

두 모집단의 비교

짝비교

- ▶ 짝비교: 두 집단이 독립이 아닐 때 평균의 차이를 확인
- ▶ 짝비교를 시행할 때의 자료 형태

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_1	Y_2	$D_2 = X_2 - Y_2$
\vdots	\vdots	\vdots	\vdots
n	X_n	Y_n	$D_n = X_n - Y_n$

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i,$$

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n - 1}$$

평균차 δ 에 대한 추론(표본의 크기가 작을 때)

- D_1, D_2, \dots, D_n 을 $N(\delta, \sigma_D^2)$ 로부터 추출한 표본이라고 할 때,
- 모평균 δ 에 대한 $100(1 - \alpha)\%$ 신뢰구간

$$\bar{D} \pm t_{\frac{\alpha}{2}}(n - 1) \times \frac{s_D}{\sqrt{n}}$$

- $H_0: \delta = \delta_0$ 에 대한 검정통계량

$$t = \frac{\bar{D} - \delta_0}{s_D / \sqrt{n}}$$

귀무가설이 참일 때, t 는 자유도가 $(n - 1)$ 인 t 분포를 따른다.
(표본의 크기가 클 때에는 근사적으로 표준정규분포를 따른다.)

평균차 δ 에 대한 추론(표본의 크기가 작을 때)

환자	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
전 (x)	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
후 (y)	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74
차이 (D)	2	8	10	6	18	10	4	26	18	-8	0	32	0	-4	10

평균차 δ 에 대한 추론(표본의 크기가 작을 때)

	1	2	3	4	5	6	7	8	9
방법 A	90	86	72	65	44	52	46	38	43
방법 B	85	87	70	62	44	53	42	35	46

모비율의 차 $p_1 - p_2$ 에 대한 추론(표본의 크기가 클 때)

- ▶ 두 모집단으로부터 추출된 독립된 두 표본

	특성 A	특성 A 아님	
모집단1	X	$n_1 - X$	n_1
모집단2	Y	$n_2 - Y$	n_2

- ▶ 각 모집단에서 추출된 표본 중 특성 A의 비율

$$\widehat{p}_1 = \frac{X}{n_1}, \quad \widehat{p}_2 = \frac{Y}{n_2}$$

모집단의 비율 p_1 과 p_2 의 추정량으로 \widehat{p}_1 과 \widehat{p}_2 를 사용할 수 있다.

따라서 두 모집단의 모비율의 차 $(p_1 - p_2)$ 의 추정량: $(\widehat{p}_1 - \widehat{p}_2)$

모비율의 차 $p_1 - p_2$ 에 대한 추론(표본의 크기가 클 때)

▶ 표본의 크기 n_1, n_2 가 충분히 큰 경우에

$$X \sim N(n_1 p_1, n_1 p_1 (1 - p_1))$$

$$Y \sim N(n_2 p_2, n_2 p_2 (1 - p_2))$$

가 근사적으로 성립한다. 따라서 표본비율의 분포는

$$\widehat{p}_1 = \frac{X}{n_1} \sim N\left(p_1, \frac{p_1(1 - p_1)}{n_1}\right)$$

$$\widehat{p}_2 = \frac{Y}{n_2} \sim N\left(p_2, \frac{p_2(1 - p_2)}{n_2}\right)$$

모비율의 차 $p_1 - p_2$ 에 대한 추론(표본의 크기가 클 때)

➤ $(\widehat{p}_1 - \widehat{p}_2)$ 의 분포는 다음과 같고,

$$\widehat{p}_1 - \widehat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

표준화를 거치면 다음과 같이 표준정규분포를 따른다.

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

모비율의 차 $p_1 - p_2$ 에 대한 신뢰구간(표본의 크기가 클 때)

➤ $(p_1 - p_2)$ 의 $100(1 - \alpha)\%$ 신뢰구간

$$(\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2} \times \sqrt{\frac{\widehat{p}_1(1 - \widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1 - \widehat{p}_2)}{n_2}}$$

모비율의 차 $p_1 - p_2$ 에 대한 신뢰구간(표본의 크기가 클 때)

	발아됨	발아되지 않음	계
화학 처리된 씨앗	88	12	100
화학 처리를 하지 않은 씨앗	126	24	150

$H_0: p_1 = p_2$ 에 대한 검정(표본의 크기가 클 때)

- 표본의 개수 n_1, n_2 가 충분히 클 때, 귀무가설 $H_0: p_1 - p_2 = 0$ 에 대한 검정통계량은 다음과 같다.

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2)}{\sqrt{\widehat{p}(1 - \widehat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

이 때 $\widehat{p} = \frac{X+Y}{n_1+n_2}$ 은 귀무가설 하에서의 공통비율 p 의 추정량이고, 검정통계량의 분포는 H_0 가 맞을 때 근사적으로 $N(0,1)$ 을 따른다. 각 대립가설에 대하여 유의수준 α 를 갖는 기각역은 다음과 같다.

$H_1: p_1 > p_2$ 일 때

$$R: Z \geq z_\alpha$$

$H_1: p_1 < p_2$ 일 때

$$R: Z \leq -z_\alpha$$

$H_1: p_1 \neq p_2$ 일 때

$$R: |Z| \geq z_{\alpha/2}$$

$H_0: p_1 = p_2$ 에 대한 검정(표본의 크기가 클 때)

	생존	사망	계
리스테린 사용	34	6	40
리스테린 미사용	19	16	35