4강 단순임의추출법(2)

정보통계학과 이기재교수

학/습/목/차

1. 모총계의 추정
2. 모비율의 추정
3. 표본의 크기 결정
4. 엑셀을 활용한 실습

모총계의 추정

1. 모총계

▶ $\tau = \sum_{i=1}^{N} y_i = N\mu$ (모평균의 함수)

(예) 전국의 쌀 생산량, 가구의 연간 총 전기소비량

2. 모총계의 추정량

▶ $\hat{\tau} = N\hat{\mu} = N\bar{y}$ (모총계 추정량은 모총계에 대한 비편향추정량임)

▶ 모총계 추정량에 대한 분산 : $V(\hat{\tau}) = V(N\bar{y}) = N^2 V(\bar{y}) = N^2 \frac{N-n}{N} \frac{S^2}{n}$

▶ 분산의 추정량 : $\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}) = N^2 \hat{V}(\bar{y}) = N^2 \frac{N-n}{N} \frac{s^2}{n}$

▶ 모총계 τ의 $100(1-\alpha)\%$신뢰구간 : $\hat{\tau} \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{\tau})}$

모총계의 추정

예제 2-5

▶ 어느 지역의 인구수 추정

- $N = 6{,}000$ 구획 중 $n = 100$ 구획을 단순임의추출
  $\bar{y} = 25.2, \quad s^2 = 136.0$

- 모총계 추정값 : $\hat{\tau} = N \cdot \bar{y}$
  $= 6000 \times 25.2 = 151{,}200$(명)

- 추정분산의 값 : $\hat{V}(\hat{\tau}) = N^2 \frac{N-n}{N} \frac{s^2}{n}$
  $= 6000^2 \frac{6000-100}{6000} \frac{136.0}{100} = 48{,}144{,}000$

- 총 인구수에 대한 95% 신뢰구간
  $151{,}200 \pm 2 \times \sqrt{48{,}144{,}000} \leftrightarrow 151{,}200 \pm 1387$

학/습/목/차

1. 모총계의 추정
2. 모비율의 추정
3. 표본의 크기 결정
4. 엑셀을 활용한 실습

모비율의 추정

1. 모비율

▶ $p = \frac{1}{N}\sum_{i=1}^{N} y_i \ , \quad y_i = 0 \text{ or } 1$

(예) 후보의 지지율, 제품의 시장점유율, TV 프로그램의 시청률, 제품 불량률

2. 모비율의 추정량

▶ $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i$ (모비율 추정량은 비편향추정량임)

▶ 모비율 추정량에 대한 분산 : $V(\hat{p}) = \frac{N-n}{N}\frac{pq}{n}$, 단, $q = 1-p$

▶ 분산의 추정량 : $\hat{V}(\hat{p}) = \frac{N-n}{N}\frac{\hat{p}\hat{q}}{n-1}$, 단, $\hat{q} = 1-\hat{p}$

▶ 모비율의 $100(1-\alpha)\%$신뢰구간 : $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{p})}$

표본비율의 분산 추정식 유도

- 모비율 추정량인 표본비율 $\hat{p}$은 각 표본단위에서 관측한 값이 0 또는 1인 경우에 대한 표본평균과 동일함

- 추정량 $\hat{p}$의 분산식은 $\hat{V}(\bar{y}) = \frac{N-n}{N}\frac{s^2}{n}$ 에서 유도됨

  "각 표본단위에서 관측한 y_i이 0 이나 1의 값을 갖는 경우에 해당함"

- $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n\hat{p}\hat{q}}{n-1}$ 대입

$$\Rightarrow \hat{V}(\hat{p}) = \frac{N-n}{N}\frac{\hat{p}\hat{q}}{n-1}$$

모비율의 추정

예제 2-6

▶ 단순임의표본 1500명 중 936명이 지지하는 후보자의 지지율 추정

- 모비율 추정값 : $\hat{p} = \dfrac{\sum_{i=1}^{n} y_i}{n} = \dfrac{936}{1500} = 0.624$
- 분산 추정값 : $\hat{V}(\hat{p}) = \dfrac{N-n}{N}\dfrac{\hat{p}\hat{q}}{n-1} \quad \leftarrow \quad \dfrac{N-n}{N} \fallingdotseq 1$
- 95% 신뢰구간 : $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{p})} \leftrightarrow 0.624 \pm 0.0250$

학/습/목/차

1. 모총계의 추정
2. 모비율의 추정
3. 표본의 크기 결정
4. 엑셀을 활용한 실습

표본의 크기결정

- 기본원칙
  - ▶ 목표정도(허용오차)의 범위 내에서 가능한 한 표본크기 작게 함

    ⇨ 목표정도(target precision) : 표본조사에서 목표하는 허용오차의 한계

- 목표정도의 형태
  - ▶ 절대오차의 한계 : 추정량의 오차한계가 일정 값 이내
  - ▶ 상대오차의 한계 : 추정량의 변동계수가 일정 비율 이내

    ⇨ 표본의 크기 ⇔ 추정의 정확성 ⇔ 조사비용

절대오차의 한계 B가 주어진 경우

1. 모평균 추정

▶ $B = z_{\alpha/2}\sqrt{\frac{N-n}{N}\frac{S^2}{n}}$

▶ $n = \frac{N(z_{\alpha/2}S)^2}{NB^2 + (z_{\alpha/2}S)^2} = \frac{n_0}{1+\frac{n_0}{N}}$, 단, $n_0 = \frac{(z_{\alpha/2}S)^2}{B^2}$

⇨ S^2 의 파악 → 과거 자료에서 얻은 s^2 을 이용하는 방법

작은 규모의 예비조사를 통해 s^2 을 추정하는 방법

절대오차의 한계 B가 주어진 경우

예제 2-7

- $N = 109, \quad S = 23,263$
- $B = 10,000$ 이며, $\quad z_{0.025} \fallingdotseq 2$
- $n_0 = (2 \times 23,263)^2 / 10,000^2 \fallingdotseq 21.6$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{21.6}{1 + \frac{21.6}{109}} \fallingdotseq 18.0$$

절대오차의 한계 B가 주어진 경우

2. 모비율 추정

- $B = z_{\alpha/2}\sqrt{\frac{N-n}{N}\frac{pq}{n-1}}$
- $n = \frac{N z_{\alpha/2}^2 pq}{NB^2 + z_{\alpha/2}^2 pq} = \frac{n_0}{1+\frac{n_0}{N}}$, $n_0 = \frac{z_{\alpha/2}^2 pq}{B^2}$

절대오차의 한계 B가 주어진 경우

예제 2-8

- $N = 3000$, 예비조사 흡연률 $\hat{p} = \dfrac{30}{50} = 0.6$

- $$n_0 = \frac{4 \times \hat{p}\hat{q}}{B^2} = \frac{4 \times 0.6 \times 0.4}{0.05^2} = 384$$

- $$n = \frac{n_0}{1 + \dfrac{n_0}{N}} = \frac{384}{1 + \dfrac{384}{3000}} \fallingdotseq 340$$

절대오차의 한계 B가 주어진 경우

3. 모총계 추정

- $$B = z_{\alpha/2}\sqrt{N^2 \frac{N-n}{N}\frac{S^2}{n}}$$

- $$n = \frac{N^2(z_{\alpha/2}S)^2}{B^2 + N(z_{\alpha/2}S)^2} = \frac{n_0}{1+\frac{n_0}{N}}$$, 단, $n_0 = N^2\frac{(z_{\alpha/2}S)^2}{B^2}$

절대오차의 한계 B가 주어진 경우

예제 2-9

▶ $N = 10{,}000$, $S = 3{,}000$, $B = 5{,}000{,}000$

▶ $$n_0 = N^2 \bullet \frac{(z_{\alpha/2}S)^2}{B^2} = 10{,}000^2 \bullet \frac{(2 \times 3{,}000)^2}{(5{,}000{,}000)^2} = 144$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{144}{1 + \frac{144}{10000}} \fallingdotseq 142$$

상대오차의 한계 D가 주어진 경우

1. 모평균 추정

- 상대오차의 한계 : $D = \dfrac{B}{\mu}$
- 추정량의 상대표준오차(relative standard error)
- $RSE(\bar{y}) = \dfrac{\sqrt{V(\bar{y})}}{E(\bar{y})} = \dfrac{\sqrt{\dfrac{N-n}{N}\dfrac{S^2}{n}}}{\mu}$
- $D = z_{\alpha/2} RSE(\bar{y}) = z_{\alpha/2} \dfrac{\sqrt{\dfrac{N-n}{N}\dfrac{S^2}{n}}}{\mu}$
- $n = \dfrac{N(z_{\alpha/2}CV)^2}{ND^2 + (z_{\alpha/2}CV)^2} = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$, 단, $n_0 = \left(\dfrac{z_{\alpha/2}CV}{D}\right)^2$이고 , $CV = \dfrac{S}{\mu}$ 임

상대오차의 한계 D가 주어진 경우

예제 2-10

- $N = 10{,}000$, $\hat{C} = \dfrac{s}{\bar{y}} = 0.1$, $D = 0.025$

- $$n_0 = \left(\frac{z_{0.025} C}{D}\right)^2 = \left(\frac{2 \times 0.10}{0.025}\right)^2 = 64$$

$$n = \frac{n_0}{1 + \dfrac{n_0}{N}} = \frac{64}{1 + \dfrac{64}{10000}} \fallingdotseq 64$$

학/습/목/차

1. 모총계의 추정
2. 모비율의 추정
3. 표본의 크기 결정
4. 엑셀을 활용한 실습

〈실습하기〉에서 자세히 다룸

이 강의는

강의용 휴대폰(U-KNOU 서비스 휴대폰)으로도

다시 볼 수 있습니다.