설계는 연구 진행 단계에서 통계를 사용하는 첫 단계이다.
- 연구 자료의 질은 연구에 대한 사전 계획과 신중한 자료 수집과 관리 과정에 의해 정해진다.
- 그렇지만 사전 계획이 연구의 성공을 보장하는 것은 아니며, 다만 나쁜 디자인으로 인한 비뚤림(BIAS)이 발생하는 것을 방지하는 역할을 한다.
- 좋은 설계는 값비싼 인력과 재료의 낭비를 막고, 난해한 통계적 방법의 적용을 피하고, 해석상 논란이 될 수 있는 자료의 수집을 방지하고, 연구 자료는 정확하게 얻었으나 엉뚱한 연구가설에 대한 답을 제공하는 일을 예방한다.

Design Issues II

- 잘못된 통계적 디자인의 적용은 연구 진행 과정에서 되돌릴 수 없는 경우가 많기 때문에 시작 단계에서부터 연구의 주목적에 기초한 적절한 방법의 선택과 적용이 중요하다.
- 연구의 마지막 단계에서 주목적에 기술된 가설을 확인하기 위한 통계적 방법을 적용할 때, 연구의 설계가 잘못 되어 있어 주 가설을 검정하는 것이 불가능 하다는 것을 발견한다면, 그동안의 모든 노력이 수포로 돌아갈 수 있다.

Design Issues III

- 흔히 원 자료를 얻을 때 개개의 자료들의 측정값들이 서로 독립적으로 얻어진다는 가정에서 출발한다. 모든 자료들이 그런 것은 아니므로, 독립성의 가정을 만족하는 지 살펴보고, 만족되는 가정에 따라 적절한 통계적 디자인을 사용하여야 한다.
- 예를 들어 식물의 키와 씨 크기의 관계를 몇 단계의 분류 수준에서 구한 상관관계를 이용하여 알아보고자 한다.
- 씨 크기가 조상에 따라 어떻게 다른가를 알아보기 위하여 몇 가지 분류 수준 (taxonomic level) 에서 선택한 종들로부터 씨 크기들의 집단내 상관관계수(intraclass correlation, ρ)를 구하였다.
- 집단내 상관계수는 분류군내에서 종간 유사성을 측정하는 의미를 지니며, 집단의 크기가 작을수록 커진다.
- 유사종에 속하는 씨 크기들은 상관성 때문에 연구의 검정력(power)을 떨어뜨리는 결과를 초래할 수 있다.

Design Issues IV

- 집단내 상관계수가 커질수록, 표본크기, 즉 관측해야 할 씨 크기들의 개수가 커진다. 통계학에서 관측치들이 유사하다 (즉, 독립이 아님) 는 것은, 그들로부터 얻을 수 있는 새로운 정보의 양이 감소함을 의미한다.
- 동일한 양의 정보를 훨씬 작은 수의 독립적인 개체들로부터 얻어낼 수 있다면, 단위 개체당 소요되는 시간과 비용은 더 클 수 있으나, 전체적인 측면에서는 더 효율적일 것이다.
- 개체들이 서로 독립적이지 않은 경우 집단내 상관성을 고려하여 주가설을 검정하기에 충분한 검정력을 사전에 계산하여 충분한 수의 독립적인 개체를 연구에 포함하도록 한다.
- 만약 표본수가 작아서 검정력이 낮다면, 최종 분석 후 통계적으로 유의하지 않다는 결과가 나왔을 때, 그 원인이 표본수가 충분하지 못해서인지, 아니면 실제로 유의하지 않아서인지 구별할 수 없게 된다.
- 연구의 목적에 따라서 보다 효율적인 디자인을 선택하는 것이 중요하다.

Ethics and Sample Size Determination

- 표본크기는 연구 수행의 부담은 감소하고, 과학적 가치는 증가되도록 결정 한다.
- 표본크기가 너무 클 경우 과학적 의미가 없는 비교 효과의 크기도 통계적으로 유의한 결론을 내릴 가능성이 매우 높아지게 되므로 비윤리적이라고 할 수 있다.
- 또한 표본크기가 너무 작을 경우 연구의 과학적 가치가 다소 떨어질 우려가 있다는 점에서 비윤리적이라고 할 수 있다.
- 표본크기는 비교 연구시 연구의 검정력에 비례하고, 추정 연구시 추정값의 신뢰 구간 폭에 반비례한다.
- 연구를 설계하는 단계에서 표본크기를 사전에 계산하여야 하며, 결과보고서에 표본 크기를 산정할 때 이용한 방법과 이 방법을 적용하기 위해 만족되어야 할 가정을 함께 서술한다.

Statistical Errors in Hypothesis Testing

		Truth	
		Null Hypothesis [There is no Difference (A=B)]	Alternative Hypothesis [There is Difference (A≠B)]
Test Outcome	Significant Difference Detected (A ≠ B)	Type I Error (α)	Power (1 - β)
	No Significant Difference Detected (A = B)	No Error (1 - α)	Type II Error (β)

Study Power I

- Study power is computed given
 - The number of subjects (표본수)
 - Type I error (α , 제1종 오류의 크기)
 - Important effect size (비교효과 크기)
 - 과학적으로 의미가 있는 효과의 경험적 크기를 사용한다.
 - 이에 관한 정보가 없다면 기존의 연구에서 보고된 크기를 사용한다.
 - 실제적으로는 위의 검정력을 결정짓는 세 가지 값을 변화시켜 가면서 연구를 실제로 수행할 때 발생할 수 있는 다양한 시나리오를 고려하여 검정력을 결정한다.

Study Power II

- 검정력이 낮은 연구

- 검정력이 낮은 연구는 연구의 가치가 제한되어 있다. 하지만, 선행 연구가 드문 분야의 경우, 작은 규모의 연구가 체계적으로 문서화 되어 있다면, 전향적 메타 분석에 그 연구 결과를 포함하여 분석할 수 있으므로, 비록 검정력이 낮은 연구라 할지라도 의미가 있다.
- 또한 초기 단계의 연구로서, 임의 배정된 그룹간 비교에 관한 연구가 아니라 할지라도, 연구의 목적이 잘 정의되어 있고 검정력이 충분히 크다면 다음 단계의 비교 연구를 설계하는데 도움이 될 수 있으므로 연구의 가치가 있다고 하겠다.

비교그룹 배정 I

- 두 그룹 이상에 대하여 반응변수의 특성을 비교하고자 하는 연구에서 각 그룹에 배정하는 표본을 선택할 때, 주관성이 개입되어서는 안 된다.
- 연구 결과가 기대하는 방향으로 잘 나올 수 있도록 어떤 샘플을 특정한 그룹에 배정 하거나, 혹은 연구자가 어떤 샘플이 어떤 그룹에 배정되었는지를 알고 있다면 기대되는 결과를 만들기 위하여 반응값을 얻고 평가하는 과정에서 주관성이 개입될 수 있으며, 이는 연구의 객관성 확보에 저해되는 요인이다.
- 무작위 그룹 배정(group randomization)은 자료의 수집과정과 평가단계에서 나타나는 비뚤림(bias)을 방지할 수 있다.
- 또한 표본 중 어떤 샘플이 어떤 그룹에 배정되었는지에 관한 정보를 연구자 혹은 분석자가 모르게 하는 것을 맹검(blinding)이라고 하며 이는 연구의 객관성을 이루기 위한 중요한 조건이다. 이는 특히 사람을 대상으로 하는 임상시험과 같은 연구에서 중요한 원칙이다.

비교그룹 배정 II

■ 단순 무작위 배정 예

- 40개의 표본을 두 그룹에 배정하기로 한다면, 난수표나 난수 발생 컴퓨터 프로그램을 이용하여 0-4의 수가 나오면 A군에, 5-9의 수가 나오면 B군에 배정하는 방법을 취한다.
- 이 방법은 표본 수가 커지면 각 그룹에 배정되는 수가 비슷하게 되지만 그렇지 않을 경우에는 표본수의 불균형이 우려된다.

■ 블록 무작위 배정 예

- 단순무작위 배정을 보완하기 위한 방법으로 이용한다.
- 크기가 4인 블록을 이용한다면, 40명을 배정하기 위하여 모두 10개의 블록을 무작위로 선택하면 된다.
- 이때 각 블록은 임의의 순서로 배정된 두개의 A와 두개의 B로 구성되어 있다. 즉, AABB, ABAB, ABBA, BBAA, BABA, BAAB 와 같은 6개의 블록 중에서 모두 10개의 블록을 무작위로 선택하고, 선택한 순서대로 그룹 배정이 이루어진다.

자료의 수집 및 관리

- 아무리 좋은 통계적 방법을 적용한다고 해도 원자료의 수집 및 관리가 잘못되었다면 그 문제점을 극복하는 방법에는 한계가 있을 수 밖에 없다.
- 또한 잘못 얻은 자료를 확인하지 못한 채 분석하면 이상치의 존재로 말미암아 연구의 결론을 엉뚱한 쪽으로 유도할 수도 있으므로 자료의 질은 연구결과의 신뢰성과 밀접한 관련이 있다.
- 자료의 수집 및 기록은 매우 단순하고 시간이 많이 소요되는 지루한 작업이지만, 잘못 했을 경우 분석에 미치는 영향은 지대하다.
- 자료의 관리는 자료의 정확한 입력, 참가자 확인 정보의 기록 누락 여부, 무응답치(Missing Value)나 이상치 여부, 서로 관련된 값들 간의 논리적 일치성 등을 확인하고, 잘못된 부분이 있으면 수정하는 절차를 포함한다.
- 자료의 입력 및 확인 절차는 자료가 얻어진 당시 혹은 가능한 한 가까운 시점에 하는 것이 자료에 문제가 있을 경우 발견하기도 쉽고, 수정하기도 쉽다.

How to Handle Outliers

- 이상치는 다른 관측치에 비하여 훨씬 크거나 훨씬 작은 값을 갖는 관측치를 말하며, 분석 결과에 지대하게 영향을 미치기 때문에 중요하다.
- 이상치가 존재한다면 실제로 정확하게 측정되어 나온 값인지, 자료 입력을 데이터베이스에 잘못하여 생긴 것인지를 확인하여야 한다.
- 이상치를 포함한 분석은 순위에 기초한 것과 같은 로버스트(Robust)한 통계분석 방법을 적용하여야 한다. 예를 들어 평균과 표준편차를 이용하여 자료를 요약하고자 하였다면, 이상치가 있는 경우에는 평균대신에 중앙값(Median)을 이용하고, 표준편차를 4분위수범위(Inter Quartile Range)로 대체하여 사용하는 것이 좋다.
- 이상치의 존재여부가 연구결론의 방향을 바꾸는지를 확인하기 위하여 이상치를 모형에 포함한 경우와 그렇지 않은 경우의 결론을 비교하여야 한다. 만약 서로 다른 결론이 내려진다면, 그 원인을 분석하여 명시하여야 한다.

표준편차와 표준오차 I

- 표준편차는 관측된 원자료의 변동의 크기를 측정하는 반면에, 표준오차는 추정량(예: 평균, 두 평균의 차이, 회귀 계수 등)의 정확도를 측정한다.
- 표준편차는 표본을 추출한 모집단의 변동의 크기를 추정하기 위하여 정규분포뿐만 아니라 모든 분포에 사용한다. 95%의 관측치가 2표준편차 범위에 속한다고 한다면, 나머지 5%는 이 범위 밖으로 분포의 대칭성 정도에 따라 절반씩 양끝으로 흩어져 분포하거나 치우친 한 쪽으로 흩어져 분포한다.
- 표본평균을 구할 때는 이 특정한 값 자체에 관심이 있는 것이 아니라, 이와 같은 표본평균을 동일한 모집단에서 추출한 다른 표본들을 이용하여 구할 때, 그 값들의 경향에 관심이 있다. 이 표본평균들은 모두 다른 값을 가지겠지만 그 값들이 변동하는 정도를 추정할 수 있다면, 표본평균을 이용하여 모집단의 평균을 추정하는데 이용할 수 있다. 표준오차는 표본평균의 정확도를 측정하기 위한 변동의 크기를 나타내는 것으로 이해할 수 있다.

표준편차와 표준오차 II

- 표준오차(SE)와 표준편차(SD)는 $SE=SD/\sqrt{(\text{표본크기})}$ 의 관계를 가지므로, 표준오차는 표본크기가 커질수록 감소하며, 표준편차는 표본크기의 변화에 상관없는 특성을 가지고 있다.
- 어떤 연구자들은 표본크기를 매우 크게 하여 표준오차를 감소시킴으로서 통계적으로 매우 유의한 결과에 이르도록 하는데 이는 연구의 윤리성에 어긋나는 행위이다. 그러므로 통계적으로 유의한 차이를 얻었을 때는, 그 차이가 해당 연구 분야에서 의미가 있는 차이인지를 살펴보는 것이 중요하다.
- 표준오차는 신뢰구간을 구하는 데 가장 유용하게 사용된다. 표본크기가 큰 경우, 95% 신뢰구간은 평균을 중심으로 양끝으로 $1.96 \times SE$ 만큼 떨어진 곳에 구간의 상한과 하한이 주어진다.

표준편차와 표준오차 III

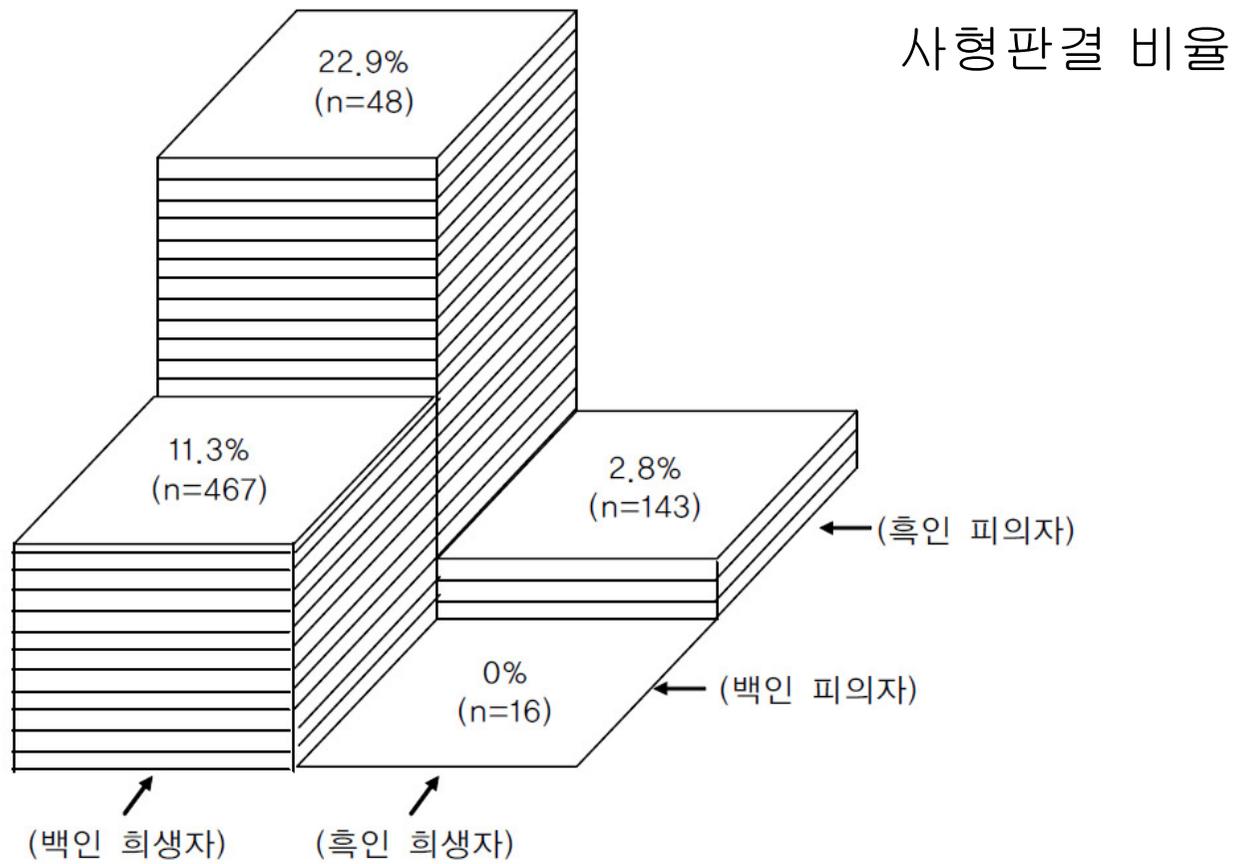
- 많은 연구에서 \pm 부호를 이용하여 표본평균을 표준편차나 표준오차와 연계하여 나타내고 있다. 예를 들어, 69.4 ± 9.3 kg이라고 나타낸다면, 9.3이 표준편차인지 표준오차인지 명시하지 않는다면 구별할 수 없을 것이다.
- Altman에 따르면 2002년도에 등재된 88편의 논문을 조사한 결과 14%의 논문이 표준 편차와 오차 여부를 명시하지 않았다고 한다. British Medical Journal과 많은 다른 학술지들에서는 표준편차와 표준오차의 인용여부를 확실히 하고 \pm 부호를 이용한 표기법을 이용하지 않도록 권고하고 있다.

Simpson's Paradox

- 자료 : 인종 특성에 따라 살인피의자의 사형판결 여부
- 조사대상 : 1976 ~ 1987년에 플로리다에서 살인죄로 기소된 674명의 피의자.
- 목적 : 희생자의 인종별로 피의자 인종에 따른 사형판결이 달라지는가?
- 결과 : 흑인 피의자가 사형판결을 받은 비율은 희생자의 인종에 관계없이 백인 피의자보다 더 높음을 알 수 있다.
- 출처 : 범주형자료분석개론(정광모등.)

희생자 인종(Z)	피의자 인종(X)	사형판결(Y)		예 비율
		예	아니오	
백인	백인	53	414	11.3
	흑인	11	37	22.9
흑인	백인	0	16	0.0
	흑인	4	139	2.8
총계	백인	53	430	11.0
	흑인	15	176	7.9

Simpson's Paradox



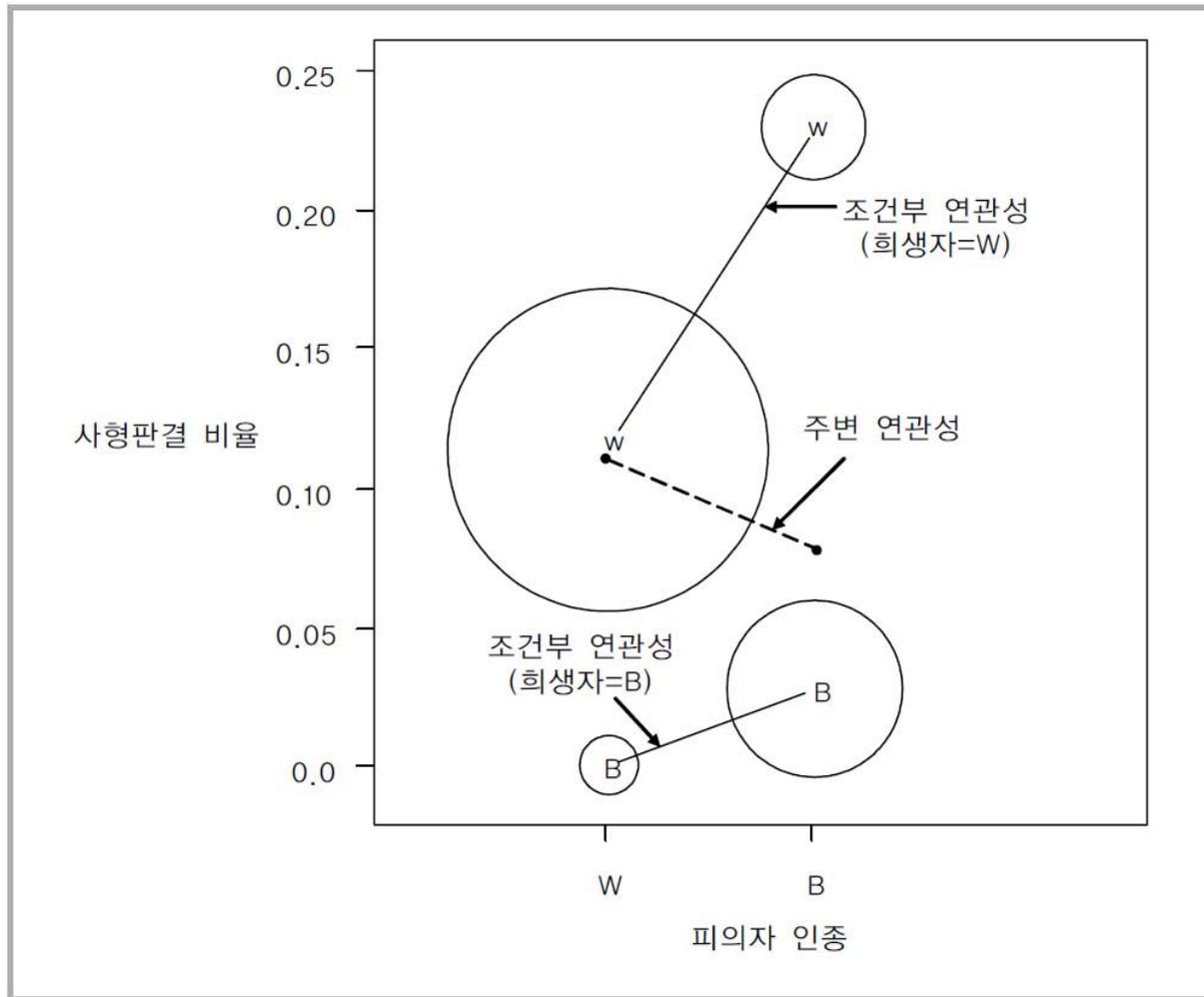
Simpson's Paradox

희생자 인종(Z)	피의자 인종(X)	사형판결(Y)		예 비율
		예	아니오	
백인	백인	53	414	11.3
	흑인	11	37	22.9
흑인	백인	0	16	0.0
	흑인	4	139	2.8
총계	백인	53	430	11.0
	흑인	15	176	7.9

피의자 인종(X)	희생자 인종 (Z)		백인(Z) 비율
	백인	흑인	
백인	467	16	0.97
흑인	48	143	0.25
총계	515	159	674

희생자 인종(Z)	사형판결(Y)		예 비율
	예	아니오	
백인	64	451	0.12
흑인	4	155	0.03
총계	68	606	674

Simpson's Paradox



〈그림 3.2〉 희생자 인종의 조건에서 피의자 인종별 사형판결 비율

From P-values To Interval Estimation I (Getting Past the Statistical Referee)

- 최근에는 학술지 논문의 동료심사(peer review) 과정에서 통계학자들의 참여가 증가함에 따라 통계학자와 학술지 편집자들 간의 교류가 활발해지고 있다. Sorenson et al에 따르면, Health Education Research 학술지에 투고한 논문의 2/3가 자료 분석의 오류 때문에 기각되었다고 한다.
- 이를 극복하기 위한 한 방법으로 P-value를 $P > 0.05$, $P < 0.01$ 과 같이 적지 말고 있는 그대로 기록하는 것이 좋다.
- 예를 들어 P-value가 0.06이라면, 자료의 검정력이 다소 낮거나 혹은 실제로 통계적으로 유의하진 않지만 해당 연구 분야에서 의미가 있는 차이가 있을 수 있지만, P-value가 0.80 이라면 사실상 의미가 없는 차이일 것이다. 따라서, $P > 0.05$ 라고 기록한다면, 그 값이 어느 정도 크기인지 알 수 없어 매우 제한된 정보만을 제시하는 셈이다.
- P-value가 통계적으로 유의할 경우와 그렇지 않은 경우 모두 그에 따른 설명이 따라야 한다.

From P-values To Interval Estimation II

- P-value는 직관적으로 이해하기 쉽지 않은 점이 있으므로 P-value만 단독으로 제시하는 것 보다는 보다 많은 정보를 제공해 주는 신뢰구간(confidence interval)을 사용하는 것이 바람직하다.
- 예를 들어 두 그룹의 비교에서 그룹간 차이가 있다고 할 때, P-value로는 어떤 그룹이 큰 값을 가지는지 알 수 없지만, 신뢰구간은 그 크기와 정확도까지를 보여준다.
- 신뢰구간은 표본에서 얻은 통계량(예: 평균, 상관계수, 오즈비)의 정확도를 나타내는 방법으로, 신뢰구간의 폭이 좁을수록 표본통계량의 정확도가 높다.
- P-value와 신뢰구간은 서로 연결되어 있는데, 예를 들어, 두 그룹 평균비교 연구에서, 처리법 효과에 따른 차이가 유의수준 5%에서 유의하다면 차이에 대한 95% 신뢰구간은 구간 내에 영을 포함하지 않는다.

분석결과의 작성 I

- 분석결과의 작성은 그 방법에 따라 연구자들의 결과 해석에 영향을 준다는 점에서 중요하다. 평균을 이용한 분석결과를 제시할 때는 변동의 크기도 함께 제시하는 것이 분석의 정확도를 표현하는 중요한 요소이나 간과되거나, 혹은 표준편차와 표준오차의 개념이 혼동되어 사용되는 경우가 있다.

분석결과의 작성 II

- 표준편차는 다양한 형태로 제시되는데 흔히 ‘평균은 ... 30 mg 이다.(SD±4.6 mg)’ 처럼 평균과 함께 괄호안에 ± 부호와 함께 제시된다. 표준편차는 늘 양의 값을 갖는데, 이 표현은 마치 표준편차가 ±4.6 인 것으로 오해될 소지가 있다.
- 이 표현은 평균을 중심으로 양쪽으로 1 SD 만큼의 폭을 갖는 범위가 (25.4 mg, 34.6 mg) 처럼 주어진다라는 의미를 갖고 있다. 이 범위는 대략적으로 68%의 관측값이 속하는 구간을 의미한다.
- 만약 95%의 관측값이 속하는 구간의 크기를 구하는 것이 관심이라면 ‘평균 ± 2 SD’와 같은 형식으로 나타내는 것이 바람직하다. 이제 구간은 (20.8 mg, 39.2 mg)과 같이 주어지고, 구간폭은 이전의 구간과 비교하여 2 배가 되었다.
- 이는 실질적으로 널리 쓰이는 95% 신뢰구간과 일치하므로 ‘평균 ± 2 SD’ 가 ‘평균 ± 1 SD’ 보다 권장되는 표현이다. 이와 같은 구간은 관측값들이 정규분포에 따른다는 가정이 필요하다.

분석결과와의 작성 III

- 정규분포 가정이 만족되지 못할 경우엔, 앞에서 설명한 바와 같이 변동의 크기를 분산대신에 4분위수 범위를 이용하여 나타내는 것이 바람직하다. 같은 표현방법을 표준오차의 경우에도 적용한다.
- 평균을 제시하는 방법도 분포의 모양을 고려하여 선택한다. 좌우대칭인 분포이면 전체 관측치의 합을 관측치의 개수로 나눈 산술평균을 이용하고, 어느 한쪽으로 치우친 모양이면 중앙값을 이용하는 것이 바람직하며, 관측값이 변화하는 양상이 곱의 형태이면 기하평균(Geometric Mean)을 이용하는 것이 좋다.
- 논문의 분석 방법을 기술하는 절에서는 결과를 보고하는 절에 실린 양적 정보를 얻기 위해 사용한 통계적 방법에 관하여 상세히 기록하는 것이 좋다. 결과를 보고할 때는 통계적 용어보다는 가능하면 보다 단순한 표현을 사용한다.

분석결과의 작성 IV (예제)

Data From a Hypothetical Randomized Controlled Trial of Treatment A Compared With Standard Care

Outcome	No. (%) of Patients		Risk Ratio (95% Confidence Interval)
	Treatment A (N = 800)	Standard Care (N = 1200)	
Death	18 (2.3)	54 (4.5)	0.50 (0.30 to 0.85)
Hospitalization	100 (12.5)	230 (19.2)	0.65 (0.52 to 0.81)
Treatment complications*			
None	563 (75.1)	615 (55.0)	Reference
1	113 (15.1)	280 (25.0)	0.53 (0.44 to 0.65)
2 or more	74 (9.9)	223 (19.9)	0.44 (0.34 to 0.56)

*Data on complications were missing for 6.3% of the treatment A patients and 6.8% of the standard care patients.

분석결과의 해석 I

(P-value)

- 올바른 통계적 방법으로 분석한 경우라도, 결과를 정확하게 해석하지 않는다면 잘못된 연구결론을 내릴 수 있고 뒤따르는 연구들을 잘못된 방향으로 이끌 수 있다.
- 가설검정 후 분석 결과를 해석하기 위하여 유의수준을 이용하는데 일반적으로 1%, 5%, 10%를 널리 사용한다. 유의수준 5%에서 P-value의 값이 0.04 라면, 증명하고자 하는 대립가설이 통계적으로 유의하다는 결론을 내릴 수 있다. 만약 P-value의 값이 0.06 이라면, 이는 대립가설이 통계적으로 유의하지 않다는 반대 결론을 의미하는 것보다는 P-value의 값이 0.04인 경우와 같은 방향으로 해석하는 것이 바람직하다.

분석결과의 해석 II

(통계적 유의성)

- 통계적으로 유의하지 않은 결과는 귀무가설을 기각하기에는 설득력이 부족하다는 것을 의미할 뿐이다. 그러므로 유의하지 않은 결과를 귀무가설에 근거한 결론은 ‘중요하지 않다’ 혹은 ‘존재하지 않는다’를 의미하는 것으로 해석하는 것은 옳지 않다. 모든 유의하지 않은 결과는 ‘그 연구의 중요성이 전혀 없다.’는 것과 동일시하는 것은 분명히 잘못된 생각이다.
- 반대로, 연구결과에 대한 충분한 검토를 거치지 않고, 유의하게 관측된 차이를 실제 존재하는 것으로 결론을 내리는 것도 매우 위험한 생각이다. 그러므로, 관측된 차이에 관한 신뢰구간을 함께 제시하는 것이 유의하지 않은 결과를 보고할 때 특히 더 중요하다.

분석결과의 해석 III

(표본크기와 검정력)

- 표본크기가 작아서 검정력이 낮은 연구는 통계적으로 유의하지 않지만, 과학적으로 중요한 효과를 보여줄 수 있다.
- 이 경우에는 효과가 없다는 결론을 내리기 보다는, 가능하다면 추가적인 자료를 얻은 후 다시 분석하는 것이 바람직하다.
- 연구의 결론이 통계적으로 유의하지 않은 결과를 보여줄 때는 수집한 연구 자료에 기반한 표본 크기를 다시 산정하여 검정력의 크기를 확인한 후 결과를 해석하는 것이 좋다.

분석결과의 해석 IV

(연관성과 인과관계)

- 흔히 해석의 오류를 범할 수 있는 예로 연관성 (association) 과 인과관계(causal relationship)를 들 수 있다.
- 제 3의 감추어진 변수가 어떤 두 변수와 각각 연관되어 있다면, 위의 두 변수간에 연관성이 없음에도 불구하고 있는 것처럼 오해할 수 있다.
- 예를 들어, 시간과 연관이 있는 이혼율과 휘발유가격은 연관성이 있는 것처럼 보이지만, 사실은 시간과 각각 연관되어 있어서 오해한 잘못된 연관성 예이다.
- 두 변수가 연관성이 있다고 해서 모두 인과관계를 의미하는 것은 아니므로 두 변수의 관계를 신중하게 해석하여야 한다.
- 인과관계의 추론은 현재 수행하고 있는 연구만으로 할 수 있는 것이 아니며, 기존의 연구 자료와 사전지식을 이용하여 가능하다.

결론 I

■ 연구의 정직성과 통계

- 최근에는 통계 비전문가들도 사용자 편의성을 고려하여 설계된 통계 패키지를 이용한 분석이 용이해짐에 따라 통계의 접근성이 개선되고 있는 추세이다.
- 반면, 사용자들의 통계적 기본지식이 충분하지 못하여, 분석과정에서 실수로 잘못된 결과를 보고할 수 있는 가능성은 높아졌다. 통계적 방법을 적용할 때는 반드시 그 방법을 적용하기 위한 가정이 만족되는지 확인한 후 사용하여야 하지만, 통계 패키지는 이러한 가정을 확인해 주는 절차는 제공하지 않기 때문이다.
- 연구자들이나 학생들이 통계 교육을 통하여 통계의 개념과 사용 방법을 이해하게 된다면 이러한 문제들은 줄어들 것이다.
- 또한 통계 원리에 대해서 잘 모른다는 사실이 부적합한 디자인이나 부적절한 분석의 면죄부가 되어서는 안 될 것이다.

결론 II

■ 올바른 자료/결과 처리

- 통계의 올바른 사용법은 연구의 과학성과 윤리성을 높이기 위한 최소한의 지침이다. 통계적 방법을 올바르게 이용하면 연구자들이 관심의 대상이 되는 연구결과에 영향을 미치는 원인을 찾는 데 강력한 수단이 될 것이다.
- 여기서 다루지 않은 다른 고급 통계적 기법들도 연구의 목적과 자료의 성격에 합당하도록, 신중하게 가정을 확인한 후 적용하여야 한다.

References

- **Design Issues:** Statistical and design issues in studies of groups: accounting for within group correlation, P Cummings and T D Koepsell, *Inj. Prev.*, 2002, 8, 6–7
- **Ethics and Sample Size Determination:** Ethics and Sample Size, Peter Bacchetti, Leslie E. Wolf, Mark R. Segal, and Charles E. McCulloch, *Am J Epidemiol* 2005;161:105–110
- **Study Power:**
 - The continuing unethical conduct of underpowered clinical trials, Scott D. Halpern, Jason H. T. Karlawish, Jesse A. Berlin., *JAMA* 2002, 288, 358–362.
 - Statistical power analysis for the behavioural sciences, Cohen J., 1988 Hillsdale, NJ: Lawrence Erlbaum
- **Simpson's Paradox:** Introduction to Categorical Data Analysis, Alan Agresti
- **From P-values to Interval Estimation:** Getting past the statistical referee: moving away from the P-values and towards interval estimation, Oxford University Press, AS Rigby, 1999, 713–715
- **Reporting Statistical Information:** Reporting statistical information in medical journal articles, P Cummings, *ARCH PEDIATR ADOLESC MED*, 2003, Vol. 157, 321–324