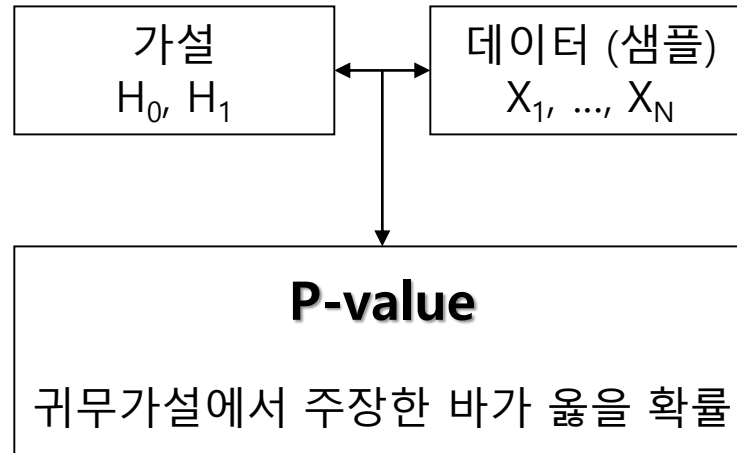


P-Value



- P-Value: 확률값, 0과 1사이로 표준화된 지표
- **귀무가설이 참이라는 가정 아래 얻은 통계량이 귀무가설을 얼마나 지지하는지를 나타낸 확률**
- 귀무가설을 채택할지 기각할지 기준으로 사용할 수 있는 값

P-Value

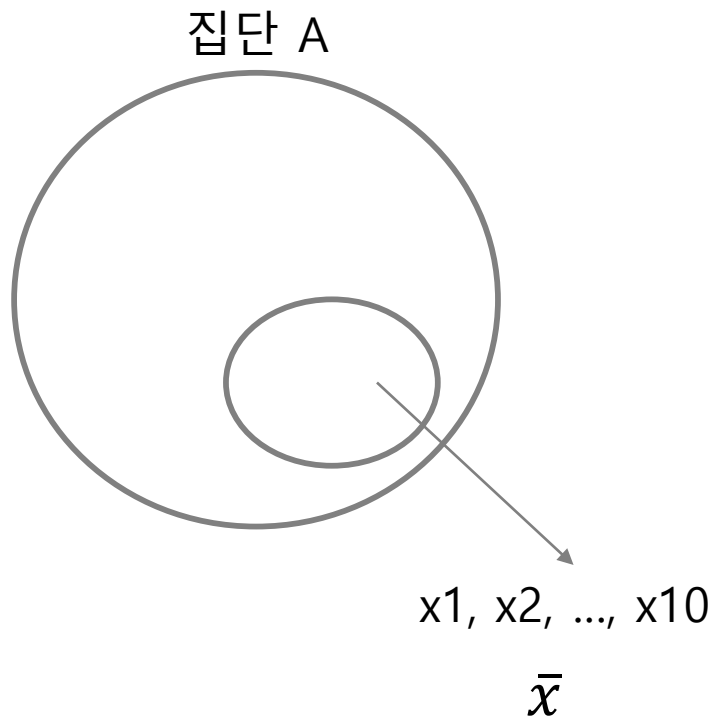
- 작은 P-value → 귀무가설이 참일 확률이 적어짐



- P-value < 0.05 (0.01): 귀무가설이 틀릴 확률이 매우 큼
- P-value > 0.1: 귀무가설이 틀리지 않을 확률이 매우 큼

검정통계량

검정통계량: 귀무가설이 참이라는 가정 아래 얻은 통계량



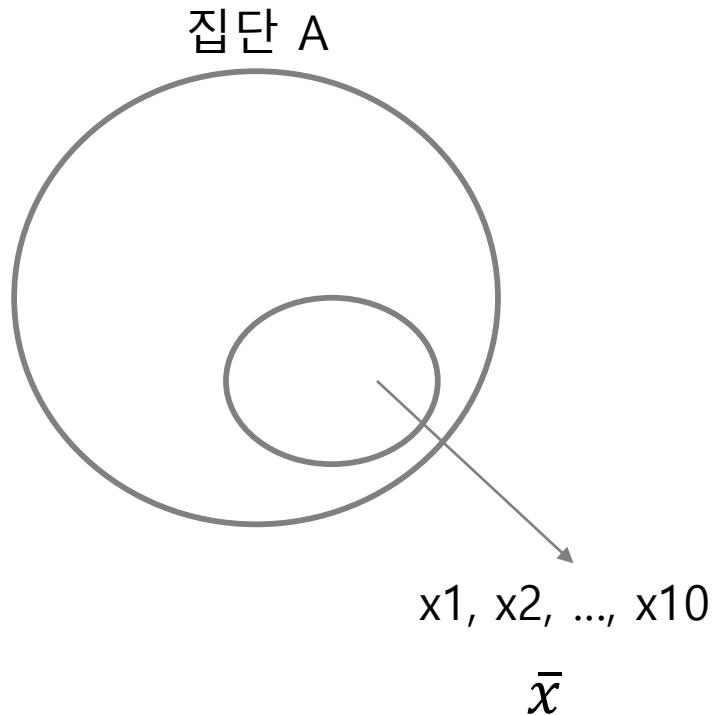
H_0 : 집단 A의 평균이 10

H_1 : 집단 A의 평균이 10이 아님

검정통계량: $T = \bar{x} - 10$

P-value = $P(Y > T), Y \sim \text{분포}$

검정통계량



H_0 : 집단 A의 평균이 10

H_1 : 집단 A의 평균이 10이 아님

검정통계량: $T = \bar{x} - 10$

P-value = $P(Y > T), Y \sim \text{분포}$

P-Value

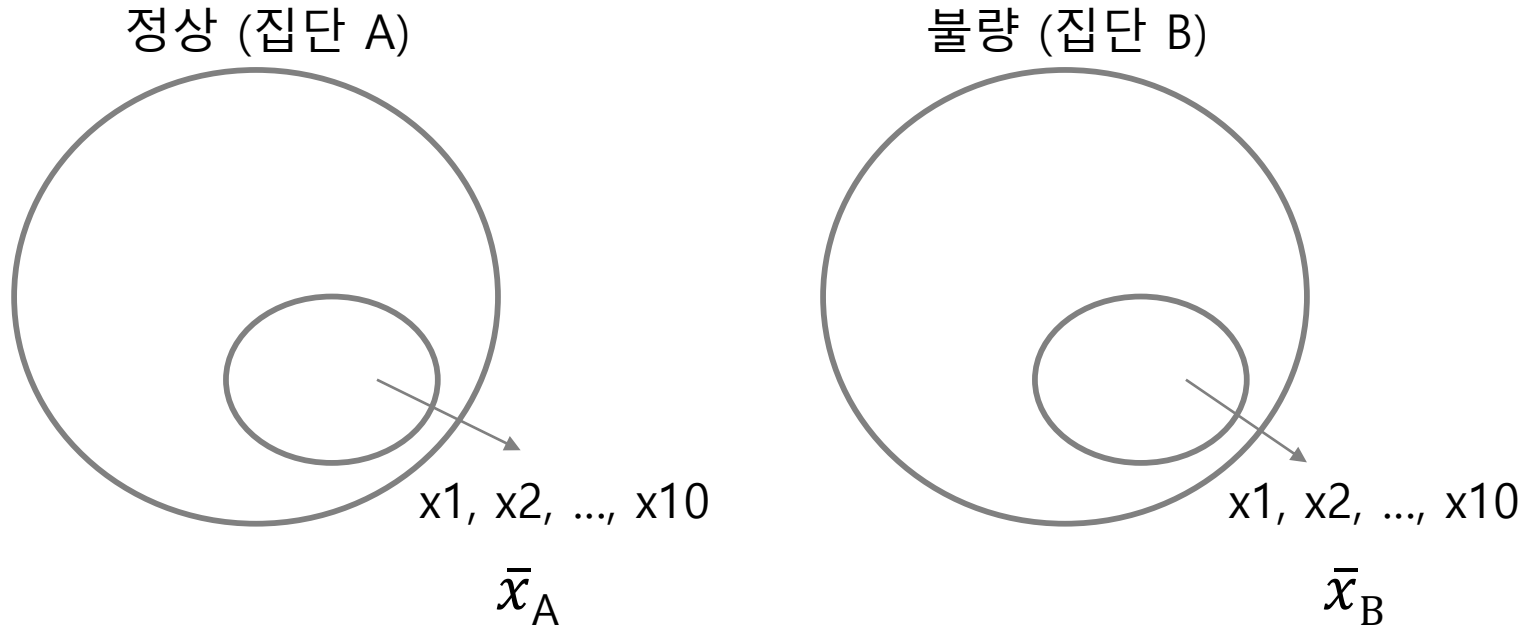
- $n=15, \bar{X}=10.6, s=1.61$

$$H_0 : \mu = 10 \text{ vs. } H_1 : \mu \neq 10$$

$$t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = \frac{10.6 - 10}{1.61 / \sqrt{15}} = 1.44$$

$$\begin{aligned} \text{p-value} &= 2 \times P(X \geq 1.44), \text{ where } X \sim t(15-1) \\ &= 2 \times 0.086 \\ &= 0.176 \end{aligned}$$

검정통계량



H_0 : 정상집단과 불량집단의 평균이 같음

H_1 : 정상집단과 불량집단의 평균이 다름

검정통계량: $T = (\bar{x}_A - \bar{x}_B) - 0$

P-value = $P(Y > T)$, $Y \sim$ 분포

두 집단의 비교 예제

정상	X
불량	75
	85
	90
	88
	96
	78
	82
	84
정상	65
	70
	68
	74
	80
	84
	70
	75
불량	64
	60

변수 x는 정상과 불량상태에서 좌측과 같은 데이터를 생성하였다. 정상상태에서의 값이 불량상태에서의 값보다 크다고 할 수 있는가?

정상상태에서의 평균값을 μ_A , 불량상태에서의 평균값을 μ_B 라고 하자.

$$H_0: \mu_A = \mu_B \text{ vs } H_1: \mu_A > \mu_B$$

$$T^* = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{S_A^2/n_A + S_B^2/n_B}} = \frac{(84.75 - 71) - 0}{\sqrt{44.78/8 + 54.67/10}} = 4.134$$

$$P\text{-value} = P(X > 4.134) = 0.003, X \sim t(16 = 8 + 10 - 2)$$

귀무가설 기각. 정상상태에서의 값이 불량상태에서의 값보다 크다. 즉, 변수 x는 정상과 불량을 나누는 데 중요한 변수이다.

선형회귀모델 기울기에 대한 가설검정

- 알려지지 않은 파라미터에 대한 가설을 세우고 이를 검정
- 일종오류 α 하에서 기울기가 0인지 아닌지 검정

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

$$t^* = \frac{\hat{\beta}_1 - 0}{sd\{\hat{\beta}_1\}}$$

If $|t^*| > t_{\alpha/2, n-2}$, we reject H_0

P-value = $2 \cdot P(T > |t^*|)$ where $T \sim t(n-2)$

선형회귀모델에서의 분산분석

판매원 수 (X_1)	광고비 (X_2)	월간 매출액 (Y)
14	37	850
16	43	970
13	38	730
10	42	940
18	36	920
17	33	830
16	40	940
15	35	900
11	34	760
10	29	710

Source	DF	SS	MS	F	P-value
Model	2	54809.18	27404.59	7.540	0
Error	7	25440.82	3634.40		
Total	9	80250.00			

$$H_0: \beta_1 = \beta_2 = 0 \text{ vs } H_1: \text{적어도 하나의 } \beta \neq 0$$

$$F^* = \frac{MSR}{MSE} = \frac{54809.18/2}{25440.82/7} = \frac{27404.59}{3634.40} = 7.540$$

$$\text{P-value} = P(Y \geq 7.540) \approx 0, \text{ where } Y \sim F(2,7)$$

적어도 하나의 $\beta \neq 0$ (판매원 수 혹은 광고비 혹은 두가지 모두 유의미)