

시장조사론

조사의 종류 / 1주차 1차시

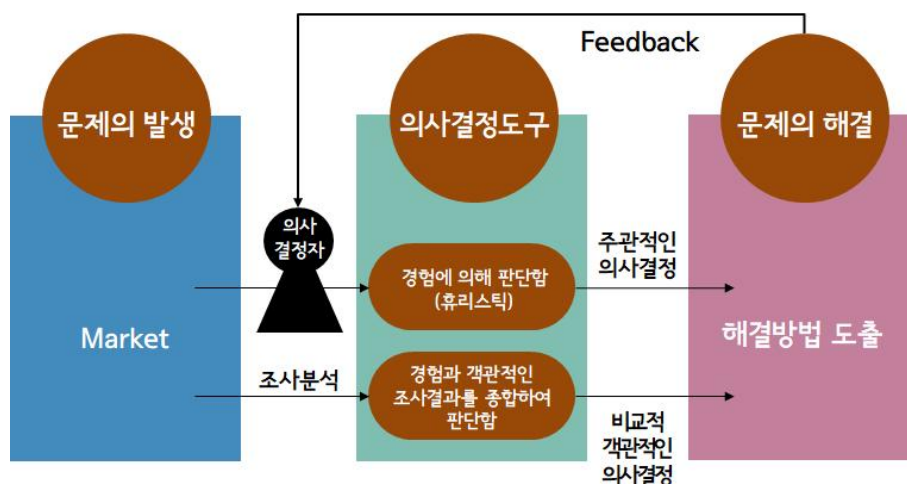
1. 조사의 종류와 조사유형의 선택

1) 경영의사결정을 위해 사용하는 방법

- 의사결정자는 문제가 발생하면 문제해결을 위하여 다양한 방법을 사용
- 사용되는 의사결정방법
 - 경험에 의한 휴리스틱
 - 객관적인 조사결과를 종합하여 판단하는 방법

2) 반복적인 경영의사결정의 결과

- 의사결정 후 적용된 결과는 새로운 경험으로 축적
 - 문제가 발생되면 새로운 조사결과와 함께 의사결정에 활용
 - 이러한 과정의 연속으로 경영의사결정의 질 향상
- 정과정



2. 탐색조사

1) 경영의사결정을 위한 단순화 전략과 휴리스틱

- 대부분의 경영의사결정자들은 문제에 대한 한정된 정보와 상황적 제한 그리고 인간으로서 한정된 인지적 능력으로 의사결정을 해야 하므로 여러 가지 단순화 전략을 활용
 - 휴리스틱은 짧은 시간과 적은 비용을 사용하는 용이한 의사결정방법
 - 휴리스틱을 통해 최선의 선택보다는 비교적 적은 시간과 노력으로 신속하고 합리적인 의사결정을 유도하는 것이 가능

2) 휴리스틱의 종류

구분	내용
가용성 휴리스틱	<ul style="list-style-type: none"> - 기억 속에서 얼마나 잘 떠오르는가 하는 정도에 따라 어떤 사건의 발생가능성과 빈도를 추정하는 경향에 기반한 단순화 전략 - 정보가 머릿속에 떠오르는 정도는 객관적인 빈도 외에 빈도와 전혀 관계없는 많은 요소로부터 영향을 받을 수 있음 - 빈도와 관계없이 단순히 생각해 내는 방법의 용이성에 의존하기 때문에 잘못된 판단을 내리기 쉬움 예) 비행기와 자동차의 안전성 비교, 첫 문자가 r인 단어와 세번째 문자가 r인 단어의 수
대표성 휴리스틱	<ul style="list-style-type: none"> - 상품 혹은 사건에 대하여 평가할 때, 평가대상을 특정한 부류로 분류하여 평가하는 단순화 전략 - 사람을 평가할 경우에 사용하면 극히 비난받을 만한 많은 차별적인 요소를 포함 예) 인종, 출신지역, 장애유무 등에 따른 평가
기준설정과 조정 휴리스틱	<ul style="list-style-type: none"> - 판단을 할 때 개인적인 가치기준점을 먼저 설정해 놓고, 최종적인 결정을 위하여 점차 조정하여 가는 단순화 전략 - 최초의 값에 따라 추정하고자 하는 값의 대부분이 결정됨 - 최초의 값이 판단기준으로 전혀 의미를 갖지 못하는 경우에도 최초의 값에서 벗어나지 못함 예) (1X2X3X4X5X6X7X)과 (7X6X5X4X3X2X1)의 추정값 비교

3. 기술조사

1) 객관적인 정보의 요구

- 객관적인 정보를 바탕으로 하는 정확하고 효과적인 경영의사결정의 요구가 증대되는 배경
 - 경쟁심화
 - 기술변화의 가속화로 상품의 라이프사이클 감소
 - 디지털화로 인하여 가용한 정보의 양 증가
 - 정보통신의 발달로 고객취향의 다양성 증가
- 경험에 의한 판단의 한계 : 경험은 과거일 뿐이므로 현재의 급변하는 환경변화에 그대로 적용하기에는 부족함
 - 추가적인 조사를 통하여 문제의 해결에 도움이 되는 객관적인 정보를 추출하여 활용해야 함

2) 조사를 통한 객관적인 정보의 제공

- 객관적인 정보를 제공해 주는 가장 유용한 수단은 조사이므로 객관적인 조사의 필요성은 점차 증가

- 예를 들어, 신상품을 개발할 경우 소비자가 어떤 상품을 원하고, 그것을 만들 수 있는 현재 기술수준은 어떠한지, 또한 그것이 상품화 되었을 경우 시장성은 있는가 등에 대하여 조사하면 비교적 객관적인 정보가 추출가능

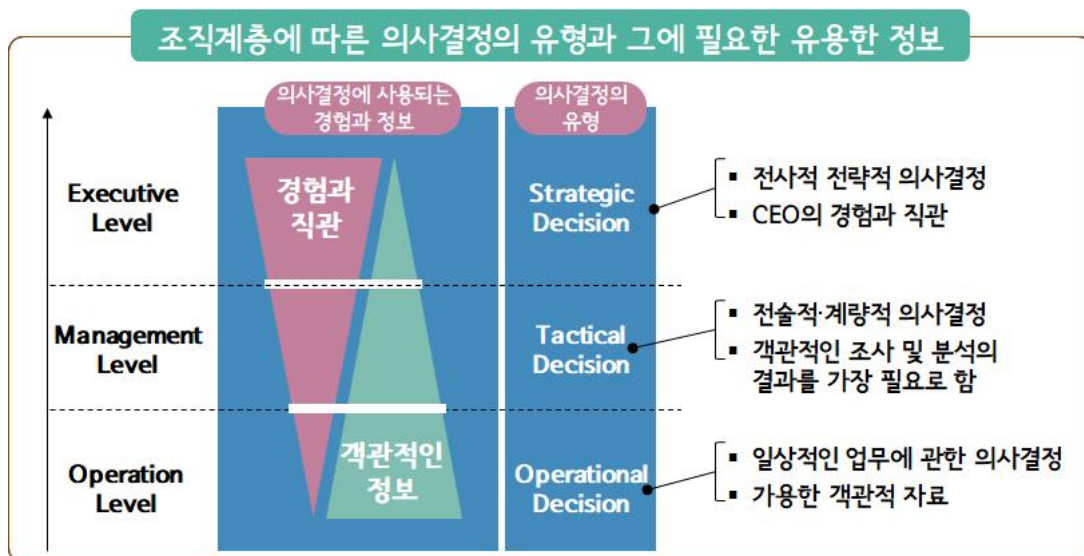
4. 인과조사

1) 조직계층과 정보의 활용

- 일반적으로 기업의 의사결정에서 활용되는 정보는 조직계층별로 달라짐
- 조직의 상위로 올라갈수록 경험과 직관에 의한 휴리스틱이 많이 사용되며, 조직 하부로 내려갈수록 객관적인 정보가 주로 사용

2) 조직계층에 따른 필요한 정보

- 조직의 상부층 : 경험에 의한 직관을 이용하여 의사결정을 하는 가운데 객관적인 정보를 참조
- 조직의 중간 관리자층 : 상부의 계획을 실행으로 옮기기 위하여 보다 세부적인 정보가 필요하기 때문에 객관적인 조사 및 분석 결과를 가장 많이 필요로 함
- 조직의 하부층: 주관적인 판단에 따른 의사결정은 감소하고 더욱 객관적이며, 궁극적으로 특별한 조사를 하지 않아도 손쉽게 얻을 수 있는 정보만을 활용한 일상적인 의사결정이 많음



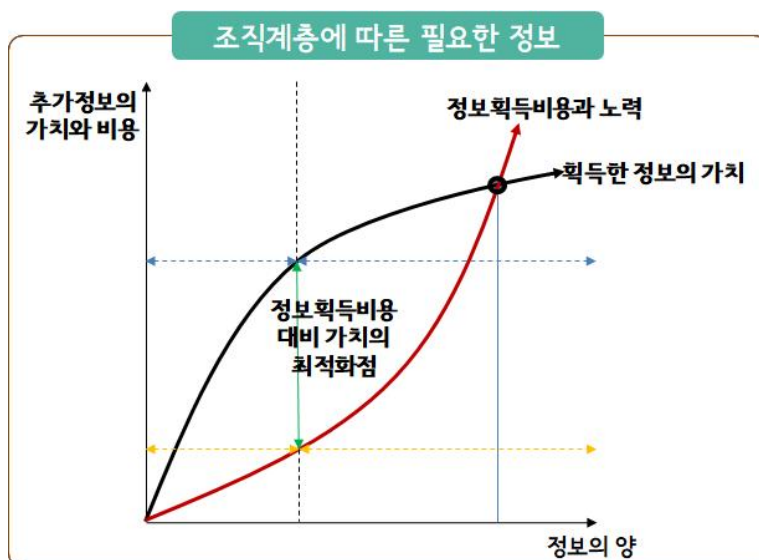
5. 정보획득의 비용과 정보의 가치

1) 정보의 획득과 획득된 정보의 가치와의 관계

- 의사결정의 질을 향상시키기 위해서는 충분하고 정확한 정보가 필요하나 새로운 정보를 수집하는 일은 추가적인 노력과 비용을 유발
 - 많은 정보가 축적됨에 따라 그 이상의 추가적인 정보를 얻기 위한 작업은 더욱더 많은 노력과 비용이 요구됨
 - 한계정보비용이 체증
- 반면에 추가적으로 구한 정보로부터 얻을 수 있는 효과, 즉 정보의 가치는 정보가 축적되어감에 따라 감소
 - 한계정보효용은 체감

2) 정보획득비용과 가치곡선에 대한 이해

- 두 곡선(비용곡선과 가치곡선)이 만나는 점 이내
 - 의사결정의 질 향상에 기여하는 정보의 가치가 이를 수집하기 위한 비용보다 높은 상태
- 두 곡선(비용곡선과 가치곡선)이 만나는 점 이상
 - 추가적으로 얻는 정보의 가치보다 이를 수집하기 위한 비용이 더 크므로 정보를 수집하는 것은 무의미
- 두 곡선(비용곡선과 가치곡선)간의 사이가 가장 크게 벌어진 점
 - 비용 대비 정보의 효과가 가장 크므로 적어도 이 수준까지는 정보를 수집하여 의사결정에 활용하는 것이 바람직함
- 급격하게 증가하는 정보획득비용과 노력에 비하여 추가적인 정보로부터 얻을 수 있는 가치는 오히려 줄기 때문에 어떠한 의사결정도 완전한 정보가 제공된 상태에서는 이루어질 수 없다는 것을 보여줌
- 정보획득비용과 가치곡선



실험설계 / 1주차 2차시

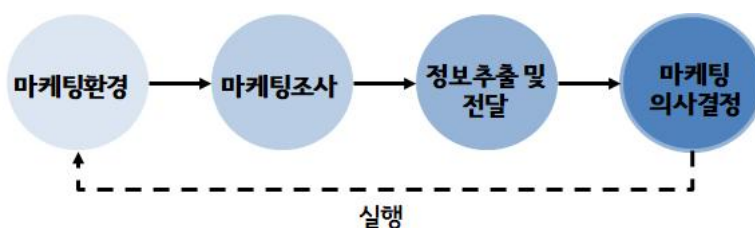
1. 마케팅 조사의 정의

1) 마케팅 조사의 다양한 개념

- 미국마케팅(AMA: American Marketing Association)에서는 마케팅 조사를
"제품과 서비스를 마케팅하는데 관련된 문제에 대해서 정확하고 객관적이며 체계적인
방법으로 자료를 수집·기록·분석하는 일련의 활동
- "Marketing research is the systematic gathering, recording, and analyzing of data about
problems relation to the marketing of goods and service"
- 궁극적으로 기업의 전략적 의사결정을 위해 이루어지는 목적(목표) 지향적인 활동
- 마케팅 의사결정을 위한 정보의 획득을 목적으로 자료를 체계적으로 수집·분석·해석하는
객관적이고 공식적인 과정
- 의사결정에 필요한 유용한 정보는 구체적인 의사결정의 필요와 욕구에 따라 다름

2) 마케팅 조사의 정의와 실시

- 마케팅 조사의 정의
 - 의사결정자의 정보욕구를 진단하고, 그에 따른 정보와 관련된 변수들을 선정하여, 유효하고
신뢰성 있는 자료를 수집·기록·분석하는 일련의 활동
- 마케팅 조사 실시
 - 거시적 환경, 자사, 경쟁사, 소비자의 행동 등 기업을 둘러싼 다양한 환경요소에 대하여
보다 명확하게 조사하고 분석함으로써 마케팅 의사결정에 도움이 되는 신뢰성과 타당성이
높은 정보를 제공하기 위하여 실시
- 마케팅 조사(marketing research)의 활용



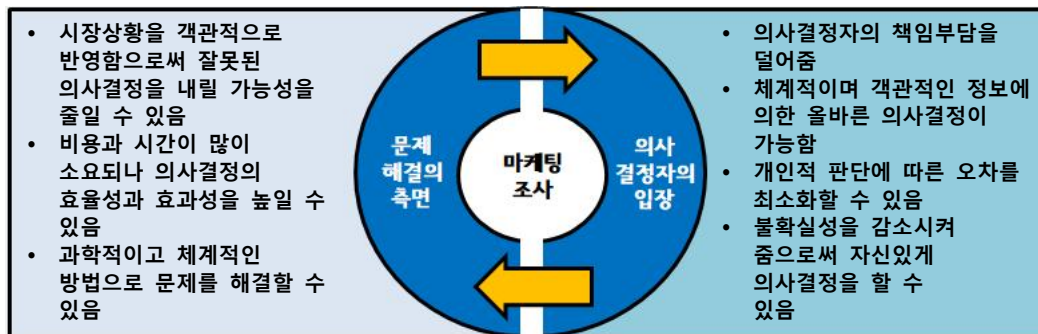
2. 마케팅 조사의 의의

1) 의의

- 마케팅 조사
 - 기업의 경영활동은 많은 수의 연속적인 의사결정을 내리는 것이기 때문에 의사결정자는
의사결정을 내리는 문제에 부딪치게 됨

- 의사결정을 내리는 문제를 해결하는 방법
 - 의사결정자의 경험이나 직관을 근거로 해서 판단하는 의사결정
 - 중요사항에 대하여 객관적으로 조사하여 얻은 결과에 자신의 경험과 직관을 적용하여 판단하는 의사결정
- 대부분 기업의 의사결정은 해당 기업이 수행해 온 경험이나 축적된 데이터베이스부터 추출한 정보를 바탕으로 판단하는 휴리스틱에 준하는 방법이 주로 사용
- 경쟁이 심화되고 급격한 환경변화로 점차 의사결정의 중요성이 높아지고, 내부 구성원 의견과 경험만을 바탕으로 하는 판단에 한계에 다다름
- 기업이 지속적으로 성장발전해 나가기 위해서는 올바른 의사결정이 필요하고, 이를 위해서는 보다 객관적이며 정확한 정보가 필요
- 객관적인 정보를 추가적으로 획득하기 위해 기업은 많은 조사를 필요로 하며, 이러한 조사 중에서 가장 대표적인 조사가 바로 마케팅조사임

■ 마케팅조사의 역할



■ 문제해결 측면에서의 마케팅 조사

- 마케팅 조사를 실시하여 의사결정을 하는 것은 주관적인 판단만으로 의사결정하는 것보다 비용과 시간이 다소 많이 소요됨
- 마케팅 조사를 통한 의사결정은 문제해결의 실마리를 제공하는 실제적인 시장상황을 반영하여 의사결정함으로써 의사결정의 실패 가능성을 줄이고, 의사결정의 효율성과 효과성을 증진시키는 효과를 가짐
- 의사결정자는 마케팅 조사를 실행하기에 앞서서 마케팅 조사를 통해 얻을 수 있는 정보와 조사를 실행하는 과정에 필요로 하는 비용을 정확하게 파악해야 할 필요
- 마케팅 조사를 통해 획득할 수 있는 정보가 의사결정에 어느 정도 도움을 줄 수 있을지에 대해 정확하게 판단하고 합당한 경우 (획득할 수 있는 정보의 가치 > 획득하는데 소요되는 시간과 비용)에만 조사를 실시

■ 의사결정자 입장에서의 마케팅 조사

- 의사결정에 대한 책임을 가지는 의사결정자에게 마케팅 조사는 의사결정자로 하여금 책임에 따른 부담을 덜고, 보다 체계적이며 객관적인 방법으로 의사결정하도록 유도
- 마케팅 조사는 의사결정자의 개인적 판단에서 오는 편견과 이에 따른 오차를 최소화하고, 시장에 대한 객관적이며 정확한 정보를 수집하여 불확실성에 따른 위험을 최소화시킴

- 의사결정자는 마케팅 조사를 통해 비교적 정확한 정보를 이용하여 의사결정을 하게 됨으로써, 자신의 결정에 확신을 가지고 이를 적극적으로 추진함

3. 마케팅 조사의 영역 및 목적

1) 마케팅 조사의 목적

- 마케팅은 기업의 실제적인 마케팅활동을 지원함으로 목표로 함으로써 고객의 태도 및 행동이 기업의 마케팅활동에 어떠한 영향을 미치고 있는가에 대하여 연구하는 학문
- 마케팅 조사는 모든 마케팅 영역에서 필요로 하는 정보를 획득하여 활용하는 대표적인 수단
- 마케팅 조사는 실무영역과 학문영역 두 가지 영역에서 서로 다른 목적을 가지고 실시
- 학문영역의 마케팅 조사 vs 실무영역의 마케팅 조사

구분	학문영역의 마케팅 조사	실무영역의 마케팅 조사
주체	- 학교나 연구실에서 실시	- 기업에서 실시
영역	- 현상의 검증 - 새로운 분석방법에 대한 연구 - 사례의 검증 등	- 자사분석 - 고객분석 - 시장분석 - 경쟁분석 - 마케팅 믹스분석 등
초점	- 체계적이며 시스템적인 분석	- 급변하는 시장상황을 정확하게 이해하고 분석하기 위해 상황에 맞는 창의적이며 즉흥적인 대응능력
목적	- 현상검증 - 새로운 마케팅모형과 이론개발 - 문제도출과 문제해결을 위한 대안제시 - 과학적인 절차에 의한 의사결정 - 위험에 대한 태도가 적극적임	- 실제 시장현황 파악 - 창의적 아이디어 창출 및 과학적 검증 - 과학적이며 창의적인 마케팅 전략수립 - 관리자와의 협의에 의한 의사결정 - 위험에 대한 태도가 소극적임
중요사항	- 과학적 절차에 의한 실행	- 마케팅 담당자와 조사자 간의 지속적이며 효과적인 협의

2) 마케팅 조사의 영역

- 학문영역의 마케팅 조사
 - 대부분 이론(theory)과 과학(science)에 기초- 주로 현상을 연구하여 새로운 마케팅모형이나 이론을 개발하고, 사례를 검증하는 데 활용
 - 경우에 따라서 문제를 명확하게 설명하고, 문제해결을 위한 대안을 제시
 - 과학적이면서 이론과 관련된 내용을 중요시 - 불확실성과 위험이 큰 상황에 직면하는 실무영역보다 위험이나 모험에 대하여 보다 적극적이고 자유로운 자세로 실행

- 실무영역의 마케팅 조사
 - 과학뿐만 아니라 창의성을 강조하는 예술적인 감각과 능력에 기반
 - 계획하고 있는 창의적인 아이디어를 현실적인 시장상황 속에서 과학적으로 검증하거나, 새로운 창조적인 아이디어를 산출하기 위한 방법으로 마케팅 조사를 활용
 - 시장, 고객, 자사, 경쟁사 및 마케팅 믹스를 분석하는 것 뿐만 아니라 이를 통해서 현황을 파악하고 문제해결을 위한 창의적인 마케팅전략을 수립하는 것이 궁극적인 목적
 - 상황에 따라 유연하게 대처해야 하므로 마케팅 담당자와 조사자 간의 지속적이며 효과적인 협의가 필요
- 학문영역의 마케팅조사와 실무영역의 마케팅 조사 간의 관계
 - 실무자들이 직면하고 있는 문제들을 해결하기 위해서는 학문영역의 이론과 분석방법 및 조사기법에 대한 지식이 필요
 - 학문영역의 마케팅 조사는 실제 마케팅현상에 대한 정보와 기초자료를 바탕으로 한 연구를 기반으로 하기 때문에 실무영역에서의 경험과 지식이 필요
 - 두 영역은 서로 별개가 아닌 상호보완적인 관계

4. 마케팅 조사의 문제점과 오남용

1) 마케팅 조사의 문제점

- 마케팅조사는 결코 완전한 의사결정을 보장하는 것이 아니며, 단지 의사결정과정에서 수반되는 불확실성을 합리적인 수준으로 감소시켜 주는 역할
- 마케팅조사에 전적으로 의존하는 것은 오히려 그릇된 의사결정을 유도하거나 마케팅조사의 남용 내지는 오용을 유발할 수 있음

(1) 마케팅 조사의 현실적 문제점

- 마케팅 조사는 정확한 판단력과 다양한 통계적 기법을 필요로 할 뿐만 아니라 풍부한 마케팅 지식과 실무적 경험을 활용해야 하는 매우 복잡한 작업
 - 마케팅 조사가 가지는 현실적인 문제점을 파악하고 이에 대한 대처방안을 마련하는 것이 필요
 - 마케팅조사를 담당하고 있는 많은 실무자들조차 마케팅조사에 관한 지식과 경험이 부족
 - ☞ 조사가 잘못 수행되거나 분석이 틀려도 이를 찾아내거나 이해하지 못하고 있는 경우가 많음
 - ☞ 잘못된 조사부분이나 분석부분을 찾아주고 수정할 수 있는 전문가의 도움이 필요(장기적인 관점에서 마케팅조사의 발전에 가장 큰 저해요인 중 하나)

- 능력을 갖춘 조사자라 할지라도 기업이 당면하고 있는 마케팅문제의 본질을 정확하게 파악하지 못하고, 자의적인 판단에 따라 독자적으로 조사를 수행하는 경우가 많음
 - 조사자는 마케팅관리자와 지속적인 대화채널을 유지해야 함
 - 특히 독선적인 조사를 지양하여 정확하게 마케팅 문제를 찾아내고 충분하게 이해해야 함
- 마케팅 조사와 관련하여 현실적으로 가장 심각한 문제는 조사를 통해 얻은 마케팅 문제에 대한 해결방안이 이를 사용할 마케팅관리자의 관심을 끌지 못하여 무시되는 경우가 많음
 - 조사자가 너무 난해한 기법이나 복잡한 내용으로 의사결정자인 마케팅 관리자를 혼동시키는 경우에 흔히 발생
 - 조사자는 정보이용자의 능력과 욕구에 맞추어 분석된 정보를 효과적으로 전달해야 함(마케팅관리자의 추가적인 요구사항에 따라 상세한 내용이나 사용기법에 대하여 알기 쉽게 설명)
- 마케팅조사에서 활용할 수 있는 다양한 기법들이 충분히 개발되어 있지 못함
 - 마케팅 조사를 통해 사회경제현상이나 인간의 행동을 파악하는 것은 매우 복잡하고 어려운 일
 - 이를 효과적으로 분석할 수 있는 방법이 지속적으로 연구 개발될 필요

(2) 마케팅 조사의 오남용

- 마케팅 관리자가 보다 나은 의사결정을 내리기 위하여 마케팅 조사를 실시함에도 불구하고 다른 목적을 가지고 조사를 행함으로써, 의도적으로 마케팅 조사를 오용하거나 남용하고 또한 오류를 범하기도 함
 - 가장 보편적인 오류는 조사자가 특정한 입장을 지지하기 위하여 마케팅 조사를 실시하는 경우
 - ☞ 조사자가 특정한 사전적인 신념 또는 선입관과 일치하는 자료만을 선택적으로 수집하거나 더 나아가서 자료를 조작하기도 함(가장 심각한 오류)
 - 특정한 목적이 없이 다른 기업이 실시하기 때문에 구색을 갖추기 위하여 실시하는 경우
 - ☞ 특정한 목적이 결여되어 조사결과가 의사결정에 유용하지 못함
 - 이미 내린 그릇된 의사결정에 대하여 사후에 면죄부를 얻기 위하여 실시
 - ☞ 이미 내린 결정을 정당화하기 위하여 사후에 실시하는 경우
 - 홍보나 촉진용으로 실시
 - 예) 일부 광고대행사나 매체사들은 고객들에게 자신들이 과학적이고 현대적인 기법을 사용한다는 인상을 주기 위하여 마케팅 조사를 이용하는 경우가 있음(고의적으로 복잡한 방법을 사용하고, 난해한 통계용어와 마케팅 전문용어를 지나치게 남발하여 고객을 혼동시킴)

- 사전적으로 마케팅활동의 결과로 나타날 수 있는 실패에 대한 책임을 회피하기 위한 근거자료를 마련하기 위하여 실시
 - 이 밖에 많은 조사들이 조사문제의 본질을 잘못 파악하거나, 관련된 변수를 잘못 선정함으로써 조사결과의 가치를 의심하게 만드는 오류를 범함
- 예) 우리 국민이 해외에서 소비하는 경비(여러 곳을 방문하여 지출규모 큼)와 외국인이 한국에서 쓰는 경비(한 곳만 방문하여 지출규모 작음)를 단순비교하는 오류
- (3) 마케팅 조사가 오용 또는 남용되는 경우
 - 특정 의견이나 입장을 지지하기 위하여
 - 특정한 목적 없이 구색만을 갖추기 위하여
 - 이미 내린 결정을 정당화하기 위하여
 - 대외 홍보나 촉진 수단으로 사용하기 위하여
 - 앞으로의 마케팅활동에 대한 책임회피 근거를 마련하기 위하여

조사의 종류 / 2주차 1차시

1. 조사의 종류와 조사유형의 선택

1) 조사의 종류

- 효과적인 의사결정을 위한 정보추출에 사용되는 조사
 - 3가지 대표적인 종류로 구분
 - 탐색조사(문헌조사, 사례조사, 전문가 의견조사, 표적집단 면접법)
 - ☞ 기업의 문제점과 기회 파악
 - ☞ 조사문제의 보다 명확한 규명
 - ☞ 의사결정에 유용한 변수들을 찾아내고 이들 간의 관계에 대한 예비지식 획득
 - ☞ 기업의 다양한 문제와 기회들 간의 우선순위 파악
 - 기술조사(횡단조사, 시계열조사, 패널조사)
 - ☞ 의사결정과 관련된 상황파악과 특정 사건의 발생빈도 조사
 - ☞ 의사결정에 영향을 미치는 변수 간의 상호관계 파악
 - ☞ 특정 값을 예측
 - 인과조사(원시실험설계, 순수실험설계, 유사실험설계)
 - ☞ 특정 현상 간의 인과관계는 어떠한가?
 - ☞ 특정 현상을 구체적으로 정확하게 이해 · 설명 · 예측할 수 있는가?
- 조사유형의 선택 시 고려요소
 - ① 조사목적과 자료의 가용성
 - ② 조사자 자신의 경험과 친숙도
 - ③ 시간 및 자원의 가용성
 - 이를 고려하여 조사를 결정
 - 조사목적이 구체적이지 못하고 이용할 자료가 명확하지 않은 경우
 - 탐색조사 → 기술조사나 인과조사
 - ☞ 탐색조사 결과보다 정확한 정보획득 목적
 - 조사목적이나 구체적인 자료가 명확한 경우
 - 탐색조사 없이 인과조사나 기술조사 실시
 - ☞ 인과조사는 인과관계 검증이 필요한 경우에 실시

2) 조사유형의 선택



3) 조사유형의 특징

조사유형 특징	탐색조사	기술조사	인과조사
주요 조사내용	- 가설설정을 위한 탐색중심의 질적조사	- 특정 시점의 자료를 이용한 가설검증	- 실험을 통한 인과관계 가설검증
조사방법	- 문헌조사 - 전문가 의견조사 - 사례조사 - 표적집단면접법(FGI)	- 횡단조사나 시계열조사 - 패널조사 - 서베이조사	- 구체적 인과관계 검정을 위한 실험설계
질문의 구조	- 질적인 정보중심 - 비구조화된 질문 - 비공식화된 양식 - 가설이 없기 때문에 시행착오식 조사 진행	- 계량적 정보중심 - 구조화된 질문 - 공식화된 양식 - 통계적 분석을 전제로 함	- 계량적 정보중심 - 정밀한 관찰 - 시간적 흐름에 따른 원인-결과분석
조사결과	- 가설탐색에 활용됨 - 일반화에 한계가 있음	- 통계적 분석방법에 의한 객관성 확보 - 가설검정 및 명확한 가설에 기초한 문제해결 방안을 제시함	- 구체적 대안들에 대한 인과구조적 테스트 결과제시 - 의사결정 대안들에 대한 효과분석 결과제시

2. 탐색조사

1) 탐색조사의 정의 및 목적

- 조사의 초기단계에서 조사에 대한 아이디어와 통찰력을 얻기 위해 주로 사용
- 하나의 조사라기 보다는 다른 조사를 수행하기 위한 선행단계로서 실행되는 것이 대부분

(1) 특징

- 당면한 의사결정문제를 보다 분명하게 파악
- 의사결정에 관련된 변수를 찾아내어 문제를 분석하기 위한 명확한 가설을 세우는 데 활용
- 특정한 조사를 수행하기 전에 시행되는 예비조사

- 조사의 초기단계에서 수행되며 상황에 따라 가변성이 높기 때문에 탐색조사를 보다 효율적으로 실시하고 보다 의미있는 정보를 얻기 위해서는 조사자의 예리한 시각과 독창적인 사고가 요구됨

(2) 목적

- 기업이 현재 문제점과 앞으로의 기회를 파악하기 위하여
- 의사결정에 영향을 미치는 유용한 변수들을 파악하고 이들 간의 관계에 대한 예비지식을 얻기 위하여
- 기업의 다양한 문제와 기회들 간의 우선순위를 파악하기 위하여
 - 기업에는 다양한 문제와 기회가 동시에 존재하며 이를 동시에 해결하기 보다는 중대한 영향을 미치고 있는 문제와 기회들을 우선적으로 선택하여 해결해 나가는 것이 필요함
- 조사문제를 보다 명확하게 규명하기 위하여
 - 조사문제와 목적이 명확하게 규명되어야 효과적으로 조사를 계획하고 수행하는 것이 가능함

2) 탐색조사의 종류

종류	내용
문헌조사	<ul style="list-style-type: none"> - 문제규명 및 가설설정을 위한 가장 경제적이며 신속한 탐색조사 방법으로서 기존의 문헌을 조사하는 방법 - 마케팅환경과 마케팅문제에 관한 2차자료 조사 예) 신문이나 잡지, 학술연구지, 정부보고서 및 경제학, 경영학, 심리학, 사회학 등에 관한 2차자료
전문가 의견조사	<ul style="list-style-type: none"> - 해당 문제에 관하여 전문적인 견해와 경험을 가지고 있는 전문가들로부터 정보를 얻어내어 새로운 아이디어를 찾고 문제해결과정에 대한 조언을 구하기 위한 방법 - 문헌조사의 보조적 방법으로 사용됨 - 정형화된 질문보다는 융통성 있는 질문이 필요함
사례조사	<ul style="list-style-type: none"> - 연구자들이 현재 직면하고 있는 상황과 유사한 사례들을 찾아서 분석하는 조사방법 - 문제에 관련된 변수들 간의 관계를 명확히 규명하는 데에 효과적임 - 사례와 문제발생 상황에 대한 면밀한 검토분석이 필요함 - 문헌조사가 일반적 성격의 조사인 반면에, 사례조사는 구체적인 성격의 조사임
표적집단면접법 (FGI: Focus Group Interview)	<ul style="list-style-type: none"> - 전문지식을 보유한 조사자가 소수의 응답자 집단(6명~12명)을 대상으로 특정한 주제에 대하여 자유로운 토론을 벌여 필요한 정보를 획득하는 방법 - 토론 주제의 제시 및 면접집단을 통제하며 토론을 진행할 사회자가 필요하기 때문에 진행자는 토론주제에 대한 기술능력이나 지식과 경험이 풍부해야 함 - 전문가의견조사는 해당 분야에 전문지식을 가진 사람을 대상으로 하는 반면에, FGI는 표적 소비자집단이 되는 일반인들을 대상으로 진행함

■ 표적집단 면접법의 장점과 단점

- 장점 ➔ 기술조사와 인과조사의 기초조사로 활용됨
 - 개별면접보다 유용한 정보를 획득할 수 있음
 - 솔직하고 정확한 의견표명이 가능함
- 단점 ➔ 객관적이고 전문적인 표적집단의 선정이 필요함
 - 사회자의 능력과 경험에 따라 조사결과가 크게 영향을 받음
 - 비계량적인 조사이므로 조사결과에 대한 분석과 해석이 난해함
 - 조사결과의 일반화가 어려움

3. 기술조사

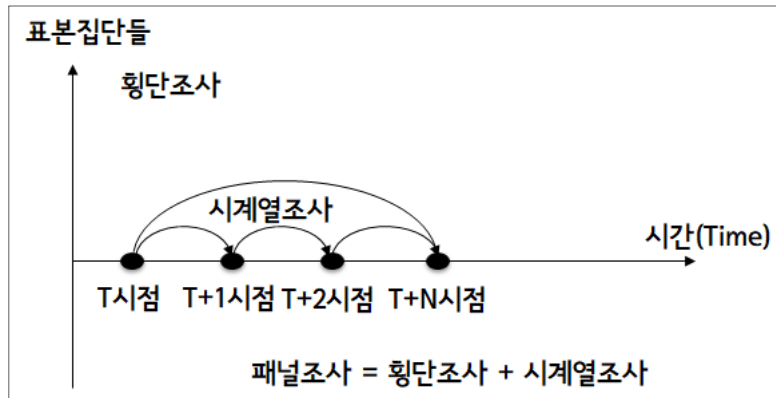
1) 기술조사의 정의 및 목적

- 의사결정에 영향을 미치는 변수들 간의 상호관계를 파악하고 상황변화에 따라 각 변수들에 대한 응답자의 반응이 어떻게 달라지는가를 예측하는 데에 사용되는 조사방법
 - 세부적인 내용에 대한 조사이므로 탐색조사와는 달리 조사목적이 구체적으로 설정되고 조사의 대상이 되는 관련변수들 간의 관계가 어느 정도 파악된 상태에서 실행
 - 목적
 - 의사결정과 관련된 상황에 대한 정확한 파악과 특정 사건의 발생빈도를 조사
 - 의사결정에 영향을 미치는 변수들 간의 상호관계를 파악
 - 특정한 변수값을 예측
- 예) 월별로 상품판매량의 변화를 조사하여 수요를 예측하는 경우

종류	내용
횡단조사	<ul style="list-style-type: none"> - 가장 보편적으로 사용되는 조사방법으로, 여러 조사 대상들을 정해진 한 시점에서 조사하고 분석하는 방법 - 상이한 특성을 가진 집단들을 측정한 값들을 서로 비교함으로써 그 차이를 분석하고 규명하는 조사 - 비교적 큰 규모의 표본을 조사할 경우에 사용함 <p>예) 서베이 조사 : 응답자로 하여금 특정 조사주제에 관한 질문에 답하게 함으로써 실증적으로 자료를 수집하고 분석하는 체계적인 방법</p>
시계열조사	<ul style="list-style-type: none"> - 정해진 조사대상에 관한 특정한 변수값을 여러 시점에 걸쳐 조사하여 변수값의 변화와 그 발생요인을 분석하는 방법 <p>예) 시간의 변화에 따른 주가지수 및 환율의 변동을 측정함</p>
패널조사	<ul style="list-style-type: none"> - 집단으로 구성된 조사대상인 패널(panel)에 대하여 여러 시점에 걸쳐 동일한 현상에 대해 지속적으로 반복 측정하여 조사하는 방법 - 비교적 작은 규모의 조사에서 사용함 - 패널(panel)

	<ul style="list-style-type: none"> ■ 공급자나 거래자 또는 소비자 등으로 구성된 고정된 표본집단 - 패널자료 ■ 패널의 특징과 변화를 시간경과에 따라 여러 차례 반복적으로 측정하여 수집된 자료
--	--

■ 횡단조사, 시계열조사, 패널조사의 관계



4. 인과조사

1) 인과조사의 개관

- 특정한 현상 간의 인과관계를 밝히기 위한 조사
 - 특정 현상을 구체적으로 정확하게 이해하여 설명하거나 예측하고자 하는 경우에 사용
 - 경영자가 의사결정을 할 때 의사결정에 따른 결과를 예상하고 가능한 한 확신을 가지고 의사결정을 하고자 함
 - 의사결정의 원인과 실행결과 간의 연관관계에 대하여 과학적으로 조사하고 분석한 결과가 필요함
- 예) 광고비와 매출액과의 관계를 밝히는 정확한 관측자료를 활용하여 의사결정을 하고자
광고비와 매출액 간의 관계를 규명하는 인과조사를 실시
- 목표 매출액을 달성하기 위한 적절한 수준의 광고비를 결정

2) 인과관계 성립조건

- 원인과 결과가 모두 발생하여야 함
 - 원인으로 추정되는 변수와 결과로 추정되는 변수가 모두 존재하며 이들 간에는 연관성이 있어야 함
- 예) 규모나 모든 여건이 비슷한 상태에서 레스토랑의 서비스 수준이 매출액에 영향을 미치도록 한다면, 원인이 되는 서비스 수준과 결과가 되는 매출액 모두가 발생해야 함

- 원인으로 추측되는 현상이 결과로 추측되는 현상보다 먼저 발생해야 함
 - 레스토랑의 서비스 수준과 매출액이 인과관계가 있다고 가정 할 경우
 - 매출액 증가 후에 서비스 수준이 향상되는 것이 아니라, 직원교육 등을 통하여 서비스 수준이 먼저 향상된 후에 레스토랑의 매출액이 증가해야 하는 논리적인 시간적 우선순위가 성립해야 함
- 외생변수는 엄격하게 통제되거나 제거되어야 함
 - 분석하고자 하는 인과관계와 관련된 원인과 결과변수 외에 이들 간의 관계에 영향을 미치는 다른 변수들을 외생변수라고 함
 - 외생변수의 영향이 없어야 원인과 결과변수 간의 명확한 관계가 파악가능함
 - 예) 서비스 수준 이외에 레스토랑의 매출액에 영향을 주는 요인들로 음식 맛, 청결도, 가격, 분위기 등과 같은 변수들이 있으며 서비스 수준과 레스토랑 매출액 간의 정확한 인과관계를 조사하기 위해서는 이러한 외생변수들은 엄격하게 배제되거나 통제되어야 함
 - 실제 상황에서는 인과관계를 측정하고자 하는 변수 이외의 외생변수들을 완벽하게 통제하기가 어려우므로 인과조사는 실무보다는 주로 학문적인 연구에서 많이 사용됨

실험설계 / 2주차 2차시

1. 실험설계의 개념

1) 실험설계의 정의

- 관심요인(변수)들 간의 관계를 명확하게 규명하기 위하여 효과적인 조사와 분석을 구상하고 계획하는 것
- 조사자는 관찰하고자 하는 현상을 명확하게 정하고 이와 관련된 요인(변수)들이 무엇인지를 파악해야 하며 이들 요인들 간의 관계를 정확하게 분석하기 위한 전반적인 조사 혹은 실험계획과 구상, 즉 실험설계가 필요

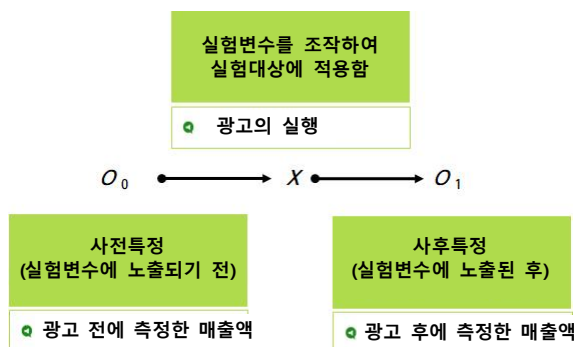
2) 실험설계의 전제조건

- 실험변수의 조작가능
 - 결과가 되는 변수값에 어떠한 변화가 있는지를 파악하기 위하여 실험결과에 영향을 주는 변수나 요인들을 인위적으로 변화시키거나 조작하는 것을 말함
 - 소비자의 제품선택에 용기의 크기가 어떠한 영향을 미치는가를 분석하기 위하여 서로 다른 크기의 용기를 제시함으로써 용기의 크기를 조작함
- 외생변수의 통제 및 제거
 - 실험변수와 결과변수 이외의 변수로서, 결과변수에 영향을 미칠 수 있는 변수들을 통제하거나 제거함
 - 콜라의 맛이 소비자의 제품선택에 미치는 영향분석을 위해서 맛 이외의 변수(브랜드명 등)들을 제거함
- 실험대상의 무작위화
 - 표본이 모집단을 대표할 수 있도록 함으로써 실험결과를 일반화하기 위함
 - 광고효과 측정에 관한 실험에 참가한 소비자들은 표적집단 소비자들 전체를 대표할 수 있도록 무작위로 선정되어야 함

3) 실험설계의 기본모형

■ 실험설계

실험집단의 선정 >> 실험변수의 조작 >> 실험효과의 측정



- 'O'
- 시간변화에 따른 실험대상의 결과변수를 측정한 값
- 'O₀'
- 실험대상이 노출되기 전에 결과변수를 사전에 측정한 값
- 'O₁'
- 실험대상이 조작된 실험변수에 노출된 후에 결과변수를 측정한 값으로 사후에 측정한 값

2. 실험설계의 기본요소

1) 가설

■ 정의

- 2개 변수들 간의 잠정적인 관계를 나타내는 문장으로 실험자가 실험을 통하여 밝히고자 하는 내용을 문장으로 요약한 것
- 예) 광고는 매출액 증대에 긍정적인 영향을 미침(선언문 형식)

■ 좋은 가설이 되기 위한 요건

- 가설에 사용된 요인(변수) 간의 관계가 논리적이고 간결하고 명확하게 표현되어야 함
- 요인(변수)값들은 측정하여 계량화가 가능해야 함
- 내용이 실험을 통하여 검증가능해야 함
- 입증된 결과는 어느정도 일반화가 가능하여 폭넓게 적용할 수 있어야 함

2) 변수

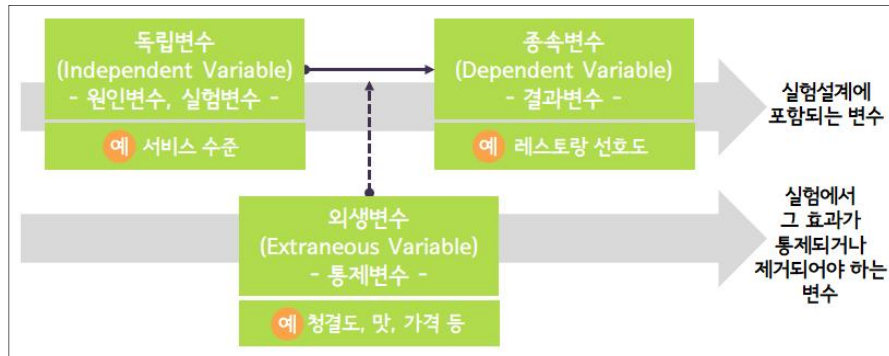
■ 정의

- 관심대상인 개체(item)의 속성이나 특성을 측정하여 기록한 것

■ 종류

- 실험변수(독립변수)
 - 결과변수(종속변수)의 값에 영향을 미쳐 종속변수가 특정한 값을 갖게 되는 원인이 된다고 가정한 변수
- 결과변수(종속변수)
 - 실험변수(독립변수)의 영향을 받아 그 값이 변할 것이라고 가정한 변수
- 외생변수(통제변수)
 - 일종의 독립변수이나 실험변수가 아니므로 그 값이 종속변수에 미치는 영향이 제거되거나 통제되어야 하는 변수

■ 변수들 간의 관계/예시



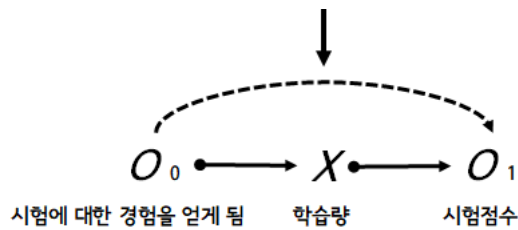
3) 외생변수

■ 정의

- 실험변수가 아니면서 결과변수에 영향을 주는 일종의 독립변수
 - 최대한으로 그 영향이 제거되거나 상쇄될 수 있도록 실험을 설계해야 함
- 예) 종업원이 친절(실험변수) 할수록 패밀리 레스토랑의 매출액(결과변수)이 증가한다는 가설을 실증적으로 검정하는 경우, 종업원의 친절도를 나타내는 변수 외에 레스토랑의 규모와 위치, 음식의 맛, 청결도, 가격수준 등과 같은 매출액에 영향을 미치는 변수인 외생변수는 철저하게 통제되어야 함

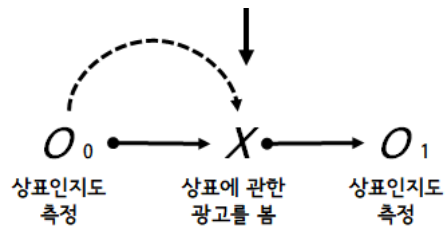
■ 종류

- 우발적 사건
 - 실험자가 전혀 예상치 못한 사건이 발생하여 실험결과에 큰 영향을 미치는 것
 - 예) 맥주시장에서 광고가 매출액에 어떠한 영향을 미치는가를 실험하는 도중 실험대상인 A맥주를 생산하는 기업에 환경오염과 연관된 우발적인 사건이 발생되어 광고와 우발적인 사건이 매출액에 동시에 영향을 미치게 되는 경우
- 성숙효과
 - 실험기간 중에 실험집단의 특성변화가 결과변수에 영향을 미치는 것
 - 예) 어린이들이 공부시간이 시험점수에 어떠한 영향을 미치는 가를 알아보기 위하여 매년 동일한 집단을 대상으로 실험을 실시할 때 어린이들의 지적수준이 크게 향상되어 같은 시간을 공부해도 1년 전보다 높은 점수가 나오는 경우(지적수준의 향상 → 성숙효과 발생)
- 시험효과
 - 실험대상자가 과거 유사한 실험에 참가해 본 경험을 이용해서 응답하는 경우
- 주시험 효과
 - 실험변수의 효과와는 관계 없이 결과변수에 대하여 동일한 측정을 반복함으로써 결과변수값의 변화에 미치는 효과
 - '한 번 시험을 치른 경험' 이 비슷한 환경에서 다시 치는 시험의 점수에 미치는 영향



- 상호작용시험효과

- 실험변수를 가하기 전에 먼저 결과변수를 측정함으로써, 후에 다시 측정할 결과변수에 대하여 생각하게 됨으로써 실험변수 자체의 영향이 달라지게 되는 효과
- '상표인지도 측정으로 높아진 관심'이 광고에 대한 관심을 증대시킴으로 인하여 실험변수의 효과가 더 높아질 수 있음



- 측정방법의 변화

- 실험을 진행하는 관찰자가 바뀌거나 설문지와 같은 측정방법의 변화가 측정결과에 미치는 영향
예) 고객만족지수의 변화를 측정하여 기업의 서비스 수준의 향상 여부를 평가하고자 하는 경우, 만족도 측정시 과거에 사용한 질문이나 질문방식 등 설문지 내용이 바뀌거나 면접자가 교체되는 경우

- 표본의 편중

- 실험변수 이외에 실험의 대상이 되는 집단 간의 차이가 결과변수에 영향을 미치게 되는 것
예) 휘발유 가격인하효과에 대한 민감도를 조사하는 경우, 고소득층 집단보다 저소득층 집단이 가격변화에 훨씬 민감하게 반응하는 경우

- 통계적 회귀

- 실험대상에 대한 측정을 반복할수록 결과변수값이 실험대상의 평균값으로 회귀하는 경향이 있어, 실험변수의 효과를 정확하게 측정하기가 어려워지는 현상
예) 평균기록이 9초인 100m 육상선수가 감기몸살로 10초의 기록이 나온 후 웨이트트레이닝을 열심히 하여 다시 9초대 기록이 달성한 경우
(기록달성은 웨이트 트레이닝 보다는 평균기록으로 회귀한 것으로볼 수 있음)

- 실험대 상의 소멸
 - 실험대상이 실험기간 동안 실험대상에서 이탈함으로써 실험결과가 왜곡되는 현상
 예) 금연 프로그램과 청소년의 흡연율에 어떠한 영향을 미치는지를 분석할 때
 실험대상인 청소년이 금연프로그램 시작 후 성인이 되어 실험대상에서 제외되는
 경우, 판촉효과에 따른 소비자의 구매패턴을 측정하기 위해 패널조사를 실시할 때
 패널에 포함된 실험대상자가 다른 지역으로 이사를 가서 제외되는 경우 등

4) 외생변수의 통제

■ 외생변수를 통제하는 방법

- 제거(Elimination)
 - 실험에 영향을 미칠 수 있는 외생변수를 완전히 제거함
 예) 특정한 경험이 실험에 영향을 미칠 수 있는 경우 그러한 경험을 하거나 혹은 하지
 못한 사람들만을 대상으로 실험을 실시함
- 상쇄(Counter Balancing)
 - 나의 실험집단에 2개 이상의 실험변수가 가해질 경우 실험변수를 번갈아 적용함으로써
 순서에 의한 효과를 상쇄함
 예) 두 개의 광고노출 - 광고순서를 바꿈으로써 순서에 의한 효과제거
- 무작위화(Randomization)
 - 외생변수의 효과를 알 수 없는 경우, 실험대상이 되는 실험집단과 통제집단을
 모집단에서 무작위로 추출하여 외생변수의 영향력을 제거하는 방법
- 균형화(Matching)
 - 외생변수를 사전에 아는 경우 외생변수가 실험대상이 되는 각 집단에 균등하게 영향을
 미칠 수 있도록 실험집단과 통제집단을 선정하여 외생변수의 효과를 통제하는 방법
 ☞ 외생변수가 학력일 경우, 각각의 조사대상이 되는 각 집단이 동일학력분포를 갖도록
 실험집단과 통제집단을 선정함

3. 실험설계의 종류

1) 실험설계란?

- 실험설계는 관심요인(변수)들 간의 관계를 명확하게 규명하기 위하여 효과적인 조사와
 분석을 구상하고 계획하는 것

- 실험결과와 일반화에 영향을 미치는 외생변수의 완벽한 통제와 모집단을 정확하게 대표할 수 있는 표본을 선정하는 것이 가장 중요하지만 거의 불가능함
- 조사의 효율성을 고려하여 주어진 상황에서 가장 효과적인 통제수준과 표본추출과정을 결정

구분기준	원시실험설계	순수실험설계	유사실험설계
실험변수의 노출시점 및 대상에 대한 통제	없음	있음	없음
실험결과와 측정시점 및 대상에 대한 통제	없음	있음	있음
대상선정의 무작위화	없음	있음	없음

- 실험설계에 사용되는 기호

부호	의미
R	실험참가자를 각각의 실험집단에 무작위로 할당함(Randomization)
EG	실험집단(Experimental Group)
CG	통제집단(Control Group)
X	실험대상자를 실험변수에 노출함(eXpose)
O	결과변수를 측정한 값(Outcome)
E	실험효과(Effect)

2) 원시실험설계

- 실험상 외생변수의 통제가 거의 이루어지지 않음
- 인과관계를 비롯해 변수 간의 관계를 정확하게 조사하기에 부적합한 실험설계 가설을 검정하기 보다는 문제를 도출하거나 문제를 명확하게 규명하기 위한 순수실험설계를 수행하기 전에 실행하는 일종의 탐색조사로 많이 활용됨
- 종류

구분	내용	사례
단일집단 사후 실험설계	<ul style="list-style-type: none"> - 실험대상을 실험변수에 노출시킨 후 사후에 실험결과를 측정함 - 실험과 관련된 변수와 실험대상에 대한 통제가 이루어지지 않음 - 실험대상을 실험자에 의하여 임의로 선정됨 - $EG : X \rightarrow O_1$ - $E = O_1$ 	<ul style="list-style-type: none"> - TV 선거방송의 효과를 조사하기 위한 사후 실험설계 - 선거방송 후 방송을 시청한 유권자들 중에서 임의로 300명을 선정하여 선거방송 내용을 얼마나 기억하고

구분	내용	사례
		있는가를 조사함으로써 선거방송의 효과를 측정함
단일집단 사전 사후 실험설계	<ul style="list-style-type: none"> - 사전측정과 사후측정의 차이를 규명함으로써 실험변수의 효과를 측정하는 방법 - EG : $O_0 \rightarrow X \rightarrow O_1$ - $E = O_1 - O_0$ 	<ul style="list-style-type: none"> - TV 선거방송의 효과를 조사하기 위한 사전사후 측정설계 - 선거방송 전 임의로 실험대상을 선정하여 특정 후보에 대한 지지율을 조사하고, 방송 후 그 후보에 대한 지지율을 조사하여 그 차이를 분석함으로써 TV 선거방송의 효과를 측정하는 방법 - 지지율의 변화가 TV 선거방송에 의한 것인지 아니면 다른 요인에 의한 것인지를 정확하게 판단하기 어려움
집단비교설계	<ul style="list-style-type: none"> - 실험변수에 노출된 집단과 노출되지 않은 집단으로 구분하여 실험변수의 효과를 측정하는 방법 - EG : $O_0 \rightarrow X \rightarrow O_1$ - CG : O_2 - $E = O_1 - O_2$ 	<ul style="list-style-type: none"> - TV 선거방송의 효과를 조사하기 위한 집단비교설계 - 실험집단(EG)은 선거방송에 노출시키고 통제집단(CG)은 선거방송에 노출시키지 않은 상태에서 집단간의 지지율 차이로 실험의 효과(E)를 측정

3) 순수실험설계

- 실험대상의 무작위추출과 외생변수에 대한 철저한 통제를 전제로 하는 실험설계방법

- 실제 외생변수의 통제가 어려운 상업적 조사보다는 학문적 조사에 주로 활용

- 종류

구분	내용
통제집단 사전사후 실험설계	<ul style="list-style-type: none"> - 실험대상을 무작위로 실험대상과 통제집단으로 할당함 - 실험변수 도입 전과 후에 실험집단 간의 결과변수값을 측정 - 통제집단에 대해서는 실험변수를 도입하지 않고, 실험집단과 같은 시기에 결과변수값을 사전과 사후에 2번 측정함
	<ul style="list-style-type: none"> - EG : $R \ O_1 \rightarrow X \rightarrow O_2$ - CG : $R \ O_3 \ O_4$ - $E = (O_2 - O_1) - (O_4 - O_3)$

통제집단 사후 실험설계	- 통제집단 사후실험설계는 사전측정을 하지 않는 실험설계방법 - 상호작용시험효과를 제거하고 표본의 무작위 추출로 외생변수의 영향력을 제거함
	- EG : $R \ X \rightarrow O_1$ - CG : $R \ O_2$ - $E = O_2 - O_1$
솔로몬 4집단 설계	- 통제집단 사전사후 실험설계와 통제집단 사후 실험설계를 결합한 형태로서 모든 외생변수의 통제가 가능한 실험설계방법
	- EG : $R \ O_1 \rightarrow X \rightarrow O_2$ - CG : $R \ O_3 \ O_4$ - EG : $R \ X \rightarrow O_5$ - CG : $R \ O_6$ - $E = [O_5 - 1/2(O_1+O_3)] - [O_6 - 1/2(O_1+O_3)]$

4) 유사실험설계

■ 순수실험설계

- 여러 가지 실험조건이 까다롭기 때문에 실제상황에서는 손쉽게 계획하고 실행하기 어려움

■ 원시실험설계


- 무작위화를 무시하여 비교적 현실적이나 외생변수의 통제가 거의 이루어지지 않는 단점을 가지고 있음
 - 서로 상반되는 문제점들을 가지고 있는 실험설계방법들을 서로 절충하여 극단적인 문제를 극복하기 위한 실험설계방법

- 실제상황에서는 순수실험설계에서 요구되는 조건을 완벽하게 갖추기 어려우므로 실험자가 인위적으로 실험상황을 조작하는 것

순수실험설계	유사실험설계
- 실험변수의 노출시점과 대상 그리고 측정시점을 통제 - 실험집단과 통제집단을 모두 무작위로 선정	- 측정시점과 대상에 대한 통제만 가능

■ 종류

구분	내용
2집단 사전사후 실험설계	- 조사대상을 실험집단과 통제집단으로 분류할 수 있으나 실험대상자를 각 집단에 무작위로 배정할 수 없을 경우에 행하는 실험설계방법(nonequivalent control group design)
	- EG : $O_1 \rightarrow X \rightarrow O_2$ - CG : $O_3 \ O_4$ - $E = (O_2 - O_1) - (O_4 - O_3)$

시계열 실험설계	<ul style="list-style-type: none"> - 실험조건상 조사대상을 실험집단과 통제집단으로 분류할 수 없을 경우 사용하는 실험설계 방법 - 동일한 실험대상에 대하여 실험변수를 도입하기 전과 후의 반응을 일정한 간격을 두고 반복적으로 결과 변수값을 측정하여 이들 간의 차이로 실험효과를 측정하는 방법
	- $O_1 \rightarrow O_2 \rightarrow O_3 \rightarrow X \rightarrow O_4 \rightarrow O_5 \rightarrow O_6$
반복 실험설계	<ul style="list-style-type: none"> - 조사대상을 실험집단과 통제집단으로 분류할 수 없고, 실험변수의 효과가 일시적이거나 변화될 가능성이 높을 경우, 동일한 대상에 대하여 일정기간을 두고 반복측정하는 실험설계방법 - 실험변수를 반복적으로 적용하여 그 결과를 비교함으로써 실험효과를 측정하는 실험설계방법
	- $O_1 \quad X \quad O_2 \quad O_3 \quad X \quad O_4$ 

4. 실험의 타당성

1) 실험의 타당성이란?

- 인과조사분석을 통하여 도출된 실험결과가 실험설계의 목적에 얼마나 정확하게 부합되느냐를 나타내는 척도로 내적 타당성과 외적 타당성이 있음

(1) 내적타당성

- 실험효과를 어느 정도까지 정확하게 설계하고 수행하였는가를 나타내는 지표
- 외생변수가 얼마나 효과적으로 통제되었는가?
- 사례
 - '종업원이 친절(실험변수)할수록 레스토랑의 매출액(결과변수)이 증가한다'
 - ☞ 친절도 이외에 모든 외생변수들은 제거하거나 통제되어야 함

(2) 외적타당성

- 실험에서 도출된 결과를 실제상황에 적용할 경우 어느 정도 적합한가를 나타내는 지표
- 실험결과를 일반화하여 실제 사회현상에 확대 적용시킬 수 있는가?
- 실험이 현실과 유사하게 설계되면 될수록 외적타당성은 높아져 실험실 결과를 일반화하여 실제 사회에 확대 적용할 수 있음

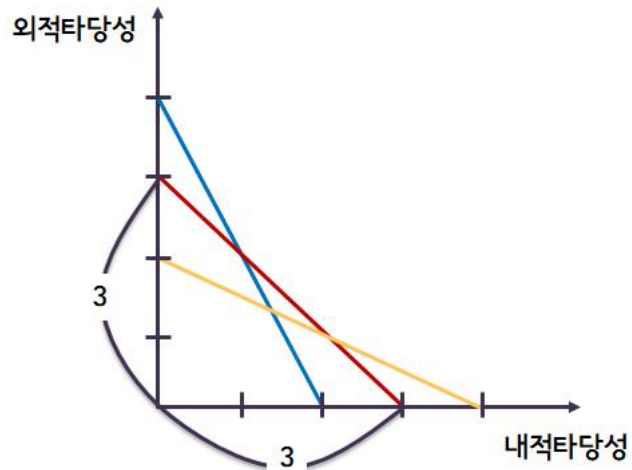
(3) 내적타당성과 외적타당성의 관계

- 외적 타당성과 내적 타당성은 서로 트레이드 오프(trade-off) 관계에 있음
- 내적 타당성을 높이기 위하여 실험조건을 엄격히 통제한다면 현실과 동떨어져 실험결과를 일반화시켜 실제상황에 적용시킬 수 있는 외적 타당성은 낮아짐

- 반대로 현실에 가깝게 실험을 설계하여 외적 타당성을 높이면 표본의 무작위화나 외생변수의 통제가 제대로 이루어지지 않기 때문에 실험 자체의 내적타당성이 낮아져서 신뢰성이 떨어짐

☞ 조사자는 내적 타당성과 외적 타당성을 상황에 맞게 적절히 조정하여 적당한 수준으로 실험을 설계하고 조사를 수행해야 함

■ 외적타당성과 내적타당성 간의 관계: 트레이드 오프 관계



마케팅조사의 실행과정 : 문제의 발생과 인식 / 3주차 1차시

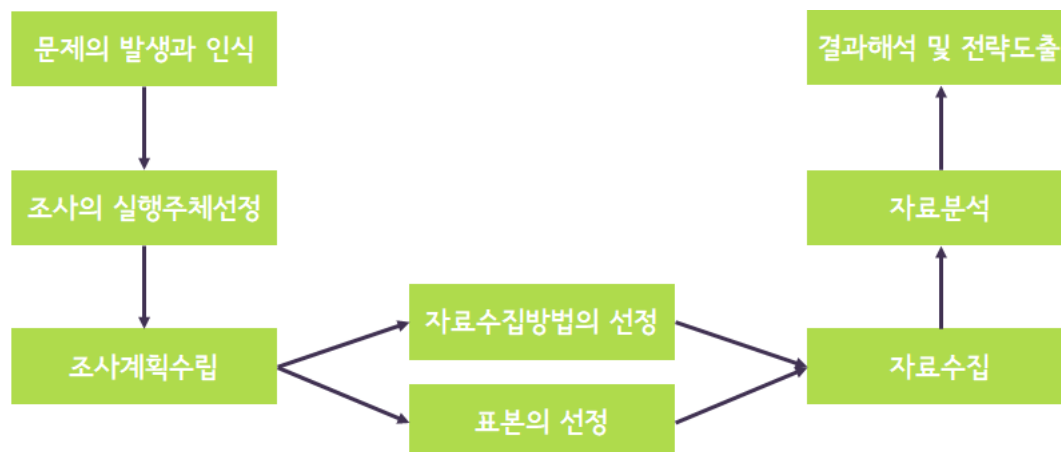
1. 마케팅조사의 실행과정의 개념

1) 마케팅조사의 실행과정

■ 마케팅조사의 질

- 통계학과 수학 그리고 경제원리나 행동과학의 이론을 비롯하여 컴퓨터의 성능과 조사자의 능력 및 경험 등 다양한 요인들에 의하여 크게 영향을 받음
 - 이 중에서 조사자의 능력이 가장 중요하지만 조사자의 개인적 특성과 능력은 달라도 대부분의 조사는 8단계의 과정을 통해 수행됨

■ 마케팅조사 실행과정 8단계



① 문제의 발생과 인식

- 의사결정과 관련된 문제 발생
- 조사의 실행 여부 결정
 - 대부분의 마케팅 조사는 문제의 발생과 인식으로부터 시작
 - 기업에서는 신제품 개발, 시장점유율 하락, 매출액 감소, 경쟁업체 진출 등과 같은 마케팅의사결정의 대상이 되는 문제발생

② 조사의 실행주체선정

- 자체조사 혹은 외부업체에 의뢰
 - 일단 조사하기로 결정하였다면, 실행주체를 선정
 - 조사문제의 성격과 기업의 여건에 따라 기업 내부의 마케팅부서나 외부 전문조사업체들 중에서 조사를 실시하기에 가장 적합한 실행주체를 선정
 - 복수의 조사업체로부터 연구제안서를 제출받아 조사의 목적과 조사에 필요한 시간과 비용, 그리고 조사업체의 규모나 경험 등을 기준으로 심사하여 조사업체를 최종적으로 선정

	기업 내부의 마케팅 부서	외부 전문조사업체
장점	해당 산업분야에 대하여 전문적인 지식과 경험을 보유	마케팅조사에 필요한 조사방법이나 분석방법에 대하여 전문적인 지식보유

③ 조사계획수립
<ul style="list-style-type: none"> ▪ 조사범위와 내용에 대한 명확한 규명 ▪ 조사시간과 비용 합의 ▪ 조사의 종류 결정 ▪ 자료를 수집하는 방법 ▪ 조사대상 선정방법 등 ▪ 구체적인 조사계획 수립
<ul style="list-style-type: none"> • 조사담당자는 조사의 범위와 내용, 조사에 필요한 시간과 비용에 대해 합의하고, 합의결과를 바탕으로 구체적인 실행계획을 수립 • 실행계획에는 조사의 종류와 자료를 수집하는 방법, 조사대상을 선정하는 표본추출방법 등이 포함
④ 자료수집방법의 선정
<ul style="list-style-type: none"> ▪ 기존의 자료를 검토하여 자료수집방법을 결정함
<ul style="list-style-type: none"> • 문헌조사를 통한 2차 자료수집 또는 인터뷰나 설문조사를 통한 2차 자료수집 등 다양한 수집방법을 결정
⑤ 표본의 선정
<ul style="list-style-type: none"> ▪ 기존의 자료를 검토하여 자료수집방법을 결정함
<ul style="list-style-type: none"> • 자료수집방법에서 1차자료를 수집하고자 할 경우에는 조사의 목적을 비롯하여 가용한 시간과 비용 등을 고려해서 적합한 표본의 크기와 표본추출방법 등을 결정
⑥ 자료수집
<ul style="list-style-type: none"> ▪ 표본의 크기와 표본선정방법을 결정함
<ul style="list-style-type: none"> • 자료수집과정에서 발생하는 오차(표본오차와 비표본오차)를 최소화하는 방법으로 자료수집
⑦ 자료분석
<ul style="list-style-type: none"> ▪ 표본오류와 비표본 오류를 최소화할 수 있는 방법으로 자료를 수집함
<ul style="list-style-type: none"> • 수집된 자료를 코딩하여 파일로 만든 다음, 컴퓨터 통계패키지를 이용하여 분석
⑦ 자료분석
<ul style="list-style-type: none"> ▪ 분석결과를 해석하고 이를 바탕으로 유용한 마케팅전략을 도출함 ▪ 의사결정자가 이해하기 쉽고 활용하기 용이하게 보고서를 작성함
<ul style="list-style-type: none"> • 분석한 결과를 해석하여 정보를 추출하고, 이를 바탕으로 의사결정에 활용될 수 있는 전략을 도출하고, 이를 종합하여 보고서를 작성함으로써 조사를 종료

2. 문제의 발생과 인식

1) 문제의 발생과 인식과의 관계

■ 마케팅조사의 출발

- 기업경영에서 발생할 수 있는 전반적인 마케팅의 문제점들을 명확하게 인식하고 규명하는 것부터 시작

■ 당면한 마케팅 문제를 명확하게 이해해야만 어떠한 내용의 의사결정이 필요함

■ 의사결정을 위하여 어떠한 정보가 필요한지를 명확하게 파악가능

■ 의사결정에 필요한 정보를 얻기 위하여 어떠한 조사를 수행하여야 할지에 대하여 판단

■ 마케팅문제에 대한 이해를 바탕으로 조사 문제를 명확하게 정의하는 일은 조사방향을 결정짓고 결과의 활용에 이르기까지 조사활동 전체에 상당한 영향을 미침

- 이 단계가 잘못되면 전체 조사활동이 무의미하게 되며, 심지어는 그릇된 정보를 제공함으로써 잘못된 의사결정을 내리게 되므로 이 단계가 마케팅 조사에서 가장 중요함

■ 마케팅조사를 필요로 하는 분석대상

- 자사를 비롯하여 고객, 경쟁, 시장, 마케팅 믹스 분석 등이 있음

2) 문제의 발생과 마케팅조사 문제제기

■ 명확한 문제인식 및 마케팅조사의 목표 및 방향설정

- 당면한 마케팅문제에 대한 명확한 규명
- 마케팅 문제가 야기된 전반적인 배경 분석
- 2차 자료분석을 통하여 필요한 정보에 대한 판단



■ 문제발생의 단계에서는 기업이 처한 마케팅 문제 자체뿐 아니라 문제가 발생하게 된 배경에 대해서도 철저하게 파악할 필요가 있음

- 문제의 발생배경이 문제해결의 실마리가 되며 문제해결을 위한 방향을 정확하게 찾아낼 수 있음

■ 문제발생과 인식단계에서 일반적으로 거론되는 사항(경영계획과 통제에 필요한 정보유형)

구분	세부항목	내용
경영조건	Customer (고객)	<ul style="list-style-type: none"> - 구매자는 어떤 사람인가(who)? - 그들은 어디에서 구입하는가(where)? - 그들은 어떻게 구입하는가(how)? - 그들은 언제 구입하는가(when)? - 그들은 어떠한 것을 구입하는가(what)? - 그들은 얼마나 구입하는가(how much)? - 그들이 구매하는 동기는 무엇인가(why)? - 이상의 구매행위나 특성이 미래에는 어떻게 변화할 것인가? - 고객들은 어느 정도 만족하는가? - 고객들이 어느 정도 유지되는가?
	company & competitor (회사와 경쟁자)	<ul style="list-style-type: none"> - 제품계열별 시장점유율 - 고객유형별 매출액 - 판매지역별 매출액 - 고객별, 거래처별 자사의 이미지 및 상품 이미지 - 경쟁자의 특성(마케팅 프로그램, 경쟁전략, 경쟁자의 지원) - 마케팅능력, 생산능력, 재무능력, 기술능력 - 내부환경의 미래 동향
	external environment (외부환경)	<ul style="list-style-type: none"> - 경제여건 및 동향 - 정부규제 및 정책방향 - 환경, 안전, 소비자 관심 - 기술동향 - 정치적 분위기
경영성과의 측정		<ul style="list-style-type: none"> - 상품계열별 매출액은 어떠한가? - 상품계열별 시장점유율은 얼마나 되는가? - 고객유형별과 판매지역별 매출액과 시장점유율은 어떠한가? - 자사의 이미지, 상품이미지가 고객별 그리고 거래처별로 어떠한가? - 자사의 촉진활동에 대한 인지도는 어떠한가? - 자사 상표의 회상수준과 인지도는 어느 정도인가? - 지역별, 고객별, 소매상 유형별 거래비율은 어떠한가? - 자사 상품가격(시장)의 변화 정도는 어떠한가?
마케팅의 4P	product(상품)	<ul style="list-style-type: none"> - 어떤 상품특성 및 효용이 중요한가? - 상품을 어떻게 차별화해야 하는가? - 경쟁상품과 비교해서 어떻게 인식되고 있는가? - 어떤 세분시장을 목표로 삼아야 하는가? - A/S 및 품질보증은 얼마나 중요한가? - 상품품목이나 계열을 변화시킬 필요가 있는가? - 포장(디자인)은 얼마나 중요한가? - 상품별 구매예측은 어떻게 하는가? - 상품에 대한 시장테스트는 어떻게 하는가?

	product(상품)	<ul style="list-style-type: none"> - 어떤 상품특성 및 효용이 중요한가? - 상품을 어떻게 차별화해야 하는가? - 경쟁상품과 비교해서 어떻게 인식되고 있는가? - 어떤 세분시장을 목표로 삼아야 하는가? - A/S 및 품질보증은 얼마나 중요한가? - 상품품목이나 계열을 변화시킬 필요가 있는가? - 포장(디자인)은 얼마나 중요한가? - 상품별 구매예측은 어떻게 하는가? - 상품에 대한 시장테스트는 어떻게 하는가?
	place(유통)	<ul style="list-style-type: none"> - 당사 상품을 어떤 유형의 중간상이 취급하도록 하는 것이 가? - 자사 상품을 취급하는 거래처의 동기 및 태도는 어떠한가? - 도매 및 소매를 지역별로 얼마나 깊숙이 침투해야 하는가? - 상품별 마진율은 얼마나 적당한가? - 물적 유통체계는 어떻게 설계해야 하는 것이 좋은가?
	price(가격)	<ul style="list-style-type: none"> - 수요탄력성은 어떠한가? - 어떤 가격정책이 적절한가? - 상품계열별 가격조정은 어떻게 되어야 하나? - 상품품목별 가격변화는 어느 정도가 적당한가? - 경쟁자의 가격 위협에 어떻게 대처하는 것이 좋은가? - 가격이 구매자에게 얼마나 중요한가? - 제품별 마진에 대한 분석은 어떻게 하여야 하는가?
	Promotion (촉진)	<ul style="list-style-type: none"> - 촉진예산의 적정규모는 얼마인가? - 수요를 자극함에 있어서 판매촉진·광고·인적판매의 효과는 어떠한가? - 촉진믹스는 어떻게 하는 것이 효과적인가? - 촉진수단별 효율성은 어떠한가? - 메시지와 매체는 어떻게 선정해야 하는가? - 판매지역의 선정은 어떻게 해야 하는가? - 판매자의 효율성은 어떠한가?

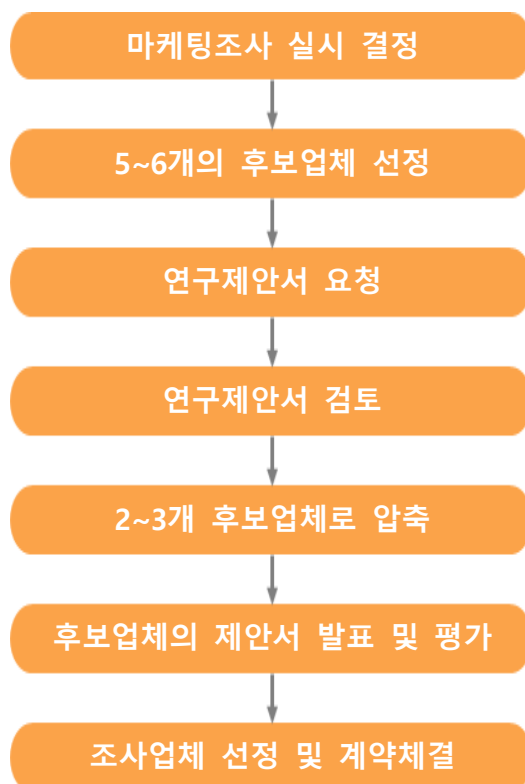
마케팅조사의 실행과정 : 실행주체의 선정과 계획수립 / 3주차 2차시

1. 마케팅조사의 실행업체 선정

1) 실행업체 선정

- 문제가 파악된 후에 마케팅담당자는 조사기획서를 작성하여 상부의 결재를 받음
 - 조사계획서
 - 구체적인 조사목적과 조사를 통해 얻을 수 있는 정보와 기대효과, 조사예산, 조사기간 등이 포함됨
- 마케팅 조사에 관한 결재 후 조사를 실행할 조사업체를 선정하는 단계로 진입함
- 마케팅 담당자는 조사를 가장 성공적으로 수행할 수 있을 것으로 판단되는 전문조사업체를 5~6개 선정하여 연구제안서를 제출하도록 요청한 후 평가하여 약 2~3개의 업체를 선정하고, 선정된 업체들로 하여금 조사진행을 포함한 조사제안서를 발표해 줄 것을 요청함
- 발표를 통하여 가장 효과적으로 조사를 수행하고 필요한 정보를 제공할 수 있을지 객관적으로 평가
 - 가장 높은 점수를 받은 조사업체를 선정한 후 조사와 관련된 계약을 체결함

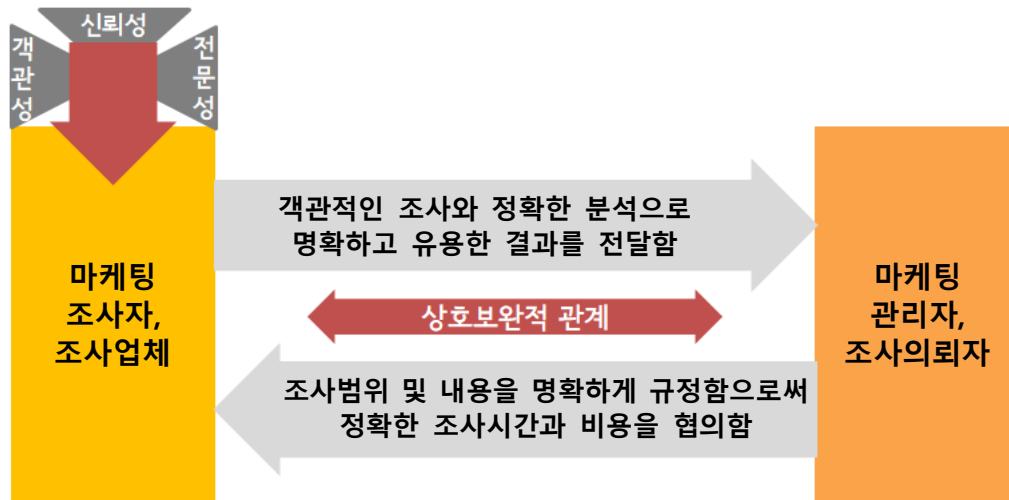
2) 마케팅조사업체 선정과정



2. 조사의뢰자와 마케팅조사업체의 관계

1) 조사 의뢰자와 조사업체의 관계

- 조사를 의뢰받은 마케팅조사자 혹은 조사업체는 자신도 기업의 마케팅 의사결정 과정에 참여를 한다는 책임의식을 가지고 능동적으로 조사를 계획하고 실행해야 함
- 성공적인 마케팅 조사가 이루어지기 위해서는 마케팅조사자(조사업체)와 의뢰한 기업의 조사관리자 사이에는 원활한 커뮤니케이션이 필요함



2) 마케팅 담당자와 조사업체가 지켜야 하는 기본적인 조건

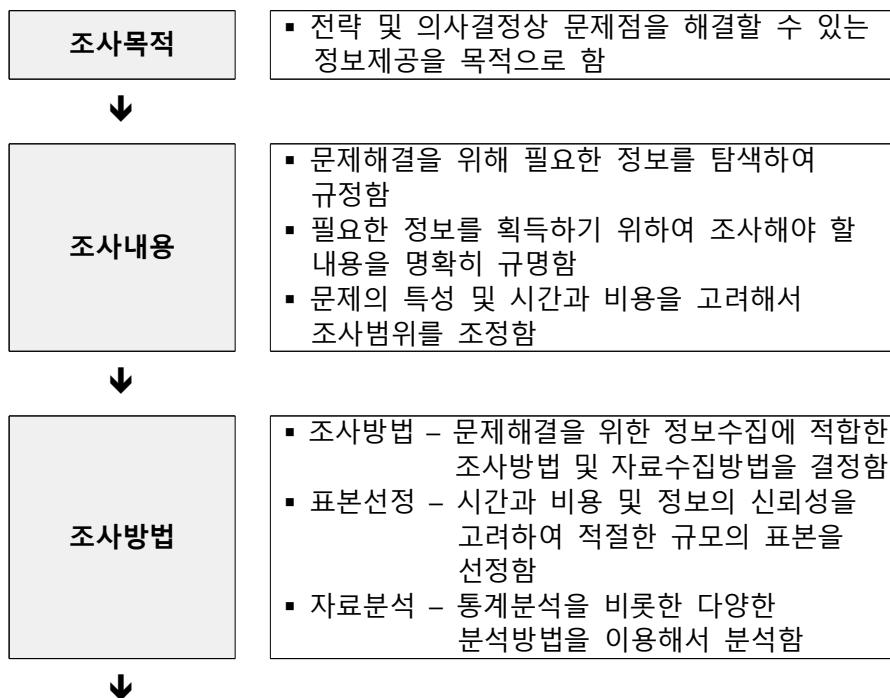
- 마케팅관리자는 조사를 의뢰할 때 처음부터 조사의 범위나 내용을 명확하게 규정해 주어야 함
 - 조사가 시작되면, 조사방법이나 내용을 중간에 수정하기가 매우 어려움
- 조사에 대한 예비지식을 비롯하여 조사에 필요한 시간과 비용에 대해서는 조사관리자와 마케팅조사자 간에 사전에 서로 협의하고 동의해야 함
 - 추가적인 비용발생이나 지체된 시간에 대하여 책임의 소재를 명확히 하여 분쟁가능성을 제거해야 하며 특히 마케팅조사는 타이밍이 중요하기 때문에 지연된 조사는 의미가 없을 수 있음
- 마케팅조사자는 전문적인 용어나 복잡한 수식을 사용함으로써 마케팅 관리자가 조사내용을 명확하게 이해하지 못하도록 하는 경우가 없도록 해야 함
 - 조사 결과를 쉽고 명확하게 전달해야 할 필요가 있음

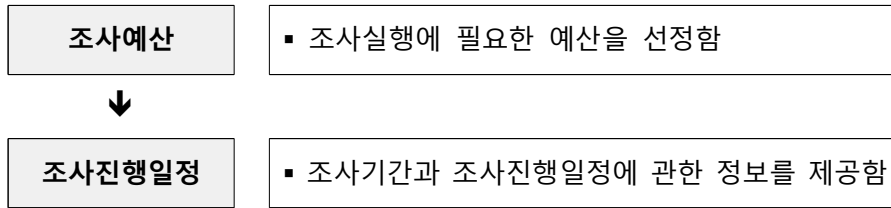
3) 마케팅담당자와 조사담당자의 금기사항

구분	마케팅담당자	조사담당자
내용	<ul style="list-style-type: none"> - 조사에 대한 명확한 목적을 감추는 행위 - 발생될 수 있는 문제점을 알면서도 조사담당자에게 알리지 않는 행위 - 조사의 내용과 상호합의하에 작성된 설문지를 계속 수정하는 행위 - 실사의 시작을 지연시키면서 보고서의 제출기일 엄수를 강력하게 요구하는 행위 - 보고서에 필요 이상의 가능한 많은 표들을 작성해 줄 것을 강요하는 행위 - 조사비 지불을 지체하는 행위 	<ul style="list-style-type: none"> - 자신이 잘 아는 조사방법을 이용하기 위해 조사의 목적을 임의로 수정하는 행위 - 마케팅담당자가 요구한 수정사항을 제대로 반영하지 않는 행위 - 정상적으로 조사가 진행되지 않음에도 불구하고 이를 감추는 행위 - 보고서제출이 지체될 경우 마케팅관리자와의 접촉을 의도적으로 피하는 행위 - 약간의 내용수정으로 높은 조사비를 요구하는 행위 - 최종보고서 제출 전에 계산서를 먼저 보내는 행위

3. 마케팅조사 계획수립

- 조사업체가 선정되면 조사의뢰자와 조사업체는 서로 협의하여 조사의 목적, 조사 내용의 범위, 조사방법, 조사일정 등을 포함하여 전반적인 조사계획을 수립
- 조사의뢰기업에서 제공한 조사의뢰서와 과거 진행한 연구보고서 등 관련된 2차 자료들을 토대로 조사계획을 수립





1) 조사목적

- 조사자는 조사의뢰서와 조사담당자로부터 얻은 정보, 문제와 관련된 자료들을 종합적으로 검토하여 조사목적을 결정
- 조사의뢰기업의 마케팅 전략이나 의사결정상의 문제점을 정확하게 파악하고 이를 효과적으로 해결할 수 있는 방향으로 조사목적을 설정

2) 조사내용

- 마케팅문제를 해결하는 데에 필요한 정보가 어떠한 것들이고, 그러한 정보를 얻기 위해서는 어떤 내용을 조사해야 하는가를 포함한 조사내용 전반에 대한 범위를 결정
- 발생한 문제와 관련하여 가능한 한 많은 사항들을 조사하면 할수록 정보의 정확성과 그 활용도는 커지나, 조사에서 필요한 시간과 비용 또는 증가함으로 문제의 특성과 상황을 고려하여 적당한 수준으로 조사의 범위를 한정하여 구체적인 조사내용을 결정

3) 조사방법

- 필요한 정보를 효과적으로 획득하기 위한 조사방법이나 자료수집방법, 표본선 정방법 및 조사된 자료를 분석하는 방법 등을 결정

(1) 조사방법

- 가능한 여러 가지 조사방법 중에서 당면한 문제해결에 꼭 필요한 자료를 정확하게 수집하고 분석할 수 있는 가장 적합한 조사방법을 선택
- 자료수집 방법과 표본선정방법 등을 고려하여 결정

(2) 표본선정

- 해결하고자 하는 문제와 관련되어 유용한 정보를 제공해 줄 수 있는 대상 전체, 즉 모집단을 가장 정확하고 효과적으로 대표할 수 있는 표본을 선정하여 조사하는 것
예) 신세대 젊은이들을 대상으로 하는 새로운 모형의 휴대폰을 개발하기 위하여 신세대 소비자의 반응을 조사하는 경우: 휴대폰을 사용하는 신세대 전체가 모집단이 되며, 표본은 모집단 전체를 조사할 수 없는 경우 신세대 전체를 대표할 수 있는 소수의 집단
- 편중되지 않고 전체 모집단을 대표할 수 있는 표본을 추출해서 조사해야만 조사 결과를 일반화할 수 있음
 - 표본추출방법이 중요함

- 직접 조사하고자 하는 표본의 수가 증가할수록 필요로 하는 조사시간과 비용이 비례해서 증가
 - 신뢰할 수 있는 정보를 추출하기 위해서는 최소한으로 필요로 하면서 주어진 조사비용과 시간에 의해서 적합한 규모로 표본의 크기를 정해야 함

(3) 자료의 분석

- 수집된 자료의 성격에 따라 사용할 수 있는 통계와 분석방법이 달라지기 때문에 자료수집 전에 분석방법을 미리 정하고 자료수집을 해야 함

4) 조사예산

- 가용한 조사예산과 조사의 시간과 질 등을 감안하여 가장 합리적인 방법으로 조사예산 계획을 수립

- 조사예산에 포함되는 항목

- 간접연구경비, 연구활동비, 인건비 및 연구과제 참여 경비
- 시약 및 소모성 재료비
- 기자재 임차료, 기기 구입비
- 도서구입비(참고문헌 및 자료수집비), 인쇄비
- 국내 여비, 해외 여비, 회의비
- 전산처리비, 기술정보비
- 공과금 및 잡비
- 기타

5) 조사일정 조사계획을 비롯하여 실시과정과 분석 및 보고서 작성기간 등을 고려하여

- 진행일정에 관한 구체적인 계획을 수립

마케팅조사의 실행과정 : 자료수집방법 선정 / 4주차 1차시

1. 의사소통법 1

1) 1차 자료 vs. 2차 자료

■ 1차 자료

- 기존의 2차 자료로부터 얻을 수는 없으나 현재의 의사결정에 꼭 필요하다고 판단되어, 조사자가 직접 조사하여 수집하는 자료

■ 2차 자료

- 현재 의사결정에 활용할 목적으로 조사된 것이 아니라, 과거에 다른 목적으로 조사되어 수집된 자료로서 현재의 의사결정에도 어느 정도 활용가능한 자료

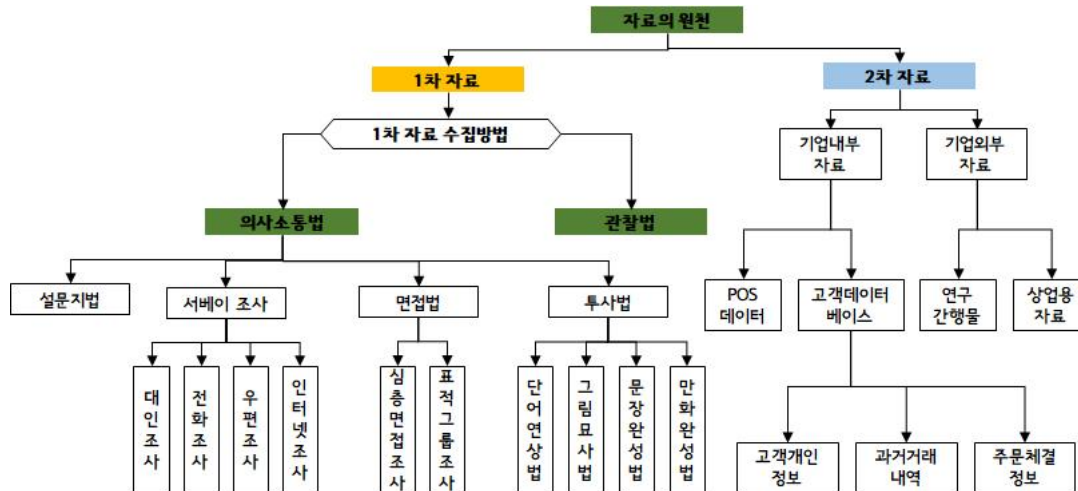
■ 1차 자료

- 의사결정과 직접적으로 연관됨
 - 분석결과를 직접 활용할 수 있는 반면에 2차 자료보다 수집과정에서 많은 시간과 노력 그리고 비용을 필요로 함
 - 모든 가용한 2차 자료를 먼저 확인한 후에 부족한 부분에 대해서만 1차 자료를 수집

■ 1차 자료와 2차 자료의 장점 및 단점

구분	2차 자료	1차 자료
장점	- 2차 자료는 1차 자료에 비하여 수집과정에서 필요로 하는 시간, 노력, 비용이 적게 듭	- 1차 자료는 의사결정과 직접적으로 연관되기 때문에 그 분석결과를 직접 활용할 수 있음
단점	- 시간이 흐름에 따라 정확성이 떨어지며 현재의 의사결정에서 필요로 하는 핵심적인 정보를 제공하지 못하는 경우가 많으므로, 그 적합성을 평가하여 의사결정에 활용해야 함	- 1차 자료는 2차 자료에 비하여 수집하는 데에 많은 시간, 노력, 비용을 필요로 하기 때문에 2차 자료를 활용할 수 있는지를 우선적으로 파악한 후 부족한 부분에 대해서만 1차 자료를 수집하는 것이 바람직함

■ 자료수집방법 : 1차 자료와 2차 자료



- 1차 자료의 수집방법

- 응답자와의 커뮤니케이션을 통하여 자료를 수집하는 의사소통법과 어떤 현상 및 사건에 대해서 제3자의 입장에서 관찰하는 관찰법이 있음

- 자료수집방법 선택

- 조사목적과 얻고자 하는 자료의 특성 및 가용시간과 예산을 고려해서 가장 적합한 자료수집방법을 선택하여 사용해야 함

2. 의사소통법 2

1) 의사소통법이란?

■ 정의

- 응답자가 가지고 있는 특정한 정보에 대하여 설문지나 직접적인 대화를 통하여 조사하는 방법
- 조사자가 작성한 설문지나 직접적인 대화를 통한 질문에 대하여 응답자가 답변함으로써 자신이 가지고 있는 정보를 제공하는 방식

■ 장점 vs 단점

장점	단점
<ul style="list-style-type: none"> - 관찰법은 응답자가 조사목적에 맞는 행동을 할 때까지 기다려야 하나, 의사소통법은 조사자가 원하는 정보를 질문을 통하여 신속하게 얻을 수 있음 예) 선호하는 향수브랜드를 조사할 때 관찰법은 구입할 때까지 기다려야 하지만 의사소통법은 질문을 통해 	<ul style="list-style-type: none"> - 응답자가 질문내용을 잘 모르고 답변하게 되면 조사를 그르치는 결과를 낳을 수 있음 예) 맥주를 마셔본 경험이 없는 미성년자에게 “당신은 H맥주의 어느 맛을 가장 선호하십니까?” 질문을 하여 정확한 답변을 하지 못하거나, “당신은 부부싸움을 한 달에 몇 번 정도 하십니까?”

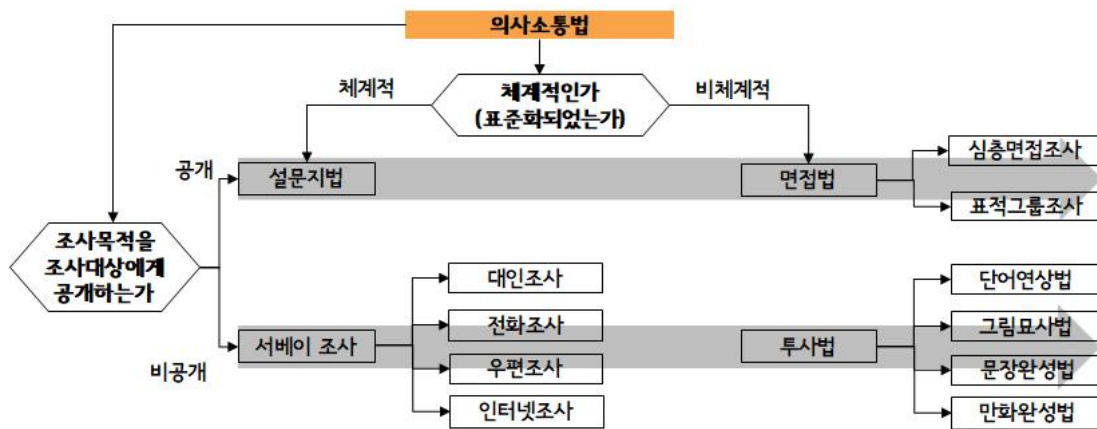
바로 확인가능	질문을 하여 응답자가 고의로 답변을 회피하도록 만들
---------	---------------------------------

■ 종류 및 구분

조사목적의 공개여부	자료수집과정의 표준성과 체계성	체계적 (표준화)	비체계적 (비표준화)
공개		설문지법	면접법
비공개		서베이법	투사법

2) 의사소통법의 분류

■ 의사통법의 분류



■ 설문지법

- 가장 대표적인 1차 자료 수집방법으로 응답자의 답변을 요하는 일련의 질문들로 구성된 설문지를 이용하여 조사하는 방법임
- 장점 vs 단점

장점	단점
- 조사과정에서 융통성을 발휘할 수 있으며 자료처리 및 결과해석과 분석이 비교적 객관적이고 용이함	- 설문지가 잘못 작성되었을 경우에는 일단 분석이 어렵고, 의사결정에 도움을 주는 정보를 제공할 수 없거나 그릇된 정보를 제공할 수 있음

■ 서베이법

- 조사에서 가장 많이 활용되는 방법
 - 대인조사, 전화조사, 우편조사, 인터넷 조사 등을 이용하여 응답자들에게 연구주제와 관련된 질문에 응답하도록 함으로써 자료를 체계적으로 수집하고 분석하는 조사설계방법임

- 서베이 매체에 따른 종류
 - 대인조사 : 조사자가 응답자를 직접 만나 조사하여 자료를 수집하는 방법
 - 전화조사 : 조사자가 전화를 이용하여 응답자로부터 직접자료를 수집하는 방법
 - 우편조사 : 응답자가 설문지를 우편으로 받아서, 응답한 후 조사자에게 발송하는 방법
 - 인터넷 조사 : 인터넷을 이용하여 온라인으로 조사하는 방법
- 서베이법을 이용한 조사의 실행매체 특성에 따른 상대적 비교

	대인조사	전화조사	우편조사	인터넷조사
수집정보의 양	고	저	중	중
수집정보의 질	중	저	저	저
제공정보의 양	고	저	중	중
제공정보의 질	고	중	저	저
정보수집 시간	저	저	중	중
정보수집 비용	고	저	중	중
응답률	고	중	저	저
표본의 대표성	중	고	중	중
의사소통	쌍방향	쌍방향	쌍방향	쌍방향

■ 면접법

- 조사의 목적을 공개하면서 자료를 수집하는 방법으로 응답자의 자유로운 의사표현을 조사자가 기록하는 식의 조사방법을 말함
- 종류

구분	심층면접법		표적집단면접법
내용	- 조사자가 주어진 질문이나 주제에 대하여 한 명의 응답자와 자유롭게 대화하면서 심도 있게 조사하는 방법		- 심층면접법의 변형으로 6~12명의 응답자가 주어진 주제에 대하여 자유롭게 토론하도록 하고, 토론과정에서 조사자가 유용한 정보를 추출하여 얻는 방법
	장점	단점	
	- 응답자의 의견을 폭넓게 반영하여 문제점을 심도있게 파악할 수 있으며, 그 결과를 설문지 설계에 효과적으로 이용할 수 있음	- 비용과 시간이 많이 들고, 조사원의 능력에 따라 조사의 결과가 달라질 수 있음	

■ 투사법

- 심리학에서 많이 사용되는 방법
 - 응답자가 조사목적은 모르는 상태에서 응답자의 내면에 있는 신념이나 태도 등을 단어연상법, 문장완성법, 그림묘사법, 만화완성법 등과 같은 다양한 동기유발방법을 이용해서 조사하는 방법

- 종류

- 문장완성법 : 미완성인 문장을 제시하여 응답자에게 나머지 문장을 완성하도록 하는 방법
- 만화완성법 : 특정 상황에 대한 만화들을 제시하여 응답자가 그 만화들을 서로 연결되도록 순서를 정함으로써 만화를 완성하도록 하는 방법
- 그림묘사법 : 응답자에게 그림을 제시하고 그 그림이 무엇을 묘사하고 있다고 생각하는지를 묻는 방법
- 단어연상법 : 한 단어를 제시하고 응답자가 그 단어로부터 연상되는 단어들을 순서대로 나열하도록 하는 방법

- 단어연상법의 예문

- 김치의 상품화로부터 연상될 수 있는 배추, 각종 양념, 숙성 정도, 가격 등과 같은 시험단어들에 대하여, 응답자들이 연상해 내는 단어들의 순서와 반응시간 등을 측정하여 조사에 활용
 - ☞ 가장 먼저 연상하고 반응시간이 짧은 단어일수록 해당 단어에 대한 응답자의 감정적 몰입이 큼

- 문장완성법의 예

- DSLR 구입시 가장 중요하게 생각하는 점은 "()입니다"라는 문장을 제시하고 공란을 메우도록 하여 조사하는 방법

3 관찰법

1) 관찰법이란?

■ 정의

- 질문과 답변을 통하여 정보를 수집하는 것이 아니라, 응답자의 행동과 태도를 조사자가 관찰하고 기록함으로써 정보를 수집하는 방법

■ 장점 vs 단점

장점	단점
<ul style="list-style-type: none"> - 의사소통법보다 비교적 정확한 정보를 얻을 수 있음 	<ul style="list-style-type: none"> - 응답자와 커뮤니케이션을 통해서만 알아낼 수 있는 인지, 신념, 선호도와 같은 심리적 특성은 관찰할 수 없을 뿐만 아니라, 개인의 사적인 활동도 관찰하기 곤란함 - 관찰하고자 하는 행동패턴이 장기간 지속되며, 자주 발생하지 않는 경우라면, 예상보다 많은 비용과 시간이 소요될 수 있음

■ 필요시 의사소통법을 이용한 자료수집방법과 서로 보완적으로 활용하는 것이 바람직함

(1) 관찰사실의 공개 여부에 따른 분류

- 공개적 관찰
 - 관찰대상이 관찰되고 있음을 인지하고 있는 상태에서 관찰하는 방법
 - 관찰대상이 관찰을 의식하여 평소와 다른 행동을 할 수 있는 문제발생
- 비공개적 관찰
 - 관찰대상에게 관찰에 대해서 사전에 알려주지 않은 상태에서 관찰하는 방법
 - 관찰대상의 자연스러운 행동을 관찰할 수 있으나 조사가 관찰대상에게 알려질 경우 윤리적인 책임문제가 발생

(2) 관찰형식의 체계성에 따른 분류

- 체계적 관찰
 - 미리 관찰표를 작성하여 정해진 형식에 따라 관찰대상의 행동을 관찰하는 방법
- 비체계적 관찰
 - 관찰대상의 행동을 사전에 예측하기 어려워 준비된 관찰표 없이 관찰하는 방법

(3) 관찰상황의 인위성 여부에 따른 분류

- 자연적 관찰
 - 조작된 상황이 아닌 자연스러운 상황에서 관찰대상의 행동을 관찰하는 방법
- 인위적 관찰
 - 실험설계와 같이 특정한 상황을 인위적으로 만들어 이에 대한 관찰 대상의 반응을 측정하는 방법
 - 예) 시음회 등을 통하여 새로 출시한 음료수에 대한 잠재고객들의 반응을 살피는 것

(4) 관찰시점에 따른 분류

- 직접 관찰
 - 관찰대상의 행동이 일어난 시점에서 직접 관찰하는 방법
 - 예) 제과점에서 소비자가 어떤 빵을 구매하는가를 직접 보고 관찰함
- 간접 관찰
 - 관찰대상의 행동이 끝난 후에 그 흔적을 관찰하는 방법
 - 예) 하루 일과 후 진열대를 관찰하여 어느 빵이 얼마나 관찰되었는가를 파악하는 식으로 관찰함

(5) 관찰도구에 따른 분류

- 사람의 직접관찰
 - 조사자가 직접 눈으로 관찰하는 방법
- 기계를 이용한 관찰
 - 사람이 직접 관찰하기 어려운 경우, 특정한 도구를 이용하여 관찰하는 방법
 - 카메라

- 모션픽처카메라
 - ☞ 고정된 CCTV에 사용되는 카메라로 슈퍼마켓이나 대형할인매장에서 소비자들의 구매행동을 관찰하는 데 사용
- 오디오미터
 - ☞ TV 시청률 조사를 목적으로 TV가 켜져 있는 시간과 시청하고 있는 채널을 자동으로 기록하는 장치
- 아이카메라
 - ☞ 관찰대상이 책이나 TV, 인터넷 광고 등을 볼 때, 눈동자의 움직임을 추적하여, 소비자가 어떠한 순서로 광고를 보며 어떤 광고에 얼마나 머무는지 측정할 수 있는 도구
- 퓨필로미터
 - ☞ 동공의 크기를 측정
- 싸이코갈바노미터
 - ☞ 자극에 대한 관찰대상의 반응정도를 측정

마케팅조사의 실행과정 : 표본선정과 오류 / 4주차 2차시

1. 전수조사와 표본조사

1) 표본의 개념

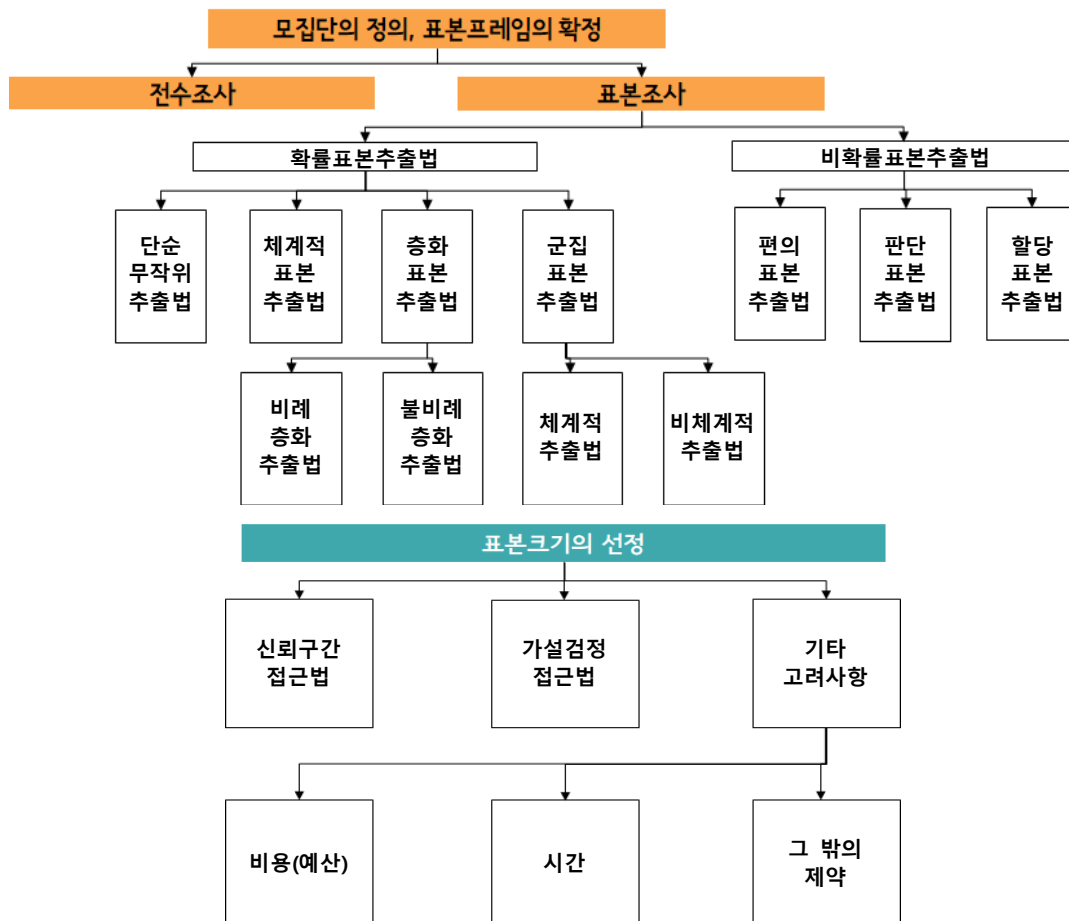
■ 표본선정

- 조사대상을 선정하는 것으로서 의사결정에 필요한 정보를 제공해 줄 수 있는 대상을 선정하는 것

■ 표본

- 모집단을 대표할 수 있도록 선정되어야 하는 것은 표본으로부터 얻은 정보는 전체 모집단의 정보를 대변할 수 있어야 함을 의미

2) 표본선정



3) 전수조사 vs 표본조사

전수조사	표본조사
- 모집단 구성원 전체에 대해 조사하는 방법	- 모집단을 대표할 수 있는 일부 대상을 표본으로 선정하여 조사하는 방법 - 대부분의 사회과학분야의 조사는 표본조사이며, 이 경우 표본을 조사하여 얻은 정보로 모집단의 정보를 추정함

4) 표본조사를 실시하는 이유

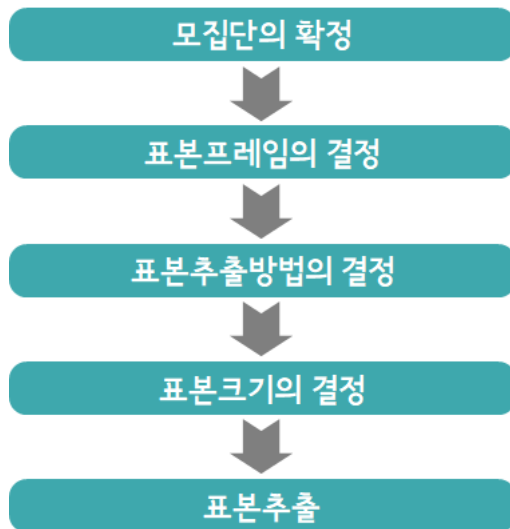
- 일반적으로 전수조사는 많은 비용과 시간을 필요로 함
- 경우에 따라서는 전수조사 자체가 아예 불가능한 경우
 - 예를 들어, 전자제품을 생산하는 회사에서 생산되는 TV의 견고함 정도를 시험하기 위해 3M 높이에서 떨어뜨려 보는 시험을 할 때 모든 TV를 떨어뜨려 볼 수 없음으로 전수조사 불가능
- 전수조사 과정에서 발생하는 비표본오류 때문에 전수조사가 오히려 표본조사보다 부정확한 결과를 산출할 가능성
 - 너무 많은 조사로 인하여 조사원이 실제 조사하지 않고 응답자의 예상되는 응답을 대신 처리하는 것이 발생
 - 잘못 작성된 설문지로부터 발생하는 오류 또는 조사 과정에서 면접원의 커뮤니케이션 오류나 특정한 조사상황에서 발생하는 오류, 수집된 자료를 분석하기 위한 코딩(Coding) 과정에서 발생하는 오류 등 조사 및 자료수집 과정에서 발생하는 오류인 비표본 오류가 증가

2. 표본추출과정

1) 표본추출의 중요성

- 표본조사를 실시하는 경우, 모집단을 정확하게 대표할 수 있는 표본을 선정·조사해서 이로부터 얻는 표본통계량(statistic) 값으로 모집단의 모수(parameter)를 추론함
- 표본통계량으로 모집단의 모수를 추론하는 과정에서 적지 않은 오류가 발생하는데 이 중 가장 큰 오류가 표본오류(sampleing error)임
 - 표본오류를 줄이기 위해서는 모집단을 가장 정확하게 대표할 수 있는 표본을 추출하는 것이 중요함

2) 과정



■ 모집단의 확정

- 모집단
 - 조사자가 조사하고자 하는 관심의 대상이 되는 사람, 기업, 상품, 지역 등과 같은 집단의 전체 집합체(set)
- 모집단을 정의하고 확정하는 것은 조사자가 필요로 하는 정보를 제공해 줄 수 있는 원천을 결정하는 것이므로 매우 중요함
- 일반적으로 모집단은 조사대상의 인구통계학적 특성과 지역 및 시간 개념들을 고려해서 구체적으로 규정해야 함
 - 예를 들어, K대학교에서 학교행정서비스 만족도를 조사하는 경우에 조사대상이 되는 모집단은 특정일(20XX년 ○월 ○일) 현재 K대학교의 모든 재학생임

■ 표본프레임의 선정

- 모집단이 정해진 후 조사자는 실제적인 표본추출의 틀이 되는 표본프레임(Sample frame)을 설정
- 표본프레임
 - 모집단에 속하는 연구대상이나 표본단위가 포함된 목록을 의미하며, 일반적으로 조사자는 표본프레임으로부터 최종적인 표본을 추출함
- 예를 들어, K대학교에서 학교행정서비스 만족도를 조사하는 경우에 표본프레임은 정일(20XX년 ○월 ○일) 현재 K대학교 학적부에 재학생으로 등록되어 있는 모든 학생들의 목록임
- 모집단이 한정적일 경우 모집단과 거의 일치하는 표본프레임을 얻을 수 있으나 현실적으로 대규모의 일반인들을 대상으로 하는 대부분의 조사에서는 모집단과 정확하게 일치하는 표본프레임을 얻는다는 것은 거의 불가능

- 표본프레임 오류(불포함 오류)
 - 모집단과 표본프레임이 일치하지 않을 경우에 발생하는 오류
- 표본추출에 이용되는 표본프레임 자체가 존재하지 않는 경우
 - 예를 들어, 0월 0일부터 0월 0일 사이에 00백화점을 방문한 고객들을 모집단으로 규정하는 경우 모집단 자체가 유동적이어서 표본프레임은 존재하지 않음
→ 상업적 조사에서는 표본프레임을 필요로 하지 않는 비확률표본추출방법이 주로 이용

■ 표본추출방법의 결정

- 표본추출방법에는 확률표본추출방법과 비확률표본추출방법으로 구분

구분	확률표본추출방법	비확률표본추출방법
정의	<ul style="list-style-type: none"> - 모집단 구성원의 명단이 기재된 표본프레임을 이용하여 표본을 추출함으로써 모집단의 개별 구성원이 표본으로 선택될 확률을 미리 알 수 있는 상태에서 표본을 추출하는 방법 	<ul style="list-style-type: none"> - 표본프레임이 없어 모집단에 속한 각각의 구성원들이 표본으로 선택될 확률을 모르는 상태에서 표본을 선정하는 방법
특성	<ul style="list-style-type: none"> - 표본추출로 인하여 발생하는 오류에 대한 추정이 가능함 	<ul style="list-style-type: none"> - 표본이 모집단을 어느 정도 대표하는가 하는 표본의 모집단 대표성 추정이 어려움 - 상업적 조사에 많이 활용됨

- 비교적 정확한 표본프레임의 입수가 가능하다면 비확률표본추출방법보다는 확률표본추출방법을 이용하는 것이 바람직함
- 모집단의 규모가 클 경우 확률표본추출방법은 비용과 시간 측면에서 비효율적일 뿐만 아니라 수행하기도 어려움
 - 상업적 조사에서는 비확률표본추출방법을 선호
- 비확률표본추출방법이 확률표본추출방법보다 반드시 열등한 것은 아님
 - 예를 들어, 신기술 등 전문적인 지식이 필요한 경우에 이를 보유한 대상들만을 조사하는 판단표본추출법을 이용하는 것이 보다 효과적일 수 있음
- 확률표본추출방법

	종류	내용	
가	단순무작위추출법	정의	<ul style="list-style-type: none"> - 표본프레임에 들어 있는 각각의 표본에 대해 일련번호를 부여하고, 이를 이용해서 일정수의 표본을 무작위(random)로 추출하는 방법으로서, 확률표본추출방법 중 가장 기본적인 방법
		과정	<ul style="list-style-type: none"> - 모집단을 대표할 수 있는 표본프레임을 확보 - 표본프레임 내의 모든 표본들에 대하여 일련번호 부여 - 엑셀과 같은 프로그램을 이용하여 일련번호로 나타나는 각각의 표본에게 유니폼 분포(Uniform distribution)하는 난수를 부여 - 난수에 의하여 정렬한 다음, 필요한 표본의 수만큼 일련번호들을 표본으로 추출
		사용하지 않는 이유	<ul style="list-style-type: none"> - 모집단의 규모가 매우 큰 집단을 대상으로 하는 경우 표본프레임이 존재하지 않거나 존재하더라도 입수하기 어려움 - 표본프레임 입수 후 각 표본에 번호를 부여하고 컴퓨터 프로그램을 이용하여 추출하는 것도 복잡함 - 전체 모집단을 제대로 반영하지 못하는 경우 발생
나	체계적추출법	정의	<ul style="list-style-type: none"> - 추출단위에 일련번호를 부여하고 이를 등간격으로 나눈 후 첫 구간에서 하나의 번호를 무작위로 선택한 다음, 등간격으로 떨어져 있는 번호들을 계속해서 추출하여 나가는 방법
		체계성의 추출	<ul style="list-style-type: none"> - 체계적 규칙을 만들어 냄 (조사자의 주관, 난수표를 활용함) 예) 100명으로 구성된 모집단에서 10명을 표본으로 추출하는 경우

다	층화표본 추출법	정의	- 모집단을 특정한 기준에 따라 서로 상이한 소집단으로 나누고 이들 각각의 소집단으로부터 빈도에 따라 적절한 수의 표본을 무작위로 추출하는 방법	
		장점	- 무작위로 표본을 추출할 경우 잘못하면 표본이 편중될 수 있는 단점을 보완 - 단, 집단의 구분하는 변수의 선정에 매우 신중하게 고려해야 함	
		종류	비례층화 추출법	- 소집단의 크기에 비례하도록 표본의 수를 할당하여 추출 - 학년별 학생수에 비례하여 각 학년별로 표본을 할당하여 추출함
			불비례층화 추출법	- 집단의 크기를 고려하지 않고 각 집단별로 비슷한 수의 표본을 추출 - 일반적으로 불비례로 표본을 추출하는 이유는 조사문제에 따라 구성원 수가 작은 집단이 어떠한 경우에는 매우 중요할 수 있기 때문 - 학년별 학생수에 관계없이 각 학년별로 동일한 숫자의 표본을 할당하여 추출함
			다단계층화 표본추출법	- 여러 단계에 걸쳐 집단을 구분하고 각 단계별로 표본을 추출하는 방법
		할당표본 추출법과 차이	- 할당표본추출법은 각 소집단내에서 표본을 추출하는 경우 편의에 의해서 추출하는 비확률적인 방법을 사용하나, 층화표본추출법은 각 소집단 내에서의 표본추출은 확률을 고려한 단순무작위표본추출법으로 이루어짐	
라	군집표본 추출법	정의	- 모집단을 군집으로 분류한 다음, 일정수의 군집을 무작위적으로 추출하여 추출된 군집 내의 모든 구성원들을 조사하는 표본추출방법 예) 대학생의 라이프스타일에 관하여 조사하는 경우 전국에서 몇 개의 대학을 선정하고 이들로부터 다시 몇 개의 학과와 학년을 선정하여, 해당되는 학생들을 모두 조사하는 방법 - 대표적인 방법으로 지역표본추출법이 있음	

		장점	- 조사에 소요되는 시간이나 경비를 줄일 수 있음
		단점	- 선출된 집단 내 표본들이 서로 동질적이면 선택된 소집단이 모집단을 충분히 대표한다고 볼 수 없음 - 군집별 구성원들의 수가 다르기 때문에 표본의 수가 사전계획에서 설정한 표본의 수와 정확하게 일치하지 못하는 문제발생

- 비확률표본추출방법

	종류	내용	
가	편의표본 추출법	정의	- 가장 간단한 형태의 표본추출방법으로서 임의로 선정한 지역과 시간대에 조사자가 임의로 원하는 사람들을 표본으로 선택하는 방법 예) TV 뉴스에서 오후 7시에 서울 명동을 지나는 행인들을 대상으로 유통시장 개방에 대한 의견을 물어보는 경우
		장점	- 표본추출비용이 거의 들지 않고 절차가 매우 간단함
		단점	- 추출된 표본이 모집단을 정확하게 대표한다고 보기 어려움 - 응답에 협조하는 대상자들만을 조사하게 됨으로써 응답을 거부하는 사람들의 특성이 전혀 반영되지 못하는 문제점 발생 → 실제 마케팅 조사에서는 거의 사용하지 않음
나	판단표본 추출법	정의	- 조사문제를 잘 알고 있거나 모집단의 의견을 효과적으로 반영할 수 있을 것으로 판단되는 특정 집단을 표본으로 선정하여 조사하는 방법 예) 해외 의료관광에 대한 전문지식을 가진 대상을 임의로 판단하여 표본을 선정
		장점	- 적은 비용으로 의미 있는 자료를 수집할 수 있음으로 표적집단 면접법과 함께 본 조사를 실시하기 전에 실시하는 사전조사에서 많이 활용
		단점	- 모집단의 특성을 정확하게 대표하지 못할 수 있음
		스노우볼 표본추출법	- 효과적으로 활용하기 위해서는 조사자는 조사문제에 대하여 전문적인 지식을 가지고 있는 표본을 정확하게 찾아낼 수 있어야 함 - 이를 충분하게 확보하지 못한 경우 먼저 접촉한 대상자들에게 자신들과 유사한 지식을 가진 다른 대상자들을 소개해 줄 것을 부탁해서 새로운 표본을 발굴하는 방법

다	할당표본 추출법	정의	- 미리 정해놓은 분류기준에 의해 전체표본을 몇 개의 집단으로 나누고 각 집단별로 필요한 만큼의 조사대상을 추출하는 방법 예) 자사 상품에 대한 만족도 조사에서 전체 구매고객의 30%가 30대 연령층이라면, 조사자는 전체표본 중 30%를 30대의 연령층에 할당
		과정	- 특성에 따라 세분된 각 집단에서의 표본선정은 조사자의 주관적 판단 혹은 편의표본추출법에 의하여 이루어짐 - 상업적 조사에서 보편적으로 사용하며 조사자는 나이, 성별, 소득에 따라 적당하게 표본을 할당하여 조사
		장점	- 비확률표본추출방법에 속하나, 조사자가 집단분류기준을 효과적으로 선정하여 이를 잘 적용한다면 모집단의 특성을 충분히 반영할 수 있는 표본을 선정 → 상업적 마케팅 조사에 널리 활용되는 이유

■ 표본크기의 결정

- 표본의 대표성과 조사에 필요한 시간과 비용 그리고 조사목적과 조사방법 등을 전반적으로 고려하여 결정

① 조사목적과 조사방법

- 단순한 소비자 반응에 따른 효과적인 마케팅 캠페인에 대한 의사결정인지 아니면 더 중요한 의사결정인 경우 표본의 수를 증가시킬 수 있음
- 사전조사(표적집단면접법 : 3~4개의 집단)인지 본조사(설문조사 : 1,000명~2,000명)인지의 여부

② 조사예산과 조사기간

- 가장 커다란 제약조건으로 표본의 크기가 커질수록 모집단의 특성을 보다 정확하게 반영하는 정보를 표본으로부터 얻을 수 있으나 조사비용과 조사기간은 상대적으로 증가

☞ 가용한 예산과 시간에 맞추어 표본의 크기 조정

③ 신뢰구간접근법

- 표본조사를 통하여 어느 정도로 신뢰할 수 있는 정보를 얻기 원하는가에 따라 표본의 크기를 결정
☞ 허용오차의 크기에 따라 필요한 표본의 수를 결정
- 표본의 수를 결정하는 과정: 표본조사 결과로 얻은 정보와 모집단의 모수와의 차이를 나타내는 허용오차 결정
- 허용오차의 구간 내에 진정한 모집단의 모수가 존재하리라 확신하는 정보를 나타내는 신뢰수준 결정

- 적합한 표본의 수를 계산하는 통계모형을 이용해 표본의 크기 산출(제8장 추정과 가설검정에서 자세히 설명)

3. 표본오류

1) 표본오류의 개념

■ 표본오류

- 모집단을 대표하지 못하는 표본을 추출함으로써 발생하는 오류

■ 표본오류의 크기는 표본추출방법이나 추출과정 그리고 추출한 표본의 크기에 따라 달라짐

2) 표본의 종류

■ 오류의 종류

- 표본오류(sampling error)
 - 충분하지 않거나 대표성이 없는 표본을 잘못 추출함으로써 발생하는 오류
- 비표본오류(Non-Sampling Error)
 - 표본오류 이외의 모든 오류

4. 비표본오류

1) 비표본오류의 개념

■ 비표본오류

- 표본오류를 제외한 나머지 모든 오류를 말하며, 주로 자료의 측정과 수집과정에서 발생하는 오류

2) 비표본 오류의 종류

■ 비표본오류 (non-sampling error) : 표본오류 이외의 모든 오류

- 관찰오류
 - 조사현장의 오류 : 조사과정에서 조사자의 잘못이나 조사상황에 따라서 발생하는 오류
 - 자료기록 및 처리의 오류 : 정확한 응답이나 행동을 한 결과를 조사자가 잘못 기록하거나, 자료를 분석하기 위해서 코딩하는 과정에서 발생하는 오류
- 비관찰오류
 - 불포함오류 : 표본프레임과 모집단의 불일치로 인하여 발생하는 오류
 - 무응답오류 : 선정된 표본 중에서 응답을 얻어내지 못하여 생기는 오류

5. 표본크기와 오류와의 관계

- 표본의 크기는 표본오류와 반비례하고 비표본오류와 비례함
- 표본의 크기에 비례해서 조사비용과 조사시간이 증가하기 때문에 조사자는 의사결정에 필요한 신뢰수준과 허용할 수 있는 오차의 크기에 따라 적당한 수준으로 표본의 크기를 정해야 함

마케팅조사의 실행과정 : 자료수집 / 5주차 1차시

1. 설문지 이용한 자료수집시 고려사항

■ 설문지

- 조사목적에 맞는 유용한 자료를 수집하기 위하여 가장 많이 사용되는 효과적인 수단

■ 설문지를 이용한 자료수집과정에서 고려사항

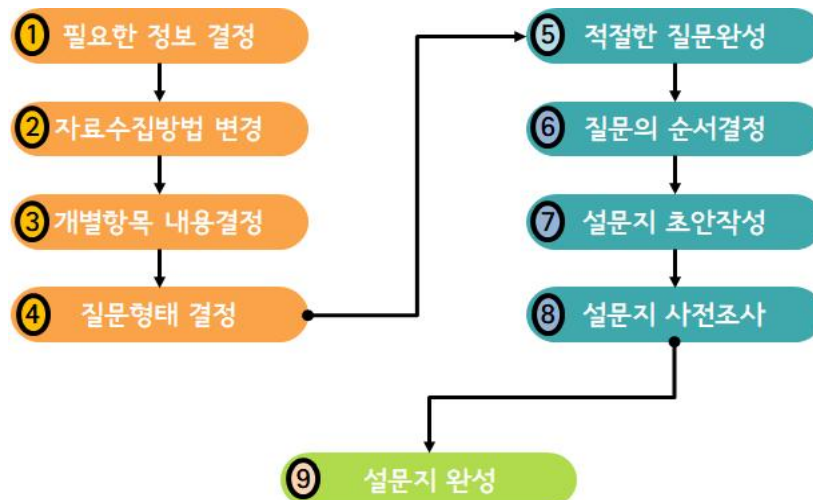
- 응답자가 질문에 답변하는 과정에서 발생할 수 있는 오류, 즉 편의(bias)를 최소화하도록 설문지의 각 문항을 간결하고 명확하게 작성
- 애매한 표현을 가능한 한 회피
- 설문문항에서 사용하는 척도와 측정방법에 따라 응답자로부터 얻을 수 있는 정보의 양과 질이 달라지므로 조사자는 조사의 목적과 내용에 적합한 척도를 사용하여 설문문항을 작성하고 설문지를 효과적으로 구성해야 보다 정확하고 다양한 정보를 얻을 수 있음

2. 설문지 작성

1) 설문지 작성절차

■ 설문지를 작성하는 일반적인 과정

- 작성절차에 따라 설문지를 작성하면, 설문지 작성과정에서 빈번하게 발생하는 오류를 사전에 인식하고, 예방함으로써 보다 효과적으로 설문지를 작성



■ 필요한 정보결정

- 설문조사를 통해 어떠한 정보를 얻고자 하는가와 어떻게 하면 정확하게 그 정보만을 얻을 수 있는가를 결정해야 함
- 불필요한 내용은 시간과 비용의 낭비를 유발하고 응답자에게 부담을 주기 때문에 다른 중요한 내용의 문항조차 제대로 조사할 수 없게 만들

■ 자료수집방법의 선정

- 설문지를 이용한 자료수집방법으로는 면접원이 응답자를 직접 만나 조사하거나 전화, 우편, 인터넷과 같은 매체를 이용하여 조사하는 방법
- 매체선택에 따라서 조사에 소요되는 시간과 비용, 설문내용이 달라지므로 조사하는 필요한 정보와 시간 및 가용한 경비 등을 신중하게 고려하여 수집방법을 선택

■ 개별항목 내용결정

- 고려하는 사항

항목	세부내용
이 질문은 꼭 필요한 것인가?	- 조사자는 문제해결에 꼭 필요한 질문들만 포함될 수 있도록 충분히 검토해서 설문지를 작성해야 함
응답자는 조사자가 필요로 하는 정보를 알고 있는가?	- 설문항목에 대하여 응답자가 답변하기에 충분한 정보를 가지고 있는가를 사전에 검토해 보아야 함 예) 호텔 레스토랑에 대한 소비자의 인식 조사를 위하여 호텔 레스토랑들을 비교한 선호도를 조사하려는 경우, 가장 중요한 사항은 비교하고자 하는 호텔 레스토랑을 모두 방문한 경험이 있는 응답자만을 선정하여 조사해야 함
응답자가 그 정보를 제공해 줄 수 있는가?	- 조사자는 응답자가 응답에 필요한 충분한 정보를 가지고 있다 하더라도 과연 그 정보를 정확하게 제공해 줄 수 있는지를 고려하여 설문지를 작성해야 함 - 일반적으로 응답자들은 개인의 사적인 정보나 사회적으로 민감한 주제에 관한 질문에 대해서는 매우 예민하게 생각하여 이를 경계함 예) 자신의 소득과 연봉수준에 대한 질문이나 자신의 신체사이즈나 학력 등에 관한 질문에 대해서는 응답자는 매우 민감하게 반응하여 불쾌감을 표출할 수 있음

■ 질문형태 결정

- 설문지에 사용되는 질문형태는 크게 개방형 질문(open-ended questions)과 선택형 질문(closed-end question)으로 구분

1	개방형 질문
정의	- 응답자가 생각하고 있는 답변을 자유롭게 표현하도록 하는 방법이고, 선택형 질문은 이미 조사자에 의해 마련되어 있는 항목들을 응답자로 하여금 선택하도록 하는 방법
종류	- 제한이 있는 개방형 질문 예) “호텔 레스토랑을 이용하실 때 가장 중요하게 생각하시는 요인 한 가지만 말씀해 주십시오” - 제한이 없는 개방형 질문 예) “호텔 레스토랑을 이용하실 때 가장 중요하게 생각하시는 요인들을 생각하시는 대로 모두 써 주십시오”
장점	- 설문으로 얻고자 하는 정보가 복잡하고 다양해서 조사자가 사전에 구체적인 설문항목을 만들기 어려운 경우에 유용하게 사용 - 일반 응답자의 주관적인 의견을 통해 얻은 정보를 추가적인 조사 등에 효과적으로 활용
단점	- 응답자가 답하기 어렵고, 특히 이를 이용한 자료분석이 매우 어려움

2	선택형 질문
정의	- 이미 조사자에 의해 마련되어 있는 항목들을 응답자로 하여금 선택하도록 하는 방법 - 선택항목의 수에 따라 다지선다형 질문과 양자택일형 질문으로 구분
종류	다지선다형 질문
	양자택일형 질문
장점	- 질문에서 주어진 여러 응답 항목들 중 하나 또는 몇 개를 선택하도록 되어 있는 질문형태 예) 귀하께서 주로 이용하시는 포털사이트는 어느 것입니까? ① 다음(www.daum.net) ② 네이버(www.naver.com) ③ 네이트(www.nate.com)
단점	- 응답항목이 명확하여 신속한 응답이 가능 - 응답자들에게 응답해야 할 내용을 정확하게 제공함으로써 답변을 얻기 매우 용이
	- 질문내용에 대한 응답항목이 ‘예, 아니오’ 혹은 ‘해당됨, 해당되지 않음’ 등과 같이 두 가지 항목 중 한 가지 항목만을 선택하도록 구성되어 있는 질문형태 예) 귀하는 해외의료관광을 다녀온 경험이 있습니까? ① 예 ② 아니오
	- 다지선다형보다 더욱 응답이 쉽고 자료의 처리가 매우 간단
	- 2개의 극단적인 질문을 제시해 주고 이 중에 하나만을 선택하도록 함으로써 중간 정도 값을 택하길 원하는 응답자의 의견을 무시 → 중요한 정보가 누락될 가능성

■ 적절한 질문완성

- 질문을 완성할 때 고려사항

I. 가능한 한 전문용어의 사용을 삼가하도록 해야 함

- 설문지 작성에 있어서 조사자가 범할 수 있는 가장 큰 오류는 응답자가 설문과 관련된 전문지식에 대해서 조사자와 비슷한 수준에 있다는 착각(조사자들이 평상시 자주 사용하는 용어를 일반 사람들도 쉽게 알고 있으리라는 것)에서부터 발생
☞ 조사자들이 가장 흔히 사용하는 '변수'와 '속성' 등

II. 개방형 질문과는 달리 다지선다형과 같은 선택형 질문의 경우에는 응답자들이 생각할 수 있는 모든 응답내용을 선택항목에 포함할 수 있도록 해야 함

- 조사자가 제시하지 못한 응답내용도 자유롭게 응답할 수 있도록 맨 마지막 항목에 기타란을 별도로 제시해 두는 것도 좋은 방법

III. 다지선다형 질문에서 응답항목들 간에 내용이 중복되어서는 안 되고, 응답항목들은 서로 명확하게 구분되어야 함

- 응답하려는 한 가지 생각에 대하여 두 가지 이상의 답이 가능하도록 응답내용의 내용이 서로 중복되어서는 안 되고, 어느 한 가지 항목에는 반드시 해당될 수 있도록 질문을 구성
- 다음과 같은 질문에서 질문에 쓰인 '전혀', '거의', '가끔'이라는 형용사는 그 뜻이 명확하게 구분되지 않아서 개인마다 서로 받아들이는 기준이 달라서 정확한 응답을 얻기 어려움

※ 응답항목들 간에 내용이 중복된 잘못된 질문

- 귀하께서는 이메일을 얼마나 자주 이용하십니까?

- ① 전혀 이용하지 않는다 ② 거의 이용하지 않는다 ③ 가끔 이용한다
④ 자주 이용한다 ⑤ 매일 이용한다

IV. 하나의 질문문항으로 두 가지 내용을 동시에 질문해서는 안 됨

- 예를 들어, 호텔 레스토랑에 대한 음식과 서비스에 대한 소비자들의 의견을 조사하려는 경우 "우리 호텔 레스토랑의 음식 맛과 서비스에 대해 얼마나 만족하십니까?"와 같이 질문을 해서는 안 됨

☞ 질문에 응답해야 하는 응답자는 음식에 대해서 평가해야 하는 것인지, 서비스에 대해 평가해야 하는 것인지 또는 음식과 서비스를 총체적으로 평가해야 하는 것인지에 대해 확신하지 못하고 갈등을 불러일으킬 수 있음

V. 대답하기 곤란한 질문을 직접 물어보거나, 개인의 사적인 정보나 사회적으로 민감한 주제에 대해 너무 자세하게 물어서는 안 됨

- 응답자의 개인신상에 관한 정보(연령, 몸무게, 학력수준, 소득수준 등)나 사회적으로 민감한 주제에 대하여 정확한 정보를 수집하기는 쉽지 않기 때문에 이러한 사항들에 대한 정보를 수집해야 하는 조사자는 각별히 신경써서 설문을 구성해야 함

VI. 조사자가 원하는 답을 유도하는 질문을 해서는 안 됨

- 설문지에 설문내용을 설명하기 위하여 서술하는 경우 응답자의 답을 유도하기 위하여 자신의 의견을 서술해서는 안 됨.

VII. 어떠한 상황이든 조사자가 임의로 가정하고 설문지 내용을 구성해서는 안 됨

- "만일 귀하께서 농구경기를 관람하실 수 있는 티켓이 두 장 생긴다면 어떻게 하시겠습니까?"하는 질문을 할 경우, 평상시에 농구에 대해 별관심이 없는 사람이라면 이러한 질문에 대해서는 별로 응답하고 싶지 않을 수 있음

■ 질문순서 결정

- 질문의 순서를 정할 때 고려사항

I. 설문지의 처음 질문은 설문지 전체 내용을 짐작할 수 있는 질문이 바람직함

- 대부분의 응답자는 처음에는 냉소적이며 경계적인 태도를 보이기 때문에 처음 질문은 기본적인 상식적인 수준에서 대답할 수 있는 내용이어야 경계감을 완화시킬 수 있고 설문 전체 내용을 대략적으로 짐작할 수 있어야 함

☞ "귀하께서는 호텔 레스토랑을 이용하신 경험이 있으십니까?"하는 질문으로 조사가 시작되면 응답자는 호텔 레스토랑에 관한 설문임을 짐작할 수 있음

II. 응답자가 쉽게 대답할 수 있는 질문은 전반부에 배치하고 응답하기 어려운 질문들은 후반부에 배치하는 것이 바람직함

- 예를 들어, 호텔 레스토랑 세분화에 관하여 조사를 하고자 하는 경우, 호텔 레스토랑에 대한 경험 유무와 같이 손쉽게 대답할 수 있는 질문은 앞부분에 포함시키고, 각 호텔 레스토랑을 속성별로 평가하거나 호텔 레스토랑들 간의 차이점과 특정 서비스에 대한 만족도를 물어보는 것과 같이 응답자의 신중한 사고와 판단을 요구하는 질문들은 가급적 뒤쪽에 배치해야 함

III. 응답자의 신상에 관한 질문, 즉 인구통계학적인 질문들은 가능한 한 설문지의 맨 뒤에 배치하는 것이 바람직

- 대부분의 설문지는 인구통계학적인 질문(성별, 연령, 결혼여부, 직업, 학력, 교육수준, 소득수준 등)인 개인신상에 관한 질문들을 포함하고 있음
- 이러한 인구통계학적인 자료들을 통해 응답자들을 특성별로 쉽게 분류하여 조사된 표본이 모집단을 얼마나 효과적으로 대표하고 있는가를 검토할 수 있게 됨
- 이러한 질문을 맨 앞에서 하게 되면 응답자의 경계심을 높이고 경우에 따라서는 자존심을 건드리는 결과를 초래하기 때문에 가능한 맨 뒤에 해야 함

■ 설문지 초안완성

- 설문지 초안을 작성할 때 고려사항

I. 설문지는 양면인쇄를 한 소책자로 만들어서 조사진행이 용이하도록 해야 함

II. 설문지 표지에는 조사목적과 조사기관을 명시하고 응답자에 대한 비밀보장 내용과 응답자가 조사내용에 관하여 문의를 하거나 조사원에 대한 신원확인을 요구할 경우 이를 확인할 수 있도록 연락처를 기재해야 함

III. 설문조사를 진행하는 과정에서 단계별로 면접자나 응답자가 주의해야 할 사항들은 확실하게 박스로 표기하여 쉽게 알아볼 수 있도록 해야 함

■ 설문지 사전조사

- 설문지 초안까지 작성되면 일단 가상의 응답자 몇 명을 선정하여 설문지에 대한 사전조사를 실시하여 설문지에 잘못된 문제점을 미리 찾아내어 수정함으로써 본조사를 보다 원활하고 효과적으로 수행할 수 있도록 해야 함

■ 설문지 완성

- 본조사에 앞서 반복적인 사전조사를 통해 설문응답 과정에서 발생할 수 있는 문제점들을 미리 찾아내어 수정함으로써 설문지를 최종완성 함

마케팅조사의 실행과정 : 자료분석 / 5주차 2차시

1. 자료수집

- 자료수집방법이 결정된 후에는 추출된 표본을 대상으로 직접 자료를 수집함.
- 특히, 설문지를 이용해서 자료를 수집하는 경우에는 조사원의 선발과 교육을 비롯하여 자료수집 현장에서의 통제 및 자료의 검증 등에 대해 세심하게 주의할 필요

1) 조사원의 선발과 교육

- 자료수집단계에서 가장 먼저 해야 할 일은 유능한 조사원을 선발하고, 이들에게 자료수집에 필요한 교육을 실시하는 것
- 조사원들의 개인적인 능력과 이들에 대한 교육과 통제는 조사의 질에 중요한 영향을 줌
- 조사원을 선발하고 교육하는데 고려해야 할 사항
 - 어떠한 기준으로 조사원을 선발할 것인가?
 - 설문지 내용 및 조사방법에 대한 교육은 어떻게 할 것인가?
 - 조사원들에 대한 조사할당은 어떻게 할 것인가?
 - 조사원들이 응답자로부터 설문에 관한 질문을 받았을 경우 어떻게 답변해야 하는가?
 - 조사원들이 조사를 성실하게 시행하였다는 것을 어떻게 확인할 것인가?
 - 수집된 설문지에서 문제가 발생하였을 경우 어떻게 처리할 것인가?

2) 자료의 수집과 통제

- 조사원들이 자료를 수집하는 동안 조사자는 실행을 책임지고 있는 관리자들과 함께 현장에서 발생하는 문제점들에 대하여 지속적으로 관리하고 통제해야 함
- 조사원들이 지시된 사항에 따라 행동하고 있는지를 현장조사하며 예기치 못한 상황이 발생했을 경우에는 즉각적으로 대응할 수 있어야 함
 - 예를 들어, 축구나 야구경기를 관람하는 관객들을 대상으로 설문조사를 실시하는 경우, 갑자기 비가 오거나 날씨가 추워진 경우에는 정상적으로 조사를 실시할 수 없게 되므로 조사자는 실외에서 설문지를 받는 것을 지양하도록 하고, 부득이하게 실외에서 조사를 강행하는 경우에는 비나 추위를 피할 수 있는 추가적인 장비를 마련한 다음 조사를 진행

3) 자료의 검증

- 검증과정은 조사원들이 조사를 착실하게 수행하고 응답자는 질문에 대하여 성실하게 답변하였는가를 검토하는 과정
- 검토하는 기준
 - 설문조사는 실제로 잘 수행되었는가?
 - 조사자는 조사원별로 전체 설문지의 약 20% 정도를 무작위로 선택하여 응답자에게 전화를 걸어 조사가 실제로 실시되었는가를 확인
 - 설문지의 주요 항목들에 대하여 기입이 정확하게 잘 되었는가?

2. 자료분석

- 자료분석은 수집된 자료를 편집하고 코딩한 후 이를 분석하는 것
- 일반적으로 조사방법은 조사계획단계에서 이미 결정되며 필요에 따라서는 자료수집 후에도 가능한 한도 내에서 새로운 방법을 선정하여 분석하는 것도 가능

1) 자료분석절차

- 편집(editing)
 - 수집된 자료를 판독 가능성, 일관성, 완전성, 등을 고려해서 사용가 능한 설문지를 검토하고 선별하여 정리하는 작업
- 코딩(coding)
 - 자료를 통계패키지나 컴퓨터소프트웨어를 이용해서 분석할 수 있도록 응답자의 반응을 구체적인 수치값로 나타내기 위한 규칙을 정하고 이에 따라 자료를 입력하는 방법
- 분석(analysis)
 - 일반적으로 분석하고자 하는 독립변수 와 종속변수의 수 및 척도의 종류, 그리고 유용한 자료의 크기 등을 고려하여 적합한 분석방법을 선택함

3. 결과해석 및 전략도출

- 조사과정은 조사결과를 보고서로 작성함으로써 종결됨
- 보고서는 조사를 통하여 얻은 정보와 이를 바탕으로 고안된 마케팅 전략을 포함
- 분석결과로부터 얻은 정보를 활용하여 마케팅 전략의 시사점을 도출하고, 이를 효과적으로 제시하는 것을 포함

1) 보고서 작성시 유의사항

- 좋은 보고서의 요건
 - 정확성, 학문성, 현실성, 의사결정의 명료성, 해석 용이성
- 보고서 작성은 분석결과에 대한 통계적 이해뿐만 아니라, 마케팅에 관한 충분한 이론과 실무경험 및 감각을 필요로 함
- 보고서는 의사결정자가 이해하기 쉽고 그 결과를 용이하게 활용할 수 있도록 작성되어야 함(고객인 의사결정자가 손쉽게 이해하고 활용할 수 있도록 작성되어야 함)
- 보고서는 의사결정의 축에 있는 다양한 계층의 관리자들이 활용하기 편하게 몇 개의 보고서로 나누어 작성하는 것이 바람직함

4. 인터넷 조사

1) 오프라인 마케팅 조사절차와 온라인 마케팅 조사

■ 오프라인 마케팅조사

- 마케팅조사 문제제기 → 마케팅조사 방향설정 → 마케팅조사 계획수립
→ 자료수집방법선정 or 표본선정 → 자료수집 → 자료분석 → 결과해석 및 전략도출

■ 온라인 마케팅조사

- 마케팅조사 문제제기 → 마케팅조사 방향설정 → 마케팅조사 계획수립
→ 인터넷을 통한 표본선정, 자료수집, 자료분석 → 결과해석 및 전략도출

■ 온라인 마케팅 조사의 장점

- 표본 및 자료수집방법의 선정, 자료의 수집과 분석은 오프라인에서와 달리 인터넷을 통해 실시간으로 수행 가능하므로 기존의 오프라인 마케팅조사에 비해 시간단축과 비용절감의 효과가 큼

■ 온라인 마케팅 조사의 단점

- 인터넷이라는 매체가 가지는 한계점에 따라 그 적용범위도 적지 않은 제약이 있음
 - 직접 대면하여 하는 조사가 아닌 만큼 응답자의 적극적인 참여를 장담하기 어려우며, 많은 시간을 요하는 복잡한 설문은 수행하기 어렵다는 한계가 있음
 - 표본을 선정하는 과정에서도 인터넷을 사용하는 사람이어야 한다는 제약과 함께 응답자의 인적사항을 확인하는 것이 용이하지 않음

■ 성공적인 조사활동을 수행하기 위해서는 이러한 인터넷 마케팅 조사의 한계점과 더불어 비용과 속도에 따른 장점을 명확하게 이해하고 이를 효과적으로 이용할 수 있는 방법을 강구해야 함

2) 인터넷 서베이 조사

■ 인터넷을 통하여 조사하면 기존의 오프라인 조사보다 빠르고 저렴한 가격으로 수행할 수 있으며 응답자의 위치에 따른 제약없이 지역적으로 넓게 퍼져 있는 응답자를 동일한 비용으로 확보할 수 있는 장점

■ 온라인 서베이 조사는 웹페이지를 매개로 하는 상호작용을 통해 이루어짐

- ① 응답자 스스로 설문조사 웹페이지에 접속하도록 하는 방법
- ② 전자메일을 이용하여 설문응답을 받는 방법
- ③ 전자메일을 클릭하면 설문 웹페이지로 이동하여 응답하도록 하는 방법 등을 활용하여 실시

■ 인터넷 서베이 조사는 시스템과의 실시간 상호작용을 통하여 이루어지기 때문에 응답결과는 바로 데이터베이스에 축적되어 이에 대한 즉각적인 분석이 가능

■ 웹의 특징인 멀티미디어와 링크기능을 적절히 활용하면, 종이나 전화를 이용하는 설문에서 불가능했던 다양한 질문도 가능

- 예를 들어, 어려운 개념이나 기능에 대한 질문은 그림이나 동영상을 이용하여 설명할 수 있으며 응답경로가 복잡하고 다양한 계층구조를 갖는 설문도 웹의 링크기능을 활용함으로써 응답자를 효과적으로 유도하여 편리하게 답변할 수 있도록 설계가능

■ 온라인 서베이 조사의 장점

- 설문 개발시간이 빠름
- 실시간 리포팅이 가능함
- 비용이 저렴함
- 개인화된 질문과 자료 공급이 용이함
- 설문답변이 용이함
- 잘못된 응답은 즉시 체크가 됨

3) 온라인 표적집단 조사

- 기존의 표적집단조사에서 가장 큰 문제는 참석자들의 지리적인 제약으로 표적집단 조사 참여 대상자가 여러 지역에 넓게 흩어져 있는 경우 그 대상자들을 한 장소에 모으기 위해서는 많은 시간과 비용을 필요로 함

- 온라인 표적집단조사(on-line focus group interview)를 활용하면, 이러한 지리적인 제약을 어느 정도 해소가능

■ 온라인 표적집단조사의 장점

- 가장 큰 장점은 조사시간과 비용을 절약하는 것이 가능
 - 비용대비 효과가 크며 인터넷이 가능한 지역이면 세계 어느 곳에서도 쉽게 접근할 수 있기 때문에 표적집단조사 의 참석자들을 보다 다양하게 선정
- 보다 정직한 인터뷰를 실현
 - 서로 대면한 상태에서 실시하는 오프라인상 인터뷰에서는 상대방의 행동과 의견에 영향을 받아 자신의 생각을 바꾸거나 간혹 정직한 발언을 하지 않는 경우도 있으나 온라인상에서는 대면하지 않은 상태에서 인터뷰가 진행되므로 자신의 신념과 생각을 정확하게 표현할 수 있어 보다 효과적인 표적집단조사를 수행

자료와 측정 / 6주차 1차시

1. 자료수집과정

1) 과정

■ 분석

- 자료로부터 시작하여 효과적인 의사결정에 필요한 정보를 추출하기 위해서는 자료가 필요함

■ 자료

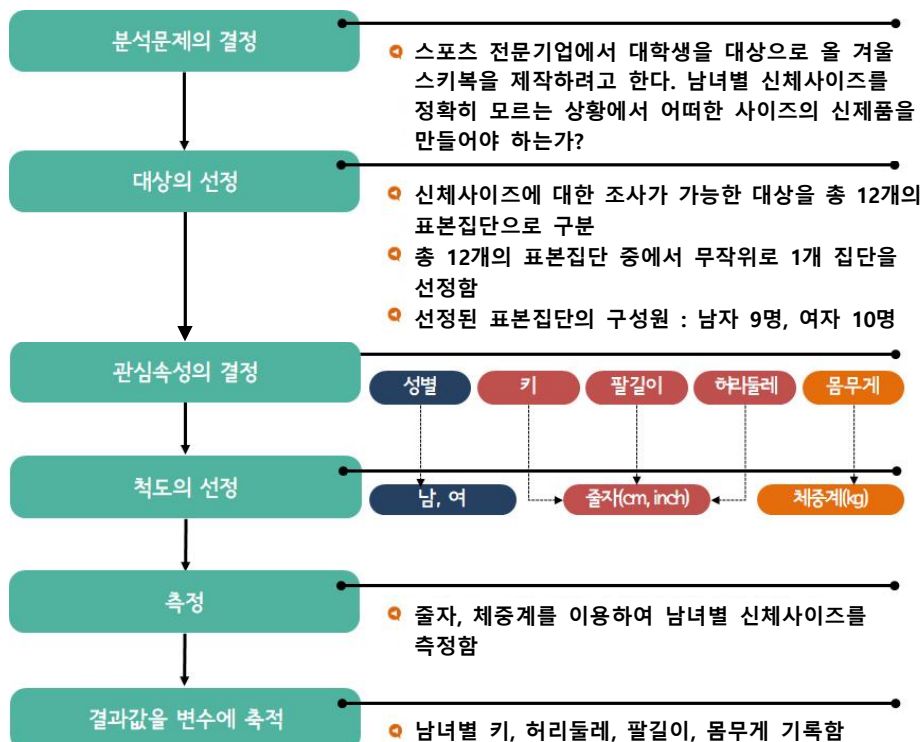
- 관심대상을 선정한 후 그 대상의 특정한 속성을 척도를 이용해서 측정한 다음 이를 변수값으로 축적함으로써 만들어짐

■ 자료수집

- 의사결정문제에 대한 명확한 인식으로부터 시작됨
- 의사결정문제가 무엇이고 효과적인 의사결정을 위해서는 어떠한 정보가 필요한지를 판단하고 확인한 후 이를 얻기 위하여 필요한 자료가 무엇인지를 결정함

■ 자료수집과정

- 자료수집대상을 선정하고 수집대상의 어떠한 속성을 측정할 것인지, 즉 관심속성을 결정하는 것으로부터 시작되며 속성에 맞는 적절한 척도를 선택하여 측정하고 그 값을 변수에 축적하는 식으로 진행



- 측정
 - 관심대상의 속성을 관찰하여 변수값으로 저장하는 과정
- 자료
 - 그 결과로 생성된 변수값들을 총칭하여 자료라고 함
- 척도
 - 측정된 값들의 특성에 따라 측정도구나 측정규칙을 구분하는 것
 - 예를 들어, 줄자와 체중계 같은 일정한 규칙이 있는 도구가 척도이며 이러한 도구를 가지고 조사대상의 관심속성을 관찰하여 그 기록을 남기는 행위가 측정이며 이를 변수값으로 저장한 것이 자료임

2. 자료의 개념과 유형

1) 자료의 개념

- 사람, 물건, 조건, 상황 등을 묘사하는 기본적인 사실의 집합으로 일반적으로 정보로 가공되기 전의 상태

2) 자료의 유형

구분	질적자료	양적자료
의미	숫자가 아닌 형태의 자료	숫자로 표시된 형태의 자료
예	성별, 안경착용 유무, 직업, 혈액형 등	000cm, 00kg 등

3) 양적자료의 유형

구분	변수	상수
의미	2개 이상의 서로 다른 값을 갖는 양적자료	오직 한 가지의 수치값만을 갖는 양적자료
예	남녀공학인 B대학교 학생의 성별처럼 남자와 여자를 나타내는 서로 다른 숫자를 변수값으로 가지는 경우	A여자대학교 학생의 성별을 숫자로 나타낼 때 모든 학생이 동일한 하나의 숫자만을 가지는 경우

4) 변수의 유형

구분	이산변수	연속변수
의미	명확히 구분되는 정수값만을 갖는 변수	연속적으로 이어지는 값
예	성별을 남자는'1' 여자는'2'라는 정수값을 가지고 구분한 경우, C자동차 회사에서 판매한 자동차 대수	키, 몸무게, 체온 등과 같이 매우 정밀한 자료 측정하면 정수와 같이 명확하게 구분되는 수치가 아니라 소수점 이하로 길게 이어지는 연속적인 값

- 이산변수와 연속변수를 명확하게 구분하기는 어려움
 - 이산변수
 - 일반적으로 컴퓨터나 자동차 판매대수처럼 정수값을 갖는 것
 - 연속변수
 - 몸무게나 온도 등과 같이 소수점 이하 연속적으로 이어지는 값을 갖는 것

5) 사례

■ 질적자료 → 척도선택 → 자료의 측정 → 양적자료

- 질적자료

- 성별
- 안경착용 유무
- 직업, 체온(높다, 낮다)
- 키(크다, 작다)

- 척도선택

- 명목척도
- 서열척도
- 등간척도
- 비율척도

- 자료의 측정

- 관찰을 통해 성별 파악
- 안경착용 여부의 관찰
- 온도계를 통한 체온측정
- 줄자를 이용한 키의 측정

- 양적자료

- 변수 → 이산변수와 연속변수

☞ 이산변수

- ★ 성별(0, 1)
- ★ 안경착용 유무(1, 0)
- ★ 작년 자동차 판매대수
- ★ A대학교 학생수

☞ 연속변수

- ★ 체온(36.5°C)
- ★ 키(175cm)

- 상수

☞ 여자대학교 학생들의 성별(모두 같음, 0)

- ✓ 사물, 소비자, 기업, 경제 등과 같은 관심대상의 속성을 척도를 이용하여 측정한 후 변수값으로 축적함
- ✓ 질적자료는 척도를 이용하여 측정됨으로써 양적자료로 변환되어 변수값으로 축적됨

3. 측정의 개념과 방법

1) 측정의 개념

- 관찰대상이 가지고 있는 속성의 값을 일정한 규칙으로 만들어진 도구(척도)를 이용하여 계량화하는 것, 즉 관찰대상의 속성에 대하여 기술적으로 수치를 부여하는 행위로 개체가 가지고 있는 속성을 관찰하여 변수값으로 표현하는 것을 의미함
 - 관찰대상의 속성값을 계량화하기 위해 적합한 척도를 이용하여 면접이나 설문 등의 커뮤니케이션 방법으로 그 값을 측정하여 자료로 기록하는 과정임
- 예) 키나 몸무게를 재서 그 크기를 숫자를 기록하거나 시험을 통하여 학생의 지식을 계량화하여 점수로 기록하는 것
- 측정은 관찰대상이 가지고 있는 속성의 값을 수치로 표현하는 것으로써 대부분의 측정결과는 숫자로 나타냄

2) 측정의 방법

- 측정도구가 필요하며 키를 측정하기 위한 자, 몸무게를 재기 위한 저울, 소비자의 성향을 조사하기 위한 설문지 등을 측정도구라 할 수 있으며 각각의 측정도구들은 나름대로의 규칙이나 측정단위를 정하여 사용함
- 예) 키를 측정하기 위한 자는 cm나 inch를 측정단위로 사용하고 저울은 kg이나 lb를 측정단위로 사용하고 있으며, 설문지 내의 모든 설문항목들도 나름대로 측정규칙을 사용하는데 이러한 규칙을 척도라고 함

척도와 변수 / 6주차 2차시

1. 척도의 의미와 유형

1) 척도의 의미

- 자료
 - 수치화 여부에 따라 질적자료와 양적자료로 구분되며 질적자료를 양적자료로 전환시키는 데에 사용되는 규칙이 척도임
 - 예) '몸무게가 무겁다', '키가 크다', '참 예쁘다'는 식의 주관적인 표현을 객관적인 도구를 이용하여 측정하거나 평가하여 수치화하면 양적자료가 됨
- 척도에는 측정된 결과로 나타나는 수치의 특성에 따라 명목척도, 서열척도, 등간척도, 비율척도 등 4가지 종류가 있음
- 척도를 이용하여 생성된 양적자료에는 변수(variable)와 상수(contrast)가 있으며 상수는 변화하지 않는 하나의 수치값만을 갖고, 변수는 서로 다른 2개 이상의 수치값을 갖음.
- 변수는 질적변수와 양적변수로 나누어짐
 - 질적변수
 - 명목척도와 서열척도 등 질적척도로 측정되어 산술적인 사칙연산, 즉(+, -, x, ÷) 하는 것이 의미가 없는 변수
 - 양적변수
 - 등간척도와 비율척도로 측정되어 사칙연산의 일부 혹은 모두가 의미 있는 변수임

2) 척도의 유형

- 명목척도
 - 관찰대상이 갖는 속성에 따라 관찰대상을 상호 배타적인 범주로 구분하는 것으로서, 관찰대상을 단순히 범주로 구분하기 위한 목적으로 숫자를 사용하는 척도를 의미
 - 예) 설문지를 이용하여 분석하는 경우에 남자는 1로 여자는 0으로 표기하기도 하는데, 이러한 숫자는 성별을 구분하기 위하여 사용하는 단순한 기호이며 숫자로서의 양적 의미는 없고, 관찰대상을 서로 구별하기 위하여 사용된 단순한 기호임.
- 서열척도
 - 관찰대상이 가지고 있는 속성의 크기를 측정하여 크기 순서대로 대상의 순위를 나타내는 것으로서 명목척도와 마찬가지로 대상을 서로 구분할 수 있을 뿐만 아니라, 속성의 크기에 따라 대상의 순서를 정할 수 있는 척도를 의미
 - 관찰대상의 속성을 비교하여 단순히 상대적인 순위만 구분할 뿐, 각 서열 간의 차이 정도를 측정하지 않기 때문에 서열 간의 구체적인 차이 정도에 관한 내용까지는 포함하지 않고 단순히 순서에 관한 정보만을 나타냄
 - 예) 올림픽에서 1위, 2위, 3위는 순위가 중요한 것이지 1위와 2위 또는 2위와 3위 간의 점수 차이는 중요하지 않음

- 서열 간의 차이, 즉 간격에 대해서는 측정하지 않고 오직 대상들의 순서에 관한 정보만을 측정하는 데에 사용되는 척도임
- 순위정보뿐만 아니라 서열척도보다 단순한 척도인 명목척도가 가지고 있는 대상을 서로 구분하는 정보도 포함하고 있음
- 예시

귀하께서 직장을 선택하실 때 중요하다고 생각하는
직장의 특성을 보기에서 골라 순서대로 말씀해 주십시오.

첫째() 둘째() 셋째() 넷째()

보기

(1) 안전성 (2) 발전성 (3) 수익성 (4) 장래성

■ 등간척도

- 속성크기에 따른 관찰대상의 서열뿐 아니라 대상들 간에 어느 정도 차이가 있는가에 관한 정보도 포함하는 것으로 관찰대상이 가지고 있는 속성의 상대적 크기를 측정하여 대상 간에 서로 비교할 수 있도록 하는 척도
- 대상이 지닌 속성의 상대적 크기와 차이를 측정하기 위하여 균일한 간격, 즉 등간격을 두고 분할한 길이를 이용하여 측정하는 척도
 - 등간척도를 이용하여 대상이 가지고 있는 속성의 크기를 측정하고자 할 경우에는 등간격을 표시하는 길이를 나타내는 그림을 제시하여 측정
- 예시

귀하께서 스마트폰을 사용하시면서 느끼시는 혜택에 관한 질문입니다.
다음 1점부터 5점 중에서 각 문항에 동의하시는 정도에 따라 점수를 주시면 됩니다.

	매우 그렇지 않다	보통이다	매우 그렇다
① 업무(일/학업)진행속도를 향상시킨다	①-----②-----③-----④-----⑤		
② 업무(일/학업)성과를 향상시킨다	①-----②-----③-----④-----⑤		
③ 업무(일/학업)의 생산성을 향상시킨다	①-----②-----③-----④-----⑤		
④ 업무(일/학업)의 효율성을 향상시킨다	①-----②-----③-----④-----⑤		
⑤ 업무(일/학업)에 유용하다	①-----②-----③-----④-----⑤		

- 일정한 길이의 선을 등간격으로 나누고 상대적 크기를 글이나 숫자로 표시한 척도이므로 등간척도를 이용하면 관심대상이 가지고 있는 속성의 크기에 따른 서열뿐만 아니라 크기의 상대적인 차이도 측정가능함
 - 상대적 차이를 나타내는 숫자를 더하거나 빼는(+, -) 산술적 계산도 가능하므로 측정된 값들의 평균을 구하여 사용할 수 있음

- 장점

- 명목척도와 서열척도가 제공할 수 있는 정보를 모두 포함하여 관찰대상이 가지고 있는 속성들의 상대적 크기를 나타내는 정보를 측정할 수 있음

- 단점

- 속성의 절대적 크기는 측정할 수 없음으로 사칙연산 중에서 가감(+, -)의 계산은 가능해도 비율처럼 곱하거나 나누는 승제(\times , \div)의 계산은 불가능함
예) 대표적인 등간척도에 속하는 온도의 경우 절대적인 기준이 없음. 섭씨에서 물이 얼음으로 변화는 온도를 0도로 하여 기준점이 있는 듯이 생각되나 물 이외의 다른 모든 물질들이 있음을 고려할 때, 하나의 특정한 물질에 불과한 물의 빙고점은 모든 물질에 적용할 수 있는 절대적인 기준점이 되지 못함.

■ 비율척도

- 절대적인 기준(일반적으로 '0')이 존재하는 척도로 절대적인 기준을 가지고 속성의 상대적 크기 비교는 물론 절대적 크기까지 측정할 수 있도록 비율의 개념이 추가된 척도

예) 무게에 있어서 절대적인 기준은 아무런 무게가 없는 0이 되며 모든 관측대상의 무게는 이러한 절대 기준을 바탕으로 평가가능함. 산의 높이에 있어서 절대적인 기준은 해수면이 0이 되며 백두산의 높이는 2,750m이고 한라산의 높이가 1,950m이므로 한라산은 백두산의 높이에 약 71%에 해당된다고 볼 수 있음
(즉, 백두산의 높이에 0.71을 곱하면 한라산의 높이를 구할 수 있음)

- 비율척도로 측정된 값들은 서로 곱하거나 나누는 것이 가능함

- 4가지 척도 가운데 가장 많은 정보(순서에 관한 정보, 대상 간의 상대적인 크기를 비교할 수 있는 정보, 서로 다른 상황에서 각각 독립적으로 측정된 값들 간에도 절대적인 기준을 바탕으로 그 크기를 비교할 수 있는 비율정보)를 포함함

- 등간척도

- 절대적인 기준이 없어 같은 조사과정에서 일정한 크기의 척도로 측정된 값들 간의 상대적인 크기 비교만 가능하고 서로 다른 조사에서 측정된 값들이면 직접 비교하기는 어려움

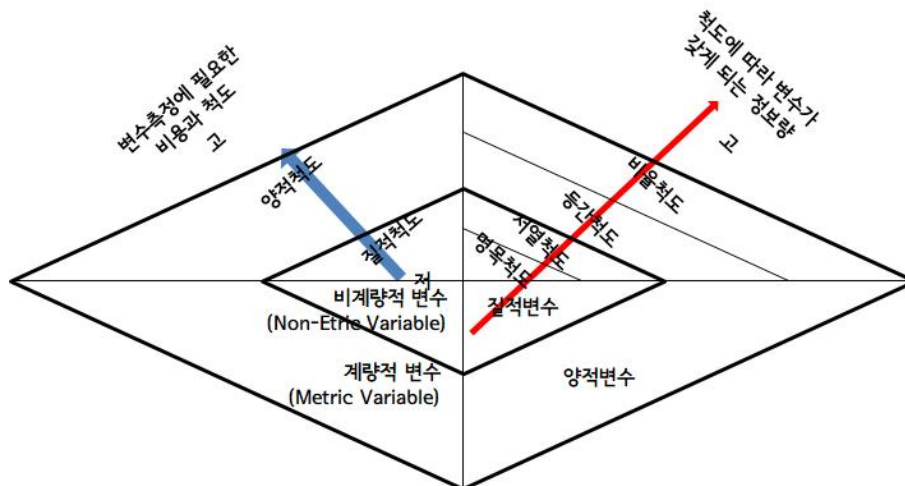
- 예시

패밀리 레스토랑 이용에 관한 문항입니다.

1. 귀하께서 지난 3개월간 월 평균 몇 번이나 패밀리 레스토랑을 이용하셨습니다?
한 달에 ()번
2. 위 패밀리 레스토랑을 이용하실 때 1인당 기준으로 어느 정도의 비용을 지출하십니까?
1인당 ()만 원

3) 척도의 특징

- 비율척도, 등간척도, 서열척도, 명목척도 순으로 많은 정보를 포함
 - 많은 정보를 얻을 수 있는 척도를 사용할수록 응답자나 조사자에게 부담을 주게 되어 실제로 측정하고자 하는 정확한 값을 얻기가 어려워짐
 - 보다 많은 정보를 얻을 수 있는 척도로 측정하기 위해서는 보다 많은 노력과 시간 그리고 비용이 요구됨
- 조사자가 필요로 하는 정확한 정보를 얻기 위해서는 응답자에게 너무 큰 부담을 주지 않는 적당한 수준에서 정보를 추출할 수 있어야 함
 - 특히 설문지 등을 이용하여 정보를 얻고자 하는 경우에는 이러한 정보의 양과 정확성의 트레이드 오프(Trade-off) 관계를 고려하여 가장 정확한 정보를 효과적으로 얻을 수 있는 척도를 사용해서 측정해야 함



- ☞ 연구자는 응답자 부담과 연구내용, 그리고 얻고자 하는 정보량을 고려하여 적합한 척도를 선택해야 함

2. 변수의 의미와 유형

1) 변수의 의미

- 척도를 이용하여 관심대상이 되는 개체(Item)의 속성을 측정한 값을 대표하여 일컫는 말로 연구자가 관심을 가지고 있는 대상의 속성을 척도로 측정하여 수치로 계량화하는 규칙
- 연구대상이 되는 개체가 가지고 있는 속성 중에 연구자가 관심을 갖는 속성을 척도로 측정하여 수치로 표현한 것으로 변수는 그 값으로 2개 이상의 서로 다른 값들, 즉 다양하게 변하는 값들을 가져야 함
 - 변수가 갖는 일정한 값들이 모두 하나의 숫자라면 이것은 변수가 아니라 상수임
 - 예) K대학교의 학생들의 성별을 측정하고 그 값을 수치로 나타내는 경우 남성은 1로 여성은 0으로 나타내는 규칙의 변수를 생각할 수 있고 또한 남성은 61로 여성은 20으로 나타내는 또 다른 규칙의 변수를 고려할 수 있음
- 변수는 관심대상의 속성을 측정하여 그 값을 수치로 나타내는 하나의 규칙으로 보는 것이 보다 이해하기 쉽고 정확한 표현임

2) 변수의 유형

- 연산가능여부에 따라

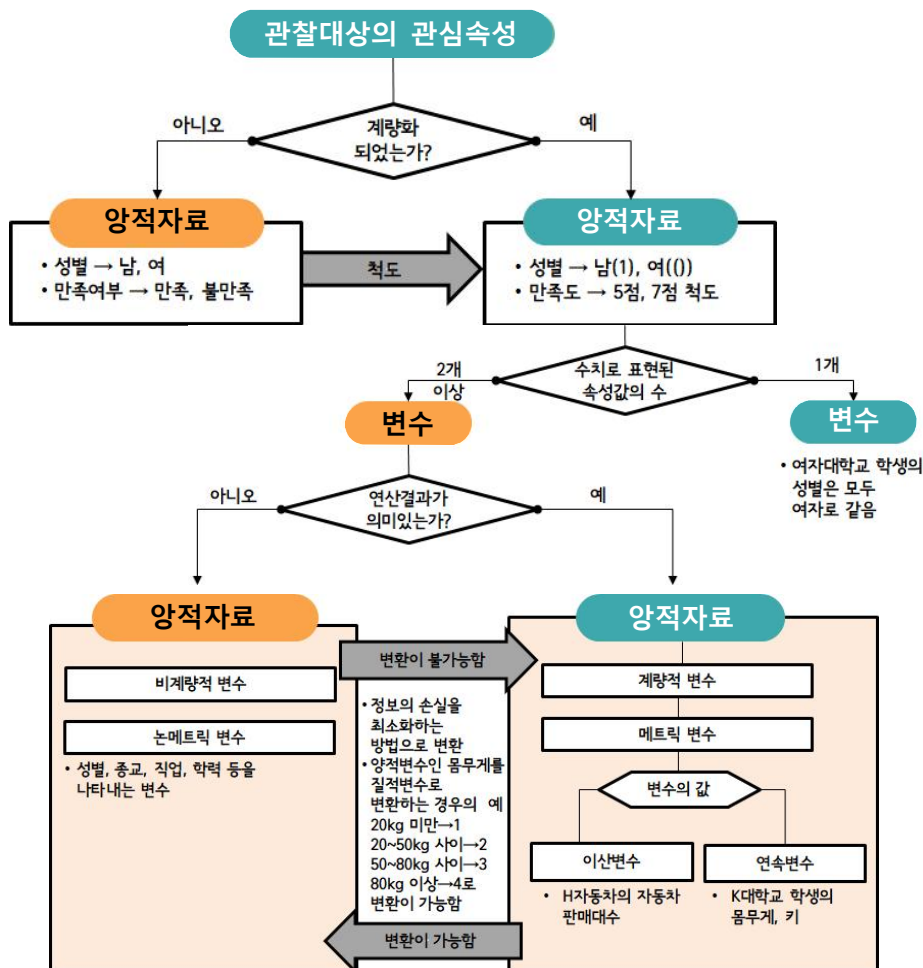
구분	질적변수	양적변수	
의미	<ul style="list-style-type: none"> - 속성의 값을 나타내는 수치의 크기가 의미없는 변수로서 비계량적 변수 혹은 논메트릭(non-metric) 변수라고 함. - 가감이 가능 	<ul style="list-style-type: none"> - 측정한 속성값을 연산이 가능한 의미 있는 수치로 나타내는 변수로서 계량적 변수 혹은 메트릭(metric)변수라고 함. - 가감승제가 가능 	
		이산변수	연속변수
		학생수나 자동차 판매대수와 같이 정수값만을 취할 수 있는 변수	몸무게나 키와 같이 연속적인 모든 실수값을 취할 수 있는 변수
예	성별을 나타내는 변수의 경우, 남성을 1로 하고 여성을 2라는 수치로 표시, 이외에도 종교, 직업, 학력 등을 구분하기 위하여 숫자로 나타내는 경우	만족도, 유사도, 시험점수, 몸무게, 판매금액 등	
사용 척도	명목척도, 서열척도	등간척도, 비율척도	

- 대부분의 경우 양적변수를 질적변수로 바꾸는 것은 가능함
- 예) 몸무게가 80kg 이상 사람이 무겁다하여 1로 표시하고, 80kg 이하인 사람을 0으로 표시하는 것은 양적변수를 질적변수로 변화시킨 것으로 볼 수 있음
- 그러나 질적변수를 양적변수로 바꾸는 것은 거의 불가능하며, 가능한 경우에도 매우 제한적이며, 변화에 따른 적지 않은 오차발생을 감수해야 함

- 사용되는 용도에 따라

구분	독립변수(independent variable)	종속변수(dependent variable)
내용	독립적으로 변하는 변수로 조사자에 의해 조작이 가능한 변수	독립변수에 종속되어 그 값을 갖게 되는 변수로 독립변수값에 따라 수동적으로 값이 주어지는 변수
예	상품이미지, 품질, 서비스 수준, 가격	상품에 대한 만족도
사용하는 상황에 따라	원인, 설명, 예측변수	결과, 목적, 타깃변수

3) 자료와 변수 및 상수



확률과 확률변수 / 7주차 1차시

1. 확률의 기본개념

1) 기본개념

■ 통계분석

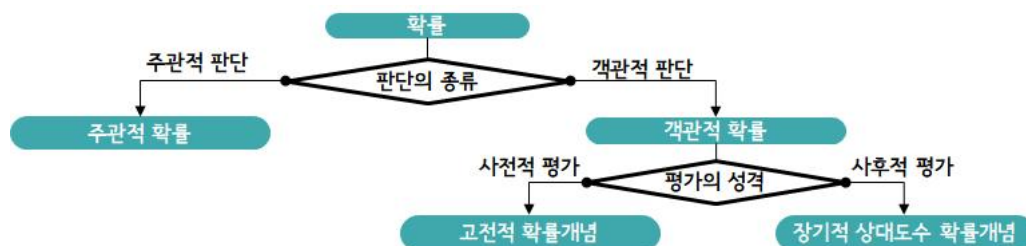
- 모집단 전체 자료를 분석하는 것이 아니라 모집단으로부터 추출한 표본의 자료를 분석하기 때문에 확률은 통계학이나 조사방법론에서 매우 중요한 개념임
 - 모집단으로부터 추출한 표본을 분석하여 얻은 표본통계량으로 모집단의 특성을 나타내는 모수를 추정하고 모수에 대한 가설을 검정함
- 모집단으로부터 어느 표본이 선정되는가에 따라 표본통계량값이 달라짐
 - 모집단으로부터 추출된 표본의 특성을 나타내는 통계량은 확정적인 값이 아니라 선정된 표본에 따라 값이 달라지는 즉 확률이 개념을 포함한 값이 될 수 밖에 없음

■ 확률에 대한 이해가 필요한 이유

- 통계분석에서 사용되는 표본은 모집단으로부터 추출할 수 있는 여러 가능한 표본 중의 하나에 불과하므로 표본을 선정하는 행위는 확률개념을 자동적으로 포함하기 때문에 확률을 충분히 이해해야 추정 및 가설검정과 같은 통계분석을 정확하게 이해할 수 있음.

2) 확률의 정의

- 경험 혹은 실험의 결과로 특정한 사건(event)이나 결과가 발생할 가능성
- 확률은 우리의 일상생활에서 항상 사용되며, 대부분의 의사결정은 확률 혹은 확률에 대한 판단을 필요로 함
- 확률은 크게 주관적 확률과 객관적 확률로 구분



- 주관적 확률

- 경험과 직관에 의한 주관적인 판단
 - ✓ 내일 비가 올 확률은 어느 정도일까?
 - ✓ 한 60%쯤 되지 않을까?
 - ✓ 지훈이가 내일 아침수업에 늦을 확률은 얼마나 될까?
 - ✓ 그 가능성은 10% 정도라고 생각해

- 고전적 확률개념

- 사전 확률: 각각의 결과가 나올 확률은 객관적으로 미리 판단할 수 있는 경우
 - ✓ 룰렛 게임에서 사망할 확률은 얼마지?
 - ✓ 그야 1/6이지. 6발 들어가는 리볼버에 총알은 1개이니까
 - ✓ 동전을 던져 앞면이 나올 확률은 1/2이다.

- 장기적 상대도수 확률개념

- 사후 확률 : 같은 실험을 수없이 많이 반복했을 때 특정사건이 발생할 수 있는 상대적 가능성
 - ✓ 동전을 던지는 실험을 반복적으로 무수히 실행할 경우 앞면이 나올 확률은 1/2이라 할 수 있나?
 - ✓ 할 수 있다. 마찬가지로 정육면체인 주사위를 반복적으로 무수히 많이 던질 경우 5가 나올 확률은 1/6이다.

주관적 확률	객관적 확률	
<ul style="list-style-type: none"> - 의사결정자의 지식, 정보 및 경험을 바탕으로 한 주관적 판단에 의하여 결정되는 확률 - 객관적인 정보나 경험이 제한된 상황에서 어떤 사건이 발생할 가능성을 예측해야 할 경우, 주로 주관적 확률이 사용됨 - 판단하는 사람에 따라 상이 - 평가자의 지식이나 정보 그리고 경험이 충분하게 축적됨에 따라 객관성을 띄게 될 가능성이 점차 높아짐 	<ul style="list-style-type: none"> - 특정 사건이 발생할 가능성이 객관적으로 명확하여 누구나 쉽게 알 수 있는 경우나, 동일한 실험을 무수히 반복적으로 수행할 경우 특정한 사건이 발생할 수 있는 확률 	
	고전적 확률개념	장기적 상대도수 확률개념
	<ul style="list-style-type: none"> - 사전적 평가에 의한 확률 - 사전적 확률개념은 모든 결과가 발생할 가능성이 명확하고, 각 결과는 서로 배타적일 때, 즉 서로 다른 2개 이상의 결과가 동시에 발생할 수 없는 경우에 사용 - 실제적인 경험에 의한 것이 아니고, 특정한 결과가 발생할 가능성을 직관적이며 객관적으로 판단하여 추정한 확률 	<ul style="list-style-type: none"> - 사후적 평가에 의한 확률 - 실제 실험이나 경험에 근거한 사후 확률개념임 - 사후적 확률개념은 동일한 실험을 반복적으로 수행할 때 특정 사건이 발생할 수 있는 상대적 빈도로 정의됨 - 반복적으로 실험을 시행하는 경우, 어느 특정한 값에 수렴하는 확률값을 얻을 수 있음. - 상대도수의 극한치를 추정하거나 한 사건이 발생할 기대상대도수를 구하여 확률산출

3) 실험과 확률

■ 실험과 표본공간

- 실험

- 어떤 행위의 결과를 관찰하고 측정하여 그 결과에 대해 구체적인 값을 부여하는 것(어떤 일을 수행하고 그 결과로 나온 값을 기록하여 자료로 남기는 행위)

예) 동전 던지기를 예를 들면,

동전을 던져서 나온 결과가 앞면인지 아니면 뒷면인지 확인하여 기록하는 행위

- 표본점

- 한 번의 실험으로부터 얻을 수 있는 결과로 단일사상 혹은 단일사건으로 불림

예) 주사위를 한 번 던져서 나온 값이 3이면 이것이 바로 표본점

- 표본공간

- 실험을 통해서 얻을 수 있는 모든 가능한 표점들의 집합

예) 주사위를 던져 나온 앞면의 숫자를 기록하는 실험의 표본공간을(1,2,3,4,5,6)임

■ 확률과 표본공간

- 조사를 위하여 모집단으로부터 표본을 추출하는 것도 일종의 실험행위이므로

모집단으로부터 추출된 표본은 하나의 표본점으로서, 표본으로 얻을 수 있는 수없이 많은 표본점 중의 하나

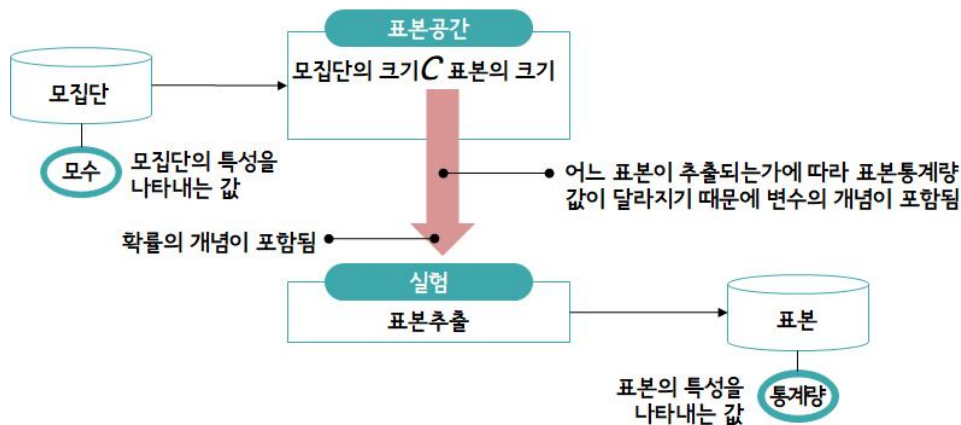
- 특정한 표본이 추출될 확률은 '1/표본 공간상의 표본점의 수'가 됨

- 표본을 추출하여 얻는 표본으로부터 얻게 되는 모든 정보에는 자연스럽게 확률의 개념이 포함됨

2. 확률과 확률변수

1) 확률변수의 정의

- 경험이나 실험결과 특정한 사건이 일어날 수 있는 가능성인 확률과 관심대상의 속성을 척도를 이용하여 측정한 값인 변수의 개념이 결합된 것
- 모집단으로부터 추출한 표본의 평균이나 분산과 같은 통계량은 확률과 변수의 개념을 모두 포함하고 있는 대표적인 확률변수
- 대부분의 통계분석은 모집단으로부터 추출된 표본을 측정하여 얻은 값들을 분석하는 것이므로 분석결과로 얻게 되는 표본통계량들은 모두 확률변수에 해당됨



3. 확률변수의 정의

1) 확률변수의 정의

- 표본공간상에 나타나는 모든 표본점들이 수치를 부여하는 규칙

2) 사례

- 2개의 동전을 던질 때 나올 수 있는 경우의 수, 즉 표본점은 총 4가지임

표본공간: $S = \{(H, H), (H, T), (T, H), (T, T)\}$

확률: $P = \{1/4, 1/4, 1/4, 1/4\}$

확률변수 X 는 동전의 앞면은 1, 뒷면은 0으로 하여 두 개의 동전을 던져 얻은 결과를 합한 값으로 정의함

표본공간	확률변수(X)의 값(x)	x값을 가질 확률
(T, T)	0	1/4
(H, T), (T, H)	1	$1/4 + 1/4 = 1/2$
(H, H)	2	1/4

모집단에서 추출된 표본의 특성을 나타내는 통계량은 확률변수임

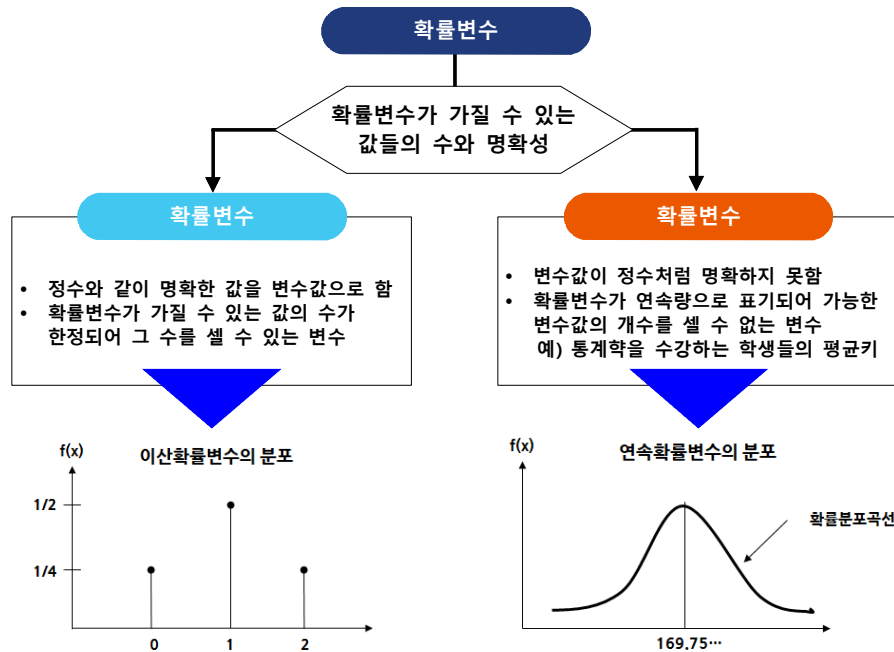
확률변수는 확률값과 연동하여 특정한 값을 갖게 되는 변수를 말함

- 표본공간상의 표본점들을 일정한 규칙을 사용하여 수치값으로 전환하는 방법은 수없이 많음으로 하나의 표본공간으로부터 수없이 많은 확률변수들이 생성될 수 있음

예) ① 확률변수 X 는 주사위를 던져 주사위 눈이 4이하이면 0, 주사위 눈이 5 이상이면 1로 하는 규칙, ② 확률변수 Y 는 주사위를 던져 주사위 눈이 홀수이면 1, 주사위 눈이 짝수이면 2로 하는 규칙 등

4. 확률변수의 종류

- 변수가 이산변수와 연속변수로 나누어지는 것과 같이 확률변수도 이산확률변수와 연속확률변수로 나누어짐
- 확률변수가 가질 수 있는 값과 그 값을 가질 확률을 표시한 것이 확률분포임



확률분포 / 7주차 2차시

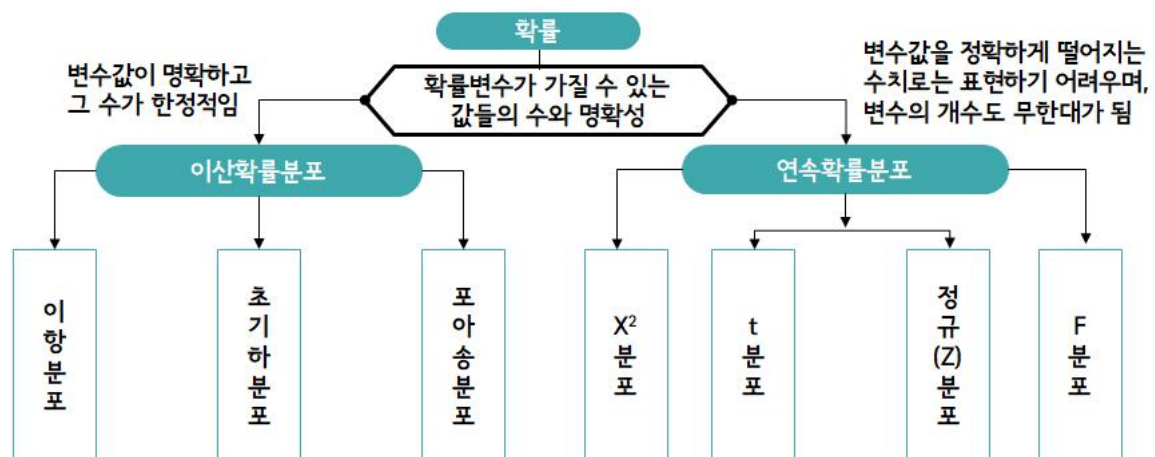
1. 확률분포의 정의

1) 정의

- 확률변수가 특정한 값을 가질 확률, 즉 상대적 가능성을 나타낸 것으로 모든 가능한 확률변수 값과 그 값이 발생할 가능성인 확률값을 도수분포표나 그래프로 나타낸 것
- 확률변수가 가질 수 있는 값과 그 값을 가질 확률값을 요약해서 알기 쉽게 표현한 것
- 확률변수
 - 그래프로 확률분포를 나타내는 경우,
 - 확률분포를 나타내는 함수를 확률함수(probability function)라고 함

2. 확률분포의 종류

- 확률분포는 확률변수의 종류에 따라 이산확률분포와 연속확률분포로 나누어짐
- 이산확률분포
 - 이산확률변수의 확률분포
- 연속확률분포
 - 연속확률변수의 확률분포
- 무수히 많은 확률분포 중에서 분포의 특성을 파악하여 함수로 나타낼 수 있는 대표적인 확률분포들이 있음



- 대부분의 통계는 그 특성을 알 수 있는 이러한 분포들을 이용해서 추정 및 가설검정을 수행함

1) 정규분포

- 연속확률분포 중에서 가장 대표적인 분포로서 통계학과 조사방법론에서 가장 많이 사용되는 확률분포
- 사람들의 몸무게나 키 그리고 지능지수(IQ) 등과 같은 대부분의 변수들은 정규분포를 따름. 이러한 모집단으로부터 추출한 평균(\bar{x})은 관측치의 수가 많아질수록 정규분포에 더욱 가까워짐
- 정규분포의 가장 큰 특징은 좌우대칭인 종모양을 이루고 있음
- 조건
 - ① 종모양의 중앙이 볼록할수록 정규분포에 더 가까워짐
 - ② 정규분포는 평균(μ)을 중심으로 좌우 1개의 표준편차($\mu \pm \sigma$) 안에 측정된 관측치, 즉 확률변수값들의 68.26%가 포함되어 있어야 하고, 좌우 2개의 표준편차($\mu \pm 2\sigma$) 안에는 95.44%의 확률변수값들이 있어야 함
 - ③ 좌우 3개의 표준편차($\mu \pm 3\sigma$) 안에는 99.74%에 이르는 대부분의 값들이 있음
 - 이러한 조건들이 반드시 일치해야만 정규분포한다고 할 수는 없으나 세 번째 조건, 즉 평균을 중심으로 양쪽 3개씩의 표준편차($\mu \pm 3\sigma$) 안에 대부분의 관측치, 즉 변수값들이 포함되어야 한다는 것은 비교적 엄격한 조건으로 인식됨
- 특징
 - ① 정규분포는 평균(μ)을 중심으로 좌우대칭인 종모양을 하고 있음
 - ② 확률밀도함수곡선과 X축 사이의 전체 면적의 합은 1이 됨
 - ③ 확률변수 X가 취할 수 있는 값의 구간은 $-\infty < x < +\infty$ 이므로 정규분포곡선도 이 구간에 존재하는 것으로 가정함
- 정규분포의 위치와 모양을 결정하는 모수는 평균(μ)과 분산(σ^2)이며 정규분포는 평균(μ)을 중심으로 좌우대칭인 종모양의 분포이기 때문에 평균값에 따라 분포가 좌우로 이동을 함
- 변수값들이 평균에 가까이 분포하고 있는지 아니면 넓게 퍼져 있는지는 분산으로 알 수 있음
 - 분산이 커서 관측치들이 평균을 중심으로 넓게 퍼져 있으면 분포는 평균을 중심으로 평평한 구름모양이 되고 분산이 작으면 변수값들이 평균 주위에 몰려 평균을 중심으로 비교적 뾰족한 종모양을 띄게 됨.
- 일반적으로 평균이 μ 이고 분산이 σ^2 인 정규분포는 $N(\mu, \sigma^2)$ 으로 표기함
 - 예) $X \sim N(5, 16)$ 이라는 표기는 확률변수 X는 평균이 5이고 분산이 16인 정규분포를 한다는 의미임

2) 표준정규분포

- 정규분포는 평균(μ)과 분산(σ^2)의 크기에 따라 여러 가지 형태를 가지므로 서로 다른 두 개의 분포를 비교하거나 확률분포상의 면적의 크기를 계산하여 확률을 알아내는 것은 쉽지 않음
 - 평균과 분산에 따라 변하는 정규분포의 모양과 측정단위에 무관하게 어떠한 정규분포이든 특정구간 사이의 확률값을 구하는 방법이 필요

- 표준화된 측정단위를 사용하여 모든 정규분포의 기준이 되는 정규분포가 필요하므로 표준편차를 변수값의 단위로 하여 만들어진 정규분포를 표준정규분포라고 함
- 표준정규분포($N(0, 1)$)는 평균이 0이고 분산이 1인 정규분포로서 표준편차를 변수값의 단위로 사용하기 때문에 모든 정규분포는 표준정규분포로의 전환이 가능함
- 표준정규분포상 특정 구간의 확률값들은 미리 계산되어 대부분의 통계학이나 조사방법론 교재에 부록으로 제시됨
- 일반적인 정규분포상의 구간을 표준정규분포상의 구간으로 바꾼 다음 제시된 표를 찾아봄으로써 해당 구간의 확률값을 손쉽게 구할 수 있음
- 표준정규분포는 평균이 0이고 분산이 1이기 때문에 표준정규분포의 확률밀도 함수식에는 평균과 분산이 별도로 표시되지 않으며 일반정규분포와 달리 평균과 분산에 대해 독립적임
- 표준정규분포는 확률변수를 Z로 표시하기 때문에 Z분포라고 함.

3) t분포

- 좌우대칭이라는 점은 정규분포와 같으나 정규분포보다 팽팽하고 두터운 꼬리모양을 가진 구름모양의 분포
- t분포의 특징
 - ① 표준정규분포와 같이 평균인 0을 중심으로 좌우대칭인 낮은 구름이나 종모양
 - ② 자유도($n-1$), 즉 표본평균(\bar{x})을 구하는 데에 사용된 관측치의 수(n)에 따라 모양이 변함
 - ③ 일반적으로 t분포의 모양은 자유도($n-1$)가 적을수록 표준정규분포보다 더 평평하고 두터운 꼬리모양을 가짐
 - ④ 자유도($n-1$)가 증가함에 따라 분산은 1에 접근하여, 표본의 크기(n)가 30개 이상이면 t분포는 표준정규분포와 거의 동일한 모양의 분포를 이룸
- 표본의 크기(n)가 30개 이상일 경우 확률분포 t의 분포는 표준정규분포와 거의 동일하기 때문에 t분포 대신에 Z분포를 사용하여 확률값을 구할 수 있으나 30개 미만인 경우 Z분포를 이용하여 통계분석을 할 수 없음
 - t분포의 특수한 형태가 Z분포이므로 t분포가 Z분포를 포함한다고 볼 수 있음
 - 이러한 이유로 통계분석에서 Z분포보다는 t분포가 사용됨

4) 카이제곱(χ^2)분포

- 모집단분산에 대한 가설검정이나 교차분석에서 유용하게 사용되는 분석 평균(μ)이나 비율(π)에 대한 검정에는 t분포를 사용하나, 분산(σ^2)에 대한 검정을 할 경우에는 카이제곱(χ^2)분포나 F분포를 이용함
- 단일 모집단으로부터 추출한 표본분산(s^2)과 모집단분산(σ^2)의 비를 나타내는 확률변수는 카이제곱(χ^2)분포를 하고 서로 다른 두 모집단으로부터 추출한 표본분산의 비를 나타내는 확률변수는 F분포를 함.
- 명목척도로 측정한 변수들 간의 상호 연관성을 분석하는 경우에도 카이제곱(χ^2)분포가 사용됨.

■ 특징

- ① 연속확률분포로서 확률변수는 항상 양의 값만을 가짐
- ② 오른쪽꼬리를 가진 비대칭분포임
- ③ 카이제곱분포는 t분포와 마찬가지로 자유도에 의해서 모양이 결정됨
- ④ 카이제곱분포도 t분포처럼 자유도가 커짐에 따라 정규분포에 접근

5) F분포

- 카이제곱 분포와 유사한 모양의 분포를 가지는 비대칭 연속확률분포로 2개의 카이제곱분포값을 각각의 자유도로 나눈 평균 카이제곱값의 비를 변수값으로 하는 확률변수의 분포
- 정규분포를 이루며 서로 독립적인 두 모집단에서 각각 표본을 추출한 경우, 두 집단의 표본분산의 비(S_1^2/S_2^2)를 모집단분산의 비(σ_1^2/σ_2^2)로 나눈 값을 변수값으로 하는 확률변수는 F분포를 함
- 2개의 표본분산이 사용되기 때문에 2개 표본의 자유도에 따라 F분포의 모양이 결정됨(즉, 2개의 자유도가 F분포의 모양을 결정하는 파라미터가 됨)
- 일반적으로 F분포는 오른쪽꼬리를 갖는 비대칭 모양의 분포이나, 자유도(df)가 증가함에 따라 점차 정규분포 모양을 띄게 됨
- F분포하는 확률변수는 항상 양(+)의 값만을 가짐

3. 중심극한정리

1) 정의

- 모집단 전체를 조사하고 분석하는 것은 상대적으로 많은 시간과 비용을 필요로 하기 때문에 대부분의 통계분석은 모집단으로부터 표본을 추출하여 분석
 - 표본을 분석하여 얻은 정보를 이용하여 모집단의 모수를 추정하고 모수에 대한 가설을 검정하여 의사결정을 함
- 통계량 혹은 추정량
 - 추출된 표본에 대하여 조사하고 분석하면, 표본의 특성을 요약하여 설명하는 다양하고도 많은 값들을 얻을 수 있는데 표본의 특성을 나타내는 값들을 통계량 혹은 추정량이라고 함
 - 표본분포는 이러한 통계량의 분포를 총칭하는 말
- 여러 가지 표본통계량 중에서 가장 대표적인 통계량은 표본의 중심성향을 나타내는 표본평균(\bar{x})이나 표본비율(p)임.
- 표본평균(\bar{x})이나 표본비율(p)의 분포는 표본으로 추출되는 관측치의 수, 즉 표본의 크기(n)가 증가함에 따라 완만한 구름모양에서 점차 종모양의 정규분포에 접근표본의 크기(n)가 증가함에 따라 표본평균(\bar{x})이나 비율(p)의 분포가 정규분포에 근접해 가는 현상
- 일반적으로 표본크기(n)가 30개 이상이면, 약간의 예외적인 경우를 제외하고 모집단 분포와 관계없이 표본평균(\bar{x})의 분포는 거의 정규분포에 근접하는 분포를 함

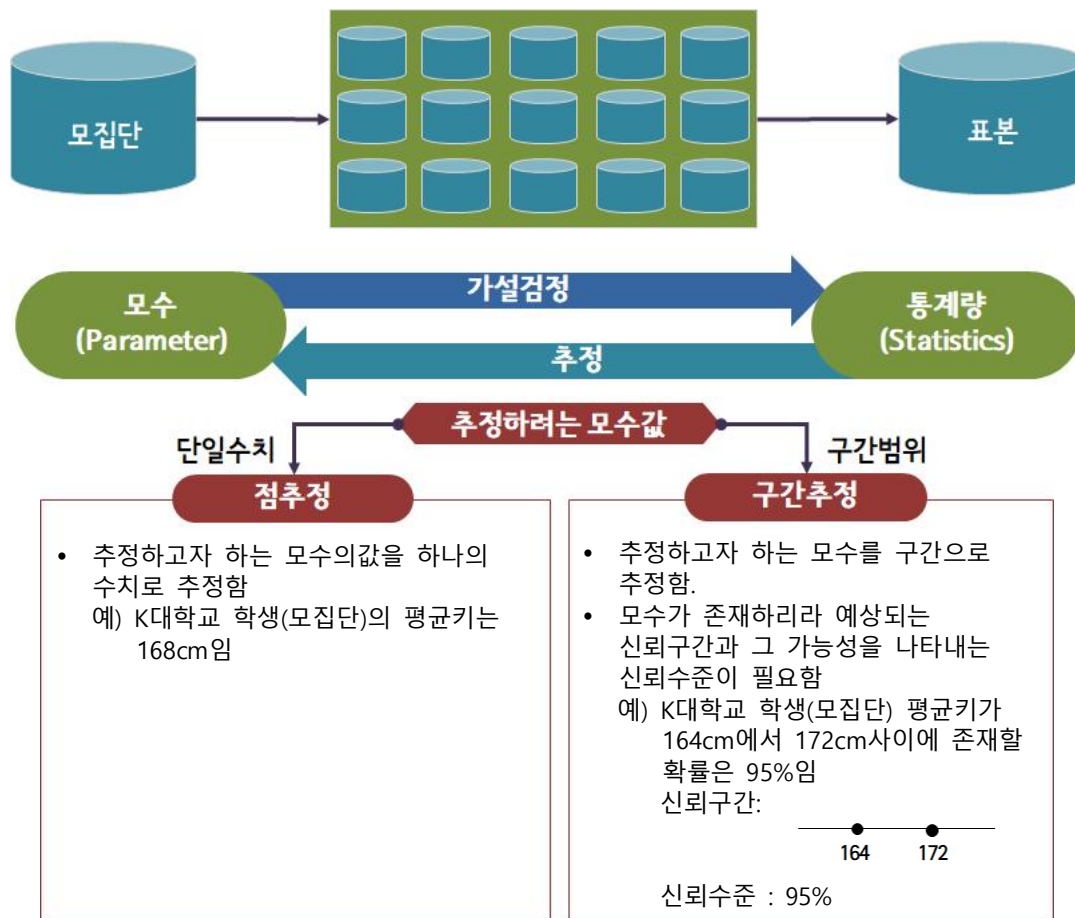
- 2) 자료의 표준화
- 표본평균(\bar{x})이 정규분포를 하여도 정규분포상의 특정 구간 내의 값을 가질 확률을 직접 구하는 것은 불가능함
 - 표본평균(\bar{x})을 자유도가 $n-1$ 인 t 분포상의 값으로 표준화하여 확률값을 계산 자유도가 $n-1$ 인 t 분포상의 확률값은 미리 계산하여 t 분포의 확률표를 사용하면 쉽게 구할 수 있음

추정 / 9주차 1차시

1. 추정의 기본개념

1) 추정의 정의

- 통계에는 기술통계와 추론통계가 있으며, 추론통계는 표본의 특성을 나타내는 통계량으로 모집단의 특성을 나타내는 모수를 추론하는 통계
- 모수를 추정하는 방법



- 모집단에 대한 어떠한 정보가 없는 상태에서, 모집단을 대표할 수 있는 표본을 추출하여, 추출된 표본의 통계량을 구한 다음, 이를 이용해서 모집단의 모수를 예측하는 것, 표본의 통계량으로 모집단의 모수를 추정하는 것
- 일반적으로 모집단은 크기가 너무 크기 때문에 전수조사를 하기 위해서는 많은 비용과 시간이 소요되며, 어떤 모집단은 전수조사가 불가능한 경우도 있음
 - 대부분의 경우 모집단을 대표할 수 있는 표본을 추출하여 그 표본의 통계량으로 모집단의 모수를 추정함

2) 추정의 종류

■ 점추정

- 추정하고자 하는 모수를 하나의 수치로 추정하는 것으로 알려져 있지 않은 모수를 추정하는 경우, 모수의 진정한 값이라고 추측되는 단일수치로 모수를 추정하는 것
- 점추정의 경우 모수를 추정하기 위한 통계량을 추정량이라고 함
 - 모평균(μ)의 추정량은 표본평균(\bar{x})이 되고 모분산의 추정량(σ^2)은 표본분산(s^2)이 됨
 - 구체적인 수치로 표시된 특정한 표본추정량 값을 추정치라고 함
- 예) 선거에서 A후보의 지지율을 예측하고자 할 경우, 표본조사를 통해 서유권자가 A후보를 지지하는 지지율이 36%로 나왔다면, 'A후보에 대한전체 유권자의 지지율은 36%이다'라고 하나의 값으로 추정하는 것이 점추정임
- 점추정의 한계
 - 점추정의 경우 선택된 추정량은 모수와 정확하게 일치하는 경우는 거의 없음
 - 하지만 이러한 정보도 없으면, 지지율이 얼마나 될 것인지 전혀 알 수 없음으로 점추정이 정확하지는 않아도 실제 얼마나 될 것인지를 짐작하여 예측하는 데 도움을 줌

■ 구간추정

- 추정하고자 하는 모수가 특정한 구간 내에 위치할 확률이 어느 정도라고 추정하는 것으로 일반적으로 사용되는 대부분의 추정은 구간추정임
- 예) 선거 때 TV방송이나 신문에서 '여론조사 결과 A후보의 지지율은 36%이고, 오차범위는 95%(0.95) 신뢰수준에서 ± 3 이다'라고 말하는 경우
- 구간추정을 위해서는 신뢰수준과 신뢰구간이 필요함

구분	내용
사례	여론조사 결과 A후보의 지지율은 36%이고, 오차범위는 35%(0.95) 신뢰수준에서 ± 3 임
신뢰수준	95%
신뢰구간	오차범위 ± 3 을 고려한 33%~39%에 이르는 구간

- 신뢰구간은 모수가 포함될 가능성이 있는 구간을 말하여, 신뢰수준과 비례하여 구간의 크기가 변화게 됨
 - 표본의 크기(n)가 일정한 경우 신뢰수준이 높아지면 신뢰구간이 넓어지고, 신뢰수준이 낮아지면 신뢰구간이 좁아지게 됨
- 일반적으로 신뢰구간은 하한값과 상한값으로 표기하며, 신뢰구간의 하한을 신뢰하한, 신뢰구간의 상한을 신뢰상한이라고 함
- 신뢰구간 내에 모수가 포함될 확률이 바로 신뢰수준이며 이는 전체 1에서 허용오차(α)를 뺀 값($1-\alpha$)임
- 대부분의 추정에서는 추정치가 모수와 일치하지 않았을 경우에 발생할 수 있는 오차에 대한 정보가 없는 점추정보다는 어느 정도 오차를 허용하는 범위내에 모수가 존재할 가능성에 대한 정보를 충분히 활용할 수 있는 구간추정이 보편적으로 사용됨

2. 구간추정

1) 신뢰수준과 신뢰구간

■ 신뢰수준

- 신뢰구간에 모수가 위치할 것이라 믿는 확률을 말하는 것으로 연구자의 믿음을 의미
 - 일반적으로 신뢰수준은 $100 \times (1 - \alpha)\%$ 로 나타내며, α 는 허용오차 수준으로 100%에서 신뢰수준을 뺀 값으로, 진정한 모수가 신뢰구간 밖에 존재할 가능성을 나타냄
예) 95%의 신뢰수준이라 하면 α 는 5%가 되며, 이는 진정한 모수가 신뢰구간 내에 있으리라고 95%는 신뢰할 수 있으나 이러한 믿음이 잘못될 가능성이 5% 정도 된다는 것을 의미함
 - 허용오차수준(α)과 신뢰수준($1 - \alpha$)의 합이 1이 되기 때문에 신뢰수준과 허용오차수준은 서로 반비례함
 - 모수가 신뢰구간내에 있을 확률은 신뢰수준($1 - \alpha$)이며, 신뢰구간을 벗어나 신뢰구간 밖에 있을 확률이 허용오차수준(α)임

■ 신뢰구간

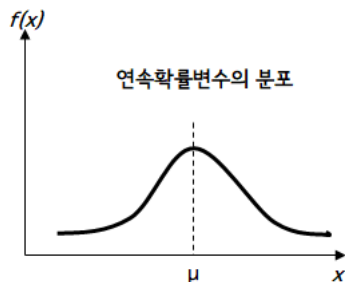
- 실제 모수가 있을 것이라고 예상되는 구간을 말하며, 일반적으로 하한값과 상한값으로 표기
 - 신뢰구간의 크기는 신뢰수준에 비례함
예) A후보의 지지율을 추정하는 경우, A후보의 진정한 지지율이 좁은 신뢰구간보다는 넓은 신뢰구간 내에 있다고 보다 자신있게 확신할 수 있음
 - 즉 35%~40%에 A후보의 지지율이 있을 것이라고는 정확하게 확신하기 어렵지만, 1%~99% 사이에 A후보의 지지율이 있을 것이라고는 거의 100% 확신할 수 있음
 - 신뢰수준이 높아지면 신뢰구간의 폭도 넓어지며, 신뢰수준이 낮아지면 신뢰구간의 폭도 좁아지게 됨

2) 모집단평균의 구간추정

- 모집단 평균(μ)을 표본평균(\bar{x})을 이용하여 구간 추정하는 경우, 중심극한정리에 의해서 모집단의 분포와는 관계없이 표본의 크기(n)가 증가함에 따라 모집단 평균(μ)을 중심으로 구름모양의 분포나 종모양의 정규분포를 보임
- 모집단의 평균이 존재하리라 예상하는 구간, 즉 신뢰구간은 표본평균(\bar{x})을 중심으로 형성되며 신뢰구간의 크기는 신뢰수준($1 - \alpha$)과 표본평균(\bar{x})의 표준편차, 즉 표준오차에 의하여 결정됨
 - 표본평균(\bar{x})의 분산에 루트를 씌운 값, 즉 제곱근값임
- 표본평균(\bar{x})의 분산은 표본의 크기(n)가 커질수록 점차 작아짐
- 표본평균(\bar{x})이 표준편차인 표준오차도 작아짐
- 표준오차가 작아짐에 따라 신뢰구간도 좁아짐
 - 즉, 표본의 크기(n)가 커짐에 따라 표본평균(\bar{x})의 표준편차인 표준오차가 작아져서 같은 신뢰수준하에서 신뢰구간은 좁아지게 됨

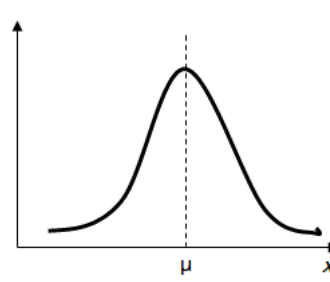
■ 표본의 크기(n)에 따른 표본평균의 분포

(표본을 10명으로 추출한 경우)



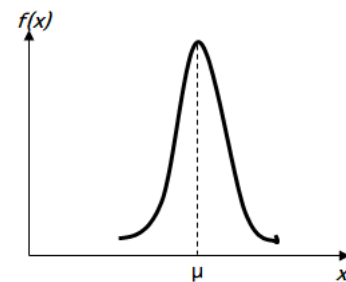
- 표본평균의 분산이 큼

(표본을 100명으로 추출한 경우)



- 표본평균의 분산이 작아짐

(표본을 999명으로 추출한 경우)



- 표본평균의 분산이 매우 작아 0에 근접함

표본의 크기



■ 신뢰수준은 모집단의 모수에 대한 추정이 잘못될 가능성의 정도, 즉 연구자가 허용할 수 있는 허용오차(α)에 따라 결정됨

- 신뢰수준($1-\alpha$)이 결정되면 표본평균(\bar{x})이 Z분포 혹은 t분포한다는 가정하여 표본평균(\bar{x})을 중심으로 하여 좌우 임계치까지 면적인 신뢰수준이($1-\alpha$)이 되도록 신뢰구간을 결정하게 됨

■ 정규분포의 경우 집단의 평균을 중심으로 좌우 1σ (표준편차) 내의 확률은 68.26%이고 좌우 2σ (표준편차) 내의 확률은 95.44%이며, 또한 좌우 3σ (표준편차) 내에는 99.74%에 이르는 대부분의 값들이 있어야 함

- 68.26%, 95.44%, 99.74%는 신뢰수준에 해당하여, 1σ (표준편차), 2σ (표준편차), 3σ (표준편차)는 이러한 신뢰수준을 확보하기 위하여 평균을 중심으로 좌우 몇 개의 표준편차가 필요한지를 나타내는 값임

■ 확률변수를 표준화하여 Z값이나 t값으로 치환할 경우, 구하고자 하는 신뢰수준을 확보하기 위하여 평균을 중심으로 좌우로 몇 개의 표준편차가 필요한지를 나타내는 값이 Z분포(표준정규분포)나 t분포상의 임계치가 됨

예) 95%의 신뢰수준의 범위를 나타내는 Z분포상의 임계치는 1.96이 됨

- 정규분포의 경우 평균을 중심으로 좌우로 1.96개의 표준편차만큼 떨어진 구간 내에 확률변수값이 존재할 확률이 95%라는 것을 의미함
- 표본평균(\bar{x})이 정규분포를 하는 경우 추출한 특정 표본의 평균을 중심으로 좌우로 1.96개씩의 표본평균(\bar{x})의 표준편차(표준오차)만큼 떨어진 구간내에 진정한 모집단의 평균이 존재할 확률이 95%라고 해석함

■ 모집단 분산에 대한 추정도 모집단 평균과 모집단 비율을 추정하듯이 표본분산을 이용해서 추정할 수 있음

■ 하지만 표본분산은 표본평균(\bar{x})이나 비율과 달리 정규분포를 하지 않기 때문에 다른 분포를 이용해서 구간추정을 해야 함

- 표본의 자유도($n-1$)에 표본분산을 곱한 값을 모집단 분산으로 나눈 값은 자유도가 $n-1$ 인 카이제곱 분포를 함으로 모집단 분산에 대한 구간추정은 카이제곱분포를 이용해서 구함

3. 표본의 크기

1) 표본의 크기의 정의

- 효과적인 추정을 하기 위해서는 모집단으로부터 표본을 추출할 경우 표본의 크기(n)을 얼마로 하여야 하는가?
- 설문조사를 한다면 몇 부나 조사해야 하는가?
- 표본의 크기 중요성
 - 표본의 크기(n)가 커질수록 표본평균(\bar{x})의 분산은 줄어들며, 표본평균(\bar{x})의 분포는 점차적으로 모집단 평균 주위에 밀집하지만 표본의 크기(n)이 커지면 커질수록 거의 전수조사에 가까우므로 조사과정에서 많은 시간과 노력 그리고 비용이 필요하게 됨
 - 추정의 정확도와 조사비용은 서로 트레이드 오프(trade-off)관계에 있음
- 추정의 목적 및 조사 여건과 상황을 고려하여 적절한 수로 표본의 크기를 결정하는 것이 바람직함
- 표본의 크기(n)을 결정하는 데 영향을 미치는 요인으로 연구자가 용납할 수 있는 최대 허용오차(e)와 신뢰수준이 있음
- 허용오차(e) vs 허용오차수준(α)

허용오차(e)	허용오차수준(α)
표본으로부터 얻은 추정량(\bar{x} , p)과 추정하고자 하는 모집단의 모수(μ , π)와의 차이 즉, 실제오차를 의미	실제 모수가 신뢰구간 밖에 존재할 확률 즉, 1에서 신뢰수준($1-\alpha$)을 뺀 확률값을 의미
예) '여론 조사 결과 A후보의 지지율은 36%이고 오차범위는 90% 신뢰수준($1-\alpha$)에서 ± 3 이다'라고 할 경우 신뢰구간을 33%~39%가 되고, 최대 허용오차(e)는 신뢰구간 크기의 절반인 3%가 되며, 오차수준(α)는 10%가 됨	

- 신뢰구간과 최대허용오차(e)는 신뢰수준($1-\alpha$)과 비례 관계에 있으며, 표본의 크기(n)와는 반비례관계에 있음
- 최대허용오차(e)를 정해놓은 상태에서, 즉 신뢰구간의 크기를 고정시킨 상태에서 신뢰수준을 높이하고자 한다면 표준오차의 크기를 줄여야 함
- 표준오차의 크기를 줄이기 위해서는 표본의 크기(n)를 늘려야 함
 - 적절한 표본의 크기(n)를 결정하기 위해서는 신뢰수준($1-\alpha$)과 최대허용오차(e)의 크기에 대한 정보가 필요함
- 대부분의 경우 최대허용오차(e)와 신뢰수준($1-\alpha$)은 특정한 공식이나 이론에 의해서 결정된다고 보기보다는 조사자가 조사목적이나 상황 그리고 과거경험 등을 고려하여 적절하게 결정함

- 표본의 크기를 구하는 공식은 다음과 같이 최대허용오차(e)의 크기와 신뢰구간의 절반이 동일하다는 등식을 표본의 크기(n)에 대하여 정리하면 산출됨

$$n = \frac{\left[t\left(\frac{\alpha}{2}, n-1\right) \right]^2}{e^2} = \frac{\left[Z\left(\frac{\alpha}{2}\right) \right]^2 S^2}{e^2}$$

- 공식에서 모집단의 분산(σ^2)이나 표본의 분산(s^2)이 필요하지만 미리 알 수 는 없기 때문에 과거의 경험치를 이용하거나 모집단이 정규분포를 한다는 가정하에 모집단으로부터 소수의 표본을 추출
 - 그 최대값과 최소값의 차이를 6개의 표준편차(6σ)로 간주하여, 최대값과 최소값의 차이를 6으로 나눈 값을 표준편차로 사용하는 방법이 있음

확률분포 / 9주차 2차시

1. 가설

1) 가설의 정의

두 개 이상의 변수 또는 현상 간의 관계를 검정 가능한 형태로 서술한 문장으로 과학적 조사에 의하여 검정이 가능한 사실

- 연구자가 새로이 주장하여 검정하고자 하는 연구가설(대립가설)과 이에 반하여 기존에 일반적인 사실로 받아들여지고 있는 내용을 나타내는 귀무가설이 있음
- 귀무가설은 H_0 로 표기하고, 연구가설은 H_1 혹은 H_2 표시하며 통상적으로 연구가설이 먼저 설정되면, 귀무가설은 연구가설에서 주장하는 내용과 반대되는 모든 내용을 포함하는 것이 됨
- 귀무가설을 기각하고 연구가설이 채택되기 위해서는 기존의 지식을 나타내는 귀무가설이 분명하게 잘못되었음을 입증하여야 함

2. 가설검정

1) 가설검정의 기본개념

■ 가설검정이란?

- 표본을 추출하여 얻은 표본통계량으로 모집단의 모수에 대한 새로운 예상과 가설 등이 옳다고 판단할 수 있는지를 평가하는 것

예) 학생의 평균키가 168cm로 알려져 있음. 그러나 연구자가 생각하기엔 요즘은 학생들의 영양상태도 좋아지고 어릴 때부터 발육도 빠르고 해서 'K대학교 학생의 평균키가 168cm보다 클 것이다'라고 생각하게 되었음.

- 이 경우 연구자가 입증하고자 하는 연구가설은 'K대학교 학생의 평균키는 168cm보다 크다'이며, 귀무가설은 'K대학교 학생의 평균키는 168cm보다 작거나 같다' 혹은 'K대학교 학생의 평균키는 168cm 이다'가 됨

■ 가설검정

- 모집단에 대해서 알려진 정보나 상식을 나타내는 귀무가설이 옳다는 전제하에 이루어짐
 - 표본조사 결과로부터 얻은 통계량 값이 귀무가설이 옳다는 전제하에서 쉽게 얻을 수 없는 매우 극단적이고 예외적인 값이 나오면, 무언가 귀무가설이 주장하는 바가 잘못되었음을 짐작할 수 있고 이러한 경우야 비로소 귀무가설을 기각하고 연구가설을 채택함

■ 귀무가설의 기각과 채택

- 모집단의 모수가 옳다는 전제하에 표본통계량 값이 나올 가능성, 즉 확률을 가지고 판단하여야 하며 표본의 크기(n)에 따라서도 달라지기 때문에 확률과 표본의 크기(n)를 동시에 고려한 확률값을 가지고 판단해야 함

■ 가설검정에서 사용되는 용어

구분	내용
유의수준(α)	- 귀무가설이 옳다는 전제하에서 어느 정도 극단적인 표본 통계량 값이 나오면 귀무가설이 잘못되었다고 판단하여 귀무가설을 기각할 최대한의 확률
임계치	- 검정의 종류(양측, 단측)와 유의수준(α)을 고려해서 산출한 값으로 가설의 채택 여부를 결정짓는 경계값
검정통계량	- 표본으로부터 추출한 통계량이나 검정에 사용할 분포에 따라 그에 맞는 값으로 치환한 통계량
p값(p-value)	- 표본으로부터 얻은 통계량 혹은 이를 치환한 검정통계량의 절대값보다 더 큰 절대값을 또다른 표본으로부터 얻을 수 있는 확률(컴퓨터 패키지는 항상 양측검정을 기준으로 p값을 산출함)

2) 가설검정의 절차

■ 가설설정

- 가설검정의 절차는 먼저 연구자가 관심을 가지고 있는 일이나 사건 등에 대해 문제제기를 하면서 시작됨
 - 문제가 파악되면 연구자는 이를 바탕으로 새로운 입증하고자 하는 내용을 연구가설로 설정하고 반대가 되는 나머지 내용을 포함하는 연구가설의 여집합을 중심으로 귀무가설을 설정함
 - 가설 설정하면서 중요한 것은 양측검정을 하기 위한 가설로 할 것인지 단측검정을 위한 가설로 할 것인지를 결정하는 일
- 양측검정 vs 단측검정
 - 양측검정
 - ☞ 귀무가설을 기각하고 연구가설을 채택할 수 있는 영역이 양쪽에 있는 경우
 - ☞ 통계량으로 얻은 값이 예외적으로 크거나 작아서 귀무가설에서 주장하는
 - ☞ 값과 통계적으로 유의하게 다르면 귀무가설을 기각하고 연구가설을 채택하는 식의 검정
 - ☞ '모수는 어느 특정한 값인 a가 아니다'가 되고 귀무가설은 '모수는 특정한 값인 a이다'라는 식으로 표현
 - 예) 연구가설을 'B회사의 통조림의 무게는 400g이 아니다'로 한 경우

- 단측검정

- ☞ 연구가설을 채택할 수 있는 영역이 한쪽에만 있는 경우
- ☞ 즉 '다르다(\neq)'가 아니라 '크다($>$)' 나 '작다($<$)'와 같이 한쪽만을 검정
- ☞ 연구가설이 주장하는 방향으로 매우 극단적인 임계치 밖의 통계량 값을 표본으로부터 얻게 되면 연구가설을 채택하고 귀무가설을 기각
- ☞ 연구가설에서 모수가 특정한 값보다 크다고 주장하면 표본통계량값이 임계치보다 더 큰지를 판단하여 검정(오른쪽꼬리검정)하고, 반대로 모수가 특정한 값보다 작다고 주장하면 표본통계량값이 작은 쪽의 임계치보다 더 작은지를 판단하여 검정(왼쪽꼬리검정)함
- 예) 연구가설을 'B회사의 통조림의 무게는 400g이 안 된다.
즉, 400g보다 작다'로 한 경우

- 유의수준 결정

- 가설검정에서 귀무가설을 기각시키고 연구가설을 채택할 확률을 유의수준이라고 하며, α 로 표시함
- 유의수준(α)은 표본으로부터 구한 통계량 값이 귀무가설이 옳다는 전제하에서 어느 정도로 얻기 힘든 극단적인 값인가를 나타내는 확률수준
- 예) 유의수준(α)을 0.05로 하여 가설검정한다는 것은 모집단으로부터 추출한 표본통계량값이 귀무가설이 옳다는 전제하에서는 매우 얻기 힘든, 즉 이러한 통계량값을 얻게 될 확률이 5% 미만인 극단적인 값일 경우에만 귀무가설이 잘못되었다고 판단하겠다는 것을 의미함

- 임계치 산출

- 양측검정이나 단측검정과 같은 검정의 종류와 유의수준(α)을 고려하여 가설의 채택여부 판단에 기준이 되는 임계치를 산출
- 양측검정인 경우
 - 유의수준 α 를 반으로 나누어 분포의 좌우꼬리로부터 $\alpha/2$ 에 해당하는 확률을 얻을 수 있는 경계값들인 2개의 임계치를 산출함
- 단측검정인 경우
 - 연구가설에 따라 분포의 좌우 한쪽으로부터 전체 유의수준 α 에 해당하는 확률을 할당하게 함으로써 임계치는 연구가설에서 주장하는 한쪽에만 설정됨
- 연구가설의 채택여부는 임계치와 표본으로부터 얻은 통계량값을 비교하여 판단함
- 임계치와 통계량을 같은 단위로 통일하여 서로 쉽게 비교할 수 있도록 해야 함

- 검정방법

평균에 대한 가설검정	분산에 대한 가설검정
- 표준편차를 단위로 하는 t분포상의 t값이 임계치로 사용	- 카이제곱분포(χ^2)나 F분포상의 카이제곱값이나 F값이 임계치로 사용
- 표본통계량을 그대로 사용하여 검정하기 보다는 이를 표준화하여 t통계량값으로 변환시킨 다음 이 값을 t분포상의 임계치와 비교하여 검정	- 표본으로부터 얻은 통계량을 카이제곱(χ^2) 통계량 값이나 F통계량 값으로 치환한 다음 이를 임계치와 비교하여 검정

■ 가설의 채택여부 결정

- 검정의 종류(양측, 단측)와 유의수준(α)에 따라 임계치가 결정되면, 이를 표본으로부터 얻은 통계량값과 비교하여 채택할 가설과 기각할 가설을 결정함
- 검정통계량 t값 혹은 p값(p-value)을 이용한 가설검정

검정통계량 t값을 이용한 가설검정에서 연구가설을 채택하는 경우		p값(p-value)을 이용한 가설검정에서 연구가설을 채택하는 경우	
양측검정	$ \text{검정통계량 } t\text{값} > \text{임계치 } t\text{값} $	양측검정	$p\text{값}(p\text{-value}) < \text{유의수준}(\alpha)$
단측검정	오른쪽 꼬리검정 $\text{검정통계량 } t\text{값} > \text{임계치 } t\text{값}$	단측검정	통계패키지에서 얻은 p값의 $1/2 < \text{유의수준}(\alpha)$
	왼쪽 꼬리검정 $\text{검정통계량 } t\text{값} < \text{임계치 } t\text{값}$		

- 가설검정의 4단계

① 1단계 : 가설점정

- 연구가설 설정 입증하고자 하는 내용
- 귀무가설 설정 입증하고자 하는 내용의 여집합 혹은 등식
- 양측, 단측검정 결정

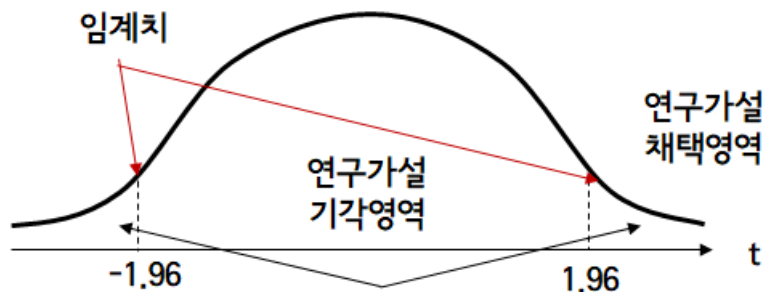


② 2단계 : 유의수준 결정

- 연구자의 목적과 중요성을 고려하여 유의수준을 결정함
- 유의수준이 작을수록 보다 엄격한 기준으로 수행된 가설검정임
- 일반적으로 유의수준은 0.05나 0.1 미만으로 설정함

③ 3단계 : 임계치 산출

- 유의수준 $\alpha=0.05$
- ☞ 양측검정의 경우



- 임계치의 확률이 좌우 각각 $\alpha/2(=0.025)$ 씩임

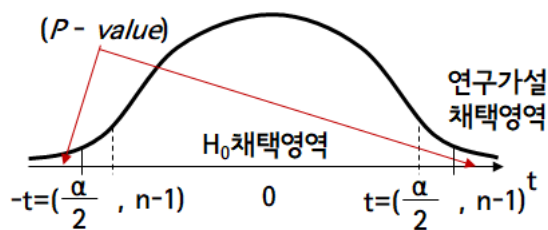
④ 4단계 : 가설채택 여부 결정

- 통계량값 혹은 통계량을 치환한 검정통계량값을 구한 다음 임계치와 비교하여 연구가설의 채택여부를 결정함
- ☞ 통계량을 t 값으로 치환하여 검정통계량 t 값을 구함

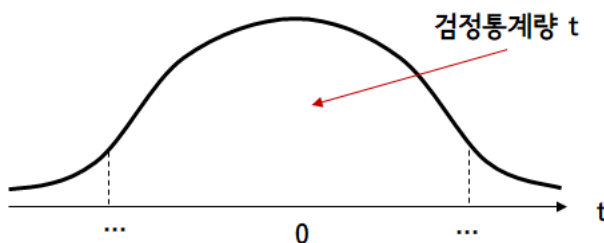
$$t = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

- P값(P - value)

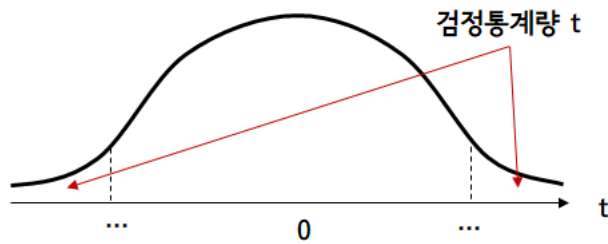
- ☞ 양측검정의 경우 현재의(검정) 통계량 값보다 더 극단적인 값을 다른 표본으로부터 얻을 수 있는 확률, 양쪽 끝 두 면적의 합



- ☞ 연구가설 기각

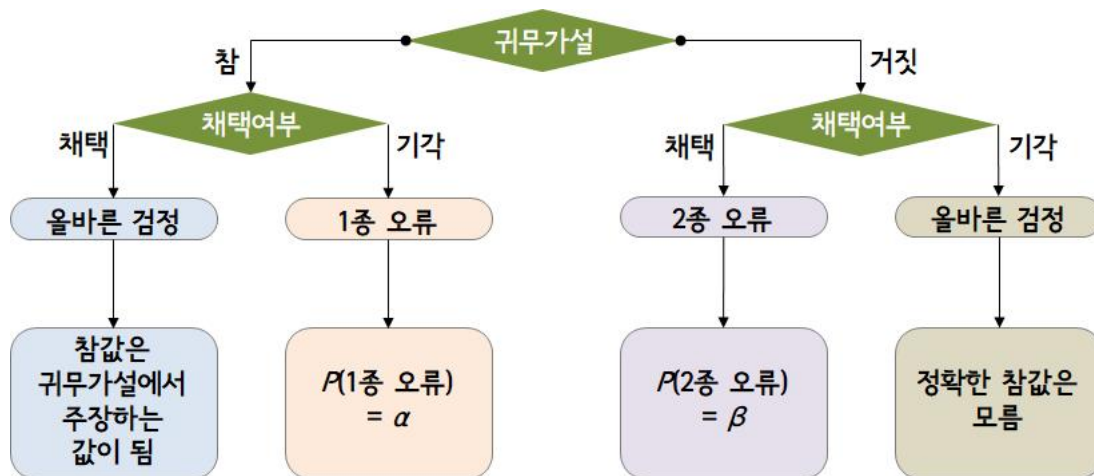


연구가설 채택



3) 가설검정의 오류

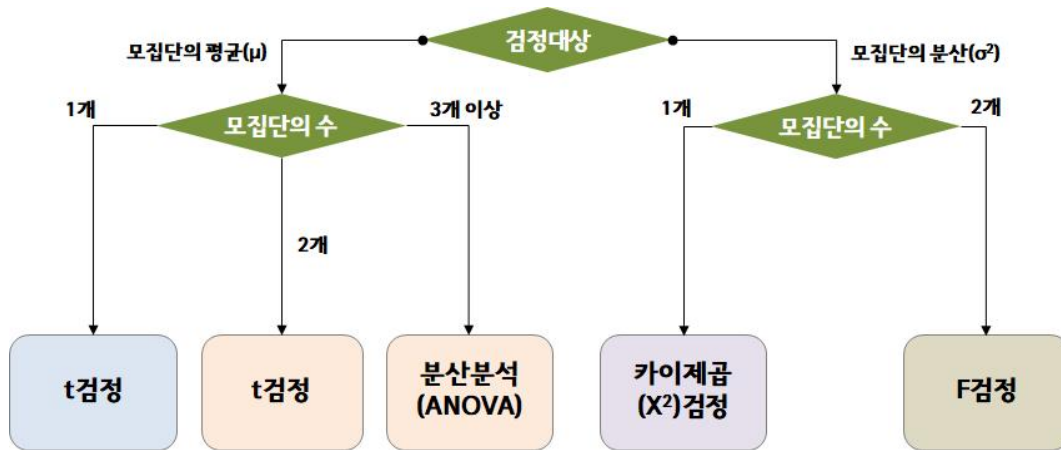
■ 가설검정의 오류유형



1종 오류	2종 오류
<ul style="list-style-type: none"> - 실제로는 귀무가설이 옳음에도 불구하고 매우 예외적인 표본이 추출되어 귀무가설을 기각하고 연구가설을 잘못 채택함으로써 발생하는 오류로 기호로는 α로 표시 예) 실제로 K대학교 학생의 평균키는 168cm임에도 불구하고, 표본추출과정에서 잘못되어 키가 큰 학생들이나 혹은 여학생들만을 표본으로 추출함으로써 K대학교 학생의 평균키가 168cm라는 귀무가설을 기각하는 경우 - 허용오차수준 : 귀무가설이 옳음에도 불구하고 귀무가설을 기각하는 오류를 범하게 되는 것을 허용하는 수준 	<ul style="list-style-type: none"> - 연구가설이 옳고 귀무가설이 틀림에도 불구하고, 귀무가설을 기각하지 못하고 귀무가설을 채택하는 오류로 기호로는 β로 표시 - 실제 K대학교의 평균키가 168cm가 아닌대도 불구하고, 검정결과 168cm라고 주장하는 귀무가설의 주장을 기각하지 못하는 경우 - 검정력 : 실제로 연구가설이 옳은 경우 잘못된 귀무가설을 기각할 수 있는 확률로 전체 확률값 1에서 2종 오류를 확률을 뺀 값($1-\beta$)

3. 대표적인 가설검정의 종류

1) 가설검정



모집단 평균에 대한 가설검정 / 10 주차 1차시

1. 단일 모집단평균에 대한 가설검정

■ 평균에 대한 가설검정

모집단의 수	1~2개	3개 이상
가설검정	t검정	분산분석(ANOVA)

■ 단일 모집단평균에 대한 가설검정

- 모집단의 평균이 어떠하다는 가설이 맞는지를 검정하는 것
 - 검정을 위해서는 모집단으로부터 표본을 추출하여 그 표본의 평균 즉, 통계량값을 가설에서 주장하고 있는 값과 비교하여 검정함

■ 2개 모집단평균에 대한 가설검정

- 2개 모집단의 평균이 서로 다른지 혹은 어느 것이 더 큰지에 대하여 가설을 검정하는 것
 - 모집단에서 각각 뽑은 표본의 평균값들을 서로 비교하여 가설을 검정함

■ 쌍체비교

- 동일한 표본을 대상으로 실험을 실시하기 전에 측정한 값과 실험 후에 측정한 값과의 차이가 어떠한가를 분석하여 검정하는 것

예) 다이어트 프로그램의 효과를 파악하고자 하는 경우 A, B, C, D라는 사람을 표본으로 뽑아서 몸무게를 측정하고 다이어트 프로그램을 수행한 후에 이들 A, B, C, D의 몸무게를 다시 측정하여 각 표본의 몸무게 변동값들을 분석하여 다이어트 프로그램의 효과를 검정하는 것

1) 가설과 검정의 종류

- 단일 모집단 평균에 대한 가설검정은 모집단의 평균(μ)은 어느 특정한 값(a)이라는 믿음에 대해 연구자가 이의를 제기함으로써 시작됨
 - 연구자가 주장하는 연구가설은 ' $\mu \neq a$ '라고 설정되며, 이에 대응되는 귀무가설은 ' $\mu = a$ '가 됨

예) 종로제과점에서는 아이스크림을 생산하고 있는데 그 무게가 200g이 되도록 만들어 왔음. 그러나 이 아이스크림의 무게가 200g이 아닌 것 같다는 소비자들의 불만이 있었음

연구가설	귀무가설	검정종류
종로제과점에서 만드는 아이스크림은 200g이 아니다($\mu \neq a$).	종로제과점에서 만드는 아이스크림은 200g이다($\mu = a$).	표본조사 결과 표본평균(\bar{x})이 200g보다 현저히 작거나 큰 경우 귀무가설을 기각할 수 있는 양측검정
종로제과점에서 만드는 아이스크림은 200g보다 작다($\mu < a$).	종로제과점에서 만드는 아이스크림은 200g보다 크거나 같다($\mu \geq a$).	표본통계량(\bar{x})이 200g보다 현저하게 작아야만 귀무가설을 기각할 수 있는 왼쪽 꼬리검정
종로제과점에서 만드는 아이스크림은 200g보다 크다($\mu > a$).	종로제과점에서 만드는 아이스크림은 200g보다 작거나 같다($\mu \leq a$).	표본통계량(\bar{x})이 200g보다 현저하게 커야만 귀무가설을 기각할 수 있는 오른쪽 꼬리검정

2) 유의수준

- 가설검정의 통계적 유의수준(α), 즉 귀무가설을 기각하고 연구가설을 채택할 확률을 정함
- 유의수준(α)이 결정되면 다음으로 검정의 종류(양측, 왼쪽꼬리, 오른쪽 꼬리)에 따른 임계치를 산출하고, 이를 기준으로 표본조사를 통해 얻은 통계량 값이 어디에 위치하는가에 따라 가설의 채택여부를 검정함
- 유의수준(α)이 작을수록 검정통계량 값도 더욱 작게 설정된 임계치보다 더 크거나 더 작아야 귀무가설을 기각할 수 있으므로 작은 유의수준(α)의 검정에서 채택된 연구가설일수록 그 결과를 신뢰할 수 있음

3) 검정기준에 따른 검정방법

- 표본으로부터 얻은 통계량을 그대로 검정통계량으로 사용하여 검정
 - 예) 키인 경우 cm, 몸무게인 경우 kg으로 표시된 측정값, 즉 표본을 측정한 단위로 표시된 통계량값을 직접 검정통계량으로 사용하기 때문에 이해하기는 쉬우나, 임계치를 구하는 과정이 복잡하고, 특히 검정통계량 값만으로는 검정 결과를 미리 짐작하기가 어려움
- 중심극한정리에 의하여 표본평균(\bar{x})이 $t(Z)$ 분포한다는 가정하에 측정된 통계량
- 값(\bar{x})을 표준화하여, 즉 $t(Z)$ 값으로 치환한 값을 검정통계량으로 하여 검정하는 방법
 - 표본을 측정한 단위(예를 들어, cm 또는 kg)를 표준편차의 단위(σ)로 바꾼 다음 $t(Z)$ 분포를 이용하여 검정하는 것
 - 가장 큰 장점
 - 대부분의 경우 일반적으로 사용되는 유의수준(α)에 해당하는 임계치 t 값을 잘 알고 있기 때문에 치환된 t 통계량 값만으로도 연구가설의 채택 여부를 판단할 수 있다는 점
 - 예) 유의수준(α)을 0.05로 하는 양측검정에서의 임계치는 ± 1.96 정도가 됨으로써 만일 표본통계량 t 값이 +1.96 보다 크거나 -1.96보다 작으면 귀무가설을 기각할 수 있음

- 표본통계량값에 상응하는 p값(p-value)을 구하여, p값이 유의수준(α)보다 작을 경우에는 연구가설을 채택하고 반대로 크면 기각하는 식으로 검정하는 방법
 - p값을 이용한 검정의 경우에도 통계량 p값을 구하기 위해서는 일반적으로 통계량 t값을 산출해야 하기 때문에 두 번째 방법과 유사한 방법임
 - p값은 현재의 표본으로부터 구한 통계량값보다 절대값으로 비교하여 더 큰통계량 값을 갖는 표본을 추출할 수 있는 확률임
 - p값을 이용하여 검정하는 경우에는 p값이 유의수준(α)보다 크면 연구가설이 기각되고 반대로 작으면 연구가설이 채택됨
 - 즉, p값이 작을수록 귀무가설을 기각하고 연구가설을 채택할 가능성이 높아짐

4) 가설검정

- 표본평균(\bar{x})을 검정통계량으로 사용하는 검정 - 양측검정인 경우, 연구가설은 $\mu \neq a$ 가 되고, 귀무가설은 $\mu = a$ 가 됨
 - 주어진 유의수준(α)을 이용하여 표본평균(\bar{x})과 같은 단위의 값으로 표현되는 임계치를 구함
 - 조사한 표본평균(\bar{x})이 좌우 임계치 밖에 위치하면 연구가설을 채택하고 안쪽에 위치하면 연구가설을 기각함 - 임계치를 구하는 방법
 - 유의수준(α)과 표본평균의 표준편차인 표준오차($\bar{x} \frac{S}{\sqrt{n}}$)에 의해서 결정
- 표본평균(\bar{x})을 표준화한 t값을 검정통계량으로 사용하는 검정
 - t값을 구하는 방법
 - 표본평균(\bar{x})에서 모집단평균(μ)을 빼서 얻은 값을 다시 표본평균의 표준편차인 표준오차($\bar{x} \frac{S}{\sqrt{n}}$)로 나눔
 - 좌우측 임계치를 나타내는 c_1 과 c_2 과도 표준화하여 t분포상의 값으로 치환하면, t분포상의 좌우측 임계치를 나타내는 t_{c1} 과 t_{c2} 값들을 구할 수 있음
 - 표본평균(\bar{x})을 t값으로 치환하여 얻은 통계량 값 값, 즉 검정통계량 값이 좌우 임계치 밖에 위치하면 연구가설을 채택하고 안쪽에 위치하면 연구가설을 기각함
 - 통계량 t값으로 가설검정을 하는 이유
 - 표본평균(\bar{x})값을 이용하여 가설검정을 하는 경우 표본의 평균(\bar{x})값만으로는 연구가설의 채택 여부를 판단하기 어려우나 이를 표준화한 t통계량값으로도 어느 정도 검정이 가능하기 때문임

■ p값(p - value)을 이용한 가설검정

- p값이 유의수준(α)보다 크면 연구가설을 기각하고, 유의수준(α)보다 작으면 연구가설을 채택함

예) 유의수준(α)이 0.05일 경우, p값이 0.03이라면 연구가설을 채택함

- 통계 패키지를 이용하여 분석하는 경우에는 통계패키지가 자동으로 p값을 계산해서 제시하는데, 항상 양측검정을 전제로 하여 계산한 p값이므로 단측검정에 활용하고자 하는 경우에는 p값의 절반만을 실제 p값으로 생각하여 이를 유의수준(α)과 비교해서 검정해야 함

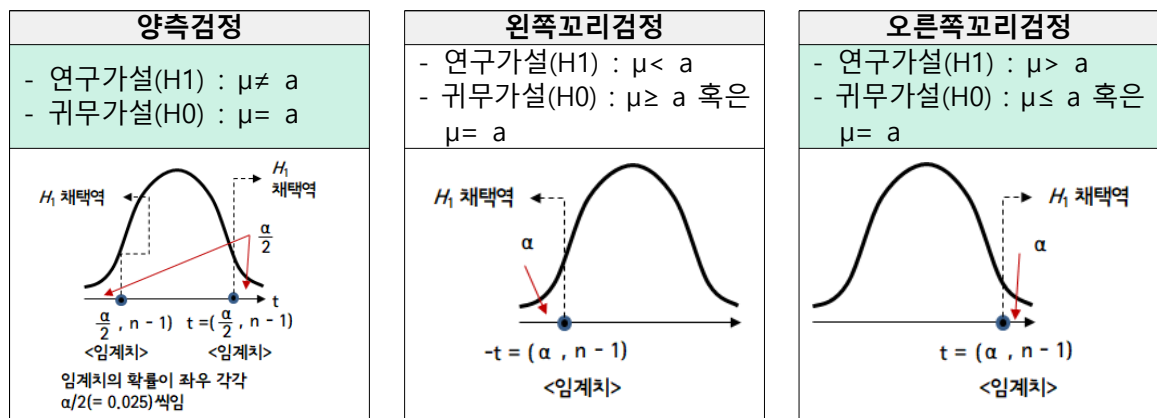
5) 단일 모집단평균에 대한 가설검정 절차

■ 1단계 : 가설검정

- 연구자가 규명하고자 하는 연구가설(H_1 : 대립가설)을 설정함
- 연구가설의 반대 혹은 모집단의 평균이 특정한 값과 같음을 나타내는 등식을 귀무가설(H_0)로 함

■ 2단계 : 유의수준(α)과 검정의 종류에 따른 임계치 산출 및 연구가설의 채택영역 설정

- 설정된 가설을 보고 검정의 종류(양측검정, 왼쪽꼬리 검정, 오른쪽 꼬리검정)를 결정함



■ 3단계 : 검정통계량(t값) 산출

- 표본평균에서 귀무가설의 모평균값을 뺀 차이를 표준오차로 나누어 검정통계량 t값을 구함

■ 4단계 : 임계치와 검정통계량을 비교하여 연구가설의 채택여부를 결정함

- 표본평균의 검정통계량값이 임계치 보다 작거나 크면(좌우 임계치 바깥쪽에 위치) 연구가설 채택함

2. 쌍체비교

1) 의미와 특징

■ 의미

일정한 표본을 대상으로 특정한 사건을 전후로 두 번 측정하여 얻은 값들 간에 유의한 차이가 있는지 검정하는 분석

■ 특징

- 단일모집단 평균검정과 차이점은 귀무가설에서 모집단의 평균이 항상 0을 중심으로 설정됨
- 독립적인 두 모집단평균에 대한 검정보다 효율적임
 - 실험에 참여하지 않은 집단과 실험에 참여한 집단을 서로 독립적인 두 모집단으로 보고 이들을 대상으로 무작위로 표본추출하여 조사한 표본 평균값들 간에 서로 유의한 차이가 검정되지 않는 경우에도 쌍체비교를 하게 되면 유의한 차이가 검정되는 경우가 있음
- 추가적인 노력으로 보다 정확하고 많은 정보를 가지고 가설을 검정하기 때문에 서로 독립적인 두 모집단의 평균을 검정하는 방법보다 더 효과적인 방법임

■ 사례

p대학교에서 운영하는 6개월 영어 학습 프로그램이 있음.
 영어 학습 프로그램의 효과를 측정하기 위한 방법으로는 학습 프로그램에 참여하지 않은 집단과 참여하였던 집단을 대상으로 각각 표본을 추출하여 영어 실력을 테스트하고 성적의 평균값을 비교함으로써 프로그램의 효과를 분석할 수 있음.
 그러나 일정한 표본, 예를 들어 100명의 학생들을 대상으로 영어 학습 프로그램에 참여하기 전에 테스트하여 성적을 기록하고, 영어 학습 프로그램이 끝나고 난 후에 이들 100명에 대하여 다시 테스트하여 그 차이의 크기로 프로그램의 효과를 비교·분석하면 보다 정확하게 학습 프로그램의 효과를 파악할 수 있음.
 즉, 프로그램에 참여하기 전에 측정한 동일한 학생의 실력을 프로그램 참여 후에 다시 측정하여 그 점수 차이를 분석함으로써 영어 학습 프로그램의 효과를 보다 정확하게 측정할 수 있음

2) 절차

- 단일 모집단에서 추출한 하나의 표본을 대상으로 실험 전과 후에 각각 측정하여 그 차이값들의 평균을 단일 모집단의 평균에 대한 가설검정과 같은 방법으로 검정함
- 연구가설은 '차이값의 평균이 0이 아니다($\mu_d \neq 0$)', '차이값이 0보다 크다($\mu_d > 0$)' 혹은 '차이값이 0보다 작다($\mu_d < 0$)'가 됨
- 차이값의 표본평균(\bar{x}_d), 차이값의 표본분산(S_d^2), 검정통계량 t값의 계산에
- 사용되는 표준오차는 차이값의 표본분산인 S_d^2 를 표본의 크기(n)으로 나눈 값에 루트를 씌운 제곱근 값이 됨
 - 단일 모집단평균검정과 동일한 방법으로 검정함

3. 두 모집단평균에 대한 가설검정

1) 의미와 특징

■ 의미

모집단이 하나가 아니고 두 개이며, 이들 모집단들은 서로 연계된 정보가 없어 쌍체비교가 불가능한 경우 실행하는 검정방법임. 즉 각각의 모집단에서 서로 독립적으로 표본을 추출해서 구한 표본평균(\bar{x})들을 서로 비교하여 검정하는 것

■ 가설설정

- 두 모집단 사이에 평균이 다른지 비교 → 양측검정

연구가설	귀무가설
'A반과 B반의 통계학 시험점수는 다르다'($\mu_A \neq \mu_B$)	'A반과 B반의 통계학 시험점수는 같다'($\mu_A = \mu_B$)

- 두 모집단간 평균이 크거나 작은지를 비교 → 단측검정(왼쪽꼬리검정, 오른쪽꼬리검정)

연구가설	귀무가설
'A반의 통계학 시험점수는 B반의 통계학 시험점수보다 높다'($\mu_A > \mu_B$)	'A반의 통계학 시험점수는 B반의 통계학 시험점수보다 낮다'($\mu_A < \mu_B$), 또는 'A반과 B반의 통계학 시험점수는 같다'($\mu_A = \mu_B$)

- 두 모집단평균에 대한 가설검정의 경우에는 구체적인 값이 없이 두 모집단의 평균에 대한 비교만이 가설로 설정됨

2) 가설검정 절차

- 연구가설은 $\mu_A - \mu_B \neq 0$ 이 되고 귀무가설은 $\mu_A - \mu_B = 0$ 이 됨

- 만일 $\mu_A - \mu_B$ 를 하나의 값으로 생각하여 단순히 μ 라 하면, 연구가설은 $\mu \neq 0$ 이 되고, 귀무가설은 $\mu = 0$ 이 됨으로써, 단일 모집단평균에 대한 가설검정과 같아짐 표준오차를 구하는 방법
- 두 모집단의 분산이 서로 다른 경우에는 두 모집단에서 추출한 각각의 표본으로부터 구한 표본분산을 이용하는 방법
- 두 모집단의 분산이 서로 같은 경우에는 과 2개 표본을 합하여 하나의 표본처럼 생각하고 이로부터 통합된 표본분산을 구하는 방법이 있음

- 표준오차가 계산되면, 표준오차의 크기와 유의수준을 고려하여 임계치를 구함

- 두 모집단의 차이를 나타내는 통계량 즉 A반의 표본평균(\bar{x}_A)과 B반의 표본평균(\bar{x}_B)의 차이값(a')을 이용하여 임계치와 비교하여 가설검정을 수행함

모집단 분산에 대한 가설검정 / 10주차 2차시

1. 단일 모집단 분산에 대한 가설검정

■ 분산

- 변수값들의 분포가 퍼져 있는 정도를 나타내는 말로서, 경우에 따라서는 분산이 평균보다 더 중요한 경우도 있음

예) 새로 개발된 수면제의 약효가 지속되는 시간이라든지, 전구의 수명과 같은 경우에는 평균수명도 중요하지만, 평균을 중심으로 각 제품마다 변동이 얼마나 심한지도 중요한 관심의 대상이 됨.

수면제의 평균 약효시간이 제각각이면 안심하고 사용하기가 곤란하여 마찬가지로 전구수명도 수명의 편차가 심하면 제품을 신뢰하고 사용하기가 어려움

■ 모집단분산에 대한 가설검정

- 모집단을 구성하는 값들이 평균으로부터 퍼져 있는 정도를 나타내는 분산에 관심을 가지고 이에 대해 가설을 세우고 이를 검정하는 것

■ 모집단분산에 대한 가설검정의 방식

모집단의 수	1개	2개
검정방식	카이제곱(χ^2) 검정	F검정

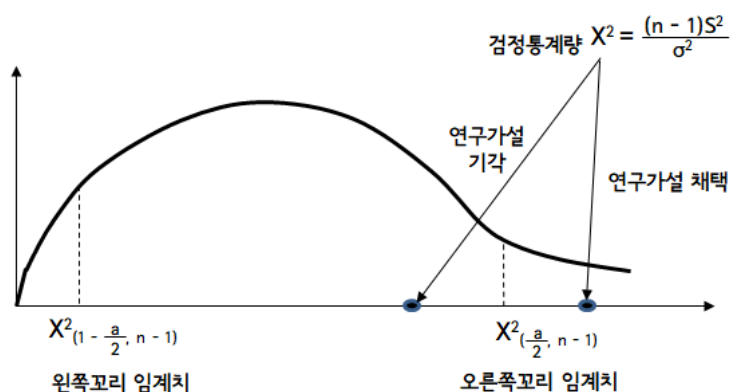
- 단일 모집단분산(σ^2)에 대한 가설검정은 단일 모집단평균(μ)에 대한 가설검정과 유사하며 즉, 평균에 대한 가설검정에서의 평균(μ) 대신에 분산(σ^2)이 사용된다는 점만 다름

연구가설	귀무가설
'모집단의 분산이 어떤 특정한 값(b)와 다르다'($\sigma^2 \neq b$)나 혹은 '특정한 값보다 크다'($\sigma^2 > b$)'나 '작다'($\sigma^2 < b$)'	'모집단의 분산이 어떤 특정한 값(b)이 같다'($\sigma^2 = b$)나 혹은 '특정한 값보다 작거나 같다'($\sigma^2 \leq b$)'나 '크거나 같다'($\sigma^2 \geq b$)'

1) 단일 모집단분산에 대한 가설검정 절차

- 카이제곱(χ^2)분포상의 임계치는 유의수준(α)과 자유도($df=n-1$), 검정의 종류(양측검정, 왼쪽꼬리검정, 오른쪽꼬리검정)에 의하여 결정됨
- 단일 모집단분산에 대한 가설검정에는 카이제곱(χ^2)분포가 사용됨
- 카이제곱 분포는 t분포처럼 자유도(df)에 따라 그 모양이 달라지나 일반적으로 오른쪽꼬리모양의 비대칭분포임

- 항상 양의 값만을 갖는 (χ^2)분포는 t분포와 다르게 좌우대칭이 아니라는 점에 유의하여 검토해야 함
 - 양측검정인 경우
 - 오른쪽 임계치는 $\chi^2_{\left(\frac{\alpha}{2}, n-1\right)}$
 - 왼쪽임계치는 이와 대칭값인 $-\chi^2_{\left(\frac{\alpha}{2}, n-1\right)}$ 이 아니라 $\chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}$ 이 됨
- 카이제곱(χ^2)분포에 있는 확률값은 항상 특정한 χ^2 값으로부터 무한대까지의 면적을 나타냄
 - 임계치를 구하는 경우에는 임계치로부터 오른쪽 무한대까지의 면적을 나타내는 확률값을 고려해서 구해야 함
 - 양측검정인 경우
 - 카이제곱(χ^2)분포의 전체 면적이 1이고 오른쪽 임계치로부터 무한대까지의 면적을 나타내는 확률이 $\frac{\alpha}{2}$
 - 왼쪽 임계치로부터 무한대까지 의 면적을 나타내는 확률은 $1-\frac{\alpha}{2}$ 가 됨
 - 전체에서 $1-\frac{\alpha}{2}$ 의 확률을 제외한 0으로부터 왼쪽임계치까지의 면적으로 표시되는 확률은 $\frac{\alpha}{2}$
 - 왼쪽 임계치는 $\chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}$ 이 됨
- 임계치가 구해지면 표본의 분산(S^2)과 모집단의 분산(σ^2), 자유도(df=n-1)를 이용하여 검점에 사용되는 검정통계량인 χ^2 값은 $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ 을 이용하여 구함



2. 두 모집단 분산에 대한 가설검정

- 2개의 모집단으로부터 각각 표본을 추출하여 이들의 분산을 비교함으로써 모집단 간 분산의 차이를 검정하는 것
- 두 모집단분산에 대한 가설검정은 두 모집단의 평균에 대한 검정과 마찬가지로, 두 모집단의 분산이 같은지 아니면 어느 쪽의 분산이 더 큰지를 검정하는 것
- 연구가설은 '모집단 A와 B의 분산이 다르다($\sigma^2_A \neq \sigma^2_B$)'가 되거나 '어느 한 쪽이 크다($\sigma^2_A > \sigma^2_B$)'는 식으로 표현함
- 검정과정은 두 모집단의 평균에 대한 검정과정과 동일하며 다만 검정에 사용되는 분포와 통계량만 다름

- 두 모집단으로부터 추출된 표본분산의 비를 변수값으로 하는 확률변수를 확률변수 F로 나타내며 이 확률변수 F는 F분포함

예) 모집단 A에서 추출한 표본의 분산을 S^2_A 라 하고, 모집단 B에서 추출된 표본의 분산이

S^2_B 라고 하면, 두 표본분산의 비($\frac{S^2_A}{S^2_B}$ OR $\frac{S^2_B}{S^2_A}$)를 나타내는 통계량, 즉 확률변수 F는

F분포함 → 두 모집단분산에 대한 가설검정은 F분포를 이용함

1) 비대칭 F분포의 특성

- F분포는 모집단으로부터 추출한 표본이 2개이기 때문에 2개 표본의 자유도에 따라 분포의 모양이 변하는 좌우 비대칭 모양의 분포임
- 두 표본분산의 비($\frac{S^2_A}{S^2_B}$ OR $\frac{S^2_B}{S^2_A}$)로 표시되는 확률변수 F의 분포는 분자와 분모에 해당하는 이들 2개 표본분산의 자유도 값에 의하여 분포의 모양이 확정됨
- 어느 표본의 분산이 분자나 분모가 되는가에 따라 통계량값 F값이 달라지며 이에 따른 F분포의 모양도 다르게 나타남
- 분자와 분모에 따라 결정되는 자유도의 순서는 F분포의 모양을 결정짓는 매우 중요한 요인임
- F분포는 카이제곱(χ^2)분포와 마찬가지로 특정한 F값으로부터 무한대에 이르는 면적으로 표시되는 확률값들을 제시함
- 임계치는 두 표본분산의 자유도에 의하여 확정된 F분포와 유의수준(α), 검정종류(양측 또는 단측)에 따라 결정됨
- 좌우 비대칭인 F분포를 이용하여 확률값을 구하거나, 주어진 확률값에 해당하는 F값을 구하는 방식은 카이제곱분포를 이용하여 확률값이나 χ^2 값을 구하는 방법과 같음

2) 두 모집단분산에 대한 가설검정 절차

- 단일 모집단분산에 대한 가설검정에서 사용된 카이제곱(χ^2)분포와는 달리 F분포를 사용함
- F분포는 모집단이 2개이고, 이에 따라 표본도 2개이기 때문에 2개 표본의 자유도(n_A-1, n_B-1)에 따라 그 모양이 달라지나 일반적으로 오른쪽 꼬리가 긴 비대칭모양을 가짐
- F분포상의 임계치는 유의수준(α)과 2개의 자유도(n_A-1, n_B-1), 검정의 종류(양측검정, 왼쪽꼬리검정, 오른쪽꼬리검정)에 의하여 결정됨
- 항상 양의 값만을 갖는 F분포는 t분포와 다르게 χ^2 분포와 같이 오른쪽 꼬리모양의 비대칭분포라는 점에 유의하여 검정해야 함
 - 양측검정인 경우
 - 오른쪽 임계치는 $F\left(\frac{\alpha}{2}, n_A-1, n_B-1\right)$
 - 왼쪽임계치는 이와 대칭값인 $-F\left(-\frac{\alpha}{2}, n_A-1, n_B-1\right)$ 이 아니라 $F\left(\frac{\alpha}{2}, n_B-1, n_A-1\right)$ 이 됨
 - 왼쪽임계치의 확률값이 0.5보다 작은 F값을 이용해서 표현하면 $\frac{1}{F\left(\frac{\alpha}{2}, n_B-1, n_A-1\right)}$ 이 됨
- 일반적으로 F분포에 있는 확률값은 특정한 F값으로부터 무한대까지의 면적을 나타냄
 - 임계치를 구하는 경우에는 임계치로부터 오른쪽 무한대까지의 면적을 나타내는 확률값을 고려해서 구해야 함
 - 양측검정인 경우
 - ☞ F분포의 전체 면적이 1이고 왼쪽 임계치로부터 무한대까지의 면적을 나타내는 확률이 $1 - \frac{\alpha}{2}$ 가 됨
 - ☞ 전체에서 $1 - \frac{\alpha}{2}$ 의 확률을 제외한 0으로부터 왼쪽임계치까지의 면적으로 표시되는 확률은 $\frac{\alpha}{2}$
 - ☞ 왼쪽 임계치는 $F\left(\frac{\alpha}{2}, n_A-1, n_B-1\right)$ 이 됨
 - 임계치가 구해지면 두 모집단으로부터 추출된 표본분산의 비율, 즉 모집단 A에서 추출한 표본분산(S_A^2)와 모집단 B에서 추출한 표본분산(S_B^2)를 이용하여 검정에 사용되는 검정통계량 $F\text{값}\left(\frac{S_A^2}{S_B^2}\right)$ 을 구함
 - 이렇게 구한 검정통계량 F값이 양측검정인 경우 왼쪽 임계치보다 작거나 오른쪽 임계치보다 크면 연구가설을 채택하고, 이들 임계치 사이의 값이면 연구가설을 기각함

타당성 / 11주차 1차시

1. 타당성의 개념

1) 타당성이란?

측정을 위해 개발한 도구를 사용하여 측정하고자 하는 개념이나 속성을 얼마나 정확하게 측정할 수 있는가를 나타내는 지표

- 아무리 정확한 측정대상을 선정했다 하여도 잘못된 측정도구나 측정방법을 사용하여 측정했다면 그 측정값들은 의미가 없음
- 타당성은 측정도구와 측정방법이 측정하고자 하는 것을 측정하기에 타당한가를 의미함
- 측정하려는 개념이나 속성을 정확하게 측정할 수 있는 측정도구와 방법을 사용해서 측정하는 경우 이 측정은 타당성이 높음
- 타당성
 - 사용하고자 하는 측정도구와 방법이 측정하고자 하는 것을 얼마나 정확하게 측정할 수 있는가를 말함
 - 측정하려고 하는 개념을 조작적으로 어떻게 잘 정의하느냐에 관한 측정의 조작적 정의의 개념에 따라서도 상당한 영향을 받음
- 측정하고자 하는 개념이나 속성을 하나의 조작적 정의를 이용해서 표현하고 측정할 수 있음
 - 하나의 조작적 정의에 의한 측정결과만으로는 측정하고자 하는 개념을 어느 정도로 정확하게 측정하였는지 쉽게 판단하기 어렵기 때문에 몇 개의 조작적 정의와 측정을 병행하는 것이 바람직함

2. 타당성의 종류

- 타당성에는 외적타당성과 내적타당성이 있음

외적타당성	내적타당성
측정하여 얻은 실험이나 연구 결과를 실제 상황에 과연 어느 정도까지 적용할 수 있는가	실험이나 연구에서 실제로 측정하고자 하는 것을 얼마나 정확하게 측정하였는가

- 내적타당성과 외적타당성은 트레이드 오프관계임
 - 내적타당성을 높이하고자 하면(모든 외생적 요인들을 완벽하게 통제된 상태에서 실험과 연구실시) 외적타당성은 낮아지고
 - 현실 세계에 적용하기 어려움
 - 외적타당성을 높이하고자 하면(현실세계와 유사한 상황에서 측정) 내적타당성은 낮아짐
 - 모든 외생적 요인을 제대로 통제하지 못한 경우 발생

1) 내적타당성

실험이나 연구에서 다른 외생적 요인에 의한 영향을 배제함으로써 연구자가 연구하고자 하는 내용을 얼마나 정확하게 측정하여 분석할 수 있는가를 나타내는 지표

예) 인과관계를 연구하는 경우, 실험이나 연구에서 측정된 결과가 과연 어느 정도까지 정확하게 원인이 되는 변수의 변화에 의해서 발생한 것인가를 나타내는 지표를 의미함

■ 내적타당성을 높이기 위한 방법

- 가능한 한 원인이 되는 변수의 작용에 의한 효과만을 정확히 추출할 수 있는 조사설계와 외생변수들에 대한 철저한 통제나 제거가 필요함

■ 내적타당성의 유형

- 내용타당성

- 측정을 위해 개발한 도구가 측정하고자 하는 대상의 정확한 속성값을 얼마나 포괄적으로 포함하고 있는가의 정도를 나타내는 타당성

예) S회사가 연말에 개개인의 기여도를 평가하여 성과급을 지급하는 경우, 전반기에 실적이 부진하였으나 후반기 마지막 달에 높은 실적을 보인 직원에서 높은 성과급을 지급하고 전반기에 실적은 좋았으나 후반기 부진한 직원에게 낮은 성과급을 지급한다면 한 해 동안의 성과를 전반적으로 평가하는 데 타당성이 결여됨

- 일반적으로 연구자나 측정자의 주관적인 판단에 의해서 평가

- 기준에 의한 타당성

- 특정한 평가기준에 의한 측정의 효과성을 제시하는 방법으로 이에 대한 대표적인 예로는 예측타당성이 있음
- 측정방법이 미래의 결과를 예측하는 데에 어느 정도 정확한가로 측정의 타당성을 평가
- 입사시험 점수와 입사 후에 평가하게 되는 업무능력 평가 점수와 높은 상관관계가 있어야 함

- 개념타당성

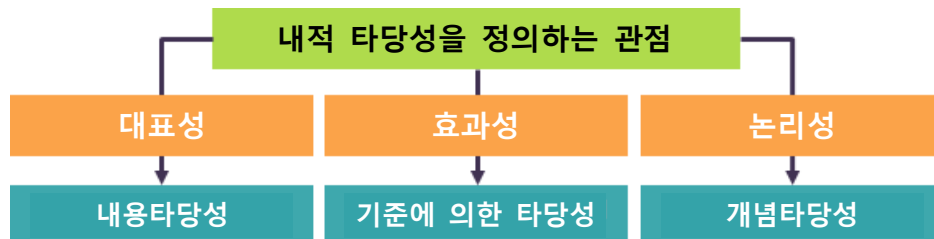
- 내용타당성이나 기준에 의한 타당성으로 설명하기 어려운 감정과 같은 추상적인 개념이나 속성을 측정도구가 얼마나 적절하게 측정하였는가를 나타내는 타당성
- 측정방법이 다른 측정방법들과 비교해서 개념적으로나 논리적으로 얼마나 정확한가를 검증함으로써 파악할 수 있는 타당성

예) 실제로 판매원의 직무만족과 매출 간에 어느 정도 높은 상관관계가 있는 것으로 측정된 경우 사용된 측정방법은 높은 개념타당성을 가짐

• 이해타당성

- ☞ 특정한 개념을 어떻게 이해하고 있는가에 관한 것
- ☞ 특정 개념과 유사한 다른 여러 개념들이 존재하는 경우, 이들 개념들을 모두 측정할 수 있는 측정 방법일수록 이해타당성이 높음

- 집중타당성
 - ☞ 동일한 개념을 서로 상이한 측정방법으로 측정한 결과값들 사이에는 높은 상관관계가 있어야 함
- 판별타당성
 - ☞ 서로 상이한 개념을 동일한 측정방법으로 측정할 경우 결과값들 간에 상관관계가 낮아야 한다는 것



- 내용타당성
 - 측정을 위해 개발한 도구가 측정하고자 하는 대상의 정확한 속성값을 얼마나 포괄적으로 포함하고 있는가를 나타내는 정도
 - 예) 한 과목의 학점은 포괄적인 평가를 바탕으로 산출해야 타당성이 높음
- 기준에 의한 타당성
 - 특정한 평가기준에 의한 측정의 효과성을 제시하는 방법
 - ☞ 예측타당성
 - 미래의 결과를 예측하는 데에 어느 정도 정확한가로 판단
 - 예) 수능점수가 학습능력을 예측하는 지표로 얼마나 정확한가?
- 개념타당성
 - 추상적인 개념이나 속성을 측정도구가 얼마나 적절하게 측정하였는가를 나타내는 타당성
 - ☞ 타당성을 측정하는 방법
 - ★ 이해타당성
 - 서로 유사한 여러 개념들을 모두 측정할 수 있는 측정도구 일수록 이해타당성이 높다고 평가
 - 측정도구의 필요성
 - ★ 집중타당성
 - 동일한 개념을 서로 상이한 측정도구를 이용해서 측정한 결과값들 간의 상관관계가 높을수록 집중타당성이 높다고 평가
 - ★ 판별타당성
 - 서로 상이한 개념을 동일한 측정도구를 사용해서 측정한 결과값들 간에 상관관계가 낮으면 낮을수록 판별타당성이 높다고 평가

2) 외적타당성

실험을 통해 도출된 결과를 이 실험과 유사한 다른 상황이나 현실세계에서도 적용할 수 있는가, 즉 일반화가 가능한가에 관한 타당성

- 마케팅 조사에서는 어느 집단을 표본으로 선정하는가, 즉 어떤 집단을 실험대 상으로 선택하는가는 외적타당성에 큰 영향을 미침
- 외적타당성을 극대화하기 위해서는 실험이나 연구상황을 실제상황과 가능한 한 동일하게 설계하여 실험과 연구를 수행하도록 해야 함
- 외적타당성을 높이려고 하면 상대적으로 내적타당성이 낮아지는 경향이 있으므로
 - 이들 간에 서로 트레이드 오프 관계에 있음

3. 타당성을 높이는 방법

1) 연구분야에 대한 충분한 지식을 습득해야 함

- 주로 잘못된 연구가설의 설정이나 개념에 대한 부정확한 정의, 설문지 작성과정에서의 오류 등으로 인해 다양한 측정오류가 발생함으로써 연구의 타당성이 떨어지게 됨
- 연구의 타당성을 높이기 위해서는 이러한 문제로 발생할 수 있는 오류를 줄일 수 있는 방법을 고안해야 함
 - 이론적 지식의 부족으로 발생

2) 가능하면 선행연구에서 이미 타당성이 입증된 측정방법을 이용해서 연구해야 함

- 기존의 선행연구를 통해서 어느 정도 타당성이 있는 것으로 인정되는 개념에 대한 정의 및 설문 등을 이용하는 방법

3) 사전조사를 충분히 활용해야 함

- 하나의 추상적인 개념을 여러 항목으로 측정하는 경우
 - 사전조사를 통해서 항목들에 대한 요인분석이나 상관분석을 통하여 상관관계가 낮은 항목들을 제거시킨 후, 상관관계가 높은 변수들만을 하나의 추상적인 개념을 측정한 값으로 활용하는 것이 바람직함
- 본조사 전에 미리 사전조사를 통하여 추상적인 개념을 측정할 항목들을 미리 조정하여 선정함으로써 연구의 측정 타당성을 높임

신뢰성 / 11주차 2차시

1. 신뢰성의 정의

1) 신뢰성의 개념

- 측정하고자 하는 현상이나 대상을 얼마나 일관성 있게 측정하였는가를 나타내는 것으로서 안정성, 일관성, 예측가능성, 정확성이라고도 표현
- 신뢰성은 어떠한 측정대상을 반복적으로 측정하는 경우 동일한 결과값을 얼마나 많이 얻을 수 있는지를 나타내는 지표로서 여러 번 반복 측정해서 얻은 측정결과값들의 분산으로 신뢰성을 판단할 수 있음
 - 예) 농구경기에서 자유투를 실시하는 경우, 성공률이 낮은 선수는 자유투 실행시 신뢰성이 떨어지고, 자유투 성공률이 90%이상 되는 선수는 신뢰성이 높다고 평가하는 것과 같음
- 신뢰성을 측정하는 방법
 - 여러 번 반복해서 측정한 조사결과나 실험결과들이 얼마나 일치하는가를 나타내는 일치성 정도에 따라 신뢰성을 판단하는 방법
 - 동일한 내용에 대해 여러 번 반복해서 조사나 실험을 해서 얻은 결과들 간에 상관관계가 크면 클수록 사용된 측정도구와 측정방법의 신뢰성이 높다고 판단하는 방법
 - 하나의 측정도구나 방법을 이용하여 얻은 측정치가 측정하고자 하는 개념이나 속성을 충분히 대변할 수 있는 진실된 측정치인지를 판단하여 신뢰성을 판단하는 방법
 - 측정한 결과값 전체에서 오차범위를 제외한 설명력 있는 부분만을 측정하여 그 값을 설명력 계수로 나누는 방법
 - 측정한 결과로 얻은 값들의 오차범위로 신뢰성을 파악하는 방법
 - 추정의 경우, 허용오차범위가 넓으면 넓을수록 그에 반응하는 신뢰수준은 반대로 줄어들게 됨
 - 측정의 경우에도 측정오차가 크면 클수록 측정도구와 측정방법의 신뢰성은 떨어지게 됨

2) 신뢰성분석의 이론

- 변수의 측정값
 - 실제값
 - 측정하고자 하는 변수의 값을 정확하게 측정한 값이며 측정하고자 하는 동일 변수에 대해 여러 번 무한대로 반복 측정하여 얻은 값들의 평균값임

- 오차

- 측정하고자 하는 변수의 실제값을 정확하게 측정하지 못함으로써 발생하는 오차
- 측정과정에서 발생할 수 있는 다양한 오차를 의미함.
- 오차는 다른 변수값이나 오차들과 연관되어 있지 않고
☞ 즉 서로 독립적이라고 가정함

■ 신뢰성이란?

- 여러 사람들에게 동일한 항목에 대해 반복해서 측정하는 경우 나타나는 측정값들의 일관성 정도를 의미함
- 단순히 변수의 측정값에서 실제값이 차지하는 비율로 파악하기 보다는 이들 측정값들의 분산으로 정의하는 것이 바람직함

■ 하나의 측정변수에 대한 신뢰도 계수는 측정변수의 총 분산 중에서 실제값들의 분산을 나타내는 진분산이 차지하는 비율을 계산함

■ 진분산

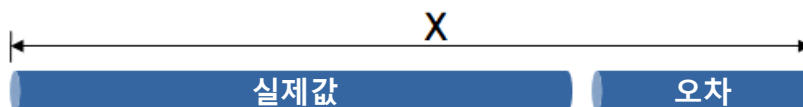
- 측정하고자 하는 값을 서로 병렬적인(parallel) 2개의 측정도구(변수)를 사용해서 측정한 변수값들 간의 상관관계수 값에다가 진분산을 측정하고자 하는 어느 한 변수의 총분산(분산)을 곱한 값이 됨

■ 진분산[$\text{var}(t)$]을 총분산[$\text{var}(X)$]으로 나눈 신뢰도계수(cx)

- 서로 다른 2개의 측정도구로 측정한 변수값들 간의 상관관계수 값이 된다
예) 2개의 서로 다른 질문 문항으로 고객만족도를 측정하였다면, 이들 각 문항으로 측정한 변수값들의 신뢰도는 이들 두 변수 간의 상관관계수값이 됨

① 변수의 측정값

- 변수 측정값(X)=실제값 + 오차



- 실제값
 - ☞ 직접적으로 측정이 불가능한 잠재되어 있는 값 무한 번 반복 측정하는 경우 측정된 값들의 평균값 오차
- 오차
 - ☞ 실제값과 측정값의 차이
 - ☞ 양(+) 혹은 음(-)의 값을 가지며 분포는 0을 중심으로 대칭
 - ☞ 오차의 기대값은 0이며, 오차는 모든 다른 변수들과 독립적임
 - ☞ 오차가 적을수록 신뢰성이 높다고 할 수 있음

② 신뢰도 계수란?

- 신뢰성은 여러 응답자에 의해 반복 측정되었을 경우 측정된 값들의 일관성을 의미함
 - ☞ 측정값의 분산을 이용하여 신뢰성을 평가함
- 신뢰도 계수
 - ☞ 총분산에서 진분산이 차지하는 비율
 - ☞ 신뢰도계수 $CX = \text{진분산}(V_t) / \text{총분산}(V)$
- 변수 X의 값
 - ☞ 실제값(t) + 오차(e)
- 변수 X의 분산
 - ☞ 진분산(V_t) + 오차분산(V_e)

③ 병렬관계에 있는 두 변수의 상관계수

- 신뢰도 계수 값은 병렬관계(Parallel)에 있는 두 측정변수들 간 상관계수값과 동일함
- 병렬관계
 - ☞ 무한 번 측정을 하게 되면 두 변수의 평균과 분산이 서로 일치하게 됨, 이때 평균은 실제값이 됨
 - ☞ $X = t + e, X' = t + e', T = t$ and $\sigma_x^2 = \sigma_x'^2$ 이면, 두 변수는 서로 병렬관계에 있음

2. 신뢰성 측정방법

1) 측정방법

■ 재검사법

- 동일한 측정대상에 대하여 동일한 측정도구를 통해 일정한 시간간격을 두고 반복적으로 측정하여 그 결과값들을 서로 비교 분석하는 방법
- 반복해서 측정한 결과값들 간의 차이가 크면 이때 사용한 측정도구나 방법은 신뢰성이 낮다고 평가하고, 반대로 차이가 작으면 측정도구나 방법이 믿을 수 있어 신뢰성이 높다고 평가함
- 서로 다른 측정도구들을 비교하거나 실제 현상에 적용시키는 데에 매우 용이
- 측정시간의 간격이 크면 클수록 신뢰성이 낮아진다는 한계점을 극복하기 어려울 뿐 아니라, 외생변수의 영향을 파악하기 어려움

■ 반분법

- 측정할 때 사용할 수 있는 도구나 방법들을 2개의 임의의 집단으로 나눈 다음, 이들을 이용하여 측정한 결과가 집단별로 서로 얼마나 다른지를 비교하여 측정도구나 방법의 신뢰도를 분석하는 방법임

예) 창의력을 측정하기 위하여 개발된 10개의 설문문항이 있는 경우,

이들을 임의로 구분하여 5개 문항씩 2개의 집단으로 나눈 다음, 동일한 설문대상자의 응답을 받음

- 집단에 속한 5개 응답들의 평균이 집단 간 차이가 작으면 작을수록 이때 사용한 10개 문항으로 구성된 측정방법은 신뢰성이 높은 것으로 판단함
 - 개별응답자의 집단평균값들 간의 상관분석을 통해 얻은 상관계수값의 크기로 측정방법의 신뢰성을 추정하는 방법
- 집단 평균값들 간의 상관관계를 조사해서 상관계수값의 크기로 측정방법의 신뢰성을 측정하는 방법

■ 내적 일관성

- 동일한 개념을 측정하기 위해 여러 개의 항목을 이용하는 경우에 신뢰성을 측정하는 방법
- 측정도구나 항목을 임의의 2개 집단으로 나누어 신뢰성을 평가하는 반분법과 달리 측정도구나 항목을 개별적으로 취급하여 신뢰성을 파악하는 방법임
- 일반적으로 크론바하 알파계수가 0.6에서 0.7 이상이 되면 측정항목들은 비교적 신뢰성이 높다고 판단함
 - 여러 항목들 중에 알파계수값을 떨어지게 하는 항목은 제외시키고, 알파계수값을 높게 유지시키는 항목들만을 선별해서 사용함으로써 측정도구의 신뢰성을 어느 정도까지 유지하고 조정함

$$\text{크론바흐 알파계수}(\alpha) = \frac{n}{n-1} \left[1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right]$$

n : 문항수, σ_i^2 : 각 문항의 분산, σ_t^2 : 문항합계점수의 분산

- 신뢰성분석에서 가장 많이 쓰이는 측정방법임

3. 신뢰성을 향상시키는 방법

1) 측정항목의 모호성을 제거

- 측정문항을 분명하고 정확하게 작성해야 함
- 문항이 모호하게 표현된 경우, 응답자는 무엇을 의미하는지를 명확하게 파악하지 못한 채 임의로 해석하고 판단하여 잘못된 답을 할 가능성이 높음
- 신뢰성을 높이기 위한 방법
 - 측정항목들을 응답자가 명확하게 이해할 수 있도록 작성하고 가능한 한 전문용어나 어려운 단어는 피하도록 해야 함

2) 측정항목수를 늘림

- 측정항목수가 많으면 많을수록 측정의 신뢰도는 증가하게 됨
- 측정의 경우에도 측정항목의 수가 늘어나면 늘어날수록 측정항목의 평균은 측정하고자 하는 대상의 실제값에 더욱 가까운 값이 되기 때문에 보다 신뢰성이 높은 측정이 가능하게 됨

- 3) 응답자가 모르는 내용은 측정하지 말아야 함
 - 측정항목은 응답자가 알고 있어서 답변할 수 있는 내용으로 작성하는 것이 중요함
 - 응답자들이 모르는 내용을 질문하게 되면 응답자들은 단순히 이미지나 추측만으로 응답할 수 밖에 없음
 - 정확한 측정값을 얻기 어렵고, 이에 따라 측정의 신뢰성이 떨어지게 됨
- 4) 검증된 측정방법을 사용
 - 가능한 한 이미 다른 연구에서 검증된 측정도구나 측정항목을 사용함

4. 타당성과 신뢰성의 차이

- 1) 차이점
 - 타당성은 측정도구나 방법에 관한 내용이고, 신뢰성은 타당성보다는 포괄적이어서 타당성을 포함하는 측정 그 자체나 측정자 혹은 측정과정에 관한 내용임
 - 신뢰성과 타당성을 구분하는 데 중요한 개념은 체계적 오차와 비체계적 오차임

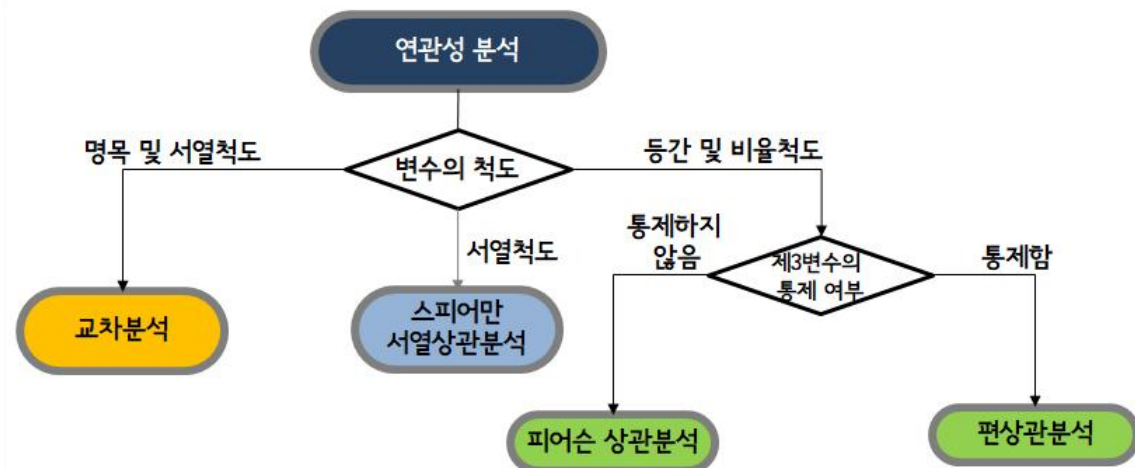
체계적 오차	비체계적 오차
<ul style="list-style-type: none"> - 체계적 오차는 타당성과 관련된 개념으로 어느 대상을 측정해도 항상 동일한 크기와 방향으로 발생하는 오차를 말함 - 측정도구가 잘못되어 측정할 때마다 항상 동일한 크기와 방향으로 발생하는 오차를 체계적 오차라 하며, 이러한 체계적 오차가 발생하는 이유는 측정도구의 타당성이 결여되었기 때문이라고 볼 수 있음 예) 목욕탕 체중계의 바늘이 잘못되어 항상 2kg이 더 나가도록 되어 있다면, 이 체중계로는 어느 누가 몸무게를 측정해도 실제 몸무게보다 항상 2kg이 더 나가는 것으로 측정됨 	<ul style="list-style-type: none"> - 비체계적 오차는 신뢰성과 관련된 개념으로 측정자의 컨디션이나 감정의 변동, 측정자의 능력 등의 차이로 인해 발생하는 무작위적인 오차라 할 수 있음 - 측정도구와 관계없이 측정상황에 따라 발생할 수 있는 오차를 말함 - 측정도구와 관계없이 측정을 수행하는 상황이나 측정자와 관련된 오차를 말함 예) 아무리 좋은 총을 가지고 사격해도 사수의 실력이 부족하면 총알이 과녁을 통과한 점들은 넓게 분산되어 나타남

상관분석 / 12주차 1차시

1. 상관분석의 정의

1) 연관성 분석의 정의와 종류

- 개념: 2개 변수들 간의 연관성을 파악하는 분석방법
- 종류



2) 상관분석의 정의

변수들 간의 연관성을 파악하기 위해 사용하는 분석기법 중의 하나로 변수 간의 선형관계 정도를 분석하는 통계기법

- 상관분석은 두 변수 간의 공분산(Covariance)분석으로부터 시작함
 - 공분산은 동시에 2개의 변수값들을 갖는 개별 관측치들이 각 변수의 평균으로부터 어느 정도 산포되어 있는가를 나타내는 지표임
- 공분산이 척도 크기에 의한 영향을 받지 않도록 각 변수를 그 변수의 표준편차로 나누어 표준화시킨 변수값을 표준화된 공분산이라고 하는데, 이 값이 바로 상관계수(Correlation Coefficient)가 됨

3) 모집단의 상관분석과 표본의 상관분석

- 모집단을 대상으로 하는 상관계수와 표본을 대상으로 하는 상관계수 간에는 큰 차이가 없으며 분산을 구하는 과정에서 사용되는 자유도의 차이만 있음

2. 공분산을 이용한 상관관계분석

1) 공분산의 의미

- 두 변수 간의 공통된 분포를 나타내는 분산을 일반분산과 구분하여 공분산(Covariance)이라 함
- 이러한 공분산은 두 변수 간의 선형적 연관성을 나타내는 대표적인 지표임
- 공분산은 일반분산과 유사한 개념으로 관측치들이 평균으로부터 산포되어 있는 정도를 나타내는 값임
- 일반분산
 - 하나의 변수에 대하여 측정한 관측치가 변수의 평균으로부터 떨어져 있는 정도를 나타내는 편차를 제곱한 값의 평균, 즉 편차제곱의 평균임
- 공분산
 - 하나가 아닌 2개의 변수값을 동시에 갖는 관측치의 각 변수에 대한 평균과의 편차를 구한 다음 이들을 서로 곱한 편차곱의 평균임
 - 예) 통계를 수강하는 60명 학생들의 키와 몸무게를 측정하여 키와 몸무게의 공분산을 구하고자 하는 경우, 공분산은 학생 개인의 몸무게와 반의 평균 몸무게 간의 편차($x_i - \bar{x}$) 개인의 키와 반의 평균 키 간의 편차($y_i - \bar{y}$)를 곱한 값들의 평균이 됨
- 공분산이 양수로서 매우 클 경우에는 두 변수는 양(+)의 선형관계가 있고, 반대로 음수로서 매우 작을 경우 두 변수는 음(-)의 선형관계가 있다고 판단할 수 있음.
즉, 공분산이 클수록 변수 간의 양의 선형관계가 커서, 한 변수가 증가할 때 다른 변수도 같이 증가하고, 반대로 공분산이 작을수록 변수 간의 음의 선형관계가 성립되어 한 변수가 증가하면 다른 변수는 오히려 감소한다는 것을 의미함

2) 측정단위에 따라 변하는 공분산

- 일반적으로 두 변수 간의 선형관계는 한 변수가 또 다른 변수와 연관되어 같이 변화하는 정도를 나타내는 공분산의 크기에 따라 결정됨
- 하지만 공분산의 크기는 두 변수를 측정한 척도크기 즉, 측정단위에 따라 크게 달라짐
 - 예) 키와 몸무게의 공분산을 구하고자 하는 경우, 키의 단위를 m와 cm 중에 어느 것을 단위로 하는가에 따라서 공분산이 달라짐
 - 공분산값만으로 두 변수 간의 연관성에 대하여 단정적으로 평가하는 것은 쉽지 않음
 - 두 변수 간의 선형관계를 측정하되 측정단위에 따라 그 값이 변하지 않는 절대적인 기준이 될 수 있는 단위가 필요함
 - 이러한 절대적인 기준의 단위로는 표준편차가 있음
- 공분산을 구할 때 사용하는 편차를 해당 변수의 표준편차로 나누어 줌으로서 편차를 표준편차 단위로 바꿀 수 있음
 - 편차를 표준편차 단위로 바꾸어 사용함으로써 측정단위에 따른 편차의 크기 변화를 제거할 수 있음

- 표준화된 편차값들을 사용하여 공분산을 구함으로써 측정단위에 따라 크기가 변하지 않는 표준화된 공분산을 구할 수 있음
 - 이러한 과정을 거쳐 구한 표준화된 공분산을 상관계수라고 함

3. 상관계수를 통한 상관관계 파악

1) 상관계수의 의미

두 변수 간의 선형적인 관계 정도와 방향을 수학적 수치로 정량화하여 표시하는 지수로 표준화된 공분산임

- 공분산을 구할 때 사용하는 각 변수의 편차를 그 변수의 표준편차로 나누어 편차를 표준편차 단위로 바꾼 다음 이들을 서로 곱한 값들의 평균을 공분산으로 함
- 측정단위에 따라 변하지 않는 두 변수 간의 선형관계를 파악할 수 있는 표준화된 공분산을 도출할 수 있는 데 상관계수라고 함
- 두 변수의 표준화된 편차곱의 평균으로 구한 상관계수를 피어슨의 적률상관계수 혹은 간단히 피어슨 상관계수라고 함

2) 상관계수의 특징

- 두 변수 간의 선형관계 정도와 방향을 수학적 수치로 정량화하여 표시하는 지수로서 -1에서 +1 사이의 값을 가짐
 - 상관계수가 +1에 혹은 -1일 경우에 두 변수는 완전히 양과 음의 상관관계를 갖는 반면에, 상관계수가 0에 가까울 때 두 변수는 선형이 아닌 다른 관계이거나 서로 독립적이어서 무관한 관계에 있다고 볼 수 있음
- 계수를 구하는 데 사용된 어느 한 변수 또는 두 변수의 값에 일정하게 0이 아닌 상수를 더하거나 양의 상수를 곱하여도 그 값이 변하지 않음
 - 하지만 어느 한 변수에만 음의 상수를 곱하면 상관계수의 크기는 일정하나 방향 즉, 부호가 바뀜
- 상관계수의 절대값 크기는 변수들 간의 선형관계 정도를 나타내는 지표가 됨
 - 음(-)의 상관계수라도 그 값이 -1에 가까우면 두 변수 간의 선형관계가 매우 높다고 할 수 있음

3) 상관계수 해석 및 유의성 검정

- 상관계수의 해석
 - 연구의 성격과 자료의 특성에 따라 상관정도를 평가하는 기준이 달라짐
 - 실제 경제현상 등을 측정한 변수들 간의 상관계수는 비교적 낮고, 설문 등을 이용하여 얻은 변수들 간의 상관계수는 상대적으로 높음

- 상관계수의 절대값과 해석

상관계수(r)의 절대값	r값의 해석
0.2 이하	관계가 거의 없음
0.2~0.4	낮은 상관관계
0.4~0.6	비교적 높은 상관관계
0.6~0.8	높은 상관관계
0.8 이상	매우 높은 상관관계

■ 상관계수의 유의성 검정

- 일반적으로 t분포를 이용하여 상관계수의 통계적 유의성을 검정함
- 표본으로부터 얻은 상관계수가 통계적으로 유의한가를 판단하기 위하여 연구가설과 귀무가설을 설정하여 계수값의 유의성을 검정함
- 가설을 검정하기 위해서는 t통계량값을 사용하며, 이때 t통계량은 자유도가 $n-2$ 인 t분포를 함
- 임계치 및 검정통계량(t값)을 산출하여 임계치와 비교하여 검정통계량(t값)이 연구가설의 채택영역인지 기각영역인지를 확인함
- 연구가설이 채택되면, 상관분석을 통하여 구한 상관계수값이 통계적으로 유의하다고 판단함

4. 서열척도로 측정된 변수들 간의 상관분석

- 일반적으로 상관분석은 등간척도나 비율척도를 이용하여 측정된 변수들 간의 선형관계를 파악하는 데 사용되나 비모수 통계분석을 이용하면 서열척도로 측정된 변수들 간의 연관성도 파악이 가능함
- 서열척도로 측정된 변수 간의 상관관계는 스피어만의 서열상관계수나 켄달의 타우기법으로 산출하는 서열상관계수를 이용함
 - 이 가운데 스피어만 서열상관계수가 많이 사용됨
 - 스피어만의 상관계수
 - 관심대상이 되는 개별 구성원이나 객체들을 두 개의 서로 다른 특성이나 관점에 대해 평가하여 순위를 매긴 순위변수들 간의 선형관계를 나타내는 계수

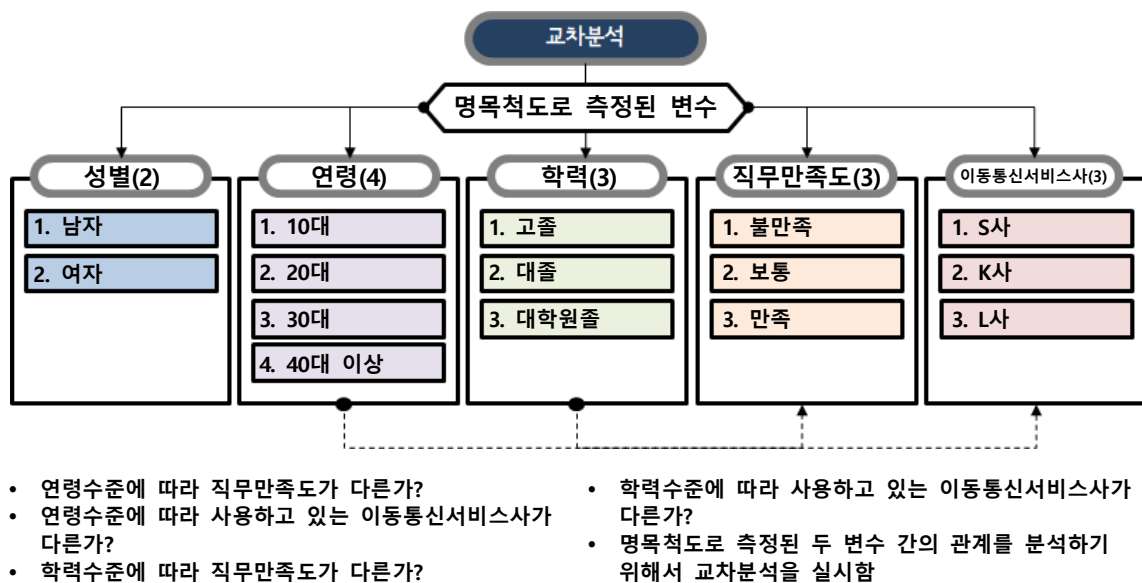
5. 편상관분석

- 두 변수 간의 상관관계를 나타내는 상관계수에서 제3의 변수의 영향으로 인하여 발생할 수 있는 부분을 제거한 상태에서 두 변수 간의 순수한 상관관계를 분석하는 방법을 의미함
- 제3의 변수를 통제하고 분석하는 이유
 - 제3의 변수가 두 변수 간의 관계에 미치는 영향을 통제함으로써 두 변수 간의 순수한, 즉 정확한 상관관계를 파악하기 위해서 함
- 제3의 통제변수를 Z라 할 때, 두 변수 X와 Y 간의 편상관계수는 Z가 X와 Y에 미치는 선형효과를 제거시킨 뒤 남은 잔차 간의 상관계수를 편상관계라 함

교차분석 / 12주차 2차시

1. 교차분석의 정의

- 명목척도나 서열척도로 측정된 범주형 변수(categorical variables)들 간의 연관성을 분석하는 방법
- 예시



- 교차분석
 - 변수들 간의 범주를 교차시켜 얻은 각각의 셀에 해당하는 관측치 빈도를 나타내는 빈도교차표를 이용하여 두 변수 간의 상호독립성이나 관련성 정도를 분석하는 기법임

2. 교차분석의 알고리즘

- 두 범주형 변수의 범주들을 이용하여 빈도교차표를 만들
 - 교차표에서 행을 i 로, 열을 j 로 놓고 i 와 j 로 구분되는 각 셀에 해당하는 관측치의 수 즉, 빈도를 f_{ij} 로 표시하여 빈도교차표를 구함
- 빈도교차표가 완성되면, 이를 이용하여 두 변수가 서로 독립적인지를 분석함
 - 두 변수가 서로 독립적이라는 전제하에서 한계확률($P_{i.}$, $P_{.j}$)을 이용하면 어느 한 관측치가 특정한 셀에 해당될 확률을 구할 수 있는데 이를 결합확률이라고 함
- 두 개의 변수가 서로 독립적이라는 가정하에서 어느 한 관측치가 특정한 셀에 해당될 확률을 나타내는 결합확률을 이용하면 전체 관측치(N)를 각 셀에 배분할 경우 각각의 셀에 기대되는 관측치의 수, 즉 기대빈도를 산출함

- 만일 실제로 측정되어 할당된 관측치의 수(f_{ij})가 두 변수가 서로 독립적이라는 가정하에 기대되는 관측치의 수(E_{ij})와 차이가 많이 날 경우, 두 변수가 서로 독립적이라는 가정이 잘못되었음을 알 수 있음
- 교차분석은 각 셀의 기대빈도(E_{ij})와 실제로 관측된 빈도(f_{ij})간의 차이를 이용하여 두 변수 간의 독립성 여부를 분석하는 것임
- 독립성 여부를 검정하는 것은 두 빈도 간의 차이의 제곱을 기대빈도(E_{ij})로 나눈 값을 각각의 셀에 대하여 구한 다음, 이들을 모두 더한 값을 카이제곱(χ^2) 검정통계량값으로 사용하여 분석함
 - 이러한 검정통계량값은 두 변수의 범주 수에서 각각 1을 뺀 후 서로 곱한 값($(i-1)(j-1)$)을 자유도로 하는 카이제곱(χ^2)분포를 함
 - 검정통계량을 이용하여 가설검정한 결과 그 차이가 통계적으로 유의하게 나오면 두 변수가 서로 독립적이지 못하다, 즉 연관성이 있다고 판단할 수 있음

3. 교차분석 절차

- 1단계
 - 분류기준이 되는 두 변수의 수준을 정하고 이들 두 변수의 수준에 의해 구분되는 각셀에 해당하는 관측치의 수를 조사함으로써 두 개의 범주형 변수들을 이용한 빈도교차표를 완성함
- 2단계
 - 분석하고자 하는 두 변수 간의 관계에 대한 가설을 설정함

연구가설	귀무가설
<ul style="list-style-type: none"> - 행변수와 열변수는 서로 연관되어 있다. - 분류기준이 되는 두 변수는 서로 독립적이지 않다. 	<ul style="list-style-type: none"> - 행변수는 열변수와 서로 연관되어 있지 않다. - 분류기준이 되는 두 변수는 서로 독립적이다.

- 3단계
 - 두 변수가 서로 독립적이라는 가정하에 각 셀의 기대빈도를 구하여 기대빈도교차표를 작성함
- 4단계
 - 빈도교차표상의 실제빈도와 기대빈도의 차이를 제곱한 값을 기대빈도로 나눈 값을 모든 셀에 대하여 구한 다음 이들을 모두 더하여 카이제곱 검정통계량(χ^2)값을 구함
- 5단계
 - 자유도($(i-1)(j-1)$)에 의하여 정의되는 카이제곱(χ^2)분포상에서 검정통계량 χ^2 값과 유의수준(α)에 해당하는 임계치를 비교하여 가설검정을 수행함

4. 교차분석 사례

1) 조사내용

S 이동통신업체에서는 서울시에서 거주지역별로 사용하는 이동통신서비스사가 다른지를 알아보기 위하여 288명의 휴대폰 사용고객을 대상으로 조사하여 자료를 수집하였다. 수집된 자료를 이용하여 거주지역(강북, 강서, 강동, 강남)과 이동통신서비스사(S사, K사, L사) 사이에 연관성이 있는지를 교차분석하였다.

2) 분석과정

- 우선 분석을 위하여 조사한 288명을 거주지역과 사용하고 있는 이동통신서비스사에 따라 구분하여 빈도교차표를 작성함
- 빈도교차표의 행빈도합과 열빈도합을 이용하여, 만일 거주지역과 서비스사를 나타내는 두 변수가 명확하게 서로 독립적이라면 발생할 수 있는 기대빈도교차표를 구함
 - 각 셀의 기대빈도는 행빈도합과 열빈도합을 곱하여 구한 값을 총빈도합으로 나누어 구할 수 있음
 - 예) 거주지역이 강남이면서 S이동통신서비스를 사용하고 있는 고객의 기대빈도는 S사의 서비스를 사용하고 있는 전체 고객의 수(119명)와 강남에 거주하고 있는 전체 고객의 수(63명)를 곱한 다음 이를 전체 응답자수(228명)으로 나누어 구하면 해당 셀의 기대빈도(26명)를 구할 수 있음
- 빈도교차표와 기대빈도교차표를 이용하여 각 셀의 카이제곱값을 계산한 다음 이들을 모두 더하여 검정통계량(χ^2)값을 구함
 - 예) 거주지역이 강남이면서 S이동통신서비스를 사용하고 있는 고객의 실제빈도는 42명이고 기대빈도가 26명일 경우, 실제빈도(42명)와 기대빈도(26명)의 차이(16명)를 제곱한 값을 기대빈도(26명)로 나누면 9.8이라는 카이제곱(χ^2)값을 구할 수 있음
 - 이러한 방법으로 모든 셀의 카이제곱(χ^2)값을 구한 다음 모두 더하면 검정통계량 χ^2 값(27.22)을 구할 수 있음
- 자유도($df=3 \times 2$)에 의해서 결정된 카이제곱(χ^2) 분포상에서 유의수준($\alpha=0.05$)에 해당하는 오른쪽꼬리검정 임계치를 구하면 임계치는 12.60이 됨
 - 임계치와 검정통계량값을 비교하면, 검정통계량 χ^2 값(27.22)이 임계치(12.60)보다 크기 때문에 유의하므로 독립성을 주장하는 귀무가설을 기각하고 연구가설을 채택함
 - 서울시에 거주하는 이동통신서비스 고객의 경우 거주지역과 이용하는 이동통신서비스사는 서로 독립적이지 않다고 판단함

분산분석의 개념 / 13주차 1차시

1. 분산분석의 정의

- 분산분석
 - 3개 이상 집단간 평균의 차이를 검정하는 분석방법
- 집단이 2개인 경우에는 t검정으로 집단간 평균차이를 검정할 수 있으나 3개 이상인 경우 t검정으로 집단간 평균 차이를 한 번에 분석하기는 어려움
 - 집단이 3개 이상인 경우 집단간 평균차이를 동시에 비교 검정할 수 있는 방법
- 독립변수로 구분되는 각각의 집단에 속한 관측치들의 평균, 종속변수값의 평균들이 통계적으로 유의하게 차이가 있는지를 분석하는 것

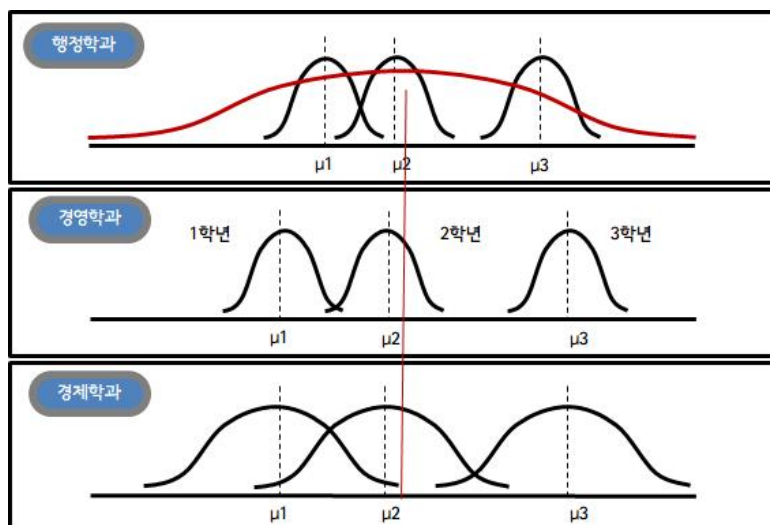
2. 분산분석의 목적

- 1) 집단 간의 평균들을 비교하는 데 왜 분산을 이용하여 분석하고 이를 분산분석이라 하는가?
 - 집단간 평균들이 서로 다르다는 것은 평균들이 서로 멀리 떨어져 퍼져 있는 정도를 나타내는 분산이 매우 크다는 것을 의미함
 - 집단의 평균들이 서로 멀리 떨어져 있어 이들의 분산이 크면 클수록 집단 간의 평균들이 서로 다를 수 있음
 - 집단간 평균들의 분산이 클수록 평균들이 서로 다르기 때문에 집단 평균들의 분산으로 집단간 평균들이 서로 어느 정도 다른지를 파악할 수 있음
- 2) 그러면 집단 평균들의 분산이 어느 정도 커야만 집단간 평균들이 서로 다르다고
 - 판단할 수 있는가?
 - 단순히 숫자로 표시된 분산값만으로는 집단간 평균들이 서로 어느 정도 다른지를 판단하기 어려움
 - 평균값들의 차이를 정확하게 파악하기 위해서는 분산의 크기 정도를 평가할 수 있는 상대적인 기준이 필요함
 - 상대적인 평가기준
 - 각 집단 내에 속해 있는 관측치들이 해당 집단의 평균으로부터 어느 정도 퍼져 있느냐를 나타내는 집단내 분산, 다른 말로 집단내 평균제곱이 사용될 수 있음
 - 집단내 평균제곱의 크기와 집단간 평균의 분산을 나타내는 집단간 평균제곱의 크기를 비교하여 봄으로써 집단간 평균제곱의 크기를 상대적으로 평가할 수 있음
 - 집단내 관측치가 그 집단의 평균으로부터 퍼져 있는 분산, 즉 집단내 평균제곱에 비하여 집단간 평균제곱의 크기가 상대적으로 크면 클수록 집단의 평균들이 넓게 퍼져 있어 서로 명확하게 다르라고 판단할 수 있음

- 평균제곱값의 상대적 크기는 두 값의 비를 이용해서 파악할 수 있으며 이때 사용하는 평균제곱값들을 표본으로부터 구했을 경우 표본분산의 비는 통계량 F값이 됨
- 통계량 F값을 검정통계량으로 하여 집단간 평균의 차이가 통계적으로 유의한지를 검정함
여러 집단간 평균들이 서로 다른 집단간 분산과 집단내 분산들을 이용하여 비교하고 판단하기 때문에 분산분석이라고 함

3) 분산분석의 사례

- 행정학과, 경영학과, 경제학과에서 통계학 시험을 치른 후, 학년별로 학생들을 분류하여 점수를 그래프로 표현함
- 행정학과의 각 학년별 분산은 경영학과와 같으나 학년별 평균은 다르며, 반면에 경영학과와 경제학과는 학년별 평균점수들은 서로 같으나, 학년별 분산은 서로 다름
 - 경영학과는 행정학과와 학년별 분산은 같으나 평균은 다르며, 반면에 경제학과와는 학년별 평균점수는 같으나 분산이 서로 다름
- 각 과의 학년별 평균이 서로 통계적으로 유의하게 다른지 아니면 다르지 않은지를 판단하는 요인을 살펴보면, 행정학과와 경영학과의 경우에 두 학과의 학년내 분산이 모두 동일하므로 학년별 평균 점수의 차이 정도에 따라 통계적 유의성 정도가 달라짐
 - 집단별 평균값의 차이가 크면 클수록, 즉 집단의 평균들이 서로 떨어져 있어 집단간 분산이 크면 클수록 집단간 평균차이가 명확하여 그 차이가 통계적으로 유의하다고 판단할 수 있음
- 반면에 경영학과와 경제학과를 비교해 보면, 두 학과의 학년별 평균점수가 동일하므로 각 학과의 학년별 평균이 통계적으로 유의하게 다른지에 대한 판단은 학년별 점수의 분포인 집단내 분산으로 판단함
 - 즉, 집단내 분산이 클수록 집단별 평균 차이가 통계적으로 유의하게 다를 가능성이 낮아지고, 반대로 집단내 분산이 작을수록 집단별 평균이 통계적으로 유의하게 다를 가능성이 높아지게 됨



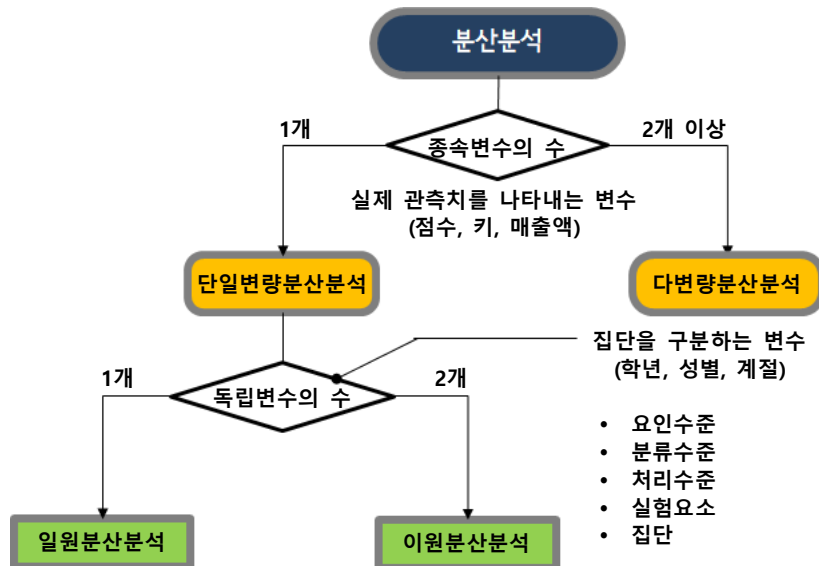
- 집단간 평균의 차이는 같아도, 즉 평균들의 분산은 같아도 각 집단내에 속한 관측치의 분산 정도에 따라 집단간 평균의 차이가 통계적으로 유의할 수 있고 유의하지 않을 수 있음
 - 집단간 평균의 차이에 대한 검정이 단순히 평균들의 차이에 따라 결정되는 것이 아니라, 집단간 평균들의 분산과 집단내 관측치들의 분산을 비교하여 집단 간 평균치를 검정함
 - 즉, 분산값들을 이용해서 집단간 평균차이를 검정하므로 분산분석임

3. 분산분석의 가정

- 분산분석에서 집단을 구분하는 변수, 즉 독립변수를 요인이라 하고, 이러한 요인의 수준 혹은 범주수에 따라 종속변수는 몇 개의 집단으로 분류됨
- 분산분석에서 독립변수 역할을 하는 요인
 - 집단을 구분할 수 있는 명목척도나 서열척도로 측정된 값이어야 하며, 종속변수는 등간척도나 비율척도로 측정된 값이어야 함
- 가정
 - 처리효과 즉 집단을 나누는 독립변수의 개별 수준에 대응하는 모집단은 정규분포해야 하며 서로 동일한 분산을 가져야 함
 - 예) 앞의 사례에서 분산분석을 하고자 하는 학과 내의 학년별 통계학 성적은 정규분포해야 하며 이들의 분산은 서로 같아야 함(표본추출을 하는 모집단의 경우도 동일)
 - 표본으로 추출된 각 집단에 속한 관측치들은 무작위로 추출된 것이며 서로 독립적이어야 함
 - 추출된 관측치들은 서로 연관성이 없어야 함

4. 분산분석의 종류

- 일반적으로 분산분석은 독립변수와 종속변수의 수에 따라 분류됨
- 독립변수가 늘어날수록 종속변수를 구분하는 차원 또한 복잡해지기 때문에 분석결과를 해석하기가 어려워져서 3개 이상의 독립변수를 사용하는 경우는 극히 드물



일원분산분석과 이원분산분석 / 13주차 2차시

1. 일원분산분석의 개념

■ 일원분산분석

- 집단을 구분하는 독립변수가 1개인 경우 집단간 종속변수의 평균차이를 분석하는 방법

■ 관측치가 전체 평균으로부터 떨어져 있는 거리($y_{ij} - \bar{y}$)

- 전체 표본평균(\bar{y})에서 이 관측치가 속해 있는 집단의 표본평균(\bar{y}_i)까지의 거리($\bar{y}_i - \bar{y}$)와 관측치가 속한 집단의 표본평균에서부터 관측치까지의 거리($y_{ij} - \bar{y}_i$)로 구성됨

■ 경우에 따라 좌변이나 우변의 항들이 음의 값을 가질 수도 있으므로 '거리'라는 표현을 적합하지 않음으로 양수만을 나타내는 '거리' 대신에 양수와 음수 모든 값을 나타낼 수 있는 '편차'를 사용

■ 좌변의 총편차

- 우변의 집단간 편차와 집단내 편차의 합으로 구성

$$(y_{ij} - \bar{y}) = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

■ 분산분석

- 하나의 관측지만을 대상으로 하여 이루어지는 것이 아니라 전체 관측치를 대상으로 하기 때문에 분산분석의 대상이 되는 모든 관측치의 편차제곱값을 더하여 전체 편차제곱합을 구해야 함

■ 전체 관측치

- 모두 g 개의 집단으로 구분되고, 각각의 집단(j)에는 n_j 개의 관측치가 있다고 가정할 경우, 양변을 제공한 값을 모든 관측치, 즉 집단(j)에 따라 n_j 개씩의 관측치를 가지고 있는 g 개의 집단을 모두 더하면 다음과 같은 식을 얻을 수 있으며 이 경우 우변을 제공하여 나올 수 있는 나머지 값들은 모두 0이 되어 상쇄됨

■ 총제곱합(SST)

- 집단간 제곱합(SSB)과 집단내 제곱합(SSW)을 더한 값이 됨

$$SST = SSB + SSW$$

■ 분산분석

- 집단간 제곱합(SSB)을 집단내 제곱합(SSW)과 그 크기를 비교함으로써 집단간 평균의 차이가 통계적으로 유의한지를 분석함

■ 제곱합을 그대로 사용하기 보다는 이들이 관측치의 수(n)나 집단의 수(g)에 의하여 받을 수 있는 영향을 제거한 평균제곱들을 구하여 분석함으로써 보다 정확한 결과를 얻을 수 있음

- 각 제곱합을 해당 자유도로 나눈 평균제곱값, 즉 분산을 구한 후 이들 간의 크기를 비교하는 것도 바람직

■ 평균제곱을 구하기 위해서는 자유도를 알아야 함

- 분산분석의 자유도는 다음과 같음

총제곱합의 자유도	집단간 제곱합의 자유도	집단내 제곱합의 자유도
n-1	g-1	n-g
n= 전체 관측치의 수, g = 집단수		

■ 각각의 제곱합을 해당 자유도로 나누면 평균제곱들을 구할 수 있음

- 총평균제곱(MST), 총분산
 - 총제곱합(SST)을 자유도인(n-1)로 나눔
 - 집단 간 평균제곱(MSB)
 - 집단 간 제곱합(SSB)을 자유도인(g-1)로 나눔
 - 집단내 평균제곱(MSW)
 - 집단내 제곱합(SSW)을 자유도인(n-g)로 나눔
- 일원분산분석표 완성

원천	제곱합(SS)	자유도(df)	평균제곱(MS)	F
집단간	$SSB = \sum_y \sum_i (\bar{y} - \bar{y})^2$	(g-1)	$MSB = \frac{SSB}{g-1}$	$\frac{MSB}{MSW}$
집단내	$SSW = \sum_y \sum_i (y_{ij} - \bar{y}_i)^2$	(n-g)	$MSW = \frac{SSW}{n-g}$	
총(합계)	$SST = \sum_y \sum_i (y_{ij} - \bar{y})^2$	(n-1)		

■ 집단간 평균제곱(MSB)을 집단내 평균제곱(MSW)으로 나눈 통계량 F값으로 집단간 평균의 분산이 집단내 관측치들의 분산보다 통계적으로 유의할 만큼 큰 것인지 아닌지를 검정함

- 검정결과 F값의 크기가 통계적으로 유의할 정도로 크면 집단의 평균 간에 유의한 차이가 있다고 판단할 수 있음
- 분산분석에서 가설검증은 '집단간 평균에 차이가 있다'는 연구가설을 전제로함.
귀무가설과 연구가설의 채택과 기각영역을 결정하는 F값은 허용오차(α)와 2개의 자유도에 의해 정의되는 F분포에 의하여 결정됨
- 즉, 2개의 자유도, 분자인 집단간 평균제곱(MSB)의 자유도(g-1)와 분모인 집단내 평균제곱(MSW)의 자유도(n-g)에 따라 그 모양이 결정됨
- 허용오차(α)와 집단의 수(g), 전체 표본의 수(n)에 의하여 임계치가 결정되고, 표본으로부터 얻은 검정통계량 F값이 임계치를 중심으로 어느 쪽에 위치하는가에 따라 연구가설의 채택여부 결정함

■ 검정결과

임계치 < F 검정통계량	임계치 > F 검정통계량
연구가설 채택	연구가설 기각

2. 일원분산분석의 사례

사례내용	■ 학과별로 대학생들의 주량에 차이가 있는지를 분석하기 위하여 3명의 건축공학과 학생과 4명의 영문학과 학생 그리고 5명의 경영학과 학생에게 소주주량을 물어 도표로 정리함					
	■ 학과별로 평균 소주주량에 차이가 있는지 유의수준(α)을 5%로 하여 검정함.					
	응답자	건축공학과		영문학과	경영학과	
	1	0.5		2.0	2.5	
	2	2.0		1.0	3.0	
	3	1.0		0.5	2.0	
	4			1.5	1.0	
	5				3.0	
평균	=1.17		=1.25	=2.30		
일원분산분석표	원천	제곱합(SS)		자유도(df)	평균제곱(MS)	F
	집단간	$\sum_y \sum_i (\bar{y} - \bar{y})^2$ SSB==3.45		g-1=2	$\frac{SSB}{g-1}$ MSB==1.725	$\frac{MSB}{MSW}$ =2.794
	집단내	$\sum_y \sum_i (y_{ij} - \bar{y}_i)^2$ SSW==5.217		n-g=9	$\frac{SSW}{n-g}$ MSW==0.580	
	총(합계)	$\sum_y \sum_i (y_{ij} - \bar{y})^2$ SST==8.67		n-1=11	MST=0.788	
가설검정	연구가설			귀무가설		
	■ ‘세 집단의 평균이 모두 같지는 않다’ ■ ‘적어도 하나는 다르다’			■ ‘세 집단의 평균이 모두 같다’		
	■ 유의수준(α) = 0.05 ■ 임계치: F(α, g-1, n-g) = F(0.05, 2, 9) = 4.26 ■ 검정통계량 F값 = 2.974 ■ 검정통계량(2.974)이 임계치(4.26)보다 작기 때문에 연구가설을 기각하고 귀무가설을 채택함 ■ 학과별로 평균 소주주량의 차이는 5% 유의수준에서 검정해 본 결과 통계적으로 유의하지 않음 ■ 학과별로 평균 소주주량에는 큰 차이가 없다고 판단됨					

3. 주효과 검정만이 가능한 이원분산분석

- 2개의 독립변수에 의하여 구분되는 각 집단(셀)에 오직 1개씩의 관측치만 있는 경우의 분석
- 주효과 검정만이 가능한 이원분산분석의 분산분석표

원천	제곱합(SST)	자유도(df)	평균제곱(MS)	F비
요인 i	SSB _i	c-1	MSB _i	$\frac{MSB_i}{MSW}$
요인 j	SSB _j	g-1	MSB _j	$\frac{MSB_j}{MSW}$
오차	SSE	(c-1)(g-1)	MSE	
총(합계)	SST	cg-1		

- 주효과 검정만이 가능한 이원분산분석에서는 2가지 요인에 대한 가설검정을 별도로 해야 함
 - 요인 i에 의하여 구분된 집단 간에 통계적으로 유의한 평균 차이가 있는지를 이원분산분석에서와 같이 검정함
 - 요인 j에 의하여 구분된 집단 간에 통계적으로 유의한 평균 차이가 있는지를 이원분산분석에서와 같이 검정함
- 가설에 대한 검정
 - 2개의 자유도로 검정되는 F분포상에서 표본으로부터 구한 검정통계량 F값과 유의수준(α)에 의한 임계치 F값과 비교하여 검정함
- 검정통계량 F값이 임계치보다 클 경우 각 요인에 따라 구분되는 집단간 평균이 모두 같지는 않다는 연구가설을 채택함

4. 주효과와 상호작용효과 검정이 가능한 이원분산분석

- 2개의 독립변수로 구분되는 세분집단(셀) 중에 2개 이상의 관측치가 있는 집단(셀)이 있는 경우에만 상호작용효과 검정이 가능함
- 주효과와 상호작용효과 검정이 가능한 이원분산분석의 분산분석표

원천	제곱합(SST)	자유도(df)	평균제곱(MS)	F비
요인 i	SSB_i	$c-1$	MSB_i	$\frac{MSB_i}{MSE}$
요인 j	SSB_j	$g-1$	MSB_j	$\frac{MSB_j}{MSE}$
요인 i와 요인 j의 상호작용	SSB_{ij}	$(c-1)(g-1)$	MSB_{ij}	$\frac{MSB_{ij}}{MSE}$
오차	SSE	$\sum_i^c \sum_j^g (h_{ij} - 1)$	MSE	
총(합계)	SST	$\sum_i^c \sum_j^g h_{ij} - 1$		
c : 요인 i에 의해 구분되는 집단의 수 g : 요인 j에 의해 구분되는 집단의 수 h : 요인 i와 요인 j로 구분되는 집단의 관측치 수				

- 주효과와 상호작용효과에 대한 검정이 가능한 이원분산분석에서는 2가지 요인을 다루고 있으므로 2가지 요인에 대한 검정 이외에 요인 i와 요인 j의 상호작용효과에 대한 검정까지 별도로 해야 함
 - 요인i에 의하여 구분되는 집단간 평균에 차이가 있는지를 일원분산분석에서와 같은 방법으로 검정함
 - 요인j에 의하여 구분되는 집단간 평균에 차이가 있는지를 일원분산분석에서와 같은 방법으로 검정함
 - 요인i와 요인j에 의해 구분되는 집단간 평균 차이에 상호작용효과가 존재하는지를 검정함
- 상호작용효과 검정이 가능한 이원분산분석의 가설검정은
 - '각 요인에 따라 구분되는 집단간 평균에 차이가 있다'와 '요인i와 요인j에 의하여 구분되는 집단간 평균 차이에 상호작용효과가 존재한다'가 연구가설로 설정됨
- 2개의 자유도로 결정되는 F분포상에서 유의수준(α)에 따라 결정되는 임계치 F값과 표본으로부터 구한 검정통계량 F값을 비교하여 검정통계량 F값이 임계치보다 클 경우 연구가설을 채택함

회귀분석의 개념과 단순회귀분석 / 14주차 1차시

1. 회귀와 회귀분석

- 회귀
 - 한 바퀴 돌아서 다시 제자리로 오는 것을 의미함
 - 평균으로 되돌아간다는 것
- 회귀선
 - 세분화된 각 집단의 평균을 통과하는 직선을 그릴 수 있는데 이것이 바로 회귀선임
 - 주어진 X값의 각 점에 대응하는 Y값들의 평균을 관통하는 직선을 의미함
- 회귀분석
 - 회귀선을 도출하기 위한 분석방법이 회귀분석임
 - 변수와 변수간의 관계를 나타내는 회귀선을 도출하여 이를 바탕으로 독립변수 X값이 주어졌을 경우 이에 대응하는 종속변수 Y값을 예측하거나 종속변수에 대한 독립변수의 영향을 설명하는 데에 사용되는 분석방법임
- 독립변수가 종속변수에 미치는 영향력의 크기를 파악하여 독립변수의 특정한 값에 대응하는 종속변수값을 예측하는 선형모형을 산출하는 방법
- 회귀분석을 필요한 문제의 예
 - 매출액은 광고횟수에 따라 어떻게 변하는가?
 - 제조환경을 개선함에 따라 생산량은 어느 정도 증가할 것인가?
 - 담배판매량과 폐암환자수와의 관계는 어떠한가?

2. 회귀분석의 종류

- 회귀분석은 독립변수의 수에 따라 구분
 - 독립변수가 1개이면 단순회귀분석이라 함
 - 2개 이상이면 다중회귀분석이라 함
- 명목척도나 서열척도로 측정된 변수를 독립변수로 사용하고자 하는 경우
 - 이항변수로 표시된 몇 개의 더미변수를 이용하여 분석할 수 있음
 - 이러한 회귀분석을 등간척도나 비율척도로 측정한 변수를 독립변수로 이용하는 일반회귀분석과 구분하여 더미변수를 이용한 회귀분석이라 함
- 독립변수와 종속변수 간의 관계에 따라 선형회귀분석과 비선형회귀분석으로 나누어짐

3. 회귀분석의 전제조건

- 1) 특정한 독립변수값(x)에 해당하는 종속변수값(y)들은 정규분포를 해야 하며, 모든 정규분포의 분산은 동일해야 함
 - 회귀분석을 통하여 신뢰할 수 있는 모형을 도출할 수 있으며, 이때의 회귀선은 모든 정규분포의 중심인 평균을 관통하는 선이 됨
 - 독립변수(x) 값에 따라 결정되는 집단별 종속변수(y_i)들의 평균이 되는 회귀식의 예측값들은 서로 달라도 분산은 모두 같아야 하는 등분산성을 만족시켜야 함
 - 만일 독립변수값(x_i)에 대응하는 종속변수값(y_i)들이 정규분포하지 않고 제 각각 나름대로 분포를 한다면 이러한 자료를 이용한 회귀분석의 결과로 얻게 되는 회귀식에 대하여 신뢰하기 어려움
 - 실제로 대부분의 경우 자료의 정규분포나 등분산성을 사전에 정확하게 파악하고 회귀분석한다는 것은 거의 불가능하기 때문에 이러한 가정은 명목적도처럼 특별한 척도로 측정한 변수가 아니면, 묵시적으로 이러한 전제조건이 성립된다는 가정하에 회귀분석을 실시
- 2) 종속변수값(y_i)들은 통계적으로 서로 독립적이어야 함
 - 일반적으로 회귀분석은 독립변수값(x_i)의 변화에 따른 종속변수값(y_i)의 변화를 분석하는 것임
 - 그러나 독립변수값(x_i)의 변화에 관계없이 종속변수값(y_i)들끼리 서로 영향받아 변한다면 올바른 회귀모형을 도출할 수 없음
- 3) 다중회귀분석의 경우 독립변수들 간에 다중공선성(multicollinearity)이 존재하지 않아야 함

다중공선성
<ul style="list-style-type: none"> ■ 독립변수들 간의 상관관계로 하나의 독립변수가 다른 독립변수에 미치는 영향이 클 경우에 다중공선성이 존재함 ■ 다중공선성이 존재하는 경우 <ul style="list-style-type: none"> ◆ 최소자승법에 의해 추정된 회귀모형의 표준오차값이 커지는 경향이 있음으로 회귀모형이 잘못 추정될 가능성이 높음 ■ 다중공선성이 클수록 추정된 회귀모형의 예측력은 그만큼 감소하게 됨 예) 커피의 가격과 맛이 선호도에 미치는 영향에 대해서 분석을 할 경우, 가격과 맛이 서로 연관성이 높다면 정확한 회귀모형을 도출하기 어려움 ■ 다중공선성을 진단하는 방법 <ul style="list-style-type: none"> ◆ 공차한계를 이용하는 방법 <ul style="list-style-type: none"> ● 공차한계란 여러 개의 독립변수들 중에 하나의 독립변수를 종속변수로 하고 나머지 다른 독립변수들을 독립변수로 한 회귀분석에서 모형의 설명력을 나타내는 결정계수(R^2)값을 구한 다음 이 값을 1에서 뺀 값($1-R^2$)임 ● 공차한계가 커진다는 것은 하나의 독립변수가 다른 독립변수들에 의해 설명되는 변량이 작다는 것을 의미함 ● 공차의 한계가 1에 가까워질수록 독립변수들 간에 다중공선성이 존재하지 않는다고 판단할 수 있음 ◆ 분산확대지수에 의한 방법 <ul style="list-style-type: none"> ● 이 값은 공차한계와 역수관계에 있기 때문에 공차한계가 커서 1에 접근할수록 분산확대지수 또는 1에 근접하게 됨 ● 분산확대지수가 1에 가까워질수록 독립변수들간에 다중공선성이 없는 것으로 판단하며 1을 초과하여 큰 값을 갖게 되면 독립변수들 간에 높은 다중공선성이 존재하는 것으로 판단함

4. 최소자승법을 이용한 회귀분석 모형의 도출

- 단순회귀분석
 - 독립변수가 하나일 경우에 종속변수와의 관계를 분석하여 독립변수가 종속변수에 미치는 영향을 분석
- 두 변수간의 관계를 파악하는 방법
 - 두 변수 간의 관계를 나타내는 산점도(개별 관측치들을 X축과 Y축으로 구성된 좌표 위에 점으로 표현된 그래프)를 그려보고 회귀분석을 통해 선형관계를 파악
→ 회귀선 도출
- 최소자승법
 - 회귀선 도출에 사용되는 가장 대표적인 방법
- 흩어져 있는 개별 관측치들과 산출될 회귀선과의 거리, 다른 말로 잔차(residual)를 최소화시킬 수 있는 회귀선의 식을 도출
- 관측치(y_i) 값과 회귀선(\hat{y}_i)과의 거리인 잔차(ϵ_i)제곱의 합을 최소화하는 직선식을 찾는 방법
- 잔차의 합($\sum_{i=1}^n \epsilon_i$)을 최소화하는 것이 아니고, 잔차제곱의 합($\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$)을 최소화하는 직선식을 구하는 방법임

5. 회귀선의 설명력

- 결정계수
 - 회귀선이 설명력을 나타내는 지수
 - 평균값으로 관측치의 값을 추정하였을 경우에 발생하는 편차를 추정된 회귀선이 어느 정도 줄일 수 있는가?
 - 즉, 전체 편차 중에 회귀선이 설명하여 줄일 수 있는 부분의 비를 말함
- 편차
 - 관측치가 평균으로부터 떨어져 있는 정도, 즉 평균과의 차이를 말함
- 잔차
 - 평균이 아니라 회귀식 등으로 추정된 값과의 차이를 말함
 - 편차의 일부분이라 생각함
- 오차
 - 편차와 달리 예측하기 위하여 추정된 값과 실제값의 차이를 말함
 - 평균을 생각할 필요 없이 단순히 예측값 혹은 기댓값과 실제값의 차이
즉, 예측값이 정확하지 못한 정도를 나타냄
- 총편차
 - 회귀선으로 설명할 수 있는 편차와 회귀선으로 설명할 수 없어 남아 있는 편차
즉, 잔차의 합이 됨

- 총제곱합(SST)
 - 편차와 마찬가지로 회귀선에 의해 설명되지 못한 편차의 제곱합(SSE)과 회귀선에 의해 설명되는 편차의 제곱합(SSR)으로 구성됨
 - $SST = SSE + SSR$
- 회귀식의 설명력은 총제곱합(SST) 중에서 회귀선에 의해 설명되는 제곱합(SSR)이 차지하는 비율을 말함

6. 회귀식 모형의 적합도분석

- 회귀식을 도출하는 데에 사용된 자료가 모집단이 아니고 표본인 경우에 표본으로 추정된 회귀모형이 어느 정도 믿을만한가에 대한 분석과 검증임
- 회귀식의 통계적 유의성 검정을 회귀모형의 적합도분석이라하며 적합도분석의 대표적인 방법

추정값의 표준오차를 이용한 방법	분산분석표를 이용한 방법
<ul style="list-style-type: none"> ■ 추정값의 표준오차 <ul style="list-style-type: none"> ◆ 실제값이 회귀식으로 추정한 값으로부터 떨어져 있는 잔차를 제곱한 값의 평균 즉, 잔차평균제곱의 제곱근값을 의미 ■ 추정값의 표준오차가 작을수록 추정된 회귀식이 적합하다고 판단 ■ 표준오차의 크기를 평가할 수 있는 객관적인 기준이 없기 때문에 표준오차값으로 회귀선의 적합도를 판단하는 것은 쉽지 않음 	<ul style="list-style-type: none"> ■ 객관적으로 도출된 회귀식이 통계적으로 유의한가를 평가하는 방법 ■ 회귀선의 설명력(R^2)이 아무리 높아도 통계적으로 유의하지 않으면 일반화하여 사용하기 어려움 ■ 분산분석에서와 같은 방법으로 회귀식의 통계적 유의성을 검정함

■ 분산분석표의 예

원천	제곱합 (SS)	자유도 (df)	평균제곱	검정통계량 F	유의확률 (p-value)
회귀	1039.026	1	1039.026	48.581	.000
잔차	384.974	18	21.387		
총(합계)	1424.000	19			
회귀 자유도 = $(k + 1) - 1$, 잔차자유도 = $n - (k + 1)$, 총자유도 = $n - 1$, k = 독립변수의 수, n = 관측치					

- 회귀식의 통계적 유의성 검정
 - 검정통계량 F값으로 회귀식의 유의성을 검정함
 - 회귀선이 평균으로부터 떨어져 있는 거리제곱의 평균인 회귀평균제곱과 실제 관측치가 회귀선으로부터 떨어져 있는 잔차제곱의 평균인 잔차평균제곱의 비율로 표시되는 검정통계량 F값으로 회귀식(모형)의 통계적 유의성을 검정함
 - 검정통계량 F값이 임계치 보다 크면, 추정된 회귀식이 통계적으로 유의한 것으로 판단함

다중회귀분석과 비선형회귀분석 / 14주차 2차시

1. 다중회귀분석의 기본개념

- 일반적인 사회현상은 하나의 변수로 설명되거나 예측되는 경우는 거의 없기 때문에 2개 이상의 독립변수를 사용하여 종속변수의 분산을 보다 효과적이고 정확하게 설명하고 예측하는 모형을 개발하는 경우를 고려해 볼 수 있음
 - 2개 이상 복수상의 독립변수들을 이용한 회귀분석
- 회귀식을 도출하고 결정계수(R²)를 구하여, 산출된 회귀식의 통계적 유의성 검정은 모두 단순회귀분석의 경우와 같은 방법에 의해서 이루어짐
- 다중회귀분석에서 도출된 회귀식(모형)의 통계적 유의성 검증
 - 분산분석표를 이용하여 수행됨(분산분석표의 예)

원천	제곱합 (SS)	자유도 (df)	평균 제곱	검정통계량F	유의확률 (p-value)
회귀	1039.026	1	1039.026	48.581	.000
잔차	384.974	18	21.387		
총(합계)	1424.000	19			
회귀 자유도 = (k + 1) - 1, 잔차 자유도 = n - (k + 1), 총자유도 = n - 1, k = 독립변수의 수, n = 관측치					

2. 독립변수의 유의성과 중요성 평가

- 회귀모형에 대한 통계적 유의성 검정이 끝나면 회귀계수를 살펴보아야 함
- 회귀계수표
 - 각의 독립변수에 대한 비표준화 계수, 표준화 계수, t값과 유의확률 p값(p-value)이 나타남

모형	비표준화계수		표준화 계수	t	유의확률 (p-value)
	B	표준오차	베타		
(상수)	122.865	5.424		22.653	.000
가격	.234	.066	.496	3.548	.001
이미지	.390	.166	.471	3.370	.002
회귀식: 만족도 = 122.865 + 0.234X1(가격) + 0.390X2(이미지)					

- t값과 유의확률 p값으로 개별 독립변수의 통계적 유의성을 검정할 수 있음
- 각각의 독립변수에 대한 통계적 유의성 검정이 끝나면 비표준화계수와 표준화계수를 살펴보아야 함

- 비표준화계수와 표준화계수의 차이는 분석을 위한 입력자료의 표준화 여부로 구분됨

표준화계수	비표준화계수
표준화한 자료를 입력자료로 하는 회귀분석에서 얻게 되는 계수	표준화하기 전의 원래 자료를 이용한 회귀분석 결과로 얻게 되는 계수

- 회귀식을 이용해서 종속변수값을 예측하기 위해서는 측정된 독립변수값을 회귀식에 직접 대입하여 종속변수값을 구할 수 있어야 함
 - 실제 예측에 활용할 회귀식은 비표준화계수로 만드는 것이 일반적임
- 통계적으로 유의한 독립변수들만으로 형성된 회귀모형이 결정되면, 여러 개의 독립변수 중에 어느 변수가 종속변수에 더 큰 영향을 미치는가를 파악할 필요가 있음
 - 일반적으로 회귀계수의 크기로 어느 변수가 더 중요한지 파악해야 함
 - 측정단위와 분포의 평균에 의해 영향을 받지 않는 표준화 계수의 크기로 독립변수가 종속변수에 미치는 상대적 영향력을 살펴보아야 함
 - 표준화계수 : 독립변수의 측정 단위와 분포의 평균에 따라 영향받지 않도록 독립변수들을 표준화한 자료로 회귀분석하여 얻게 되는 계수
 - 표준화계수값들의 크기를 서로 비교함으로써 여러 개의 독립변수 중에서 어느 변수가 종속변수에 더 많은 영향을 미치는지를 파악할 수 있음

3. 유용한 독립변수의 선택방법

- 모든 변수들을 독립변수로 사용하여 회귀분석하는 것은 바람직하지 못한 이유
 - 독립변수의 수가 증가함에 따라 자료를 수집하고 관리하는 데에도 보다 많은 시간, 노력, 비용을 필요로 함
 - 독립변수들 간의 연관성, 즉 다중공선성과 같은 문제가 발생할 수도 있음
 - 가장 좋은 회귀모형
 - ☞ 가장 적은 수의 독립변수로 종속변수의 분산을 가장 많이 설명할 수 있는 모형
- 일반적으로 독립변수들로 설명할 수 있는 종속변수 분산의 크기는 종속변수의 수에 비례함
 - 독립변수의 수들로 설명할 수 있는 종속변수의 분산도 증가하지만 추가적으로 사용하게 되는 독립변수가 설명할 수 있는 종속변수의 분산의 양은 점차적으로 감소할 수 밖에 없음
 - 가능하면 가장 유용한 독립변수들만을 선정하여 모형을 개발해야 함
- 유용한 독립변수들을 선정하기 위해서는 여러 개의 독립변수들 중에 어느 변수가 중요하고 설명력이 높은지를 판단할 수 있는 방법이 필요함
 - 유용한 독립변수들을 선택하는 데에 사용할 수 있는 대표적인 방법

선정방식	내용
입력(enter)	■ 분석자가 임의로 지정한 모든 변수들을 독립변수로 사용하여 분석하는 방법
전진선택 (forward selection)	■ 여러 개의 독립변수들 중에서 가장 중요한 변수의 순으로 하나씩 선택해 가는 방법

	<ul style="list-style-type: none"> ■ 선정과정 <ol style="list-style-type: none"> 1) 개별 독립변수 하나씩만을 독립변수로 사용하여 회귀분석 실시 2) F통계량 값이 가장 큰 모형을 찾아 이 모형에서 사용된 회귀분석 결과로 얻게 되는 모형들 중에 통계적 유의성이 가장 높음 즉, 변수를 가장 먼저 선택함 3) 아직 선택되지 않은 변수 하나씩만을 더 포함하는 회귀모형들 중에서 F값이 가장 큰 모형을 찾음
후진제거 (backward selection)	<ul style="list-style-type: none"> ■ 전체 가용한 독립변수들 중에서 불필요한 변수를 하나씩 제거해 나가는 방식 ■ 사용가능한 모든 변수들을 독립변수로 하는 회귀분석에서 시작하며 분석결과 모든 변수들이 통계적으로 유의하면 이들 모두를 사용한 회귀모형을 제시하고 분석을 종료함. ■ 전체변수들을 이용한 회귀분석 결과, 유의하지 않은 변수들로 포함되어 있으면, 전체 변수에서 돌아가면서 하나씩의 변수를 제외한 나머지 독립변수들을 이용하여 회귀분석을 실시
단계선택 (stepwise selection)	<ul style="list-style-type: none"> ■ 전진선택방식과 후진제거방식을 결합한 형태 ■ 먼저 전진선택방식을 이용하여 변수를 하나씩 선택해나감과 동시에 이미 선택되어 있는 변수들 중에 다중공선성이 높게 나타나는 변수들을 후진제거방식으로 제거해 나가는 방법임
제거 (remove)	<ul style="list-style-type: none"> ■ 다른 방식들과는 달리 단독으로 사용할 수 있는 방법 ■ 먼저 다른 선택방식을 사용해서 선택된 변수들 중 연구자가 원하는 특정 변수를 제거하고자 할 경우 사용

4. 더미변수를 이용한 회귀분석

- 명목척도로 측정된 변수를 부득이 사용해야 하는 경우에도 명목척도를 나타내 변수를 0과 1의 값만을 갖는 한 개 혹은 몇 개의 이항변수로 바꾸어 회귀분석에 활용할 수 있음
 - 이항변수
 - 실제로 측정된 변수가 아니고 명목척도로 측정된 변수값들을 서로 구분할 수 있도록 하기 위하여 가상적으로 만든 변수라 하여 이들을 더미변수(dummy variable)라 부름
 - 이러한 변수들을 이용한 회귀분석을 일반 회귀분석과 구분하여 더미변수를 이용한 회귀분석이라고 함
- 가능한 한 적은 수의 독립변수를 사용하여 회귀모형을 개발하는 것이 분석자료의 자유도를 높여, 보다 신뢰할 수 있는 모형을 만들 수 있기 때문에 더미변수를 최소한으로 설정하는 것이 좋음
 - 명목변수의 수준(level) 수가 m개 라면 최소한 그리고 최적의 더미변수의 수는 m-1개가 됨

- 더미변수를 이용하여 회귀분석 하는 또 하나의 이유
 - 명목척도로 측정된 독립변수의 영향을 제거함으로써 남은 독립변수들과 종속변수의 관계를 보다 정확하게 측정할 수 있음

5. 비선형회귀분석에서 독립변수의 변환

- 비선형회귀분석
 - 독립변수와 종속변수 중에 어느 하나 혹은 모두를 변화하여 두 변수 간의 관계가 선형관계가 되도록 한 다음에 수행하는 분석임
- 독립변수의 수가 2개 이상인 다중회귀분석의 경우
 - 종속변수를 변환하는 것은 쉽지 않음으로 대부분의 비선형회귀분석에서는 종속변수보다는 독립변수를 변화시키는 것을 우선으로 함

6. 비선형회귀분석에서 종속변수의 변환

- 일반적으로 독립변수를 변환하여 변환된 독립변수가 종속변수와 선형관계가 되도록 조정함
 - 경우에 따라서는 단순회귀분석에서 독립변수보다 종속변수를 변환하는 것이 바람직한 경우가 있음
- 예) 회귀분석에서는 각각 독립변수에 대응하는 모든 종속변수의 오차는 평균이 0이고 분산이 모두 일정한 정규분포를 한다고 가정하고 분석하지만 이러한 가정을 충족시키지 못하는 경우에는 종속변수와 선형관계가 되도록 독립변수를 아무리 변환해도 오차의 등분산성에 대한 가정을 위배하게 되는 문제는 해결되지 않음
- 경우에 따라서 독립변수 X 대신 종속변수 Y 를 변환하면, 오차의 등분산성에 대한 가정을 위배하게 되는 문제를 해결할 수 있음

7. 비선형회귀분석에서 독립변수와 종속변수의 변환

- 일반적으로 독립변수와 종속변수를 동시에 변환시키는 것은 쉽지 않으나 두 변수 간의 관계를 보다 정확하게 설명할 수 있으면 동시에 같이 변화시켜도 좋음
 - 이러한 경우 변수 간의 산점도와 오차분포를 반복적으로 그려보면서, 두 변수 간의 관계를 가장 효과적으로 설명할 수 있도록 두 변수를 변화시킨 다음 회귀분석을 수행함
- 예) 광고액(A)과 판촉비용(P)으로 매출액(S)을 예측하는 모형을 개발하고자 하는 경우
- 간단히 생각하면 단순한 다중회귀분석으로 구할 수 있음
- 이에 대한 기존의 많은 연구결과, 광고와 판촉은 서로 시너지 효과가 있어서 이들 간에는 단순한 선형관계보다 $S = \alpha A^\beta P^\gamma$ 와 같은 관계가 더 적합한 것으로 밝혀지면 모형에서 필요로 하는 α, β, γ 를 구해야만 매출액을 예측할 수 있는 구체적인 모형을 개발할 수 있음