표본론 졸업시험 2021.5.22

[1] "모평균에 대한 비편향추정량(unbiased estimator)을 얻으려면 모집단의 원소가 추출될 확률이 모두 같아지도록 표본을 추출해야 한다"라는 주장에 대해 논하시오.

[2] 선거결과 예측을 위한 어떤 전화조사에서,

> "전화는 전화번호부에서 선택해 걸지 않고 랜덤다이얼링 방식으로 걸었다. 전화를 받지 않으면 다른 시간대에 최대 두 번까지 더 전화를 걸었다. 응답 협조를 거부하는 경우나 중간에 통화가 끊긴 경우 무응답으로 간주하였다. 전화조사의 응답률은 10%이었는데, 계획한 표본 크기 1000명을 확보하기 위해 같은 방법으로 추가로 전화를 걸어 목표한 표본 크기를 다 조사하였으며, 이 때 인구총조사에서 알려진 연령별, 남녀 비율을 맞추어서 조사하였다. 95% 신뢰수준에서 이 조사의 오차의 한계는 약 3.2%이다"

라고 발표하였다. 오차의 한계 3.2%는 어떻게 계산된 것인가를 설명하고, 이 오차의 한계를 얼마나 믿을 수 있는지에 대해서도 논하시오. 오차의 한계에 대한 신빙성을 논할 때 다음 핵심어가 들어가도록 하시오: **할당표집**, **표집오차, 무응답층, 편향**

[3] 변수 y에 대한 모집단 총합을 추정하고자 할 때, y의 관측값만 이용하여 추정하지 않고 또 다른 변수 x를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)을 적용하기도 한다. 이 방법들이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건(들)이 만족될 것 같은 구체적 예를 들어 보시오.

[4] 층화표집(stratified random sampling)과 군집표집(cluster sampling)을 비교하여 설명하고, 이 두 방법이 혼합된 표집설계의 예를 들어 보시오.

[5] 재학 중인 숭실대학교 학생을 대표할 수 있는 표본 500명을 학기 중에 추출하는 방법을 제시하시오. 단, 대면수업이 정상적으로 진행되고 있다고 가정하시오.

표본론 졸업시험 2020.12.28

[1] 19대 대통령선거 결과 예측을 위해 2017년 5월 1일 한국갤럽조사연구소에서 1015명을 대상으로 실시한 전화조사에서 자유한국당 홍준표 후보 지지율이 16%, 국민의당 안철수 후보 지지율이 20%이었다.

[1-1] 만약 1015명을 단순확률표본(simple random sample)이라고 가정할 수 있다면, 조사 당시 두 후보의 지지율에 유의한 차이가 있다고 할 수 있는가?

[1-2] 조사한 1015명을 단순확률표본으로 가정할 수 있는지에 대해 논하시오,

[2] 표본크기가 n인 선거예측조사에서 한 후보의 지지율에 대한 오차의 한계를 대략 $1/\sqrt{n}$, 두 후보의 지지율 차이에 대한 오차의 한계를 대략 $2/\sqrt{n}$로 둘 수 있는 이유를 각각 설명해 보시오. 신뢰수준은 95%이다. 예를 들어 19대 대선에 관한 여론조사에서 1100명을 조사했을 경우에 95% 신뢰수준에서 오차의 한계를 $1/\sqrt{1100}$ 인 3%라고 한다. 그리고 A후보와 B후보의 지지율 차이가 이 값의 2배인 6% 이하이면 오차범위 내 경합, 6%를 벗어나면 오차의 한계를 벗어난 차이를 보인다고 흔히 발표하거나 판단한다. 그 근거가 무엇인지를 설명하시오.

[3] 이단계군집추출(two-stage cluster sampling)로 표본추출하는 방법을 구체적 사례를 들어 설명하시오. 그리고 이 사례에서 이단계군집추출이 다른 추출방법과 비교했을 때 어떤 장점을 갖는지를 설명하시오.

[4] 변수 y에 대한 모집단 총합을 추정하고자 할 때, y의 관측값만 이용하여 추정하지 않고 또 다른 변수 x를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)을 적용하기도 한다. 이 방법들이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건(들)이 만족될 것 같은 구체적 예를 들어 보시오.

[5] 재학 중인 숭실대학교 학생을 대표할 수 있는 표본 500명을 학기 중에 추출하는 방법을 제시하시오. 단, 대면수업이 정상적으로 진행되고 있다고 가정하시오.

표본론 졸업시험 2020.6.27

[1] 할당표집(quota sampling)이 무엇이며 어떤 경우에 적절한 표집 방법이 되지 않는지에 대해 논하시오.

[2] '층화추출'(stratified sampling)과 '네이먼 배분'(Neyman allocation)에 대해 각각 설명하시오.

[3] 표본크기가 n인 선거예측조사에서 한 후보의 지지율에 대한 오차의 한계를 대략 $1/\sqrt{n}$, 두 후보의 지지율 차이에 대한 오차의 한계를 대략 $2/\sqrt{n}$로 둘 수 있는 이유를 각각 설명해 보시오. 신뢰수준은 95%이다. 예를 들어 19대 대선에 관한 여론조사에서 1100명을 조사했을 경우에 95% 신뢰수준에서 오차의 한계를 $1/\sqrt{1100}$ 인 3%라고 한다. 그리고 A후보와 B후보의 지지율 차이가 이 값의 2배인 6% 이하이면 오차범위 내 경합, 6%를 벗어나면 오차의 한계를 벗어난 차이를 보인다고 흔히 발표하거나 판단한다. 그 근거가 무엇인지를 설명하시오.

[4] 변수 y에 대한 모집단 총합을 추정하고자 할 때, y의 관측값만 이용하여 추정하지 않고 또 다른 변수 x를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건(들)이 만족될 것 같은 구체적 예를 들어 보시오.

[5] 재학 중인 숭실대학교 학생을 대표할 수 있는 표본 500명을 학기 중에 추출하는 방법을 제시하시오. 단, 대면수업이 정상적으로 진행되고 있다고 가정하시오.

표본론 졸업시험 2019.9.7

[1] 선거결과 예측을 위한 어떤 전화조사에서, "전화는 전화번호부에서 선택해 걸지 않고 랜덤다이얼링 방식으로 걸었다. 그리고 전화를 받지 않을 경우 다른 시간대에 두 번까지 더 전화를 걸었는데 그래도 전화를 받지 않으면 무응답으로 간주하였다. 응답 협조를 거부하는 경우도 무응답으로 간주하였다. 전화조사의 응답률은 10%이었는데, 계획한 표본 크기 1000명을 확보하기 위해 같은 방법으로 추가로 전화를 걸어 목표한 표본 크기를 다 조사하였으며, 이 때 인구총조사에서 알려진 연령별, 남녀 비율을 맞추어서 조사하였다. 조사 후 95% 신뢰수준에서 오차의 한계는 약 3.2%이다"라고 발표하였다. 오차의 한계 3.2%는 어떻게 계산된 것인가를 설명하고, 이 오차의 한계를 얼마나 믿을 수 있는지에 대해서도 논하시오. 오차의 한계에 대한 신빙성을 논할 때 다음 핵심어가 들어가도록 하시오: **표집오차, 무응답층, 편향**

[2] 19대 대통령선거 결과 예측을 위해 2017년 5월 1일 한국갤럽조사연구소에서 1015명을 대상으로 실시한 전화조사에서 자유한국당 홍준표 후보 지지율이 16%, 국민의당 안철수 후보 지지율이 20%이었다. 만약 1015명을 단순확률표본(simple random sample)이라고 가정할 수 있다면, 조사 당시 두 후보의 지지율에 유의한 차이가 있다고 할 수 있는가?

[3] 표본을 추출할 때 반드시 모집단 모든 원소가 추출될 확률이 같아지도록 해야 하는 것은 아니다. 층화확률표집(stratified random sampling)에서 모집단 원소가 추출될 확률이 모두 같아지지는 않는 추출방법 예를 들고, 이렇게 얻은 표본으로부터 단순평균 $\sum_{i=1}^{L}\sum_{j=1}^{n_i} y_{ij}/n$을 구해 모평균을 추정하면 안 되는 이유를 설명하시오. 위 식에서 y_{ij}는 i번째 층에서 얻은 j번째 관측값을 나타내며, n_i는 i번째 층에서 추출한 표본의 크기이다. 그리고 $n = \sum_{i=1}^{L} n_i$ 이다.

[4] 이단계군집표집(two-stage cluster sampling) 방법을 구체적 예로써 설명하시오.

[5] 층화표집(stratified random sampling)과 군집표집(cluster sampling)을 비교하여 설명하고, 이 두 방법이 혼합된 표집설계의 예를 들어 보시오.

표본론 졸업시험 2019.3.9

[1] "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 추출될 확률이 모두 같도록 하는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[2] 변수 y에 대한 총합을 추정하고자 할 때, y의 관측값만 이용하여 추정하는 것에 비해, 또 다른 변수 x를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건(들)이 만족될 것 같은 구체적 예를 들어 보시오.

[3] 이단계군집표집(two-stage cluster sampling) 방법을 구체적 예로써 설명하시오.

[4] 표본크기가 n인 선거예측조사에서 한 후보의 지지율에 대한 오차의 한계를 대략 $1/\sqrt{n}$, 두 후보의 지지율 차이에 대한 오차의 한계를 대략 $2/\sqrt{n}$로 둘 수 있는 이유를 각각 설명해 보시오. 신뢰수준은 95%이다. 예를 들어 19대 대선에 관한 여론조사에서 1100명을 조사했을 경우에 95% 신뢰수준에서 오차의 한계를 $1/\sqrt{1100}$ 인 3%라고 한다. 그리고 A후보와 B후보의 지지율 차이가 이 값의 2배인 6% 이하이면 오차범위 내 경합, 6%를 벗어나면 오차의 한계를 벗어난 차이를 보인다고 흔히 발표하거나 판단한다. 그 근거가 무엇인지를 설명해보시오.

[5] 낮은 응답률을 보이는 여론조사의 신뢰도에 관한 문제이다. 2018년 2월 중앙선거여론조사심의위원회에 등록된 전국 단위 대선 및 정당 지지율 여론조사 66개를 분석한 결과 응답률 최저는 3.1%, 최고는 23.9%였으며 평균은 약 9.2%였다. 낮은 응답률을 보인 조사에 대한 다음 두 주장의 타당성에 대해 논하시오.

A: "1천 명을 조사하기 위해 1만 명에게 물었는지, 2만 명에게 물었는지는 신뢰도와는 직결되지 않는 문제이다"

B: "여론조사를 하면 90% 이상이 응답을 하지 않는다. 지금 여론조사라는 것은 경향성만 보는 것이지 큰 의미가 없다"

표본론 졸업시험 2018.3.10

[1] (10점) "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[2] (10점) 계통추출(systematic sampling)이 단순확률추출(simple random sampling)과 비슷한 추출방법이 되는 사례와, 단순확률추출보다 더 나은 추출방법이 되는 사례를 각각 들고, 왜 그렇게 생각하는지 이유를 밝히시오.

[3] (10점) 층화추출(stratified random sampling)과 군집추출(cluster sampling)을 비교하여 설명하고, 이 두 방법이 혼합된 표집설계의 예를 들어 보시오.

[4] (20점) 표본크기가 n인 선거예측조사에서 한 후보의 지지율에 대한 오차의 한계를 대략 $1/\sqrt{n}$, 두 후보의 지지율 차이에 대한 오차의 한계를 대략 $2/\sqrt{n}$로 둘 수 있는 이유를 각각 설명해 보시오. 신뢰수준은 95%이다. 예를 들어 19대 대선에 관한 여론조사에서 1100명을 조사했을 경우에 95% 신뢰수준에서 오차의 한계를 $1/\sqrt{1100}$ 인 3%라고 한다. 그리고 A후보와 B후보의 지지율 차이가 이 값의 2배인 6% 이하이면 오차범위 내 경합, 6%를 벗어나면 오차의 한계를 벗어난 차이를 보인다고 흔히 발표하거나 판단한다. 그 근거가 무엇인지를 설명해보시오.

표본론 졸업시험 2017.5.20

[1] (10점) 할당표집(quota sampling)이 무엇이며 어떤 경우에 문제가 되는지에 대해 논하시오.

[2] (10점) "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[3] (10점) 이단계군집표집(two-stage cluster sampling)의 표본추출방법을 구체적 예로써 설명하시오.

[4] (20점) 표본크기가 n인 선거예측조사에서 한 후보의 지지율에 대한 오차의 한계를 대략 $1/\sqrt{n}$, 두 후보의 지지율 차이에 대한 오차의 한계를 대략 $2/\sqrt{n}$로 둘 수 있는 이유를 각각 설명해 보시오. 신뢰수준은 95%이다. 예를 들어 19대 대선에 관한 여론조사에서 1100명을 조사했을 경우에 95% 신뢰수준에서 오차의 한계를 $1/\sqrt{1100}$ 인 3%라고 한다. 그리고 A후보와 B후보의 지지율 차이가 이의 2배인 6% 이하이면 오차범위 내 경합, 6%를 벗어나면 오차의 한계를 벗어난 차이를 보인다고 흔히 발표하거나 판단한다. 그 근거가 무엇인지를 설명해보시오.

표본론 졸업종합시험

2017.3

[1] (15점) 1994년 미국 대통령 선거 결과 예측을 위해 Gallup에서 1579명을 대상으로 실시한 조사에서 Clinton의 지지율이 43%, Bush의 지지율이 36%이었다. Gallup은 Clinton이 우세하다고 발표하였다. 만약 1579명이 단순확률추출(simple random sampling) 되었다고 가정할 수 있다면 이 발표는 타당한가?

[2] (15점) 층화추출(stratified random sampling)과 군집추출(cluster sampling)을 비교하여 설명하고, 이 두 방법이 혼합된 표집설계의 예를 들어 보시오.

[3] (20점) 낮은 응답률을 보이는 여론조사의 신뢰도에 관한 문제이다. 지난 2월 중앙선거여론조사심의위원회에 등록된 전국 단위 대선 및 정당 지지율 여론조사 66개를 분석한 결과 응답률 최저는 3.1%, 최고는 23.9%였으며 평균은 약 9.2%였다. 낮은 응답률을 보인 조사에 대한 다음 두 사람의 주장에 대해 논하시오.

A: “1천 명을 뽑아내기 위해 1만 명에게 물었는지, 2만 명에게 물었는지는 신뢰도와는 직결되지 않는 문제이다”

B: “여론조사를 하면 90% 이상이 응답을 하지 않는다. 지금 여론조사라는 것은 경향성만 보는 것이지 큰 의미가 없다”

표본론 졸업시험 2016.3.12.

[1] (10점) 할당표집(quota sampling)이 무엇이며 어떤 경우에 문제가 되는지에 대해 논하시오.

[2] (10점) 층화추출(stratified random sampling)과 군집추출(cluster sampling)을 비교하여 설명하시오.

[3] (10점) 이번 학기에 재학 중인 숭실대학교 학생을 대표할 수 있는 표본 500명을 추출하는 방법을 제시하시오.

[4] (20점) 선거결과 예측을 위한 어떤 전화조사에서, "전화는 전화번호부에서 선택해 걸지 않고 랜덤다이얼링 방식으로 걸었다. 그리고 전화를 받지 않을 경우 다른 시간대에 두 번까지 더 전화를 걸었는데 그래도 전화를 받지 않으면 무응답으로 간주하였다. 응답 협조를 거부하는 경우도 무응답으로 간주하였다. 전화조사의 응답률은 10%이었는데, 계획한 표본크기 1000명을 확보하기 위해 같은 방법으로 다른 번호로 전화를 걸어 목표한 표본의 크기를 다 조사하였으며, 이 때 인구총조사에서 알려진 연령별, 남녀 비율을 맞추어서 조사하였다. 조사 후 95% 신뢰수준에서 오차의 한계는 약 3.2%이다"라고 발표하였다. 오차의 한계 3.2%는 어떻게 계산된 것인가를 설명하고 이 조사결과를 얼마나 믿을 수 있는 것인가에 대해서도 논하시오.

표본론 졸업시험 2014.3.8

[1] (10점) 단순확률추출(simple random sampling)과 비교했을 때 층화추출(stratified random sampling)이 갖는 장점을 설명하시오.

[2] (10점) "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같아지는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[3] (10점) 숭실대학교 학생을 대표할 수 있는 표본 500명을 추출하는 방법에 대해 설명하시오.

[4] (20점) 선거결과 예측을 위한 어떤 전화조사에서, "전화는 전화번호부에서 선택해 걸지 않고 랜덤다이얼링 방식으로 걸었다. 그리고 전화를 받지 않을 경우 다른 시간대에 두 번까지 더 전화를 걸었는데 그래도 전화를 받지 않으면 무응답으로 간주하였다. 응답 협조를 거부하는 경우도 무응답으로 간주하였다. 전화조사의 응답률은 10%이었는데, 계획한 표본크기 1000명을 확보하기 위해 같은 방법으로 다른 번호로 전화를 걸어 목표한 표본의 크기를 다 조사하였으며, 이 때 인구총조사에서 알려진 연령별, 남녀 비율을 맞추어서 조사하였다. 조사 후 95% 신뢰수준에서 오차의 한계는 약 3.2%이다"라고 발표하였다. 오차의 한계 3.2%는 어떻게 계산된 것인가를 설명하고 이 조사결과를 얼마나 믿을 수 있는 것인지에 대해서도 논하시오.

표본론 졸업시험 2012.3.17

[1] (10점) 이단계군집추출(two-stage cluster sampling)의 표본추출방법을 구체적 예로써 설명하시오.

[2] (10점) "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[3] (10점) 모집단의 변수 y에 대한 총합을 추정하고자 할 때, y의 관측값만 이용하여 추정하는 것에 비해, 또 다른 변수 x를 이용하는 방법인 비추정(ratio estimation)이나 회귀추정(regression estimation)이 더 나은 방법이 되기 위한 조건(들)은 무엇인가? 그리고 이러한 조건들이 만족될 것 같은 구체적 예를 들어 보시오.

[4] (20점) 선거결과 예측을 위한 어떤 전화조사에서, "전화는 전화번호부에서 선택해 걸지 않고 랜덤다이얼링 방식으로 걸었다. 그리고 전화를 받지 않을 경우 다른 시간대에 두 번까지 더 전화를 걸었는데 그래도 전화를 받지 않으면 무응답으로 간주하였다. 응답 협조를 거부하는 경우도 무응답으로 간주하였다. 전화조사의 응답률은 10%이었는데, 계획한 표본크기 1000명을 확보하기 위해 같은 방법으로 다른 번호로 전화를 걸어 목표한 표본의 크기를 다 조사하였으며, 이 때 인구총조사에서 알려진 연령별, 남녀 비율을 맞추어서 조사하였다. 조사 후 95% 신뢰수준에서 오차의 한계는 약 3.2%이다"라고 발표하였다. 오차의 한계 3.2%는 어떻게 계산된 것인가를 설명하고 이 조사결과를 얼마나 믿을 수 있는 것인지에 대해서도 논하시오.

표본론 졸업시험 **2011.03**

[1] (10점) 조사 과정에서 무응답(nonresponse)은 어떤 경우에 문제가 되는가? 무응답으로 인한 오차를 줄이기 위한 방안에 대해서도 설명하시오.

[2] (10점) 전화조사를 통해 "민간의료보험을 도입하는 것에 찬성하십니까"를 물었다. 전체 응답자 1000명 중에서 여자의 비율이 0.8, 남자의 비율이 0.2 이었다. 1000명 중에서 찬성이라고 답한 비율이 $\hat{p}$이었고, 성별로 분류하여 살펴본 결과 여자와 남자의 찬성 비율이 각각 $\hat{p}_1$과 $\hat{p}_2$ 이었다. 전체 인구의 남녀 비율이 같은 점을 고려하여 모집단의 찬성 비율을 $\hat{p}$ 으로 추정하지 않고 $0.5\hat{p}_1 + 0.5\hat{p}_2$ 으로 추정하였다. 찬성 비율에 관한 두 추정량 $\hat{p}$와 $0.5\hat{p}_1 + 0.5\hat{p}_2$의 타당성에 대해 각각 논하시오.

[3] (10점) "모평균에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 추출될 확률이 모두 같도록 해야 한다"라는 주장의 타당성에 대해 논하시오.

[4] (10점) 층화확률추출(stratified random sampling)과 군집추출(cluster sampling)을 비교하여 설명하시오.

[5] (10점) 2011-1학기 재학 중인 숭실대학교 학생을 모집단으로 정의했을 때 약 400명의 표본을 추출하는 방법을 제시하시오.

표본론 졸업시험 **2010.03**

[1] (10점) "전수조사(census)를 하면 오차가 생기지 않는다"는 주장의 타당성에 대해 논하시오.

[2] (10점) 조사 과정에서 무응답(nonresponse)은 왜 문제가 되는가?

[3] (10점) "모평균에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 추출될 확률이 모두 같도록 해야 한다"라는 주장의 타당성에 대해 논하시오.

[4] (10점) 할당표집(quota sampling)에 대해 예를 들어 설명하고 어떤 경우에 문제가 되는지에 대해 논하시오.

[5] (10점) 2010-1학기 재학 중인 숭실대학교 학생을 모집단으로 정의했을 때 표본을 추출하는 방법을 제시하시오.

표본론 졸업종합시험 **2009.03.21**

[1] (15점) 조사에서 생기는 오차의 원인(sources of errors in surveys) 중에 무응답(nonresponse)이 있다. 무응답에 의한 오차(nonresponse error)가 무엇이며 왜 문제가 되는지, 그리고 대처방안이 무엇인지에 대해 각각 설명하시오.

[2] (20점) 크기 N인 모집단에서 n개의 표본을 추출하고자 한다. 모집단을 크기가 각각 $N_1, ..., N_L$인 L개의 군집으로 나눈 다음, K개의 군집을 단순확률추출(simple random sampling)하고, 추출된 군집에서 각각 $n_1,...,n_K$개의 표본을 추출하려고 한다. $n = n_1 + \cdots + n_K$.

(2-1) 모집단의 각 원소가 추출될 확률이 모두 같도록 하려면 $n_1,...,n_K$를 어떻게 정해야 하는가?

(2-2) 앞에서 제시한 방법대로 하면 모집단의 원소가 추출될 확률이 모두 같게 된다는 사실을 수리적으로 밝히시오.

(2-3) "모평균에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 추출될 확률이 모두 같도록 해야 한다"라는 주장의 타당성에 대해 논하시오.

(2-4) 모집단의 원소가 추출될 확률이 모두 같도록 하면 어떤 장점이 있는가?

[3] (15점) 2009-1학기 재학 중인 숭실대학교 학생을 모집단으로 정의했을 때 표본을 추출하는 방법을 제시하시오.

표본론 졸업시험 2008.03.15

[1] (20점) 숭실대학교 학생들의 이성교제에 관한 의식 조사를 위해 200명의 표본을 추출하려고 한다. 숭실대학교 남학생 중에서 100명, 여학생 중에서 100명을 단순임의추출(simple random sampling) 하였다고 하자.

(a) 이 추출방법에 대해 A 학생이 "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 써야만 하는데, 숭실대학교 남녀비율이 1:1이 아니므로 이 표본추출방법은 옳지 않다"고 지적하였다. 이 주장의 타당성에 대해 논하시오.

(b) A 학생의 주장대로 모든 원소가 뽑힐 확률이 같도록 했을 때 어떤 장점이 있는지 설명하시오.

[2] (20점) 크기 N인 모집단에서 n개의 표본을 추출하고자 한다. 모집단을 크기가 각각 $N_1, ..., N_L$인 L개의 군집으로 나눈 다음, 군집의 크기에 비례하는 확률로 K개의 군집을 추출하였다.

(a) 추출된 K개의 군집에서 다시 군집의 크기에 비례해 $n_1, ..., n_K$개씩의 표본을 추출하였다. 이 때 i번째 군집에 속하는 원소가 추출될 확률을 구하시오. (참고: $N = N_1 + \cdots + N_L$, $n = n_1 + \cdots + n_K$)

(b) 모집단의 모든 원소가 추출될 확률이 같아지도록 하려면 두 번째 단계에서 어떻게 추출하여야 하는가?

[3] (10점) N명의 모집단을 응답을 하는 N_1명의 층과 무응답자 N_2명의 층으로 나눌 수 있다고 하자. 모집단에서 n명을 단순임의추출하여 이 중 n_1명으로부터 응답을 얻었고 이로부터 모평균 μ의 추정량 $\bar{y}_1$를 얻었다. $\bar{y}_1$의 편향(bias)을 두 층의 평균 μ_1, μ_2와 층의 크기 비율인 $W_1 = N_1/N$, $W_2 = N_2/N$ 을 써서 나타내시오. 그리고 이 식으로부터 $\bar{y}_1$이 μ의 비편향 추정량(unbiased estimator)가 되기 위한 조건을 밝히시오.

표본론: 2007 졸업종합시험

[1] 조사에서 생기는 오차의 원인(sources of errors in surveys) 중에 무응답(nonresponse)이 있다. 무응답에 의한 오차(nonresponse error)가 무엇이며 왜 문제가 되는지, 그리고 대처방안이 무엇인지에 대해 설명하시오.

[2] "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 쓰는 것이 좋다"라는 주장에 대해 논하시오.

[3] 크기 N인 모집단에서 n개의 표본을 추출하고자 한다. 모집단을 크기가 각각 $N_1, ..., N_L$인 L개의 군집으로 나눈 다음, 군집의 크기에 비례하는 확률로 K개의 군집을 추출하였다.

(a) 추출된 K개의 군집에서 다시 군집의 크기에 비례해 $n_1, ..., n_K$개씩의 표본을 추출하였다 ($n_i \propto N_i$). 이 때 i번째 군집에 속하는 원소가 추출될 확률을 구하시오. (참고: $N = N_1 + \cdots + N_L,\ n = n_1 + \cdots + n_K$)

(b) 모집단의 모든 원소가 추출될 확률이 같아지도록 하려면 두 번째 단계에서 어떻게 추출하여야 하는가?

표본론: 2006 졸업종합시험

[1] 어느 시청에서 주민들을 대상으로 행정서비스에 대한 주민들의 의견을 수렴하기 위해 표본조사를 실시하고자 한다.

a) 어느 특정한 날 시청 민원실에 들린 주민 중 10%에 해당하는 주민을 표본으로 추출하는 방법에 대해 설명하고, 이렇게 추출한 표본이 민원실을 이용하는 전체 주민을 대표할 수 있는 표본이 되기 위한 가정을 설명하시오.

b) 표본으로 선정된 주민들을 대상으로 다음 질문을 했다.

다음 중 가장 시급하게 보완되어야 할 공공시설은 무엇입니까?
① 도서관 ② 수영장 ③ 보육시설 ④ 공원 ⑤ 기타

응답자 중 성별에 따라 위 문항에 대한 응답에 차이가 있는지 검정하고자 한다. 자료가 얻어졌다고 가정하고 검정방법에 대해 설명하시오.

[2] "모집단 특성에 관한 비편향 추정량(unbiased estimator)을 얻기 위해서는 모집단의 원소가 뽑힐 확률이 모두 같도록 하는 표본추출방법을 써야만 한다"는 주장에 대해 논하시오.

[3] 숭실대학교 학생을 대상으로 '병역거부자의 대체복무 허용'에 관한 찬반의견 조사를 하고 싶다. 크기비례추출법에 의한 이단계군집표집(two-stage cluster sampling with PPS)을 한다면 어떻게 표본을 추출해야 하는지를 설명하고, 층화확률추출(또는 층화확률표집, 보다 정확히 말하면 층화일단확률표집(one-stage stratified random sampling))과 비교하여 장단점을 논하시오.

표본론: 2002 졸업종합시험

[1] 표집조사(sampling survey)에서 발생하는 오차를 크게 표집오차(sampling error)와 비표집오차(nonsampling error)로 나눌 수 있다.

1) 비표집오차의 원인들과 대처방안들에 대해 설명하시오.

2) "전수조사(census)를 하면 오차가 없다"라는 주장에 대해 논하시오.

[2] 어느 시청에서 주민들을 대상으로 행정서비스에 대한 주민들의 의견을 수렴하기 위해 표본조사를 실시하고자 한다.

1) 어느 특정한 날 시청 민원실에 들린 주민 중 10%에 해당하는 주민을 표본으로 추출하는 방법에 대해 설명하시오.

2) 표본으로 선정된 주민들을 대상으로 다음 질문을 했다.

다음 중 가장 시급하게 보완되어야 할 공공시설은 무엇입니까?
① 도서관 ② 수영장 ③ 보육시설 ④ 공원 ⑤ 기타

응답자 중 남녀 성별에 따라 응답에 차이가 있는지 알아보기 위한 검정방법에 대해 설명하시오.

[3] 심리학자가 성인 정신장애자들의 자극에 대한 평균반응시간을 추정하고자 한다. 남녀별로 반응시간에 차이가 있다고 알려져 있어 성별로 층화추출(stratified random sampling)을 하기로 하였다. 전부 96명 중에서 남자가 43명이며, 지난 연구로부터 남자들의 반응시간은 5초에서 20초이고, 여자들은 3초에서 14초임이 알려져 있다고 하자. 표본을 뽑는 비용은 남녀별로 차이가 없으며, 최적할당(optimum allocation)을 하기로 하였다. 오차의 한계(bound on the error of estimation)를 1초로 하고 싶다면, 각 층에서 몇 명씩을 추출해야 하는지 결정하시오

[4] 군집표집(cluster sampling)이 단순확률표집(simple random sampling)에 비해 갖는 장단점을 논하고, 군집표집이 단순확률표집보다 더 '좋은' 표집방법이 되는 구체적 예를 들어보시오. 그리고 그 예에서 왜 군집표집이 더 '좋다'고 생각하는지 이유를 밝히시오.