

데이터로 배우는 통계학

자연과학대학 통계학과
장원철 교수

신뢰구간과 가설검정

1. 순열검정과 P값

팔짱을 낄 때 어느 팔이 위로 올라오는가?

- 팔짱을 낄 때 어느 팔이 위로 올라오는가? 영국의 통계학자 데이비드 스피겔헬터는 이 질문에 54명의 대학원생에게 물어본 결과 다음과 같은 대답을 얻었다.

	여성	남성	총합
왼팔이 위로	5	17	22
오른팔이 위로	9	23	32
총합	14	40	54

[성별 팔짱 끼기 방식 선호도]
(The Art of Statistics, p260)

팔짱을 낄 때 어느 팔이 위로 올라오는가?

- 앞서서와 같이 변수(성별/팔)의 값에 따라 거기에 해당하는 관측치의 개수를 표시하는 유형의 표를 분할표(contingency table)라고 한다.
- 결과를 살펴보면 과반수 이상($32/54=59\%$)이 오른팔을 위에 놓았는데 오른팔을 위에 놓은 비율은 남성($23/40=57\%$)보다 여성($9/14=64\%$)이 더 높았다.

팔짱을 낄 때 어느 팔이 위로 올라오는가?

- 여기서 우리가 궁금한 것은 성별에 따라 팔짱 끼기의 패턴이 다른지 유무이다. 구체적으로 여기서 귀무가설은 “성별에 따라 팔짱 끼는 방식에 차이가 없다.”이다.
- 정리된 결과로 여성이 남성보다 오른팔을 위에 올리는 비율은 7% 높지만, 과연 이 차이가 우연히 변동성에 의해서 생긴 것인지 혹은 실제로 여성이 오른팔을 올리는 걸 선호하는지 알아내야 한다.

팔짱을 낄 때 어느 팔이 위로 올라오는가?

- 이 질문에 답변하기 위해서 법정 시스템을 다시 생각해보자.
- 검사의 입장에서 피고가 무죄라면 이렇게 수많은 증거들을 수집하기 힘들다는 것을 강조할 것이다.
- 가설검정에서는 귀무가설이 참이라면 이런 데이터를 관측하기 힘들다는 것을 보이면 된다.
- 그렇다면 남녀별로 팔짱 끼기 차이가 없다는 가정하에서 어떻게 7% 차이가 관측하기 힘든지 여부를 판단할 수 있을까? 이 문제에 대한 답은 순열검정(permutation test)를 이용하면 알 수 있다.

순열검정 (Permutation Test)

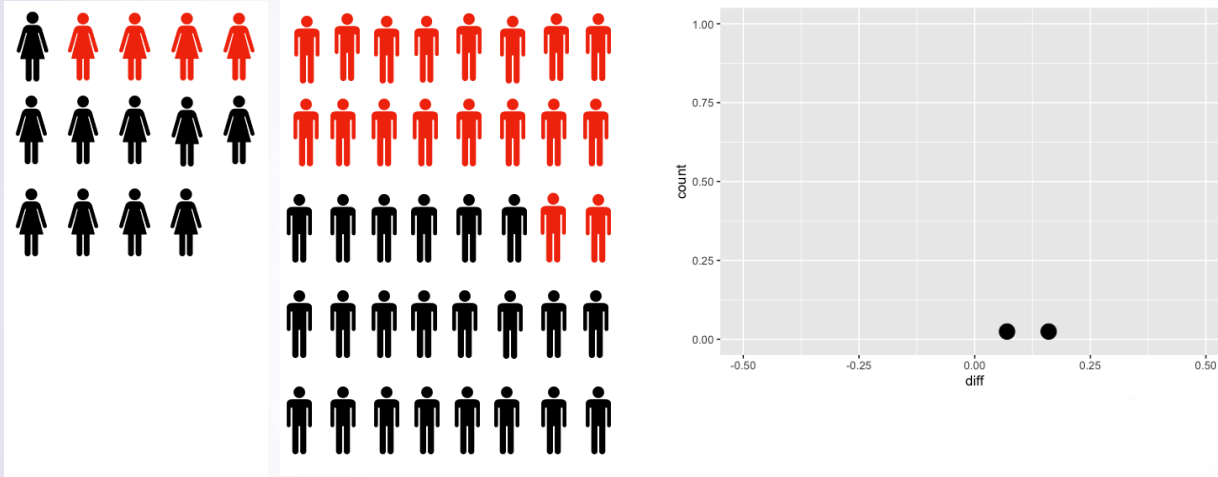


- 남녀별 오른손을 위에 놓는 비율의 차

$$\frac{9}{14} - \frac{23}{40} = 0.07$$

- 왼쪽 그림에서 빨간색은 왼쪽을 위에 올린 경우를, 검은색은 오른손을 위에 올린 경우를 의미한다.
- 만약 성별 여부와 팔의 선호도와 관련이 없다면 22개의 빨간색과 32개의 검은색이 성별에 관계없이 아무렇게나 섞일 수 있고 그 경우 검은색의 비율의 차이가 얼마나 되는지 알아보자.

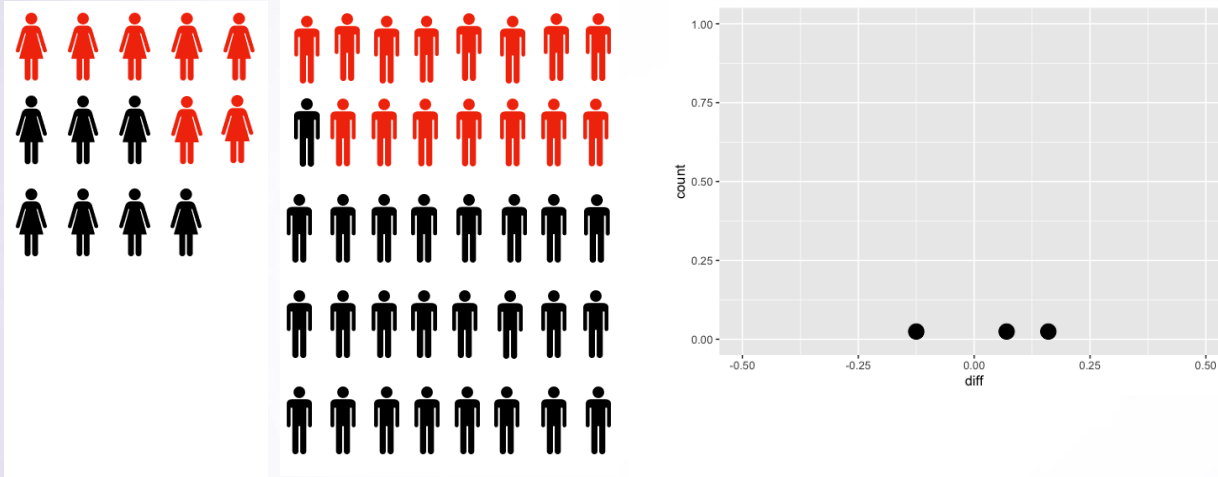
순열검정 (Permutation Test)



○ 남녀별 오른손을 위에 놓는 비율의 차

$$\frac{10}{14} - \frac{22}{40} = 0.16$$

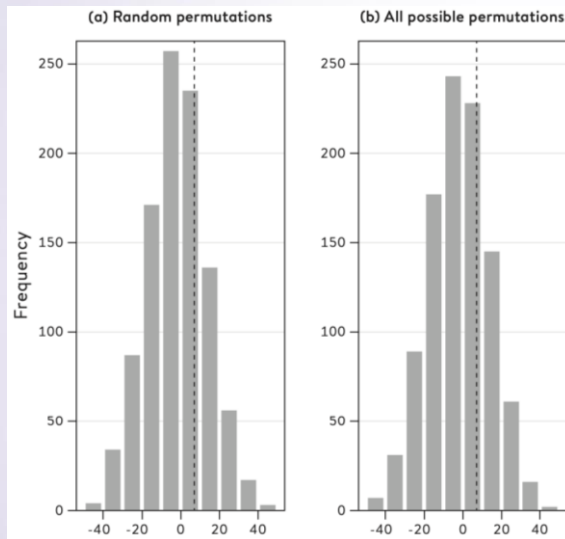
순열검정 (Permutation Test)



○ 남녀별 오른손을 위에 놓는 비율의 차

$$\frac{7}{14} - \frac{25}{40} = -0.125$$

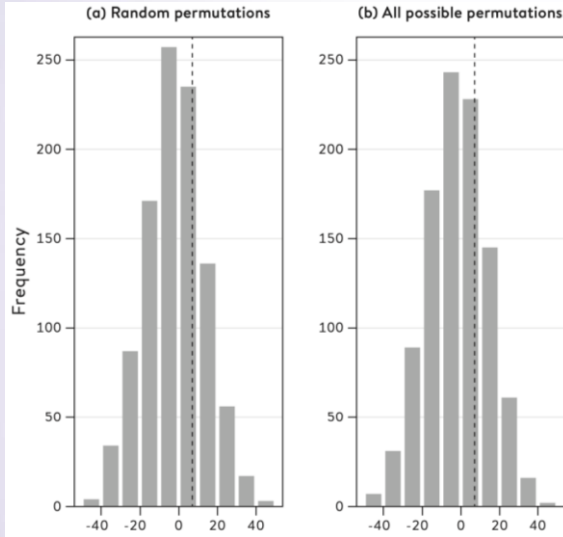
순열검정 (Permutation Test)



[남녀간 팔짱 낄 때 오른손이 위에 오는 비율의 차이]
(The Art of Statistics, p263)

- 앞에서와 같은 실험을 1,000번 반복한 후 비율의 차이를 히스토그램으로 그린 결과가 왼쪽 그림 (a)이다. 가운데 점선이 실제 설문 결과였던 7%를 나타낸다.
- 즉 이 분포는 **비율의 차이**에 관한 **표본분포**이다.

순열검정 (Permutation Test)



- 왼쪽 그림 (b)는 이 비율의 차이의 확률분포가 초기하 분포 (hypergeometric distribution)이라는 점을 이용해 그린 그림이다.

[남녀간 팔짱 낄 때 오른손이 위에 오는 비율의 차이]
(The Art of Statistics, p263)

P-value와 유의수준

- P-value는 귀무가설이 참이라는 전제하에 우리가 관측한 검정통계량의 값이나 혹은 그보다 더 극단적인 값을 얻을 확률을 의미한다.
- 팔짱 끼기 예제에서는 남녀간 오른팔이 위에 있는 비율의 차이가 0.07 이상인 경우에 해당하는 확률이 p-value에 해당한다.
- 이 경우 그림에서 점선 오른쪽 꼬리 부분의 넓이를 계산하면 p-value가 0.45임을 알 수 있다.
- P-value가 주어진 기준값(유의수준)보다 작을 경우 검정통계량의 값이 극단적이라고 이야기한다.

대립가설 (단측검정과 양측검정)

- 대립가설은 우리가 증명하고자 하는 가설을 의미한다.
- 팔짱 끼기 방식에서 대립가설이 “성별로 팔짱 끼는 방식이 다르다.”라고 한다면 검정통계량의 값은 사실 음수도 나올 수 있기 때문에 앞의 p-value 계산 시 검정통계량의 값이 -0.07 보다 작을 확률도 계산해야한다. 이러한 검정방식을 양측검정이라고 하고 이 경우 p-value는 0.90 이 된다.

대립가설 (단측검정과 양측검정)

- 하지만 대립가설이 “여성이 남성보다 팔짱을 낄 때 오른팔을 올려놓는 것을 선호한다.”라면 p -value는 0.45가 된다. 이러한 검정방식을 단측검정이라고 한다.
- 하지만 이 예제에서 여성이 특정 팔을 선호할 과학적인 근거는 전혀 없다. 즉 가설검정에서는 특별한 경우가 아니라면 양측검정을 사용하는 것을 원칙으로 한다.

양측검정과 p-value

- 단측검정을 사용하면 양측검정을 사용할 때 보다 p-value가 절반밖에 되지 않는다는 것을 알 수 있다.
- 그래서 연구자들은 단측검정을 사용하는 것을 선호한다.
- 하지만 단측검정은 과학적인 근거나 선행연구 결과가 있지 않는 한 사용하지 않는 것이 맞다.
- 신약개발 시 제약회사는 항상 신약이 병을 낮게 할 것이라고 믿지만 반대로 결과가 나올 수 있다고 생각해서 미국 FDA에서는 항상 양측검정으로 신약시판 승인을 한다.

가설검정의 절차

1. 가설설정

- 귀무가설: 현 상태에 대한 잠정적 가정
- 대립가설: 우리가 알고 싶은 것

2. 검정통계량: 법정 시스템에서 증거와 같은 역할. 검정통계량의 값이 극단적이라면 귀무가설이 사실인지 여부에 의문을 표시하게 된다.

3. 검정통계량의 표본분포: 귀무가설이 참이라는 가정하에 검정통계량의 분포로 이를 통해서 검정통계량의 값이 극단인지 여부를 판별할 수 있다.

가설검정의 절차

4. P-value 계산

- 1) 양측검정인지 단측검정인지 여부에 따라 검정통계량의 분포를 이용하여 P-value를 계산한다.
- 2) P-value가 주어진 유의수준 (보통 0.05)보다 작은 경우 귀무가설하에서 이러한 결과를 얻을 가능성이 극단적(즉 5% 이하)이라고 생각하고 귀무가설을 기각한다.

5. 결론: 원래 질문의 문맥에 맞게 결론을 제시한다.



오늘의 강의 요점

- 순열검정
- P-value
- 단측검정과 양측검정



○ 출처

#1~2 D. Spiegelhater, (2019), The Art of Statistics, Penguin Random House

신뢰구간과 가설검정

2. 카이제곱검정과 t-검정

검정통계량

- 가설검정에서 일반적으로 검정통계량은 다음과 같은 형태를 가진다.
 - $(\text{관측치} - \text{귀무가설하에서 기댓값}) / \text{표준오차}$
 - $[(\text{관측치} - \text{귀무가설하에서 기댓값}) / \text{표준오차}]^2$
- 귀무가설하에서 이런 검정통계량의 분포는 표준정규분포, 카이제곱분포, t-분포를 따르며 카이제곱분포와 t-분포의 경우 **자유도**를 모수로 가진다.
- 자유도가 커지면 두 분포를 정규분포의 모습과 비슷해진다.

팔짱 끼는 방식이 성별로 차이가 있는가?

- 팔짱 끼는 방식에 관한 문제에서 순열검정을 이용해서 가설검정을 실시했다.
- 보다 일반적인 방식은 카이제곱 검정을 이용한 독립성검정이다.
- 우선 귀무가설하에서 분할표의 각 칸에서의 기댓값을 계산해보자.
 - 예를 들면 귀무가설하에서 여성 중 왼팔이 위로 가는 사람의 기대비율은 (전체 여성비율) \times (전체 인원중 왼팔이 위로 가는 사람의 비율)로 생각할 수 있다.
 - 여기서 이 칸의 기댓횟수는 위의 비율에 전체 인원을 곱한 것이다. 즉 $(14/54) \times (22/54) \times 54 = 5.7$

팔짱 끼는 방식이 성별로 차이가 있는가?

- 팔짱을 낄 때 어느 팔이 위로 올라오는가? 영국의 통계학자 데이비드 스피겔헬터는 이 질문에 54명의 대학원생에게 물어본 결과 다음과 같은 대답을 얻었다.

	여성	남성	총합
왼팔이 위로	5(5.7)	17(6.3)	22
오른팔이 위로	9(8.3)	23(23.7)	32
총합	14	40	54

[성별 팔짱끼기 방식 선호도의 실제 관측값과 기댓값]
(The Art of Statistics, p269)

카이제곱검정

- 카 이 제 곱 검 정 에서 검 정 통 계 량 은 각 칸 에서 $(\text{관측값} - \text{기댓값})^2 / \text{관측값}$ 을 계산하고 이 값을 모두 합한 것으로 정의된다.
- P-value를 구하기 위해서 귀무가설하에서 검정통계량의 분포를 알아야 하는데 이 분포는 카이제곱분포를 따른다.

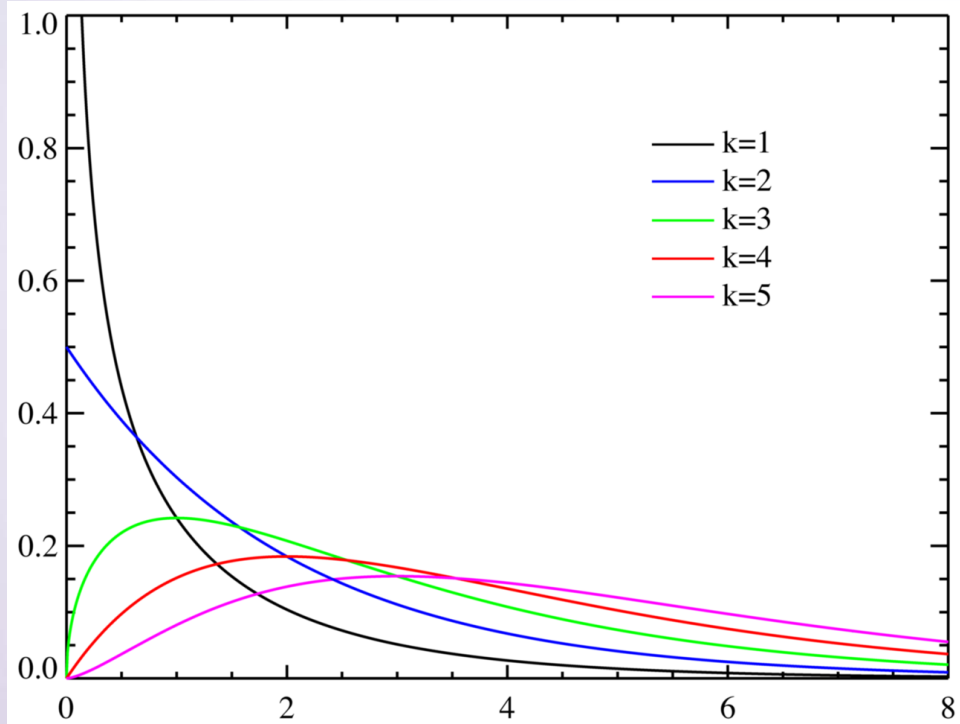
카이제곱분포

○ 카이제곱분포는 다음과 같은 성질을 가진다.

- 정의: 표준정규분포를 따르는 확률변수의 제곱이 자유도가 1인 카이제곱분포를 따른다.
- 서로 독립인 카이제곱분포의 합은 역시 카이제곱분포를 따르며 이 경우 자유도는 합치기 전 각각 확률변수의 자유도의 합과 같다. 즉

$$\chi^2(df_1) + \chi^2(df_2) = \chi^2(df_1 + df_2)$$

카이제곱분포



[자유도가 k 인 카이제곱분포]
(Wikipedia)

카이제곱검정

- 귀무가설하에서 검정통계량의 분포의 카이제곱분포를 따르며 자유도는 (행의 갯수 - 1) \times (열의 갯수 - 1) 이다.
- 팔짱 끼기 예제에서 검정통계량의 값은 0.2이며 자유도가 1 이 카이제곱 분포를 따른다.
- 이 경우 p-value는 0.90이다. 따라서 우리는 귀무가설을 기각할 수 없다. 즉 성별로 팔짱 끼는 방식에 차이가 있다고 결론 낼 수 없다.



아버지의 키를 통제한다면 어머니의 키는 아들의 키와 관련이 있는가?

- 앞에서 배운 다중회귀분석을 이용하여 아버지와 어머니의 키를 예측변수로 해서 아들의 키를 예측해보자.
- R을 이용하여 다중회귀분석을 실시하면 기울기와 절편에 대한 추정치와 추정치의 표준오차를 제공해준다.
- 여기서 귀무가설은 “기울기(절편)가 0이다.”이며 대립가설은 “기울기가 0이 아니다.” 이다.
- 검정통계량은 $(\text{추정치} - 0) / \text{표준오차}$ 로 계산할 수 있으며 귀무가설하에서 t 분포를 따른다.

아버지의 키를 통제한다면 어머니의 키는 아들의 키와 관련이 있는가?

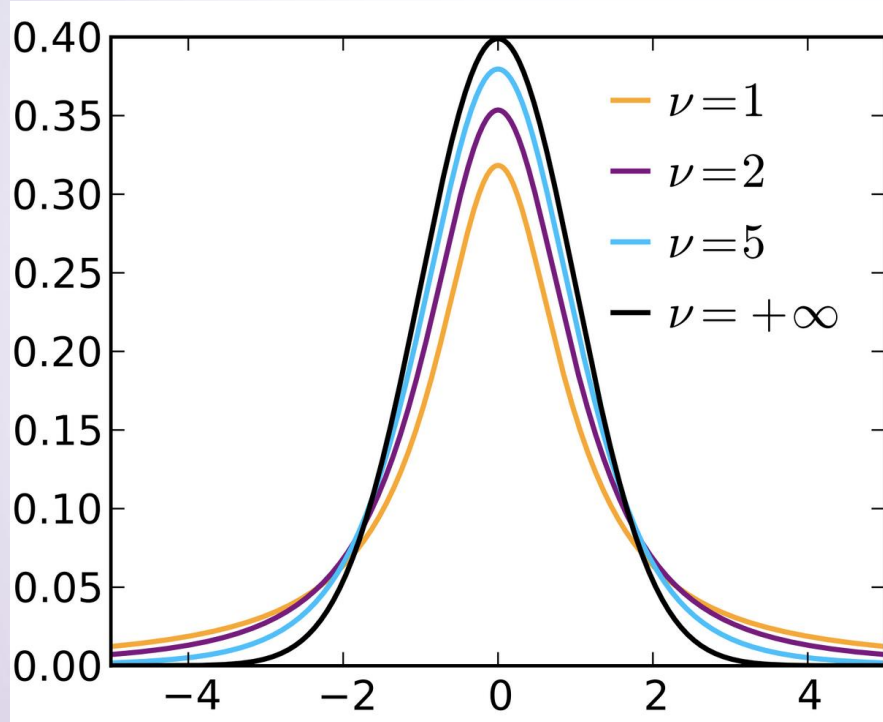
	추정치	표준오차	t값	p-value
절편	69.22882	0.10664	649.168	$< 2 \cdot 10^{-16}$ ***
어머니의 키	0.33355	0.04600	7.252	$1.74 \cdot 10^{-12}$ ***
아버지의 키	0.41175	0.04668	8.820	$< 2 \cdot 10^{-16}$ ***

[골턴의 자료를 이용한 다중회귀분석 결과,
여기서 ***는 p-value가 0.001보다 작다는 것을 의미한다.]
(The Art of Statistics, p276)

t-분포

- t-분포는 student t-분포의 줄임말로 맥주회사 기네스사에서 일하던 윌리엄 고셋이 1908년에 제안하였는데 회사의 방침에 따라 본명을 사용할 수 없었던 고셋이 “student”라는 필명을 사용하여 제안하였다.
- t-분포는 정규분포와 비슷한 모양으로 0을 중심으로 대칭이며 꼬리 부분이 보다 두터운 모양을 가진다.
- t-분포의 모수는 자유도 하나이며 자유도가 커질수록 정규분포와 흡사한 모양을 가지게 된다.

t-분포



[자유도가 ν 인 t-분포]
(Wikipedia)



아버지의 키를 통제한다면 어머니의 키는 아들의 키와 관련이 있는가?

- 원래 주어진 질문에 답변하기 위해서는 어머니의 키 변수의 기울기가 0인지 여부를 검정해야 하는데 p-value들이 매우 작으므로 귀무가설을 기각할 수 있다.
- 즉 결론은 “아버지의 키가 같은 집단에서 어머니의 키가 클수록 아들의 키도 큰 경향이 있다”.



오늘의 강의 요점

- 카이제곱검정
- t-검정

○ 출처

#1 D. Spiegelhater, (2019), The Art of Statistics, Penguin Random House

#2 Wikipedia <https://bit.ly/2Jbfceu>

#3 D. Spiegelhater, (2019), The Art of Statistics, Penguin Random House

#4 Wikipedia <https://bit.ly/2J9F36B>

신뢰구간과 가설검정

3. 다중비교와 검정력

다중검정이란?

- 유의수준으로 흔히 사용하는 0.05와 0.01은 정확한 p-value를 구하는 것이 쉽지 않았을 때 통계학자 피셔가 임의로 제시한 값이었다.
- 하지만 수많은 가설검정을 시행해서 매번 p-value가 0.05보다 작은지를 알아본다는 것은 무엇을 의미하는 것일까?
- 귀무가설이 참일 경우 p-value가 0.05보다 작을 확률은 5%이며 이 경우 귀무가설을 기각하여 잘못된 결론(false-positive)에 이르게 된다.

다중검정이란?

- 만약 귀무가설이 참인 두 번의 가설검정에서 적어도 한번은 p-value가 0.05보다 작을 확률은 $1 - (\text{두 번 모두 p-value가 0.05보다 클 경우}) = 1 - 0.95^2 = 0.0975$ 로 거의 10%에 육박한다.
- 만약 귀무가설의 개수가 10개라면 유의수준 0.05에서 우리가 한 번이라도 잘못 결론을 내릴 확률은 40%나 된다.
- 이런 다중검정(multiple testing)의 문제는 연구자들이 데이터를 여러 개로 쪼개서 각각에 대해 가설 검정을 한 후 유의미한 결과를 발표할 때 생기는 문제이다.

죽은 연어가 산 연구자를 놀린다!

- fMRI와 같은 뇌 영상 이미지는 일반적으로 수백만 개의 voxel로 이루어져 있다.
- 뇌 영상분석에서는 주어진 자극에 따라 어느 특정 뇌 부위의 활동이 변화하는지에 관심이 있다 이 경우 voxel별로 주어진 자극에 따른 뇌 활동의 변화를 검정통계량으로 나타낸 후 어 voxel별로 가설검정을 실시한다.

죽은 연어가 산 연구자를 놀린다!

- 2009년 미국의 연구자들이 피실험자에게 감정이 드러난 사람들의 사진을 보여주면서 뇌 활동의 변화 여부를 찾아내기 위해 fMRI 자료를 이용해 각 voxel별로 유의수준 0.001에서 가설 검정을 실시한 결과 총 8,064개의 영역 중 16개가 유의미한 반응을 보였다.
- 이 실험에서 피실험자는 죽은 연어였다!

Bonferroni 교정

- 앞의 실험에서 귀무가설이 참일 경우 voxel의 p-value값이 0.001보다 작을 확률은 0.001이다. 따라서 8,000개 voxel에서 가설검정을 할 경우 평균적으로 8개 정도의 거짓 양성 (false-positive - 귀무가설이 참인데 귀무가설을 기각하는 경우)이 발생할 수 있다.
- 이러한 문제를 피하기 위해서는 유의수준을 일반적 기준인 0.05보다 많이 낮추어야 한다.
- 가장 보편적인 p-value 교정 방법은 Bonferroni 교정이다. 즉 기준이 되는 유의수준을 일반적 유의수준 0.05를 전체 가설검정의 횟수로 나누는 것이다.

Bonferroni 교정

- 예를 들어 죽은 연어 실험에서는 p -value 가 $0.05/8064=0.0000062$ 약 16만 분의 1보다 작은 경우 귀무가설을 기각한다.
- 이 경우 한 번이라도 잘못 귀무가설을 기각할 확률은 0.05보다 작다는 것을 증명할 수 있다.

오발견율(False Discovery Rate)

- 하지만 Bonferroni 교정을 사용할 경우 상대적으로 귀무가설을 너무 기각하지 않는다는 단점이 있다. 즉 100만 개의 귀무가설 중에 한 번이라도 잘못 기각할 확률을 0.05 이하로 하는 경우와 1,000개중 한 번이라도 잘못 기각할 확률을 0.05로 하는 것을 생각해본다면 전자의 경우 훨씬 엄격한 기준잣대를 사용하고 있다는 것을 알 수 있다.
- 그래서 나온 제안된 다른 교정방법은 오발견율(False Discovery Rate)을 통제하는 것이다.

오발견율(False Discovery Rate)

- 여기서 오발견율은 전체기각된 가설검정의 개수 중 잘못된 기각한 경우(false-positive)의 비율을 이야기한다.
- 두 가지 교정방법 중 선호하는 방식은 총 100개를 기각했을 때 10개를 잘못된 기각한 경우와 50개를 기각했을 때 6개를 잘못된 기각한 것 중 어느 것이 더 심각한 실수인지에 따라 선택할 수 있다.

제1종의 오류와 제2종의 오류

	귀무가설을 기각하지 않음 (무죄라고 판결)	귀무가설을 기각함 (유죄라고 판결)
귀무가설이 참일 때 (피고가 무죄)	무죄인 사람을 사회로 돌려보냄	제1종의 오류 무죄인 사람을 감옥으로 보냄
대립가설이 참일 때 (피고가 유죄)	제2종의 오류 범죄자를 사회로 돌려보냄	유죄인 사람을 감옥으로 보냄

[제1종의 오류와 제2종의 오류]

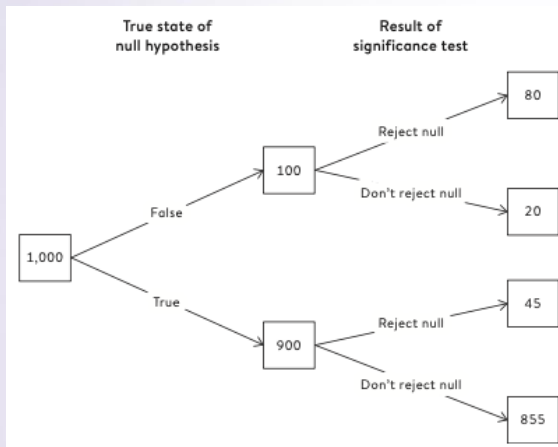
유의수준과 검정력

- 일반적으로 유의수준은 α 로, 제2종의 오류를 범할 확률을 β 로 표시한다. 여기서 $1-\beta$ 를 **검정력**이라고 한다.
- 우리의 가설검정 전략은 제1종의 오류를 범할 확률을 α 이하로 하고 검정력을 최대로 하는 결정을 찾는 것이다.
- 흔히 사용하는 유의수준 $\alpha=0.05$ 이며 검정력은 0.80이다. 즉 새로운 신약이 치료효과가 있는지 검정하기 위해서는 치료효과가 없을 때 귀무가설을 기각할 확률은 0.05 이하이어야 하고 치료가 효과적일 때 치료효과가 있다(귀무가설 기각)라고 결론 내릴 확률이 80%이어야 한다는 의미이다.

효과크기와 표본 수

- 검정력은 표본의 개수가 커질수록 증가한다.
- 따라서 우리는 원하는 검정력을 얻기 위해서는 어느 정도의 표본이 필요한지 미리 계산할 수 있다.
- 물론 이 경우 효과크기(effect size)를 미리 정의해야 한다. 효과크기는 비교하고자 하는 대상의 기대변화량이라고 생각할 수 있다. 만약 두 그룹값의 특정변수의 값을 비교한다면 그룹간 변수값의 차이를 (한 그룹의)표준편차로 나누어준 것을 의미한다.
- 이러한 정보(유의수준, 검정력, 효과크기)를 바탕으로 우리는 표본 수를 미리 결정할 수 있다.

P-value를 둘러싼 논란



[1,000개 가상 가설검정에 관한 기대숫자나무]
(The Art of Statistics, p296)

- 왼쪽 그림은 1,000개의 가상의 연구에 관한 가설검정결과를 기대 숫자나무로 제시하였다.
- 각각의 연구에서 유의수준이 0.05이고 검정력은 0.80이며 전체 연구 중 10%에 해당하는 경우만 귀무가설이 거짓이라고 가정하자.
- 이 경우 전체연구 중 귀무가설을 기각하는 경우는 총 125개이며 이 중 45개는 잘못 기각한 경우이다.

출판된 연구결과의 대부분은 틀렸다!

- 2015년 스탠포드 의대 교수인 존 이오아니디스는 “출판된 대부분의 연구결과는 틀렸다”라는 논문을 유명학술지 “PLoS Medicine”에 발표함.
- 2016년 미국통계학회에서 p-value의 오용을 막기 위한 다음과 같은 6가지 원칙을 제시
 1. P-value는 데이터가 가정된 특정모형(즉 귀무가설)과 얼마나 양립할 수 없는지(귀무가설을 기각)를 나타내기 위한 값이다.
 2. P-value는 귀무가설이 참일 확률을 의미하지 않는다.

미국통계학회의 p-value에 관한 6가지 원칙

3. 과학, 정책, 기업에서 주요의사결정을 p-value가 유의수준보다 작은지 여부로만 판단해서는 안된다.
 - 위의 원칙은 검정력이 작고 신뢰구간이 넓은 소규모 연구에서 특히 유의해야 한다.
4. 투명한 분석절차와 이에 관한 완벽한 보고를 할 경우만 p-value를 사용한 추론이 의미가 있다.
 - 예를 들면 전체 가설검정의 횟수 등이 정확히 보고되어야 한다.

미국통계학회의 p-value에 관한 6가지 원칙

5. P-value 자체 (혹은 통계적 유의미)가 결과의 중요성 또는 효과크기를 의미하지 않는다.

→ 몇 년 전 몇몇 치약이 유해성분을 포함한다고 소동이 일어난 적이 있다. 하지만 대부분의 치약은 이러한 유해성분을 아주 소량 포함하고 있기 때문에 실제 인체에 유해한 영향을 주기 위해서는 엄청난 양의 치약 성분이 인체에 들어가야만 가능했다. 즉 효과 크기가 작은 경우이기 때문에 실제적 유의성(practical significance)은 없다고 보는 것이 맞다.

미국통계학회의 p-value에 관한 6가지 원칙

6. P-value 자체가 특정 모형이나 가설에 대한 증거의 좋은 척도는 아니다. 예를 들면 p-value가 0.05보다 아주 조금 작은 것은 귀무가설에 반하는 약한 증거에 불과하다.

→ 그래서 미국통계학회에서는 많이 사용하는 유의수준을 0.05에서 0.005로 낮출 것을 권장한다. 이 경우 앞의 기대숫자나무에서 귀무가설을 기각하는 경우는 84.5개로 감소하지만 이중 귀무가설을 잘못 기각할 경우는 4.5개로 줄어든다. 따라서 전체 기각한 귀무가설 중 잘못 기각한 비율은 5%에 불과하다.

오늘의 강의 요점

○ 다중검정

→ Bonferroni 교정

→ False discovery rate

○ 검정력과 표본 수, 그리고 효과크기

○ 미국통계학회의 p-value에 관한 6가지 원칙

○ 출처

#1 D. Spiegelhater, (2019), The Art of Statistics, Penguin Random House

가설검정의 여러 문제

Lab 10. 사례연구

성별에 따라 여자친구 최애멤버는 다르다?

- 여자친구 팬클럽 버디 회원 중 300명을 무작위로 추출하여 여자친구 멤버들에 대한 선호도 조사를 하였다. 회원들의 성별에 따라 멤버들의 선호도 차이가 있는지 여부를 알고 싶다. 선호도 조사 결과는 다음과 같다.

성별	소원	예린	은하	유주	신비	엄지	합계
남	20	33	30	22	22	23	150
여	19	34	27	22	32	16	150
합계	39	67	57	44	54	39	300

성별에 따라 여자친구 최애멤버는 다르다?

```
library(DT)
gfriend.tab1 <- matrix(c(20,19,33,34,30,27,22,22,22,32,23,16), nrow=2)
dimnames(gfriend.tab1) <- list(sex=c("male", "female"), member=c("소원", "예린", "은하", "유주", "신비", "엄지"))
datatable(gfriend.tab1)
```

Show 10 entries

Search:

	소원	예린	은하	유주	신비	엄지
male	20	33	30	22	22	23
female	19	34	27	22	32	16

Showing 1 to 2 of 2 entries

Previous

1

Next

성별에 따라 여자친구 최애멤버는 다르다?

성별에 따라 선호하는 멤버가 차이가 있는지 알아보기 위해 카이제곱 검정을 시행하자.

```
chisq.test(gfriend.tab1)
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  gfriend.tab1  
## X-squared = 3.3067, df = 5, p-value = 0.6528
```

급성백혈병의 2가지 종류

- 급성 백혈병은 세포의 기원에 따라 급성 골수성 백혈병(AML)과 급성 림프모구 백혈병(ALL)으로 나뉜다.
- Golub et al.(1999)는 마이크로 어레이 데이터를 이용하여 두 종류의 급성 백혈병을 분류하는 방법에 대해서 연구하였다.
- R package multtest에서는 Golub et al. 에서 사용된 급성 백혈병 마이크로 어레이 자료중 골수에서 추출한 38(AML 14명, ALL 24명)명의 자료를 제공하고 있으며 각 표본은 3,501개의 유전자의 발현정도를 나타내고 있다.
- 우리의 목표는 AML과 ALL의 분류를 쉽게 하는 biomarker에 해당하는 유전자를 찾는 것이다.

Biomarker와 다중검정

- 이러한 biomarker를 찾기 위해 각각의 유전자 별로 2표본 검정을 실시한 후 p값을 얻을 수 있다.
- 하지만 3,000개 이상의 가설검정을 동시에 하기 때문에 다중 검정에 관한 문제가 발생한다.
- 따라서 Bonferroni 보정, 혹은 FDR 통제를 통해 통계적으로 유의미하게 발현 정도가 차이가 나는 유전자의 목록을 찾고자 한다.

다중비교

Multtest package를 설치하고 Golub et al.(1999)의 급성백혈병 자료를 불러온다.

```
if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")
```

```
BiocManager::install("multtest")
```

```
## Bioconductor version 3.12 (BiocManager 1.30.10), R 4.0.3 (2020-10-10)
```

```
## Installing package(s) 'multtest'
```

```
##  
## The downloaded binary packages are in  
## /var/folders/rx/nksb_6pd39g1fh40b2j9bcg00000gn/T//RtmpkDP0e3/downloaded_packages
```

```
## Old packages: 'AnalyzeFMRI'
```

```
library(multtest)
```

다중비교

2표본 t-검정을 유전자별로 실시하고 p값을 구하자.

```
teststat = mt.teststat(golub, golub.c1)
rawp = 2 * (1 - pnorm(abs(teststat)))
```

Bonferroni 교정과 FDR control 방법을 시행하자,

```
adjusted = mt.rawp2adjp(rawp, c("Bonferroni", "BH"))
```

Bonferroni 교정방법과 FDR control을 통해 귀무가설을 기각한 횟수

```
sum(adjusted$adj[,2] < 0.05) # Bonferroni
```

```
## [1] 228
```

```
sum(adjusted$adj[,3] < 0.05) # FDR
```

```
## [1] 883
```



오늘의 강의 요점

- 카이제곱검정

- 다중비교

 - Bonferroni 교정

 - FDR