

실험계획법 실습 QUIZ

문1. 표본과 모집단의 차이? 그리고 여기서 조심해야 할 부분은 무엇이 있는 지 적어보세요.

모집단: 얻고자 하는 정보와 관련 있는 모든 개체로부터 얻을 수 있는 모든 관측 값들의 집합.

표본: 모집단의 일부분이다.

여기서 표본을 뽑을 때 모집단을 대표 할 수 있어야 함.

예를 들어서, 인하대학교 학생의 평균 주량을 측정하고 싶을 때, 단순히 통계학과 학생만

분석한다면 문제가 있지 않을까? 물론 높게 나올 수 있겠죠?

만약 소주 3병이 나왔다고 쳤을 때, 인하대학교 학생이 모두 그럴까?

문2. Data분석에 있어서 변수(Variable)의 종류는 크게 어떻게 분류가 되는가?

범주형 변수(순위형 & 일반 범주형): 성별, 학점 등

연속형 변수: 키, 몸무게 등

문3. 문2에서 생각한 변수에 있어서 자신이 생각하기에 가장 맨 처음으로 해야 할 분석이나 구해야 할 값은 무엇이라고 생각하는가?

범주형 변수: 빈도분석 -> 이상한 값 관측 가능

예: 남, 여, 여성, 남자, M, F, L, H 맞, 여, ㅈ

연속형 변수: 기초통계량(평균, 표준편차, 최대, 최소 등) -> 이상한 값 관측 가능

예: 표준편차 > 평균 ? -> 최대 값이나 최소 값이 이상한 경우에 발생

점수로 예를 들면 100점 만점인데, 150점으로 적힌 경우

따라서, 범주형 변수는 빈도분석 연속형 변수는 기초통계량을 꼭 구하기!!

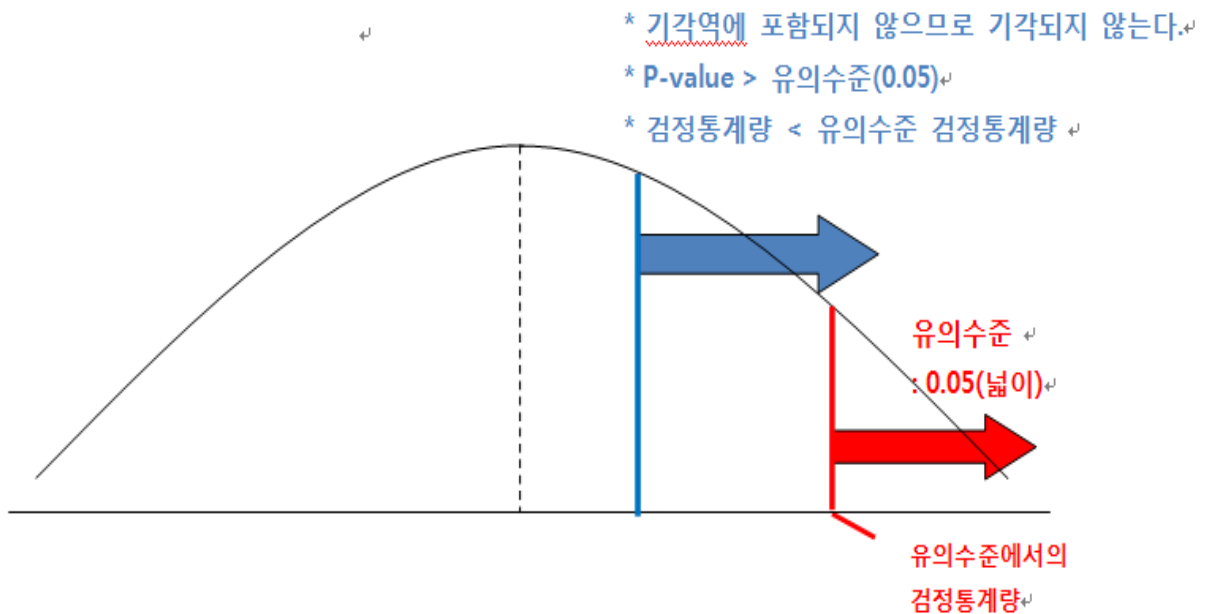
문4. (대립가설과 귀무가설) 및 유의수준에 대해서 서술을 먼저 하고, 문1에서 생각한 변수에 대해서 각각 또는 혼합해서 사용되는 검정 방법을 간략하게 적어보세요.

대립가설(H1): 우리가 주장하고 싶은 가설

귀무가설(H0): 대립가설의 반대

유의 수준: 주어진 검정통계량의 관측치로부터 H0를 기각하게 하는 최소의 유의수준

Tip! 그림으로 이해



가설 검정을 할 때, 만약 의미가 있다면, 어떤 의미가 있는 지 꼭 분석!!

의미가 있다. -> 차이가 있다.

평균비교: 단일(일)표본, 독립(이)표본 비교, 대응(쌍체) 표본, 분산분석

교차분석: 두 범주형 변수 비교 성별에 따라 결혼여부에 차이가 있는가?

상관분석: 두 연속형 변수간의 직선관계의 존재 여부 확인 꼭!! 산점도 확인

그 외 회귀분석, 로지스틱 회귀분석 등등 있어요

다음 문제들은 자신이 원하는 통계프로그램을 이용하여, 진행하세요.

문5. DATA1.csv에서 얻어 낼 수 있는 정보를 구하면서,

남녀에 따라, 키(Height)와 몸무게(Weight)의 차이가 나는 지 검정해보세요.(0.05기준)

ID(고유 번호), Sex(성별), Age(나이), Height(키), Weight(몸무게)

1. 문자형: 성별 빈도분석
 2. 연속형: Age, Height, Weight 요약통계량 또는 분포분석(히스토그램, 상자그림)
 3. T-검정: 등분산이 만족하는지 보고, p-value 선택(둘다 등분산 만족)
- 기타. 일원분산분석: 실험계획법과 연관이 많음. 한번 언급!!

문6. DATA2.csv에서 성별에 따른 결혼 여부의 차이가 있는 지 검정해보세요.(0.05기준)

Sex(성별), Marriage(결혼여부)

1. 문자형: 성별 및 결혼여부 빈도분석
2. 교차분석: 카이제곱 검정

문7. DATA3.csv에서 단순선형회귀를

$y = \beta_0 + \beta_1 * x + \varepsilon$ 에서,

β_1 이 유의한 경우 계수 값이랑 모델의 R-Square값을 적어주세요.(hint: 한번 더 고민!!)

1. 먼저 회귀분석 실시
2. 상관분석 후, 이상한 점 찾고, 제거 후, 실시
3. 기존(1.87806, 0.6846) & 제거(1.95419, 0.9822)

문8. DATA4_1.csv와 DATA4_2.CSV를 결합하여 다음과 같은 데이터를 만들어보세요.

subject_number	subject	+	subject_number	professor	=	professor	subject_number	subject
1	통계소프트웨어 및 실습		1	전홍석(교수님)		전홍석(교수님)	1	통계소프트웨어 및 실습
2	실험계획법		2	전홍석(교수님)		전홍석(교수님)	2	실험계획법
3	자료분석		3	박현진(교수님)		박현진(교수님)	3	자료분석
4	데이터마이닝		4	박현진(교수님)		박현진(교수님)	4	데이터마이닝
5	기초통계1		5	황진수(교수님)		황진수(교수님)	5	기초통계1
6	확률론		6	김진경(교수님)		김진경(교수님)	6	확률론
7	통계수학		7	김진경(교수님)		김진경(교수님)	7	통계수학

질의 빌더 이용(SQL) 대신