

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (40점)

100명의 학생이 두 개의 선택과목 중 한 과목을 선택하여 시험을 친 결과가 다음과 같다.

구 분	인 원	평 균	분 산
선택과목 1	40	80	100
선택과목 2	60	70	25

- 1) 과목 선택을 무시하고 전체 학생의 평균과 분산을 계산하라
- 2) ‘선택과목 1’의 답안지 중 1장이 잘못 채점되어 실제로 10점인데 90점으로 계산되었다. 실제 ‘선택과목 1’의 평균과 분산을 계산하라.
- 3) ‘선택과목 1’의 난이도가 ‘선택과목 2’ 보다 낮은 것으로 보인다. 이 경우 ‘선택과목 2’를 선택한 학생들이 불이익을 받지 않도록 난이도를 조정하는 방안을 제시하라.

제 2 문 (60점)

1. 우리나라 국민들을 대상으로 정부에서 입안한 새로운 교육정책에 대해 찬성하는 사람들의 비율 p 를 파악하고자 한다. 이와 관련해 두개의 조사기관에서 각각 독립적으로 랜덤하게 표본을 추출하여 여론조사가 수행되었다. 조사기관 '갑'에서는 800명을 랜덤하게 추출하여 조사한 결과 찬성률 $\hat{p}_1 = 480/800 = 0.6$ 을 얻었고, 다른 조사기관 '을'에서는 랜덤하게 추출된 200명을 조사하여 찬성률 $\hat{p}_2 = 80/200 = 0.4$ 를 얻었다. 정책 담당자는 두 조사기관에서 얻은 결과를 종합하여 좀 더 신뢰할 수 있는 찬성률 p 를 추정하기 위해 다음과 같은 두 가지 추정방법을 고려하고자 한다.(40점)

① $\hat{p} = (0.6 + 0.4)/2 = 0.5$

② $\tilde{p} = (480 + 80)/(800 + 200) = 0.56$

- 1) 두 가지 추정방법에 대해 각각 편향(bias)이 발생하는지 밝혀라.
- 2) 두 가지 추정방법에 대해 추정량의 분산을 구하라.
- 3) 제시된 \hat{p} 과 \tilde{p} 는 \hat{p}_1 과 \hat{p}_2 의 가중평균 형식의 추정량이다 이런 가중평균 추정량 중 가장 효율적인 추정량을 구하는 방법에 대해 설명하라.

2. 여러 개의 변수가 주어진 경우 이들 사이의 관계를 알아보고자 한다.(20점)

- 1) 편상관계수(partial correlation coefficient)를 정의하고 그 의미를 설명하라.
- 2) 회귀분석에서 편상관계수가 어떻게 활용되는지 설명하라.