



SPSS + DataIN을 활용한 통계분석 기본과정

와이즈인컴퍼니 공공솔루션 CSAP 인증서 획득 기념!

연간멤버십 오픈 ALL PASS

전과목 수강 29만원

10

31

일 까지

통계분석 + 데이터사이언스
+ 클릭기반 AI분 강의

총 545 강좌

전과목
(오프라인 1회 수강료비용)

29만원

수강기간
1년 + 1년

2년

데이터캠퍼스의 모든 강의를 1회 수강료 비용으로 대방출

오픈 기념 이벤트 10월 31일 까지 (11월 이후 수강권은 99만원 /1년으로 변경)

타 교육기관 오프라인 1회 평균비용이 수백만원대

데이터캠퍼스

29만원

전 과목
(총 545강좌)

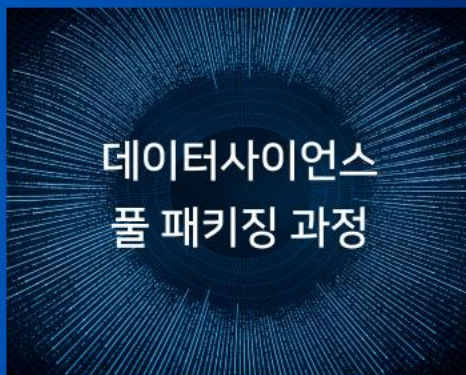
가장 저렴한 비용
통계분석과 빅데이터분석 마스터, 29만원!



- 교수를 가르치는 교수의 명강의!
- 사회과학의 모든 연구방법을 아우르는, 국내 유일의 강의!

강의구성 (9과목, 총 231 강좌 패키지)

SPSS 기본, SPSS 고급회귀, SPSS 다변량, SPSS 비모수통계, AMOS 구조방정식모델분석, HLM 다중선형모형, STATA 메타분석, STATA 패널데이터분석, STATA시계열분석



- 빅데이터 분석 전문가를 위한 풀패키지 과정!
- R과 Python의 모든 분석을 동시에 마스터 한다!
- 데이터분석 핵심+텍스트마이닝+머신러닝+딥러닝 학습 가능

강의구성 (8과목, 총 210 강좌 패키지)

Python : 핵심, 텍스트마이닝, 머신러닝, 딥러닝
R : 데이터분석과 시각화, 텍스트마이닝, 머신러닝, 딥러닝



- 코딩 없이 머신러닝을 완성하는 최고의 방법!
- 오픈소스 ORANGE 활용 AI 분석의 국내 유일의 강의!
- 학생, 직장인도 취업과 실무에 활용할 수 있는 최고의 커리큘럼!

강의구성 (4과목, 총 104강좌 패키지)

ORNAGE와 기본핵심 마스터, 지도학습 마스터, 비지도학습 마스터, 텍스트와 이미지분석 마스터

구매 방법



구매창 바로가기

위 사이트에서 바로 구매 가능

CONTENTS



통계분석 기본과정

1. 자료와 분석방법 간의 관계 마스터하기
2. 통계결과 해석을 위한 통계량 이해하기
3. 분석실습1: 기술통계와 차이분석
4. 분석실습2: 상관과 영향관계분석

Part
01

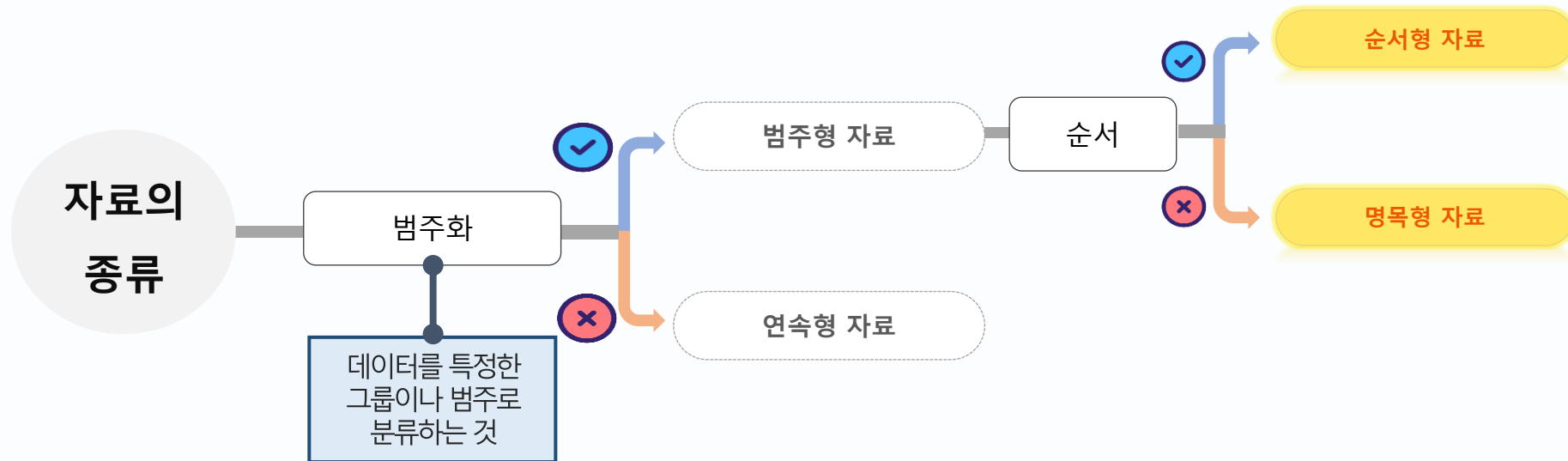
자료와 분석방법 간의 관계 마스터하기



1. 자료의 이해

가. 자료의 정의

- 자료는 수집되고, 요약되고, 분석, 해석된 사실이나 수치
- 통계분석 방법은 자료의 속성과 분석 목적에 따라 달라짐
- 분석하고자 하는 자료가 분석방법에서 가정한 조건을 얼마나 만족하는지에 따라 분석방법의 적절성이 결정
→ 자료의 속성에 따른 분류 필요



1. 자료의 이해

나. 자료의 종류

종류	세부 항목	정의	예시
범주형 자료 : 수치로 측정 불가능한 자료	명목형 자료	<ul style="list-style-type: none"> 범주 간에 순서가 없는 데이터 주로 빈도 분석을 통해 범주별로 빈도를 계산하고 범주 간 비교를 수행함 	성별, 지역 등
	순서형 자료	<ul style="list-style-type: none"> 범주 간에 순서가 있는 데이터 순위와 같은 상대적인 비교에 사용됨 	<ul style="list-style-type: none"> 학력 수준(고졸, 대졸, 석사, 박사) 만족도(매우 불만족, 불만족, 보통, 만족, 매우 만족) 등
연속형 자료		<ul style="list-style-type: none"> 수치적이며 무한한 값을 가질 수 있는 데이터 연속적인 스케일을 가지며 측정 가능한 정밀도가 있음 평균, 분산, 표준편차, 상관 분석, 회귀 분석 등의 통계 분석 기법을 적용하는데 주로 사용됨 	나이, 키, 온도, 판매량, 수익 등

1. 자료의 이해

다. 변수의 정의

- 변수란 연구대상의 경험적 속성을 나타내며 그 속성에 계량적인 수치를 부여할 수 있는 개념
- 독립변수 : 종속변수의 변화를 가져오거나 영향을 미치는 원인 변수로써, 결과를 예측을 하게 하거나 차이를 설명하기 위해 사용되는 예측 변수
- 종속변수 : 독립변수의 영향으로 나타나는 결과가 되는 결과 변수로써, 독립변수의 변화에 따라 변화되는 변수

차이분석에서 독립변수와 종속변수

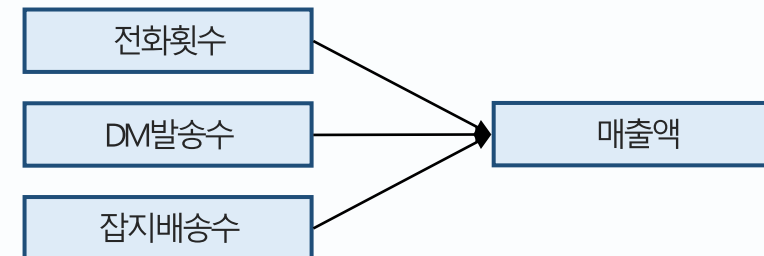
- 성별이 무엇이나에 따라 결과(지지정당)의 명수/비율이 달라짐
- 독립변수: 성별
- 종속변수: 지지정당

구분	A당	B당	C당	D당
전체	120	50	30	200명

구분	A당	B당	C당	D당
전체	120	50	30	200명
남	80	15	5	100명
여	40	35	25	100명

영향관계에서 독립변수와 종속변수

- 어떤 판촉활동이 매출에 더 밀접한가(영향을 미치는가)를 파악
- 각 판촉활동 수에 따라 매출액이 달라짐
- 독립변수: 전화횟수, DM발송수, 잡지배송수
- 종속변수: 매출액



1. 자료의 이해

라. 척도와 분석방법

변수의 형태(척도)와 분석방법 간의 관계		독립변수	
		범주형 자료	연속형 자료
종속변수	범주형 자료	교차분석	로지스틱 회귀분석 프로빗 분석 판별분석 군집분석
	연속형 자료	t-test 분산분석(ANOVA)	상관관계분석 회귀분석 (구조방정식모델분석)

Part
02

통계결과 해석을 위한 통계량 이해하기

1. 가설과 통계량

가. 가설의 개념

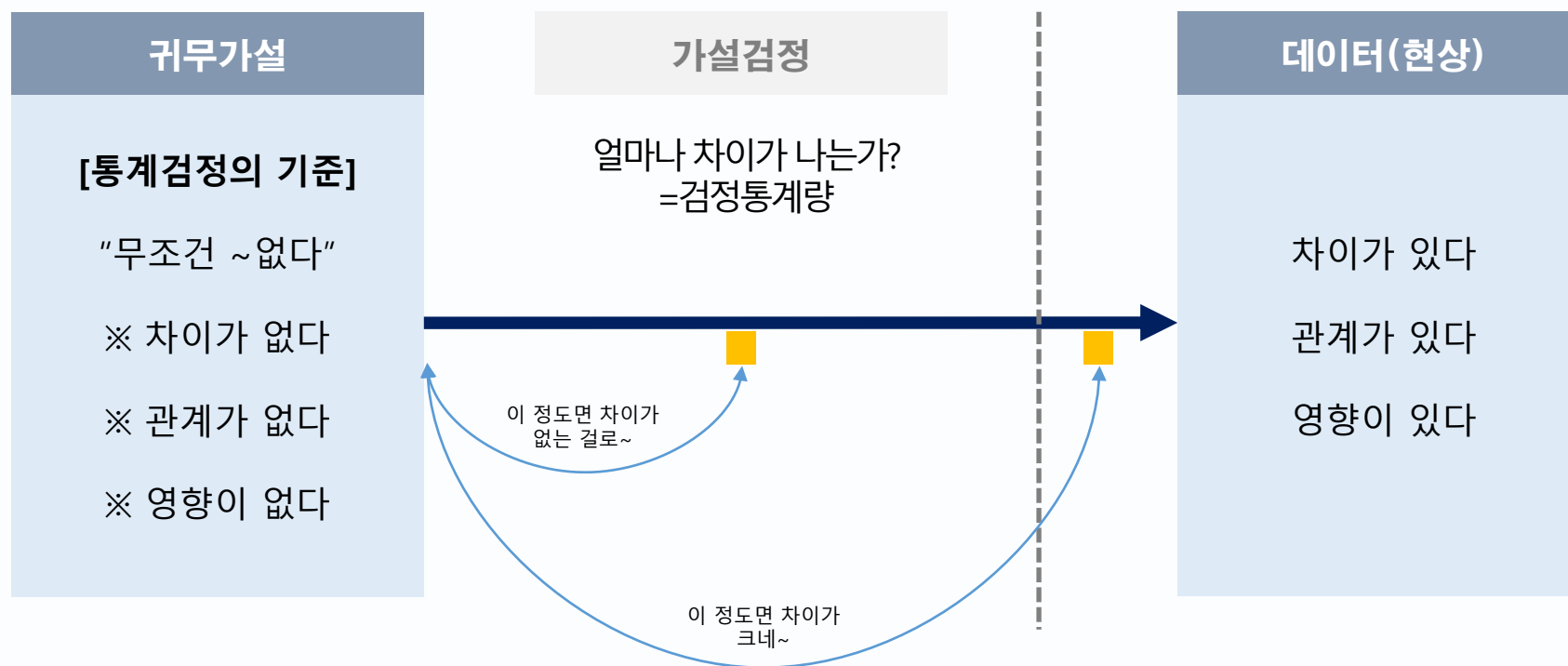
- 가설(hypothesis)이란 '아직 경험적으로 검증되지 않은 일종의 예비이론으로서 둘 혹은 그 이상의 변인들 간의 추측적 진술로서 독립변수와 종속변수의 관계'
- 가설은 귀무가설(null hypothesis)과 대립가설(alternative hypothesis)로 나뉨

가설의 유형		
유형	가설	가설적 표현의 예시
차이의 가설	귀무가설(영가설)	성별에 따라서 행복에 차이가 없다(없을 것이다)
	대립가설(연구가설)	성별에 따라서 행복에 차이가 있다(있을 것이다)
관계의 가설	귀무가설(영가설)	소득과 행복은 상관관계가 없다(없을 것이다)
	대립가설(연구가설)	소득과 행복은 상관관계가 있다(있을 것이다)
영향의 가설	귀무가설(영가설)	소득은 행복에 영향이 없다(영향을 미치지 않는다)
	대립가설(연구가설)	소득은 행복에 영향이 있다(영향을 미친다)

1. 가설과 통계량

나. 귀무가설과 대립가설

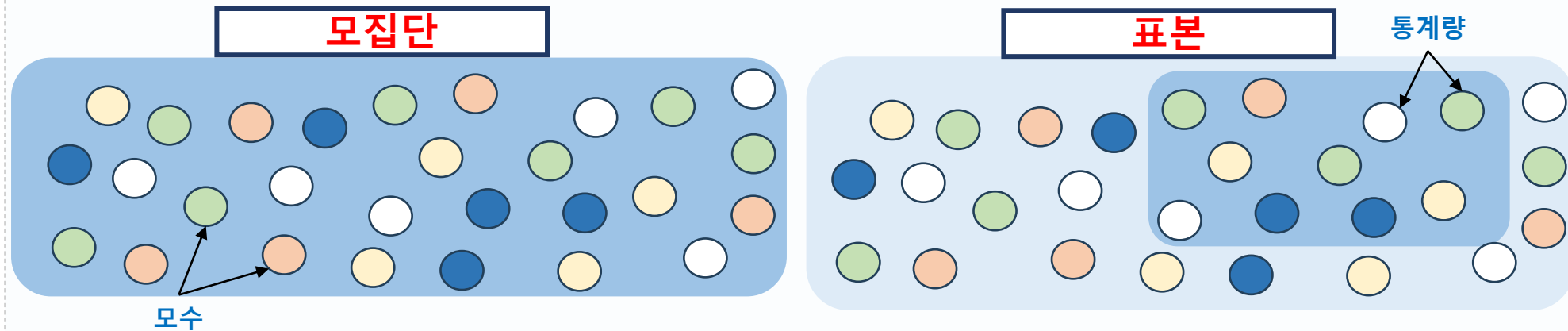
- 귀무가설: 자신이 주장하고자 하는 연구가설과 논리적으로 반대의 입장을 취하는 가설 (기존과 차이가 없음)
- 대립가설(연구가설): 연구자가 주장하고자 하는 가설로 연구문제에 대한 잠정적인 해답



1. 가설과 통계량

다. 모집단과 표본

- 모집단: 어떤 정보를 얻고자 하는 전체 대상 또는 전체 집합
- 표본: 모집단으로부터 추출된 모집단의 부분 집합



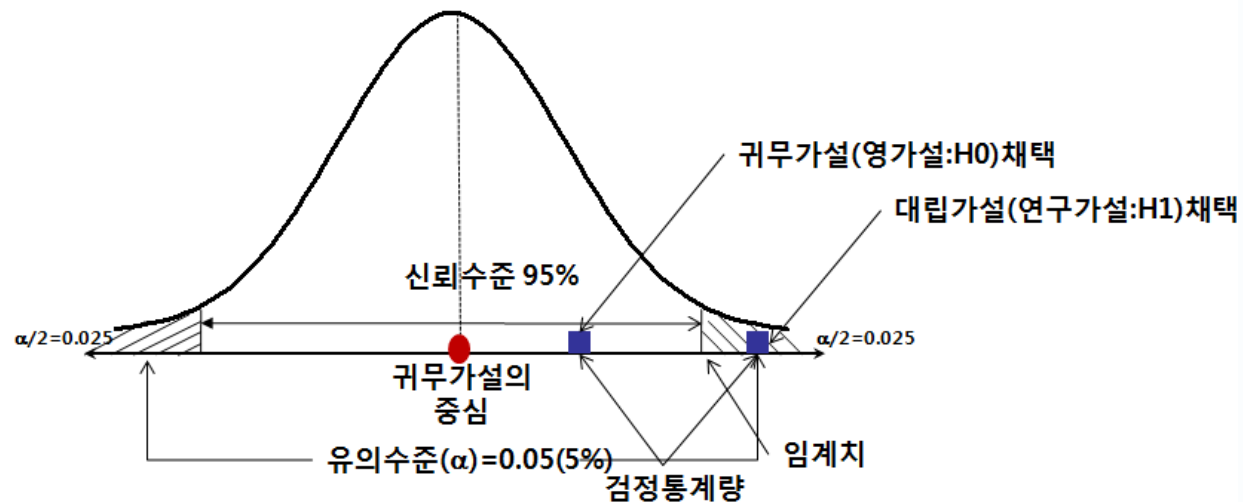
전수조사가 가능하다면, '추정'이 필요 없음
일부 표본으로 얻어진 추정결과에는 오차가 발생할 수 있음(표본오차)

1. 가설과 통계량

라. 가설검정

- 가설검정(hypothesis test) : 표본으로부터 얻은 사실에 근거하여, 모집단에 이 가설이 맞는지 틀리는지를 통계적으로 검정하는 분석방법
- 가설을 검정한다: 귀무가설과 표본에서 수집된 데이터 간의 차이 정도를 판단하여 귀무가설을 채택할지 혹은 기각할지를 판단하는 것
- 검정통계량: 실제 현상(수집된 데이터)이 판단의 기준이 되는 귀무가설(~없다)과 얼마나 차이가 나타나는지를 계산한 값
- 신뢰수준(confidence level) : 귀무가설을 여전히 신뢰(채택)할만한 구간

가설과
통계량의
개념



1. 가설과 통계량

라. 가설검정

- 1종 오류(type 1 error) : 귀무가설의 잘못된 기각 (연구가설의 잘못된 채택 → 설레발)
- 2종 오류(type 2 error) : 귀무가설의 잘못된 채택 (연구가설의 잘못된 기각 → 현상의 간과)

통계검정과 오류의 유형

추정 실제	귀무가설이 참일 경우	귀무가설이 거짓일 경우
	귀무가설을 기각 (연구가설을 지지)	1종 오류(α) 올바른 추론
귀무가설을 기각하지 않음 (연구가설을 배척)	올바른 추론	2종 오류(β)

1. 가설과 통계량

마. 신뢰구간

- 신뢰수준에 따른 표본평균 추론의 용인 가능한 범위

결과1: 대선 후보의 지지율을 파악하기 위해 전국 성인남녀 1,000명을 조사한 결과 지지율은 A후보: 45%, B후보 41%이며, 본 조사는 신뢰수준 95%, 표집오차 3.0%로 나타났다. 두 A후보가 B후보를 앞선다고 말할 수 있는가?

결과2: 대선 후보의 지지율을 파악하기 위해 전국 성인남녀 1,000명을 조사한 결과 지지율은 A후보: 47%, B후보 40%이며, 본 조사는 신뢰수준 95%, 표집오차 3.0%로 나타났다. 두 A후보가 B후보를 앞선다고 말할 수 있는가?

신뢰구간에 대한 해석		
후보	결과1	결과2
A후보	42% ~48%	44% ~50%
B후보	38%~ 44%	37%~ 43%
의미	다른 표본(혹은 모집단 전체)을 조사하면 결과가 뒤바뀔 수도 있다	다른 표본(혹은 모집단 전체)을 조사해도 결과는 뒤바뀌지 않는다
통계적 표현	귀무가설(영가설)을 채택	귀무가설(영가설)을 기각

1. 가설과 통계량

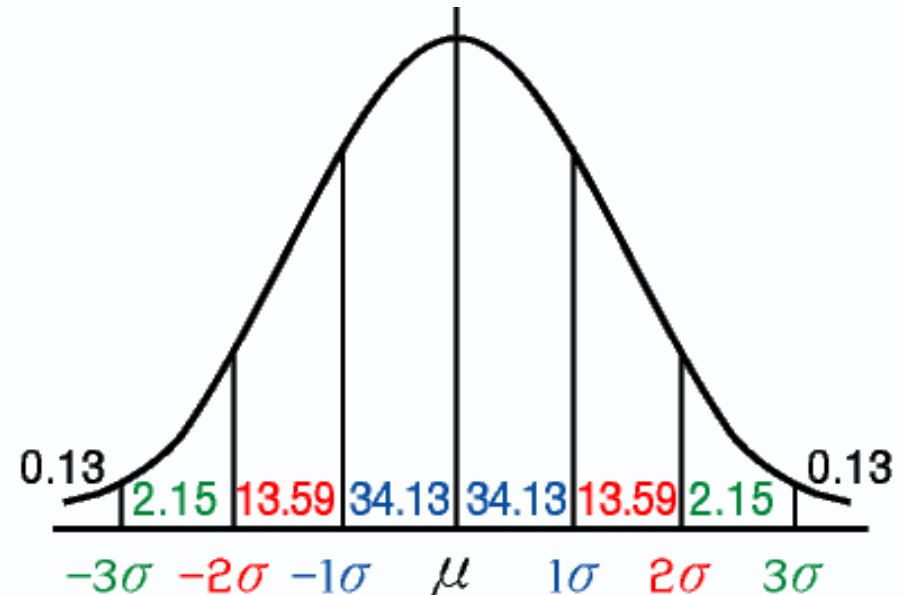
바. 분석방법별 검정통계량

분석방법	독립 - 종속	귀무가설(H0)/대립가설(H1)	검정통계량	p
교차분석	범주-범주	H0: 독립변수에 따라 종속변수에는 차이가 없다. H1: 독립변수에 따라 종속변수는 차이가 있다.	χ^2	$P < 0.05$ → 대립
독립표본 t-test	범주-연속	H0: 독립변수에 따라 종속변수에는 차이가 없다. H1: 독립변수에 따라 종속변수는 차이가 있다.	t-value	$P < 0.05$ → 대립
분산분석	범주-연속	H0: 독립변수에 따라 종속변수에는 차이가 없다. H1: 독립변수에 따라 종속변수는 차이가 있다.	F-value	$P < 0.05$ → 대립
상관관계 분석	연속-연속	H0: 독립변수와 종속변수는 상관관계가 없다. H1: 독립변수와 종속변수는 상관관계가 있다.	Pearson's r	$P < 0.05$ → 대립

2. 정규분포와 정규성 검정

가. 정규분포의 개념

- 특정 값의 출현 비율을 그렸을 때,
중심(평균값)을 기준으로 좌우 대칭 형태(종 모양)
- 평균과 중앙값은 같고, 분포의 중앙에 있음
- 연속확률분포 중 가장 대표적인 분포



평균을 중심으로 ($\mu-1\sigma$, $\mu+1\sigma$) 내에
전체 사례수의 68.26%가 존재

$\mu \pm 1\sigma$
68.26%

평균을 중심으로 ($\mu-2\sigma$, $\mu+2\sigma$) 내에
전체 사례수의 95.44%가 존재

$\mu \pm 2\sigma$
95.44%

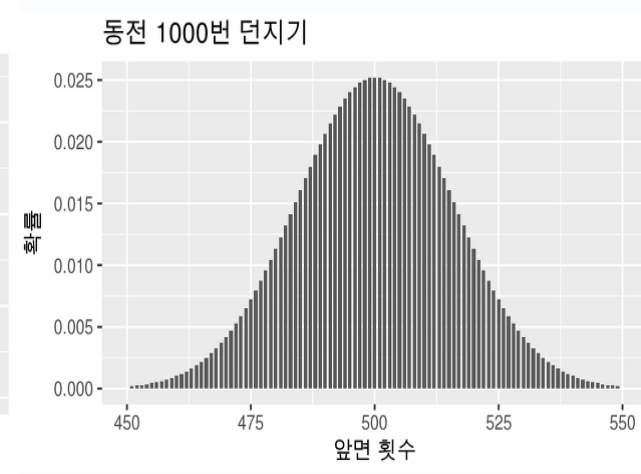
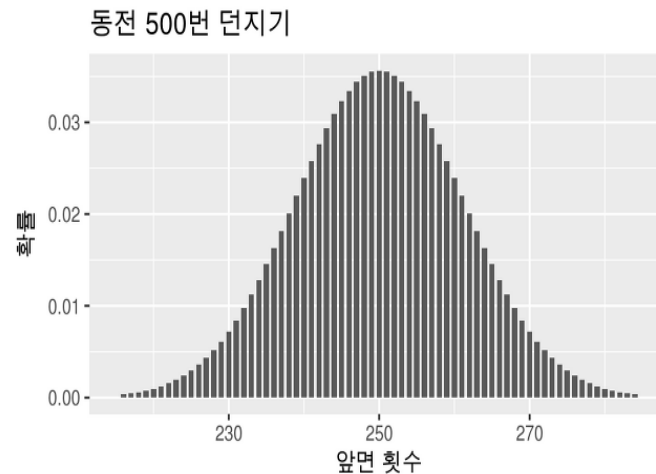
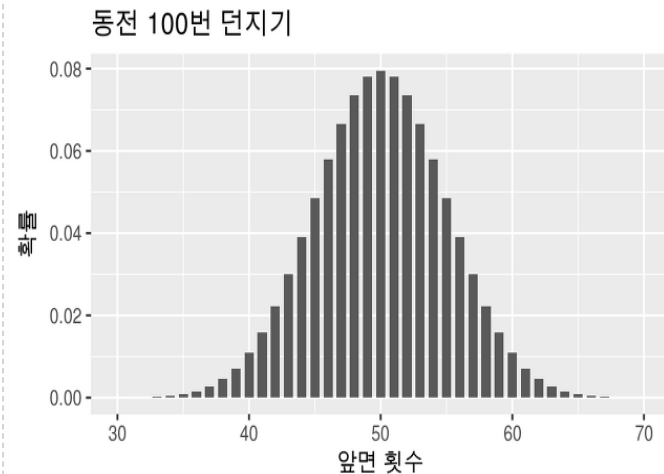
평균을 중심으로 ($\mu-3\sigma$, $\mu+3\sigma$) 내에
전체 사례수의 99.74%가 존재

$\mu \pm 3\sigma$
99.74%

2. 정규분포와 정규성 검정

나. 정규분포 예시

- 사람이 동전(앞, 뒤만 존재)을 던졌을 때 앞면이 나오는 경우는 초반(실험 횟수가 적을 경우)에는 정규 분포와 다른 경우가 나올 수 있다.
- 횟수가 늘어날수록 앞면과 뒷면이 나오는 경우의 수가 비슷해지고, 무한히 많이 시행한다면 50%에 수렴
- 동전을 100번 던졌을 경우에도 어느정도 정규 분포와 비슷한 모양이 나오지만, 횟수가 500번, 1000번으로 증가할수록 정규 분포와 더욱 유사해진다.

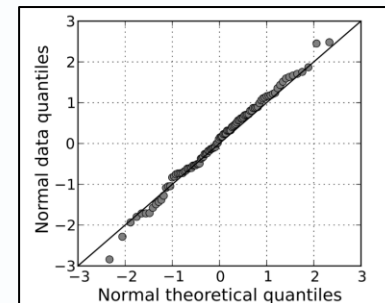


2. 정규분포와 정규성 검정

다. 정규성 검정

- 관측값들이 정규분포를 따르는 모집단에서 추출되었는지를 검정하는 것
- 정규성 검정을 하는 이유: 모수통계 검증 방법을 사용하기 위해서(t-검정, 분산분석, Pearson 상관관계분석, 회귀분석)
- 모수통계: 관측 값들이 정규분포를 가지고 있다고 간주할 때 활용하는 통계

종류	특성
샤피로-윌크 검정 (Shapiro-Wilks test)	<ul style="list-style-type: none"> • 표본수(n)가 2000 미만인 데이터셋에 적합한 정규성 검정
콜모고로프-스미르노프 검정 (Kolmogorov-Smirnov test)	<ul style="list-style-type: none"> • 표본수(n)가 2000 초과인 데이터셋에 적합한 정규성 검정
왜도와 첨도	<ul style="list-style-type: none"> • 왜도의 절댓값 < 2, 첨도의 절댓값 < 7 일때, 정규성 만족
QQ 플롯 (Quantile-Quantile plot)	<ul style="list-style-type: none"> • 데이터셋이 정규분포를 따르는지 판단하는 시각적 분석 방법 • 분석할 데이터 종류가 많지 않다면, QQ plot을 통해 시각적으로 확인해보는게 가장 간단하며 직관적



데이터가 정규분포를 따르는 QQ 플롯

Part
03

분석실습1: 기술통계와 차이분석

0.1 SPSS 설치

1. https://www.ibm.com/kr-ko/products/spss-statistics?lot=1&mhsrsrc=ibmsearch_a&mhq=spss

위의 링크로 접속

IBM

제품 ▾ 솔루션 ▾ 컨설팅 지원 ▾ 더 둘러보기 ▾

IBM SPSS IBM SPSS Statistics 기능 ▾ 에디션 ▾ 가격 지원 ▾ 자원 ▾

12개월 구독 시 10% 할인

IBM SPSS Statistics

비즈니스 및 연구 문제 해결에 도움이 되는 데이터 인사이트 발견

지금 구매하기(월 USD 99*) → SPSS Statistics 무료 체험 → ① 클릭

Explore the power of statistical analysis in your organization (4:53)

0.1 SPSS 설치

2. 회원가입 및 로그인

3. 다운로드

IBM | My IBM | 프로파일 | 청구

내 제품

Q 제품 이름 검색

제품명별로 정렬

필터링된 제품 목록을 표시합니다. 모두 표시

체험판
1 오퍼링

Σ IBM SPSS Statistics Subscription Trial
Upgrade options

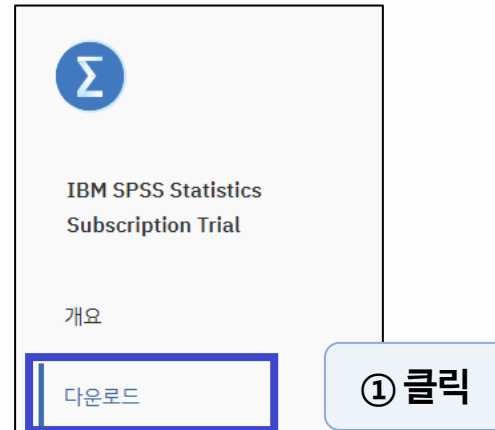
만료일: 2024년 5월 9일
사용 중

다운로드 관리

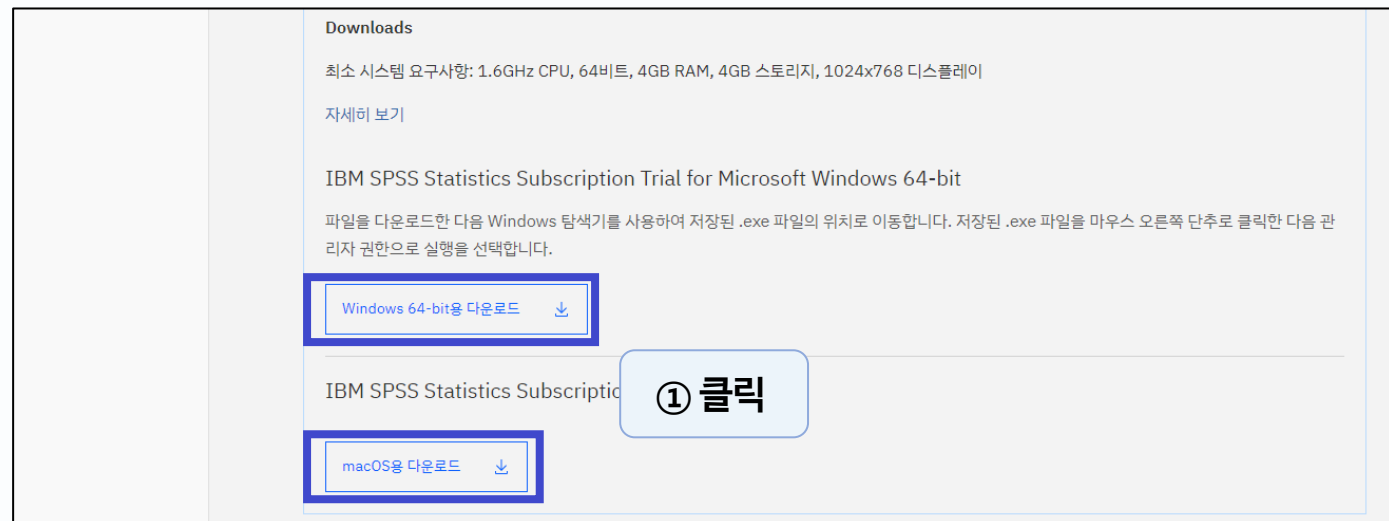
① 클릭

0.1 SPSS 설치

4. 개요 → 다운로드 클릭



5. 스크롤 내려서 각자 컴퓨터 용으로 다운 받기



0.1 SPSS 자료등록

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래프(G) 유틸리티(U) 확장(X) 창

새로 만들기(N) >
열기(O) >
데이터 가져오기(D) >
+ 일반 열기
복원점 열기(P)...
닫기(C) Ctrl+F4
저장(S) Ctrl+S
다른 이름으로 저장(A)...
모든 데이터 저장(L)
복원점 저장(S)...
내보내기(T) >
파일을 읽기 전용으로 표시(K)
저장된 파일로 되돌리기(E)
+ Make Variable Catalog
☑ 파일에 자동 복구 사용(E)
데이터 세트 이름 변경(M)...
데이터 파일 정보 표시(I) >
데이터 캐쉬(H)...
프로세서 중단 Ctrl+Period
인쇄 미리보기(V)
인쇄(P)... Ctrl+P

데이터베이스(B) >
Excel...
CSV 데이터...
텍스트 데이터(T)...
SAS...
Stata...
Cognos Business Intelligence(O)...
Cognos TM1...

① 클릭

데이터 열기

찾아보기(I): 4월 강의안_최종

4월 17일_(1)통계분석기본_data.xlsx
4월 17일_(1)통계분석기본_문항속성.xlsx
4월 19일_path1.xls
4월 19일_path2.xls
4월 19일_sem1.xls

② 파일 선택

파일 이름(N): 4월 17일_(1)통계분석기본_data.xlsx 열기(O)
파일 유형(T): Excel(*.xls, *.xlsx, *.xlsm) 취소
인코딩(E):

워크시트(K): Sheet1 [A1:H251]
범위(N):

☑ 데이터 첫 행에서 변수 이름 읽어오기(V)
☑ 데이터 유형을 결정하는 값의 파센트(E) 95
☑ 숨겨진 행과 열 무시(I)
☐ 문자열 값에서 선행 공백 제거(M)
☐ 문자열 값에서 후행 공백 제거(G)
미리보기(W)

③ 불러올 영역 설정

	직원번호	서비스...	작년...	고객만	성별	연령	직업	소득
1	4	5	3	4	1	32	1	2
2	4	2	5	4	1	57	1	1
3	4	5	4	1	1	36	1	5
4	4	4	3	2	1	44	1	2
5	3	3	2	4	1	46	1	3
6	4	4	3	2	1	55	1	4
7	5	4	3	2	1	35	1	2

최종 데이터 유형은 모든 데이터를 기준으로 하므로, 처음 200개 데이터 행을 기준으로 하는 미리보기에서는 이와 다를 수 있습니다. 미리보기에는 처음 500개 열만 표시됩니다.

확인 불러넣기(P) 재설정(R) 취소 도움말

0.1 SPSS 자료등록

파일(F) 편집(E) 보기(V) 데이터(D) 변환(I) 분석(A) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

43 : 직원만족도 서비스만족도 작년_고객만족도 고객만족도 성별 연령 직업 소득

1 4 5 3 4 1 32 1 2
2 4 2 5 4 1 57 1 1
3 4 5 4 4 1 36 1 5
4 4 4 3 2 1 44 1 2
5 3 3 2 4 1 46 1 3
6 4 4 3 2 1 55 1 4
7 5 4 3 2 1 35 1 2
8 4 4 3 3 1 61 1 3
9 2 3 1 2 1 26 1 5
10 5 4 3 4 1 24 7 2
11 3 3 2 4 1 32 2 2
12 4 4 3 2 1 42 1 1
13 2 3 3 3 1 48 4 1
14 2 2 1 2 1 55 1 2
15 2 2 2 1 1 36 1 1
16 3 3 2 3 1 62 4 3
17 3 2 1 4 1 42 2 2
18 4 3 2 4 1 41 2 2
19 3 3 3 3 1 44 3 3
20 4 4 3 4 1 40 2 2
21 4 4 1 4 1 30 2 3
22 4 4 3 3 1 48 7 2
23 4 4 3 4 1 35 7 2
24 3 3 3 1 1 31 2 3
25 4 4 2 3 1 37 4 4
26 3 3 2 4 1 46 2 2
27 4 4 1 3 1 43 1 2
28 3 4 2 1 1 41 1 1
29 4 4 3 4 1 40 2 2
30 4 3 2 3 1 32 3 2
31 4 3 3 3 1 33 3 3
32 4 3 1 4 1 44 2 4
33 3 3 4 1 1 30 2 4
34 4 3 4 1 1 34 2 3
35 3 4 1 1 1 52 2 4

개요 데이터 보기 변수 보기

이름	유형	너비	소수점자리	레이블	값	값측간	열	맞춤	측도	요약
1 직원만족	숫자	2	0	영화관 직원들은 친절한 서비스를 제공한다	지정없음	지정없음	12	오른쪽	명목형	인적
2 서비스만족	숫자	2	0	영화관 서비스에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
3 작년_고객만족도	숫자	2	0	내가 이용한 영화관에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
4 고객만족도	숫자	2	0	내가 이용한 영화관에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
5 성별	숫자	2	0	귀하의 성별은?	지정없음	지정없음	12	오른쪽	명목형	인적
6 연령	숫자	3	0	귀하의 연령은?	지정없음	지정없음	12	오른쪽	척도	인적
7 직업	숫자	2	0	귀하의 직업은?	지정없음	지정없음	12	오른쪽	명목형	인적
8 소득	숫자	2	0	귀하의 가계 월평균 소득은?	지정없음	지정없음	12	오른쪽	명목형	인적

레이블	값	값측간	열	맞춤	측도	요약
영화관 직원들은 친절한 서비스를 제공한다	지정없음	지정없음	12	오른쪽	명목형	인적
영화관 서비스에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
내가 이용한 영화관에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
내가 이용한 영화관에 대해 전반적으로 만족한다	지정없음	지정없음	12	오른쪽	명목형	인적
귀하의 성별은?	지정없음	지정없음	12	오른쪽	명목형	인적
귀하의 연령은?	지정없음	지정없음	12	오른쪽	척도	인적
귀하의 직업은?	지정없음	지정없음	12	오른쪽	명목형	인적
귀하의 가계 월평균 소득은?	지정없음	지정없음	12	오른쪽	명목형	인적

값 레이블(V)

맞춤법(S)...

값 레이블(V):

값(U)	레이블(L)
1	전혀 그렇지 않다
2	그렇지 않다
3	보통이다
4	그렇다
5	매우 그렇다

0.2 DataIN 자료등록

1. 엑셀 데이터 준비

	A	B	C	D	E	F	G	H
1	직원친절만족	서비스만족	작년_고객만족	고객만족도	성별	연령	직업	소득
2	4	5	3	4	1	32	1	2
3	4	2	5	4	1	57	1	1
4	4	5	4	4	1	36	1	5
5	4	4	3	2	1	44	1	2
6	3	3	2	4	1	46	1	3
7	4	4	3	2	1	55	1	4
8	5	4	3	2	1	35	1	2
9	4	4	3	3	1	61	1	3
10	2	3	1	2	1	26	1	5
11	5	4	3	4	1	24	7	2
12	3	3	2	4	1	32	2	2
13	4	4	3	2	1	42	1	1
14	2	3	3	3	1	48	4	1
15	2	2	1	2	1	55	1	2
16	2	2	2	1	1	36	1	1
17	3	3	2	3	1	62	4	3
18	3	2	1	4	1	42	2	2
19	4	3	2	4	1	41	2	2
20	3	3	3	3	1	44	3	3
21	4	4	3	4	1	40	2	2
22	4	4	1	4	1	30	2	3
23	4	4	3	3	1	48	7	2
24	4	4	3	4	1	35	7	2
25	3	3	3	1	1	31	2	3
26	4	4	2	3	1	37	4	4
27	3	3	2	4	1	46	2	2
28	4	4	1	3	1	43	1	2
29	3	4	2	1	1	41	1	1
30	4	4	3	4	1	40	2	2
31	4	3	2	3	1	32	3	2
32	4	3	3	3	1	33	3	3

2. 속성 데이터 준비

보기형 → 범주형

	A	B	C	D	E	F	G
1	문항명	문항설명	문항유형	보기개수	보기1	보기2	보기3
2	직원친절만족	영화관 직원들은 친절한 서비스를 제공한다	보기형	5	전혀 그렇지 않다	그렇지 않다	보통이다
3	서비스만족	영화관 서비스에 대해 전반적으로 만족한다	보기형	5	전혀 그렇지 않다	그렇지 않다	보통이다
4	작년_고객만족도	내가 이용한 영화관에 대해 전반적으로 만족한다	숫자형				
5	고객만족도	내가 이용한 영화관에 대해 전반적으로 만족한다	숫자형				
6	성별	귀하의 성별은?	보기형	2	남성	여성	
7	연령	귀하의 연령은?	숫자형				
8	직업	귀하의 직업은?	보기형	7	학생	회사원/공무원	자영업/사업
9	소득	귀하의 가계 월평균 소득은?	보기형	5	100만원 미만	100~200만 이내	200~300만 이내

숫자형 → 연속형

0.2 DataIN 자료등록

1. DataIN 자료등록

W 데이터인

설문

분석

다면평가 Adv.

통합리포팅

응답보채기

문의

+

새분석생성

0

파일보관함

5

분석리스트

새 프로젝트 추가

ProjectTitle

Q

◇

전체

📄

☐ 분석

📄

📄

조절회귀

16문항

433명

2024-04-09 20:15:44

☐ 분석

📄

📄

이항로지스틱회귀분석

7문항

700명

2024-04-08 20:35:02

☐ 분석

📄

📄

매개회귀

8문항

375명

파일데이터등록

설문데이터등록

등록하기

통계분석 기본

데이터파일

샘플 다운로드

파일선택

통계분석기본_data.xlsx

문항속성파일

샘플 다운로드

파일선택

통계분석기본_문항속성.xlsx

TIP

- ✓ '샘플 다운로드' 버튼 클릭 후 예시 파일 양식을 참고하시기 바랍니다.
- ✓ 머리글, 번호는 입력해야 합니다.
- ✓ 다중 입력형인 경우 [문항 이름_1.문항 이름_2.문항 이름_3] 형태로 입력하시기 바랍니다.

0.2 DataIN 자료등록

2 자료등록 완료 확인

전체

분석

통계분석 기본

8문항

250명

2024-04-12 13:40:06

분석

조절회귀

16문항

433명

2024-04-09 20:15:44

분석

이항로지스틱회귀분석

7문항

700명

2024-04-08 20:35:02

분석

매개회귀

8문항

375명

2024-04-08 20:24:56

저장하기

① 신규 프로젝트 등록됨

TIP

- 문항 속성 파일 업로드 기능을 통해 일괄 업로드를 할 수 있습니다.
- Sample을 참고하여 문항 속성파일을 업로드 하주세요.
- 셀이 병합되어 있거나, 공백이 포함된 경우 정상적으로 업로드 되지 않을 수 있습니다.
- 문항 속성이 기존에 존재하는 경우 기존의 속성은 모두 삭제 되고, 새롭게 등록이 됩니다.

문항속성 샘플 다운로드

문항속성 파일선택

선택된 파일이 없습니다.

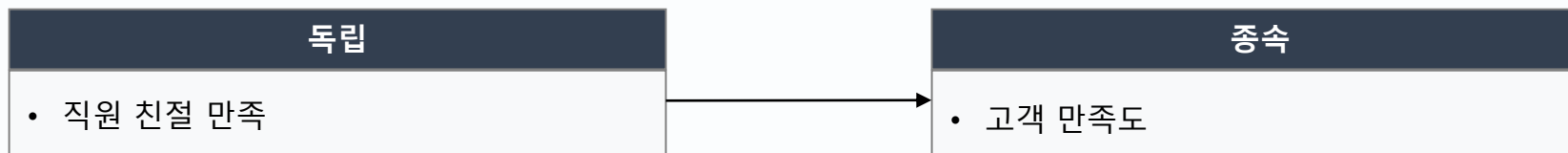
② 문항별 속성이 보여짐

1~10 10 개씩 보기

no	문항명	문항설명	유형
1	직원친절만족	영화관 직원들은 친절한 서비스를 제공한다	보기형
	<div>최소값: 1 최대값: 5</div> <div>전혀 그렇지 않다</div> <div>그렇지 않다</div> <div>보통이다</div>		<div>보기 한번에 볼여넣기</div>

0.3 자료 설명

1. 연구 모형



2. 가설 및 분석 방법

가설/기본분석	분석방법	도출정보
응답자(표본) 특성 정리	빈도분석	응답자 특성 정리
주요 변수(만족도 등) 점수(수준) 정리	기술통계분석	전체 만족도 파악
H1:고객의 성별, 연령대에 따라서 소득에 차이가 있을 것이다	교차분석	응답자 소득 차이여부 확인 가능
H2:고객의 성별에 따라서 고객만족도에 차이가 있을 것이다	ttest 및 ANOVA	응답자 성별 고객만족도 차이여부 확인 가능
H3:고객의 직업에 따라서 고객만족도에 차이가 있을 것이다	ttest 및 ANOVA	응답자 직업별 고객만족도도 차이여부 확인 가능
H4:직원친절만족은 고객만족도에 정(+)의 영향을 미칠 것이다	Pearson 상관관계분석	독립변수와 종속변수간의 상관 정도 파악

1. 기술통계분석

가. 분석 및 자료의 특성

- 연속형 자료의 평균/표준편차/왜도&첨도를 구하는 분석방법
- 연속형 자료(등간척도와 비율척도)가 분석의 대상

구분	명수	평균	표준편차	왜도	첨도	표준오차
평균월급	400	342.2	33.4	2.23	4.32	1.67

나. 분석목적

- 연속형 자료의 중심성(평균 등), 자료의 퍼짐(표준편차/분산 등), 치우침과 뾰족함(왜도/첨도) 등의 특성 요약을 파악
- 왜도는 절대값 2, 첨도는 절대값 7보다 작을 경우 정규성에 위배되지 않는다고 봄

1. 기술통계분석

다. SPSS 실습

● 기술통계 실행

① 분석 클릭 → ② 기술통계량 클릭 → ③ 기술통계 클릭 → ④ 변수 1개(고객만족도) 선택
→ ⑤ 옵션 클릭 → ⑥ 분포(첨도, 왜도) 선택 → ⑦ 계속 클릭 → ⑧ 확인 클릭

① 클릭

② 클릭

③ 클릭

④ 변수 1개 선택

⑤ 클릭

⑥ 클릭

⑦ 클릭

⑧ 클릭

변수(M): 고객만족도

기술통계: 옵션

☒ 평균(M) ☐ 합계(S)

산포도

☒ 표준편차(T) ☒ 최소값(N)

☐ 분산(V) ☒ 최대값(X)

☐ 범위(R) ☐ 평균의 표준오차(E)

분포

☒ 첨도(K) ☒ 왜도(W)

표시 순서

☒ 변수 목록(B)

☐ 문자순(A)

☐ 평균값 오름차순(C)

☐ 평균값 내림차순(D)

확인

불여넣기(P)

재설정(R)

계속(C)

취소

도움말

1. 기술통계분석

다. SPSS 실습

- 기술통계 결과

기술통계량									
	N	최소값	최대값	평균	표준편차	왜도		첨도	
	등계량	등계량	등계량	등계량	등계량	등계량	표준오차	등계량	표준오차
고객만족도	250	1	5	3.16	.933	-.197	.154	-.435	.307
유효 N(목록별)	250								

문항이 정상적 범위 내에 있으며
왜도/첨도 절대값 2 미만으로 정규성을 위배하지 않음

1. 기술통계분석

라. DATA IN 실습

- 기술통계 실행

① 통계분석Reporting 클릭 → ② 단변량분석 클릭 → ③ 기술통계분석 클릭
→ ④ 변수 1개(고객만족도) 선택 → ⑤ 분석하기 클릭

① 클릭

② 클릭

③ 클릭

④ 변수 선택 후 이동

⑤ 클릭

보고서Reporting 통계분석Reporting

기술통계분석

집단비교분석 척도화분석 회귀분석

빈도분석 a=3 and c*0.3 >=1

기술통계분석

교차분석

R Syntax 분석초기화 분석하기

전체할목록 8건 전체선택 선택취소

1.직원친절만족 (집단수:5)

2.서비스만족 (집단수:5)

3.작년_고객만족도... (집단수:0)

5.성별 (집단수:2)

6.연령 (집단수:0)

7.직업 (집단수:7)

선택변수 1건 영역초기화 선택삭제

4.고객만족도 (집단수:0)

1. 기술통계분석

라. DATA IN 실습

- 기술통계 결과

① 문서로 바로 저장

주요+상세 파일저장

파일저장

파일보관함저장

주요결과

상세결과

결과표 및 결과의 자동해석 제시!!

정규성 평가의 기준은 왜도 2미만, 첨도 7미만으로 한다.

변수	N	최소값	최대값	평균	표준편차	왜도	첨도
고객만족도	250	1	5	3.16	0.93	-0.20	-0.44

기술통계분석결과

고객만족도는 평균 3.16(표준편차: 0.93), 왜도 -0.20, 첨도 -0.44로 정규성을 충족하는 것으로 나타났다.

문항이 정상적 범위 내에 있으며
왜도/첨도 절대값 2 미만으로 정규성을 위배하지 않음

2. 교차분석

가. 분석 및 자료의 특성

- 명목자료를 이용하여 두 변수간의 상호관련성을 알아보고자 할 때
- 분석기법과 해석이 비교적 쉽지만, 실제로 정치·여론조사, 마케팅조사 등에서 가장 널리 사용되는 통계분석방법

구분	샤넬	루이비통	구찌	기타	전체
직장인	40%	25%	15%	20%	100%
주부	55%	30%	10%	5%	100%
학생	35%	35%	15%	15%	100%
기타	30%	40%	10%	10%	100%
전체	40%	325%	125%	125%	1000%

나. 교차분석의 가설 및 검정통계량

- 귀무가설: 소비자의 직업에 따라 선호하는 명품가방에 차이가 없다
- 대립가설: 소비자의 직업에 따라 선호하는 명품가방에 차이가 있다
- 검정통계량: χ^2 , 자유도: df, 이를 고려하여, 유의확률(p)의 계산

2. 교차분석

다. SPSS 실습

● 교차분석 실행

① 분석 클릭 → ② 기술통계량 클릭 → ③ 교차분석 클릭
→ ④ 행 변수 2개(성별, 직업) 선택 → ⑤ 열 변수 1개(소득) 선택 →

스CREENSHOT 1: SPSS Statistics Data Editor 메뉴판

- ① 클릭: '분석(A)' 메뉴 클릭
- ② 클릭: '기술통계량(E)' 하위 메뉴 클릭
- ③ 클릭: '교차분석(C)...' 메뉴 클릭

스CREENSHOT 2: 교차분석: 행과 열 대화 상자

- ④ 행 변수 2개 선택: '성별'과 '직업'을 행 변수로 선택
- ⑤ 열 변수 1개 선택: '소득'을 열 변수로 선택

대화 상자 내 다른 옵션:

- 정확(X)...
- 통계량(S)...
- 셀(E)...
- 형식(F)...
- 유형(L)...
- 북스트랩(A)...
- 표 레이어에 레이어 변수 표시(L) ☒
- 수평배열 막대형 차트 표시(B) ☐
- 교차표를 출력하지 않음(I) ☐
- 확인, 불어넣기(P), 재설정(R), 취소, 도움말

2. 교차분석

다. SPSS 실습

- 교차분석 실행

⑥ 통계량 클릭 → ⑦ 카이제곱 클릭 → ⑧ 계속 클릭
→ ⑨ 셀 클릭 → ⑩ 관측빈도 클릭 → ⑪ 행 클릭 → ⑫ 계속 클릭 → ⑬ 확인 클릭

The diagram illustrates the step-by-step process of performing a cross-tabulation analysis in SPSS, with numbered callouts (⑥ to ⑬) indicating the sequence of clicks across three different dialog boxes.

- Step ⑥:** In the main 'Cross-tabulation' dialog box, click on 'Statistics...' (통계량(S)...).
- Step ⑦:** In the 'Statistics' sub-dialog box, click on 'Chi-square' (카이제곱(χ²)).
- Step ⑧:** In the 'Statistics' sub-dialog box, click on 'Continue' (계속(C)).
- Step ⑨:** In the main 'Cross-tabulation' dialog box, click on 'Cells...' (셀(E)...).
- Step ⑩:** In the 'Cells' sub-dialog box, click on 'Observed' (관측빈도(O)).
- Step ⑪:** In the 'Cells' sub-dialog box, click on 'Row' (행(R)).
- Step ⑫:** In the 'Cells' sub-dialog box, click on 'Continue' (계속(C)).
- Step ⑬:** In the main 'Cross-tabulation' dialog box, click on 'OK' (확인).

2. 교차분석

다. SPSS 실습

● 교차분석 결과

성별 * 소득

교차분석								
		소득						
		100만원 미만	100~200만 이 내	200~300만 이 내	300~400만 이 내	400만원 이상	전체	
성별	남	빈도	32	45	37	12	7	133
		성별 중 %	24.1%	33.8%	27.8%	9.0%	5.3%	100.0%
여	빈도	27	43	25	11	11	117	
		성별 중 %	23.1%	36.8%	21.4%	9.4%	9.4%	100.0%
전체	빈도	59	88	62	23	18	250	
		성별 중 %	23.6%	35.2%	24.8%	9.2%	7.2%	100.0%

카이제곱 검정

	값	자유도	근사 유의확률 (양측검정)
Pearson 카이제곱	2.711 ^a	4	.607
우도비	2.722	4	.605
선형 대 선형결합	.277	1	.599
유효 케이스 수	250		

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 8.42입니다.

성별과 소득은 유의미한
관계가 나타나지 않음

직업 * 소득

교차분석								
			소득					
			100만원 미만	100~200만 이 내	200~300만 이 내	300~400만 이 내	400만원 이상	전체
직업	학생	빈도	37	24	14	5	8	88
		직업 중 %	42.0%	27.3%	15.9%	5.7%	9.1%	100.0%
	회사원/공무원	빈도	9	47	30	8	5	99
		직업 중 %	9.1%	47.5%	30.3%	8.1%	5.1%	100.0%
	자영업/사업	빈도	1	4	6	2	3	16
		직업 중 %	6.3%	25.0%	37.5%	12.5%	18.8%	100.0%
	전문직	빈도	4	3	3	1	1	12
		직업 중 %	33.3%	25.0%	25.0%	8.3%	8.3%	100.0%
	주부	빈도	1	2	7	5	0	15
		직업 중 %	6.7%	13.3%	46.7%	33.3%	0.0%	100.0%
	무직	빈도	1	1	1	1	0	4
		직업 중 %	25.0%	25.0%	25.0%	25.0%	0.0%	100.0%
	기타	빈도	6	7	1	1	1	16
		직업 중 %	37.5%	43.8%	6.3%	6.3%	6.3%	100.0%
	전체	빈도	59	88	62	23	18	250
		직업 중 %	23.6%	35.2%	24.8%	9.2%	7.2%	100.0%

카이제곱 검정

	값	자유도	근사 유의확률 (양측검정)
Pearson 카이제곱	64.164 ^a	24	<.001
우도비	63.256	24	<.001
선형 대 선형결합	1.116	1	.291
유효 케이스 수	250		

a. 22 셀 (62.9%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 .29입니다.

직업과 소득은
유의미한 관계가 나타남

2. 교차분석

라. DATA IN 실습

- 교차분석 실행

① 통계분석Reporting 클릭 → ② 교차분석 클릭
→ ③ 행 변수 2개(성별, 직업) 선택 → ④ 열 변수 1개(소득) 선택 → ⑤ 분석하기 클릭

① 클릭

보고서Reporting | **통계분석Reporting**

교차분석 | 집단비교분석 | 척도화분석 | 회귀분석

빈도분석 | **기술통계분석** | 교차분석

빈도+비율 | 백분율기준: ☒ 행비율(%) ☐ 열비율(%)

a=3 and c*0.3 >=1

R Syntax | 분석초기화 | **분석하기**

② 클릭

전체항목 8건

<input type="checkbox"/> 보기형 1.직원친절만족 (집단수:5)	<input type="checkbox"/> 보기형 2.서비스만족 (집단수:5)	<input type="checkbox"/> 숫자형 3.작년_고객만족도... (집단수:0)
<input type="checkbox"/> 숫자형 4.고객만족도 (집단수:0)	<input type="checkbox"/> 숫자형 6.연령 (집단수:0)	

③ 행 변수 2개 선택

행변수 2건

<input type="checkbox"/> 보기형 × 5.성별 (집단수:2)	<input type="checkbox"/> 보기형 × 7.직업 (집단수:7)
---	---

④ 열 변수 1개 선택

열변수 1건

<input type="checkbox"/> 보기형 × 8.소득 (집단수:5)

⑤ 클릭

2. 교차분석

라. DATA IN 실습

● 교차분석 결과

① 문서로 바로 저장

주요+상세 파일저장 | 파일저장 | 파일보관함저장

주요결과 | 상세결과

열 전체 비율보다 10% 이상 차이이면 진한파랑, 5~10% 미만이면 연한파랑, 0~5% 미만 또는 -0~-5% 미만이면 하얀색, -5~-10% 미만이면 연한빨강, -10% 이하이면 진한빨강으로 바탕색을 처리한다.

변수	구분	전체	소득					x ² (p)
			100만원 미만	100~200만 이내	200~300만 이내	300~400만 이내	400만원 이상	
성별	전체	250(100.0)	59(23.6)	88(35.2)	62(24.8)	23(9.2)	18(7.2)	x ² ≈ 2.711 p= 0.607
	남성	133(53.2)	32(24.1)	45(33.8)	37(27.8)	12(9.0)	7(5.3)	
	여성	117(46.8)	27(23.1)	43(36.8)	25(21.4)	11(9.4)	11(9.4)	x ² ≈ 64.164 p= 0.000***
직업	학생	88(35.2)	37(42.0)	24(27.3)	14(15.9)	5(5.7)	8(9.1)	
	회사원/공무원	99(39.6)	9(9.1)	47(47.5)	30(30.3)	8(8.1)	5(5.1)	
	자영업/사업	16(6.4)	1(6.3)	4(25.0)	6(37.5)	2(12.5)	3(18.8)	
	전문직	12(4.8)	4(33.3)	3(25.0)	3(25.0)	1(8.3)	1(8.3)	
	주부	15(6.0)	1(6.7)	2(13.3)	7(46.7)	5(33.3)	0(0.0)	
	무직	4(1.6)	1(25.0)	1(25.0)	1(25.0)	1(25.0)	0(0.0)	
	기타	16(6.4)	6(37.5)	7(43.8)	1(6.3)	1(6.3)	1(6.3)	

성별과 소득은 유의미한
관계가 나타나지 않으며,
직업과 소득은 유의미한
관계가 있는 것으로 나타남

소득에 대한 전체 응답결과, 총 250(100.0) 중 100만원 미만 59(23.6), 100~200만 이내 88(35.2), 200~300만 이내 62(24.8), 300~400만 이내 23(9.2), 400만원 이상 18(7.2)로 나타났다.

각 변수와 소득간의 교차분석 결과

성별과 소득간에는 유의미한 관계가 나타나지 않았다(x²= 2.711, p>0.05).

직업과 소득간에는 유의미한 관계가 나타났다(x²=64.164, p<0.05).

학생과 기타는 다른 집단에 비해 100만원 미만에 대한 응답비율이 높게 나타났다.

회사원/공무원은 다른 집단에 비해 100~200만 이내에 대한 응답비율이 높게 나타났다.

자영업/사업과 주부는 다른 집단에 비해 200~300만 이내에 대한 응답비율이 높게 나타났다.

주부와 무직은 다른 집단에 비해 300~400만 이내에 대한 응답비율이 높게 나타났다.

자영업/사업은 다른 집단에 비해 400만원 이상에 대한 응답비율이 높게 나타났다.

결과표 및 결과의 자동해석 제시!!

3. t-test

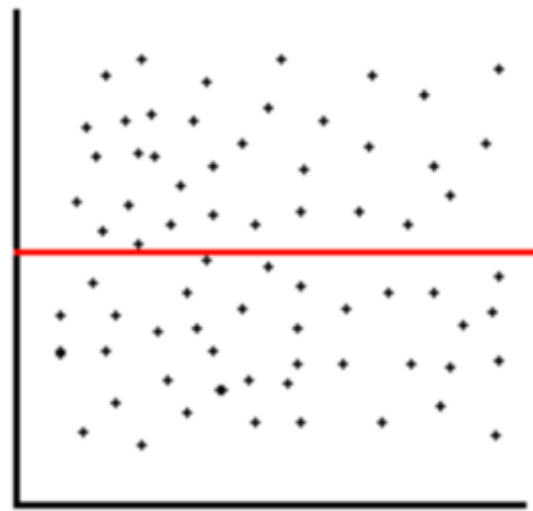
가. t-test 분석 종류

	독립표본 t-test	대응표본 t-test
정의	<ul style="list-style-type: none"> 두 집단 간의 평균 차이를 비교하는 검정 	<ul style="list-style-type: none"> 실험 이전의 집단과 실험 이후의 집단이 동일한 집단일 경우 사용하는 검정
모집단	<ul style="list-style-type: none"> 2개 	<ul style="list-style-type: none"> 1개
전제 조건	<ul style="list-style-type: none"> 독립성, 정규성, 등분산성 만족 	<ul style="list-style-type: none"> 정규성 만족
가설 예시	<ul style="list-style-type: none"> H0 : 남녀 고객간 고객만족도는 차이가 없다 	<ul style="list-style-type: none"> H0 : 작년 고객만족도와 올해 고객만족도 간에는 차이가 없다
	<ul style="list-style-type: none"> H1 : 남녀 고객간 고객만족도는 차이가 있다 	<ul style="list-style-type: none"> H1 : 작년 고객만족도와 올해 고객만족도 간에는 차이가 있다
검정통계량	<ul style="list-style-type: none"> t-value, 자유도: df 이를 고려하여 유의확률(p)의 계산 	<ul style="list-style-type: none"> t-value, 자유도: df 이를 고려하여 유의확률(p)의 계산

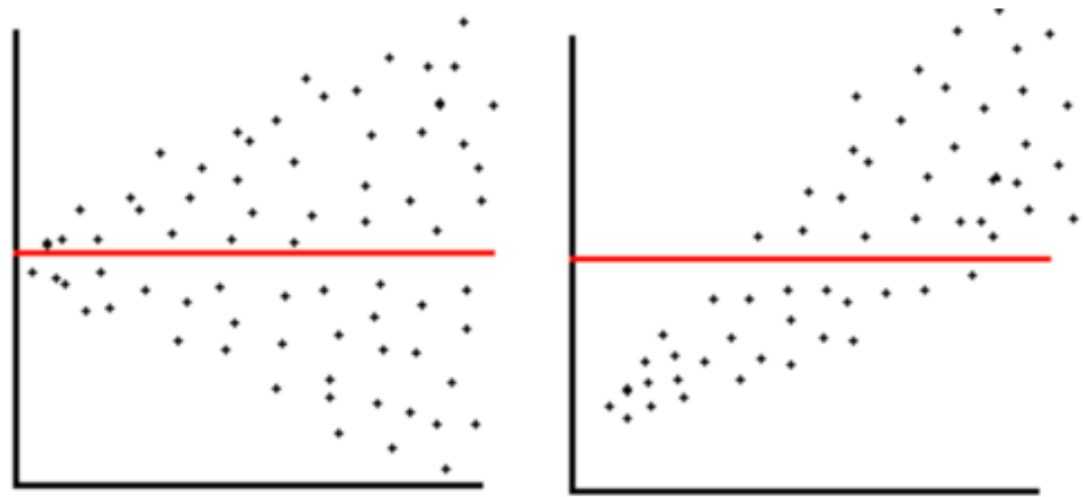
3. t-test

나. 등분산성과 이분산성

- 분산: 개별 케이스가 평균에서 이탈된 편차의 제곱합, 평균에서 개별 사례가 이탈된 정도
- 독립 t-검정의 경우 등분산성 여부에 따라 다른 값을 참조함



등분산성 충족



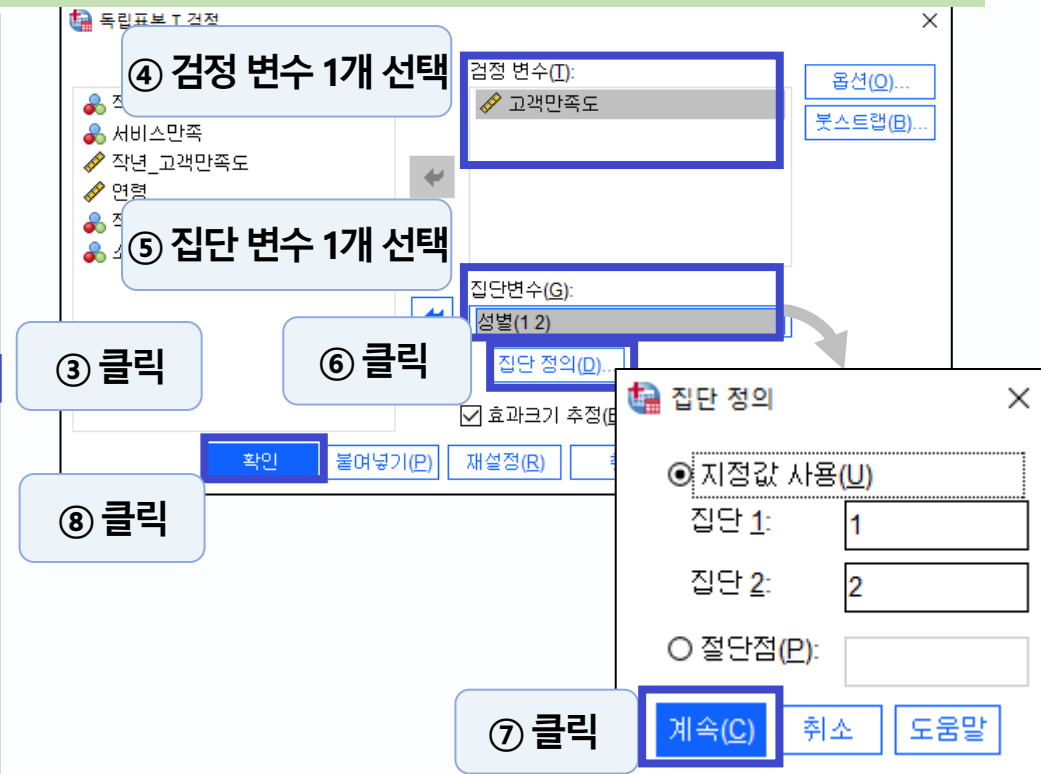
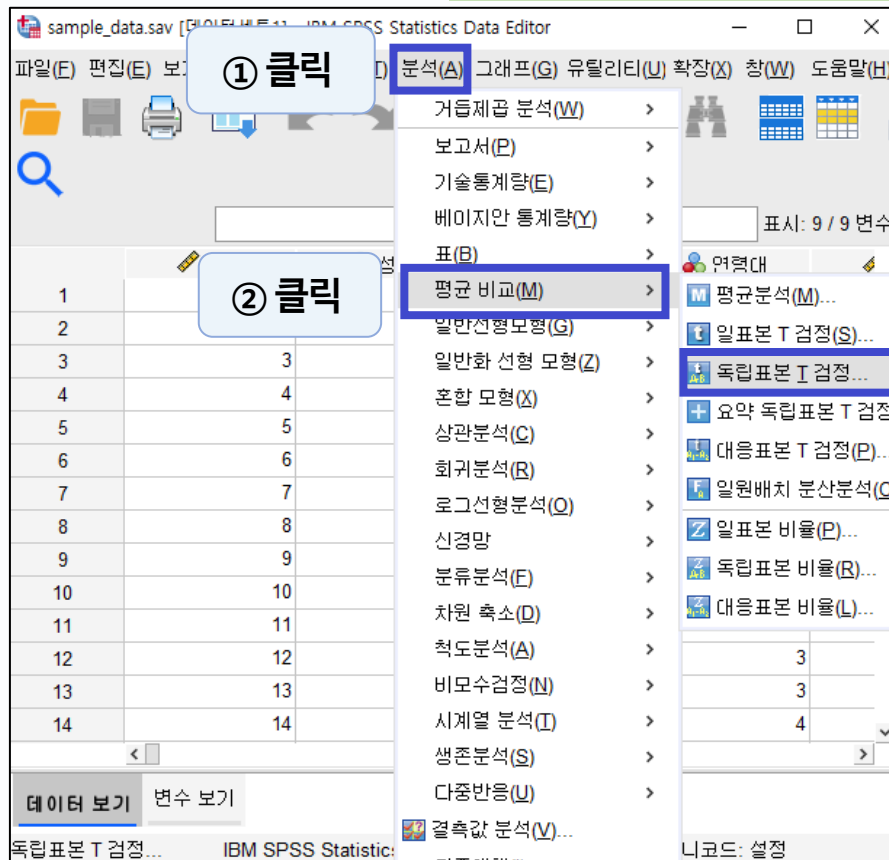
등분산성 미충족(이분산성)

3.1 독립표본 t-test

다. SPSS 실습

● 독립표본 t-test 실행

① 분석 클릭 → ② 평균비교 클릭 → ③ 독립표본 T검정 클릭 → ④ 검정 변수 1개(고객만족도) 선택 → ⑤ 집단변수 1개(성별)선택 → ⑥ 집단 정의 → ⑦ 계속 → ⑧ 확인 클릭



3.1 독립표본 t-test

다. SPSS 실습

- 독립표본 t-test 결과

집단통계량					
	성별	N	평균	표준편차	평균의 표준오차
고객만족도	남	133	3.03	.921	.080
	여	117	3.30	.931	.086

Levene 검정 결과 $p > 0.05$ 일 경우 등분산 가정					
Levene의 등분산 검정			독립표본 검정		
		F	유의확률	t	자유도
고객만족도	등분산을 가정함	2.071	.151	-2.294	248
	등분산을 가정하지 않음			-2.292	243.241

고객만족도는 성별 간
유의미한 차이가 있는 것으로 나타남

3.1 독립표본 t-test

라. DATA IN 실습

● 독립표본 t-test 실행

① 통계분석Reporting 클릭 → ② 집단비교분석 클릭 → ③ 독립T검정 클릭
→ ④ 독립변수 1개(성별) 선택 → ⑤ 종속변수 1개(고객만족도) 선택 → ⑥ 분석하기 클릭

① 클릭 통계분석Reporting

② 클릭 독립T검정

③ 클릭 독립변수 1개(성별) 선택

④ 독립변수 선택 후 이동

⑤ 종속변수 선택 후 이동

⑥ 클릭 분석하기

3.1 독립표본 t-test

라. DATA IN 실습

- 독립표본 t-test 결과

① 문서로 바로 저장

주요+상세 파일저장 ▾ 파일저장 ▾ 파일보관함저장

주요결과

상세결과

결과표 및 결과의 자동해석 제시!!

행기준 결과

	성별	N	평균	표준편차	t-value	p
고객만족도	남성	133	3.03	0.92	-2.294	0.023*
	여성	117	3.30	0.93		

* p<0.05 ** p<0.01 *** p<0.001

성별에 따른 각 변수에 대한 t-test분석 결과

고객만족도는 유의수준 0.05에서 유의한 차이가 나타났다(p<0.05). 평균을 보면 남성은 3.03, 여성은 3.30로 각각 나타났다.

고객만족도는 성별 간
유의미한 차이가 있는 것으로 나타남

3.2 대응표본 t-test

가. SPSS 실습

● 대응표본 t-test 실행

① 분석 클릭 → ② 평균비교 클릭 → ③ 대응표본 T검정 클릭
→ ④ 대응 변수 2개(고객만족도, 작년_고객만족도) 선택 → ⑤ 확인 클릭

The image shows the SPSS Statistics Data Editor and the '대응표본 T 검정(P)...' dialog box. The steps are numbered as follows:

- ① 클릭**: Click on the '분석(A)' menu.
- ② 클릭**: Click on '평균 비교(M)' in the '통계' (Statistics) submenu.
- ③ 클릭**: Click on '대응표본 T 검정(P)...' in the '통계' (Statistics) submenu.
- ④ 대응 변수 2개 선택**: Select '작년_고객만족도' and '고객만족도' in the '대응 변수(V):' list.
- ⑤ 클릭**: Click on the '확인' (OK) button.

The dialog box '대응표본 T 검정(P)...' shows the following settings:

- 대응 변수(V): 1. 작년_고객만족도, 2. 고객만족도
- 효과 크기 추정(E): ☒ 효과 크기 추정(E)
- Standardizer 계산 방법: ☒ 차이의 표준편차(S), ☐ 차이의 수정된 표준편차(C), ☐ 분산의 평균(A)
- Buttons: 옵션(O)..., 북스트랩(B)..., 확인, 불어넣기(I), 재설정(R), 취소, 도움말

3.2 대응표본 t-test

가. SPSS 실습

- 대응표본 t-test 결과

대응표본 통계량					
		평균	N	표준편차	평균의 표준오차
대응 1	작년_고객만족도	2.69	250	.955	.060
	고객만족도	3.16	250	.933	.059

대응표본 검정								
		대응차		차이의 95% 신뢰구간		t	자유도	유의확률 (양측)
		평균	표준편차	평균의 표준오차	하한	상한		
대응 1	작년_고객만족도 - 고객만족도	-.464	1.169	.074	-.610	-.318	-6.277	<.001

올해 고객만족도는 작년 고객만족도와
유의미한 차이가 있는 것으로 나타남

3.2 대응표본 t-test

나. DATA IN 실습

● 대응표본 t-test 실행

① 통계분석Reporting 클릭 → ② 집단비교분석 클릭 → ③ 대응T검정 클릭 → ④ 선택변수1 1개 (작년_고객만족도) 선택 → ⑤ 선택변수2 1개(고객만족도) 선택 → ⑥ 분석하기 클릭

① 클릭: 통계분석Reporting 클릭

② 클릭: 대응T검정 클릭

③ 클릭: 집단비교분석 클릭

④ 선택변수1 선택 후 이동: 3.작년_고객만족도... (집단수:0) 선택

⑤ 선택변수2 선택 후 이동: 4.고객만족도 (집단수:0) 선택

⑥ 클릭: 분석하기 클릭

3.2 대응표본 t-test

나. DATA IN 실습

- 대응표본 t-test 결과

- ① t-value: 값이 클수록 대응표본별 차이가 있음
- ② p(유의확률): 0.05보다 작으면 통계적으로 유의한 차이

① 문서로 바로 저장

주요+상세 파일저장

파일저장

파일보관함저장

주요결과

상세결과

결과표 및 결과의 자동해석 제시!!

행기준 결과

		N	평균	표준 편차	t-value	p
대응 1	작년_고객만족도	250	2.69	0.96	-6.277	0.000***
	고객만족도	250	3.16			

* p<0.05 ** p<0.01 *** p<0.001

각 변수간 대응표본 t-test분석 결과

작년_고객만족도와 고객만족도간에는 유의수준 0.05에서 유의한 차이가 나타났다(p<0.05). 작년_고객만족도는 평균 2.69, 고객만족도는 3.16로 나타났다.

올해 고객만족도는 작년 고객만족도와
유의미한 차이가 있는 것으로 나타남

4. 분산분석

가. 분산 분석의 종류

	일원배치 분산분석(ANOVA)	이원분산분석(Two-Way ANOVA)	공분산분석(ANCOVA)
정의	세 집단 이상간의 평균차이 분석	두 개의 독립변수에 따라 집단 간 종속변수의 평균차이 분석	외생변수를 공변량으로 처리한 후 각 그룹 사이의 종속변수 값에 차이가 있는지를 조사하는 방법
변수	<ul style="list-style-type: none"> 독립변수(3집단 이상) 종속변수 (연속형 변수) 	<ul style="list-style-type: none"> 2개의 독립변수 (범주형 변수) 종속변수 (연속형 변수) 	<ul style="list-style-type: none"> 독립변수 (범주형 변수) 종속변수 (연속형 변수) 공변량 (연속형 변수)
전제 조건	독립성, 정규성, 등분산성 만족	독립성, 정규성, 등분산성 만족	독립성, 정규성, 등분산성 만족
가설 예시	H0 : 고객의 직업에 따라 고객만족도는 차이가 없다.	H0 : 고객의 성별에 따라 고객만족도는 차이가 없다. H0 : 고객의 소득에 따라 고객만족도는 차이가 없다. H0 : 고객의 성별과 소득간의 상호작용 효과가 없을 것이다.	H0 : 고객의 소득에 따라 고객만족도는 차이가 없다.
	H1 : 고객의 직업에 따라 고객만족도는 차이가 있다.	H0 : 고객의 성별에 따라 고객만족도는 차이가 있다. H0 : 고객의 소득에 따라 고객만족도는 차이가 있다. H0 : 고객의 성별과 소득간의 상호작용 효과가 있을 것이다	H1 : 고객의 소득에 따라 고객만족도는 차이가 있다.
검정 통계량	F-value, 자유도: df 이를 고려하여 유의확률(p)의 계산	F-value, 자유도: df 이를 고려하여 유의확률(p)의 계산	F-value, 자유도: df 이를 고려하여 유의확률(p)의 계산

4. 분산분석

나. 사후분석

- 사용 목적

- ANOVA 분석 결과 집단 간의 차이가 유의하게 나타났을 때, 집단을 분리시켜 각각의 차이 정도를 보고자 할 때 사용
- t-테스트를 여러 번 진행해 어떤 집단이 차이가 있는지 확인 할 수도 있지만, 비교대상이 증가할 수록 신뢰수준이 낮아지는 문제가 발생함.
- 예를 들어, 하나의 집단에 대해서 차이가 있는지를 볼때에는 신뢰수준이 95%이지만, 신뢰수준이 95%인 두개의 집단의 결과가 동일할 확률은 $95\% * 95\% = 90.25\%$ 로 신뢰수준이 감소하게 됨
이 경우, 1종 오류가 증가하는 문제가 발생함

- 정규분포를 따르는 경우 사후검정의 분류(모수적 방법)

	표본 크기 동일	표본 크기 다를 경우
등분산	<ul style="list-style-type: none"> • Tukey • Dunnett • Duncan 	<ul style="list-style-type: none"> • Bonferroni • Scheffe
등분산 아니어도 됨		<ul style="list-style-type: none"> • Tamhane T2 • Dunnett T3 • Games-Howell

4. 분산분석

나. 사후분석

- 등분산일 경우 사용하는 사후분석 종류

	종류	특징	장점	단점
표본 크기 동일	Tukey	<ul style="list-style-type: none"> 비교 대상 표본수가 동일한 경우에만 사용 가능 모든 집단 조합에 대하여 분석 	<ul style="list-style-type: none"> 표본수가 동일한 경우 가장 많이 사용 	<ul style="list-style-type: none"> 비교 대상 표본수가 동일하여야 함 표본수가 적을수록 정확도가 낮아짐
	Dunnett	<ul style="list-style-type: none"> 하나의 집단을 기준으로 다른 집단들과 차이에 대하여 분석 양측 검정 가능 	<ul style="list-style-type: none"> 1개의 대조군과 여러 실험군과의 비교를 하는 연구에 사용 가능 Tukey보다 검정력 높음 	<ul style="list-style-type: none"> 모든 집단 조합에 대한 검정을 하지 않음
	Duncan	<ul style="list-style-type: none"> 상대적으로 엄격하지 않은 기준 인접하는 평균값들을 단계적으로 비교하는 방법 	<ul style="list-style-type: none"> 유의성을 도출하기 쉬움 	<ul style="list-style-type: none"> 1종 오류 위험 증가
표본 크기를 경우	Bonferroni	<ul style="list-style-type: none"> 모수, 비모수검정 적용 가능 Tukey보다 엄격하지만 Scheffe보다는 관대함 	<ul style="list-style-type: none"> ANOVA, 다중 t-test, 비모수 검정 등에 적용 가능 	<ul style="list-style-type: none"> 비교대상이 많아질수록 검정력이 약해짐
	Scheffe	<ul style="list-style-type: none"> 가장 보수적이고 엄격한 사후 검정 방식 	<ul style="list-style-type: none"> 엄격한 기준으로 사후 검정 실시 	<ul style="list-style-type: none"> 통계적으로 유의한 차이를 도출하기가 쉽지 않음

4. 분산분석

나. 사후분석

- 등분산이 아닐 경우 사용하는 사후분석 종류

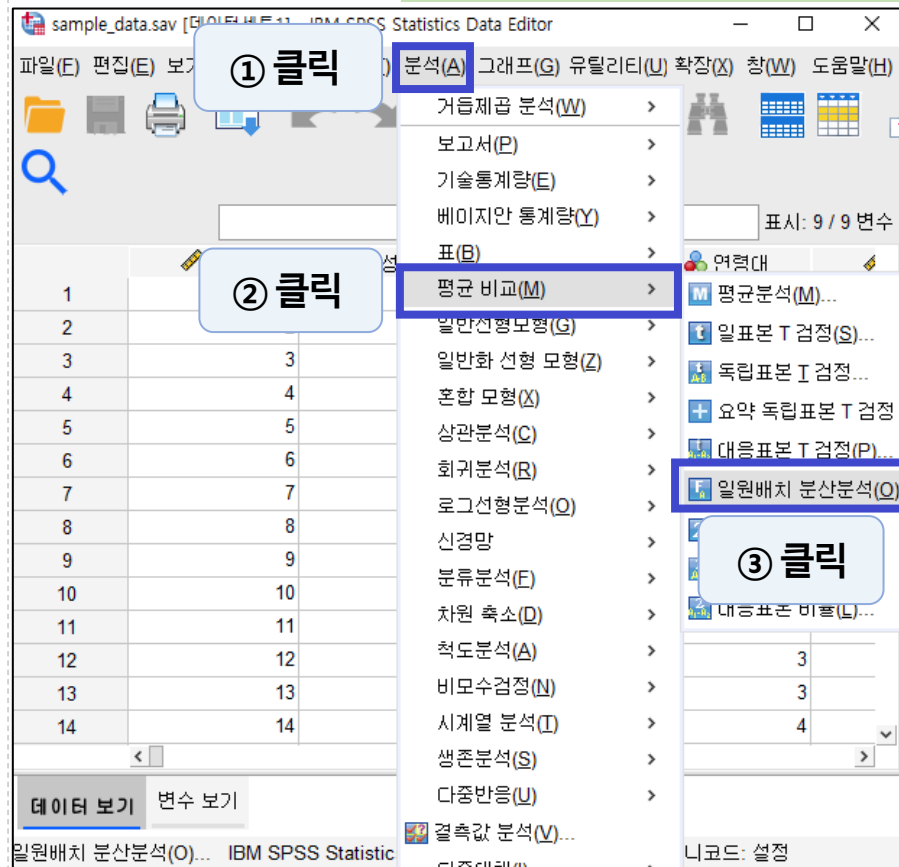
종류	특징	장점	단점
Tamhane T2	<ul style="list-style-type: none"> 유의수준 조정 및 t분포를 기준으로 분석 	<ul style="list-style-type: none"> Games-Howell보다 엄격한 기준 적용 가능 	<ul style="list-style-type: none"> 샘플이 많아질수록 1종오류 높아짐
Dunnett T3	<ul style="list-style-type: none"> 유의수준 조정 및 t분포를 기준으로 분석 	<ul style="list-style-type: none"> 집단별 표본수가 50개 미만인 경우 Games-Howell보다 검정력 우수 	<ul style="list-style-type: none"> 집단별 표본수가 50개 이상일 경우 Games-Howell보다 1종오류 높아짐
Games-Howell	<ul style="list-style-type: none"> Welch 분석 기법을 응용 	<ul style="list-style-type: none"> 집단별 표본수가 다르거나 분산의 동질성이 확보되지 않아도 적용 가능 	<ul style="list-style-type: none"> 표본수가 6개 미만일 경우 1종오류 발생률 높아짐(15개 이상 권장)

4.1 일원배치 분산분석

가. SPSS 실습

● 분산분석 실습

① 분석 클릭 → ② 평균비교 클릭 → ③ 일원배치 분산분석 클릭
→ ④ 종속 변수 1개(고객만족도) 선택 → ⑤ 요인 1개(직업)선택



4.1 일원배치 분산분석

가. SPSS 실습

● 분산분석 실습

⑥ 사후분석 클릭 → ⑦ Scheffe 클릭 → ⑧ 계속 클릭 → ⑨ 옵션 클릭 → ⑩ 기술통계 클릭
→ ⑪ 분산 동질성 검정 클릭 → ⑫ 계속 클릭 → ⑬ 확인 클릭

The image shows three sequential screenshots of the SPSS One-Way ANOVA procedure, with numbered steps indicating the sequence of actions:

- Step 6:** In the '일원배치 분산분석' (One-Way ANOVA) dialog box, the '종속변수(E):' (Dependent Variable) is '고객만족도' (Customer Satisfaction). The '요인(F):' (Factor) is '직업' (Occupation). The '확인' (OK) button is highlighted.
- Step 7:** In the '일원배치 분산분석: 사후분석 - 다중비교' (One-Way ANOVA: Post Hoc - Multiple Comparisons) sub-dialog box, the '등분산을 가정함' (Assume equal variances) section has 'Scheffe' selected. The '계속(C)' (Continue) button is highlighted.
- Step 8:** In the '일원배치 분산분석: 옵션' (One-Way ANOVA: Options) sub-dialog box, the '통계량' (Statistics) section has '기술통계(D)' (Descriptive Statistics) and '분산 동질성 검정(H)' (Test of Homogeneity of Variances) selected. The '계속(C)' (Continue) button is highlighted.
- Step 9:** In the '일원배치 분산분석' (One-Way ANOVA) dialog box, the '확인' (OK) button is highlighted.
- Step 10:** In the '일원배치 분산분석: 옵션' (One-Way ANOVA: Options) sub-dialog box, the '통계량' (Statistics) section has '기술통계(D)' (Descriptive Statistics) and '분산 동질성 검정(H)' (Test of Homogeneity of Variances) selected. The '계속(C)' (Continue) button is highlighted.
- Step 11:** In the '일원배치 분산분석: 옵션' (One-Way ANOVA: Options) sub-dialog box, the '통계량' (Statistics) section has '기술통계(D)' (Descriptive Statistics) and '분산 동질성 검정(H)' (Test of Homogeneity of Variances) selected. The '계속(C)' (Continue) button is highlighted.
- Step 12:** In the '일원배치 분산분석: 옵션' (One-Way ANOVA: Options) sub-dialog box, the '통계량' (Statistics) section has '기술통계(D)' (Descriptive Statistics) and '분산 동질성 검정(H)' (Test of Homogeneity of Variances) selected. The '계속(C)' (Continue) button is highlighted.
- Step 13:** In the '일원배치 분산분석' (One-Way ANOVA) dialog box, the '확인' (OK) button is highlighted.

4.1 일원배치 분산분석

가. SPSS 실습

- 분산분석 결과

분산의 동질성 검정					
		Levene 통계량	df1	df2	유의확률
내가 이용한 영화관에 대해 전반적으로 만족한다	평균을 기준으로 합니다.	.343	6	243	.913
	중위수를 기준으로 합니다.	.273	6	243	.949
	자유도를 수정한 상태에서 중위수를 기준으로 합니다.	.273	6	229.509	.949
	절삭평균을 기준으로 합니다.	.318	6	243	.927

유의확률이 0.05보다 크기 때문에 Levene 등분산의 귀무가설 “집단의 분산은 동일하다” 기각할 수 없음
따라서, 고객의 직업은 등분산 가정을 충족

ANOVA					
내가 이용한 영화관에 대해 전반적으로 만족한다					
	제곱합	자유도	평균제곱	F	유의확률
집단-간	66.044	6	11.007	14.495	<.001
집단-내	184.532	243	.759		
전체	250.576	249			

유의확률이 <0.001로 0.05보다 작기 때문에, 고객의 직업별 고객만족도에 유의미한 차이가 있는 것으로 나타남

4.1 일원배치 분산분석

가. SPSS 실습

● 분산분석 결과

다중비교						
종속변수: 내가 이용한 영화관에 대해 전반적으로 만족한다						
Scheffe						
(I) 귀하의 직업은?	(J) 귀하의 직업은?	평균차이(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한	상한
학생	회사원/공무원	-1.155	.128	<.001	-1.61	-.70
	자영업/사업	-.585	.237	.414	-1.43	.26
	전문직	-.898	.268	.087	-1.86	.06
	주부	-1.064	.243	.005	-1.94	-.19
	무직	-.648	.446	.908	-2.24	.95
	기타	-.773	.237	.105	-1.62	.08
회사원/공무원	학생	1.155	.128	<.001	.70	1.61
	자영업/사업	.570	.235	.438	-.27	1.41
	전문직	.258	.266	.988	-.70	1.21
	주부	.091	.241	1.000	-.77	.96
	무직	.508	.444	.971	-1.08	2.10
	기타	.383	.235	.850	-.46	1.22
자영업/사업	학생	.585	.237	.414	-.26	1.43
	회사원/공무원	-.570	.235	.438	-1.41	.27
	전문직	-.313	.333	.990	-1.50	.88
	주부	-.479	.313	.885	-1.60	.64
	무직	-.063	.487	1.000	-1.81	1.68
	기타	-.188	.308	.999	-1.29	.92

Scheffe의 사후검정 결과 $\alpha=0.05$ 에서

학생과 회사원/공무원, 학생과 주부간의 유의미한 차이가 있는 것으로 나타남.

즉, 회사원/공무원과 주부는 학생보다 고객만족도가 높다고 할 수 있음

전문직	학생	.898	.268	.087	-.06	1.86
	회사원/공무원	-.258	.266	.988	-1.21	.70
	자영업/사업	.313	.333	.990	-.88	1.50
	주부	-.167	.338	1.000	-1.37	1.04
	무직	.250	.503	1.000	-1.55	2.05
	기타	.125	.333	1.000	-1.07	1.32
주부	학생	1.064	.243	.005	.19	1.94
	회사원/공무원	-.091	.241	1.000	-.96	.77
	자영업/사업	.479	.313	.885	-.64	1.60
	전문직	.167	.338	1.000	-1.04	1.37
	무직	.417	.490	.994	-1.34	2.17
	기타	.292	.313	.990	-.83	1.41
무직	학생	.648	.446	.908	-.95	2.24
	회사원/공무원	-.508	.444	.971	-2.10	1.08
	자영업/사업	.063	.487	1.000	-1.68	1.81
	전문직	-.250	.503	1.000	-2.05	1.55
	주부	-.417	.490	.994	-2.17	1.34
	기타	-.125	.487	1.000	-1.87	1.62
기타	학생	.773	.237	.105	-.08	1.62
	회사원/공무원	-.383	.235	.850	-1.22	.46
	자영업/사업	.188	.308	.999	-.92	1.29
	전문직	-.125	.333	1.000	-1.32	1.07
	주부	-.292	.313	.990	-1.41	.83
	무직	.125	.487	1.000	-1.62	1.87

*. 평균차이는 0.05 수준에서 유의합니다.

4.1 일원배치 분산분석

나. DATA IN 실습

● 분산분석 실행

① 통계분석Reporting 클릭 → ② 집단비교분석 클릭 → ③ 일원배치분산분석 클릭 → ④ scheffe 클릭
→ ⑤ 독립변수 1개(직업) 선택 → ⑥ 종속변수 1개(고객만족도) 선택 → ⑦ 분석하기 클릭

① 클릭

② 클릭

③ 클릭

④ 클릭

⑤ 독립변수 선택 후 이동

⑥ 종속변수 선택 후 이동

⑦ 클릭

R Syntax 분석초기화 **분석하기**

전체선택 선택취소

전체항목 8건

<input type="checkbox"/> 보기형 1.직원친절만족 (집단수:5)	<input type="checkbox"/> 보기형 2.서비스만족 (집단수:5)	<input type="checkbox"/> 숫자형 3.작년_고객만족도... (집단수:0)
<input type="checkbox"/> 보기형 5.성별 (집단수:2)	<input type="checkbox"/> 숫자형 6.연령 (집단수:0)	<input type="checkbox"/> 보기형 8.소득 (집단수:5)

독립변수 1건

☐ 보기형 ×
7.직업
(집단수:7)

종속변수 1건

☐ 숫자형 ×
4.고객만족도
(집단수:0)

영역초기화 선택삭제

4.1 일원배치 분산분석

나. DATA IN 실습

- 분산분석 결과

문서로 바로 저장

주요+상세 파일저장 ▾

파일저장 ▾

파일보관할저장

주요결과

상세결과

유의확률이 0.000으로 0.05보다 작기 때문에,
고객의 직업별 고객만족도에
유의미한 차이가 있는 것으로 나타남

행기준 결과

변수	구분	N	평균	표준편차	F-value	p
	학생	88	2.60 ^b	0.84	14.495	0.000***
	회사원/공무원	99	3.76 ^a	0.92		
		16	3.19 ^{ab}	1.05		
		12	3.50 ^{ab}	0.67		
		15	3.67 ^a	0.72		
	무직	4	3.25 ^{ab}	0.96		
	기타	16	3.38 ^{ab}	0.81		

같은 집단에 속하면 같은 알파벳이 표기
집단을 공유하지 않으면 통계적 차이 존재

* p<0.05 ** p<0.01 *** p<0.001

결과표 및 결

직업별 각 변수에 차이 분산분석을 이용해 분석한 결과
고객만족도는 유의수준 0.05에서 유의한 차이가 나타났다(p<0.05). 응답평균을 보면 학생은 2.60, 회사원/공무원은 3.76, 자영업/사업은 3.19, 전문직은 3.50, 주부는 3.67, 무직은 3.25, 기타는 3.38로 각각 나타났다.

유의확률이 0.000으로 0.05보다 작기 때문에,
고객의 직업별 고객만족도에
유의미한 차이가 있는 것으로 나타남

같은 집단에 속하면 같은 알파벳이 표기
집단을 공유하지 않으면 통계적 차이 존재

결과표 및 결과의 자동해석 제시!!

Part
04

분석실습2: 상관과 영향관계분석

1. 상관관계분석

가. 분석 및 자료의 특성

- 두 연속형 변수 간의 상관 정도를 파악
- 상관계수 (-1~+1)를 통해 표준화된 상관관계를 도출
- 주로 회귀분석을 하기 전에 독립변수 ↔ 종속변수 간의 상관이 있는지 탐색적 목적

요인	공부시간	지능지수	성적
공부시간	1	.123	.435
지능지수	.123	1	.343
성적	.435	.343	1

나. 상관관계분석의 가설 및 검정통계량

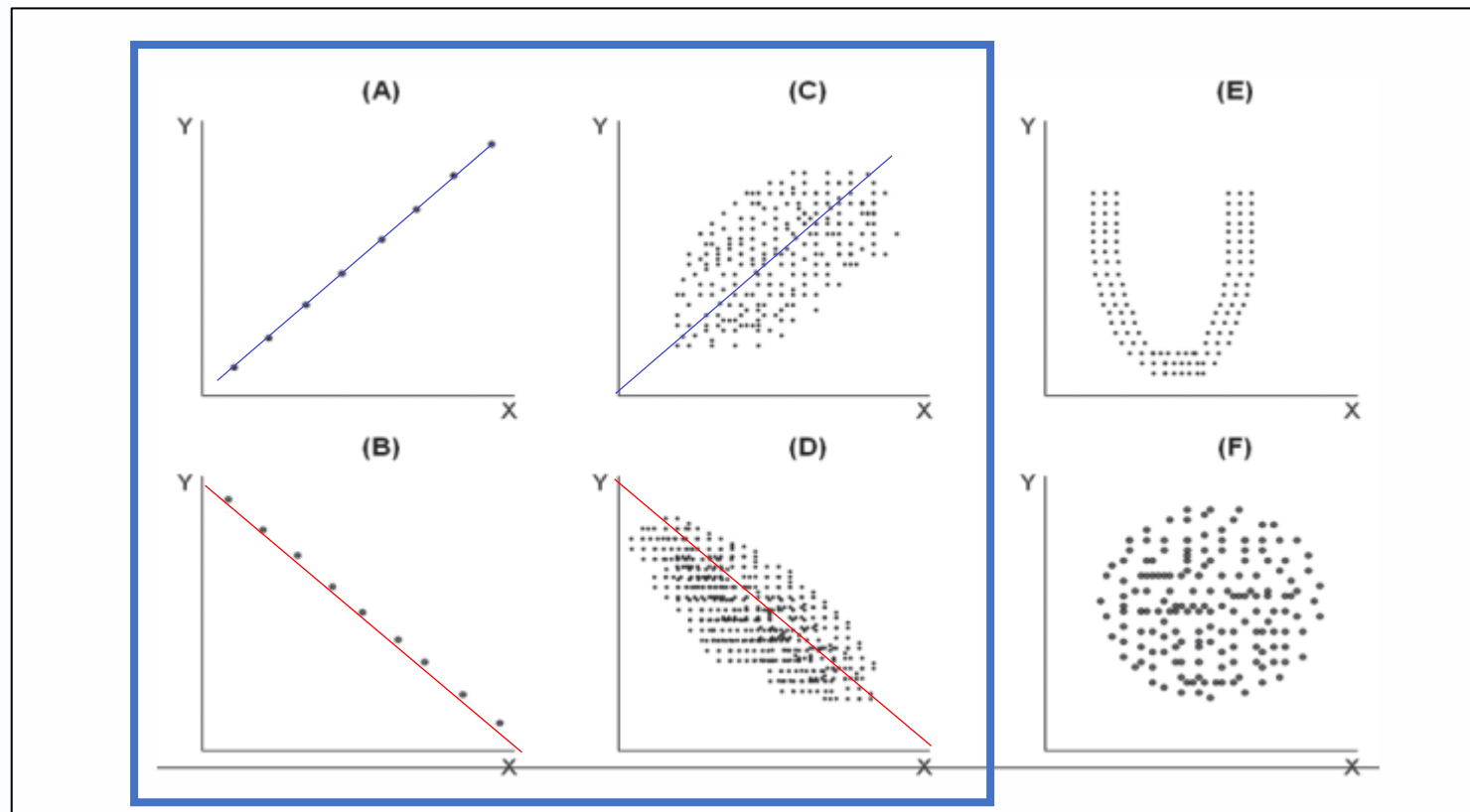
- 귀무가설: 공부시간/지능지수와 성적 간에는 상관관계가 없다.
- 대립가설: 공부시간/지능지수와 성적 간에는 상관관계가 있다.
- 검정통계량: 상관계수(r), 자료수(n)를 고려하여 유의확률(p)의 계산

1. 상관관계분석

다. 상관관계 예시

- 이 중 상관관계가 있는 그래프는 무엇일까요?

Pearson 상관관계는 선형성을 가정

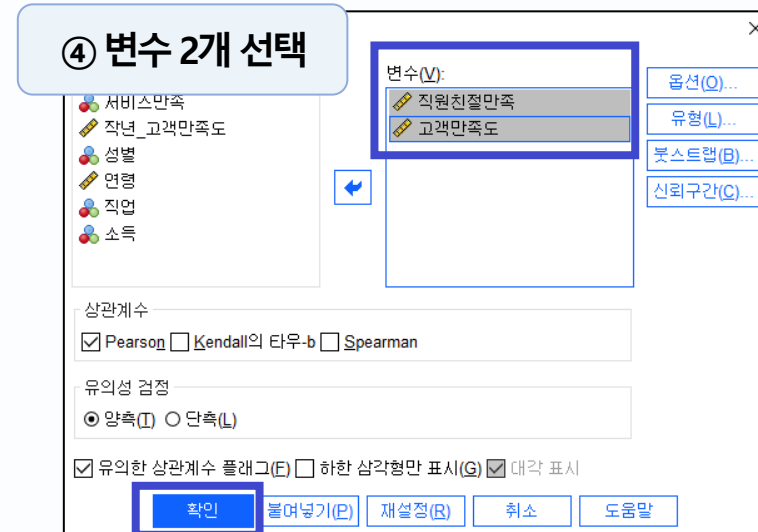
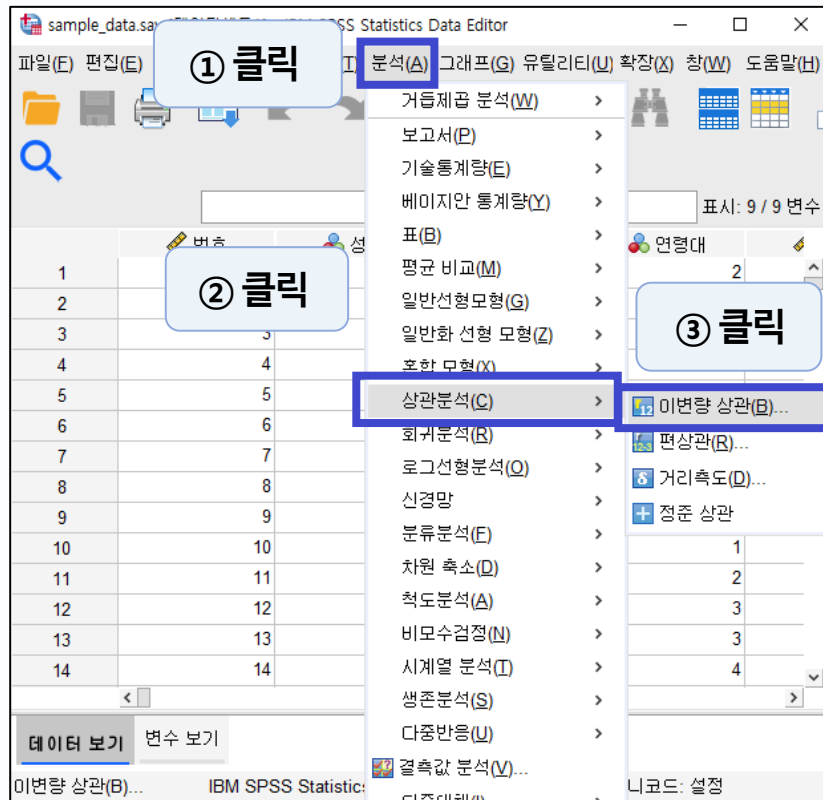


1. 상관관계분석

라. SPSS 실습

- 상관관계분석 실습

① 분석 클릭 → ② 상관분석 클릭 → ③ 이변량 상관 클릭
→ ④ 변수 2개(직원친절만족, 고객만족도) 선택 → ⑤ 확인 클릭



1. 상관관계분석

라. SPSS 실습

- 상관관계분석 결과

상관관계			
		직원친절만족	고객만족도
직원친절만족	Pearson 상관	1	.268**
	유의확률 (양측)		<.001
	N	250	250
고객만족도	Pearson 상관	.268**	1
	유의확률 (양측)	<.001	
	N	250	250

** . 상관관계가 0.01 수준에서 유의합니다(양측).

고객만족도와 직원친절만족은 상관계수가 0.268이며, 유의확률이 <.001로 상관관계를 보이고 있는 것으로 나타남

1. 상관관계분석

마. DATA IN 실습

- 상관관계분석 실행

① 통계분석Reporting 클릭 → ② 회귀분석 클릭 → ③ 상관관계분석 클릭
→ ④ 변수 2개(직원친절만족, 고객만족도) 선택 → ⑤ 분석하기 클릭

보고서Reporting | **통계분석Reporting** | ① 클릭

단변량분석 | 집단비교분석 | 척도화분석 | **상관관계분석** | ② 클릭

상관계수: ☒ Pearson ☐ Kendall ☐ Spearman | 평균/표준편차 | ③ 클릭

케이스선택 | 케이스취소 | a=3 and c*0.3 >=1

상관관계분석

- 선형회귀분석
- 위계회귀분석
- 매개효과분석
- 조절효과분석
- 이항로짓회귀분석
- 다항로짓회귀분석
- 경로분석
- 구조방정식

R Syntax | 분석초기화 | **분석하기** | ⑤ 클릭

전체할록 8건 | 전체선택

<input type="checkbox"/> 보기형 2.서비스만족 (집단수:5)	<input type="checkbox"/> 숫자형 3.작년_고객만족도... (집단수:0)	<input type="checkbox"/> 보기형 5.성별 (집단수:2)
<input type="checkbox"/> 숫자형 6.연령 (집단수:0)	<input type="checkbox"/> 보기형 7.직업 (집단수:7)	<input type="checkbox"/> 보기형 8.소득 (집단수:5)

④ 변수 선택 후 이동

1.직원친절만족
(집단수:5)

4.고객만족도
(집단수:0)

1. 상관관계분석

마. DATA IN 실습

- 상관관계분석 실행

① 문서로 바로 저장

주요+상세 파일저장 ▾ 파일저장 ▾ 파일보관함저장

주요결과

상세결과

변수	평균	표준편차	직원 친절 만족
직원 친절 만족	3.42	0.88	1
고객만족도	3.16	0.93	0.270***

* p<0.05 ** p<0.01 *** p<0.001

상관관계 분석결과

모든 변수간에 유의한 상관관계를 보이고 있는 것으로 나타났다(p<0.05).

또한 독립변수간의 상관관계는 모두 유의하지만 최대 상관계수가 0.270으로서 회귀분석 과정에서 나타날 수 있는 다중공선성 수준이 0.85보다 낮게 나타나고 있어 이러한 문제는 발생하지 않을 것으로 판단된다.

결과표 및 결과의 자동해석 제시!!

고객만족도와 직원친절만족은 상관계수가 0.270이며, 유의확률이 <.001로 상관관계를 보이고 있는 것으로 나타남



THANK YOU

(주)와이즈인컴퍼니 / 서울시 강남구 언주로 309, 기성빌딩 3층 / T 02.558.5144 / F 02.558.5146