# 7. 중간시험 리뷰 및 가설 검정 II

## 정치와 데이터분석

박상훈 (sh.park.poli@gmail.com)
강원대학교

# 오늘의 목표

.pull-left[

### 10:05-10:45

중간시험에서 오답률이 높았던 문항들 위주로 검토

### 10:55-11:40

가설 검정의 논리에 대한 개요를 복습 및 학습

### 11:55-12:40

실습과제 해설 및 질의응답

보충과제 설명

# Part I. 중간시험 다시보기

# 중간시험 리뷰

## 객관식 문제 3번

유의수준 0.05에서 실시한 가설검정 결과 $p$-값이 0.03으로 산출되었다. 이에 대한 올바른 해석은 무엇인가?

A. 귀무가설이 참일 확률이 0.03이라는 뜻이다.

B. 귀무가설이 참이라고 가정할 때, 지금과 같은 결과가 나올 확률이 0.03이라는 뜻이다.

C. 대립가설이 참일 확률이 97%라는 뜻이다.

D. 유의수준이 0.03으로 자동 조정되었다는 뜻이다.

# 중간시험 리뷰

## 객관식 문제 6번

두 사건 A와 B가 독립(independent)일 때 다음 중 옳게 성립하는 것은?

A. $P(A \cap B) = P(A) \times P(B)$

B. $P(A \cap B) = P(A) + P(B)$

C. $P(A \cup B) = P(A) \times P(B)$

D. 두 사건이 독립이라는 것은 곧 상호배타적이라는 뜻이다.

# 중간시험 리뷰

## 객관식 문제 8번

어떤 표본으로부터 추정한 모평균의 95% 신뢰구간이 10에서 20으로 주어졌다. 다음 설명 중 옳은 것을 고르시오.

A. 귀무가설 $H_0 : \mu = 0$ 는 5% 유의수준에서 기각된다.

B. 귀무가설 $H_0 : \mu = 15$는 5% 유의수준에서 기각된다.

C. 신뢰수준을 99%로 높이면 이 신뢰구간의 길이는 더 짧아질 것이다.

D. 동일한 표본으로 90% 신뢰구간을 계산하면 95% 신뢰구간보다 폭이 넓을 것이다.

E. 모평균이 10과 20 사이에 있을 확률이 95%이다.

# 중간시험 리뷰

## 객관식 9번

표본 크기($n$)가 커질수록 표준오차(SE)가 감소하는 이유로 가장 적절한 것은?

A. 표본의 평균이 모집단 평균과 점점 멀어지기 때문이다.

B. 표본 크기가 커지면 표본평균의 변동성이 작아지기 때문이다.

C. 표본 크기가 커지면 신뢰수준이 자동으로 높아지기 때문이다.

D. 표본 크기가 커질수록 표준편차 자체가 작아지기 때문이다.

# 중간시험 리뷰

## 약술형 11번

다음은 R 코드 일부이다.

이 코드가 수행하는 작업과 그 결과를 평이한 한국어로 설명하시오. (5점)

```r
scores <- c(85, 90, 78, 92, 88)
adjusted <- c()

for (i in 1:length(scores)) {
  adjusted[i] <- scores[i] + 5
}

mean(adjusted) # 예시: adjusted 객체(벡터)의 평균을 구하라.
```

# 중간시험 리뷰

## 약술형 11번

다음은 R 코드 일부이다.

이 코드가 수행하는 작업과 그 결과를 평이한 한국어로 설명하시오. (5점)

```r
# scores라는 벡터에 5명의 학생 점수(85, 90, 78, 92, 88)를 저장
scores <- c(85, 90, 78, 92, 88)
# adjusted라는 빈 벡터를 생성
adjusted <- c()
# for 반복문을 이용해 scores의 각 원소에 5점을 더한 값을 adjusted에 차례로 저장
for (i in 1:length(scores)) {
  adjusted[i] <- scores[i] + 5
}
mean(adjusted) # 예시: adjusted 객체(벡터)의 평균을 구하라.
```

# 중간시험 리뷰

## 약술형 11번

다음은 R 코드 일부이다.

이 코드가 수행하는 작업과 그 결과를 평이한 한국어로 설명하시오. (5점)

```r
# scores라는 벡터에 5명의 학생 점수(85, 90, 78, 92, 88)를 저장
scores <- c(85, 90, 78, 92, 88)
# adjusted라는 빈 벡터를 생성
adjusted <- c()
# for 반복문을 이용해 scores의 각 원소에 5점을 더한 값을 adjusted에 차례로 저장
for (i in 1:length(scores)) {
  adjusted[i] <- scores[i] + 5
}
mean(adjusted) # 예시: adjusted 객체(벡터)의 평균을 구하라.
```

```
## [1] 91.6
```

# 중간시험 리뷰

## 약술형 11번

다음 중 위 코드와 동일한 결과를 만들지 않는 표현을 고르시오. (5점)

A. `mean(scores + 5)`

B. `for (i in 1:5) { scores[i] <- scores[i] + 5 }; mean(scores)`

C. `adjusted <- scores + 5; mean(adjusted)`

D. `mean(scores) + 5`

# 중간시험 리뷰

## 약술형 11번

```r
scores <- c(85, 90, 78, 92, 88)
mean(scores + 5)
```

```
## [1] 91.6
```

```r
scores <- c(85, 90, 78, 92, 88)
for (i in 1:5) { scores[i] <- scores[i] + 5 }; mean(scores)
```

```
## [1] 91.6
```

# 중간시험 리뷰

## 약술형 11번

```r
scores <- c(85, 90, 78, 92, 88)
adjusted <- scores + 5; mean(adjusted)
```

```
## [1] 91.6
```

```r
scores <- c(85, 90, 78, 92, 88)
mean(scores) + 5
```

```
## [1] 91.6
```

# 중간시험 리뷰

## 약술형 13번

> 한 연구에서 모집단의 평균이 50이라고 알려져 있다. 연구자가 새로운 표본 30명을 조사했더니 표본평균이 55로 나타났고, 표본의 표준편차는 10이었다. 이 변화가 우연에 의한 것인지 검정하기 위해 귀무가설을 $H_0 : \mu = 50$으로 두고 분석한다고 하자. (a) 해당 표본에 대한 검정통계량 $t$ 값을 계산하시오. (b) 계산된 $t$값이 유의수준 0.05에서 유의한지 여부를 판단하고 그 근거를 간략히 설명하시오. 이때, $t$-통계량을 계산하는 공식은 다음과 같다. 단, 자유도는 고려하지 않고 주어진 통계량으로 계산된 $t$-통계량으로 판단하라. (10점)

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

# 중간시험 리뷰

## 약술형 13번

### 검정통계량 계산

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{55 - 50}{10/\sqrt{30}} = \frac{5}{1.8257} \approx 2.74$$

# 중간시험 리뷰

## 약술형 13번

### 유의성 판단

유의수준 $\alpha = 0.05$에서 양측 검정을 가정하면 임계값은 약 $1.96$이다.
계산된 $|t| = 2.74 > 1.96$이므로, 귀무가설 $H_0 : \mu = 50$을 기각한다.

**결론**: 표본평균이 55로 나타난 것은 우연에 의한 차이라 보기 어렵고, 유의수준 0.05에서 통계적으로 유의하다.

# 중간시험 리뷰

## 약술형 14번

> 어떤 조사에서 집단 A의 평균값이 80점이고 집단 B의 평균값이 75점으로 측정되었다. 두 집단의 표본 크기는 각각 30명이며, 표준편차는 두 집단 모두 5점이라고 하자. 두 집단 평균의 차이가 우연에 의한 것인지 알아보기 위해, 독립표본 t-검정의 검정통계량 $t$ 값을 계산하시오. 계산한 $t$ 값이 유의수준 0.05에서 통계적으로 유의한지도 함께 판단하고 설명하시오. 단, 자유도는 고려하지 않고 주어진 통계량으로 계산된 $t$-통계량으로 판단하라. (10점)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

# 중간시험 리뷰

## 약술형 14번

### 검정통계량 계산

주어진 값: $\bar{X}_1 = 80,\ \bar{X}_2 = 75,\ s_1 = s_2 = 5,\ n_1 = n_2 = 30.$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{80 - 75}{\sqrt{\frac{25}{30} + \frac{25}{30}}} = \frac{5}{\sqrt{\frac{50}{30}}} = \frac{5}{\sqrt{\frac{5}{3}}} \approx \frac{5}{1.29099} \approx 3.873.$$

# 중간시험 리뷰

## 약술형 14번

### 유의성 판단

유의수준 $\alpha = 0.05$에서(자유도는 고려하지 않는다는 지시에 따라) 정규 임계값 $1.96$과 비교하면 $|t| = 3.873 > 1.96$이므로 ₩textbf{통계적으로 유의}하다.

**결론**: 두 집단 평균 차이는 우연만으로 보기 어렵다.

# 중간시험 리뷰

## 약술형 15번

**제1종 오류(Type I error)**: 실제로는 귀무가설($H_0$)이 참인데 이를 잘못 기각하는 오류를 말한다. 즉, “효과가 없는데 있다”고 결론 내리는 잘못이다. 연구 해설의 관점에서 이는 연구자가 잘못된 긍정적 결과를 보고하게 되는 거짓 긍정(false positive)의 의미를 가지며, 잘못된 정책 제안이나 이론적 주장으로 이어질 수 있다.

**제2종 오류(Type II error)**란 실제로는 대립가설($H_1$)이 참인데 귀무가설을 기각하지 못하는 오류를 말한다. 즉, "효과가 있는데 없다"고 결론 내리는 잘못이다. 연구 해설의 관점에서 이는 거짓 부정(false negative)에 해당하며, 실제로 중요한 관계나 현상을 발견하지 못해 학문적 또는 정책적으로 의미 있는 사실을 놓치는 문제를 뜻한다.

# Part II. 가설 검정의 논리

# 가설 검정 II.

## 귀무가설 유의성 검정 복습

귀무가설 ($H_0$): "효과가 없다"는 가설 (예: 두 집단 평균 차이 = 0, 계수 = 0 등)

연구가설 ($H_A$): **우리가 입증하고자 하는 가설** (효과가 존재하거나 차이가 있음)

# 가설 검정 II.

## 귀무가설 유의성 검정 복습

유의수준 ($\alpha$): 귀무가설 기각의 기준 확률. 보통 0.05 (5%) 사용. 귀무가설이 참일 때 5% 미만의 희박한 결과만 인정

$p-$값: 귀무가설이 참이라고 가정할 때, 현재 관측된 결과만큼 극단적이거나 더 극단적인 결과가 나올 확률. 값이 작을수록 현재 데이터가 귀무가설과 양립하기 어려움.

- $p-$값 $< \alpha$ 이면 결과를 **유의미하다**고 보고 귀무가설 기각
- $p-$값 $\geq \alpha$ 이면 **유의하지 않다**라고 보고 귀무가설을 기각하지 못함(채택이 아님에 유의).

# 가설 검정 II.

## $p-$값과 통계적 유의성의 한계

통계적 유의미성은 **연히 나오기 어렵다**는 의미이지, 효과의 크기나 중요도를 직접 말해주지는 않음.

- 효과 크기 미제시: $p-$값은 관찰값의 희귀성을 나타낼 뿐, 효과의 크기나 중요성을 알려주지 않음. 큰 표본에서는 아주 작은 차이도 $p-$값을 유의하게 만들 수 있음.
- 실질적 중요성 착각: 통계적으로 유의미한 결과가 반드시 실질적으로 의미있는 영향이나 큰 효과를 의미하지는 않음. 예를 들어, 정치적으로 미미한 차이도 표본 수가 크면 유의하게 나올 수 있음.
- 임의적 기준: 0.05라는 유의수준 자체가 자의적. $p = 0.049$와 $p = 0.051$은 거의 차이 없지만 하나는 유의, 하나는 유의 아님으로 구분되어 버리는 문제

# 가설 검정 II.

## $p-$값과 통계적 유의성의 한계

$p-$hacking 우려: 연구자가 많은 가설을 검정하여 운 좋게 유의한 결과만 선택 보고할 위험 존재(유의확률 남용). $p-$값만 맹신하면 거짓양성 결과에 속을 수 있음.

**대안 강조**: 따라서 $p-$값과 함께 효과의 크기, 신뢰구간 등을 함께 보고 해석하는 것이 중요(통계적 유의성 + 실질적 중요성 모두 고려)

# 가설 검정 II.

## 이변량 가설검정

두 변수 간 관계를 검정할 때, 변수의 측정 수준 조합에 따라 적절한 통계기법이 달라짐.

| 종속변수 유형 / 독립변수 유형 | 범주형 IV | 연속형 IV |
|---|---|---|
| **범주형 DV** (명목/서열) | **교차분석** (카이제곱 검정) | **로지스틱 회귀** 등 (추후 심화) |
| **연속형 DV** (등간/비율) | **평균 차이 검정** (t-검정) | **상관분석** (피어슨 r, 또는 회귀) |

범주형-범주형: 분할표(교차표)를 만들어 독립성 검정 (카이제곱 테스트)

범주형-연속형: 두 집단(또는 다수 집단)의 평균을 비교하여 t-검정 (또는 ANOVA) 실시

연속형-연속형: 상관계수를 계산하고 그 유의미성을 검정 (또는 단순 회귀분석으로도 동일한 결과를 얻음)

연속형 독립변수-범주형 종속변수 조합은 로지스틱 회귀 등 비선형 모형 필요(추후에 다룸)

# 가설 검정 II.

## 이변량 가설검정

### 교차분석(카이제곱 검정)

연구질문: 민주주의 국가와 비민주주의 국가 간에 정부의 언론 검열 발생률에 차이가 있는가? (두 범주형 변수 간 관계)

- 변수: 민주주의 여부 (범주형: 민주국/비민주국), 언론검열 존재 여부 (범주형: 있음/없음)
- 귀무가설 $H_0$: 두 변수는 독립적이다 (민주국이든 아니든 검열 비율 차이 없음)
- 대립가설 $H_A$: 민주국과 비민주국의 검열 발생률에 차이가 있다 (민주국에서 검열이 덜할 것이라는 기대)

# 가설 검정 II.

## 이변량 가설검정

### 교차분석(카이제곱 검정)

```
# V-Dem 2019년 자료에서 변수 생성
vdemdata::vdem -> vdem
vdem2019 <- vdem |> dplyr::filter(year == 2019)
vdem2019 |> mutate(
  democracy = if_else(v2x_polyarchy > 0.5, "민주국", "비민주국"),
  censorship = if_else(v2mecenefm > 0, "검열 있음", "검열 없음")
) -> vdem2019
```

# 가설 검정 II.

## 이변량 가설검정

### 교차분석(카이제곱 검정)

```
# 교차표 생성
vdem2019 |> janitor::tabyl(democracy, censorship)
```

```
##   democracy 검열 없음 검열 있음
##      민주국         3        91
##    비민주국        56        29
```

# 가설 검정 II.

## 이변량 가설검정

### 교차분석(카이제곱 검정)

```r
# 카이제곱 독립성 검정
chisq.test(vdem2019$democracy, vdem2019$censorship)
```

```
## 
##      Pearson's Chi-squared test with Yates' continuity correction
## 
## data:  vdem2019$democracy and vdem2019$censorship
## X-squared = 76.58, df = 1, p-value < 2.2e-16
```

해석: $\chi^2(1) = 76.58, p < 0.001$ 수준이므로 귀무가설 기각. 민주국과 비민주국 사이에 언론 검열 발생 빈도에 유의한 차이가 있음(민주국에서 검열이 상대적으로 드물게 나타남).

# 가설 검정 II.

## 이변량 가설검정

### 평균차이 검정(t-test)

연구질문: 민주주의 국가와 비민주주의 국가의 표현의 자유 수준 평균에 차이가 있는가? (범주형 - 연속형)

- 변수: 민주주의 여부 (두 집단), 표현의 자유 지수 (0~1 연속지표)
- 귀무가설 $H_0$: 두 집단 평균이 같다 (차이=0)
- 대립가설 $H_A$: 두 집단 평균이 다르다 (민주국이 더 높을 것이라 기대)

# 가설 검정 II.

## 이변량 가설검정

### 평균차이 검정(t-test)

```
# 민주국 vs 비민주국의 표현의 자유 지수 비교 (두 집단 독립표본 t-검정)
t.test(vdem2019$v2x_freexp ~ vdem2019$democracy)
# t = 16.718, df = 107.52, p-value < 2.2e-16
# alternative hypothesis: true difference in means between group 민주국 and group 비민주국 is no
# 95 percent confidence interval:
#   0.4010025 0.5088904
# sample estimates:
#   mean in group 민주국 mean in group 비민주국
#             0.8616170             0.4066706
```

# 가설 검정 II.

## 이변량 가설검정

### 평균차이 검정(t-test)

민주국 평균 표현의 자유 지수 $\approx$ 0.86, 비민주국 $\approx$ 0.41 로 약 0.45의 차이가 관찰

검정 결과 $t \approx 15.0, p < 0.001$ 로 유의수준 0.05 하에서 유의하므로, 두 집단의 평균 차이는 통계적으로 유의미(민주국이 표현의 자유 수준이 확연히 높음).

# 가설 검정 II.

## 이변량 가설검정

### 상관계수 검정

연구질문: 한 국가의 민주주의 수준과 표현의 자유 수준은 상관관계가 있는가? (연속형 - 연속형)

- 변수: 민주주의 지수, 표현의 자유 지수, 범위 0~1 연속형
- 귀무가설 $H_0$: 두 변수 간 상관계수 $\rho = 0$ (관계 없음)
- 대립가설 $H_A$: $\rho \neq 0$ (상관관계 있음)

# 가설 검정 II.

## 이변량 가설검정

### 상관계수 검정

```
# 피어슨 상관계수 및 유의성 검정
cor.test(vdem2019$v2x_polyarchy, vdem2019$v2x_freexp)
```

```
## 
##      Pearson's product-moment correlation
## 
## data:  vdem2019$v2x_polyarchy and vdem2019$v2x_freexp
## t = 32.538, df = 177, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9013059 0.9441126
## sample estimates:
##      cor 
## 0.925615
```

# 가설 검정 II.

## 이변량 가설검정

### 상관계수 검정

Pearson 상관계수 $r \approx 0.93$ 으로 매우 강한 양의 상관관계를 보임.

- $p-$값 $< 0.001$ 수준으로 유의하므로, 민주주의 정도가 높을수록 표현의 자유도 높은 경향이 통계적으로도 유의미함(귀무가설 $\rho = 0$는 기각)

상관관계의 방향은 양(+), 정도는 강함으로 해석되고, 결정계수 $r^2 \approx 0.86$ 수준이므로 한 변수의 변동 상당 부분(86%)을 다른 변수가 설명한다고 볼 수 있음(회귀분석 파트에서 다시 다룰 개념임).

**단, 상관관계는 인과관계를 의미하진 않으며, 제3요인 통제는 이후 다변량 회귀에서 다룸.**

# 가설 검정 II.

## 회귀분석으로의 확장: $t-$통계량과 유의성

이제 다수의 독립변수가 있는 회귀분석 상황으로 넘어가도, 핵심 개념은 동일

- 각 추정된 회귀계수에 대해 "영향 없음"의 귀무가설을 검정하게 되는 것
- 개별 계수의 가설검정
  - 귀무가설 $H_0 : \beta_k = 0$ (변수 $k$의 효과 없음).
  - 대립가설 $H_A : \beta_k \neq 0$

# 가설 검정 II.

## 회귀분석으로의 확장: $t-$통계량과 유의성

$t-$통계량 계산: 추정된 계수의 표준오차 대비 크기를 나타냄. 계수 $\hat{\beta}_k$에 대하여,

$$t = \frac{\hat{\beta}_k - 0}{\mathrm{SE}(\hat{\beta}_k)} \sim t_{(n-k-1)}$$

- 여기서 $n$은 표본 크기, $k$는 독립변수 개수 (회귀식의 모수 수; 자유도는 $n-k-1$). 귀무가설이 참이면 이 비율이 $t-$분포를 따를 것

## 유의성 판단

계산된 $|t|$ 값이 매우 크면 (양측검정 기준) 영가설 하에서는 보기 힘든 극단적 값이라는 뜻이므로 관례적으로 큰 표본일 때 $|t| \gtrsim 1.96$이면 $p-$값 $\approx 0.05$ 임계수준이고, $|t| \geq 2 \sim 3$ 이상이면 $p-$값은 0.05보다 훨씬 작아지게 됨(유의미).

# 가설 검정 II.

## 신뢰구간과 임계값

신뢰구간(confidence interval): 표본 통계에 기반해 모수의 "믿을 만한 값의 범위"를 제시

- 보통 95% 신뢰구간은 "모수가 이 구간 내에 있을 것으로 기대되는 범위"를 의미
- 어떤 비율의 95% 신뢰구간이 40% ~ 50%라면, 참값이 0.4~0.5 사이에 있을 확률이 95%라는 뜻

95% 신뢰구간은 대략 추정치 $\pm 1.96\times$ 표준오차로 계산

- 여기서 1.96은 앞서 본 표준정규분포에서 95% 영역에 해당하는 임계값
- 표본 크기가 충분히 크다면 $t \approx 1.96$

# 가설 검정 II.

## 신뢰구간과 임계값

### p값과의 연결

- 귀무가설 값(예: 0)이 신뢰구간 밖에 있으면 그 효과는 유의수준 0.05에서 유의하다고 이해할 수 있음.
- 반대로 귀무가설 값이 신뢰구간 안에 포함되면 $p-$값은 0.05보다 큰 것이고 통계적으로 유의하지 않다고 볼 수 있음.

# 가설 검정 II.

## 신뢰구간과 임계값

두 접근은 동일한 검정 근거를 다른 방식으로 표현한 것

- 예를 들어 회귀계수의 95% 신뢰구간이 $[0.1, 0.5]$라면 0을 포함하지 않으므로 해당 계수의 $p < 0.05$
- 신뢰구간 $[-0.2, 0.8]$처럼 0을 포함하면 $p > 0.05$일 것으로 기대할 수 있음.
- 신뢰구간은 효과의 방향과 범위를 직관적으로 보여준다는 장점이 있어 해석에 유용

# 가설 검정 II.

## 마지막 점검

Q. $p-$값이 0.05보다 작다는 것은 무엇을 의미하는가? "영가설이 참일 때 이런 정도의 극단적 결과가 나올 확률이 5% 미만이다"라는 의미를 정확히 설명해보자.

Q. 통계적으로 유의미한 결과가 항상 연구적으로 중요하거나 큰 효과를 뜻하지는 않는다. 아주 큰 표본으로 사소한 차이를 찾아낸 경우 어떻게 해석해야 할까?

Q. 연구에서 $p-$값에만 의존하는 것의 위험은 무엇일까? $p-$값 이외에 신뢰구간, 효과크기 등을 함께 고려하면 어떤 점이 보완될까?

Q. 표본의 크기가 증가하면 (다른 조건이 동일할 때) $p-$값에 어떤 영향이 있을까?

# Part III. 실습과제 및 추가과제 논의

# 실습과제 8

어떤 사람이 "젤리빈이 여드름을 유발한다"고 주장했다. 과학자들은 이 주장을 검증하기 위해 실험을 수행했고, "젤리빈과 여드름 사이에는 유의미한 관계가 없다 ($p > 0.05$)"는 결과를 얻었다. 그러자 또 다른 사람이 "그중 특정 색깔의 젤리빈이 문제일 수도 있다"고 주장했다. 과학자들은 20가지 색깔의 젤리빈 각각에 대해 별도의 검정을 수행했다. 그 결과, 단 하나의 색깔(초록색) 젤리빈만이 $p < 0.05$로 나타났다. 다음 날 신문에는 "초록색 젤리빈이 여드름을 유발한다!는 기사가 실렸다.

# 실습과제 8

## 모집단(population), 표본(sample), 표집(sampling) 개념을 이용해, 이 실험의 구조를 설명해보자.

모집단(population): 모든 젤리빈 색깔과 여드름의 관계에 대한 전체 대상

표본(sample): 실제 실험에 참여한 사람들 또는 각 색 젤리빈을 먹은 집단

표집(sampling): 모집단에서 일부를 선택하여 각 색 젤리빈 실험에 배정한 과정

과학자들은 모집단 전체를 관찰할 수 없어, 표본을 통해 각 색에 대한 효과를 추정한 실험 구조를 가짐

# 실습과제 8

## 영가설(null hypothesis)과 대립가설(alternative hypothesis)을 각각 어떻게 설정할 수 있을까?

영가설($H_0$): 젤리빈 색깔과 여드름 사이에 아무런 관련이 없음.

대립가설($H_1$): 특정 색깔의 젤리빈이 여드름과 관련이 있음.

즉, 관계가 없다는 가설을 기각할 수 있을지를 검정하는 구조

## 과학자들이 사용한 유의수준($\alpha$ = 0.05)은 무엇을 의미하는가?

영가설이 참인데도 잘못 기각할 확률을 5%로 설정

이는 제1종 오류(Type I error)를 허용하는 수준을 의미

실제로 아무 관련이 없어도 100번 중 5번은 우연히 유의하게 나올 수 있음을 뜻함.

# 실습과제 8

## 왜 20가지 색깔을 각각 따로 검정했을 때 하나쯤은 우연히 유의하게 나올 가능성이 높아지는가?

각 검정마다 우연히 유의하게 나올 확률이 0.05

- 20번 모두 유의하지 않을 확률은 $0.95^{20} \approx 0.36$임.
- 따라서 최소 한 번 이상 유의하게 나올 확률은 $1 - 0.95^{20} \approx 0.64$임.

즉, 검정 횟수가 많아질수록 우연히 유의한 결과가 나올 가능성이 높아짐.

이러한 현상을 다중 비교 문제(multiple testing problem)라 함.

# 실습과제 8

**마지막 신문 기사의 “95% 신뢰 수준에서 초록색 젤리빈은 여드름과 관련 있다”는 주장은 왜 올바른 해석인가 잘못된 해석인가? 왜 그러한가?**

**올바른 해석이 아님.**

- 다중 검정에서 한 색만 유의하게 나온 것은 단순한 우연일 가능성이 큼.
- 유의수준 0.05는 "영가설이 참일 때 이런 결과가 나올 확률이 5% 미만임"을 의미
- 이는 "초록 젤리빈이 실제로 여드름을 유발할 확률이 95%임"을 뜻하지 않음.

**통계적 유의성(statistical significance)과 실질적 유의성(practical significance)은 다름.**

따라서 해당 기사는 통계적 결과를 잘못 해석한 사례

KNU 강원대학교
KANGWON NATIONAL UNIVERSITY

# 감사합니다!

## 궁금한 것이 있으면 언제든 연락하세요.

강사 연락처

| 연락처 | 박상훈 |
|---|---|
|  | sh.park.poli@gmail.com |
|  | sanghoon-park.com/ |
|  | 영상바이오관 405 |