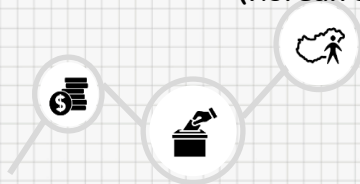


E-stat 으로 알아보는 통계분석

한국사회과학데이터센터
(Korean Social science Data Center)



KSDC DB
STATISTICS & ANALYSIS



슬라이드 번호



E-stat 3.0 ? 3



기술통계량 7



빈도분석 16



교차분석 21



독립표본 t-test 35



대응표본 t-test 46



분산분석 55



상관분석 64



회귀분석 71



결측값 지정 80



데이터 변환 85



E-stat 3.0

- ① E-stat 3.0 소개
- ② 변수 형태에 따른 통계분석 기법



E-STAT 3.0



소장 자료*에 대한
실시간 통계 분석 기능을
제공하고 있습니다.

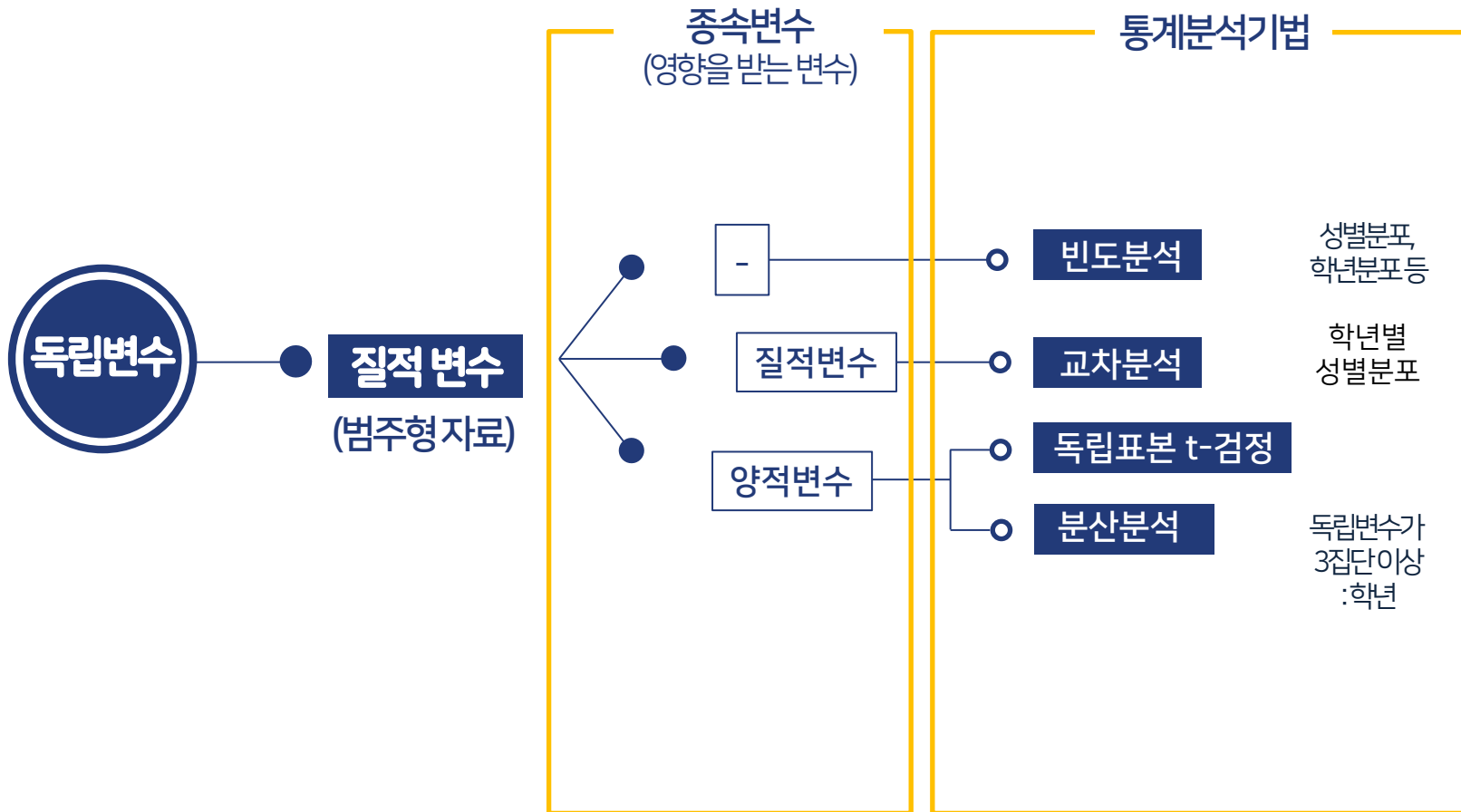
연구자 개인의 파일도
E-STAT 3.0을 통해
분석 가능하도록
지원합니다.

자료의 종류에 따라
제공하는 통계 분석이
다릅니다.

조사자료		통계자료
- 기술통계량	- 분산분석	- 기술통계량
- 그래프	- 상관분석	- 시계열 그래프
- 빈도분석	- 회귀분석	- 상관분석
- 교차분석	- 데이터 변환 및	- 회귀분석
- 독립표본 T검정	결측값 지정	
- 대응표본 T검정		

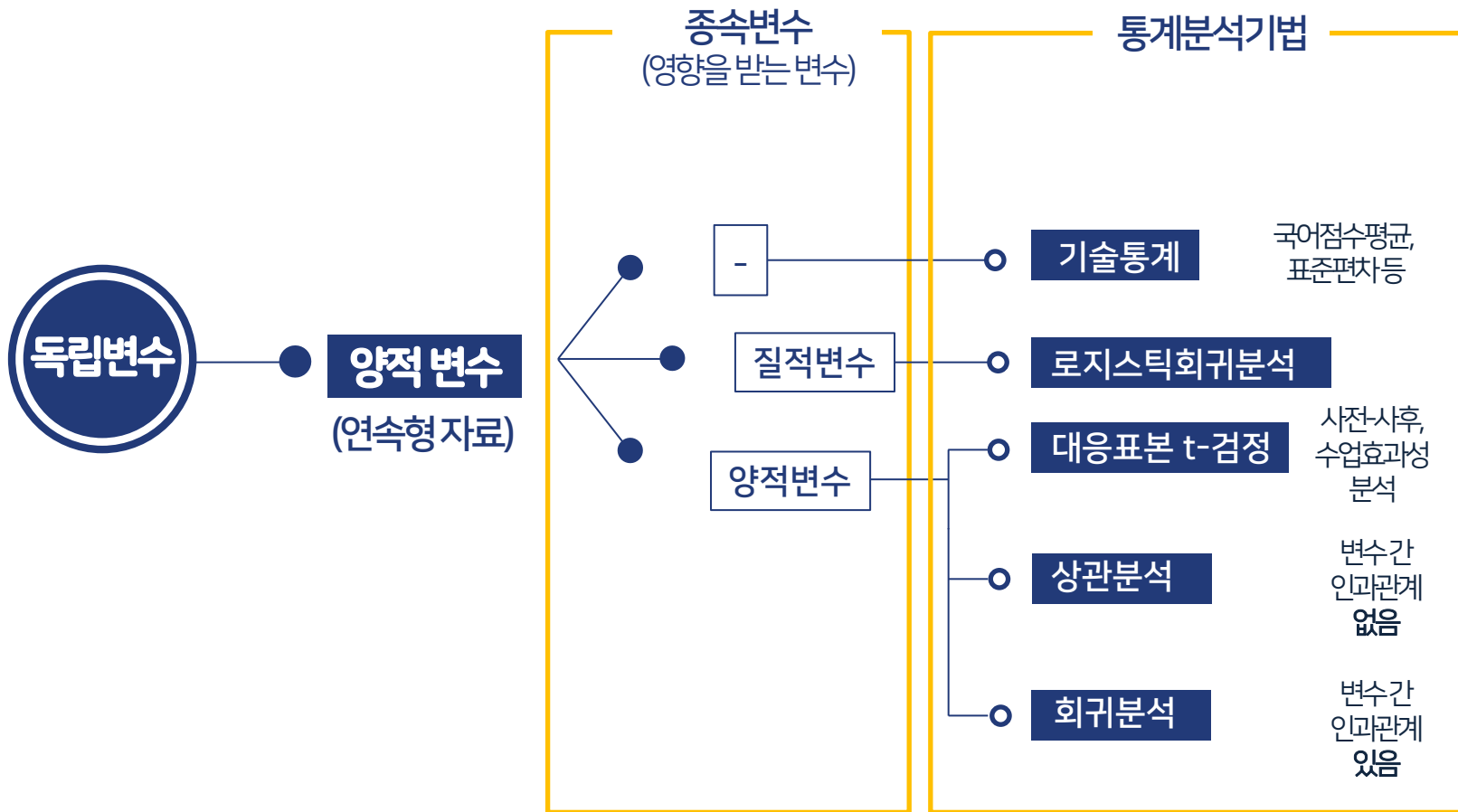


변수 형태에 따른 통계분석 기법





변수 형태에 따른 통계분석 기법





기술통계량

- ① 기초 이론
- ② E-stat 3.0 활용실습



기술통계량

기술통계량이란?

- ✓ 연속형 변수일 때 사용할 수 있는 통계기법
- ✓ 변수를 몇 개의 의미있는 수치로 요약하여, 데이터의 전반적인 특성을 파악할 수 있음
- ✓ 중심위치의 측도, 산포의 측도, 분포의 형태에 관한 측도로 나눌 수 있음

예) 평균, 표준편차, 왜도, 첨도



기술통계량

기술통계량 - 중심위치의 측도

「중심 위치를 측정하는 통계량」

평균 : 어떤 값들 집합의 적절한 특징을 나타내거나 요약하는 것을 의미

중위값 : 데이터를 크기 순으로 나열했을 경우,
가장 중앙에 위치하게 되는 데이터의 값이나 이에 해당되는 값

최빈값 : 관측치 중 가장 여러 번 나타난 값



기술통계량

기술통계량 - 산포의 척도

「표본의 특성을 나타내는 데이터의 퍼져있는 정도를 설명하는 통계량」

최대값 : 데이터에서 가장 큰 값

최소값 : 데이터에서 가장 작은 값

범위 : 최대값과 최소값의 차이. 범위 자체가 산포 정도의 척도로 이용됨

사분위수 : 데이터를 크기에 따라 순서대로 정렬했을 때,
전체 관측값을 4등분 하는 위치에 오는 값 (제2 사분위수 = 중앙값)

분산 : 변수의 흩어진 정도를 계산하는 지표
모든 관측치가 평균에서 얼마나 떨어져 있는가를 나타냄

표준편차 : 분산의 제곱근



기술통계량

기술통계량 - 분포의 형태에 관한 측도

「데이터 분포의 형태와 대칭성을 설명할 수 있는 통계량」

왜도 : 데이터의 기울어진 정도, 비대칭성을 나타내는 통계량

절대값이 클수록 비대칭성이 커짐

(+) 일 때 분포의 꼬리가 오른쪽으로 늘어지고(정적분포),

(-) 일 때 분포의 꼬리가 왼쪽으로 늘어짐(부적분포)

첨도 : 데이터의 뾰족함을 설명하는 통계량

0보다 크면 정규분포보다 뾰족하고, 0보다 작으면 정규분포보다 평평함

Ex) 정규분포는 분포가 일정하고 대칭형이므로 왜도와 첨도가 모두 '0' 임



기술통계량

E-STAT 3.0

통계 분석 > 기술통계량

통계 분석 <

기술통계량

상관분석

회귀분석

작업파일 저장

분석 종료

변수 리스트

시기

국민 총 소득(GNI:원)

선택 변수

그래프 종류 선택

없음 시계열 그래프 1 시계열 그래프 2

분석하기

도움말

좌측메뉴에서 '기술통계량'을 클릭합니다.



기술통계량

E-STAT 3.0

통계 분석 > 기술통계량

통계 분석

기술통계량

상관분석

회귀분석

작업파일 저장

분석 종료

변수 리스트

시기

국민 총 소득(GNI:원)

선택 변수

그래프 종류 선택

없음

시계열 그래프 1

시계열 그래프 2

분석하기





도움말

원하는 변수를 두 리스트 사이에 있는 화살표를 클릭하여 선택 변수로 이동 후, 분석 버튼을 클릭합니다.



기술통계량

분석결과 - E-STAT 3.0 버전을 이용하여 자료를 분석한 결과입니다.

분석결과
 분석방법 : 기술통계량
 선택변수 : 국민 총 소득(GNI 원)

	국민 총 소득(GNI 원)
관측치	66
평균	416,875.212
중위수	92,427.6
최빈수	0
범위	1,781,042.7
분산	300,265,439,966.098
최대값	1,781,091
최소값	48.3
표준편차	547,964.816
1사분위수(25%)	2,266.9
3사분위수(75%)	758,862.6
왜도	1.16
첨도	0.013

메시지

저장 버튼을 클릭하면
분석 결과를 다양한 파일 형식으로
다운로드 받을 수 있습니다.

저장

파일 형식

- Hangul File(*.hwp)
- Microsoft Excel File(*.xlsx)
- Microsoft Presentation(*.ppt)
- Adobe PDF File(*.pdf)
- JPEG Image(*.jpg)

옵션 확인 취소

평균, 표준편차, 왜도 등의 값을 확인할 수 있습니다.



기술통계량

E-STAT 3.0

통계 분석 > 기술통계량

도움말

변수 리스트

- 시기

선택 변수

- 국민 총 소득(GNI:원)

그래프 종류 선택

- 없음
- 시계열 그래프 1
- 시계열 그래프 2

분석하기

통계 자료는 기술통계량 분석메뉴에서 시계열 그래프를 그려볼 수 있습니다. 그래프1과 2는 디자인 차이가 있습니다.





빈도분석

- ① 기초 이론
- ② E-stat 3.0 활용실습



빈도분석

빈도분석이란?

- ✓ 범주형 변수일 때 사용할 수 있는 통계기법
 - ✓ 변수의 분포와 특성을 확인할 수 있음
 - ✓ 각 범주에 응답한 사람의 수 또는 각 응답 범주가 발생한 횟수를 나타냄
- Ex) 전체 응답 인원의 성별 분포. 문항 1번의 2번 보기에 몇 명의 사람이 응답했는가



빈도분석

E-STAT 3.0

통계 분석 > 빈도분석

탐색 분석

통계 분석

- 기술통계량
- 빈도분석**
- 교차분석
- 독립표본T검정
- 대응표본T검정
- 분산분석
- 상관분석
- 회귀분석

데이터 관리

작업파일 저장

분석 종료

변수 리스트

선문1. 응답자 거주지역
선문1-1. 응답자 거주지역 - 시/군/구
선문2. 조사지역 도시규모
선문3. 응답자 성별
문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?
문2. 선생님께서 이번 대통령선거에 얼마나 관심이 있었습니까?

선택 변수

그래프 종류 선택

없음 막대 그래프 원 그래프

분석하기

좌측 메뉴에서 빈도분석을 선택하고
원하는 변수를 두 리스트 사이에 있는 화살표를 클릭하여
선택 변수로 이동 후, 분석 버튼을 클릭합니다.





빈도분석

E-STAT 3.0

통계 분석 > 빈도분석

탐색 분석

통계 분석

- 기술통계량
- 빈도분석**
- 교차분석
- 독립표본T검정
- 대응표본T검정
- 분산분석
- 상관분석
- 회귀분석

데이터 관리

작업파일 저장

분석 종료

변수 리스트

선문1. 응답자 거주지역
선문1-1. 응답자 거주지역 - 시/군/구
선문2. 조사지역 도시규모
선문3. 응답자 성별
문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?
문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?

선택 변수

그래프 종류 선택

없음 막대 그래프 원 그래프

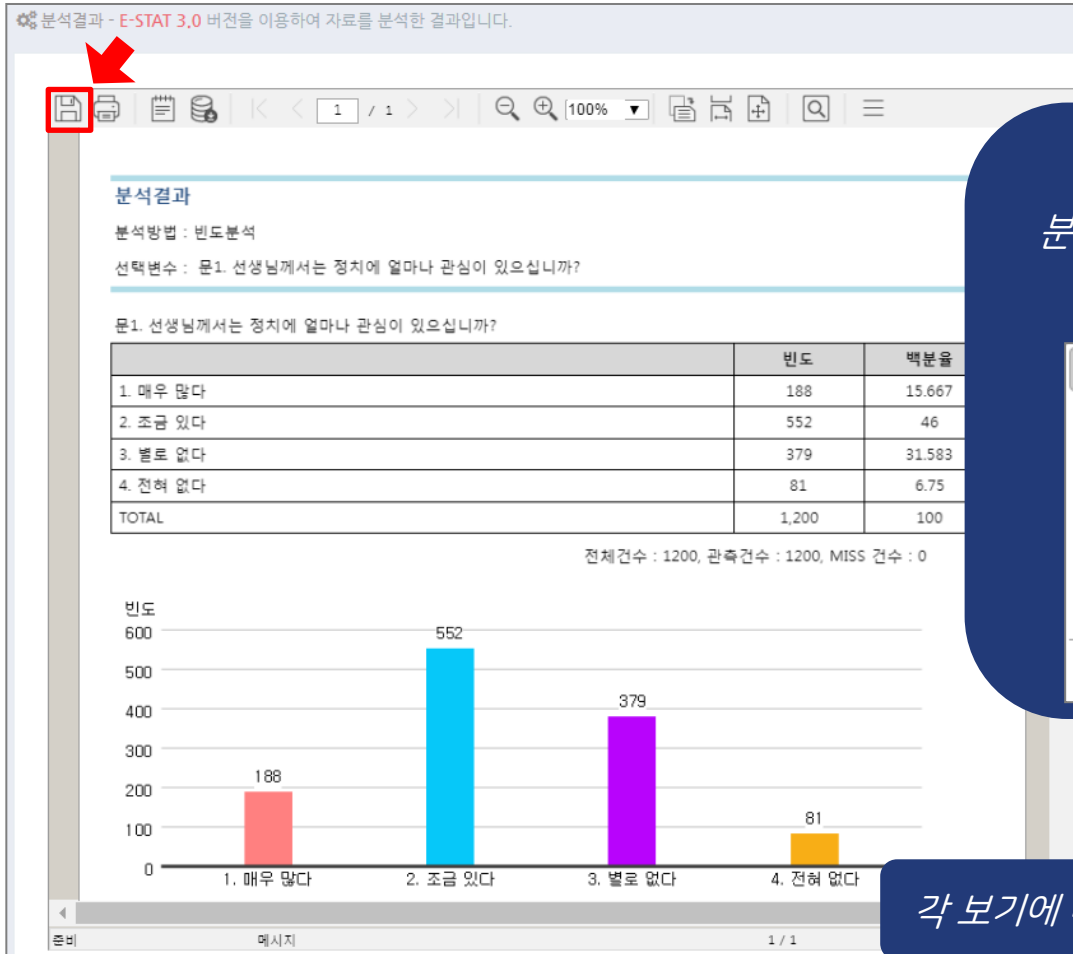
분석하기

그래프도 함께 그리고자 하는 경우에는 원하는 그래프의 모양을 선택하면 됩니다.





빈도분석



저장 버튼을 클릭하면
분석결과를 다양한 파일 형식으로
다운로드 받을 수 있습니다.

저장

파일 형식

- Hangul File (*.hwp)
- Microsoft Excel File (*.xlsx)
- Microsoft Presentation (*.ppt)
- Adobe PDF File (*.pdf)
- JPEG Image (*.jpg)

옵션 확인 취소

각 보기에 대한 빈도표를 확인할 수 있습니다.



교차분석

- ① 기초 이론
- ② E-stat 3.0 활용실습



교차분석

교차분석이란?

- ✓ 범주형 변수일 때 사용할 수 있는 통계기법
- ✓ 변수들간의 범주를 교차시켜 얻은 빈도교차표를 이용하여 해당 변수들 간의 연관성(분포, 비율)을 파악할 수 있음

Ex) 거주지 별 사용하는 통신사 분석 / 성별에 따라 선호하는 자동차 종류 분석



교차분석

교차분석 주요 용어

실제(관찰)빈도 : 실제로 나타난 빈도

기대빈도 : 예측된 빈도

행 백분율 : 행을 기준으로 표기한 백분율

열 백분율 : 열을 기준으로 표기한 백분율

전체 백분율 : 전체를 기준으로 표기한 백분율

카이제곱 검정

: 범주형 자료의 변수간 상관성을 통계적으로 검증하고자 할 때 사용하는 검정방법



교차분석

카이제곱

「카이제곱 검정의 조건」

- ✓ 범주형 자료여야 함
- ✓ 기대빈도가 5이하인 셀이 전체의 20%를 넘지 않아야 함
 - 만약 20%를 넘는다면*
 - ① 표본의 크기를 늘린다.
 - ② 항목을 그룹화하여 빈도수가 낮은 셀을 줄인다.
- ✓ 빈도표의 각 범주끼리는 독립이어야 함
- ✓ 관찰빈도가 0이나 1인 셀이 없어야 함



교차분석

카이제곱

「카이제곱 검정 가설」

- ✓ **영가설 (귀무가설)**: 행변수와 열변수는 서로 연관되어 있지 않다.
- 연구가설 (대립가설)**: 행변수와 열변수는 서로 연관되어 있다.

「카이제곱 검증과 P-value」

- ✓ 유의확률을 뜻하며, 사회과학분야에서는 보통 '0.05'를 기준으로 두고 있음
 - 즉, $P\text{-value} > 0.05$: 통계적으로 의미가 없다 / 영가설 채택
 - $P\text{-value} < 0.05$: 통계적으로 의미가 있다 / 연구가설 채택



교차분석

E-STAT 3.0

통계 분석 > 교차분석

도움말

변수 리스트

- 선문1. 응답자 거주지역
- 선문1-1. 응답자 거주지역 - 시/군/구
- 선문2. 조사지역 도시규모
- 문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?
- 문3. 선생님께서는 이번 선거결과에 대해 어떻게 생각하십니까?
- 문4. 선생님께서는 이번 대통령선거에 투표하셨습니까?

행 변수

- 선문3. 응답자 성별

열 변수

- 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

메이스 변수

좌측메뉴에서 '교차분석'을 클릭합니다.





교차분석

E-STAT 3.0

통계 분석 > 교차분석

변수 리스트

선문1. 응답자 거주지역
 선문1-1. 응답자 거주지역 - 시/군/구
 선문2. 조사지역 도시규모
 문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?
 문3. 선생님께서는 이번 선거결과에 대해 어떻게 생각하십니까?
 문4. 선생님께서는 이번 대통령선거에 투표하셨습니다가?

행 변수

선문3. 응답자 성별

열 변수

문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

베이스 변수

도움말

변수 리스트에서 '기준'이 되는 변수를 '행' 변수로
 '비율 차이를 확인하려는 변수'를 '열' 변수로 옮깁니다
 ex) '성별(선문3)'에 따른 '정치 관심도(문1)'





교차분석

분석결과 - E-STAT 3.0 버전을 이용하여 자료를 분석한 결과입니다.



빈도 교차표를 통해 두 변수의 값이 공유하고 있는 빈도수를 파악할 수 있습니다.

분석결과

분석방법 : 교차분석

열변수 : 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

행변수 : 설문3. 응답자 성별

베이스변수 전체

분석옵션 : 실제빈도, 행백분율,

메인 제목			문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?				
			매우 많다	조금 있다	별로 없다	전혀 없다	전체
설문3. 응답자 성별	남자	실제빈도	141	291	138	25	595
		행백분율	23.70%	48.91%	23.19%	4.20%	100.00%
	여자	실제빈도	47	261	241	56	605
		행백분율	7.77%	43.14%	39.83%	9.26%	100.00%
	전체	실제빈도	188	552	379	81	1200
		행백분율	15.67%	46.00%	31.58%	6.75%	100.00%



교차분석

회귀분석

- 데이터 관리
- 작업파일 저장
- 분석 종료

선문3. 응답자 성별

문1. 선생님께서는 정치에 얼마나 관심이 있으십니까

⬇ ⬆

베이스 변수

중심위치	백분율	■ Chi-square-based Measure
<input checked="" type="checkbox"/> 실제빈도	<input type="checkbox"/> 행 백분율	피어슨 카이제곱(χ^2) 검정
<input type="checkbox"/> 기대빈도	<input checked="" type="checkbox"/> 열 백분율	우도비(Likelihood Ratio) 검정
	<input type="checkbox"/> 전체 백분율	■ 그래프

공유하는 값 뿐만 아니라
E-stat 화면 하단에서 기대빈도 및 백분율,
카이제곱 검정을 확인할 수 있습니다.





교차분석

회귀분석

- 데이터 관리
- 작업파일 저장
- 분석 종료

선문3. 응답자 성별

문1. 선생님께서는 정치에 얼마나 관심이 있으십니까

⬇ ⬆

베이스 변수

이 예제의 영가설은 '성별(선문3)과 정치관심도(문1)은 연관이 없다' 이고
연구가설은 '성별(선문3)과 정치관심도(문1)은 연관이 있다' 입니다.

중심위치	백분율	■ Chi-square-based Measure
<input checked="" type="checkbox"/> 실제빈도	<input type="checkbox"/> 행 백분율	피어슨 카이제곱(χ^2) 검정
<input type="checkbox"/> 기대빈도	<input checked="" type="checkbox"/> 열 백분율	우도비(Likelihood Ratio) 검정
	<input type="checkbox"/> 전체 백분율	■ 그래프





교차분석

결과창 첫 번째 표

메인 제목			문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?				
			매우 많다	조금 있다	별로 없다	전혀 없다	전체
선문3. 응답자 성별	남자	실제빈도	141	291	138	25	595
		기대빈도	93.2	273.7	187.9	40.2	595.0
		행백분율	23.70%	48.91%	23.19%	4.20%	100.00%
		열백분율	75.00%	52.72%	36.41%	30.86%	49.58%
		전체백분율	11.75%	24.25%	11.50%	2.08%	49.58%
	여자	실제빈도	47	261	241	56	605
		기대빈도	94.8	278.3	191.1	40.8	605.0
		행백분율	7.77%	43.14%	39.83%	9.26%	100.00%
		열백분율	25.00%	47.28%	63.59%	69.14%	50.42%
		전체백분율	3.92%	21.75%	20.08%	4.67%	50.42%
	전체	실제빈도	188	552	379	81	1200
		기대빈도	188.0	552.0	379.0	81.0	1,200.0
		행백분율	15.67%	46.00%	31.58%	6.75%	100.00%
		열백분율	100.00%	100.00%	100.00%	100.00%	100.00%
전체백분율		15.67%	46.00%	31.58%	6.75%	100.00%	

- 행백분율 행의 총합이 100%
- 열백분율 열의 총합이 100%
- 전체백분율 전체 총합이 100%

빈도교차표를 이용하여
각 빈도 및 백분율을 확인할 수 있습니다.




교차분석

결과창 두 번째 표

	자유도	통계량	유의확률(P-Value)
Pearson's Chi-Square(χ^2)	3	88.41	0.000
Likelihood Ratio	3	91.25	0.000
Phi Coefficient		0.271	0.00
Cramer's V		0.27	0.00
Contingency Coefficient		0.26	

[χ^2 값과 P-value]

- 1) P-value < 0.05 : χ^2 값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : χ^2 값이 유의하지 않음 (통계적으로 의미가 없음)

 빈도교차표 하단의 표를 이용하여 카이제곱 값을 포함한 다양한 검정 값 및 P-value를 확인할 수 있습니다.



교차분석

💡 결과창 두 번째 표

	자유도	통계량	유의확률(P-Value)
Pearson's Chi-Square(χ^2)	3	88.41	0.000
Likelihood Ratio	3	91.25	0.000
Phi Coefficient		0.271	0.00
Cramer's V		0.27	0.00
Contingency Coefficient		0.26	

[χ^2 값과 P-value]

- 1) P-value < 0.05 : χ^2 값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : χ^2 값이 유의하지 않음 (통계적으로 의미가 없음)

💡
다양한 검정통계량 중에서
피어스 카이제곱 값을
주로 확인합니다.



교차분석

💡 결과창 두 번째 표

	자유도	통계량	유의확률(P-Value)
Pearson's Chi-Square(χ^2)	3	88.41	0.000
Likelihood Ratio	3	91.25	0.000
Phi Coefficient		0.271	0.00
Cramer's V		0.27	0.00
Contingency Coefficient		0.26	

[χ^2 값과 P-value]

- 1) P-value < 0.05 : χ^2 값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : χ^2 값이 유의하지 않음 (통계적으로 의미가 없음)

💡

카이제곱 검정결과
P-value < 0.05 이므로 연구가설을 채택하여
 '성별'과 '정치관심도'는
 서로 연관이 있다고 해석할 수 있습니다.



독립표본 t-검정

- ① 기초 이론
- ② E-stat 3.0 활용실습



독립표본 t-검정

독립표본 t-test에 들어가기 전에

t - 검정 이란?

- ✓ 두 집단 간의 평균차이를 비교하여 가설을 검정하는 통계기법
- ✓ 두 집단이 독립표본인 경우 : 독립표본 t-test
(ex. 남자와 여자 각 100명을 대상으로 언어능력 검사 결과 차이가 의미 있는가)
- ✓ 두 집단이 종속표본인 경우 : 대응표본 t-test
(ex. 동일한 집단을 대상으로 교육 전과 교육 후의 차이가 의미 있는가)
- cf) 집단이 3개 이상일 경우에는 분산분석(ANOVA) 를 이용
(ex. 10대, 20대, 30대의 소득수준에 대한 차이의 의미가 있는가)



독립표본 t-검정

독립표본 t - 검정 이란?

- ✓ 독립변수는 범주형 변수, 종속변수는 연속형 범주일 때 사용할 수 있는 통계기법
- ✓ 서로 독립적인 두 집단 간의 표본을 바탕으로 두 집단 간 평균 차이를 비교하여 가설을 검정하는 방법
- ✓ 영가설 (귀무가설) : 집단에 따른 평균의 차이가 없다.
- ✓ 연구가설 (대립가설) : 집단에 따른 평균의 차이가 있다.

ex) 성별에 따른 중간고사 점수의 차이가 통계적으로 의미 있는가

동일한 제품을 생산하는 A공장의 일일 생산량과 B공장의 일일 생산량의 차이가 통계적으로 의미 있는가



독립표본 t-검정

독립표본 t-검정의 조건

- ✓ 독립성 : 두 변수는 서로 독립된 값이어야 함
- ✓ 정규성 : 각 변수들은 정규분포여야 함
- ✓ 등분산성 : 독립변수에 따른 종속변수의 분산의 동질성 여부에 따라 분석 공식이 달라짐

등분산성의 충족 여부에 따라 E-stat 창에서 '등분산이 가정됨', 또는 '등분산이 가정되지 않음'의 결과값을 활용해야 함.



독립표본 t-검정

E-STAT 3.0

통계 분석 > 독립표본T검정

도움말

변수 리스트

- 선문1. 응답자 거주지역
- 선문1-1. 응답자 거주지역 - 시/군/구
- 선문2. 조사지역 도시규모
- 문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?
- 문3. 선생님께서는 이번 선거결과에 대해 어떻게 생각하십니까?
- 문4. 선생님께서는 이번 대통령선거에 투표하십니까?

선택 변수

- 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까

그룹 변수

- 선문3. 응답자 성별

검정방법

유의 수준
(유의수준은 0과 1사이의 값입니다.)

Next

좌측메뉴에서 '독립표본 T검정'을 클릭합니다.





독립표본 t-검정

E-STAT 3.0

통계 분석 > 독립표본T검정

도움말

변수 리스트

선문1. 응답자 거주지역
 선문1-1. 응답자 거주지역 - 시/군/구
 선문2. 조사지역 도시규모
 문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?
 문3. 선생님께서는 이번 선거결과에 대해 어떻게 생각하십니까?
 문4. 선생님께서는 이번 대통령선거에 투표하셨습니다?

선택 변수

문1. 선생님께서는 정치에 얼마나 관심이 있으십니까

그룹 변수

선문3. 응답자 성별

검정방법

유의 수준
 (유의수준은 0과 1사이의 값입니다.)

Next

변수리스트에서
 차이를 확인하고 싶은 변수를
 '선택변수'로,
 두 집단을 지닌 변수를
 '그룹변수'로 이동합니다.





독립표본 t-검정

E-STAT 3.0

통계 분석 > 독립표본T검정

도움말

변수 리스트

선문1. 응답자 거주지역
 선문1-1. 응답자 거주지역 - 시/군/구
 선문2. 조사지역 도시규모
 문2. 선생님께서는 이번 대통령선거에 얼마나 관심이 있었습니까?
 문3. 선생님께서는 이번 선거결과에 대해 어떻게 생각하십니까?
 문4. 선생님께서는 이번 대통령선거에 투표하십니까?

선택 변수

문1. 선생님께서는 정치에 얼마나 관심이 있으십니까

그룹 변수

선문3. 응답자 성별

이 예제에서
 영가설은 '성별과 정치관심도의 평균 차이는 의미가 없다'
 연구가설은 '성별과 정치관심도의 평균차이는 의미가 있다' 입니다.

Next



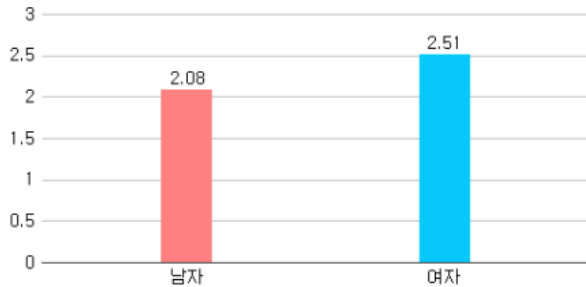


독립표본 t-검정

5. 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

* 기술통계량

그룹	빈도	평균	표준편차	표준오차	최소	최대
남자	595	2.079	0.795	0.033	1.0	4.0
여자	605	2.506	0.769	0.031	1.0	4.0



*Levene의 등분산 검정

F	유의확률 (P-value)
9.427	0.002

[Levene의 등분산 검정의 P-value]

- 1) P-value < 0.05 : 두 변수 사이의 분산이 동일하지 않다.(→등분산이 가정되지 않음)
- 2) P-value > 0.05 : 두 변수 사이의 분산이 동일하다.(→등분산이 가정됨)

*독립표본T검정

	자유도	T	유의확률 (P-value)	평균차	95.0% 하한	95.0% 상한
등분산이 가정됨	1,198	-9.453	0.0	-0.427	-0.515	-0.338
등분산이 가정되지않음	1,195	-9.451	0.0	-0.427	-0.515	-0.338

[T값과 P-value]

- 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)

Levene의 등분산 검정에서 P값이 0.05보다 크면 아래 표의 '등분산이 가정됨 사용' 0.05보다 작으면 '등분산이 가정되지 않음' 을 사용하면 됩니다.



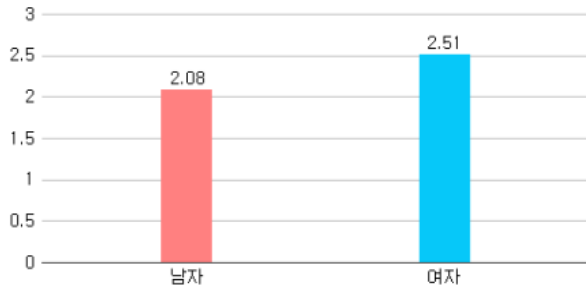


독립표본 t-검정

5. 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

* 기술통계량

그룹	빈도	평균	표준편차	표준오차	최소	최대
남자	595	2.079	0.795	0.033	1.0	4.0
여자	605	2.506	0.769	0.031	1.0	4.0



*Levene의 등분산 검정

F	유의확률 (P-value)
9.427	0.002

[Levene의 등분산 검정의 P-value]
 1) P-value < 0.05 : 두 변수 사이의 분산이 동일하지 않다.(→등분산이 가정되지 않음)
 2) P-value > 0.05 : 두 변수 사이의 분산이 동일하다.(→등분산이 가정됨)



*독립표본T검정

	자유도	T	유의확률 (P-value)	평균자	95.0% 하한	95.0% 상한
등분산이 가정됨	1,198	-9.453	0.0	-0.427	-0.515	-0.338
등분산이 가정되지않음	1,195	-9.451	0.0	-0.427	-0.515	-0.338

[T값과 P-value]
 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)

P-value < 0.05 이므로
 '등분산이 가정되지 않음' 값을 확인합니다.



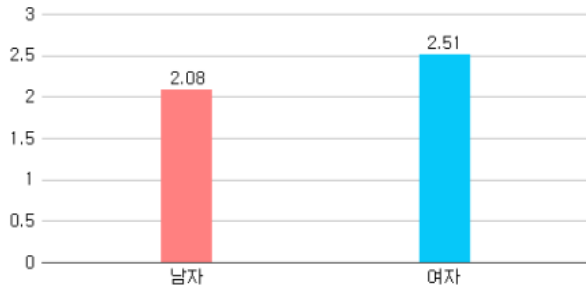


독립표본 t-검정

5. 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

* 기술통계량

그룹	빈도	평균	표준편차	표준오차	최소	최대
남자	595	2.079	0.795	0.033	1.0	4.0
여자	605	2.506	0.769	0.031	1.0	4.0



*Levene의 등분산 검정

F	유의확률 (P-value)
9.427	0.002

[Levene의 등분산 검정의 P-value]

- 1) P-value < 0.05 : 두 변수 사이의 분산이 동일하지 않다.(→등분산이 가정되지 않음)
- 2) P-value > 0.05 : 두 변수 사이의 분산이 동일하다.(→등분산이 가정됨)

*독립표본T검정

	자유도	T	유의확률 (P-value)	평균자	95.0% 하한	95.0% 상한
등분산이 가정됨	1,198	-9.453	0.0	-0.427	-0.515	-0.338
등분산이 가정되지않음	1,195	-9.451	0.0	-0.427	-0.515	-0.338

[T값과 P-value]

- 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)

P-value < 0.05 : t값이 유의함(통계적으로 의미가 있음)
P-value > 0.05 : t값이 유의하지 않음(통계적으로 의미가 없음)



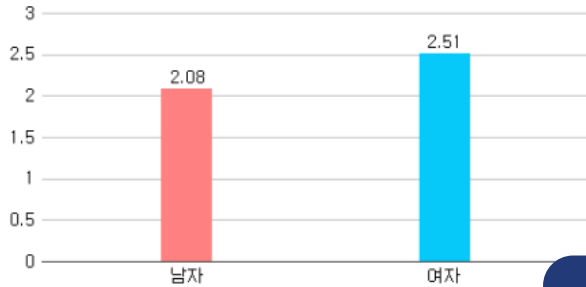


독립표본 t-검정

5. 문1. 선생님께서는 정치에 얼마나 관심이 있으십니까?

* 기술통계량

그룹	빈도	평균	표준편차	표준오차	최소	최대
남자	595	2.079	0.795	0.033	1.0	4.0
여자	605	2.506	0.769	0.031	1.0	4.0



*Levene의 등분산 검정

F	유의확률 (P-value)
9.427	0.002

독립표본 t-검정 결과 P-value < 0.05 이므로, 연구가설을 채택하여 '성별'과 '정치관심도'의 평균차이가 유의하다고 해석할 수 있습니다.

[Levene의 등분산 검정의 P-value]

- 1) P-value < 0.05 : 두 변수 사이의 분산이 동일하지 않다.(→등분산이 가정되지 않음)
- 2) P-value > 0.05 : 두 변수 사이의 분산이 동일하다.(→등분산이 가정됨)

*독립표본T검정

	자유도	T	유의확률 (P-value)	평균자	95.0% 하한	95.0% 상한
등분산이 가정됨	1,198	-9.453	0.0	-0.427	-0.515	-0.338
등분산이 가정되지않음	1,195	-9.451	0.0	-0.427	-0.515	-0.338

[T값과 P-value]

- 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)





대응표본 t-검정

- ① 기초 이론
- ② E-stat 3.0 활용실습



대응표본 t - 검정

대응표본 t-test에 들어가기 전에

t - 검정 이란?

- ✓ 두 집단 간의 평균차이를 비교하여 가설을 검정하는 통계기법
- ✓ 두 집단이 독립표본인 경우 : 독립표본 t-test
(ex. 남자와 여자 각 100명을 대상으로 언어능력 검사 결과 차이가 의미 있는가)
- ✓ 두 집단이 종속표본인 경우 : 대응표본 t-test
(ex. 동일한 집단을 대상으로 교육 전과 교육 후의 차이가 의미 있는가)
- cf) 집단이 3개 이상일 경우에는 분산분석(ANOVA) 를 이용
(ex. 10대, 20대, 30대의 소득수준에 대한 차이의 의미가 있는가)



대응표본 t - 검정

대응표본 t - 검정 이란?

- ✓ 독립변수와 종속변수가 연속형 변수 일 때 사용할 수 있는 통계기법
- ✓ 동일 집단에서 실험 전과 후의 평균 차이를 비교하여 가설을 검정하는 방법
- ✓ 광고, 약품, 교육 등이 연구 대상자에게 얼마나 효과적인지 측정할 수 있음
- ✓ 영가설 : 실험 전과 후의 평균값은 통계적으로 차이가 없다.
- ✓ 연구가설 : 실험 전과 후의 평균값은 통계적으로 차이가 있다.

ex) 광고를 보기 전과 보고 난 후의 A 제품에 대한 호감도의 차이가 의미가 있는가
영어 학습 프로그램에 대한 교육 전과 후 점수의 차이가 의미가 있는가



대응표본 t - 검정



T값과 P-value

p값은 유의확률을 뜻하며, 사회과학분야에서는 보통 '0.05'를 기준으로 두고 있음

즉, $P\text{-value} > 0.05$: t값이 유의하지 않다 / 통계적으로 의미가 없다 / 영가설 선택

$P\text{-value} < 0.05$: t값이 유의하다 / 통계적으로 의미가 있다 / 연구가설 선택

대응차이의 평균값 = 사전조사 - 사후조사

즉,

“ 이 값이 양의 값(+) 이라면 사전조사>사후조사
이 값이 음의 값(-) 이라면 사전조사<사후조사 ”



대응표본 t - 검정

💡 대응표본 t-test 분석용 파일 만들기

사전 자료

	A	B	C	D	E
1	ID	학력수준	국어선호정도	수업만족도	(사전)국어공부즐거움
2	1	4	3	4	2
3	2	4	3	2	3
4	3	3	1	3	3
5	4	5	4	5	3
6	5	4	3	4	4
7	6	4	3	4	3
8	7	2	2	3	3
9	8	3	4	3	5
10	9	5	5	4	4
11	10	4	5	4	3
12	11	4	4	5	2
13	12	5	4	4	3
14	13	3	3	3	3
15	14	2	2	2	4
16	15	2	2	2	3

사후 자료

	A	B	C	D	E
1	ID	학력수준	국어선호	수업만족도	(사후)국어공부즐거움
2	1	4	3	4	3
3	2	4	3	2	3
4	3	3	1	3	4
5	4	5	4	5	4
6	5	4	3	4	4
7	6	4	3	4	2
8	7	2	2	3	5
9	8	3	4	3	4
10	9	5	5	4	4
11	10	4	5	4	4
12	11	4	4	5	3
13	12	5	4	4	4
14	13	3	3	3	4
15	14	2	2	2	3

	A	B
1	(사전)국어공부즐거움	(사후)국어공부즐거움
2		2
3		3
4		3
5		3
6		4
7		3
8		3
9		5
10		4
11		3
12		2
13		3
14		3
15		4
16		2
17		4
18		3

사전+사후
통합자료

- ① 사전 및 사후자료에서 대응표본 t-검정용 문항을 파악한다.
- ② 사전과 사후를 매칭한 통합자료를 엑셀파일로 만든다.
(이 때, 동일 ID는 같은 행에 들어가야 하며 왼쪽이 사전, 오른쪽이 사후여야 한다.)
- ③ 'ONLINE ANALYSIS' 를 통해 엑셀파일 업로드 후 분석을 진행한다.
(이 때, 사전-사후 매칭되는 문항이 총 3쌍 이라면 1쌍씩 3번 분석을 진행한다.)



대응표본 t - 검정

E-STAT 3.0 통계 분석 > 대응표본T검정

변수 리스트

- 교육 전
- 교육 후

선택 변수

검정방법

유의 수준
(유의수준은 0과 1사이의 값입니다.)

분석하기

좌측 메뉴에서 대응표본T검정을 선택하고
원하는 변수를 두 리스트 사이에 있는 화살표를 클릭하여
선택 변수로 이동 후, 분석하기 버튼을 클릭합니다.





대응표본 t - 검정

E-STAT 3.0

통계 분석 > 대응표본t검정

도움말

변수 리스트

- 교육 전
- 교육 후

선택 변수

검정방법

유의 수준
(유의수준은 0과 1사이의 값입니다.)

분석하기

이 예제에서 영가설은 '교육 전과 후의 평균 차이는 의미가 없다'
연구가설은 '교육 전과 후의 평균차이는 의미가 있다' 입니다.

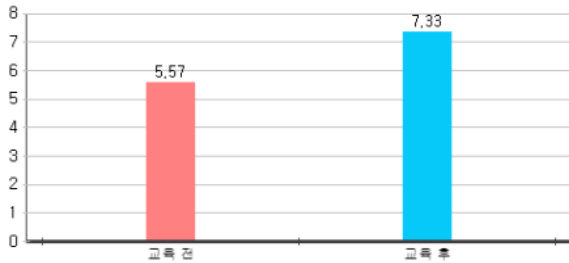




대응표본 t - 검정

대응표본 통계

	평균	빈도	표준편차
교육 전	5.567	30	1.755
교육 후	7.333	30	1.124



대응표본 t-검정 결과
P-value < 0.05 이므로, 연구가설을 채택하여
 '교육 전과 후'의 평균차이가 유의하다고 해석할 수 있습니다.



대응표본상관(Correlation)

빈도	Pearson Correlation	통계량	유의확률 (P-value)
30	0.495	3.014	0.005

대응표본검정(Paired T-TEST)

T	자유도	유의확률 (P-value)	95.0% 상한	95.0% 하한
-6.257	29.000	0.000	-2.344	-1.189

대응차이(Difference)

대응차이		
평균	표준편차	표준오차
-1.767	1.547	0.282

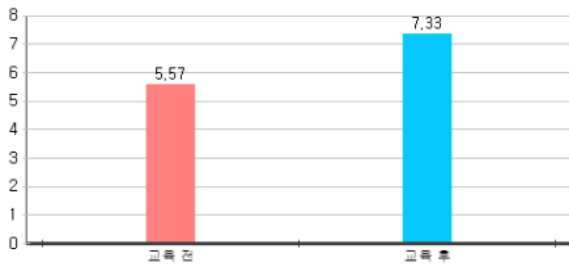
[T값과 P-value]
 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)



대응표본 t - 검정

대응표본 통계

	평균	빈도	표준편차
교육 전	5.567	30	1.755
교육 후	7.333	30	1.124



대응표본상관(Correlation)

빈도	Pearson Correlation	통계량	유의확률 (P-value)
30	0.495	3.014	0.005

대응표본검정(Paired T-TEST)

T	자유도	유의확률 (P-value)	95.0% 상한	95.0% 하한
-6.257	29.000	0.000	-2.344	-1.189

대응차이(Difference)

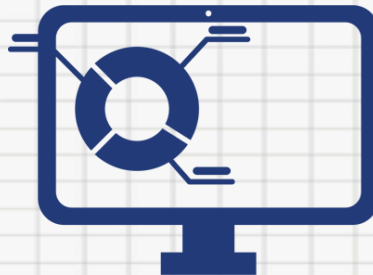
대응차이		
평균	표준편차	표준오차
-1.767	1.547	0.282

[T값과 P-value]

- 1) P-value < 0.05 : T값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : T값이 유의하지 않음 (통계적으로 의미가 없음)

대응차이의 평균값은 사전조사-사후조사를 한 값으로
음의 값이 나왔기 때문에, 사전조사보다
사후조사의 평균값이 높다는 것을 의미합니다.





분산분석

- ① 기초 이론
- ② E-stat 3.0 활용실습



분산분석

일원분산분석 (one-way ANOVA)?

- ✓ 독립변수가 범주형, 종속변수가 연속형 변수일 때 사용할 수 있는 통계기법
- ✓ 세 집단 이상의 집단간, 집단내 평균과 분산을 비교하여 가설을 검정하는 방법
- ✓ 영가설 : 집단들의 평균이 모집단 평균과 동일하다.
- ✓ 연구가설 : 집단들의 평균이 모두 같지는 않다.

ex) 우리나라 광역시의 복지 정책에 대한 만족도는 지역별로 차이가 있는가?
대학생의 학식 만족도는 학년별로 차이가 있는가?



분산분석



분산분석 가정

- 표본이 추출된 모집단의 분포는 정규분포이다.
- 표본이 추출된 모집단은 동일한 분산이나 표준편차를 갖는다.
- 서로 다른 모집단에서 추출된 표본들은 확률적으로 독립이다.



분산분석

분산분석 주요 용어

(편차) 제곱합 : 전체편차 = 집단간 편차 + 집단내 편차

편차의 총 합은 0 이므로 제곱식을 이용하게 됨

* 전체제곱합 (SST) = 집단간 편차제곱합(SSB) + 집단내 편차제곱합(SSW)

집단간 평균제곱 : 각 집단의 평균을 전체평균과의 편차를 토대로 얻은 값

집단내 평균제곱 : 집단평균으로부터 각 값의 편차를 토대로 얻은 값



분산분석



F값과 P-value

p값은 유의확률을 뜻하며, 사회과학분야에서는 보통 '0.05'를 기준으로 두고 있음

즉, $P\text{-value} > 0.05$: F값이 유의하지 않다 / 통계적으로 의미가 없다 / 영가설 채택

$P\text{-value} < 0.05$: F값이 유의하다 / 통계적으로 의미가 있다 / 연구가설 채택



분산분석

The screenshot shows the E-STAT 3.0 interface. On the left sidebar, the '분산분석' (ANOVA) option is highlighted with a red box. The main workspace is titled '통계 분석 > 분산분석'. It features a '변수 리스트' (Variable List) area at the top. Below it, two panels are visible: '종속 변수' (Dependent Variable) containing '학식만족도' and '그룹 변수' (Group Variable) containing '학년'. At the bottom, there are two control panels: '통계량' (Statistics) with a checked box for '기술 통계량' (Descriptive Statistics), and '검정방법' (Test Method) with a significance level of 0.05. A '분석하기' (Analyze) button is located at the bottom center.

좌측 메뉴에서 분산분석을 선택하고
변수리스트에서 차이를 확인하고 싶은 변수를 '종속변수'로,
세 집단 이상을 지닌 변수를 '그룹변수'로 이동합니다.





분산분석

E-STAT 3.0

통계 분석 > 분산분석

도움말

변수 리스트

이 예제에서
영가설은 '학식만족도는 학년에 따른 차이가 없다'이며,
연구가설은 '학식만족도는 학년에 따라 다르다' 입니다.

종속 변수

학식만족도

그룹 변수

학년

통계량

기술 통계량

검정방법

유의 수준
(유의수준은 0과 1사이의 값입니다.)

분석하기





분산분석

2. 학식만족도

* 기술통계량

그룹	수준	빈도	평균	표준편차	최소	최대
학년	1.0	18	7.111	1.641	3.000	9.000
	2.0	18	6.778	1.517	5.000	9.000
	3.0	18	7.000	1.455	5.000	9.000
	4.0	18	7.056	1.434	5.000	9.000

* 분산의 동질성 검정

df1 (집단간 자유도)	df2 (집단내 자유도)	F (Levene 통계량)	유의확률 (P-value)
3.000	68.000	0.020	0.996

[분산의 동질성 검정의 P-value]
 1) P-value < 0.05 : 집단외 모집단 분산이 동질하지 않다.
 2) P-value > 0.05 : 집단외 모집단 분산이 동질하다.

* 분산 분석표

Source	제곱합	자유도	평균제곱	F	유의확률 (P-value)
집단간	1.153	3	0.384	0.168	0.918
집단내	155.833	68	2.292		
합계	156.986	71			

[분산 분석표의 F값과 P-value]
 1) P-value < 0.05 : F값이 유의함 (통계적으로 의미가 있음)
 2) P-value > 0.05 : F값이 유의하지 않음 (통계적으로 의미가 없음)

동질성 검정에서 P-value를 확인합니다.
 P-Value > 0.05 : 분산이 동질함
 P-Value < 0.05 : 분산이 동질하지 않음





분산분석

2. 학식만족도

* 기술통계량

그룹	수준	빈도	평균	표준편차	최소	최대
학년	1.0	18	7.111	1.641	3.000	9.000
	2.0	18	6.778	1.517	5.000	9.000
	3.0	18	7.000	1.455	5.000	9.000
	4.0	18	7.056	1.434	5.000	9.000

* 분산의 동질성 검정

df1 (집단간 자유도)	df2 (집단내 자유도)	F (Levene 통계량)
3.000	68.000	0.020

[분산의 동질성 검정의 P-value]

- 1) P-value < 0.05 : 집단내 모집단 분산이 동질하지 않다.
- 2) P-value > 0.05 : 집단내 모집단 분산이 동질하다.

분산분석 결과 P-value > 0.05 이기 때문에,
영가설을 기각하지 않고 '학식만족도는 학년에 관계없이 동일한 수준'
이라고 해석할 수 있습니다.

* 분산 분석표

Source	제곱합	자유도	평균제곱	F	유의확률 (P-value)
집단간	1.153	3	0.384	0.168	0.918
집단내	155.833	68	2.292		
합계	156.986	71			

[분산 분석표의 F값과 P-value]

- 1) P-value < 0.05 : F값이 유의함 (통계적으로 의미가 있음)
- 2) P-value > 0.05 : F값이 유의하지 않음 (통계적으로 의미가 없음)



상관분석

- ① 기초 이론
- ② E-stat 3.0 활용실습



상관분석

상관분석이란?

- ✓ 독립변수와 종속변수가 모두 연속형 변수일 때 사용할 수 있는 통계기법
- ✓ ****상관계수를 통해 변수 간 관계성의 방향과 강도를 파악할 수 있음**
- ✓ 두 변수의 연관성만을 알 수 있고, 어떤 변수가 원인인지 파악할 수 없음

예) 키와 몸무게의 관련성, 나이와 스마트폰 사용시간의 관련성



상관분석

** 상관계수란 ? 변수간의 관련된 정도를 나타내 주는 값

상관계수(r)와 상관성 정도

± 0.9이상:매우 높은상관관계

± 0.7~0.9미만:높은상관관계

± 0.4~0.7미만:다소 높은상관관계

± 0.2~0.4미만:낮은상관관계

± 0.2미만:상관관계가 거의 없음



상관분석

상관계수가 양수일 때 : 변수간 증감의 방향이 같은 경우

상관계수가 음수일 때 : 변수간 증감의 방향이 반대인 경우

✓ 주의사항

r값은 측정치가 아니기 때문에 0.5는 0.25보다 상관도가 두 배 높다고 할 수 없고,
r값이 0.4와 0.6일 때 두 값의 차이가 0.5와 0.7의 차이와 같다고 할 수 없음



상관분석

E-STAT 3.0

통계 분석 > 상관분석

변수 리스트

시기
국민 총 소득(GNI:원)
종합병원 입원환자수

선택 변수

기술 통계량 Pearson Spearman Cronbach's alpha(reliability) 분포도

분석하기

도움말

좌측 메뉴에서 상관분석을 선택하고
변수리스트에서 연관성을 확인할 변수를 선택변수로 옮깁니다.





상관분석

* 기술통계량

변수	빈도	평균	표준편차	최소	최대
1. 국민 총 소득(GNI:원)	15	1,045,336.933	469,568.266	197,414.7	1,646,208.9
2. 종합병원,입원환자수	15	109,946.933	20,974.243	67,716.0	141,920.0

* Pearson

	국민 총 소득(GNI:원)	종합병원,입원환자수
국민 총 소득(GNI:원)		0.981 18.391 0.98
종합병원,입원환자수	0.981 18.391 0.0	

상관계수, Approximately t-Statistics, 양측 검정에 대한 p-value

상관계수
t값
P값

*상관계수값(r)과 상관성 정도

- 1) ±0.9 이상 : 매우 높은 상관관계
- 2) ±0.7~0.9 미만 : 높은 상관관계
- 3) ±0.4~0.7 미만 : 다소 높은 상관관계
- 4) ±0.2~0.4 미만 : 낮은 상관관계
- 5) ±0.2 미만 : 상관관계가 거의 없음

상관계수(r)



산출된 r값

P값



상관계수의 유의확률

(상관계수가 통계적으로 유의한지, 그렇지 않은지 판단)



상관분석

* 기술통계량

변수	빈도	평균	표준편차	최소	최대
1. 국민 총 소득(GNI:원)	15	1,045,336.933	469,568.266	197,414.7	1,646,208.9
2. 종합병원,입원환자수	15	109,946.933	20,974.243	67,716.0	141,920.0

* Pearson

	국민 총 소득(GNI:원)	종합병원,입원환자수
국민 총 소득(GNI:원)		0.981 18.391 0.0
종합병원,입원환자수	0.981 18.391 0.0	

상관계수, Approximately t-Statistics, 양측 검정에 대한 p-value

*상관계수값(r)과 상관성 정도

- 1) ±0.9 이상 : 매우 높은 상관관계
- 2) ±0.7~0.9 미만 : 높은 상관관계
- 3) ±0.4~0.7 미만 : 다소 높은 상관관계
- 4) ±0.2~0.4 미만 : 낮은 상관관계
- 5) ±0.2 미만 : 상관관계가 거의 없음

종합병원의 입원환자수와 국민총소득은
0.98만큼의 매우 높은 양의 상관성이 있으며,
이는 $P\text{-value} < 0.05$ 이므로
유의한 값이라고 해석할 수 있습니다.





회귀분석

- ① 기초 이론
- ② E-stat 3.0 활용실습



회귀분석

회귀분석이란?

- ✓ 독립변수와 종속변수가 모두 연속형 변수일 때 사용할 수 있는 통계기법
- ✓ 변수들간의 함수관계를 분석하여, 독립변수가 종속변수를 얼마나 설명하거나 예측하는지 알 수 있음
- ✓ 독립변수가 1개일 때 단순회귀, 2개 이상일 때 다중회귀
- ✓ 영가설: 회귀계수의 모수치는 0이다. (독립변수는 종속변수에 영향을 주지 않는다.)
- ✓ 연구가설: 회귀계수의 모수치는 0이 아니다.(독립변수는 종속변수에 영향을 준다.)

ex) 한 달 소득이 영화 관람 횟수에 영향을 미치는가, 온도가 아이스크림 판매량에 영향을 미치는가



회귀분석



회귀분석 주요 용어

R-square : 독립변수들이 종속변수를 얼마나 잘 설명하는가 나타내는 값
(R-square가 0.3이라면 30%를 예측할 수 있다는 의미를 지님)

즉, R-square가 1에 가까울 수록 설명력이 높고 0에 가까울수록 설명력이 낮다.

표준오차 : 회귀계수, 회귀상수의 정확성을 나타내주는 척도 / 작을수록 정확하게 예측



회귀분석

E-STAT 3.0

통계 분석 > 회귀분석

통계 분석 <

- 기술통계량
- 상관분석
- 회귀분석**
- 작업파일 저장
- 분석 종료

변수 리스트

시기

독립 변수

국민 총 소득(GNI:원)

종속 변수

해외여행객수

분석하기

도움말

좌측메뉴에서 '회귀분석'을 클릭합니다.



회귀분석

E-STAT 3.0

통계 분석 > 회귀분석

변수 리스트

시기

독립 변수

국민 총 소득(GNI:원)

종속 변수

해외여행객수

분석하기

도움말

변수리스트에서 설명하거나 예측하는 변수를 '독립변수'로, 설명이나 예측 되어지는 변수를 '종속변수'로 이동합니다.





회귀분석

이때의 영가설은 '국민 총소득은 해외여행객수에 영향을 주지 않는다' 이고
연구가설은 '국민총소득은 해외여행객수에 영향을 미친다' 입니다.





회귀분석

*Ordinary Least Squares Estimates

SSE	DFE	MSE	Root MSE	R-Square	Durbin-Waston
1,407,113.543	27.000	52,115.316	228.288	0.906	0.303

*분산분석

모형	제곱합	자유도	평균제곱	F-Statistic	P-Value	R-Square
회귀모형	13,623,488.485	1	13,623,488.485	619.640	0.000	0.906
잔차	1,407,113.543	64	21,986.149			
합계	15,030,602.028	65				

R-Square 값이 0.906 이므로 해당 독립변수는 종속변수를 90.6% 설명할 수 있다는 것을 의미합니다.



*회귀계수

변수	추정값(B)	표준화추정값(베타)	표준오차	T	유의확률 (P-value)
절편	-303.921		90.981	-3.340	0.002
국민 총 소득(GNI:원)	0.001	?	0.000	16.168	0.000

[F, T값과 P-value]
 1) P-value < 0.05 : F, T값이 유의함 (통계적으로 의미가 있음)
 2) P-value > 0.05 : F, T값이 유의하지 않음 (통계적으로 의미가 없음)



회귀분석

*Ordinary Least Squares Estimates

SSE	DFE	MSE	Root MSE	R-Square	Durbin-Waston
1,407,113.543	27.000	52,115.316	228.288	0.906	0.303

*분산분석

모형	제곱합	자유도	평균제곱	F	유의확률 (P-value)	R-Square
회귀모형	13,623,488.485	1	13,623,488.485	619.640	0.000	0.906
잔차	1,407,113.543	64	21,986.149			
합계	15,030,602.028	65				

*유의확률이 < 0.05 이므로
분석된 회귀모형은 유의하다고 판단할 수 있습니다.*

*회귀계수

변수	추정값(B)	표준화추정값(베타)	표준오차	T	유의확률 (P-value)
절편	-303.921		90.981	-3.340	0.002
국민 총 소득(GNI:원)	0.001	?	0.000	16.168	0.000

[F, T값과 P-value]
 1) P-value < 0.05 : F, T값이 유의함 (통계적으로 의미가 있음)
 2) P-value > 0.05 : F, T값이 유의하지 않음 (통계적으로 의미가 없음)



회귀분석

*Ordinary Least Squares Estimates

SSE	DFE	MSE	Root MSE	R-Square	Durbin-Waston
1,407,113.543	27.000	52,115.316	228.288	0.906	0.303

*분산분석

모형	제곱합	자유도	평균제곱	F	유의확률 (P-value)	R-Square
회귀모형	13,623,488.485	1	13,623,488.485	619.640	0.000	0.906
잔차	1,407,113.543	64	21,986.149			
합계	15,030,602.028	65				

*회귀계수

변수	추정값(B)	표준화추정값(베타)	표준오차	T	유의확률 (P-value)
절편	-303.921		90.981	-3.340	0.002
국민 총 소득(GNI:원)	0.001	?	0.000	16.168	0.000



[F, T값과 P-value]

- 1) P-value < 0.05 : F, T값이 유의함 (통계적으로)
- 2) P-value > 0.05 : F, T값이 유의하지 않음 (통계적으로)

P-value < 0.05 이므로 연구가설을 채택하여
국민 총 소득은 해외여행객수에 영향을 미친다고 볼 수 있습니다.



결측값 지정

① E-stat 3.0 활용실습



결측값 지정

분석결과

분석방법 : 빈도분석

선택변수 : 3. 요즈음 00님의 건강상태는 어떻습니까?

3. 요즈음 00님의 건강상태는 어떻습니까?

	빈도	백분율
1. 매우 건강하다	172	14.333
2. 건강한 편이다	795	66.25
3. 건강하지 않은 편이다	210	17.5
4. 전혀 건강하지 않다	13	1.083
99. 무응답	10	0.833
TOTAL	1,200	100

전체건수 : 1200, 관측건수 : 1200, MISS 건수 : 0



결측값 지정은
'무응답'과 같은 값을
분석에서 제외할 때
사용하는 기능입니다.



결측값 지정

E-STAT 3.0

데이터 관리 > 결측값 지정

도움말

결측값으로 지정할 변수를 선택해 주십시오.

인구
여가생활
정치
직장
종교
2. 모든 것을 고려할 때, 00님께서 현재 어느 정도 행복하다고 생각하십니까?
3. 요즘 00님의 건강상태는 어떨습니까?

No.	결측값 지정	선택
1.	매우 건강하다	<input type="checkbox"/>
2.	건강한 편이다	<input type="checkbox"/>
3.	건강하지 않은 편이다	<input type="checkbox"/>
4.	전혀 건강하지 않다	<input type="checkbox"/>
99.	무응답	<input checked="" type="checkbox"/>

좌측메뉴에서 '결측값 지정'을 클릭합니다.





결측값 지정

E-STAT 3.0

데이터 관리 > 결측값 지정

도움말

결측값으로 지정할 변수를 선택해 주십시오.

친구
 여가생활
 정치
 직장
 종교
 2. 모든 것을 고려할 때, 00님께서 현재 어느 정도 행복하다고 생각하십니까?
 3. 요즘 00님의 건강상태는 어떨습니까?

No.	결측값 지정	선택
1.	매우 건강하다	<input type="checkbox"/>
2.	건강한 편이다	<input type="checkbox"/>
3.	건강하지 않은 편이다	<input type="checkbox"/>
4.	전혀 건강하지 않다	<input type="checkbox"/>
99.	무응답	<input checked="" type="checkbox"/>

변수를 선택한 뒤, 하단의 값 중 결측값으로 지정할 값에 체크한 후 선택 버튼을 클릭합니다.





결측값 지정

결측값 지정 전

분석결과

분석방법 : 빈도분석

선택변수 : 3. 요즈음 00님의 건강상태는 어떻습니까?

3. 요즈음 00님의 건강상태는 어떻습니까?

	빈도	백분율
1. 매우 건강하다	172	14.333
2. 건강한 편이다	795	66.25
3. 건강하지 않은 편이다	210	17.5
4. 전혀 건강하지 않다	13	1.083
99. 무응답	10	0.833
TOTAL	1,200	100

전체건수 : 1200, 관측건수 : 1200, MISS 건수 : 0

결측값 지정 후

분석결과

분석방법 : 빈도분석

선택변수 : 3. 요즈음 00님의 건강상태는 어떻습니까?

3. 요즈음 00님의 건강상태는 어떻습니까?

	빈도	백분율
1. 매우 건강하다	172	14.454
2. 건강한 편이다	795	66.807
3. 건강하지 않은 편이다	210	17.647
4. 전혀 건강하지 않다	13	1.092
TOTAL	1,190	100

전체건수 : 1200, 관측건수 : 1190, MISS 건수 : 10



해당문항을 다시 빈도분석 해보면 선택한 값이 사라지고
TOTAL값이 이 변경된 것을 확인할 수 있습니다.



데이터 변환

① E-stat 3.0 활용실습



데이터 변환

분석결과

분석방법 : 빈도분석

선택변수 : 3. 요즈음 00님의 건강상태는 어떻습니까?

3. 요즈음 00님의 건강상태는 어떻습니까?

	빈도	백분율
1. 매우 건강하다	172	14.454
2. 건강한 편이다	795	66.807
3. 건강하지 않은 편이다	210	17.647
4. 전혀 건강하지 않다	13	1.092
TOTAL	1,190	100

전체건수 : 1200, 관측건수 : 1190, MISS 건수 : 10

데이터 변환은 특정 조건을 부여하여 새롭게 변수를 만드는 기능입니다.



➤ '건강하다' 로 병합

➤ '건강하지 않다' 로 병합



데이터 변환

E-STAT 3.0

데이터 관리 > 데이터 변환

도움말

탐색 분석

통계 분석

데이터 관리

데이터 변환

결측값 지정

작업파일 저장

분석 종료

변수 리스트

종교

2. 모든 것을 고려할 때, 00님께서는 현재 어느 정도 행복하다고 생각하십니까?
3. 요즈음 00님의 건강상태는 어떻습니까?
5. 모든 것을 고려할 때, 00님께서는 요즈음 삶에 대하여 어느 정도 만족하십니까?
6. 일반적으로 말해서 사람들을 믿을 수 있다고 생각하십니까? 아니면 인간관계에서 조심해야 한다고 보십니까?

종교나 교회단체
스포츠 레크레이션 단체

선택 변수

신규 변수명

데이터변환_건강상태

좌측메뉴에서 '데이터 변환'을 클릭합니다.





데이터 변환

E-STAT 3.0

데이터 관리 > 데이터 변환

도움말

탐색 분석
통계 분석
데이터 관리

데이터 변환
결측값 지정
작업파일 저장
분석 종료

변수 리스트

종교
2. 모든 것을 고려할 때, 00님께서 현재 어느 정도 행복하다고 생각하십니까?
3. 요즈음 00님의 건강상태는 어떻습니까?
5. 모든 것을 고려할 때, 00님께서 요즈음 삶에 대하여 어느 정도 만족하십니까?
6. 일반적으로 말해서 사람들을 믿을 수 있다고 생각하십니까? 아니면 인간관계에서 조심해야 한다고 보십니까?
종교나 교회단체
스포츠 레크리에이션 단체

선택 변수

신규 변수명

데이터변환_건강상태

변수리스트에서 문항을 선택하고
선택 변수로 옮긴 뒤
새로운 변수의 이름을 지정해주시면 됩니다.



데이터 변환

E-STAT 3.0

데이터 관리 > 데이터 변환

신규 변수명 : 데이터변환_건강상태

변수(3. 요즈음 00님의 건강상태는 어떻습니까?)의 기존값

Value :

Range : 과 사이

Range : 이하인 것

Range : 이상인 것

신규 변수의 설정

Value : Label :


변수 정보

Q8. 3. 요즈음 00님의 건강상태는 어떻습니까?(N=...)

- 1) 매우 건강하다
- 2) 건강한 편이다
- 3) 건강하지 않은 편이다
- 4) 전혀 건강하지 않다
- 99) 무응답

1. 기존값에서 병합할 값(Value)의 범위를 지정합니다

2. 신규 변수의 설정에서 새로운 값(Value)과 내용(Label)을 입력합니다.





데이터 변환

E-STAT 3.0

데이터 관리 > 데이터 변환

신규 변수명 : 데이터변환_건강상태

변수(3. 요즈음 00님의 건강상태는 어떻습니까?)의 기준값

Value :

Range : 과 사이

Range : 이하인 것

Range : 이상인 것

신규 변수의 설정

Value : Label : + 추가 × 취소

새 값	새 값 라벨	조건	삭제
1	건강하다	Q8 의 값이 1 부터 2 까지 인것	삭제

Process

변수 정보

Q8. 3. 요즈음 00님의 건강상태는 어떻습니까?
1) 매우 건강하다
2) 건강한 편이다

추가 버튼을 누르면 하단에 새로운 변수(데이터변환_건강상태)의 보기1번이 생성된 것을 확인할 수 있습니다.





데이터 변환

E-STAT 3.0

데이터 관리 > 데이터 변환

신규 변수명 : 데이터변환_건강상태

변수(3. 요즈음 00님의 건강상태는 어떻습니까?)의 기준값

Value :

Range : 과 사이

Range : 이

Range : 0

신규 변수의 설정

Value : Label :

새 값	새 값 라벨	조건	삭제
1	건강하다	Q8 의 값이 1 부터 2 까지 인것	<input type="button" value="삭제"/>
2	건강하지 않다	Q8 의 값이 3 부터 4 까지 인것	<input type="button" value="삭제"/>

같은 방식으로 보기2 번 '건강하지 않다'를 입력 후 하단의 process 버튼을 누르면 완료됩니다.





데이터 변환

E-STAT 3.0

통계 분석 > 빈도분석

도움말

변수 리스트

- Q1. 응답자의 반응
- Q2. 응답자의 사생활 보호
- Q3. 응답자의 문맹 여부
- Q4. 조사지역 특성
- Q5. 조사지역1
- 데이터변환_건강상태

선택 변수

새로 생성된 변수는 변수리스트 가장 하단에 위치하고 있습니다.



데이터 변환

분석결과

분석방법 : 빈도분석

선택변수 : 데이터변환_건강상태

데이터변환_건강상태

	빈도	백분율
1. 건강하다	967	81.261
2. 건강하지 않다	223	18.739
TOTAL	1,190	100

전체건수 : 1200, 관측건수 : 1190, MISS 건수 : 10



새 변수(데이터변환_건강상태)를
빈도 분석 해보면
기존 문항의 보기들이 병합된 것을
확인할 수 있습니다.