제9장 이변량 자료의 추정

Jinseog Kim
Dongguk University
jinseog.kim@gmail.com

2017-09-13

모평균의 차이에 대한 추정

- 독립(independent)표본 : 하나의 모집단에서 표본과 다른 모집단에서 표본이 전혀 관련이 없는 경우
  - 두개의 서로 다른 공장에서 만든 제품의 품질
  - 두 학교 학생들의 성적

모평균의 차이에 대한 추정

- 두 정규모집단에서의 랜덤표본

$$X_1, X_2, ..., X_n \sim N(\mu_1, \sigma_1^2),$$
$$Y_1, Y_2, ..., Y_m \sim N(\mu_2, \sigma_2^2)$$

- 두 모집단의 모평균(μ_1, μ_2)의 추정

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^{n} Y_i.$$

- 두 모집단의 모평균 차$(\mu_1 - \mu_2)$의 추정

$$\bar{X} - \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{n} Y_i.$$

모평균의 차이에 대한 추정 - conti.

- $\bar{X} - \bar{Y}$의 기대값 (불편성)

$$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2.$$

- $\bar{X} - \bar{Y}$의 분산

$$V(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}, \quad X_i\text{와 } Y_i\text{는 서로 독립이므로}$$

- $\bar{X} - \bar{Y}$의 분포
  1. 각각 서로 독립인 정규분포를 따를 때

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

  2. 각각 서로 독립인 분포, 정규분포가 아니지만 충분히 자료의 수가 많은 경우 (중심극한정리)

$$\bar{X} - \bar{Y} \approx N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

신뢰구간

- 표준화된 변수의 분포는 표준정규분포

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \sim N(0,1).$$

- 표준정규분포의 임계값(z_α)을 이용

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} < z_{\alpha/2}\right) = 1-\alpha.$$

- 신뢰구간

$$\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}},\ \ \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

예제 10

- 두 종류의 자동차 타이어에 대해 마모 정도의 차이 비교
- 각 종류별로 100개씩을 표본추출하여 마모될 때 까지의 주행거리 측정
  - 타이어 1 : $\bar{x}_1 = 26,400, \quad s_1^2 = 1,440,000$
  - 타이어 2 : $\bar{x}_2 = 25,100, \quad s_2^2 = 1,960,000$
- 두 종류의 타이어에 대해 평균 사용거리의 차이를 추정, 추정오차의 한계?
  - 차이의 점추정

$$\bar{x}_1 - \bar{x}_2 = 26,400 - 25,100 = 1,300$$

  - 95% 오차의 한계

$$1.96\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}} = 1.96\sqrt{34,000} = 361.4061$$

  - 95% 신뢰구간 - 타이어 1이 타이어 2에 비해 품질이 우수함

$$1,300 \pm 1.96\sqrt{34,000} = (938.594,\ 1,661.406)$$

두 개의 이항모집단 차이의 추정

- 두 이항 모집단에서의 랜덤표본

$$X_1, X_2, ..., X_n \sim Bernoulli(p_1), \text{또는 } X = \sum_{i=1}^{n} X_i \sim Bin(n, p_1)$$

$$Y_1, Y_2, ..., Y_m \sim Bernoulli(p_2), \text{또는 } Y = \sum_{i=1}^{m} X_i \sim Bin(m, p_2)$$

- 두 모집단의 모비율 (p_1, p_2)의 추정

$$\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X}{n}, \quad \hat{p}_2 = \frac{1}{m}\sum_{i=1}^{n} Y_i = \frac{Y}{m}.$$

- 두 모집단의 모비율 차$(p_1 - p_2)$의 추정

$$\hat{p}_1 - \hat{p}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{n} Y_i = X/n - Y/m.$$

두 개의 이항모집단 차이의 추정 - conti.

- $\hat{p}_1 - \hat{p}_2$의 기대값 (불편성)

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2.$$

- $\hat{p}_1 - \hat{p}_2$의 분산

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}, \quad X\text{와 } Y\text{는 서로 독립이므로}$$

- $\hat{p}_1 - \hat{p}_2$의 분포
  1. 각각 서로 독립인 정규분포를 따를 때

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}} \sim N(0,1).$$

  2. 각각 서로 독립인 분포, 충분히 자료의 수가 많은 경우 (중심극한정리)

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} \sim N(0,1).$$

신뢰구간

- 표준정규분포의 임계값(z_α)을 이용

$$P\left(-z_{\alpha/2} < \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} < z_{\alpha/2}\right) = 1 - \alpha.$$

- 신뢰구간

$$\left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}},\ \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right)$$

예제 13

살충제를 생산하는 제조업자가 자체 개발한 두 종류의 살충제 A, B의 효능을 비교하고자 한다. 동 일한 조건의 2개 방에 1,000마리의 같은 해충을 넣은 후, 첫 번째 방은 살충제 A를, 두 번째 방은 살충제 B를 같은 양만큼 사용하였다. 그 결과 두 곳의 방에서 각각 825마리와 760마리의 해충이 죽었다. 이 두 종류의 살충제에 대한 살충률의 차이를 추정하고, 이에 대한 95% 신뢰구간을 구하시오

- 신뢰수준 95% : $z_{0.05/2} = 1.96$ (정규분포표)
- 차이에 대한 점추정 :

$$\hat{p}_1 - \hat{p}_2 = 825/1000 - 760/1000 = 0.065$$

- 구간추정

$$z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} = 1.96 \times \sqrt{\frac{0.825(1-0.825)}{1000} + \frac{0.76(1-0.76)}{1000}} = 0.0354$$

$$0.065 \pm 0.0354 = (0.0296, 0.1004)$$