



제8장 기술통계분석

| 도입 사례 |

2015년 우리나라 전체 TOEIC시험(정기시험) 평균점수가 677점으로 10년 전인 2005년 평균점수 593점보다 84점 가량 향상된 것으로 나타났다. 구체적으로는 고등학생의 평균점수는 643점, 대학생의 평균은 688점, 그리고 대학원생의 평균은 701점이었다. 2015년 한 해 동안 TOEIC에 응시한 인원은 약 200여 만명으로 대략 우리나라 전체 인구 100명 중 4명이 TOEIC시험을 본 셈이다.

한편, 2015년도 한국의 국가별 TOEIC 점수 순위는 응시국 중 18위를 차지했다.

생각해 볼 문제 _____

- ① 평균 이외에 사회현상을 요약·설명해 주는 수단은 무엇이 있을까요?
- ② 현상들간의 관계를 연구할 때 각 현상의 변화 정도를 파악해야 하는 이유는 무엇일까요?



◆ 추상적인 현상간(예: 만족도와 재구매의도간)의 관계연구의 핵심

- ① 구체적인 수(만족도 및 재구매의도에 대한 설문조사결과)로 추상적인 현상을 나타내고,
- ② 수들간의 관계를 통계분석하여,
- ③ 이들 수가 대변하고 있는 추상적인 현상들간의 관계를 설명하려는 것

◆ 경영/경제 현상의 설명

◆ 기술통계(descriptive statistics)

- ◆ 수로 측정된 현상(많은 양의 정보)을 요약 · 설명하는 역할을 하는 통계학의 분야

- 자료를 표나 그림으로 나타내고, 자료 특징 반영에 필요한 값(평균, 분산, 표준편차 등)을 계산함
예) 도입사례의 지구촌 인구, 종교, 언어, 부(wealth), 건강, 에너지, 자유, 전쟁 등
관련 통계자료

◆ 추리통계(inferential statistics) * 현대통계학의 중핵을 이룸

- ◆ 기술통계학에서 정리하고 계산한 자료들을 바탕으로, 주어진 문제에 대하여 통계학적 추론을 해나가는 과정을 내용으로 함

- 명칭 사유: 부분에 포함된 제한된 정보로부터 전체에 대한 어떤 결론을 이끌어 내는
논리적 과정에 관한 이론임



제1절 기술통계분석이란 무엇인가?: 사회현상의 요약설명

- ◆ 기술통계(descriptive statistics)
 - 수로 측정된 현상을 요약 · 설명하는 역할을 하는 통계학의 분야
예) TOEIC평균점수 \Rightarrow TOEIC으로 측정한 영어실력을 요약

1. 사회현상의 설명

√ 도수(度數) : ◆ 반복되는 회수

◆ 각 등급구간(class interval)에 해당되는 응답자
(즉, 관찰된) 수

- ◆ 빈도/도수 분석(frequency analysis)
 - 현상(영어실력)에 대한 자세한 설명을 가능하게 하는 기본 분석
예) 각 TOEIC 응시자의 점수를 점수대별로 구분하여 각각의 점수대에는 몇 명이 속하는지를 정리
- ◆ 도수분포표(frequency table)
 - 빈도를 표로 정리한 것, 상대도수기둥그래프, 도수다각형, 누적도수다각형
- ◆ 도수분포곡선(frequency curve)
 - 빈도를 그래프로 정리한 것

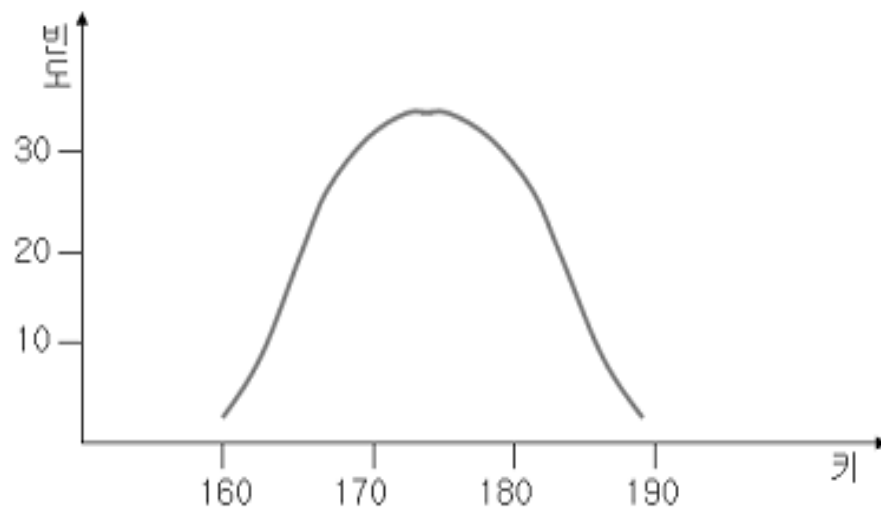


제1절 기술통계분석이란 무엇인가?: 사회현상의 요약설명

〈표 8-1〉 도수분포표

키	빈도/도수
185.1~190	10
180.1~185	20
⋮	⋮
⋮	⋮
160.1~165	10
계	100명

[그림 8-1] 도수분포곡선





2. 사회현상의 요약

◆ 현상의 요약 도구

- ① 현상을 대표하는 수
- ② 현상의 변화정도 파악

1) 자료(현상)의 요약

◆ 대표값/중심화경향(measures of location 혹은 central tendency)

- 자료를 요약(대표)하는 수, 어떠한 수치를 중심으로 자료가 분포되었는가 나타냄
평균 등

예1) 기말 경영/경제학 시험 반 평균: 100점 만점에 82점

예2) 347개 상장회사 대졸초임 평균: 2천 713만원

2) 자료(현상)의 변화정도 파악

◆ 산포도(measures of dispersion/spread 혹은 variability)

- 자료의 변화여부와 변화정도를 표현하는 도구, 자료가 어느 정도 산포되어 있는지를 나타냄

☞ 분산, 표준편차 등



제1절 기술통계분석이란 무엇인가?: 사회현상의 요약설명

〈표 8-2〉 동일평균 상이 산포도의 예 : 소비자 만족도 응답값

응답자 번호	응답값의 변화폭 0인 경우	응답값의 변화폭이 0이 아닌 경우
1	4	4
2	4	5
3	4	3
4	4	5
5	4	4
6	4	5
7	4	4
8	4	3
9	4	5
10	4	2
평균	4	4
변화폭	0	1



1. 빈도분석

◆ <표 8-3> 검토

- 228명의 대학생을 대상으로 설문조사를 실시한 결과 중 응답자의 연령(age)의 빈도를 SPSS Statistics로 분석한 결과(도수분포표)

◆ 시스템 결측값

- 전체 응답자 228명 중 3명은 연령에 관한 문항에 대답을 하지 않았고 이 것이 '시스템 결측값'이라는 행(row)에 제시됨

〈표 8-3〉 빈도분석 결과

		AGE			
		빈도	퍼센트	유효 퍼센트	누적퍼센트
유효	20,00	12	5,3	5,3	5,3
	21,00	46	20,2	20,4	25,8
	22,00	38	16,7	16,9	42,7
	23,00	33	14,5	14,7	57,3
	24,00	20	8,8	8,9	66,2
	25,00	26	11,4	11,6	77,8
	26,00	26	11,4	11,6	89,3
	27,00	12	5,3	5,3	94,7
	28,00	6	2,6	2,7	97,3
	29,00	4	1,8	1,8	99,1
	30,00	1	,4	,4	99,6
	32,00	1	,4	,4	100,0
	합계	225	98,7	100,0	
결측	시스템 결측값	3	1,3		
합계		228	100,0		



2. 대푯값: 현상의 요약

- ◆ 대푯값은 자료수집시 사용된 척도의 유형(즉, 사용된 숫자가 보유한 특징에 따라)에 따라 상이

1) 명목척도와 대푯값

◆ 최빈값(mode)

- 명목척도 측정 현상의 대푯값
- 최고의 빈도(frequency)를 가지는 값(value) 혹은 집단(category)을 지칭

〈표 8-3A〉 명목척도의 대푯값 1

응답자 성별	빈 도
남	465
여	535
합 계	1,000
최빈값	여성/535명



제2절 기술통계분석결과는 어떻게 해석하는가?

- ◆ 원칙적으로 남성과 여성응답자가 25명씩 동수로 존재하는 경우에 최빈값은 존재하지 않는 것으로 해석

〈표 8-3B〉 명목척도의 대푯값 2

	남	여	소 계
Sonata	25명	25명	50명
SM5	25명	25명	50명
소 계	50명	50명	100명



2) 서열척도와 대푯값

◆ 중앙값(median)

- 서열로 측정된 현상을 요약해 주는 대푯값
- 모든 응답치의 중앙에 해당하는 값(value)
 - 변량의 수를 n 이라 할 때, 홀수인 경우는 $(n+1)/2$ 번째 변량의 값,
 - 짝수인 경우는 가운데 있는 두 변량을 산술평균한 값으로 함

〈표 8-4A〉 서열척도 수준으로 측정한 원(original)응답자료

응답자번호	Sonata	SM5	Lotze
1	1	2	3
2	2	1	3
3	3	2	1
4	1	2	3
5	2	1	3
6	1	3	2
7	1	2	3
8	2	1	3
9	1	2	3
합 계	-	-	-



제2절 기술통계분석결과는 어떻게 해석하는가?

〈표 8-4B〉 서열척도 수준으로 측정한 원(original) 응답자료를 재구성한 자료

응답자번호	Sonata	SM5	Lotze
1	1	2	3
4	1	2	3
6	1	3	2
7	1	2	3
9	1*	2*	3*
2	2	1	3
5	2	1	3
8	2	1	3
3	3	2	1
중앙값	1위	2위	3위

☞ Sonata의 중앙값 1위, SM5의 중앙값 2위, Lotze의 중앙값 3위



3) 등간척도/비율척도와 대푯값

◆ 평균(mean)

- 자료가 등간척도 혹은 비율척도로 측정된 경우에 이러한 자료를 요약해주는 대푯값
- 모든 응답자의 응답값을 더하고 이를 모든 응답자의 수로 나누어서 구함

$$\begin{aligned}
 &\text{평균(mean)} && \text{(식 8-1)} \\
 &\text{평균}(=\bar{X}) \\
 &= (\text{응답값의 합})/(\text{응답자 수}) \\
 &= \sum X_i / N
 \end{aligned}$$

◆ <표 8-2>의 예 중 9명의 응답이 각각 4인 경우의 평균

$$\square (4 + 4 + \cdots + 4)/9 = (4 * 9)/9 = 4$$

* **성질** $E(a)=a$, $E(aX)=aE(X)$, $E(a+bX)=a+bE(X)$, 단, a , b 는 상수



3. 산포도: 현상의 변화정도

- ◆ 자료요약(descriptive summary)의 좀 더 중요한 목적은 조사의 대상 사회현상이 변하는 현상인지의 여부와, 변화한다면 변화폭은 어느 정도인지를 알아보는 것
- ◆ 변화여부/폭 판단 기준
 - 범위, 편차, 분산 등

1) 범 위(range)

- 최대응답값과 최소응답값의 차이, n 이 작은 경우가 값이 안정되므로 n 이 10(또는 20) 이하일 때 많이 쓰이며 공업제품의 품질관리에 유용, n 이 너무 크면 분포 범위의 크기는 불안정하게 됨 => 책 p.244 유용성

예 1) <표 8-2>의 column B에 제시된 응답값의 경우

☞ 범위는 3 (= 5-2)

☞ '변화하는 현상'이라고 판단 가능



2) 편 차(deviation)

▣ 개인의 응답값에서 평균을 차감한 수 ($=X_i - \bar{X}$)

☞ 편차는 응답자의 수만큼 존재

☞ 대표적인 편차는 어떻게 정해야 하는지의 문제 발생

예) 편차의 합, 편차들의 평균?

* 편차개념의 **단점** : 경우에 따라 실제로 존재하는 변화폭의 크기를 보여주지 못한다는 단점이 있어 다른 산포도 도구가 개발됨



제2절 기술통계분석결과는 어떻게 해석하는가?

〈표 8-5〉 편차, 분산, 표준편차

응답자 번호	응답값	편차(= 응답값-평균)	편차제곱
1	4	0(= 4 - 4)	0
2	5	1(= 5 - 4)	1
3	3	-1(= 3 - 4)	1
4	5	1(= 5 - 4)	1
5	4	0(= 4 - 4)	0
6	5	1(= 5 - 4)	1
7	4	0(= 4 - 4)	0
8	4	0(= 4 - 4)	0
9	5	1(= 5 - 4)	1
10	1	-3(= 1 - 4)	9
합	40(=A)	0 (=B)	14 (=C)
응답자수	10 (= D)		
평균	4 (= 40/10 = A/D)		
분산	1.4 (= 14/10 = C/D)		
표준편차	$\sqrt{1.4}$		

3) 분 산

(1) 분산의 구분

◆ 분산(variance)

- 평균으로부터 개별 응답값들이 '평균적으로' 얼마나 벌어져 있는지를 요약해 주는 도구 (확률변수가 평균을 중심으로 하여 얼마나 멀리 퍼져 있는 형태를 취하는가를 나타내 주는 지표가 됨)

- 편차제곱의 합을 응답자수(정확하게는 "응답자수 -1"로 나눔)로 나누어서 산출

☞ 분산은 하나만 존재

☞ 분산은 기술통계는 물론 추리통계에서도 중요한 역할을 수행

(식 8-2) 분산($=S^2$)

$$= (\text{편차 제곱의 합}) / (\text{응답자의 수} - 1)$$

$$= \sum (X_i - \bar{X})^2 / (N - 1)$$

$$\begin{aligned} * \text{Var}(X) &= E[(X-u)^2] = E[X^2 - 2uX + u^2] \\ &= E(X^2) - 2uE(X) + u^2 = E(X^2) - u^2 \end{aligned}$$

* 성질

• 분모를 n-1로 나누는 이유

- 표본분산은 모분산보다 변동폭이 약간 작은 경향이 있기 때문
 - > 표본분산이 모분산에 접근하도록 할 수 있기 때문(Malhotra, 2002)
- S^2 이 모분산 σ^2 의 불편추정량이 되기 위해서
- 소표본에서 모평균 μ 추정시 t 분포를 사용할 수 있기 위해서



제2절 기술통계분석결과는 어떻게 해석하는가?

◆ 분산의 장점

- 현상의 변화방향과는 상관없이 현상의 변화폭이 존재하는 양(의 제공)에 비례해서 분산의 크기가 결정됨
 - ☞ 현상의 변화폭을 대표적으로 쉽게 파악할 수 있는 도구

◆ 분산의 단점

- 분산이 편차의 '제곱'으로 표현되기 때문에 실제로 응답이 측정된 단위보다 매우 큰 숫자로 표현이 된다는 점
 - ☞ 실제 측정단위를 기준으로 어느 정도의 변화폭이 존재하는지에 대한 직관적인 답을 줄 수 없다는 점

* 분산의 구분-> 측정대상 현상의 변화 + 다른 현상의 변화 (* IQ 테스트)

◆ 모분산 (population variance) vs. 표본분산(sample variance)

◆ **총분산**(total variance)= **체계적** 분산(집단간분산/실험분산/독립변수에 의한 분산) + **비체계적** 분산(집단내분산/오차분산/불확실한 원인에 의한 분산)

* 책 pp.246-248



제2절 기술통계분석결과는 어떻게 해석하는가?

◆ 공분산(covariance)

- 하나의 변수가 변함에 따라 다른 변수가 어떤 방향으로 얼마나 변화하는가를 보여주는 도구

공분산(covariance) (식 8-3)

$$\begin{aligned}\text{공분산}(=COV_{xy}) &= \sum(X_i \text{의 편차})(Y_i \text{의 편차}) / (\text{표본크기}-1) \\ &= \sum(X_i - \bar{X})(Y_i - \bar{Y}) / (N - 1)\end{aligned}$$

(2) 분산과 공분산의 중요성

- ◆ 분산과 공분산은 현상(즉, 변수)의 변화폭 혹은 현상(변수)들간의 관계를 요약해 줌
- ◆ 분산과 공분산은 많은 통계분석기법의 척추와 같은 역할을 수행 (p.248)



4) 표준편차

- ◆ 표준편차(standard deviation)

- 분산에 제곱근을 적용하여 얻어지는 수($S = \sqrt{S^2}$)

- ◆ 표준편차의 장점

- 분산의 장점을 그대로 보유하되, 현상의 측정단위와 동일한 단위로 표현된 변화폭 설명도구

- ◆ 왜도(skewness): 정상분포곡선이 좌우로 기울어진 정도, 정상분포의 경우 값이 0임, 분포곡선이 정상분포대비 좌측으로 기울어지면 (+) 값으로, 우측으로 기울어지면 (-) 값으로 나타냄

첨도(kurtosis): 정상분포곡선이 뾰족/완만한 편인지 의미, 정상분포의 경우 이 값이 0임, 뾰족한 편일수록 (+) 값으로, 완만한 편일수록 (-) 값으로 나타냄

p. 252 & 한글파일(프린트물) 참조



제2절 기술통계분석결과는 어떻게 해석하는가?

♣ 분산과 표준편차의 유용성

- 분산은 평균으로부터의 편차 값을 이용하여 계산되므로 평균값과는 독립적으로 계산됨
 - 미분이 가능한 점 등 수학적 조작이 간편함(분산)
 - 주어진 자료의 모든 값을 사용하여 계산하며, 분산의 값은 자료의 조그마한 변화에도 아주 민감함
 - 한 그룹의 관측치들이 특정한 분포를 따를 때에 분포되어 있는 정도에 대한 훌륭한 정보를 제공할 수 있음
 - 분산은 일단 편차를 제공하는 과정을 거치므로, 편차가 큰 것일수록 제공함으로써 일종의 평가를 가중하게 함
- > 여러 가지 자료에서 오차가 커진다거나 기대 값으로부터 멀어져 위험 부담이 커질 때에, 이를 가능한 회피하고자 할 필요가 있을 때 분산과 표준편차가 유용하게 사용될 수 있을 것임



제3절 기술통계분석은 어떻게 실행하는가?

- ◆ 모든 통계분석이든 분석대상 자료를 열어 놓고 진행할 필요
- ◆ SPSS Statistics를 이용한 기술통계분석은
 - main 메뉴에서 **분석** → **기술통계량** → **빈도분석** 혹은 **분석** → **기술통계량** → **기술통계**를 선택해서 수행하게 됩니다.
- ◆ SPSS main 메뉴 중 빈도분석메뉴 혹은 기술통계메뉴 사용 가능

1. 빈도분석메뉴

1) 빈도분석메뉴 시작방법

- ◆ 분석대상 자료파일을 불러옴
- ◆ 다음 [그림 8-2A]에서와 같이, SPSS main 메뉴 중 **분석(A)** → **기술통계량(E)** → **빈도분석(F)**을 선택

제3절 기술통계분석은 어떻게 실행하는가?

[그림 8-2A] 빈도분석메뉴 시작

*intmktg.sav [데이터세트1] - IBM SPSS Statistics Data Editor

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

보고서(P) ▶
 기술통계량(E) ▶
 표(B) ▶
 평균 비교(M) ▶
 일반선형모형(G) ▶
 일반화 선형 모형(Z) ▶
 혼합 모형(X) ▶
 상관분석(C) ▶
 회귀분석(R) ▶
 로그선형분석(Q) ▶
 신경망(W) ▶
 분류분석(F) ▶
 차원 축소(D) ▶
 척도분석(A) ▶

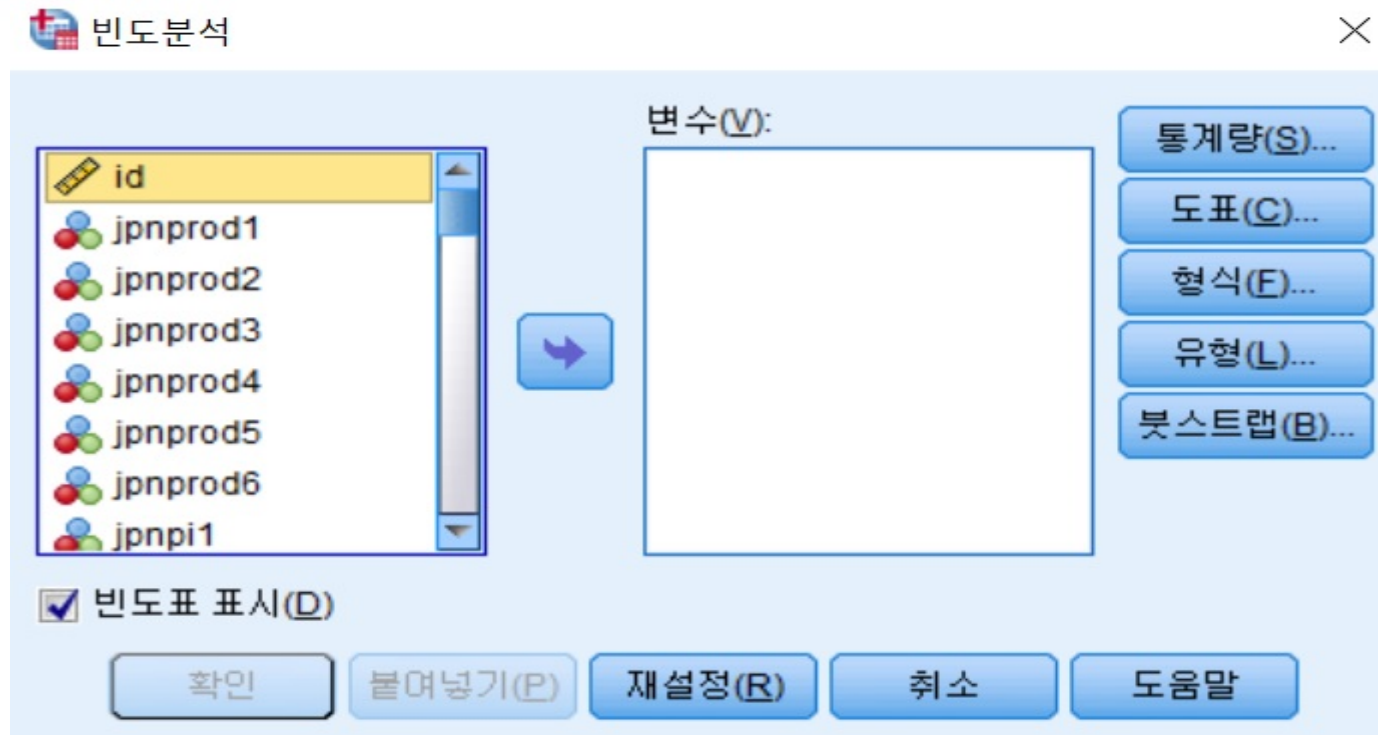
빈도분석(F)...
 기술통계(D)...
 데이터 탐색(E)...
 교차분석(C)...
 TURF 분석
 비율통계량(R)...
 P-P 도표...
 Q-Q 도표...

	id	jpnprod1	jpnprod2
1	1.00	5.00	2.00
2	2.00	4.00	3.00
3	3.00	7.00	3.00
4	4.00	5.00	2.00
5	5.00	5.00	2.00
6	6.00	3.00	2.00
7	7.00	6.00	2.00
8	8.00	6.00	3.00
9	9.00	6.00	2.00
10	10.00	6.00	2.00
11	11.00	6.00	3.00

- ◆ 위와 같은 선택을 하게 되면, [그림 8-2B]처럼 빈도분석의 초기화면이 나타나게 됨

제3절 기술통계분석은 어떻게 실행하는가?

[그림 8-2B] 빈도분석메뉴 시작



- ◆ [그림 8-2B]에서 보면 그림 아래쪽 부분에 통계량(S), 도표(C), 형식(F)의 세 개의 분석 키가 있음



제3절 기술통계분석은 어떻게 실행하는가?

〈표 8-6〉 빈도분석 분석 키 설명

종 류	내 용
통계량(S)	대푯값/중심경향(평균, 중위수, 최빈값, 합계), 산포도(표준편차, 최소값, 최대값, 분산, 범위, 평균의 표준오차) 등의 통계량을 지정하여 계산 수 있습니다.
도표(C)	막대도표, 원도표, 히스토그램 등 도표의 형식을 지정할 수 있습니다.
형식(F)	어떤 순서대로 출력할 것인지의 출력순서와 표 출력 범주 등을 설정할 수 있습니다.

제3절 기술통계분석은 어떻게 실행하는가?

- ◆ 먼저 세 개의 분석 키 중 통계량(S)을 선택하여 클릭하면 <그림 8-2C>화면을 얻을 수 있음
 - ☞ 이 화면에는 기본적으로 백분위수, 대푯값/중심경향, 산포도 및 분포의 계산과 관련된 선택을 할 수 있게 되어 있음

[그림 8-2C] 빈도분석 키 중 통계량(S) 선택시 화면

IBM SPSS Statistics Data Editor

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

표시: 51 / 51 변수

빈도분석: 통계량

백분위수 값

☐ 사분위수(Q)

☐ 절단점(U): 10 동일 크기 집단 수

☐ 백분위수(P):

추가(A) 변경(C) 제거(R)

중심경향

☐ 평균(M)

☐ 중위수(D)

☐ 최빈값(Q)

☐ 합계(S)

산포도

☐ 표준화 편차 ☐ 최소값(I)

☐ 분산(V) ☐ 최대값(X)

☐ 범위(N) ☐ S.D. 평균

분포

☐ 왜도(W)

☐ 첨도(K)

계속(C) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode: ON

- ◆ 왜도 vs. 첨도

제3절 기술통계분석은 어떻게 실행하는가?

- ◆ 둘째, 세 개의 분석 키 중 도표(C)를 선택하여 클릭하면 [그림 8-2D]와 같은 화면을 얻게 됨

☞ 빈도분석에서의 도표는 막대도표, 원 도표, 히스토그램의 3가지 형식으로 구성

[그림 8-2D] 빈도분석 키 중 도표(C) 선택시 화면

IBM SPSS Statistics Data Editor

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

표시: 51 / 51 변수

빈도분석: 도표

도표 유형

- ☒ 없음(N)
- ☐ 막대도표(B)
- ☐ 원도표(P)
- ☐ 히스토그램(H):
 - ☐ 히스토그램에 정규곡선 표시(S)

도표화 값

- ☒ 빈도(F) ☒ 퍼센트(C)

☒ 빈도표 표시(D)

확인

계속(C) 취소 도움말

통계량(S)... 도표(C)... 형식(F)... 유형(L)... 붓스트랩(B)...

도움말

	id	jpnprod1	jpnprod2	jpnprod3	jpnprod4	jpnprod5	jpnprod6	jpnpi1	jpnpi2	jpnpi3	jpnpi4	jpnpi5	jpnpi6	ethno1
1	1.00	5.00	2.00	5.00	5.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
2	2.00	4.00	3.00	4.00	4.00	5.00	5.00	3.00	2.00	3.00	2.00	2.00	2.00	2.00
3	3.00	7.00	3.00	6.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
4	4.00	5.00	2.00	6.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
5	5.00	5.00	2.00	7.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
6	6.00	3.00	2.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
7	7.00	6.00	2.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
8	8.00	6.00	3.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
9	9.00	6.00	2.00	6.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
10	10.00	6.00	2.00	6.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
11	11.00	6.00	3.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
12	12.00	5.00	2.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
13	13.00	4.00	2.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
14	14.00	5.00	2.00	7.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
15	15.00	7.00	2.00	5.00	6.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
16	16.00	5.00	4.00	6.00	7.00	5.00	5.00	4.00	4.00	5.00	2.00	2.00	3.00	4.00
17	17.00	5.00	2.00	5.00	4.00	3.00	4.00	2.00	4.00	4.00	4.00	2.00	6.00	2.00
18	18.00	6.00	3.00	6.00	6.00	5.00	5.00	2.00	1.00	3.00	3.00	2.00	4.00	2.00
19	19.00	7.00	3.00	6.00	6.00	6.00	6.00	4.00	3.00	3.00	2.00	4.00	2.00	3.00
20	20.00	7.00	1.00	7.00	7.00	6.00	7.00	5.00	2.00	3.00	4.00	3.00	5.00	5.00

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료

Unicode: ON

제3절 기술통계분석은 어떻게 실행하는가?

- ◆ 셋째, 세 개의 분석 키 중 형식(F)을 선택하여 클릭하면 [그림 8-2E]와 같은 화면을 얻게 됨
 - ☞ SPSS 빈도분석에서 형식 키는 빈도분석결과를 일정한 순서에 따라 정렬할 수 있게 하는 등의 기능을 수행해 줌

[그림 8-2E] 빈도분석 키 중 형식(C) 선택시 화면

The screenshot displays the IBM SPSS Statistics Data Editor window. The main data grid shows columns for 'id' and various product variables (jpnprod1 to jpnprod6, jpnpi1 to jpnpi6, and ethno1). A '빈도분석' (Frequency) dialog box is open, showing the '형식' (Format) tab. The '출력순서' (Output Order) section has '변수값 오름차순(A)' (Ascending) selected. The '다중 변수' (Multiple Variables) section has '변수를 비교(C)' (Compare variables) selected. Other options include '각 변수별로 출력결과를 나타냄(Q)' (Show results for each variable) and '범주가 많은 표는 출력하지 않음(I)' (Do not display tables with many categories). The '최대 범주 수(M):' (Maximum number of categories) is set to 10. Buttons for '계속(C)' (Continue), '취소' (Cancel), and '도움말' (Help) are visible.



2) 빈도분석메뉴 실행방법

- ◆ 위에서 설명한 빈도분석 관련 옵션을 설정(통계량, 도표, 형식 키의 세부 옵션 선택)한 다음에는 [그림 8-2F]의 창으로 돌아오게 됨
- ◆ 이 빈도분석 대화상자 창에서 분석하기를 원하는 변수를 마우스로 클릭하여 선택한 후 가운데에 있는 화살표 버튼을 누르거나 더블 클릭하면 오른쪽의 변수(V) 칸에 옮김
- ◆ 이 칸에 옮겨진 변수는 연구자가 선택한 빈도분석대상 변수임



제3절 기술통계분석은 어떻게 실행하는가?

[그림 8-2F] 빈도분석메뉴 실행방법

*intmktg.sav [데이터세트1] - IBM SPSS Statistics Data Editor

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

표시: 51/51 변수

	id	jpnprod1	jpnprod2	jpnprod3	jpnprod4	jpnprod5	jpnprod6	jpnpi1	jpnpi2	jpnpi3	jpnpi4	jpnpi5	jpnpi6	ethno1
1	1.00	5.00	2.00	5.00	5.00	6.00	6.00	1.00	1.00	5.00	2.00	3.00	4.00	2.00
2	2.00	4.00	3.00	4.00	4.00	5.00	5.00	3.00	2.00	3.00	2.00	2.00	2.00	2.00
3	3.00	7.00	3.00	6.00	5.00	3.00	3.00	2.00	3.00	2.00	2.00	2.00	2.00	2.00
4	4.00	5.00	2.00	6.00	6.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
5	5.00	5.00	2.00	7.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
6	6.00	3.00	2.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	7.00	6.00	2.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
8	8.00	6.00	3.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
9	9.00	6.00	2.00	6.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
10	10.00	6.00	2.00	6.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
11	11.00	6.00	3.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
12	12.00	5.00	2.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
13	13.00	4.00	2.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
14	14.00	5.00	2.00	7.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
15	15.00	7.00	2.00	5.00	5.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
16	16.00	5.00	4.00	6.00	7.00	5.00	5.00	4.00	4.00	5.00	2.00	2.00	3.00	4.00
17	17.00	5.00	2.00	5.00	4.00	3.00	4.00	2.00	4.00	4.00	4.00	2.00	6.00	2.00
18	18.00	6.00	3.00	6.00	6.00	5.00	5.00	2.00	1.00	3.00	3.00	2.00	4.00	2.00
19	19.00	7.00	3.00	6.00	6.00	6.00		4.00	3.00	3.00	2.00	4.00	2.00	3.00
20	20.00	7.00	1.00	7.00	7.00	6.00	7.00	5.00	2.00	3.00	4.00	3.00	5.00	5.00

빈도분석

변수(V):

- nb4
- mc4
- sn1
- sn2
- jnpur1
- jnpur2
- mexpense
- var00001

통계량(S)...
도표(O)...
형식(F)...
유형(L)...
북스트랩(B)...

☒ 빈도표 표시(D)

확인 불어넣기(P) 재설정(R) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON



3) 빈도분석메뉴 실행결과

- ◆ 이어서 확인버튼을 누르면 빈도분석메뉴가 실행

〈표 8-7A〉 빈도분석결과

통계량

	GENDER	AGE	JOB
N 유효	226	225	226
결측	2	3	2

〈표 8-7B〉 빈도분석결과-성별

GENDER

	빈도	퍼센트	유효퍼센트	누적퍼센트
유효 1.00	88	38.6	38.9	38.9
2.00	138	60.5	161.1	100.0
합계	226	99.1	100.0	
결측 시스템 결측값	2	.9		
합계	228	100.0		

제3절 기술통계분석은 어떻게 실행하는가?

〈표 8-7C〉 빈도분석결과-연령

		AGE			
		빈 도	퍼센트	유효퍼센트	누적퍼센트
유효	20.00	12	5.3	5.3	57.3
	21.00	46	20.2	20.4	66.2
	22.00	38	16.7	16.9	77.8
	23.00	33	14.5	14.7	89.3
	24.00	20	8.8	8.9	94.7
	25.00	26	11.4	11.6	97.3
	26.00	26	11.4	11.6	99.1
	27.00	12	5.3	5.3	99.6
	28.00	6	2.6	2.7	100.0
	29.00	4	1.8	1.8	
	30.00	1	.4	.4	
	32.00	1	.4	.4	
	합 계	225	98.7	100.0	
결측	시스템 결측값	3	1.3		
합계		228	100.0		



제3절 기술통계분석은 어떻게 실행하는가?

〈표 8-7D〉 빈도분석결과-직업

JOB

	빈 도	퍼센트	유효퍼센트	누적퍼센트
유효 1.00	165	72.4	73.0	73.0
2.00	31	13.6	13.7	86.7
3.00	29	12.7	12.8	99.6
11.00	1	.4	.4	100.0
합 계	226	99.1	100.0	
결측 시스템 결측값	2	.9		
합계	228	100.0		



2. 기술통계메뉴

☞ 위에서도 설명한 바와 같이, SPSS Statistics 기술통계도 설문조사에서 수집한 조사 대상 현상의 특징에 대한 중요한 기초정보를 제공합니다.

1) 기술통계메뉴 시작방법

◆ 우선, [그림 8-3A]처럼 분석대상 자료파일을 불러옴

제3절 기술통계분석은 어떻게 실행하는가?

[그림 8-3A] 기술통계메뉴 실행방법

*intmktg.sav [데이터세트1] - IBM SPSS Statistics Data Editor

	2	gender	age	job	mexpen...	var00001
1	00	2.00	23.00	3.00	25.00	.
2	00	2.00	22.00	3.00	20.00	.
3	00	2.00	21.00	3.00	30.00	.
4	00	2.00	21.00	1.00	20.00	.
5	00	2.00	21.00	1.00	5.00	.
6	00	2.00	20.00	1.00	10.00	.
7	00	2.00	21.00	1.00	20.00	.
8	00	2.00	25.00	1.00	35.00	.
9	00	2.00	21.00	1.00	20.00	.
10	00	1.00	21.00	1.00	30.00	.
11	00	2.00	21.00	1.00	25.00	.
12	00	2.00	21.00	1.00	25.00	.
13	00	1.00	21.00	1.00	20.00	.
14	00	2.00	21.00	1.00	20.00	.
15	00	2.00	21.00	1.00	35.00	.
16	00	2.00	21.00	1.00	25.00	.
17	00	1.00	22.00	1.00	20.00	.
18	00	2.00	21.00	3.00	25.00	.
19	00	1.00	25.00	1.00	25.00	.

제3절 기술통계분석은 어떻게 실행하는가?

- ◆ [그림 8-3 C]에서 기술통계의 대상이 되는 변수를 지정

[그림 8-3C] 기술통계메뉴 실행방법

*intmktg.sav [데이터세트1] - IBM SPSS Statistics Data Editor

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

	2	gender	age	job	mexpen...	var00001	변수	변수	변수
1	00	2.00	23.00	3.00	25.00	.			
2	00	2.00	22.00	3.00	20.00	.			
3	00								
4	00								
5	00								
6	00								
7	00								
8	00								
9	00								
10	00								
11	00								
12	00								
13	00								
14	00								
15	00								
16	00	2.00	21.00	1.00	25.00	.			
17	00	1.00	22.00	1.00	20.00	.			
18	00	2.00	21.00	3.00	25.00	.			
19	00	1.00	25.00	1.00	25.00	.			

제3절 기술통계분석은 어떻게 실행하는가?

[그림 8-3D] 기술통계메뉴 실행방법

*intmktg.sav [데이터세트1] - IBM SPSS Statistics Data Editor

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

기술통계

변수(V):

- nb4
- mc4
- sn1
- sn2
- jnpur1
- jnpur2
- mexpense
- var00001

☒ 표준화 값을 변수로 저장(Z)

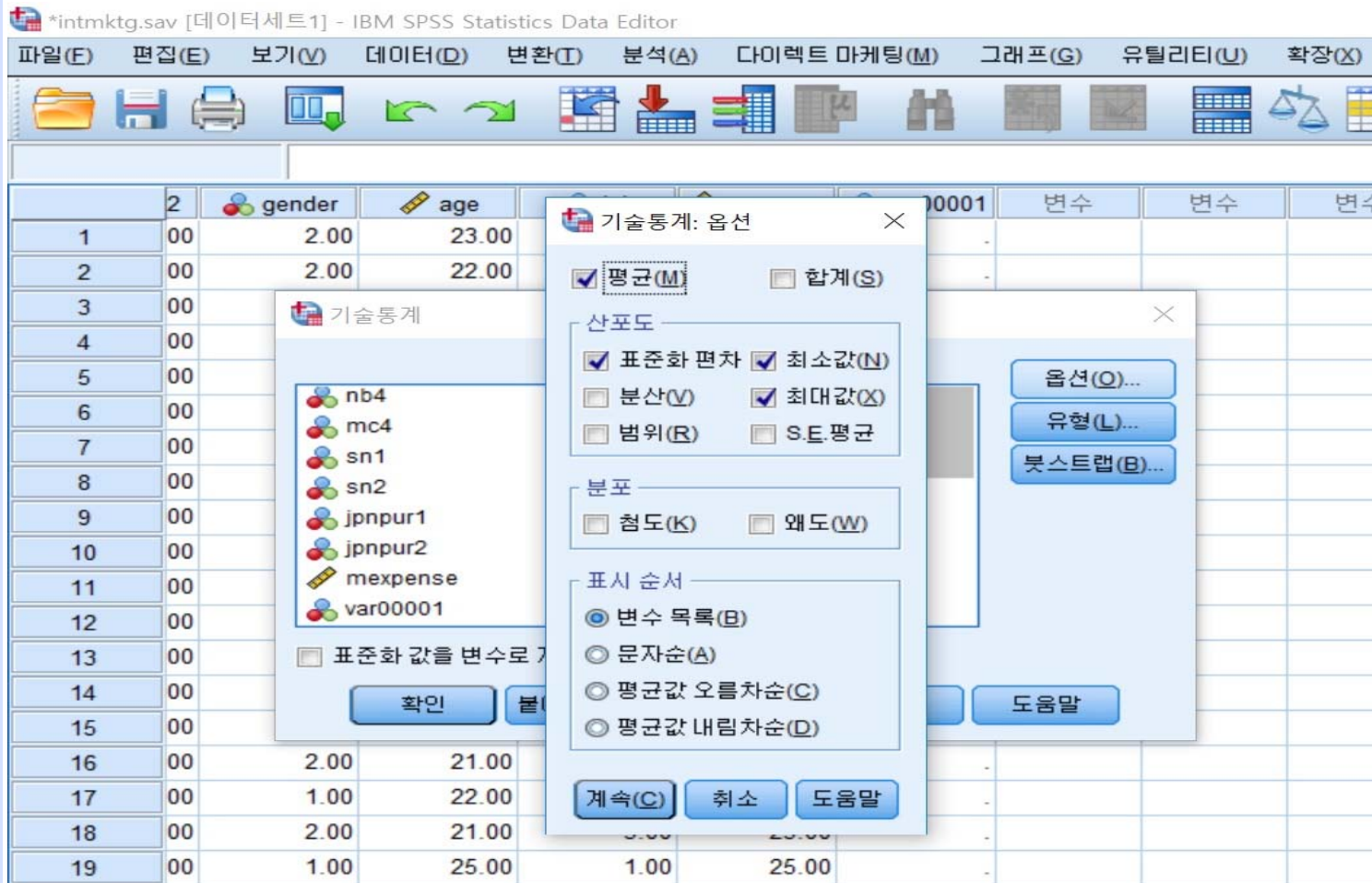
확인 불러넣기(P) 재설정(R) 취소 도움말

	2	gender	age	job	mexpen...	var00001	변수	변수	변수
1	00	2.00	23.00	3.00	25.00	.			
2	00	2.00	22.00	3.00	20.00	.			
3	00								
4	00								
5	00								
6	00								
7	00								
8	00								
9	00								
10	00								
11	00								
12	00								
13	00								
14	00								
15	00								
16	00	2.00	21.00	1.00	25.00	.			
17	00	1.00	22.00	1.00	20.00	.			
18	00	2.00	21.00	3.00	25.00	.			
19	00	1.00	25.00	1.00	25.00	.			

2) 기술통계메뉴 실행방법

- ◆ 옵션버튼을 클릭하면, [그림 8-3E]에서 볼 수 있는 옵션 창이 뜨게 됨
- ☞ 평균, 산포도, 분포, 출력순서 중 원하는 분석내용을 체크하고, 계속 버튼을 누르면 기술통계 메뉴가 실행됨

[그림 8-3E] 기술통계메뉴 실행방법



3) 기술통계메뉴 실행결과

- ◆ 기술통계메뉴가 실행되면 [그림 8-3F]와 같은 결과 창을 얻게 됨
 ➡ 결과 중 이상한 부분은 무엇?

[그림 8-3F] 기술통계메뉴 실행방법

*출력결과1 [문서1] - IBM SPSS Statistics Viewer

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 삽입(I) 형식(O) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

출력결과
 기술통계
 제목
 노트
 기술통계량

기술통계

	N	최소값	최대값	평균	표준편차
gender	226	1.00	2.00	1.6106	.48869
age	225	20.00	32.00	23.4533	2.36960
job	226	1.00	11.00	1.4381	.95133
유효 N(목록별)	225				



1) 기술통계분석과 추리통계분석

- ◆ 기술통계분석(descriptive statistics)
 - 수에 대한 설명의 목적을 가진 통계분석
- ◆ 추리통계분석(inferential statistics)
 - 표본에서의 변수간의 관계가 모집단에서도 성립하는지를 추론(모수추정 및 가설검정)하는 목적을 가진 통계분석
 - 기술통계가 추리통계의 기초를 제공



2) 모수통계분석과 비모수통계분석

- ◆ 모수통계(parametric statistics)
 - 현상이 등간척도 혹은 비율척도로 측정되어 모집단의 특성(분포, 표본오차 등)에 대한 정보가 충분하기 때문에 표본통계량으로 모수추정(parameter estimation)이 가능한 통계분석기법

- ◆ 비모수통계(non-parametric statistics)
 - 현상이 명목척도나 서열척도로 측정되어 모집단의 분포형태나 모수의 특징을 추론해 내기 어려운 경우에 사용하는 분석기법



3) 단일변량통계분석과 다변량통계분석

- ◆ 단일변량 통계분석(univariate statistics)
 - 하나의 변수를 대상으로 하는 통계분석
 - ☞ 기술통계(descriptive statistics) 중심

- ◆ 다변량 통계분석(multivariate statistics)
 - 두 개 이상의 변수들을 대상으로 하는 통계분석
 - ☞ 추측통계(inferential statistics)가 주축

〈표 8-8〉 통계분석기법의 종류

	기술통계*	추측통계	
		모수통계	비모수통계
단일변량통계	빈도분석 대푯값** 산포도	단일평균검정	run test binomial test
다변량통계	요인분석 신뢰성분석	t-test 분산분석 상관관계분석 회귀분석	χ^2 test Kendall test

* 등간/비율척도로 측정된 경우 대부분 모수통계에 속함

** 중위값, 최빈값의 경우 비모수통계에 속함