

IBM SPSS Statistics Module

IBM SPSS Statistics 19

다양한 모듈로 새로운 IBM SPSS Statistics를 경험하십시오!!

New IBM SPSS Statistics

The screenshot displays the IBM SPSS Statistics interface with several key components:

- Hypothesis Test Summary Table:**

Null Hypothesis	Test	Sig.	Decision
The distribution of Cooking ADL is the same across categories of Treatment group.	Independent-Samples Mann-Whitney U Test	.030	Reject the null hypothesis
The distribution of Housekeeping ADL is the same across categories of Treatment group.	Independent-Samples Mann-Whitney U Test	.004	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05.
- Independent-Samples Mann-Whitney U Test Results:**
 - Treatment group: N = 48, Mean Rank = 57.07
 - Treatment group: N = 54, Mean Rank = 44.91
 - U = 940.000
 - Wilcoxon W = 2,625.000
 - Test Statistic = 940.000
 - Standard Error = 139.508
 - Standardized Test Statistic = -2.165
- Marketing Campaign Modules:**
 - Improve My Marketing Campaigns:**
 - Identify the top responding postal codes (Map showing postal codes: 40024, 45, 59710, 75224, 59222).
 - Select contacts most likely to purchase (Grid of contact icons with checkmarks).
 - Compare effectiveness of campaigns (Control Package Test) (Grid of campaign icons).

> IBM SPSS Statistics Data Analysis Path



1. Planning

- SPSS Conjoint
- SPSS Complex Samples

2. Data Collection

- SPSS Data Entry

3. Data Access

- SPSS Basic

4. Data management & Preparation

- SPSS Basic
- SPSS Missing Values

7. Deployment

- SPSS PES
(Predictive Enterprise Services)

6. Reporting

- SPSS Basic

5. Data Analysis

- SPSS Basic
- SPSS Complex Samples
- SPSS Regression
- SPSS Advanced Statistics
- SPSS Categories
- SPSS Decision Trees
- SPSS Exact Tests
- SPSS Forecasting
- SPSS Korea Options pack
- SPSS Bio equiv
- SPSS Bootstrapping
- SPSS Direct Marketing



IBM SPSS Statistics의 가장 기본적인 패키지로써 분석의 전 과정을 쉽게 수행할 수 있습니다.

IBM SPSS Statistics Basic은 기업의 방대한 데이터에 쉽게 접근하여 모든 기본 분석을 수행할 수 있으며 통계수치로 표현된 데이터를 활용하기 쉬운 다양한 형태로 출력하여 결과를 표현할 수 있는 기본적인 패키지입니다.

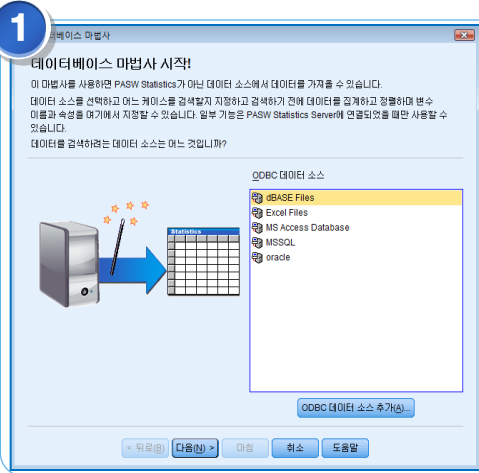
Q 특징

- 간편한 데이터 접근 및 핸들링을 통해 시간 절약
- 편리한 데이터 변환 및 정제
- 다양한 통계 분석 기능
- 누구나 쉽게 이해할 수 있는 Reporting 결과 제공
- 다양한 프로그램으로 결과 내보내기 기능
- 코드북을 통한 변수 정보 확인
- 사용자정의 대화상자 추가 기능 및 스펠링 체크 기능

Q 주요 분석

- 빈도분석/ 기술통계분석/ 데이터탐색/ 교차분석
- 집단별 평균 분석/ T-검정(T-test)
- 보고서
- 일원배치 분산분석
- 이변량 상관분석/ 편상관분석/ 거리측도
- 선형회귀분석/ 순서형회귀분석
- 곡선추정
- 비모수검정
- 다중응답분석
- 요인분석/ 판별분석
- 군집분석/ K-평균 군집분석/ 가장 가까운 이웃분석
- 척도화분석
- 신뢰도분석
- ALSCAL 다차원척도법
- 행렬연산기능

> IBM SPSS Statistics Basic

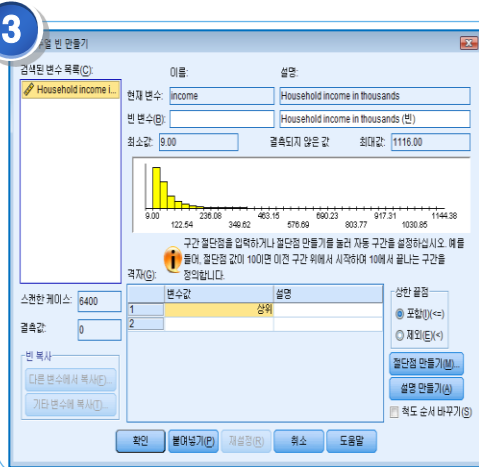
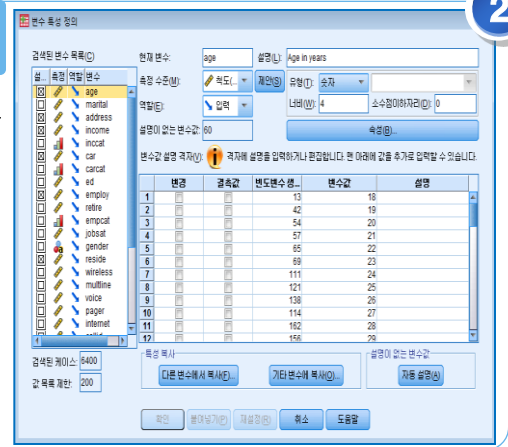


용이한 데이터 접근으로 시간을 절약할 수 있습니다.

- IBM SPSS Statistics Database Wizard를 통해 대용량 데이터를 빠르고 쉽게 불러올 수 있습니다. 또한, IBM SPSS Statistics에서는 Excel, SAS, 텍스트 파일 등 다양한 파일 형식에 대한 불러오기 기능을 제공합니다.

분석 계획 단계를 간소화할 수 있습니다.

- 변수 특성 정의 기능(Define Variable Properties)을 이용하여 모든 변수에 개별적으로 유형, 측도, 결측치, 설명 등 정의를 부여하는 일을 간편화할 수 있습니다.



비주얼 빈 만들기 기능을 통해 연속형 데이터를 쉽게 범주화할 수 있습니다.

- 연령, 소득과 같은 데이터를 특정 범주로 묶어야 할 때, 비주얼 빈 만들기(Visual Binning) 기능을 이용하여 빠르고 쉽게 범주화할 수 있습니다. 또한, 범주를 나누는 기준점을 히스토그램을 통해 쉽게 파악할 수 있습니다.

> IBM SPSS Statistics Basic

4

OLAP 큐브						
가구소득별분주합계						
	합계	N	평균	표준편차	전체 합계 퍼센트	전체 수 퍼센트
가구소득 (단위:1000)	444839.00	8400	66.4748	78.71856	100.0%	100.0%
현직장 근무년수	67624	8400	10.57	9.724	100.0%	100.0%
현주소지 거주년수	73983	8400	11.56	9.938	100.0%	100.0%
보유자동차 가격	182821.60	8400	30.1284	21.92692	100.0%	100.0%

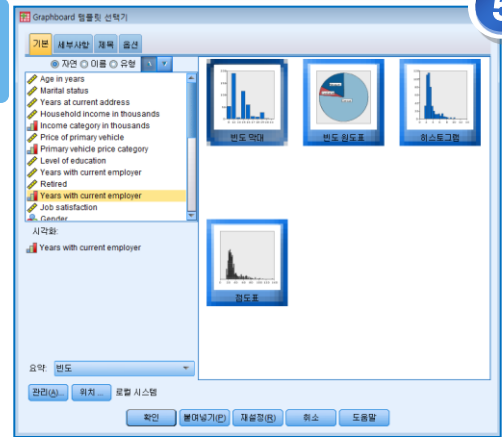
OLAP 큐브						
가구소득별분주 \$50000 - \$74000						
	합계	N	평균	표준편차	전체 합계 퍼센트	전체 수 퍼센트
가구소득 (단위:1000)	67929.00	1120	60.6509	7.08474	15.3%	17.5%
현직장 근무년수	12395	1120	11.07	7.276	18.3%	17.5%
현주소지 거주년수	13609	1120	12.15	9.117	18.4%	17.5%
보유자동차 가격	33883.30	1120	30.2619	3.60757	17.8%	17.5%

리포트 OLAP 큐브를 통해 최적의 결과를 표현할 수 있습니다.

- 범주형 그룹변수의 범주별로 연속형 요약변수들의 합계, 평균, 중위수 등의 기술통계량을 산출할 수 있습니다. 또한, IBM SPSS Statistics 피벗표 개체를 수정하여 특정 변수의 범주값들을 변경하면서 결과를 살펴볼 수 있습니다.

다양한 그래프를 통해 분석 결과를 시각적으로 표현할 수 있습니다.

- IBM SPSS Statistics Basic은 막대도표, 원도표, 상자도표 등 기본적인 그래프 뿐만 아니라 산점도, ROC도표, 품질관리(QC)도표 등 다양한 그래프를 지원하여 분석 결과를 보다 시각적으로 표현할 수 있습니다. 또한, Graphboard를 통해 변수간의 관계를 더욱 빠르고 쉽게 파악하는 강력한 시각화 기능을 제공합니다.



5

6

		연령	연도	
			2010	2011
여성	직업	고교졸	91	579
		대졸	91	866
	교육수준	고졸	64	623
		대졸	58	618
남자	직업	고교졸	11	170
		대졸	109	607
	교육수준	고졸	119	386
		대졸	76	395
연령	20-29	46	633	
	30-39	8	169	

분석 결과를 다양한 파일 형식으로 출력할 수 있습니다.

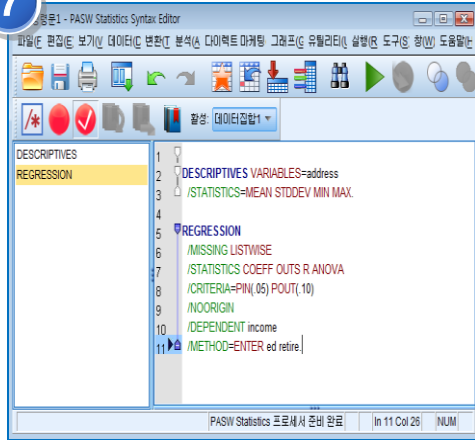
- MS Word/ Excel/ Power Point 뿐만 아니라 PDF, HTML 등 다양한 파일 형식으로 결과를 바로 출력할 수 있어 모든 사람들이 쉽게 결과물을 살펴볼 수 있습니다.

> IBM SPSS Statistics Basic

7

Syntax 편집기의 사용이 편리합니다.

- 명령어를 사용하여 작업을 저장한 후 분석을 반복하거나 Production Facility를 통해 자동화된 작업을 실행할 수 있습니다. 자동완성기능이 있으며 실행하기 전에 오류를 발견할 수 있습니다.

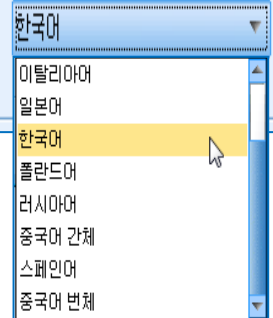


다국어어를 지원합니다.

- 사용자 인터페이스 언어를 필요에 따라 변경할 수 있습니다. 한글버전을 제공하여 S/W 접근성이 용이합니다.

사용자 인터페이스

언어(G):



8

9

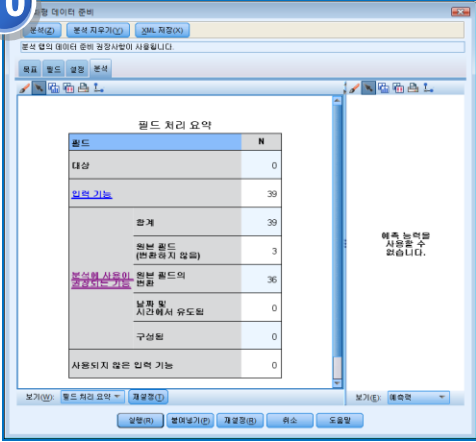
모형 생성을 위한 역할(Role) 항목을 지정할 수 있습니다.

속도	역할
척도(S)	입력
척도(S)	입력
척도(S)	대상
척도(S)	모두
순서(O)	없음
척도(S)	분할
순서(O)	분할

- 역할(Role) 항목을 추가하여 모델링 기능을 직접 지정할 수 있습니다.
- 역할(Role) 항목에는 입력(Input), 대상(Target), 모두(Both), 없음(None), 분할(Partition), 분할(Split)로 구성됩니다.

> IBM SPSS Statistics Basic

10



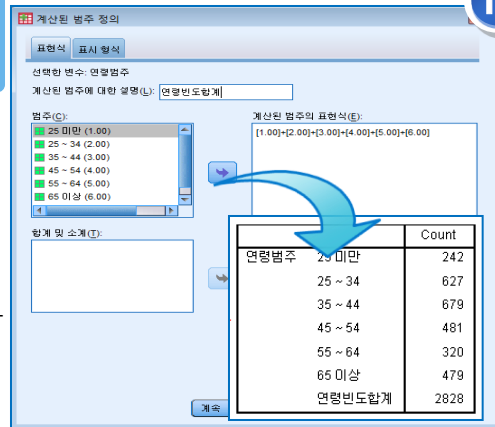
자동 데이터 준비 (Automated Data Preparation)을 수행합니다.

- 자동으로 데이터를 분석하고, 유용할 것 같지 않은 필드를 가려내어, 통계에 대한 사전 지식을 많이 보유하고 있지 않아도 모델링을 쉽게 할 수 있도록 준비합니다.
- 크게 대화형 데이터 준비, 자동 데이터 준비, 역변환 점수로 분류되어 수행됩니다.

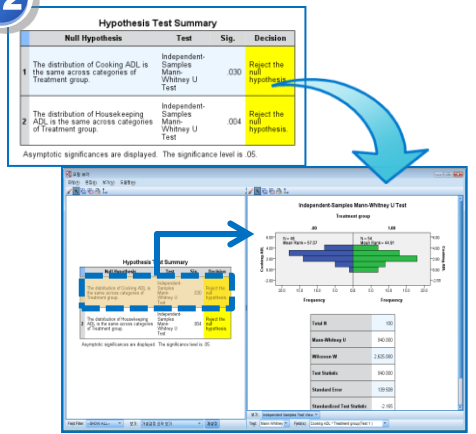
사용자 정의표(Custom Table)의 기능이 개선됩니다.

- 통계표 작성시 새로운 카테고리(범주)를 추가 할 수 있습니다. 새로운 카테고리(범주)를 생성하고 연산에 사용된 카테고리(범주) 감추기 옵션을 이용하여 감추기가 가능합니다. 이를 이용하여 재범주화 되고 단순화한 표를 작성 할 수 있습니다.

11



12



Model Viewer의 기능이 강화됩니다.

- 결과를 좀 더 쉽게 이해 할 수 있도록 Viewer의 기능이 강화됩니다.
- 자동 데이터 준비, 이단계 군집 분석, 비모수 검정에서 사용 가능합니다.

강력한 회귀분석 절차를 통해 보다 더 나은 예측을 할 수 있습니다.

SPSS Regression은 분석가가 특정 데이터 유형에 대해 가장 정확한 반응을 얻을 수 있도록 보다 폭넓은 통계량을 도출하도록 합니다. 또한, 보다 정교한 모형들을 처리하기 위해 필요한 분석기법들을 쉽게 다룰 수 있는 기능을 제공하며, 변수들간의 선형적인 관계 뿐만 아니라 복잡한 형태의 관계 또한 파악할 수 있는 모듈입니다.

 **특 징**

- 데이터의 제약을 받지 않음
- 데이터를 그룹으로 쉽게 분류
- 최적의 설명변수를 찾을 수 있음

 **주요 분석**

- 다항 로지스틱 회귀분석 (Multinomial Logistic Regression)
- 이분형 로지스틱 회귀 분석 (Binary Logistic Regression)
- 비선형 회귀 분석 (Nonlinear Regression)
- 제약된 비선형 회귀분석 (Constrained Nonlinear Regression)
- 가중 최소제곱법 (Weighted Least Square)
- 2단계 최소제곱법 (Two-stage Least Square)
- 프로빗 모형분석 (Probit Analysis)

 **활용 분야**

- Market Research: 고객 구매 패턴 분석
- Medical Research: 약물 반응 조사
- 신용도 분석
- 교육 기관 성과 측정 조사



> SPSS Regression



예측 모델 생성

- IBM SPSS Statistics Regression을 통해 일련의 예측 변수로부터 기인한 특징 및 결과를 예측할 수 있습니다. 예를 들어, 은행의 대출 담당자는 대출 전에 대출금을 제대로 갚지 않을 가능성이 있는 사람을 식별할 수 있는 특성을 알아내어 고객의 신용 위험도를 평가할 지표로 사용할 수 있습니다.

범주형 결과 예측

- 우측의 분류표는 모형에 의해 선택된 케이스의 분류 정확도가 90%임을 나타냅니다.

분류표^a

		예측		
		반응		분류정확 %
감시됨	예	아니오		
1 단계 반응	예	196	8	96.1
	아니오	18	26	59.1
전체 퍼센트				89.5

a. 절단값은 .500입니다.

2





복잡한 데이터 간의 관계를 정교하게 분석할 수 있습니다.

SPSS Advanced Statistics는 심층적인 분석을 통해 데이터에 내포되어 있는 의미를 찾고자 할 때 사용되는 모듈로써, 사용자가 기본적인 분석만으로는 불가능한 심층적인 분석과 관련된 모형 수립, 생존데이터 분석, 복잡한 관계가 있는 여러 개의 변수 분석 시 보다 정교하게 분석할 수 있습니다.

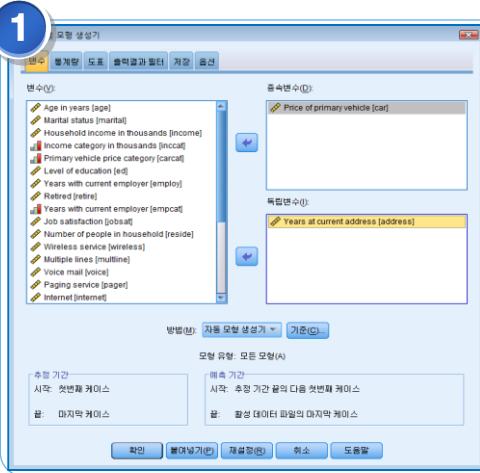
특징

- 계층 구조의 데이터 분석
- 다양하고 심층적인 결과 도출
- 매크로 기능 및 행렬 계산 기능
- 생존 분석 수행

주요 분석

- 일반선형모형(General Liner Models) : 종속변수와 여러 개의 독립변수 간의 관계를 보다 유연하게 설명할 수 있습니다.
- 고정효과 ANOVA, 임의/혼합효과 공분산분석, 공분산분석(ANCOVA), Repeated measures ANOVA and MANOVA, 다변량 공분산분석(MANCOVA), 다변량 분산분석(MANOVA), 임의/혼합효과 분산분석, 분산성분추정(VARCOMP)
- Linear Mixed Models : 관측자료들이 상관구조를 갖거나 각기 다른 분산을 가질 수 있도록 GLM을 확장한 것으로 보다 더 정확한 모델을 생성할 수 있습니다.
- 생존분석(Survival Analysis) : Kaplan-Meier, Cox Regression 등 최신 생존 분석 기법을 적용할 수 있습니다.
- 생명표 분석
- 일반화 추정 방정식(GEE)
- 로그 선형모형 분석(LOGLINEAR)
- 계보적 로그선형모형(HILOGLEANEAR)
- 일반화 로그선형모형(Generalized Loglinear)
- 다항 로짓모형 분석
- 일반화 선형모형(GENLIN)

> SPSS Advanced Statistics

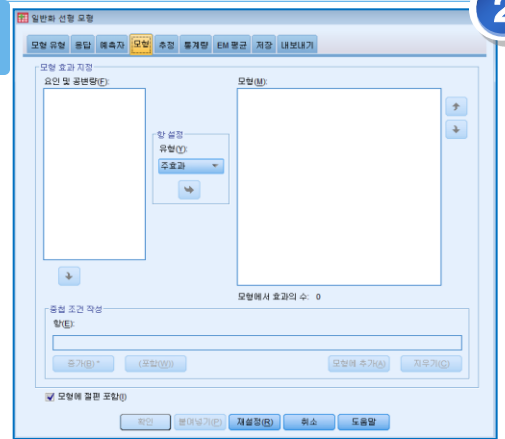


보다 정확한 모델 생성

- 나이에 따라 신체 특성 변수가 어떻게 성장하는지 그리고 그것이 성별에 따라 차이가 있는가를 테스트할 수 있습니다.

임의 효과(Random Effects) 확인

- 고정 효과와 임의 효과를 각각 지정합니다. 임의 효과 공분산 구조 유형 중 1차 자기회귀, 복합대칭, 분산성분, 척도법 동일 등에서 하나를 지정합니다.



3

다양한 분석 기법을 이용하여 결과를 도출

제3 유형 고정 효과 검정

소스	분자 df	분모 df	거짓
절편	1	25.723	75.0
gender	1	8.701	3.7
age	1	23.687	22.7

a. 종속변수: Distance (mm) from center to pteryo-maxillary fissure.

모형 차원^a

고정 효과	수준의 수	공분산 구조	모수의 수	개체 변수	개체의 수
절편	1		1		
gender	2		1		
age	1		1		
반복 효과	3	1차 자기회귀	2	subject	10
합계	7		5		

a. 종속변수: Distance (mm) from center of pituitary to pteryo-maxillary fissure.

> SPSS Forecasting

전문적인 시계열 분석을 쉽고 빠르게 수행할 수 있습니다.

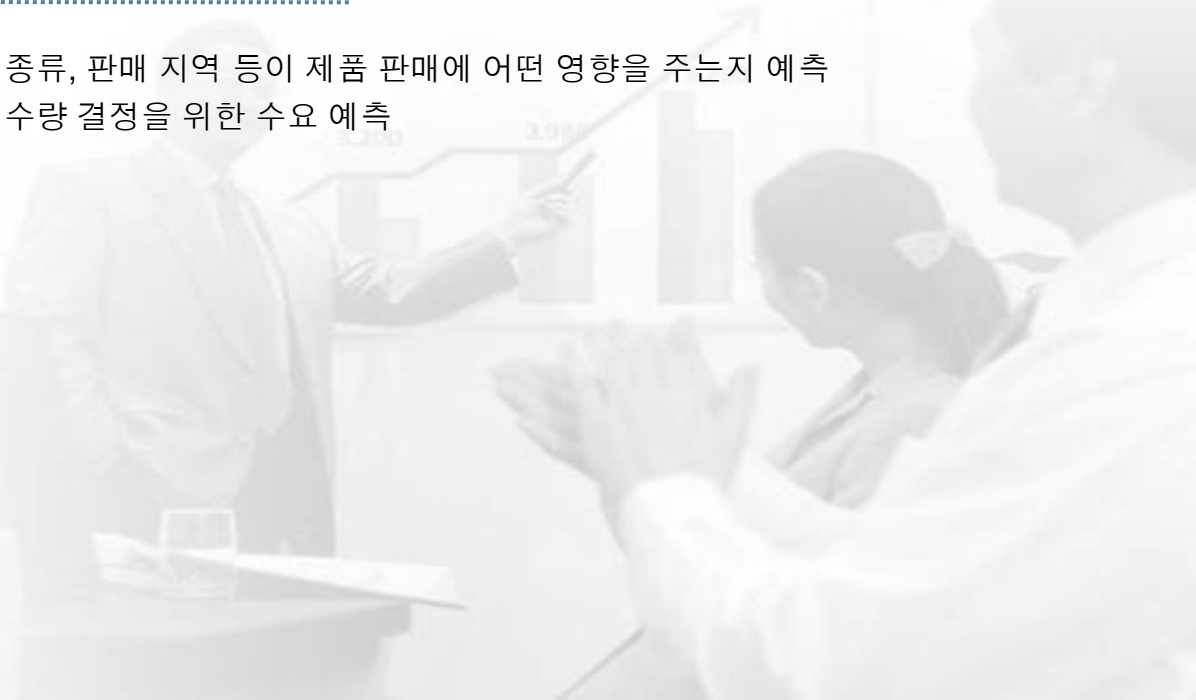
시계열 분석은 시간에 따라 축적된 정보를 분석하고, 모델을 생성하여 미래의 결과를 예측하는 가장 강력한 분석입니다. SPSS Forecasting은 일반 스프레드시트 형식과 달리 시계열 데이터를 가지고 작업하기 위해 필요한 고급 통계 분석 기법들을 제공합니다.

특징

- Expert Modeler를 사용하여 최적의 모델 도출
 - ✓ ARIMA 및 지수평활모형, 계절모형 모두 고려
 - ✓ 시계열에 영향을 주는 사건을 정의하는 독립변수 지정
 - ✓ 관측된 데이터에 가장 적절한 모형을 선택하여 제공
- 시계열 분석 지원
- 데이터 사이즈 또는 변수 개수에 상관없이 신뢰성 있는 모델을 빠르게 생성
- 모수 추정 시 영향을 주는 이상치 감지
- 신뢰구간 및 모델 적합도 그래프 생성하여 어떤 모델이 시계열 및 독립 변수에 가장 적합한 것인지 자동적으로 결정할 수 있도록 지원

활용 분야

- 제품 종류, 판매 지역 등이 제품 판매에 어떤 영향을 주는지 예측
- 생산 수량 결정을 위한 수요 예측



> SPSS Forecasting

Q 주요 분석

- ARIMA
 - ✓ 정확한 최대우도법
 - ✓ 자기회귀 이동평균
 - ✓ 신뢰구간 수준 설정
 - ✓ 시계열 내에 내재된 결측치 감지
 - ✓ Intervention 모형 분석 등
- 지수평활법
 - ✓ 단순, 승법모형, 가법모형, Holt 모형
 - ✓ 모형에 사용된 주기, 초기계절요인, 초기 일반적 추세값 설정
- 계절분해
 - ✓ 계절을 계절성분, 결합된 추세 및 순환 성분, 오차 성분으로 분해
 - ✓ 승법모형, 가법모형
- 스펙트럼
 - ✓ 시계열의 주기 동작을 식별
 - ✓ 평활에 이용가능한 스펙트럼 밀도 추정값을 구하기 위해 주기를 평활하는 방법은 Tukey-Hamming, Tukey, Parzen, Bartlett, Daniell(Unit) 중 선택
 - ✓ 주기도, 스펙트럼 밀도, 공스펙트럼 밀도, 구적 스펙트럼, 교차진폭, Gain, 제곱일관도, 위상스펙트럼 도표 제공
- 자기회귀
 - ✓ Exact Maximum-likelihood 추정법, Cochrane-Orcutt 추정법, Prais-Winsten 추정법 등
 - ✓ 임시 변수를 작업 데이터 파일에 추가
 - ✓ 가능한 통계량 : CREATE, RMV, CASEPLOT, NPPOLT, ACF, PACF, CCF, CURVEFIT

> SPSS Categories

범주형 데이터 간의 관계를 명확히 할 수 있습니다.

SPSS Categories는 사용자가 분석하고자 하는 범주형 데이터 간의 관계를 쉽게 분석할 수 있으며, 데이터 간의 관계를 더 상세하게 확인할 수 있는 모듈입니다.

특징

- 오브젝트 가중치, 데이터의 분할, 보조적인 오브젝트의 추가, 상관행렬 가능
- 다차원 범주형 데이터에 대한 시각화 및 탐색 기능
- 이차원 또는 다차원 테이블을 통해 쉽게 정보를 이해할 수 있음
- 고해상도 그래프를 통해 관계를 보다 명확히 표현할 수 있음

주요 분석

- 대응일치분석(CORRESPONDENCE)
- 다중 대응일치분석(MULTIPLE CORRESPONDENCE)
- 범주형 회귀분석(CATREG)
- 범주형 주성분분석(CATPCA)
- 비선형 정준상관분석(OVERALS)
- 다차원척도법(PROXSCAL)
- 다차원 확장(PREFSCAL)

활용 분야

- 시장 세분화 분석
- 정당의 선호도
- 생물학적인 분류
- 지역별, 판매 담당자 별 등에 의한 거래 예측
- 고객 만족도 분석

> SPSS Categories

1

행과 열 범주간의 유사성, 연관관계 등을 분석할 수 있습니다.

Staff Group	Smoking				액티브 주변
	None	Light	Medium	Heavy	
Sr Managers	4	2	3	2	11
Jr Managers	4	3	7	4	18
Sr Employees	25	10	12	4	51
Jr Employees	18	24	33	13	88
Secretaries	10	6	7	2	25
액티브 주변	61	45	62	25	193

- 대응일치분석을 통해 범주간의 차이점을 쉽게 분석할 수 있습니다. 예를 들어, 사용자는 다음과 같은 표를 통해 특정 장소에서의 직급(Staff Group)과 흡연량 간의 관계를 파악할 수 있습니다.

새로운 범주를 쉽게 추가하여 관계를 파악할 수 있습니다.

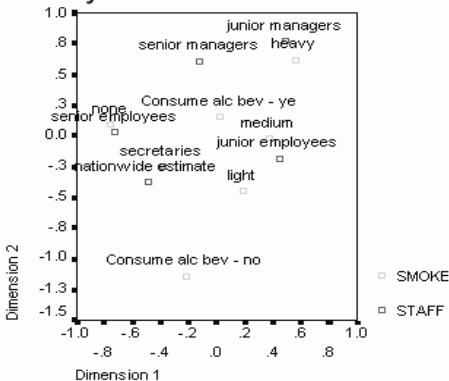
- SPSS Categories에서는 새로운 범주들을 추가시켜 분석할 수 있습니다. 대응분석에서 추가적인 범주는 기존 공간상에 재표현하여 범주간의 관계를 살펴보는 것으로 다음의 대응일치표는 Staff Group별 알코올 소비에 대한 정보를 나타내고 있습니다.

		Alcohol		Total
		Consume alc bev - no	Consume alc bev - yes	
STAFF	senior managers		11	11
	junior managers	1	17	18
	senior employees	5	46	51
	junior employees	10	78	88
	secretaries	7	18	25
Total		23	170	193

2

3

Symmetrical normalization



연관성과 관계를 파악할 수 있습니다.

- 대칭적 정규화를 사용하여 서로 다른 변수들에 대한 범주들간의 거리를 의미 있게 해석할 수 있습니다.
- 다음의 Bi-Plot을 통해 비음주자와 비흡연자 사이의 연관성은 크지 않음을 발견할 수 있으며, 상대적으로 흡연량이 많은 집단이 더 많은 음주량을 가지고 있음을 추정할 수 있습니다.

고객의 구매 결정의 요인을 파악할 수 있습니다.

SPSS Conjoint는 제품 및 서비스가 여러 특성 요인의 조합으로 구성되는 경우, 이러한 특성 가운데 고객에게 중요하게 인지되는 특성 요인이 무엇인지, 또한 상대적으로 덜 중요하게 인지되는 특성 요인은 어떤 것인지를 알아내기 위한 모듈입니다.

Q 주요 분석

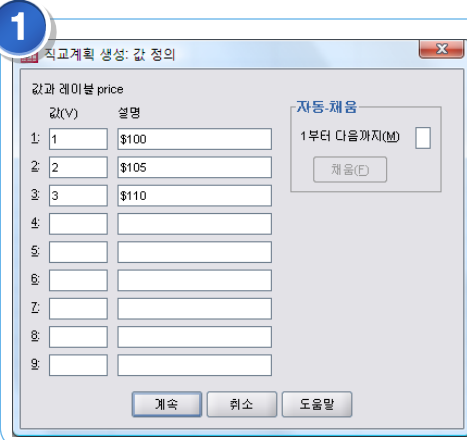
- Orthoplan(직교계획)
- 컨조인트 분석을 위한 실험카드 작성(PLANCARDS)
- 컨조인트 분석
- 컨조인트 플러스 분석

Q 활용 분야

- 기존 제품과 구상 중인 신제품에 대한 시장 점유율 측정
- 고객이 제품을 구매할 때 가장 중요하게 생각하는 특성 요인 도출
- 어떤 요인들의 조합 및 배열이 구매 결정에 영향을 미치는지 측정
- 효과적인 제품 설계/개발을 위한 마케팅 리서치 기법으로 활용



> SPSS Conjoint



직교계획(Orthoplan)을 통해 시간과 비용을 절약할 수 있습니다.

- 응답자에게 질문해야 할 직교배열을 설계할 때, SPSS Conjoint를 통해 각기 다른 제품 특성으로 조합된 최적의 직교배열을 설계함으로써 시간과 비용을 절약할 수 있습니다.

프로파일 카드 생성

- 응답자가 모든 특성 조합에 대해 각각의 순위를 쉽게 매길 수 있는 프로파일 카드를 쉽고 빠르게 생성할 수 있습니다.

프로파일 번호 1				
카드 ID	가격	좌석 사이즈	경유 횟수	음료/스낵 서비스
1	\$105	B	.00	Yes

프로파일 번호 2				
카드 ID	가격	좌석 사이즈	경유 횟수	음료/스낵 서비스
2	\$110	B	1.00	No

프로파일 번호 3				
카드 ID	가격	좌석 사이즈	경유 횟수	음료/스낵 서비스
3	\$105	C	.00	No

프로파일 번호 4				
카드 ID	가격	좌석 사이즈	경유 횟수	음료/스낵 서비스
4	\$100	B	.00	No

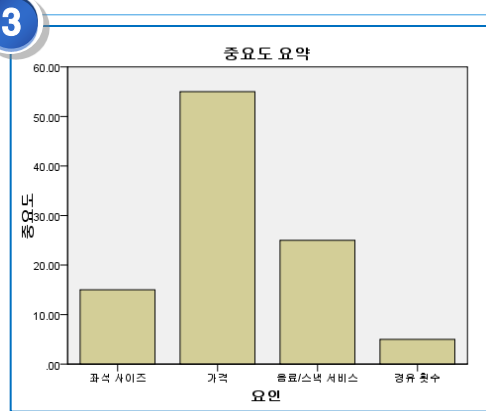


Chart 생성

- 프로파일 카드에 대한 데이터를 모아서 입력한 후, SPSS Conjoint를 이용해 선호도 및 순위에 대한 분석을 할 수 있습니다. 또한, 분석 결과를 그래프로 생성하여 기대되는 시장 점유율을 추정할 수 있습니다.

컨조인트 플러스 분석(Conjoint Plus Analysis)

- 각각의 응답자들이 중요하게 생각하는 요인을 파악할 수 있습니다. 각 평가대상의 효용값(Utility)을 출력하고, 어떤 속성의 어떤 수준이 결합하여 가장 선호하는 평가 대상에 이르는지 알 수 있습니다.

Overall Statistics			
	Utility	Imp. Val.	Correlation
package	A = .000		
	B = -.417		.657
brand	Ortho = -.600		.657
	Ortho = -.753		.657
price	\$1.15 = .100		.657
	\$1.50 = .100		.657
seat	No = .000		1.000
	Yes = 1.000		1.000
money	No = 1.000		1.000
	Yes = 3.000		1.000
(Constant)	Yes = 6.253		2.363

Imp. Val.	
package	Utility
A	.000
B	-.417
brand	Ortho = -.600
	Ortho = -.753
price	\$1.15 = .100
	\$1.50 = .100
seat	No = .000
	Yes = 1.000
money	No = 1.000
	Yes = 3.000

Coefficients	
package	Utility
A	.000
B	-.417
brand	Ortho = -.600
	Ortho = -.753
price	\$1.15 = .100
	\$1.50 = .100
seat	No = .000
	Yes = 1.000
money	No = 1.000
	Yes = 3.000

상관계수			
package	Utility	Imp. Val.	Correlation
package	A = .000		
	B = -.417		.657
brand	Ortho = -.600		.657
	Ortho = -.753		.657
price	\$1.15 = .100		.657
	\$1.50 = .100		.657
seat	No = .000		1.000
	Yes = 1.000		1.000
money	No = 1.000		1.000
	Yes = 3.000		1.000

> SPSS Exact Tests

작은 표본(Sample)만으로도 충분히 정확한 결과를 생성할 수 있습니다.

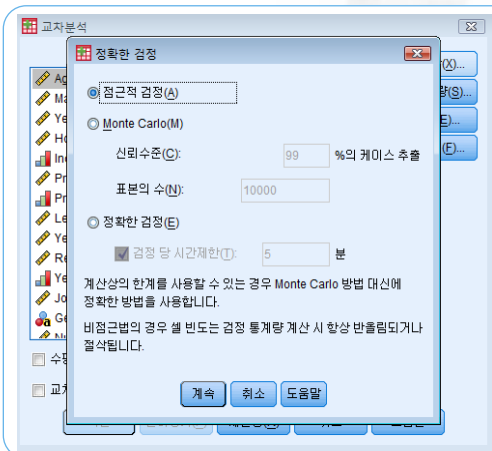
SPSS Exact Tests는 극소규모 또는 극대규모의 범주형 및 비모수 데이터로 분석을 할 경우 발생하는 문제를 해결하여 정확한 통계분석 결과를 도출할 수 있는 모듈입니다.

특징

- 적은 케이스 수로 분석 실행 가능
- 한 범주에서 높은 응답율을 보이는 변수로 분석 실행 가능
- 전통적인 30가지 이상의 Exact Tests 기법을 제공

주요 분석

- 독립/연관된 일표본, 이표본, K-표본 검정
- 적합도 검정(Goodness-of-fits-test)
- R*C contingency table의 독립성 검정
- Measures of association 검정



Exact Test를 위한 다양한 검정 옵션

- 이 출력결과는, 오직 10개의 케이스만을 가지고 있음에도 불구하고, SPSS Exact Test를 통해 유의성이 존재하는지 결정할 수 있습니다. 이처럼, 소규모의 데이터만으로도 신뢰성 있는 결과를 얻을 수 있습니다.

> SPSS Missing Values

결측값을 추정함으로써 보다 가치 있는 데이터와 모델을 생성할 수 있습니다.

SPSS Missing Values는 분석 결과에 큰 영향을 끼칠 수 있는 결측 데이터(Missing Value)의 패턴을 찾고 더 나은 모델을 생성할 수 있도록 보다 가치있는 값을 대체할 수 있는 모듈입니다.

Q 특징

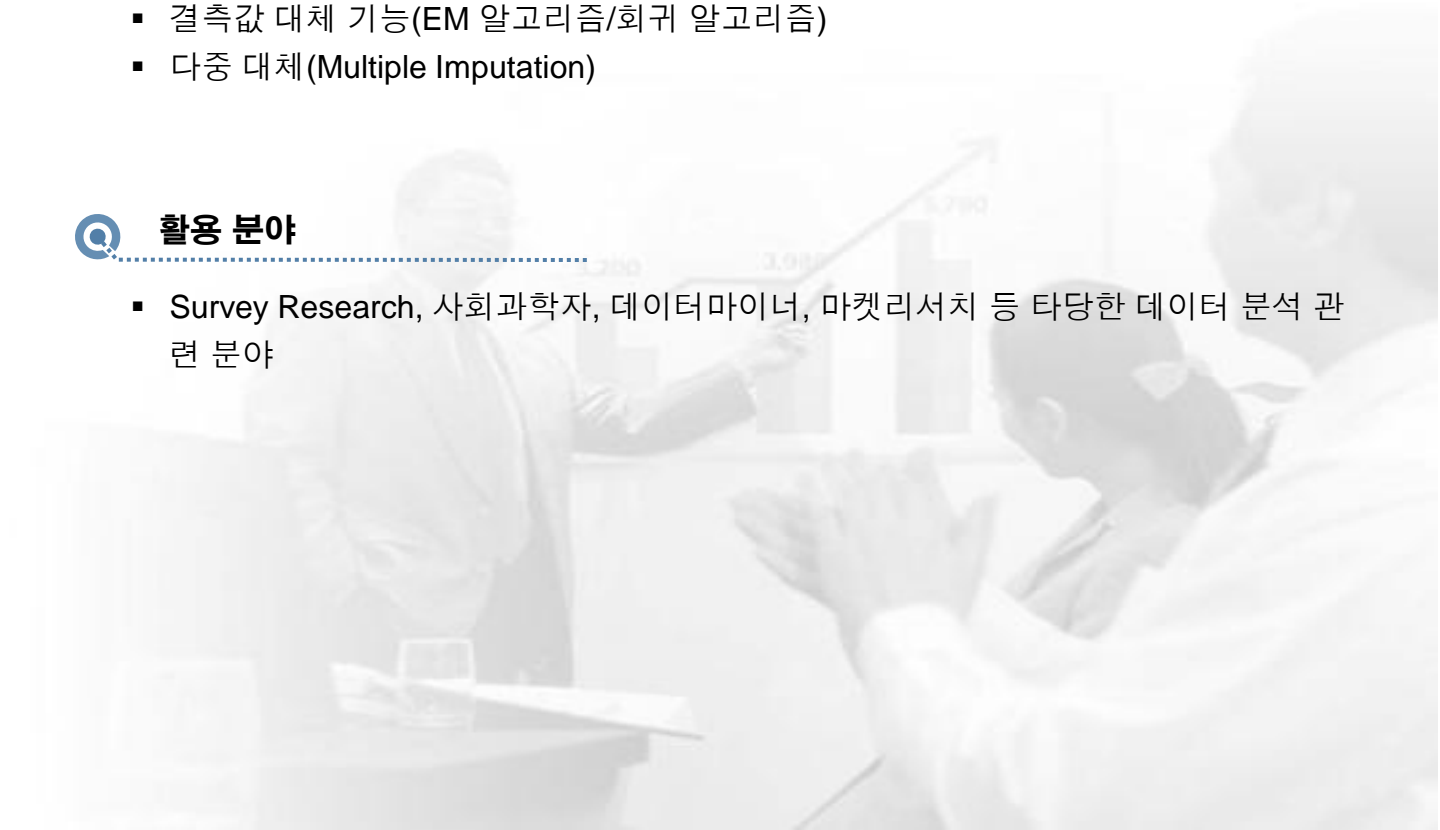
- 측값을 적절한 추정치로 대체하여 통계적으로 보다 유의한 결과 도출
- 데이터로부터 숨겨진 편향(bias)을 제거하고 보다 정확한 결론에 도달
- 발견된 결측치 패턴을 통해 문제가 되는 질문을 확인함으로써 설문 조사 질문 향상
- 추정치로 결측값을 쉽고 빠르게 교체

Q 주요 분석

- 결측 데이터 패턴 분석
- 결측값 대체 기능(EM 알고리즘/회귀 알고리즘)
- 다중 대체(Multiple Imputation)

Q 활용 분야

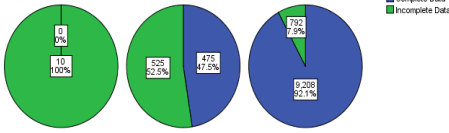
- Survey Research, 사회과학자, 데이터마이너, 마케터리서치 등 타당한 데이터 분석 관련 분야



> SPSS Missing Values

1

Overall Summary of Missing Values

Variable Summary^{a,b}

	Missing		Valid N	Mean	Std. Deviation
	N	Percent			
Household income in thousands	179	17.9%	821	71.1462	83.14424
Years at current address	150	15.0%	850	11.47	9.965
Marital status	115	11.5%	885		

a. Maximum number of variables shown: 25

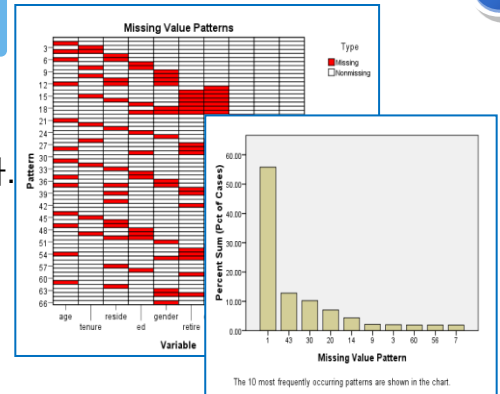
b. Minimum percentage of missing values for variable to be included: 10.0%

결측 데이터 탐색

- 먼저, 결측값이 어느 변수에 있으며 비중이 어느 정도인지를 탐색합니다. 위의 표는 결측 비율 10% 이상인 변수들을 모아서 결측 비율이 큰 순서대로 보여줍니다.

결측값 패턴 분석

- 우측 도표는 결측값 패턴을 보여줍니다. 결측 비율이 커지는 순서로 변수들이 왼쪽에서 오른쪽으로 배열됩니다. 결측값 패턴의 상대적 빈도를 출력하여 빈번히 발생하는 패턴을 확인 할 수 있습니다.



2

3

Descriptive Statistics

		tenure					
Data	Imputation	N	Mean	Std. Deviation	Minimum	Maximum	
Original Data		968	35.56	21.268	1.00	72.00	
Imputed Values	1	32	40.60	18.155	11.18	83.10	
	2	32	35.89	23.090	-1.03	79.61	
	3	32	37.56	22.469	-3.11	89.09	
	4	32	39.51	22.394	1.04	95.15	
	5	32	40.15	22.654	2.31	78.28	
Complete Data After Imputation	1	1000	35.72	21.186	1.00	83.10	
	2	1000	35.57	21.316	-1.03	79.61	
	3	1000	35.62	21.299	-3.11	89.09	
	4	1000	35.69	21.305	1.00	95.15	
	5	1000	35.71	21.317	1.00	78.28	

결측값 대체

- 대체 방법에는 자동과 수동이 있습니다. 수동인 경우엔 FCS(Fully Conditioned Specification) 상황에서 MCMC(Monte Carlo Markov Chain) 알고리즘을 적용합니다. 척도형 변수는 통상적으로 선형회귀모형이 상정되지만 대안으로 PMM(Predictive Mean Matching) 방법이 적용될 수 있습니다. 이 방법은 선형회귀모형을 쓰되 결측값에 대한 대체값을 적합모형으로부터 임의로 생성시키는 대신 예측값에 가장 가까운 관측값으로 대체하는 방법입니다.

> SPSS Complex Samples

복합 표본의 통계량을 보다 정확하고 쉽게 계산할 수 있습니다.

SPSS Complex Samples는 구조적인 복합설계에 따라 표본을 추출할 때 효과적으로 사용할 수 있으며 데이터 분석을 위한 구체적인 표본 설계 사항들이 타당한지 확인할 수 있는 모듈입니다. 다양한 샘플링 기법을 적용할 수 있으며 통계량과 표준오차를 정확하고 빠르게 계산하여 신뢰할 수 있는 표본을 얻을 수 있습니다.

Q 특징

- 서베이 데이터에 대한 정확한 분석
- 표본 추출 마법사를 통해 손쉽게 조사 계획 및 수행
- 샘플 데이터를 본 분석에 앞서 미리 분석
- 선형회귀, 분산분석(ANOVA), 공분산분석(ANCOVA) 모델 구축
- 범주형 및 수치형 결과를 정확하게 분석 및 예측
- 군집화를 통해 비용 절감

Q 주요 분석

- 층화 추출 (Stratified Sampling)
- 군집 추출 (Clustered Sampling)
- 다단계 추출 (Multistage Sampling)
- Complex Samples Descriptives(CSDESCRIPTIVES)
- Complex Sample Tabulate(CSTABULATE)
- Complex Samples General Linear Models(CSGLM)
- Complex Ordinals Selection(CSORDINAL)
- Complex Samples Logistic Regression(CSLOGISTIC)
- Complex Samples Cox Regression(CSCOXREG)

> SPSS Complex Samples

Q 서버이 데이터 분석 과정

표본선택마법사
(Sampling Plan Wizard)

분석준비마법사
(Analysis Preparation Wizard)

데이터 수집

- 표본 설계
- 표본 추출

Plan Files

데이터 분석

- 표본추출 방법 명시
- 사용자 간의 공유 및 저장 용이

Analyze data

Results

- SPSS Complex Samples를 통해 서버이 데이터에 대한 정확한 분석을 할 수 있습니다. 위와 같이 표본선택마법사와 분석준비마법사를 통해 계획을 작성하고 여러 가지 방법으로 데이터를 분석하여 그 결과를 쉽게 해석할 수 있습니다.

- **기술통계량** : 여러 변수에 대한 요약 일변량 통계량을 제공/1개 이상의 범주형 변수로 정의된 하위 그룹의 통계량 작성
- **교차 분석** : 선택된 변수들의 대응에 관한 교차분석표를 만들고 이원 통계량 출력/ 부그룹별로 통계량 출력 가능

> SPSS Decision Trees

데이터 내의 패턴과 분류를 확인하여 미래에 대한 예측을 할 수 있습니다.

SPSS Decision Tree는 데이터 내의 패턴과 관계를 파악하여 의사결정에 필요한 정보를 제공하고, 미래에 대한 예측을 돕는 모듈입니다.

특징

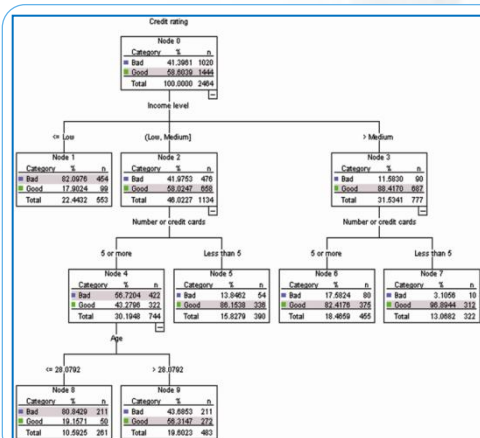
- 그룹 간의 관계를 파악하여 미래의 현상 예측
- 범주형 결과를 의사결정나무를 통해 시각적으로 명확히 설명
- 데이터 내에 있는 특정 부분집합 분석 가능

주요 분석

- IBM SPSS Statistics Syntax, SQL문, 단순텍스트 형식에서 분류/예측 규칙 생성
- CHAID, exhaustive CHAID, CART, QUEST 등 네 가지 tree-growing 알고리즘 중의 하나를 사용하여 분석 수행

활용 분야

- 고객 세분화
- 고객 이탈/유지 예측
- 제품 불량률 예측



의사결정 규칙(Rule)이 생성됩니다.

- 데이터 내에 숨어 있는 관계를 보여주는 고해상도의 의사결정나무입니다. IBM SPSS Statistics Classification Tree의 다이어그램, 테이블, 그래프를 통해 보다 쉽게 분석 결과를 해석할 수 있습니다.

> SPSS Neural Networks

데이터 내의 복잡한 관계를 보다 쉽게 파악할 수 있습니다.

SPSS Neural Networks는 비선형 데이터를 모델링할 수 있는 모듈로, 복잡하게 얽혀 있는 데이터 간의 연결 강도를 조절하여 문제를 분석합니다. 이를 통해 추론, 분류, 예측 등을 할 수 있습니다.

특징

- 비선형 데이터로 이루어진 현상 분석 시 용이
ex. 시간에 따른 날씨의 변화 등과 같이 수학적으로 해결하기 힘든 문제
- 과거의 통계학적 분석방법에 비해 분석 시간이 짧게 걸리고 비용이 저렴
- 패턴인식, 예측, 분류 등에 효과적
- 새로운 방법으로 데이터 탐색 및 예측 모델 생성

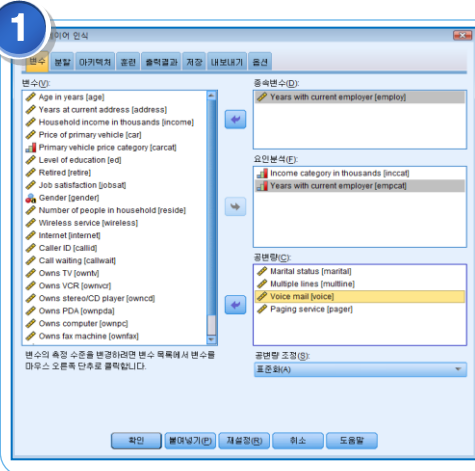
주요 분석

- 다층 퍼셉트론(Multilayer Perceptron)
- 방사형 기저함수(Radial Basis Function)

활용 분야

- 마켓 리서치
 - ✓ 고객 프로필 작성
 - ✓ 고객 성향 파악
- 데이터베이스 마케팅
 - ✓ 고객 DB 분류
 - ✓ 캠페인 최적화
- 경영 및 재정 분석
 - ✓ 신용도 분석
 - ✓ 보험 사기 감지
 - ✓ 현금 흐름 관리
- 의료 분야
 - ✓ 치료 비용 예측
 - ✓ 입원 기간 예측

> SPSS Neural Networks

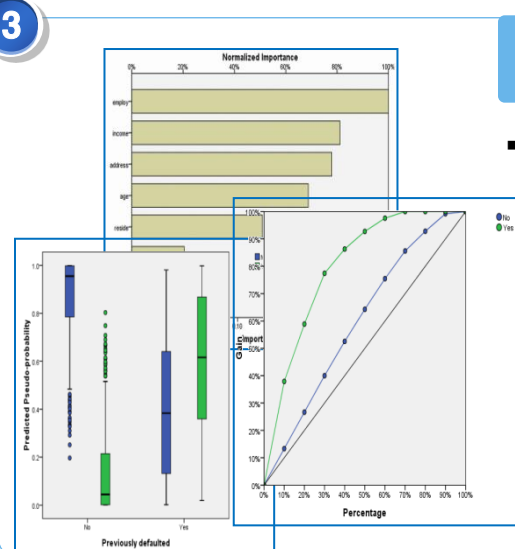
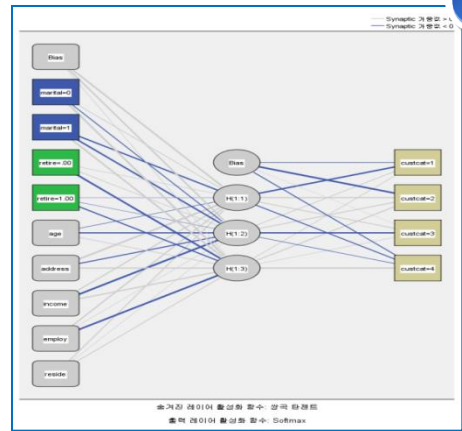


신경망의 다양한 알고리즘을 바로 선택하여 수행할 수 있습니다.

- 다층퍼셉트론(Multilayer Perceptron) 또는 방사형 기저 함수(Radial Basis Function)을 통해 모델에 포함하고 싶은 변수 및 공변량을 선택할 수 있습니다.

신경망 구조를 바로 결과로 확인할 수 있습니다.

- Input Layer부터 Output Layer까지 Network 구조를 확인할 수 있습니다.
- 중간 층인 Hidden Layer의 활성화함수를 직접 결과로 확인할 수 있습니다.



모델 정확도 평가를 위한 다양한 옵션이 있습니다.

- IBM SPSS Statistics Neural Networks 모듈을 통해 ROC 곡선, 누적 이익 도표, Lift 도표 등 다양한 그래프를 출력할 수 있습니다.

> SPSS Direct Marketing

고객 및 고객 정보를 다각도로 분석하여 마케팅 전략에 활용합니다.

고객 및 고객 정보를 보다 깊게 이해하기 위해서는 RFM(Recency, Frequency, Monetary) 분석, 군집분석 및 유망고객 프로파일 분석 등의 다양한 분석을 수행해야 합니다. 또한, 마케팅 분야의 캠페인을 개선하기 위해서는 필수적으로 응답 가능성 높은 지역 (우편번호)을 식별하거나 고객 성향 점수의 산출 및 캠페인 효과를 비교할 수 있는 테스트 단계를 거쳐야 합니다. SPSS Direct Marketing은 마케팅을 위한 고객 및 접촉 정보를 분석하여 마케팅 예산의 ROI를 극대화할 수 있습니다.

Q 주요 특징

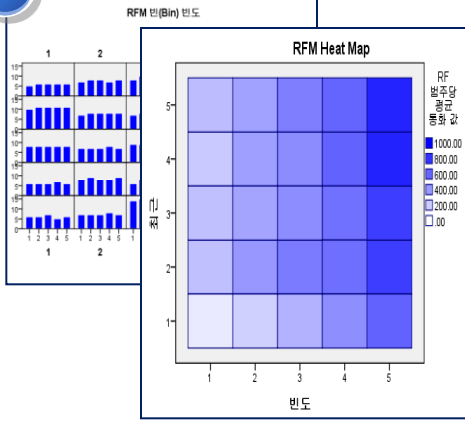
- RFM(Recency, Frequency, Monetary) 분석
- 고객 및 접촉 정보의 세분화
- 고객 및 접촉 정보의 프로파일 분석
- 구매 가능성이 높은 지역(Postal Code) 식별
- 고객 성향 점수 분석
- 캠페인 비교 테스트(통제집단 패키지 테스트)

Q 활용 분야

- 마케팅 분야
 - ✓ 고객 세분화 및 고객 가치 측정
 - ✓ 고객 성향 점수 산출
 - ✓ 고객 프로파일 분석
 - ✓ 다양한 캠페인 효과 분석 및 비교
 - ✓ 특정 프로모션에 반응할 가능성이 높은 고객 식별
 - ✓ 마케팅 전략의 개발

> SPSS Direct Marketing

1



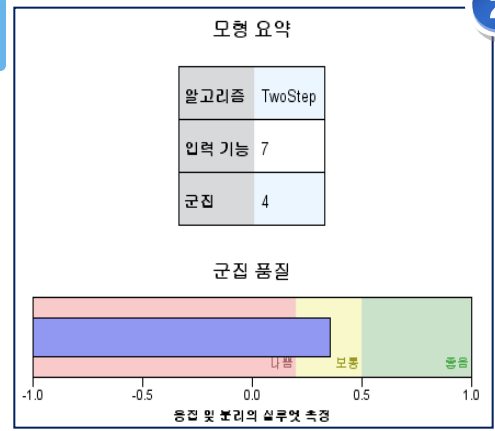
RFM 분석

- 고객 데이터 및 거래(Transaction) 데이터를 기반으로 구매시기(Recency), 구매빈도(Frequency), 구매금액(Monetary)의 3가지의 관점에서 고객의 점수를 바로 산출할 수 있습니다.

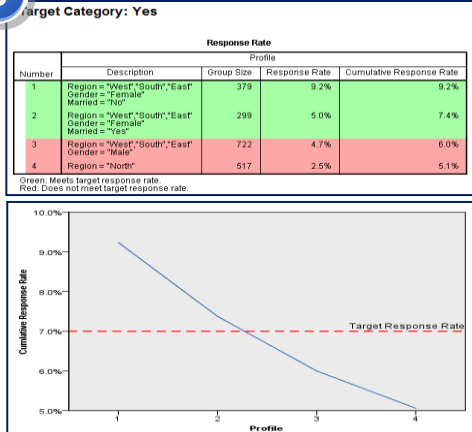
고객 및 접촉 정보의 세분화

- 고객 및 접촉 정보 데이터를 몇 개의 유사한 그룹으로 세분화하여 각 그룹의 특징을 파악합니다.

2



3

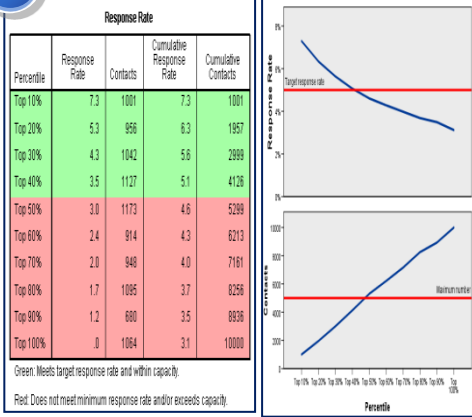


고객 및 접촉 정보의 프로파일 분석

- 마케팅 또는 캠페인 대상자를 Targeting하기 위한 고객의 특성들을 Profiling 함으로써 향후 캠페인 활동에 반영하는 분석입니다.

> SPSS Direct Marketing

4

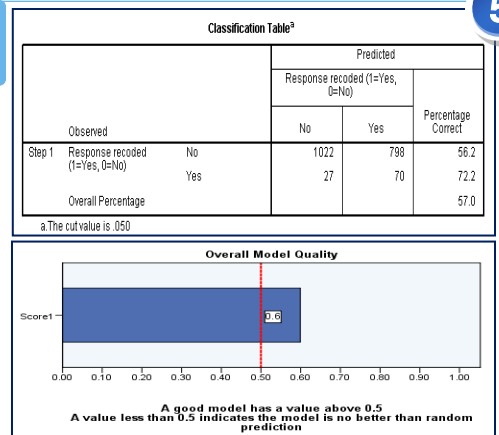


구매 가능성이 높은 지역(Postal Code) 식별

- 이전의 캠페인 활동에서 반응 비율이 높은 지역(Postal Code)을 기준으로 분석합니다.

고객 성향 점수 분석

- 고객의 성향에 따라 구매로 이어질 확률을 나타내는 점수를 생성합니다.



6

	Control Package				
		Control		Test	
		Count	Column N %	Count	Column N %
검정 필드 코딩변경(1=예 0=아니오)	0	875	96.2%	945	93.8%
	1	35	3.8%	62	6.2%

캠페인 비교 테스트 (통제집단 패키지 테스트)

- 기존의 캠페인과 새로운 캠페인의 차이를 테스트함으로써 마케팅 의사결정에 반영할 수 있습니다.

		Control Package	
		Control	Test
		(A)	(B)
검정 필드 코딩변경(1=예 0=아니오)	0	B	A
	1		

Results are based on two-sided tests with significance level 0.05. For each significant pair, the key of the category with the smaller column proportion appears under the category with the larger column proportion.

a. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.



> SPSS Bootstrapping

생성한 모형들에서 보다 안정적으로 결과를 산출할 수 있습니다.

Bootstrapping은 여러분이 수행하시는 통계 분석의 정확성을 높여주고, 생성한 모형이 안정적으로 결과를 산출하기 위한 유용한 기법입니다. 데이터로부터 표본 재추출(Re-sampling)을 빠르고 쉽게 수행함으로써 데이터에 따른 변화를 최소화하여 신뢰성을 높여주는 매우 효율적인 방법입니다.

특징

- 데이터로부터 반복적으로 표본 재추출(Re-sampling)을 수행하므로 분포를 쉽고 빠르게 추정할 수 있습니다.
- 데이터에서 이상값(Outliers)의 영향을 감소시키고 평균, 중위수, 비율, 오즈비(Odds Ratio), 상관계수, 회귀계수 등과 같은 각종 표본 추정 값들을 모집단의 모수와 비교하여 더욱 정확한 모수의 표준오차와 신뢰구간을 추정할 수 있습니다.
- 더욱 정확한 분석을 위해 수천 개의 다른 버전(version)의 데이터 셋을 생성하여 Bootstrapping을 수행할 수 있습니다.

주요 분석

- | | |
|---|---|
| <ul style="list-style-type: none"> ▪ 기술통계 프로시저 <ul style="list-style-type: none"> ✓ 기술통계 ✓ 빈도분석 ✓ 데이터탐색 ✓ 집단별 평균분석 ✓ 교차분석 ✓ 일표본/독립표본 T-검정 ✓ 상관분석의 이변량상관계수 ✓ 상관분석의 편상관계수 | <ul style="list-style-type: none"> ▪ 모델링 프로시저 <ul style="list-style-type: none"> ✓ 일원배치 분산분석 ✓ 일변량 일반선형모형 ✓ 일반화 선형모형 ✓ 판별분석 ✓ 로지스틱 회귀분석 ✓ 이분형/다항 로지스틱 회귀분석 ✓ 순서 회귀분석 ✓ 선형혼합모형 ✓ Cox 회귀모형 |
|---|---|

활용 분야

- 공공정책 조사 연구
- 사회과학 리서치
- 교육 리서치
- 각종 통계분석이 요구되는 분야

생동성 분석을 보다 쉽게 분석할 수 있습니다.

생동성시험이란 오리지널 의약품과 제네릭 의약품의 효능이 동등하다는 것을 입증하기 위해 두 의약품의 효능의 한 척도인 생체이용률의 동등성을 입증하는 통계적 시험입니다. SPSS Bio Equiv는 생동성 시험의 설계와 분석을 보다 쉽고 정확하게 수행하도록 지원하는 모듈입니다. 또한, 예제를 통해서 Monte Carlo Simulation을 실습할 수 있는 Syntax 명령문을 제공합니다.

특징

- 생동성 분석에 대한 참고자료로서 비모수 생동성 분석 및 집단 생동성 분석 지원
- 피험자 수가 같지 않을 경우에도 원활한 분석 가능
- 부적합한 결과가 나왔을 경우 원인 규명 자료로 활용 가능
- 생물학적 동등성 시험기준 제 19조 2에 따른 추가 시험 분석 지원

주요 분석

- 피험자수 계산
- 확률화
- 평균생동성 분석
- 잔차분석
- 비모수 생동성 분석
- 집단 생동성 분석
- 추가 시험 분석
- Monte Carlo Simulation Application

> SPSS Bio-Equiv

1

피험자 수 계산과 확률화를 매우 쉽게 수행

유의수준(α) 0.05

검정력($1-\beta$) 0.80 군별 최소 피험자 수

제제평균차이(δ) 0.05

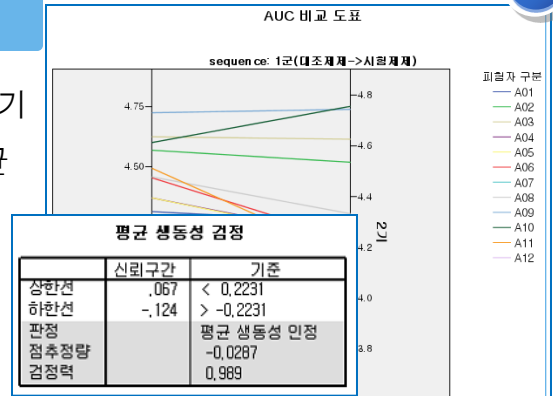
변동계수(CV) 0.2

제제 표준편차(σ) 0.2

- 유의수준, 검정력, 제제평균차이 등 시험 설계 조건을 자유롭게 설정한 후 군별 최소 피험자 수를 계산할 수 있으며 한 명의 피험자를 두 순서집단에 랜덤 배치할 수 있습니다.

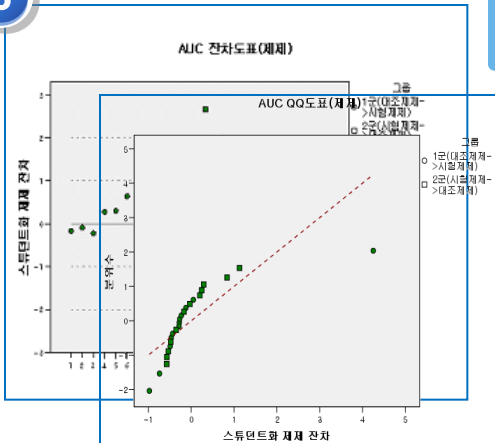
평균 생동성 분석

- AUC, Cmax 그래프 및 신뢰구간을 계산하여 생동성 기준 값인 ($\log(0.8), \log(1.25)$)와 비교하여 제제간의 평균 생동성 검정할 수 있습니다.



2

3



잔차 분석을 통하여 분석의 타당성 검토

- 분석의 가정 검토를 위해서 피험자내 잔차와 피험자간 잔차를 이용하여 잔차 도표를 출력하여 이상점 여부를 판단할 수 있고 정규성을 평가할 수 있는 QQ도표, Kolmogorov-Smirnov 검정결과가 출력됩니다.

> SPSS Bio-Equiv

4

비모수 생동성 검정

	신뢰구간	기준
상한선	.053	< 0.2231
하한선	-.096	> -0.2231
판정		비모수 생동성 인정
중앙값		-0.0120

비모수 생동성 분석

- 시험 데이터가 정규분포를 따르지 않을 경우 Wilcoxon-Mann-Whitney 방법을 활용하여 제제간의 생동성 여부의 판정을 할 수 있습니다.

변동성을 함께 고려한 생동성 평가

- 제제간의 평균과 변동성을 함께 고려한 US FDA 기준을 적용함으로써 평균 생동성 검정의 타당성을 더욱 높일 수가 있습니다.

집단 생동성 검정 (US FDA)

	값	기준
통계량	-.035	< 0.0000
판정		US FDA 기준 집단 생동성 인정

집단 생동성 검정 (Modified US FDA)

	값	기준
통계량	-.047	< 0.0000
판정		US FDA 기준 집단 생동성 인정

5

6

AUC 일관성 체크

	값	기준
군당 최소 피험자수	12	≥ 12
본시험과 추가시험 간의 F-검정	.184	$> .05$
본시험과 추가시험 간의 제제효과	.678	$> .05$
F-검정		
판정		본시험과 추가시험의 일관성 인정

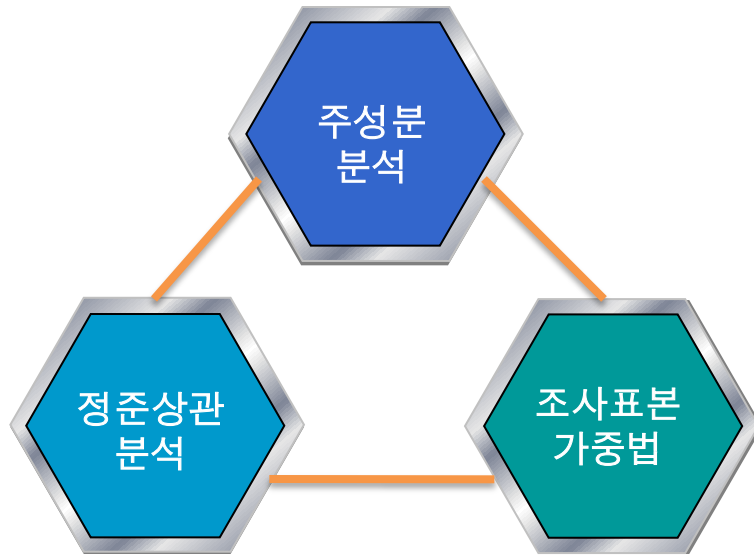
추가 시험 분석

- 생동성 시험 결과 제제간의 생동성을 입증하지 못했다면 추가시험을 실시 할 수 있습니다. 이 때, 추가된 시험을 결합하는 타당성을 확보할 수 있는 검정을 수행할 수 있습니다.

> SPSS Korea Options Pack

주성분 분석, 정준상관분석, 조사표본 가중법이 가능합니다.

SPSS Plus-Stat는 Syntax를 이용하여 불편하게 수행해왔던 몇몇 분석 기능을 대화상자를 이용하여 보다 편리하게 분석할 수 있는 모듈입니다.



특 징

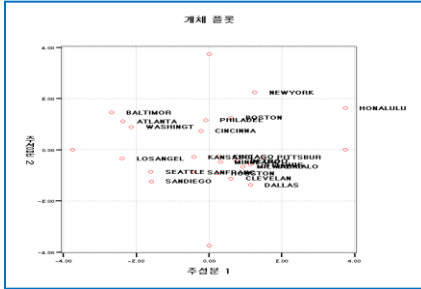
- 기존 Windows 버전에서 GUI(Graphic User Interface)에서 지원하지 않은 새로운 분석 기능 및 데이터 변환 기능 제공
- Syntax를 이용하여 불편하게 수행한 각종 기능을 편리한 대화상자로 수행

주요 분석

- 주성분 분석 - 자료의 축약
- 정준상관분석 - 변수 집단 간 연관성 파악
- 조사표본 가중법 - 조사 결과에 대한 신뢰성과 정확성을 확보

> SPSS Korea Options Pack

1



주성분 분석(Principal Component Analysis)

- 변수 플롯(variables plot)과 개체 플롯(observation plot)을 통해 변수가 저차원 주성분 공간에 놓이는 방향을 그래프로 쉽게 확인할 수 있습니다.

정준상관분석(Canonical Correlation Analysis)

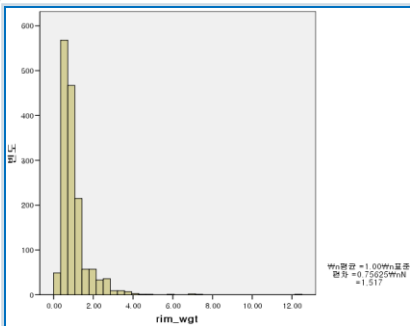
- 정준상관분석은 다수의 독립변수들과 다수의 종속변수들 사이에 존재하는 관계(패턴)를 검토하는 것입니다. 변수 군별로 개체들의 정준변량 산점도와 변수 플롯을 만들 수 있습니다.

중복도 분석(Redundancy Analysis):

자체 정준변량에 의해 설명된 변수세트-1의 분산비율	
S1_CV001	.274
S1_CV002	.172
S1_CV003	.554
변수세트-2의 정준변량에 의해 설명된 변수세트-1의 분산비율	
S2_CV001	.174
S2_CV002	.048
S2_CV003	.061
자체 정준변량에 의해 설명된 변수세트-2의 분산비율	
S2_CV001	.154
S2_CV002	.254
S2_CV003	.191
변수세트-1의 정준변량에 의해 설명된 변수세트-2의 분산비율	
S1_CV001	.104
S1_CV002	.071
S1_CV003	.021

2

3



조사표본가중법(Rim Weighting Survey Samples)

- 가중법(weighting method)은 모집단과 표본간 편차를 줄이거나 없애므로써 조사 결과의 정확성을 제고하기 위한 통계적 보정 방법입니다.