

사회조사분석사 작업형 풀이 시 유의사항

1

제공된 시험지의 코딩 양식 확인(코드북 확인)

- 문항, 변수명, 내용, 결측치 처리 등을 확인

2

원시 데이터 확인(txt)

- 변인 간의 구분자 확인 (tab, 빈공간, 특수기호 등등) 불러들이는 방법 확정

3

불러들인 후 자료 확인

- txt ➡ sav 변환 시 코드북에 맞게 입력되었는지 확인
- 불필요하게 입력된 '점(.)' 등의 행렬 삭제

4

변수명 입력

- 문항, 변수, 내용, 결측치 처리 입력

결측치(결측값) 처리란?

- 무응답을 컬럼 내의 지정된 수치로 입력
- 응답거부(98), 무응답(99) 등으로 지정

결측치(결측값) 처리 이유

- 실제 조사에서 100% 응답을 받아내기란 어려우므로, 무응답 한 부분의 경우 분석에서 자동 제거하여 분석함.
- 100% 응답이 아닐 경우 자료 자체를 제거하기도 함.

이상치와 오류값

- 이상치 : 응답은 정확하나, 동떨어진 값 (예 : 한국 성인남자 키와 몸무게를 측정하는데, 최홍만 선수의 키와 몸무게가 있을 경우)
- 오류값 : 응답범주 외의 수치가 입력된 경우(예 : 응답이 2번까지만 있는데 2 이상의 3, 4 등이 입력된 경우)

5

무응답, 응답 거부의 처리

- 무응답, 응답거부를 결측치 처리(missing value)할 것인가?
(예시 : 결측값 “9” , “99”)
- 응답 자체에서 제거 할 것인가? 선택

6

이상치, 오류값 처리

- 이상치, 오류값을 결측치 처리 할것인가?
- 응답 자체에서 제거 할것인가? 선택

7

역코딩 확인

- 긍정이 최고점이 되도록 변환 (예시 : $6-x_1$) 반드시 결측, 이상치 처리 후

8

소수점 확인

- 반올림, 올림, 몇째 자리 등

9

평균값 처리 방법

- 모두 응답한 경우에만 평균 산출 (산식 : $(x_1+x_2)/2$ 이용)
- 응답 여부에 상관없이 평균 산출 (산식 : $\text{mean}(x_1, x_2)$ 이용)

10

유의 수준 확인

- 유의수준이 0.05, 0.1, 0.01 확인
- 문제에서 주어지지 않는다면, 0.05 범위로 해석

11

반드시 앞에서부터 풀이

- 결측값 처리, 변수 삭제 등이 주어지므로 반드시 앞에서부터 문제 풀이

변수 계산의 중요 함수

- mean함수 [mean(X1,X2,X3)] 와 산식 $(X1+X2+X3)/3$ 차이점
- sum함수 [SUM(X1,X2,X3)]와 산식 $(X1+X2+X3)$ 의 차이점

ABS : 절대값

LN : 로그

SQRT : 루트

Trunc : 절사하기

- 예시 : $\text{trunc}(4.579, 0.1) = 4.5$, 소수점 첫째자리 이후 무조건 절사
 $\text{trunc}(4.579, 1) = 4$, 소수점 자리는 무조건 절사

RND : 반올림하기

- 예시: $\text{RND}(4.579, 0.1) = 4.6$, 소수점 둘째 자리에서 반올림하여 소수점 첫째자리로 만들.

<표 1> 핵심 요점 사항

분석 기법		주요 내용
교차분석		○질적 변수와 질적 변수(예 : 성별과 지역) · 귀무가설 : 두 변수간의 관련이 없다(독립이다). · 대립가설 : 두 변수간의 관련이 있다(독립이 아니다). · 검정통계량 분포 : Pearson Chi-Square ※ 통계량 ■ 카이제곱 선택, 셀 출력의 행 퍼센트
다중응답 분석		○분석 ■ 다중응답 ■ 변수군 정의
T-test	독립표본 T-검정	○독립변수의 수 : 질적 변수 1개 ○독립변수의 속성수 : 2가지 속성(예 : 남자, 여자) ○종속변수 : 양적 변수 1개 (예 : 성별에 따른 만족도 평균차이) · 귀무가설 : 두 집단의 평균에는 차이가 없다. · 대립가설 : 두 집단의 평균에는 차이가 있다. ○가정(정규성) 등분산성 검증 - 귀무가설 : 두 집단의 분산은 동일하다. - 대립가설 : 두 집단의 분산은 동일하지 않다. - 검정통계량 분포 : F ○등분산성 검증 결과로 등분산, 이분산 결과치로 해석 · 검정통계량 분포 : t
	대응표본 T-검정	○동일한 응답자의 사전, 사후 차이 분석(질적 변수) (예 : 실험 전과 후, 동일한 응답자의 영어, 수학점수) · 귀무가설 : A와 B에는 차이가 없다. · 대립가설 : A와 B에는 차이가 있다. · 검정통계량 분포 : t ※반드시 동일한 응답자의 사전, 사후 차이 분석임
일원배치 분산분석		○독립변수의 수 : 질적 변수 1개 ○독립변수의 속성수 : 3가지이상 속성 (예 : 20대, 30대, 40대) ○종속변수 : 양적 변수 1개 (예 : 연령대별에 따른 만족도 평균차이) ○가정(독립성, 정규성, 등분산성 - 등분산성 꼭 선택) · 귀무가설 : 집단수준에 따라 평균의 차이는 없다. · 대립가설 : 최소한 하나의 집단에서는 평균의 차이가 있다. · 등분산성 검정 결과 반드시 확인 · 검정통계량 분포 : F ○차이가 있을 경우 사후분석 실시
이원배치 분산분석		○독립변수의 수 : 질적 변수 2개 (성별, 연령별)

분석 기법		주요 내용
		○독립변수의 속성 수 : 각각 2가지 이상 속성 (예 : 성별 : 2개의 속성, 연령별 : 3개의 속성) ○종속변수 : 양적 변수 1개 ○가정(독립성, 정규성, 등분산성 $\Rightarrow$ 등분산성 꼭 선택) ○반복 측정 여부 확인 (원시 데이터에서 확인 후 모형에서 설정 가능) ○등분산성 검정 결과 반드시 확인 ○수정 모형 결과로 모형의 타당성 검정 ○상호작용 효과 분석(반복 측정일 경우에만) · 귀무가설 : 요인 A와 요인 B 간의 상호작용 효과가 없다. · 대립가설 : 요인 A와 요인 B 간의 상호작용 효과가 있다. · 상호작용 효과가 없으면 반복 측정이 아닌 경우로 분석 ○유의한 독립변수 선별 · 귀무가설 : 요인 A에 대해 집단의 평균 차이 없다. · 대립가설 : 최소한 하나의 요인 A에 대해 집단의 평균 차이 있다. ○사후분석은 반복 측정이 아닌 경우의 유의한 독립변수만 선택하여 분석
상관분석		○양적 변수와 양적 변수 두 변수간의 상관관계 · 귀무가설 : A와 B는 상관이 없다($\rho = 0$). · 대립가설 : A와 B는 상관이 있다($\rho \neq 0$). · 검정통계량 : 피어슨 상관계수
신뢰도 분석		○리커트 척도(예 : 5점 척도)로 긍정이 최고가 되도록 ○항목제거 시 척도 활용
회귀분석	단순/ 다중 회귀분석	○가정 : 오차항의 (독립성, 정규성, 등분산성, 평균은 0) ○단순회귀 : 양적 변수 1개의 종속변수, 양적 변수 1개의 독립변수 ○다중회귀 : 양적 변수 1개의 종속변수, 다수의 독립변수(질적 변수의 경우 더미변수화) ○분석절차 원칙 : 산점도 $\Rightarrow$ 상관분석 $\Rightarrow$ 회귀식의 유의성 검정(분산분석) F값 $\Rightarrow$ 결정계수 $\Rightarrow$ 독립변수의 유의성 판단
		1. 분산분석 결과로 모형의 유의성 검증 · 귀무가설 : 회귀식은 유의하지 않다.($\beta = 0$) · 대립가설 : 회귀식은 유의하다.($\beta \neq 0$) · 검정통계량 분포 : F
		2. 결정계수 : 0~1까지 · R^2 : 단순회귀

분석 기법		주요 내용
		· 수정된 R^2 : 다중회귀(문제에서 별다른 언급 없으면, 다중회귀라도 R^2 값을 쓴다.)
		3. 회귀식 작성 · 회귀식은 전체 투입 변수들의 비표준화 계수로 작성
		4. 회귀식 설명 · 1차 : p값으로 유의한 변수 선정 · 2차 : 유의한 변수 간에는 표준화 계수로 영향력 정도 판단 · 유의한 변수 선택은 문제에서 주어진 방법으로 진행
	더미변수 포함 회귀분석	○ 종속변수 : 양적 변수 ○ 독립변수 : 질적 변수(반드시 0, 1로 구성) 포함 · 더미 생성은 변환  코딩변경으로 · 그룹별 회귀분석 작성 방법(파일분할)과 비교 · 더미회귀분석의 장점(250쪽 참조)

※ 2010년부터 통계적 가설을 기재하라는 문제가 자주 출제되고 있으므로, 수식을 사용한 통계적 가설설정을 암기할 필요가 있다.

Ver 2.0

본 자료는 사회조사분석사 2급 준비생들을 위해 “사경환”이 2012년 3월에 제공하는 자료로 저작권자를 밝히는 경우에 한하여 누구나 게시 가능합니다.

하지만, 내용의 일부 인용, 편집하여 마치 자신의 만든 것처럼 가공 시 저작권에 위배됩니다.

특히, 다른 자료출처는 명기하면서 저작권자의 인용을 뺀 경우, 법적 조치를 받을 수 있습니다.

보다 자세한 내용은 <http://cafe.naver.com/statclub>를 참조하시기 바랍니다.

2012년 3월 12일 사경환