

〈3주차〉		
관련 능력단위요소	수업내용(관련 수행준거)	수업방법/기자재
현장별 설문조사 실시하기 (1401030103_14v2.2)	<ul style="list-style-type: none"> 설문대상자 선정 및 설문서 작성(2.1) <ul style="list-style-type: none"> 설문조사의 정의와 특징 설문조사의 절차, 오차, 신뢰성 설문조사의 방법 설문조사 관련 통계의 내용 <ul style="list-style-type: none"> 기초 통계의 이해, 용의 정의 및 개념 	이론강의

1 설문 조사

설문 조사는 주로 면접원이 조사 대상자와 면접을 통해 설문지에 응답을 받는 방식으로 이루어진다. 적절한 조사 대상자의 선정이 가능하고, 다양하고 깊이 있는 대상자 정보를 얻을 수 있어 가장 많이 활용되는 양적 조사 방법이다.

1. 설문 조사의 정의와 특징

(1) 설문 조사의 정의

사회의 여러 분야에서 일어나는 문제를 해결하기 위해서 문제에 관계하고 있는 사람 혹은 조직에 대해서 동일한 질문을 하고, 질문에 대한 대답으로서 자료를 수집하고 수집된 자료를 분석함으로써 문제 해결에 도움을 주는 정보를 얻어내는 과정

(2) 설문 조사의 특징

- (가) 문제해결을 위해 필요한 정보를 얻는 목적으로 실시
- (나) 집단에 관한 정보를 찾기 위해서 실시
- (다) 설문지를 이용하여 자료를 수집
- (라) 수집된 자료를 활용하기 위한 자료 분석

2. 설문 조사의 진행 절차

조사 문제의 규명 → 조사 설계 → 자료 수집 → 자료의 분석 및 해석 → 보고서 작성 및 발표

(1) 조사 문제의 규명

조사를 통해 해결하고자 하는 문제를 체계적으로 규명하고 정의

(2) 조사 설계

조사 수행 및 통제를 위한 청사진 수립

(가) 조사 문제의 종합적 검토

조사 목표, 조사 문제 및 연구 가설의 검토 및 평가

(나) 조사 항목의 결정

조사 문제를 해결하기 위해 필요한 정보 항목의 구체화

(다) 조사 체계의 제시

전반적 조사 절차, 조사의 유형, 자료 수집 절차, 자료 분석 기법 등의 결정

(라) 예산 편성과 조사 일정 작성

인원, 시간, 비용 고려

(마) 조사 설계의 평가

조사 설계의 현실성, 신뢰성, 타당성, 조사 결과의 일반화 가능성 등 평가

(3) 자료 수집

(가) 자료의 성격

1) 1차 자료: 직면한 문제와 직접적으로 관련되는 자료

2) 2차 자료: 다른 조사자가 다른 목적으로 수집한 자료

(나) 자료 수집 방법의 선정

1) 직접적 방법: 조사 대상자에게 조사 문제에 관하여 직접 질문 또는 자극을 주어 자료를 얻어냄

2) 간접적 방법: 조사 대상자로부터 간접적으로 자료를 얻어냄 (관찰, 흔적 조사법, 내용 분석법)

(다) 조사 대상자의 선정

1) 모집단의 확정: 조사 대상자의 성격과 범위

2) 표본의 추출: 대상, 기간, 장소에 관한 자료 수집

(라) 실사

1) 표본에 대한 접근: 조사 대상자의 협조와 승낙 획득, 참여에 대한 보상 제공

2) 조사원에 대한 통제: 조사원의 모집, 교육 및 훈련, 실사 과정에서 발생하는 문제점 파악 및 해결

(4) 자료의 분석 및 해석

(가) 자료의 정리와 편집: 자료의 정정, 보완, 삭제

(나) 자료의 코딩: 자료의 내용에 대한 일정한 숫자 부여

(다) 자료의 정확성 평가: 조사 오류, 신뢰성, 타당성

(라) 통계 분석: 가설에 적합한 분석 기법을 이용한 통계적 분석 수행

(마) 분석 결과의 해석: 통계적으로 분석한 결과의 의미 해석

(바) 시사점 도출: 의사 결정 문제의 해결 방안 추출

(5) 보고서 작성 및 발표

(가) 조사 결과와 결론 정리

(나) 보고 사항의 선정: 관리적으로 의미있고 중요한 내용 발체

(다) 목차 및 구성의 결정: 기술적인 내용은 첨부로 처리

(라) 최종 보고서 및 발표 자료 작성: 간결성과 흥미성

3. 설문 조사의 오차

(1) 표본 오차

(가) 표본의 추출과 선정 과정에서 발생하는 오류

(나) 모집단의 특성과 표본의 특성이 정확하게 일치하지 않기 때문에 야기됨

(다) 표본 오차는 표본의 크기를 증가시킴으로써 작아짐

(2) 비 표본 오차

- (가) 조사 과정 상에서 발생하는 다양한 오차
- (나) 조사 결과에 대해 표본 오차보다 더 심각한 영향을 미침
- (다) 조사 방법과 과정의 개선을 통해서만 감소시킬 수 있음
- (라) 질문에 대한 응답자의 허위 대답 및 질문의 의미 혼동, 응답자의 부재와 응답거부, 조사원의 부정, 조사 자료를 집계할 때의 계산 착오, 자료 입력 착오

4. 신뢰성과 타당성

(1) 신뢰성의 의미

- (가) 신뢰성이란 측정치의 일관성, 안정성, 예측 가능성
- (나) 측정값들의 차이가 적을 때 신뢰성 높음

(2) 신뢰성의 측정 방법

- (가) 반복 측정 신뢰성
동일한 측정 도구를 2회 반복 측정하여 두 측정 값들 간의 차이를 분석하거나 상관 계수를 추정하여 평가
- (나) 유사 측정 도구 신뢰성
동일한 개념을 측정하는 유사한 2가지 측정 도구를 이용하지만 유사한 측정 도구개발에 비용과 시간이 많이 소요됨
- (다) 항목 분할 신뢰성
다수의 측정 항목을 대등한 2개의 그룹으로 나누고, 두 그룹의 상관 관계 추정
- (라) 내적 일관성
항목들의 일관성 측정

(3) 신뢰성 개선 방안

- (가) 측정 도구 모호성 제거
모든 조사 대상자가 동일한 의미로 해석하도록 내용을 명확히 한다.
- (나) 측정 항목수 증가
신뢰성 계수 공식에서 항목을 늘리면 계수 증가
- (다) 측정 방식의 일관성 유지
- (라) 조사 대상자가 잘 모르는 질문 배제
- (마) 유사 질문의 반복 측정

(4) 타당성의 의미

측정 도구가 측정하고자 하는 개념을 정확히 반영하고 있는가 하는 정도를 나타내며 신뢰성이 아무리 높아도 타당성이 낮으면 측정 결과는 무용하다.

(5) 타당성의 개선 방안

- (가) 신뢰성을 높임
- (나) 다양한 타당성 검증 방법을 통한 측정 도구의 평가

(다) 타당성이 입증된 측정 방법과 측정 도구의 활용

② 설문 조사 방법

1. 의사 소통 방법

응답자에게 질문을 하여 자료를 획득하는 방법

(1) 대인 면접

조사자와 응답자가 대면해서 질문을 설명하고 응답을 말로 얻어내는 방법

(2) 전화 면접

전화를 이용하여 질문을 설명하고 응답을 얻어냄

(3) 우편 조사

응답자에게 설문지를 우편으로 발송하고 이를 작성하여 회송하도록 하는 방법

2. 관찰 방법

질문이 아닌 상황이나 행동을 관찰하여 기록하는 방법

(1) 장점

객관적이고 정확한 자료 수집이 가능하며, 응답자가 본인의 태도나 의견에 대하여 정확히 모르는 경우에도 조사가 가능하다.

(2) 단점

관찰이 불가능한 행동에 대한 조사가 어려우며, 행동을 관찰하여 내면적 심리를 추정하는 것이 객관성과 타당성이 결여된다.

[설문조사 수행 순서]

① 설문 대상자를 선정한다.

전체 모집단에서 표본을 선정하면 조사에 소요되는 시간 및 조사 비용이 절약되며, 제한된 시간과 예산으로 정밀한 자료 수집이 가능하다. 전수 조사가 불가능한 경우 설문 대상자를 선정하여 설문을 실시한다.

1. 모집단 확정하기

(1) 모집단을 정의한다.

통계적으로 관찰하고자 하는 대상이 되는 집단 전체를 가리켜서 모집단이라 한다.

(2) 모집단을 규정한다.

(가) 연구 대상

자료가 수집되는 대상을 정한다.

(나) 표본 단위

표본으로 추출될 수 있는 요소를 정한다.

(다) 범위

연구 대상에 대한 조작적 정의를 한다.

(라) 시간

시간적 경계를 정한다.

2. 설문 대상자 프레임 결정하기

(1) 표본 프레임을 정의한다.

연구 대상이나 표본 단위가 수록된 목록을 표본 프레임이라 한다.

(2) 표본 프레임 오차를 정의한다.

모집단과 표본 프레임의 불일치로 발생하는 오차를 말한다.

(3) 표본 프레임의 오차를 감소시킨다.

선별 질문을 통하여 부적합한 조사 대상자를 제거하고, 수집된 자료에 가중치를 적용하여 표본 집단의 특성과 모집단의 특성을 일치시킨다.

3. 설문 대상자 추출 방법 결정하기

(1) 확률적 샘플링을 구분한다.

(가) 단순 랜덤 샘플링

샘플링의 가장 기초적인 방법으로 표본의 수가 정해졌을 때, 모든 가능한 표본들이 실제 표본으로 선정되는 확률이 동일한 경우로 샘플링한다.

(나) 층화 샘플링

모집단을 구성하고 있는 층에 특정 수의 표본을 할당하는 확률적 방법으로 샘플링한다.

(다) 군집 샘플링

모집단을 구성하고 있는 군집들 중에서 일부의 군집을 랜덤으로 선택하여, 선택된 군집에 속해 있는 모든 연구 대상 또는 다시 랜덤으로 선택된 군집에 속한 연구대상의 일부를 조사하는 방법으로 샘플링한다.

(라) 체계적 샘플링

샘플링 프레임에 속한 연구 대상을 나열하고, 랜덤으로 시작하여 첫번째 연구 대상을 표본으로 선택한 다음, 매번 정해진 순번째 연구 대상을 표본으로 선택하는 확률적 방법으로 샘플링한다.

(2) 비확률적 샘플링을 구분한다.

(가) 편의 샘플링

연구자가 연구 진행에 편리하게 표본을 선택하는 방법으로 샘플링한다.

(나) 판단 샘플링

연구자가 연구 목적에 가장 적합한 표본이라고 생각하는 연구 대상을 선택하는 방법으로 샘플링한다.

(다) 쿼타 샘플링

모집단을 세그먼트로 구분하여 각 세그먼트에 대한 표본의 수를 나타내는 쿼타를 연구자의 판단 하에 결정하는 방법으로 샘플링한다.

(라) 스노우볼 샘플링

랜덤으로 연구 대상을 선택하여 초기 표본을 구성 및 인터뷰를 실시하고, 그 연구 대상에서 다른 연구 대상을 소개받아 추가 표본을 구성한다. 이를 반복하여, 원하는 표본 수를 얻을 때까지 계속 소개받아 인터뷰하는 방법으로 샘플링한다.

(3) 설문 대상자 추출 방법을 결정한다.

4. 설문 대상자 크기를 결정한다.

(1) 표본 크기는 조사 비용, 조사 기간 및 정확도와 밀접하게 관련된다.

(2) 표본 크기를 결정하는 경우 고려한다.

(가) 연구 목적

(나) 시간 제약

(다) 비용 제약

(라) 자료 분석 상 제약

(마) 표본 추출 기법 및 추정 절차 유형

(3) 표본 크기를 결정한다.

(가) 95% 신뢰도를 얻으면 좋다

(나) 표본 수는 적어도 100은 필요하다.

(다) 조사의 회수율도 고려한다.

(라) 성별이나 연령별로 분석할 경우에 표본 수가 작게 되지 않도록 한다.

② 설문서를 작성한다.

1. 설문서 작성시 주의 사항

(1) 정보의 종류: 조사문제와 가설과 관련된 정보

(2) 측정 방법: 척도의 형태

(3) 분석 내용: 어떤 문항을 어떻게 살펴본 것인가?

(4) 분석 방법: 구체적인 통계 기법

(5) 조사 결과의 방향: 출력 이미지를 통한 조사결과 추정

2. 설문서의 내용

- (1) 응답자에 대한 협조 요청
조사자와 조사 기관 소개, 조사 취지 설명, 응답 내용의 비밀 보장
- (2) 응답자 식별 자료
응답자 성명 및 주소
- (3) 지시사항
설문 작성에 필요한 지시 내용 (응답 방법, 응답 순서, 회수 방법 등)
- (4) 필요 정보 항목
- (5) 응답자의 특성
인구 통계적 특성 (성별, 교육 수준, 나이, 소득)

3. 설문서 작성하기

- (1) 목적과 범위를 확정한다.
설문 목적 및 범위를 명확히 해야 한다.
- (2) 문항의 수를 결정한다.
 - (가) 문항의 수가 많으면 응답 소요 시간이 증가하고 응답자의 불성실 응답이 우려된다.
 - (나) 문항의 수가 적으면 응답자의 내용 이해도가 저하되고 충분한 분석 도구의 가치가 저감된다.
 - (다) 문항의 수를 최소한으로 하되 충분한 분석 도구로서 설문서가 쓰일 수 있도록 결정해야 한다.
- (3) 질문의 형태를 선정한다.
 - (가) 개방형 질문
 - 1) 문항에 대해 응답자가 자유로이 대답한다.
 - 2) 결과가 통일성이 없어 통계처리가 어렵다.
 - 3) 무응답이 많고 시간이 많이 걸린다.
 - 4) 모집단의 내용을 잘 모르는 탐색적 조사에 쓰이면 좋다.
 - (나) 폐쇄형 질문
 - 1) 응답자가 선택 가능한 대답의 내용이 한정된다.
 - 2) 응답 결과가 통일성을 갖고 통계 처리가 매우 쉽다.
 - 3) 응답자의 응답 욕구를 다 충족시키지 못한다.
 - (다) 찬부식 질문
 - 1) 응답자가 질문에 대한 답으로 찬성, 반대 중에서 하나를 택하는 방법이다.
 - 2) 모든 견해를 이분법으로 구분하여 중간 견해를 무시한다.
 - 3) 찬반의 확실한 의사를 묻는 조사에서 유용하다.
 - (라) 다항 선택식 질문
 - 1) 폐쇄형 질문과 유사하나 두개 이상의 답을 선택할 수 있다.
 - 2) 하나의 선택으로 문제를 해결할 수 없는 경우 이용된다.

3) 선택의 기회가 많은 반면 많은 응답 문항을 만들어야 하고 서로 중복되지 않도록 해야 한다.

(마) 체크리스트

- 1) 다항 선택식 질문의 한 형태이다.
- 2) 많은 응답 항목 중 응답자에게 해당되는 모든 것을 선택하게 하는 방법이다.

(바) 서열식 질문

- 1) 문항에 대한 가능한 모든 대답을 열거한 다음 이들을 순서대로 나열한다.
- 2) 응답자가 순서를 정하는데 곤란을 겪을 수 있으므로 열거한 내용이 너무 많아서는 안된다.

(사) 평정식 질문

- 1) 응답 내용이 동일한 것으로 다만 그 강도를 달리하는 응답을 요구하는 방법이다.
- 2) 응답의 수에 따라 3단계, 4단계 및 5단계 평정식으로 불리운다.

(4) 질문의 어구를 구성한다.

- (가) 모호한 질문은 피한다.
- (나) 유도성 질문은 피한다.
- (다) 이중 의미의 질문은 피한다.
- (라) 응답 불균형이 예상되는 질문은 피한다.

(5) 질문의 순서를 결정한다.

- (가) 질문의 순서를 어떻게 정하느냐에 따라 응답자가 응답하는데 영향을 받게 해서는 안 된다.
- (나) 고려사항
 - 1) 첫 질문은 응답하기 간단하며 흥미를 유발할 수 있는 것으로 한다.
 - 2) 뒤의 질문이 앞 질문에 영향을 받지 않도록 한다.
 - 3) 어려운 질문은 분산하여 한다.

(6) 척도를 구성한다.

(가) 측정

연구 대상의 속성에 대하여 수치를 부여하는 과정이 필요하다.

(나) 척도의 유형

1) 명목 척도

측정 대상의 특성을 분류하는 것을 목적으로 숫자를 부여한다.

2) 서열 척도

측정 대상간의 순서 관계를 숫자로 표현한다.

3) 등간 척도

속성에 대한 수치 사이의 간격이 동일한 경우 양적인 정도의 차이를 나타낸다.

4) 비율 척도

등간 척도에 추가적으로 측정 값 간의 비율 계산이 가능하다.

(7) 무응답에 대한 대처를 한다.

(가) 신뢰감이 들도록 조사 취지를 설명한다.

(나) 휴가철, 업무 시간, 식사 시간을 피한다.

(다) 쉽고 짧게 설문 내용을 설명한다.

(라) 보상과 벌칙을 활용한다.

〈4주차〉		
관련 능력단위요소	수업내용(관련 수행준거)	수업방법/기자재
현장별 설문조사 실시하기 (1401030103_14v2.2)	<ul style="list-style-type: none"> 설문대상자 선정 및 설문서 작성(2.1) <ul style="list-style-type: none"> 설문조사 모의 실습 : 모집단 구성, 대상자 프레임 설정 설문서 작성 등 설문결과 분석(2.2) <ul style="list-style-type: none"> 설문 문항의 분석, 자료 분석 및 통계처리 조사결과 분석 실습 	이론강의, 실습/ 컴퓨터 빔프로젝터

1 기초 통계 용어

1. 모집단

관심의 대상이 되는 전체 단위의 모임을 말하며 크게 유한 모집단과 무한 모집단으로 구분한다. 우리가 일반적으로 모집단이라 함은 보통 무한 모집단을 가리킨다.

(1) 유한 모집단

일정 지역의 인구 수와 같이 모집단을 구성하는 통계 단위가 한정된 경우

(2) 무한 모집단

주사위를 던지는 실험의 경우처럼 단위의 수가 무한한 경우

2. 표본

모집단으로부터 추출한 단위의 부분이나 부분집합을 말하며 모집단을 추정하기 위하여 활용한다.

3. 모수와 통계량

(1) 모수

모집단의 특성을 나타내는 값을 말한다. 모수의 값은 고정되어 있지만 보통 미지이다.

(2) 통계량

표본 평균, 표본 분산 등과 같이 표본의 특성을 나타내는 수치를 말한다. 표본을 뽑을 때마다 표본의 특성은 다르게 나타나는 것이 일반적이다.

4. 계량형 데이터

(1) 연속량으로서 측정되는 품질 특성의 값으로 길이, 질량, 시간 등 일반적으로 측정되는 대부분의 값은 계량형 데이터에 속한다.

(2) 계량형 데이터는 계수형 데이터 보다는 공정에 대한 정보를 많이 제공해 주는 장점이 있다.

5. 계수형 데이터

(1) 불량품의 수, 결점의 수 등과 같이 개수를 세어 얻어지는 품질 특성의 값을 의미하며 다음과 같이 대 표적으로 분류 할 수 있다.

(가) 불량품 데이터

합격 및 불합격과 같이 2가지로 분류되는 데이터

(나) 결함 데이터

결함의 수를 셀 수 있는 데이터

(2) 주문서에 기입 사항이 누락된 항목수, 계약서에서 타이핑 에러의 수 등

6. 중심값의 척도

(1) 평균

(가) 모든 데이터들의 합을 데이터 갯수로 나눈 값

(나) 모든 데이터의 영향을 반영한다.

(다) 이상 데이터에 대한 영향이 크다.

(2) 중앙값

(가) 데이터를 크기순으로 나열 시 중앙에 위치한 값

(나) 데이터가 짝수일때는 중앙의 2개 데이터의 평균이 중앙값이 된다.

(다) 계산에 있어 모든 데이터가 필수적으로 포함되지는 않는다.

(라) 이상 데이터에 대한 영향이 작다.

(3) 최빈값

데이터의 집합에서 가장 빈도수가 많은 수

7. 산포의 척도

(1) 범위

산포의 가장 간단한 측정 수단으로서, 최대값과 최소값의 차이를 말한다.

(2) 분산

편차 제곱합을 데이터 개수로 나누어 데이터 한 개당 평균 편차 제곱합을 의미한다.

8. 표준 편차

(1) 정의

모집단의 중심값으로부터 개별치가 흩어져 있는 정도, 산포의 크기를 나타낸다.

(2) 활용

어떤 집단의 평균과 표준 편차를 알고 있다면 그 집단의 분포 형태 및 위치에 대한 예측이 가능하다.

9. 정규 분포

(1) 정의

(가) 모든 확률 분포 중에서 대표적이며 가장 많이 쓰이는 분포이다.

(나) 일반적으로 측정 관리되어 지는 계량형 데이터는 정규 분포를 따른다.

(다) 제조 현장에서 어떤 결과물의 평균과 표준 편차를 알고 있다면, 분포를 예측할 수 있으며 분포 형태에 의거하여 그 결과물의 품질을 예측할 수 있다.

(2) 특징

(가) 분포 형태가 평균을 중심으로 좌우 대칭형이며 종모양이다.

(나) 확률 밀도 함수를 적분시 면적은 1이다.

[수행 순서]

① 설문 결과를 분석한다.

1. 설문 결과 자료 분석 계획을 작성한다.

설문 결과에 따라 작성된 자료에 대한 분석 계획을 작성한다.

2. 설문 문항을 검토한다.

(1) 설문 문항 중에서 일부분이 불완전하지는 않은가?

(2) 응답자가 이해하지 못하는 응답 패턴이 나타나거나 자료를 수집하는 진행 절차에서 오류가 발생되지는 않았는가?

(3) 응답 결과의 분산이 0에 가깝지는 않은가? 가령, 5점 척도에서 모든 응답자가 3점을 체크하지는 않았는가?

(4) 설문지의 일부 페이지가 누락되어 있지는 않았는가?

(5) 설문 응답이 마감 날짜 이후에 제시되지는 않았는가?

(6) 응답자가 연구 설계에서 정한 연구 대상의 자격 요건에 부적절하지는 않은가?

3. 설문 결과를 편집한다.

(1) 자료의 정확성과 정밀성을 높일 목적으로 문항에 대한 응답을 검토하는 과정이다.

(2) 판독하기 어려운 응답이나, 불완전한 응답, 일관성이 없는 응답, 애매 모호한 응답, 기록이 제대로 이루어지지 않은 응답 등에 대한 문제를 검토한다.

(3) 잘못된 응답을 처리한다.

(가) 재조사

해당 응답자에게 재조사한다. (응답자 확인 가능, 표본의 수가 적을 때 사용 가능)

(나) 결측치 처리

그 값을 결측치로 처리한다.

(다) 응답자 제거

해당되는 응답자 전체의 응답 결과를 제거한다.

4. 코딩한다.

각 항목의 응답 결과를 (대부분의 경우 숫자로 표현되는) 코드로 어느 위치에, 어느 정도의 크기로 할당하여 기록할 것인지를 결정하고, 실제로 기록하는 과정이다.

5. 자료를 정제한다.

(1) 결측치를 어떻게 처리할 것인지에 대한 문제와 자료의 일관성에 대한 검토를 심화한다.

(2) 결측치 처리 방법

(가) 중간값으로 대체한다.

해당 변수의 평균값으로 대체한다.

(나) 임putation에 의한다.

결측치를 갖는 응답자의 특성을 파악하여 비슷한 유형의 응답자가 어떠한 응답을 하였는지를 분석하여, 결측치가 어떠한 값을 갖게 될 것인지를 추정하는 방법이다.

주) imputation은 데이터에서 비어 있는 값들에 어떤 값들로 채워 넣는 것을 뜻한다.

(다) 응답자를 제거한다.

결측치가 한 개라도 있는 응답자를 모두 제거시킨다.

(라) 결측치를 제거한다.

분석에서 오직 완벽한 응답 자료만을 이용한다.

6. 통계를 이용하여 자료를 조정한다.

(1) 가중치를 이용하여 조정한다.

응답자에게 중요도에 따라 가중치를 할당하는 방법으로 모집단의 대표성을 임의 조작하는 것이므로 보고서에 이유를 함께 명시하여야 한다.

(2) 변수를 재구성한다.

변수를 변환시키거나, 기존의 변수를 이용하여 새로운 변수를 정의한다.

(3) 척도를 변환한다.

다른 변수가 가지고 있는 서로 다른 측정 단위나 척도를 비교하여 분석할 수 있도록 자료를 변환시켜 준다.

7. 자료 분석 방법을 선택한다.

(1) 자료 특성에 따른 분석 방법의 분류

(가) 모수형 분석 방법

1) 단일 모집단

t-검정, Z-검정

<단일표본 t-검정, t-test>

- 한 집단에서의 평균값이 설정한 값(가설) 또는 모집단이 평균값과 차이가 있는지 분석
- 예시

(문제) A 라는 사람이 매일 아침 베이커리에 들어서 식빵을 구매합니다. 이 식빵의 무게는 500g 이라고 표시되어 있습니다. 하지만 이 A라는 사람은 자신이 구매하는 빵이 500g 이 맞는지 의구심이 들기 시작합니다. 그래서 일주일 동안 식빵을 사온 뒤 매일 무게를 기록하였고 다음과 같습니다.

495, 496, 502, 494, 499, 501, 493g

평균 497.1g , 표준편차 3.53 g

(분석) (1) 일반적 판단 : 평균값이 497.1g 이므로 500g 보다 작다.

(2) 통계적 판단

a) 단측검정(크다, 작다)

귀무가설(원하지 않는 가설) : 평균값이 500g 보다 큼

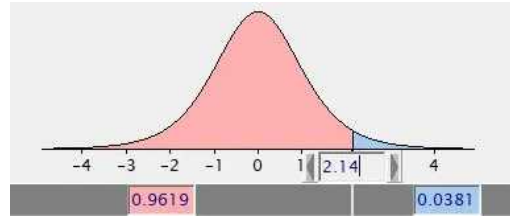
대립가설(원하는 가설) : 평균값이 500g 보다 작음

통계량 계산 : $t = -2.14$, 유의수준 $p = 0.0381 < 0.05 \rightarrow$ 귀무가설기각, 대립가설 채택

통계량 계산 근거(프로그램, 엑셀 등 활용)

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(표본집단 평균 - 모집단평균) / (분산 / $\sqrt{\text{표본수}}$)



따라서, 빵은 500g 보다 작다.

b) 양측검정 (같다. 같지않다.)

귀무가설(원하지 않는 가설) : 평균값이 500g 과 같다.

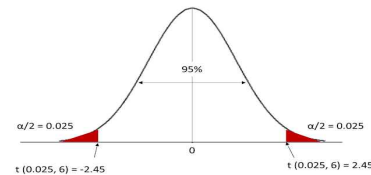
대립가설(원하는 가설) : 평균값이 500g 과 같지 않다.

통계량 계산 : $t = -2.45$, 유의수준 $p = 0.076 > 0.05 \rightarrow$ 귀무가설채택, 대립가설기각

통계량 계산 근거(프로그램, 엑셀 등 활용)

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(표본집단 평균 - 모집단평균) / (분산 / $\sqrt{\text{표본수}}$)



따라서, 빵은 500g 과 같다.

c) 통계적 판단의 중요성

- 통계적 가설을 500g 과 같다/같지않다의 문제로 보면 500g 과 같다는 결론이고, 500g 보다 작다/크다의 문제로 보면 500g보다 작다는 것이 맞음

<단일표본 Z-검정, Z-test>

- 모집단의 통계치와 표본의 통계치를 비교하는 방법으로 모집단의 분포는 정규분포로 가정함
- 예시
 - 어떤 집단의 오랜 기간동안의 수학점수의 평균은 75점, 표준편차는 15점으로 정규분포로 나타났다. 올해 100명의 학생들이 수학성적을 확인한 결과 평균 79점이 나왔다. 올해 수학점수의 평균은 기존의 수학점수의 평균과 같다고 볼수 있는가?

2) 모집단의 수가 두 개 이상인 경우

t-검정, ANOVA

<t-검정, t-test, Z-test>

- 서로 다른 두 개의 집단간의 평균값이 충분히 차이가 나는지 여부를 분석하는 기법
- 예시
 - 20대 남녀 100명을 무작위로 선택해 몸무게를 측정한 결과 남자의 평균 몸무게는 68kg, 표준편차는 10kg, 여자의 몸무게는 66kg, 표준편차는 9kg이었다. 남녀의 몸무게의 차이가 있는지 여부를 검증하라.

<ANOVA>

- ANOVA는 분산 분석으로 세 개 이상의 독립 그룹의 평균 사이에 통계학적 차이가 있는지 판단한다. 만약 데이터 그룹 내의 분산이 크다면(데이터가 평균과 멀리 떨어져 있음), 데이터에서 선택된 표본의 평균이 확률상 다를 가능성이 더욱 커진다. 그래서, ANOVA는 데이터 그룹 내의 분산을 확인할 뿐만 아니라 표본 크기와 표본 평균 사이의 차이를 고려한다.
- 예시
 - 연령, 성별 또는 소득이 당신의 매장에서 한 달 동안 지출하는 금액에 영향을 미칠까요?
이 질문에는 3가지 독립 변수와 1가지 종속 변수가 존재하기 때문에, 질문에 답하려면 당신은 서로 다른 연령 그룹(예: 0-20, 21-40, 41-70, 71 이상), 소득 계층 및 모든 성별 관련 데이터를 수집해야 하고, 그 다음 이원 분산 분석으로 종속 변수(지출액)에 이러한 변수들의 영향을 동시에 평가하고, 차이가 있는지 검증한다.

(나) 비모수형 분석 방법

1) 단일 모집단

카이 제곱 검정

2) 모집단의 수가 두 개 이상인 경우

카이 제곱 검정, ANOVA, 윌콕슨 검정

<카이 제곱 검정, χ^2 검정>

- 카이제곱검정이란 카이제곱 분포에 기초한 통계적 방법으로 관찰된 빈도(obs)가 기대되는 빈도(exp)와 의미있게 다른지의 여부를 검증하기 위해 사용되는 검증방법이다. 카이제곱검정의 목적은 데이터의 비율, 분포, 빈도 등을 검정하는 것이다. 즉, 주어진 데이터가 예상되는 특정 분포와 동일한 분포를 나타내는지에 대한 가설검정이라고 할 수 있다.
- 예시
 - 한 사회학자는 사람들의 직업의 분포가 아래와 같다고 가정하였다. <“농업 20%, 노동자 30%, 공무원 30%, 전문직 15%, 관리자 5%”> 이 가설을 검증하기 위하여 850명의 표본을 조사한 결과가 다음과 같을 때 이 사회학자의 이론은 실제 직업분포를 잘 설명하고 있다고 할 수 있는가?

(2) 분석 형태에 따른 분석 방법의 분류

(가) 의존 관계에 대한 분석 방법

1) 단일 변량 분석

ANOVA, 회귀 분석

2) 모집단의 수가 두 개 이상인 경우

MANOVA, 다변량 회귀 분석

(나) 상호 의존 관계에 대한 분석 방법

1) 변수들 사이의 의존 관계

주 성분 분석, 요인 분석

2) 관측 값들 사이의 의존 관계

군집 분석, 다차원 척도법