

3강 단순임의추출법(1)

정보통계학과 이기재교수

학/습/목/차

1. 개요

2. 단순임의표본의 추출

3. 모수의 추정

4. 엑셀을 활용한 실습

복원 vs 비복원

1. 복원추출(sampling with replacement)

- ▶ 동일한 추출단위가 표본에 반복적으로 추출되는 것이 가능한 경우

2. 비복원추출(sampling without replacement)

- ▶ 동일한 추출단위는 표본에 한번 이상 추출되지 않는 경우

〈참고〉 무한모집단 : 복원추출 \approx 비복원추출

단순임의추출법의 특징

- ▶ 가장 단순한 방법의 확률추출법으로 다른 표본추출법의 이론적인 기초가 됨
- ▶ 소규모 조사나 예비조사(pilot survey)에서 주로 사용됨
- ▶ 모집단의 어느 부분도 과대하게 또는 과소하게 반영하지 않음

☞ 예비조사는 소규모로 실시하여 조사표가 잘 작성되었는가? 조사방법과 조사과정은 적절한가? 등을 검토하기 위하여 실시됨. 표본크기를 정하고자 할 때 과거 조사 결과나 비슷한 조사가 없어 모분산을 모르는 경우에 이를 추정하기 위해서 예비조사를 실시되기도 함

비복원단순임의추출법

(simple random sampling without replacement: SRSWOR)

크기가 N 인 모집단에서 크기가 n 인 표본을 추출하고자 한다면 $\binom{N}{n}$ 가지의 추출 가능한 표본들이 존재하게 된다.



이들 $\binom{N}{n}$ 가지의 표본에 대해 추출확률을 모두 동일하게 해주는 표본 추출방법을 단순임의추출법이라고 한다.

❖ 예제 2-1

▶ 모집단 $U = \{1, 2, 3, 4\}$ 에서 크기 2인 표본을 단순임의 추출할 경우 가능한 표본

$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$



위의 6가지 표본 각각이 추출될 확률은 $1/6$ 로 동일

단순임의추출법의 예

1. 정부의 어떤 정책에 대한 여론조사
 - ▶ 각 지역별로 단순임의추출
2. 선거여론조사를 위한 전화조사
 - ▶ 단순임의추출에 의해 전화번호 선정
3. 시장조사(marketing research)
 - ▶ 자동차의 색상에 대한 잠재고객들이 선호도 조사
4. 생산 제품의 불량률조사
 - ▶ 생산 제품 중 일부를 단순임의추출하여 조사

단순임의추출법의 예

〈참고〉 임의(任意, random) 추출이란?

- ▶ 아무렇게나 추출한다는 뜻이 아님
- ▶ 추출 시 일체의 작위(作爲)를 배제한다는 뜻
- ▶ 추출자의 선택 편향을 제거하기 위한 수단



학/습/목/차

1. 개요
2. 단순임의표본의 추출
3. 모수의 추정
4. 엑셀을 활용한 실습

표본추출 방법

1

모든 추출단위들에 대해 1에서 N까지의 일련번호 부여



2

1에서 N까지의 값이 나올 확률이 $1/N$ 로 동일하게 하는 하나의 난수(random number)를 추출



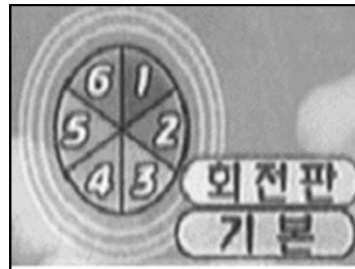
3

추출된 난수의 번호에 해당하는 추출단위를 표본으로 선택

〈참고〉 난수발생장치(randomizing device)

4245 3003 4264 1374 0180 7953 5429 2875 3683 2195
6436 2359 5116 9106 9612 3071 4479 5155 4799 1389
2836 9434 5734 8270 1097 9655 4775 9293 8255 8815
3522 5539 0220 4672 1701 2551 7408 6215 4012 6094
6126 9027 0893 9637 8688 4529 6962 7205 5796 8714
4748 4084 8138 8235 0879 9074 9672 7229 7171 5272
0376 8202 9424 0607 2583 6334 0564 8889 3565 3364
1764 7311 4305 8364 4679 6015 6942 3841 2867 9011
6707 2580 2368 1136 4363 5222 7091 6040 9280 2128
7759 8132 7014 9671 0907 4365 2966 5910 2372 2586
9337 2138 5151 7977 5395 6542 5546 6634 7451 6092

〈난수표〉



〈회전판〉



〈구슬상자〉 등

표본추출 방법

✚ 예제 2-1

▶ N = 450 명 중 n = 20 명을 단순임의추출(난수표 활용)

행 \ 열	(1)	(2)	(3)	(4)
1	10480	15011	01536	02011
2	22368	46573	25593	85393
3	24130	48360	22527	97265
4	42167	93093	06243	61680
5	37570	39975	81837	16656
6	77921	06907	11008	42751
7	99562	72905	56420	69994
8	96301	91977	05463	07972
9	89579	14342	63661	10281
10	85475	36857	53342	53988

학/습/목/차

1. 개요

2. 단순임의표본의 추출

3. 모수의 추정

4. 엑셀을 활용한 실습

모수의 추정

1. 모집단

▶ y_1, y_2, \dots, y_N 로 표시

2. 모수

▶ 모평균 : $\mu = \frac{1}{N} \sum_{i=1}^N y_i$

▶ 모총계 : $\tau = \sum_{i=1}^N y_i = N\mu$

▶ 모분산 : $S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2$

모평균의 추정

- ▶ 단순임의표본의 데이터 : y_1, y_2, \dots, y_n
- ▶ 모평균 추정량 = 표본평균 : $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

✚ 정리 2-1

1 크기 N 의 모집단에서 크기 n 인 단순임의표본(비복원)을 추출하는 경우에 표본평균 \bar{y} 는 모평균 μ 의 비편향추정량임

2 모평균의 추정량인 표본평균 \bar{y} 의 분산은 $V(\bar{y}) = \frac{N-n}{N} \frac{S^2}{n}$

모평균의 추정

❖ 예제 2-3

▶ 모집단 : { 1, 2, 3, 4 } (단위 : 백만원)

▶ \bar{y} 의 확률분포

표본	\bar{y}	$P(\bar{y})$
{ 1, 2 }	1.5	1/6
{ 1, 3 }	2.0	1/6
{ 1, 4 }	2.5	1/6
{ 2, 3 }	2.5	1/6
{ 2, 4 }	3.0	1/6
{ 3, 4 }	3.5	1/6

$$E(\bar{y}) = \sum \bar{y} \cdot P(\bar{y}) = (1.5 + 2.0 + \dots + 3.5) \cdot \frac{1}{6} = 2.5$$

$$\begin{aligned} V(\bar{y}) &= E(\bar{y} - \mu)^2 = \sum (\bar{y} - \mu)^2 \cdot P(\bar{y}) \\ &= \sum (\bar{y})^2 \cdot P(\bar{y}) - \mu^2 \\ &= \frac{5}{12} \end{aligned}$$

$$V(\bar{y}) = \frac{N-n}{N} \frac{S^2}{n} = \frac{4-2}{4} \frac{1}{2} \frac{5}{3} = \frac{5}{12}$$

분산의 추정

- ▶ $\hat{V}(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n}$ 단, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$
- ▶ 표본의 크기(n)이 커질수록 분산추정값은 줄어듦
- ▶ 유한모집단수정항(finite population correction: fpc) : $\frac{N-n}{N}$
- ▶ fpc를 무시하면 기초통계학에서의 추정량 $\hat{V}(\bar{y}) = \frac{s^2}{n}$ 과 동일

1. 모평균 μ 에 대한 $100(1-\alpha)\%$ 신뢰구간

- ▶ $(\bar{y} - z_{\alpha/2} \sqrt{\hat{V}(\bar{y})}, \bar{y} + z_{\alpha/2} \sqrt{\hat{V}(\bar{y})})$

2. $100(1-\alpha)\%$ 신뢰도에서의 오차한계

- ▶ $B = z_{\alpha/2} \sqrt{\hat{V}(\bar{y})}$

분산의 추정

예제 2-4

▶ 표본데이터

3 4 3 3 5

3 4 4 3 5

6 3 4 3 4

3 3 4 4 3

$$\begin{aligned}\hat{V}(\bar{y}) &= \frac{N-n}{N} \cdot \frac{s^2}{n} \\ &= \frac{1000-20}{1000} \cdot \frac{0.7474}{20} = 0.0366\end{aligned}$$

- 평균 가족 수에 대한 95% 신뢰구간

$$\begin{aligned}\bar{y} \pm z_{\alpha/2} \cdot \sqrt{\hat{V}(\bar{y})} \\ \leftrightarrow 3.7 \pm 2 \sqrt{0.0366} \\ \leftrightarrow 3.7 \pm 0.38\end{aligned}$$

학/습/목/차

1. 개요

2. 단순임의표본의 추출

3. 모수의 추정

4. 엑셀을 활용한 실습

↳ <실습하기>에서 자세히 다룸



Korea National Open University
이 강의는
강의용 휴대폰(U-KNOU 서비스 휴대폰)으로도
다시 볼 수 있습니다.

다시 볼 수 있습니다.