제2장 두 모집단의 비교

2.1 기본 용어의 정리

· 모집단(population): 통계분석의 대상이 되는 단위 원소의 총집합

· 랜덤추출(random sampling): 모집단을 잘 대표할 수 있도록 모집단을 구성하는 모든 원소마다 표본으로 선택될 확률이 같고, 한 원소의 선택이 다른 원소의 선택에 영향을 미치지 않도록 표본을 추출하는 방법

· 랜덤표본(random sample): 랜덤추출을 통하여 얻어진 모집단의 일부

· 모수(parameter): 모평균, 모분산, 모비율 등과 같이 모집단을 묘사하거나 규정하는데 도움이 되는 상수

· 통계량(statistic): 표본평균, 표본분산, 표본비율 등과 같이 모수를 추정하기 위하여 관측된 랜덤표본으로부터 계산되는 값

· 추정(estimation): 표본통계량을 기초로 모수에 대한 추정값을 얻는 것

· 점추정(point estimation): 단일 통계량 값으로 모수를 추정하는 것

· 구간추정(interval estimation): 추정값 자체를 표본오차까지 고려하여 구간을 만들어 제시하는 추정방식. 추정량 ± 임계값 × (추정량의 표준편차)

· 신뢰수준(confidence level): 구간추정에서 얻어진 신뢰구간이 모수 추정에 얼마나 신뢰성이 있는지를 나타내는 신뢰도

· 통계적 가설검정(statistical hypothesis testing): 자료를 기초로 실험자가 주장하고 싶은 가설의 옳고 그름을 판정하는 통계절차

· 통계적 가설(statistical hypothesis): 모집단에 대한 주장이나 추측을 말하며, 귀무가설(H_0)과 대립가설(H_1)의 두 가설을 설정하여 자료에 담긴 정보를 토대로 두 가

설 중 어느 하나로 결론을 내리는 형식을 취한다. 이때 실험자가 주장하려는 사실을 대립가설로 설정한다.

· 검정통계량(test statistic): 귀무가설 H_0의 기각 여부를 결정할 때 사용되는 통계량

· 기각역(rejection region): 귀무가설 H_0를 기각시키는 검정통계량의 영역

· 제1종 오류와 제2종 오류

통계적 결정	귀무가설 H_0	
	참	거짓
H_0 채택	옳은 결정(1-α)	제2종 오류(β)
H_1 채택	제1종 오류(α)	옳은 결정(1-β)

· 유의수준(significance level): 제1종 오류를 범하는 최대허용확률

· 유의확률(significance probability): P-value, 실제로는 귀무가설이 참인데 주어진 데이터가 우연히 대립가설을 지지할 확률

2.2 독립표본을 이용한 두 모평균 차이에 대한 추론

	모집단 1	모집단 2
모평균	μ_1	μ_2
모분산	${\sigma_1}^2$	${\sigma_2}^2$
표본의 크기	n_1	n_2
랜덤표본	$x_{11}, x_{12}, \dots, x_{1n_1}$	$x_{21}, x_{22}, \dots, x_{2n_2}$
표준평균	$\bar{x}_1 = \frac{\sum x_{1i}}{n_1}$	$\bar{x}_2 = \frac{\sum x_{2i}}{n_2}$
표본분산	$V_1 = \frac{\sum (x_{1i}-\bar{x}_1)^2}{n_1-1}$	$V_2 = \frac{\sum (x_{2i}-\bar{x}_2)^2}{n_2-1}$

가정 모집단 1은 정규분포 $N(\mu_1, \sigma^2)$를 따르고,
모집단 2는 정규분포 $N(\mu_2, \sigma^2)$를 따른다.

- 공통분산 σ^2은 다음의 합동표본분산(pooled sample variance)으로 추정함

$$s_p{}^2 = \frac{(n_1-1)V_1+(n_2-1)V_2}{n_1+n_2-2}$$

- 검정통계량 : $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ 는

자유도 $(n_1 + n_2 - 2)$ 인 t분포를 따름

2.3 짝지어진 비교

쌍	표본 1	표본 2	차이 $d = x_1 - x_2$
1	x_{11}	x_{21}	$d_1 = x_{11} - x_{21}$
2	x_{12}	x_{22}	$d_2 = x_{12} - x_{22}$
⋮	⋮	⋮	⋮
n	x_{1n}	x_{2n}	$d_n = x_{1n} - x_{2n}$

가정

$$d_1, d_2, \cdots d_n \sim N(\delta, \sigma_\delta{}^2), \qquad \delta = \mu_1 - \mu_2$$

- 귀무가설 : $H_0 : \delta = \delta_0$
- 검정통계량

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}, \ \mathrm{V_d} = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1} \qquad \text{....................} \ (2.6)$$

이라고 하는 경우 $t = \frac{\bar{d} - \delta_0}{s_d/\sqrt{n}}$ 은 자유도 $(n-1)$인 t분포를 따름
(단, $s_d = \sqrt{V_d}$)

풀이

1) 가설의 설정 $H_0: \delta = 0$ VS $H_1: \delta > 0$ $(\delta = \mu_B - \mu_A)$

2) 검정통계량의 값 $\bar{d} = 0.41$, $V_d(= s_d^2) = 0.149$, $s_d = 0.387$

$$\frac{s_d}{\sqrt{n}} = \frac{0.387}{\sqrt{10}} = 0.122$$

$$t = \frac{0.41}{0.122} = 3.4$$

3) 의사결정 $t(9; 0.05) = 1.833 < 3.4$ ➡ 귀무가설 기각

(유의 수준 5%에서 재질 B로 만든 밑창이
재질 A로 만든 밑창보다 더 많이 닳는다고 결론내림)

2.4 두 모집단에서 두 모분산비에 대한 추론

2.2절의 절차(두 평균차 검정)를 적용하기 전에
두 모분산이 같은지 먼저 검정을 해야 한다.

- 모분산이 σ_1^2인 정규분포에서 크기 n_1의 랜덤표본이 추출되고
또 모분산이 σ_2^2인 정규분포에서 크기 n_2의 두 번째 랜덤표본이 추출될 때
표본분산 V_1과 V_2는 각각 σ_1^2 과 σ_2^2의 추정량으로서
다음 통계량은 F분포를 따른다.

$$\frac{V_1/\sigma_1^2}{V_2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$