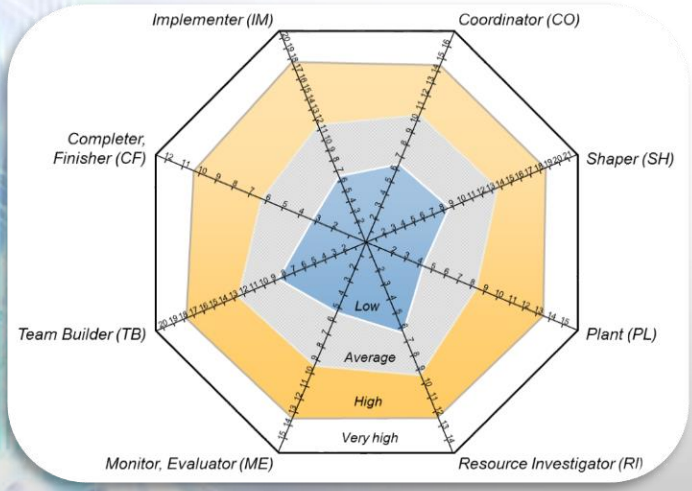
Part-III. 추론통계 분석

10. 분석절차와 통계지식
11. 기술통계 분석
12. 교차분석과 Chi-square 분석
13. 집단 간 차이 분석
14. 요인분석과 상관분석

10-1. 통계분석 절차



단계0. 연구조사
단계1. 가설설정
단계2. 유의수준 결정
단계3. 측정도구 선정
단계4. 데이터 수집
단계5. 데이터 코딩
단계6. 통계분석 수행
단계7. 결과분석

I. 통계분석 절차

- 논문/보고서 작성을 위한 통계분석 절차

1. 가설설정
2. 유의수준 결정
3. 측정도구 선정
4. 데이터 수집(설문지,웹,SNS)
5. 데이터 코딩/프로그래밍
6. 통계분석 수행(R,SPSS,SAS)
7. 결과분석(논문/보고서 작성)

단계0. 연구조사

● 가설 설정 이전의 연구조사

연구문제 선정

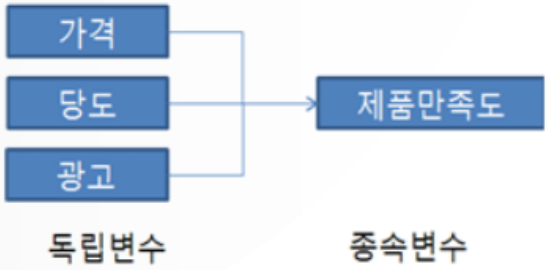
선정기준 : 연구의 독창성, 검증가능성, 결과의 실용성, 구체성, 경제성(소요비용, 노력)

예비조사

연구문제에 대한 사전 지식 획득, 변수 규명, 가설 도출 등을 위해 실시하는 조사
(예: 문헌 조사, 전문가 조사(파일럿 조사), 사례 조사 등)
☞ **사전조사** : 설문지를 작성 후 본 조사 전에 실시한다.

연구모형

연구문제에서 개념과 변수를 식별하는 과정으로 독립변수와 종속변수 형태로 기술한다.(예: '음료수 제품 만족도'에 대한 조사연구에서 연구모형은 다음과 같다.)



단계1. 가설 설정

- **가설(Hypothesis)**
  - 어떤 사실을 설명하기 위해서 설정한 가정
  - 사회 조사.연구에서 주어진 연구 문제에 대한 예측적 해답
  - 실증적인 증명에 앞서 세우는 잠정적인 진술
  - 나중에 논리적으로 검정될 수 있는 명제
  - 통계분석을 통해서 채택 또는 기각

※ 과학적 연구에서 가설의 설정은 매우 중요

단계1. 가설 설정

- **가설의 유형**
  - ① 귀무가설(영가설)

    ' 두 변수간의 관계가 없다.' 또는 '차이가 없다.'

    ✓ 부정적 형태 진술(예, H_0 : 교육수준에 따라서 사회 정책에 대한 비판적 태도에서 차이가 없다.)
  - ② 연구가설(대립가설)

    '차이가 있다.' 또는 '효과가 있다.'

    ✓ 긍정적 형태 진술(예, H_1: 영양소별 효과의 차이는 있다.)

※ 논문에서 **연구가설 제시**, 귀무가설을 통해서 가설 검정

단계2. 유의수준과 임계값 결정(1/3)

H_1 = '신약A는 A암 치료에 효과가 있다.'

H_0 = '신약A는 A암 치료에 효과가 없다.'

- **분석결과 : 생쥐 100마리를 대상으로 신약A를 투약한 결과 검정통계량의 유의확률(P=0.03)이 나왔다.**
  - 이때 귀무가설은 기각되는가? ➔ **YES**

- **사회과학분야 임계값 : α=0.05(p<0.05(5%미만))**
  - 적어도 96마리 이상 효과 ➔ **H_1 채택**
- **의.생명분야 임계값 : α=0.01(99% 신뢰도 보장)**
  - 적어도 99마리 이상 효과 ➔ **H_1 채택**

단계2. 유의수준과 임계값 결정(2/3)

- **유의수준(Signigicant level)**
  - 가설 채택 또는 기각 기준
  - 분석 결과 유의수준 이내 ➔ 가설 채택(그렇지 않으면 기각)
  - α(알파) 표시
  - 유의수준의 임계값(기준값) 결정
    - ✓ 일반 사회과학분야 : α=0.05(p<0.05) 기준
    - ✓ α=0.05 : 통계치가 모수치를 대표하는 허용 오차 5%(신뢰도 95%) (예, 100번 가운데 5번 미만 나올 확률)
  - 의생명분야 : 오차범위 최소 α=0.01(1% 오차 허용, 99% 신뢰도 확보)

단계2. 유의수준과 임계값 결정(3/3)

- **유의수준 α와 P값 관계**

α > P값 : 연구가설 채택(귀무가설 기각)
α ≦ P값 : 연구가설 기각(귀무가설 채택)

- 귀무가설(H_0) : '영양소별 효과의 차이는 없다'에서 임계값(α=0.05) 일때 가설 검정 결과 확률(p값) 0.04가 나왔다면
  p(0.04) < α(0.05) ➔ 귀무가설(영가설) 기각
- 영양소별 효과의 차이가 있을 확률이 높기 때문에 연구가설 채택
- 이때 통계적으로 유의하다라고 해석, p<0.01이면 매우 유의하다.
  p<0.05 수준이면 통계적으로 유의적인 차이를 보인다.
  '귀무가설이 의심스럽다'는 의미

단계3. 측정도구 선정

- **측정도구 선정**
  - 가설에 나오는 변수를 무엇으로 측정할 것인가를 결정하는 단계
  - 가설에 나오는 변수(변인) 추출
  - 변수의 척도를 고려 측정도구 선정

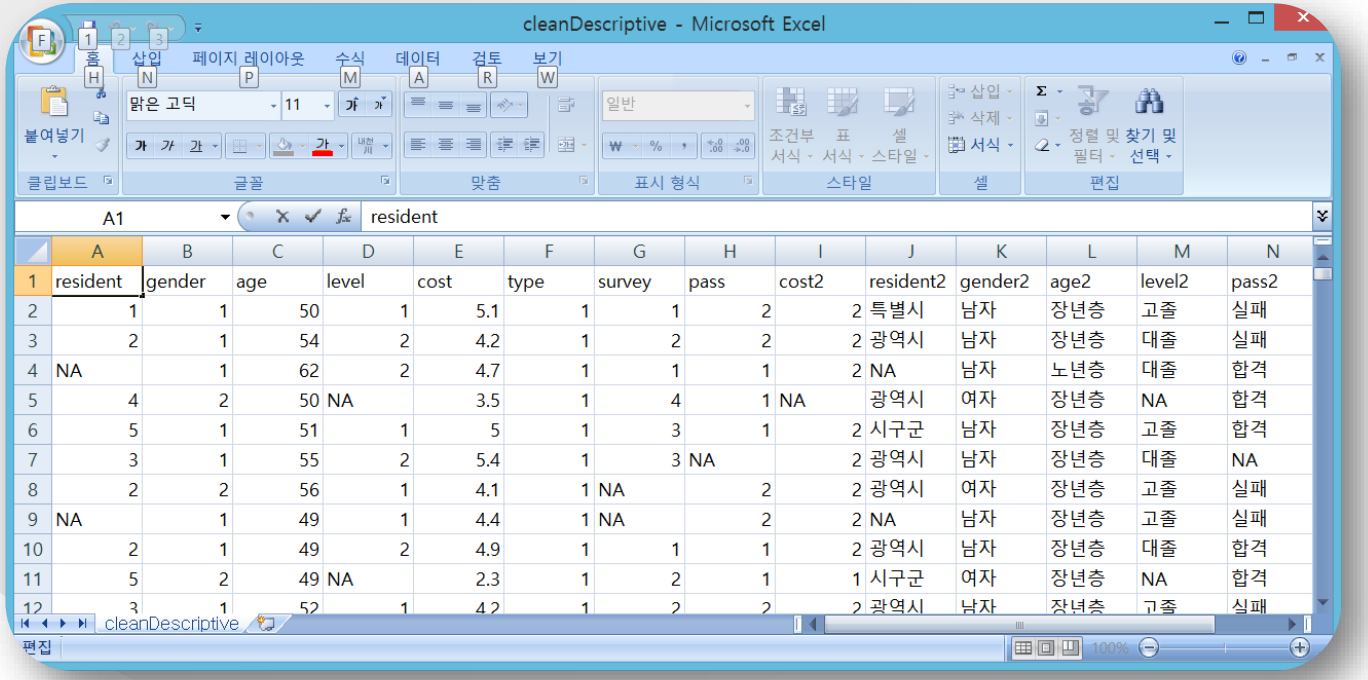
  ▶ 【척도(Scale)】 참조

단계4. 데이터 수집

- **데이터 수집(설문지 작성)**
  - 선정된 측정도구를 이용하여 설문 문항 작성 단계
  - 조사응답자 대상 설문 실시 & 회수
  - 정형/비정형 데이터 수집(DB, WEB, SNS 등)
  - 본 단계까지 완료된 경우
    - ✓연구목적과 배경, 연구모형, 연구가설까지 끝난 상태
    - ➔ **논문 50% 이상 완성**

단계5. 데이터 코딩

- **데이터(설문지) 코딩**
  - 통계분석 프로그램(Excel, R, SPSS, SAS,) 데이터 입력
  - 데이터 전처리(미 응답자, 잘못된 데이터 처리)

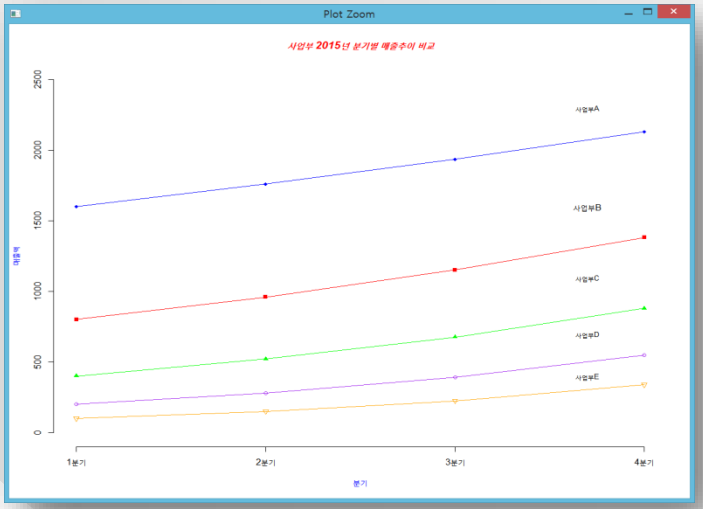


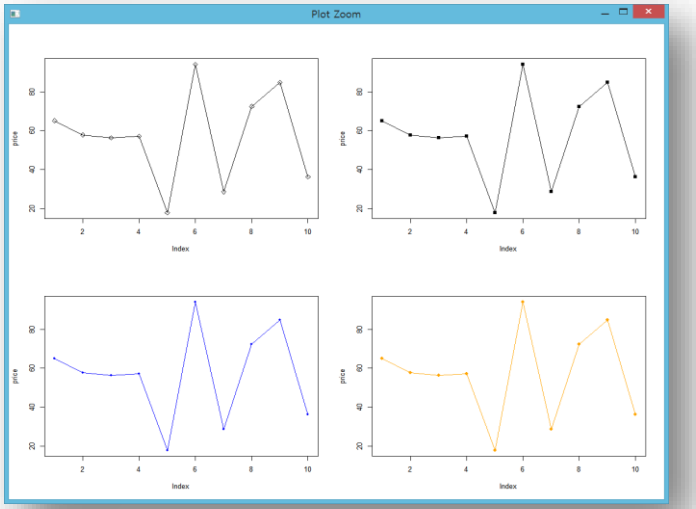
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	resident	gender	age	level	cost	type	survey	pass	cost2	resident2	gender2	age2	level2	pass2
2	1	1	50	1	5.1	1	1	2	2	특별시	남자	장년층	고졸	실패
3	2	1	54	2	4.2	1	2	2	2	광역시	남자	장년층	대졸	실패
4	NA	1	62	2	4.7	1	1	1	2	NA	남자	노년층	대졸	합격
5	4	2	50	NA	3.5	1	4	1	NA	광역시	여자	장년층	NA	합격
6	5	1	51	1	5	1	3	1	2	시구군	남자	장년층	고졸	합격
7	3	1	55	2	5.4	1	3	NA	2	광역시	남자	장년층	대졸	NA
8	2	2	56	1	4.1	1	NA	2	2	광역시	여자	장년층	고졸	실패
9	NA	1	49	1	4.4	1	NA	2	2	NA	남자	장년층	고졸	실패
10	2	1	49	2	4.9	1	1	1	2	광역시	남자	장년층	대졸	합격
11	5	2	49	NA	2.3	1	2	1	1	시구군	여자	장년층	NA	합격
12	3	1	52	1	4.2	1	2	2	2	광역시	남자	장년층	고졸	실패

단계6. 통계분석 수행

● 통계분석 수행

- ➢ 전문 통계분석 프로그램(R, SPSS, SAS) 분석 단계
- ❖ 통계분석 방법을 계획하지 않고 데이터를 수집할 경우 실패 확률 높음

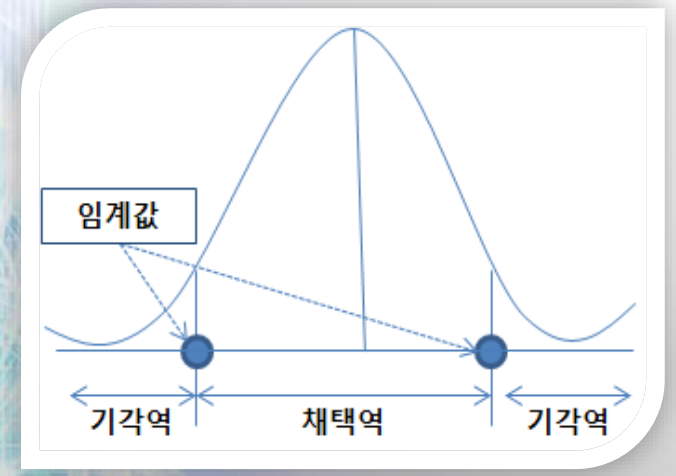




단계7. 결과분석

- **결과분석 제시**
  - 연구목적과 연구가설에 대한 분석 및 검증 단계
  - 인구통계학적 특성 반영
  - 주요 변인에 대한 기술통계량 제시
  - 연구가설에 대한 통계량 검정 및 해석
  - 연구자 의견 기술(논문/보고서 작성)

10-2. 통계 사전 지식



1) 통계학 개요
2) 모집단과 표본
3) 추정과 검정
4) 가설검정 오류
5) 검정통계량
6) 정규분포
7) 모수 & 비모수

1) 통계학 개요

- **통계학(Statistics)?**
  - ✓ 논리적 사고와 객관적인 사실에 의거, 확률 기반 인과관계 규명
  - ✓ 특히 연구목적에 의해 설정된 가설들에 대하여 분석결과가 어떤 결과를 뒷받침하고 있는지를 통계적 방법으로 검정.
  - ✓ 사회학, 경제학, 경영학, 정치학, 교육학, 공학, 의.생명 등 대부분의 모든 학문 분야에서 폭넓게 이용

구분	기술(Descriptive) 통계학	추론(Inferential) 통계학
기능	• 수집된 자료의 특성을 쉽게 파악하기 위해서 자료를 정리 및 요약	• 모집단에서 추출한 표본의 정보를 이용하여 모집단의 다양한 특성을 과학적으로 추론
방법	• 표, 그래프, 대푯값 등	• 회귀분석, T-검정, 분산분석 등

2. 모집단과 표본

① 전수조사

- 모집단내에 있는 모든 대상 조사 방법(예, 인구조사)
- 모집단의 특성 정확히 반영
- 시간과 비용이 많이 소요되는 단점

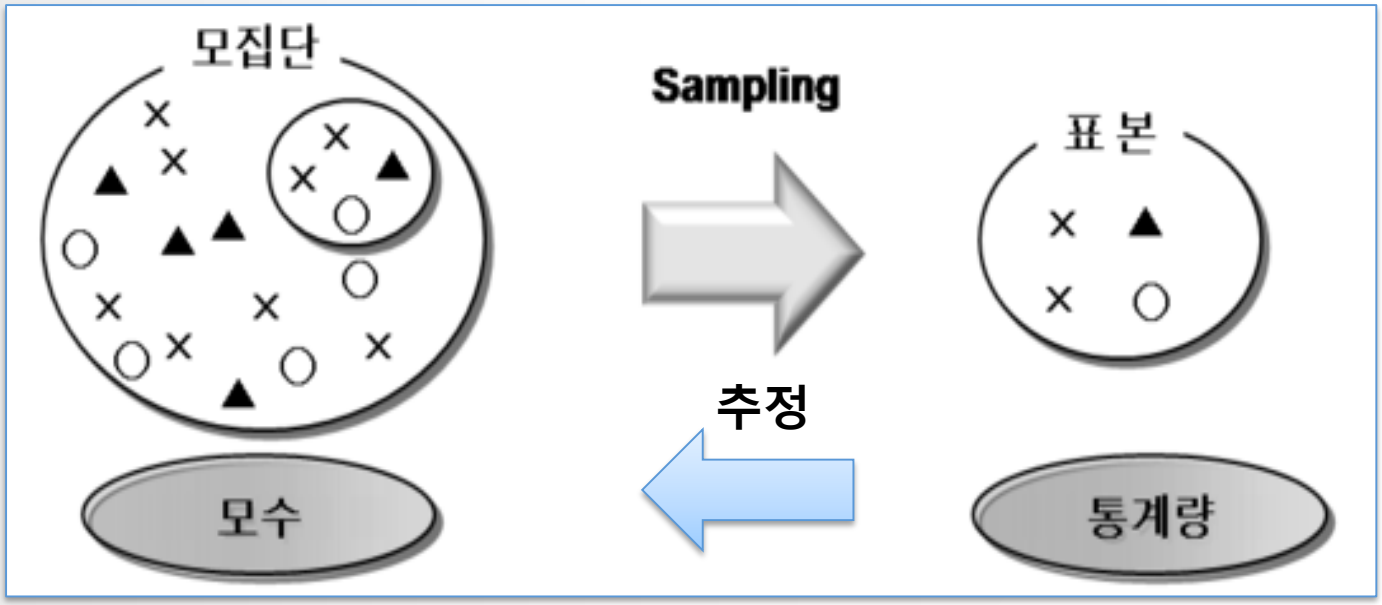
② 표본조사

- 모집단으로부터 추출된 표본을 대상으로 분석 실시

  (예, 선거 여론조사, 마케팅조사, 안전성 검사, 의생명 임상실험)
- 모집단의 특성을 반영하지 못하는 표본은 무용지물

2. 모집단과 표본

- **모집단과 표본**
  - **Sampling : 표본추출**



2. 모집단과 표본

모수와 통계량 표현

구분	모수(모집단)	통계량(표본)
의미	모집단의 특성을 나타내는 수치	표본의 특성을 나타내는 수치
표기	그리스, 로마자	영문 알파벳
평균	μ(모평균)	$\overline{X}$(표본의 평균)
표준편차	σ(모표준편차)	S (표본의 표준편차)
분산	σ^2(모분산)	S^2 (표본의 분산)
대상 수	N(사례수)	n(표본수)

2. 모집단과 표본

- **표본 추출 과정**

무작위 표본추출 (random sampling)

2. 모집단과 표본

- **표본크기 결정**
  - 유한모집단의 경우

$$n \geq \frac{N}{(\frac{e}{k})^2 \frac{N-1}{P(1-P)} + 1}$$

  - 무한모집단의 경우

$$n \geq \frac{1}{(\frac{e}{k})^2 \frac{1}{P(1-P)}}$$

N : 모집단의 크기

e : 요구정밀도

P : 모집단의 비율

k : 신뢰수준(α=0.05일 때 k=1.96)

2. 모집단과 표본

- **표본크기 결정**

① **요구정밀도 e의 결정 :** 허용가능 최대오차(**10%** 설정)

② **신뢰수준 α의 결정 :** 95% 신뢰도(α=0.05 설정)

- 95% 신로도 → α=0.05 → **k = 1.96**
- 90% 신로도 → α=0.1 → k = 1.65
- 99% 신로도 → α=0.01 → k = 2.58

N : 모집단 크기
e : 요구정밀도
P : 모집단 비율
k : 신뢰수준

③ **모집단 비율 P 예측 :** 예비조사 결과나 기존의 설문조사 결과를 기초로 P값 예측(예측 불가능한 경우 **P(찬성률) 50%** 설정)

④ **수식 계산 :** 유한 또는 무한모집단의 특성을 고려 해당 수식 적용

2. 모집단과 표본

$$n \geq \frac{N}{(\frac{e}{k})^2 \frac{N-1}{P(1-P)}+1}$$

N : 모집단 크기
e : 요구정밀도
P : 모집단 비율
k : 신뢰수준

【표본 크기 결정 예제】

- A전기 회사의 사원수가 5,000명인 경우 요구정밀도 10%, 신뢰수준 95% 일 때 표본의 크기는 얼마인가?

$$n \geq \frac{5000}{(\frac{0.1}{1.96})^2 \frac{5000-1}{0.5(1-0.5)}+1} = \frac{5000}{0.0026 \times \frac{4999}{0.25}+1} = \frac{5000}{52.9896} = 94.358 \rightarrow \textbf{94명}$$

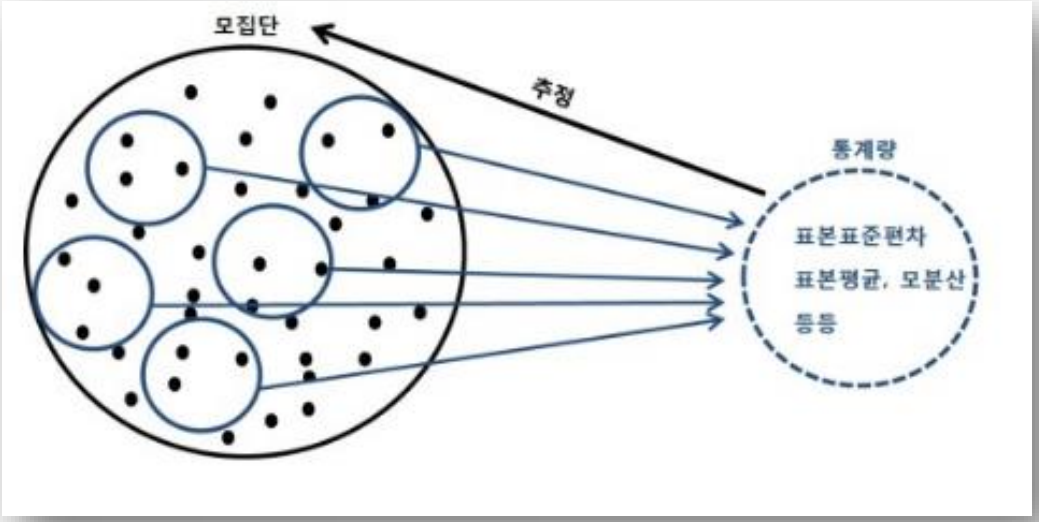
만약 사원수가 10,000명인 경우 표본의 크기는?

$$n \geq \frac{10000}{(\frac{0.1}{1.96})^2 \frac{10000-1}{0.5(1-0.5)}+1} = 95.247 \rightarrow \textbf{95명}$$

- 모집단 크기 **N = 5,000** 일 때 표본의 크기 = 94명
- 모집단 크기 **N = 10,000** 일 때 표본의 크기 = 95명

3) 추정과 검정

- **통계적 추정**
  - 모집단의 특성을 대표하는 표본을 추출하고, 이러한 표본을 이용하여 모집단의 특성을 나타내는 각종 모수(모평균, 모분산 등)를 예측하는 방법



3) 추정과 검정

● 통계적 추정

➢ 모집단의 특성을 대표하는 표본을 추출하고, 이러한 표본을 이용하여 모집단의 특성을 나타내는 각종 모수(모평균, 모분산 등)를 예측하는 방법

구분	점 추정	구간 추정
방식	• 모집단의 특성을 **하나의 값**으로 추정하는 방식 • 모평균은 25정도로 추정	• 모집단의 특성을 **적절한 구간**을 이용하여 추정하는 방식 • 모평균은 20~30 사이로 추정
특징	• 모수와 동일할 가능성이 가장 높은 하나의 값을 선택하는 방법	• 모수가 속하는 일정구간(하한값, 상한값)으로 추정(일반적으로 많이 사용)

3) 추정과 검정

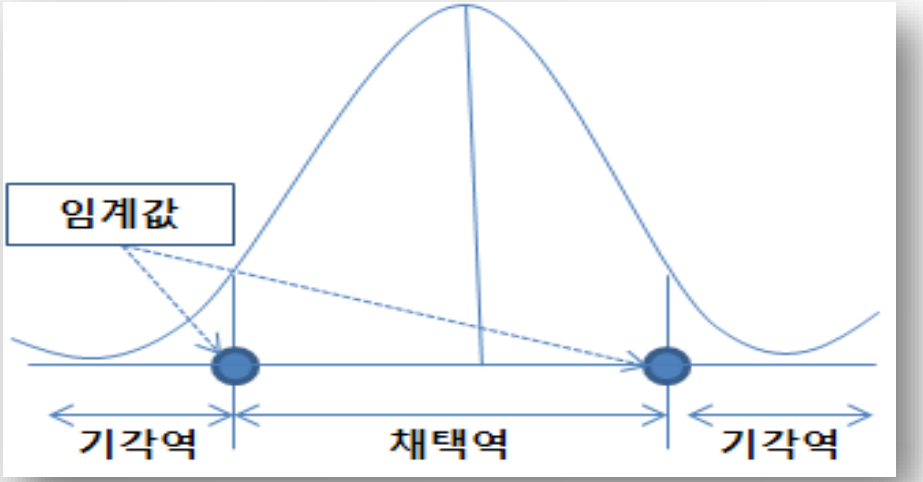
- **구간추정 주요 용어**
  - 신뢰수준(Confidence Level) : 계산된 구간이 모수를 포함할 확률 의미 (통상 90%, 95%, 99% 등으로 표현)
  - 신뢰구간(Confidence Interval) : 신뢰수준 하에서 모수를 포함하는 구간 (하한값 ~ 상한값 형식으로 표현)
  - 표본오차(Sampling Error) : 모집단에서 추출한 표본이 모집단의 특성과 정확히 일치하지 않아서 발생되는 확률의 차이

예)) 대통령 후보의 지지율 여론조사에서 모 후부의 지지율이 95% 신뢰수준에서 표본오차 ±3% 범위에서 32.4%로 조사 되었다고 가정한다면 실제 지지율은 29.4%~35.4%(-3%~+3%)사이에 나타날 수 있다는 의미이다. 여기서 95% 정도는 이 범위의 지지율을 신뢰할 수 있지만 5% 수준에서는 틀릴 수도 있는 의미이다.

➔ **신뢰수준 95%, 신뢰구간 29.4%~35.4%, 표본오차 ±3%,**

3) 추정과 검정

- **임계값에 따른 기각역과 채택역**
  - 임계값(Critical value) : 귀무가설 채택 or 기각 기준점
  - 채택역(Acceptance region) : 임계값 기준 채택(귀무가설) 범위
  - 기각역(Critical region) : 기각 범위



3) 추정과 검정

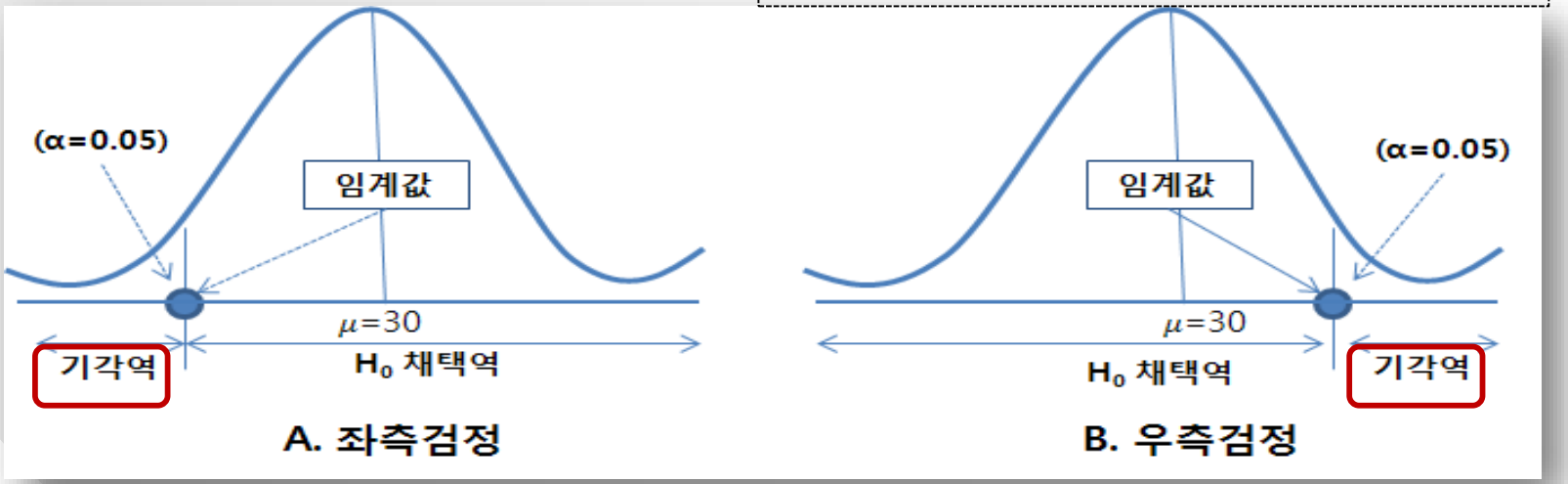
- **단측검정(1-sided test) : 방향(우열) 있는 단측가설 검정**

H_0 : 1일 생산되는 불량품의 개수는 평균 30개 이다.(μ=30)

H_1 : 1일 생산되는 불량품의 개수는 평균 30개 이하이다.(μ<30) ▶ 왼쪽 단측검정

1일 생산되는 불량품의 개수는 평균 30개 이상이다.(μ>30) ▶ 오른쪽 단측검정

연구가설이 < 또는 > 두 가지 가설 포함



왼쪽 단측검정 오른쪽 단측검정

3) 추정과 검정

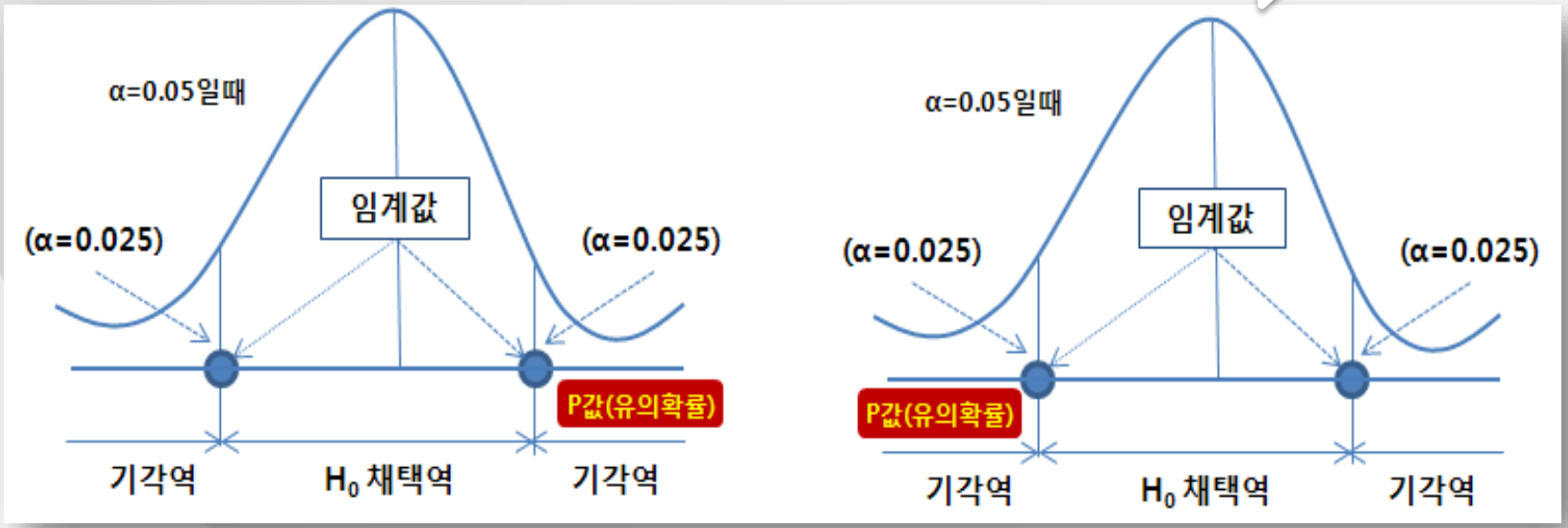
- **양측검정(2-sided test) : 방향 없는 양측가설 검정**

H_0 : 성별에 따라 만족도에 차이가 없다.(같다)

H_1 : 성별에 따라 만족도에 차이가 있다.(같지 않다)

대립가설
연구 환경에
따라 달라짐

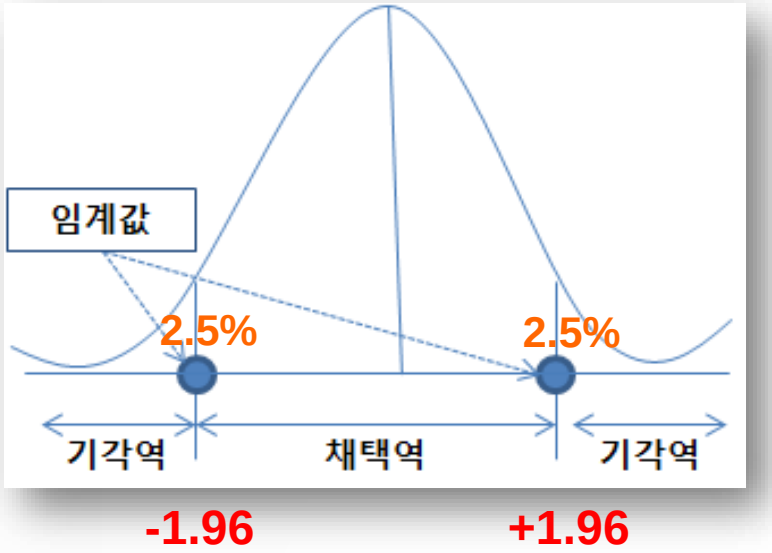
3가지 대립가설 : 같지 않다. 남자 > 여자, 남자 < 여자



3) 추정과 검정

- **유의수준 vs Z값(채택역)**

유의수준(α)/확률	정규분포 Z값(채택역)
0.5%(0.005)/99%	±2.58(양측검정)
2.5%(0.025)/95%	±1.96(양측검정)
5%(0.05)/90%	±1.64(양측검정)



3) 추정과 검정

- **T 분포표**

Z 분포 이용 : 모집단의 표준편차가 알려진 경우

T 분포 이용 : 모집단의 표준편차가 알려지지 않은 경우 표본 표준편차 이용

자유도	꼬리확률 q									
ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	[illegible]	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	[illegible]	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	[illegible]	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	[illegible]	[illegible]
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	[illegible]	[illegible]
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	[illegible]	[illegible]
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	[illegible]	[illegible]
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	[illegible]	[illegible]
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.792	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

95% 신뢰수준 경우
알파 = 0.025(좌우대칭)

4) 가설검정 오류

- **제1종 오류**
  - 귀무가설이 참인 경우 귀무가설 기각 오류
- **제2종 오류**
  - 귀무가설이 거짓인 경우 귀무가설 채택 오류

가설현황 / 검정 결과	귀무가설(H_0) 참인 경우	연구가설(H_1) 참인 경우
귀무가설(H_0) 채택	문제 없음	**제2종 오류**
연구가설(H_1) 채택	**제1종 오류**	문제 없음

- 가설검정에서 두 가지 오류 발생(모두 작은 경우가 바람직함)
- 제1종 오류가 발생되는 것을 가만해서 유의수준 정함
  (유의수준 α : 0.1, 0.05, 0.01)
- 제2종 오류를 범하지 않을 확률은 1-β = 검정력(Power of the test)

5) 검정통계량

- **검정통계량(Test statistic)**
  - 가설 검정 위해 수집된 자료를 계산한 통계량
  - 가설검정에서 기각역을 결정하는 기준이 되는 통계량
  - 유의수준 **α**의 값과 비교하여 귀무가설 기각/채택
  - 상관분석 **r**값, T검정 **t**값, 분산분석/회귀분석 **F**값, 카이제곱 $\mathbf{X^2}$값

5) 검정통계량

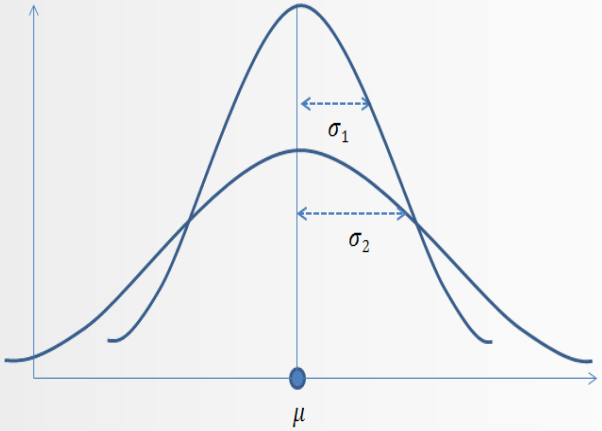
연구가설(H_1) : '학력수준에 따라 제품만족도에 차이가 있다.'를 검정하기 위해서 독립표본 T검정을 수행하였다. 이때 유의수준은 α=0.05로 결정 하였다.
검정 결과 검정통계량 t값이 10.652, 유의확률 p값이 0.012가 나왔다고 가정한다면 귀무가설은 기각되는가? 채택되는가?

검정통계량 t=10.652값은 유의확률 p=0.012이다. 유의수준 α=0.05 수준에서 귀무가설('학력수준에 따라 제품만족도에 차이가 없다.)' 기각(p<α)

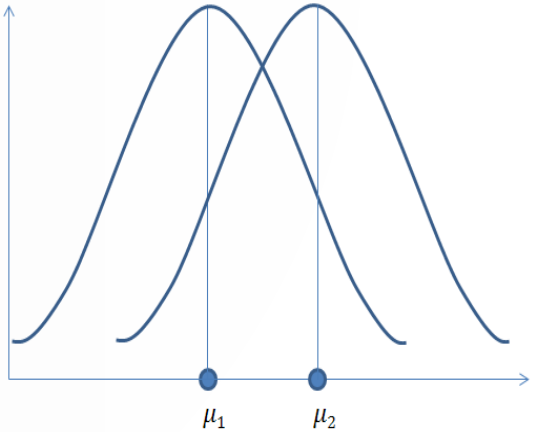
➲ 학력수준에 따라 제품만족도에 유의미한 차이가 있는 것으로 볼 수 있다.

6) 정규분포

- **정규분포(Normal Distribution)**
  - ➢ 도수분포곡선이 평균값을 중앙으로 하여 좌우대칭인 종 모양
  - ➢ K.F.가우스가 측정오차의 분포에서 중요성 강조 ➔ 가우스분포(가우스곡선)
  - ➢ 평균과 표준편차에 의해서 정규분포 모양과 위치가 결정



표준편차(σ_1, σ_2)에 따른 그래프 모양



평균($\mu 1$, $\mu 2$)에 따른 그래프 모양

6) 정규분포

● 정규분포(Normal Distribution)의 특징

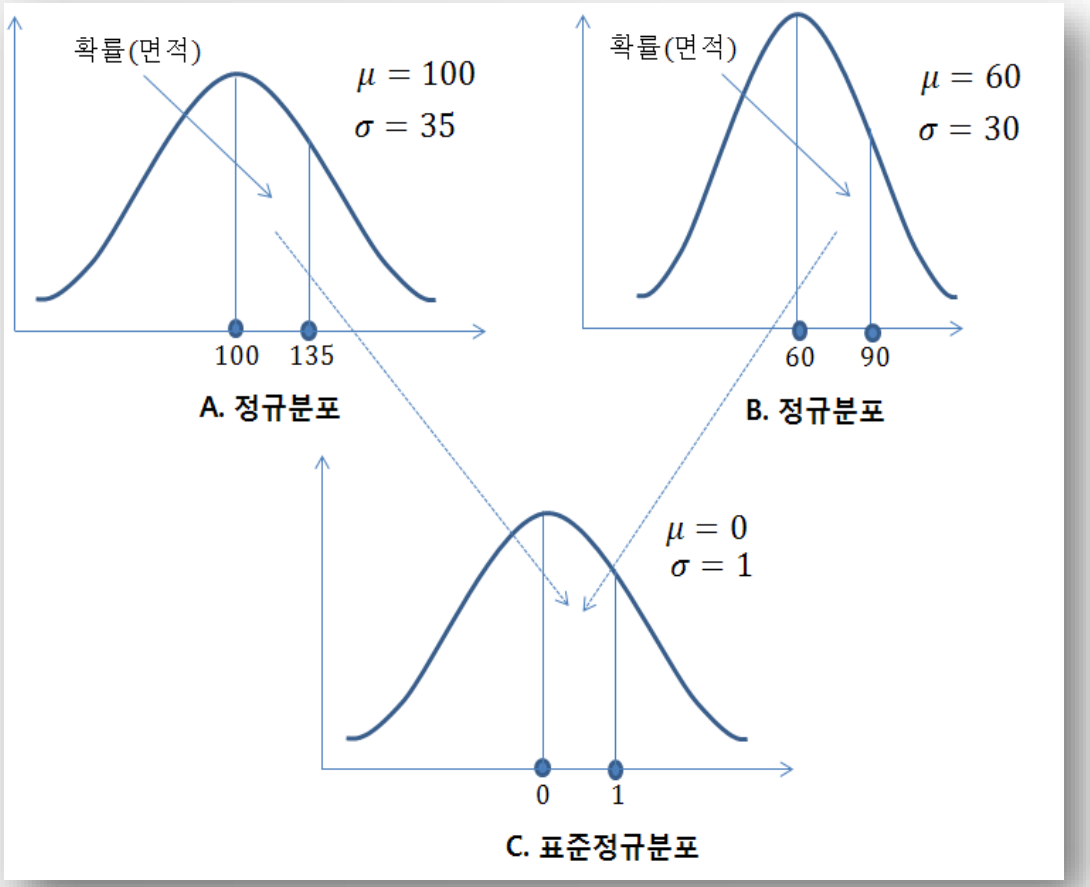
- ➢ 데이터의 분포가 평균을 중심으로 많은 데이터가 모여 있는 특성
- ➢ 대부분 정규분포를 이룬다고 가정하고, 통계분석 진행 ➔ 모수 검정
- ➢ '중심극한의 정리'에 의해서 데이터의 수가 많아질수록 정규분포를 따른다.

구분	특징
변수	• 연속 변수
분포	• 평균을 중심으로 좌우대칭인 **종 모양**
대푯값	• **평균 = 중앙값 = 최빈값**
왜도/첨도	• 왜도 = 0, 첨도 = 0(또는 3)
모양	• 표준편차(σ)에 의해서 모양이 달라진다.
위치	• 평균(μ)에 의해서 위치가 달라진다.
넓이	• 정규분포의 전체 면적은 1이다.

※ 표준정규분포 : 평균이 0이고, 표준편차가 1인 정규분포$N(0, 1^2)$

3) 추정과 검정

- 정규분포 vs 표준정규분포



6) 정규분포

- **대푯값 기술통계량**
  - 자료 전체를 대표하는 값(분포의 중심위치를 나타내는 측정치)
  - 합계(Sum), 평균(Mean)
  - 중위수(Median), 최빈수(mode), 사분위수

6) 정규분포

- **산포도 기술통계량**
  - 변량이 흩어져있는 정도(평균에 모여 있으면 산포도가 작다)

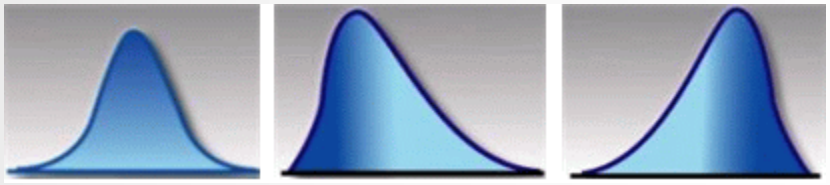
➢ 평균(μ) = $\dfrac{x_1+x_2+x_3+\ldots+x_n}{n}$

➢ 분산(σ^2) = $\dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots(x_n-\bar{x})^2}{n}$

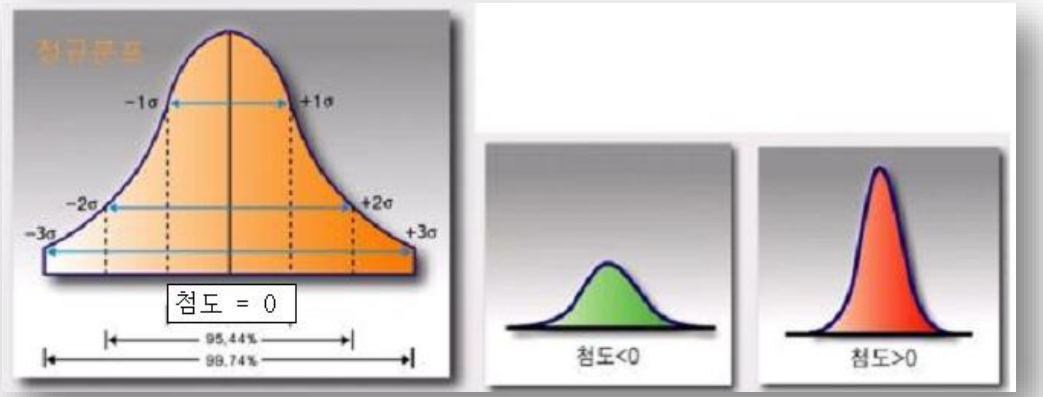
➢ 표준편차(σ) = $\sqrt{\dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots(x_n-\bar{x})^2}{n}}$

6) 정규분포

- **비대칭도 기술통계량**
  - 분포가 기울어진 방향과 정도



왜도=0 왜도 > 0 왜도 < 0

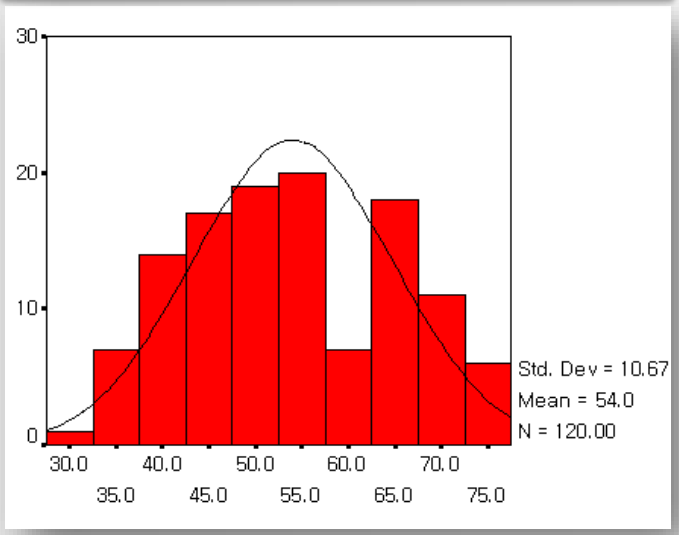


첨도=0 첨도 < 0 첨도 > 0

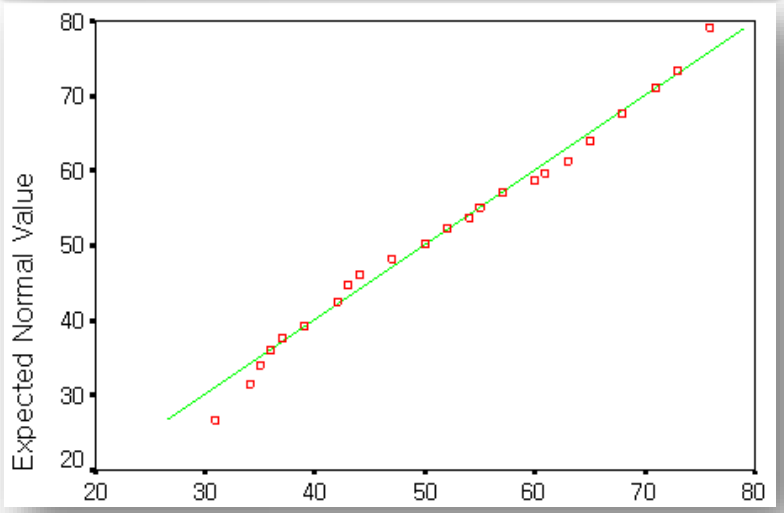
6) 정규분포

- 정규성 검정 관련 그래프

1. Graphs → Histogram



2. Graphs → Q-Q Plots



7) 모수 vs 비모수

- **모수(Parametric) 검정**
  - 관측값이 확률분포(정규분포, 이항분포 등)를 따른 경우
- **비모수(Non-parametric) 검정**
  - 관측값이 어느 특정한 확률분포를 따른다고 전제할 수 없는 경우

【중심극한정리】

- 케이스 30개 이상이면 정규분포를 따른다고 전제

➔ **모수 검정 방법 실시**

7) 모수 vs 비모수

- **모수 vs 비모수 검정 방법**

검정 방법	모수(정규분포)	비모수(비정규분포)
t검정	독립표본 t검정	윌콕슨((Wilcoxon) 검정
	대응표본 t검정	맨-휘트니(Mann-Whitney) 검정
분산분석	일원배치분산분석	크루스칼-월리스(Kruskal-Wallis)검정
관계분석	상관분석	비모수적 상관분석