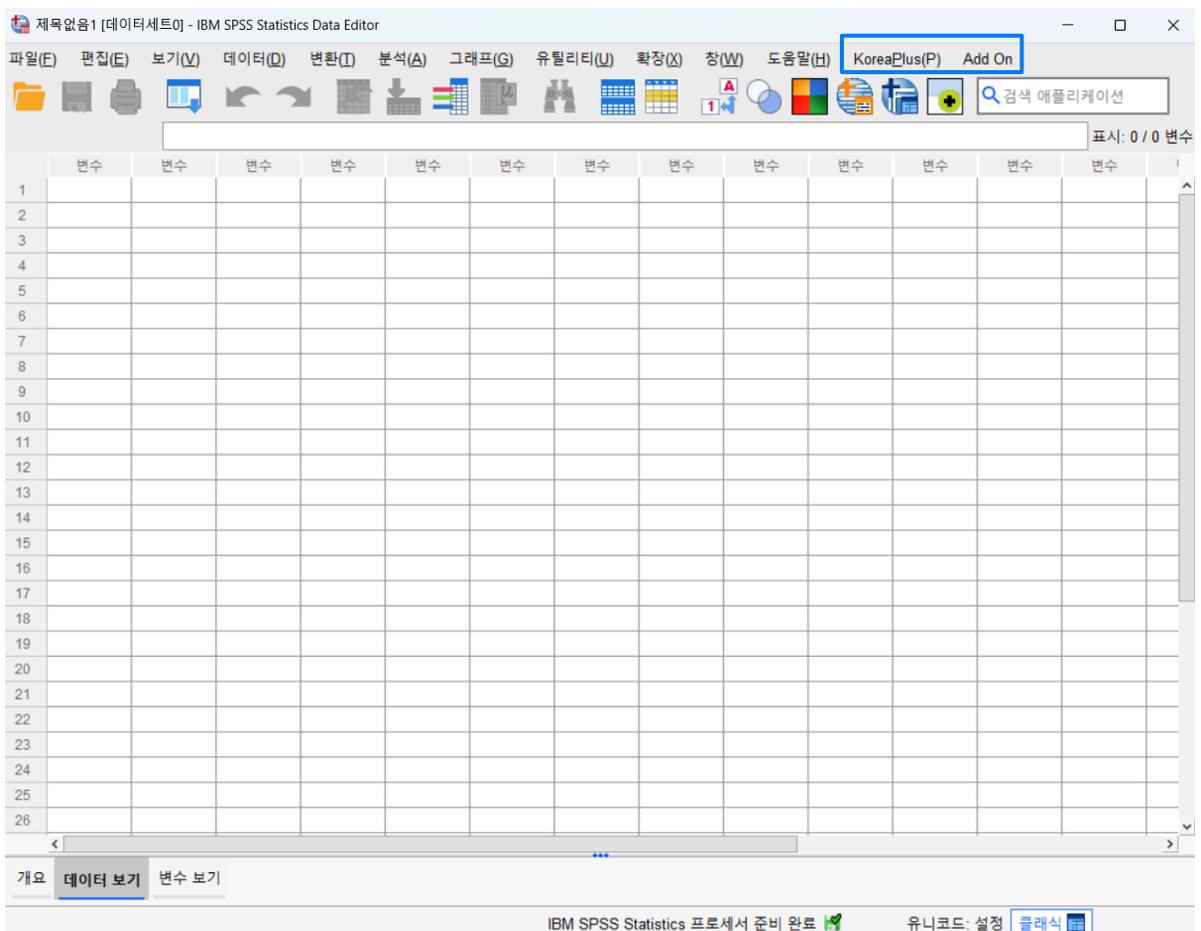


# KoreaPlus Statistics

## Embedded on IBM SPSS Statistics Professional



# KoreaPlus Statistics

## - Embedded on IBM SPSS Statistics

### Packages Professional

01

#### DATASOLUTION의 KoreaPlus Statistics

- IBM에서 소개한 KoreaPlus Statistics
- KoreaPlus Statistics 소개
- 나라장터 등록 정보 [목록정보시스템]

02

#### Embedded on IBM SPSS Statistics

- IBM SPSS Statistics 특징
- New IBM SPSS Statistics Ver.31

03

#### Packages Professional

- 패키지 구성 (Standard, Professional, Premium)
- Professional 구성 모듈
- 모듈 별 분석 기법 상세

04

#### Add-On KoreaPlus Statistics Modules

- KoreaPlus의 Add-On 기능
- AI Coach
- KoreaPlus의 분석 기능 상세
- KoreaPlus Statistics Utility 1.0

05

#### DATASOLUTION Service

- 제품 구성
- 기술지원 및 운영 서비스
- Community SPSS User Portal
- 교육지원 빅데이터러닝센터

# KoreaPlus Statistics

- Embedded on IBM SPSS Statistics Professional



## | DATASOLUTION의 KoreaPlus Statistics

- IBM에서 소개한 KoreaPlus Statistics
- KoreaPlus Statistics 소개
- 나라장터 등록 정보 [목록정보시스템]

# 데이터솔루션의 KoreaPlus Statistics

(주)데이터솔루션은 한국 IBM사와 IBM Business Partner **ESA (Embedded Solution Agreement) 계약**을 체결하였으며, 이에 따라 KoreaPlus Statistics 소프트웨어의 제품공급, 기술지원, 자문컨설팅 등 관련 업무를 하고 있습니다. IBM Analytics(SPSS포함) Sales & Tech Certification을 보유하여 소프트웨어 납품 외에 기술력을 인증 받은 파트너로서 고객 관리 및 기술지원을 대행 할 수 있는 **IBM Platinum Business Partner** 입니다. **IBM Support Provider 자격**으로 SPSS 전문 기술지원이 가능하며, SPSS와 관련된 분석 기법 및 확장 모듈을 개발하고 판매할 수 있습니다.

IBM Partner World에서 확인 할 수 있습니다.

## IBM에서 소개한 KoreaPlus Statistics

The screenshot displays the IBM Partner Plus Directory page for Datasolution, Inc. The page features the IBM logo and navigation menu at the top. The main heading is 'DATAsolution' with the sub-heading 'Datasolution, Inc.'. Below this, it states 'Datasolution is an IBM Platinum business partner, focused on advanced statistical analysis tools and solutions for today's business needs.' The 'Solutions' section is highlighted, showing 'KoreaPlus Statistics' as a featured solution. A description follows: 'KoreaPlus Statistics was developed as advanced statistical analysis tool based on IBM SPSS statistics, which is optimized for Korean users.' A 'View more' link is provided. The 'Company Overview' section is partially visible, mentioning 'Datasolution provides highly specialized solutions focused on advanced statistical analysis' and 'As a IBM Platinum Business Partner in Korea, Datasolution is certified from the IBM software platform(Watson) in order to help customers with our business solutions.' Contact information includes the address '10F Hyundai Intellex, 620 Eonju-ro, Gangnam-g Seoul, Seoul Teugbyeolai 06101 Korea, Republic of', telephone '+82 2 34677228', and website 'http://www.datasolution.kr'. The partner type is listed as 'Independent Software Vendor (ISV) VAR/Reseller/Solution Provider'.

# KoreaPlus Statistics 소개

KoreaPlus Statistics는 IBM SPSS Statistics 전 기능에 데이터솔루션에서 개발한 KoreaPlus 모듈과 기술지원 서비스, 교육 서비스가 추가 된 KoreaPlus 패키지 입니다.

## KoreaPlus Statistics 기능

### ✓ 분석 기능

+ IBM SPSS Statistics 전 분석 기능

+ KoreaPlus Modules

- Meta Analysis
- Medical Analysis +
- Conjoint (SPSS Conjoint 모듈 필요)
- Multivariate Analysis
- Text Analytics +
- CVA(Customer Value Analysis)
- Sample Size Calculation
- Scale
- Weibull Analysis
- AI (RNN기반예측분석) +
- Data Station (Data Lake)
- PSM (Propensity Score Matching)

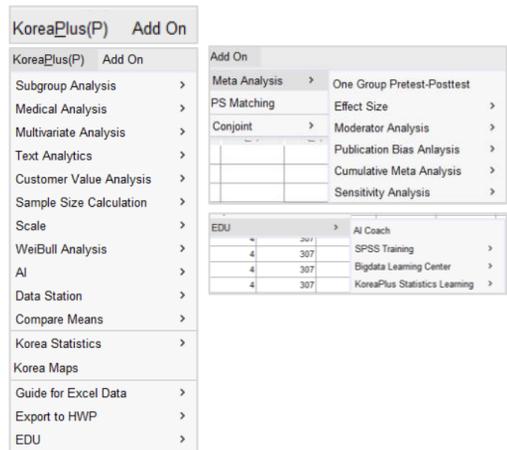
### 최신기능

- + Meta Analysis : 분석 기능 추가
- + 평균 검정 : 분석 기능 추가
- + AI : 네 가지 머신러닝 알고리즘 추가
- + Text Analytics : 동의어 사전 추가 외
- + Subgroup Analysis (Cox, Logistic, Poisson)
- + PSM : 결측치 처리(전처리 기능) 추가
- + AI Coach 기능

### ✓ 사용자 편의 기능

- KoeaPlus Statistics(국내 통계포털 연계)
- 한국 Maps 기능 추가
- SPSS Training 교육 자료 외
- Guide for Excel Data
- Output 한글(.hwp) 내보내기
- KoreaPlus 사용자 온라인 무료 교육 링크

## KoreaPlus Statistics 메뉴



### ✓ KoreaPlus 메뉴

- KoreaPlus 메뉴 생성
- Add on > PS Matching 생성

## DATASOLUTION Service

### ✓ 전문 기술 지원 서비스

- 기술지원 접속 코드 (Serial Number)

### ✓ 자체 러닝센터 교육 지원 서비스

- 빅데이터러닝센터 (On/OFF-Line)교육쿠폰

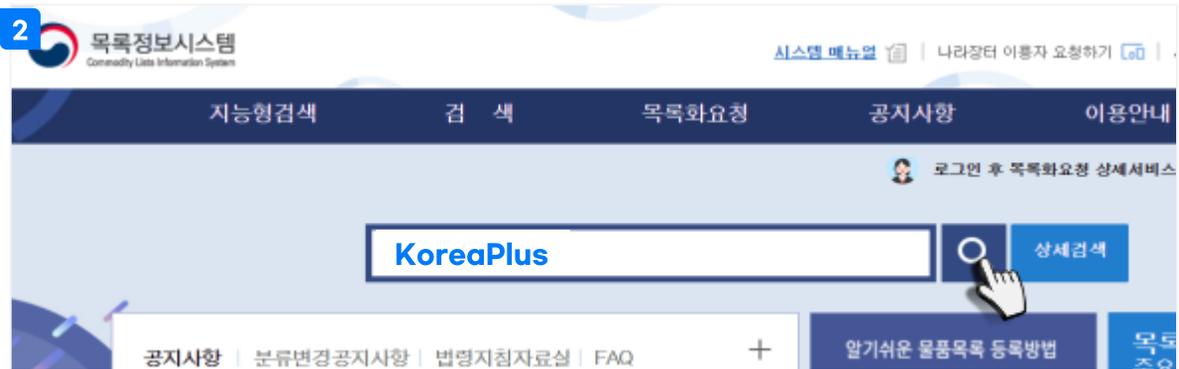
### ✓ 설치 프로그램 및 매뉴얼 서비스

- 다운로드 링크 및 이메일 인증서
- 분석 매뉴얼 (한글, 영문) PDF

# 나라장터 등록 정보

KoreaPlus Statistics는 [나라장터 - 목록정보시스템] 에 등록되어 있습니다.

- 품명 : 해석 또는 과학소프트웨어
- 물품분류번호 43232605 – 23336509
- 모델명 : KoreaPlus Statistics
- 제조업체명 : (주)데이타솔루션
- 세부품명 : 응용 과학용 소프트웨어
- 품목구분 : 일반용 품목



# KoreaPlus Statistics

- Embedded on IBM SPSS Statistics Professional



## | Embedded on IBM SPSS Statistics

- IBM SPSS Statistics 특징
- New IBM SPSS Statistics Ver.31

# IBM SPSS Statistics 특징

KoreaPlus Statistics에는 IBM SPSS Statistics가 Embedded 되어 있습니다. IBM SPSS Statistics는 전 세계에서 가장 많은 사용자를 보유하고 있는 글로벌 통계소프트웨어로, 데이터 입력 및 관리, 집계 통계 분석 등의 작업을 수행할 수 있습니다. 또한 분석 결과를 표와 그래프로 시각화할 수 있는 전문적인 통계분석 도구입니다.



## 1 비즈니스 및 연구 과제 해결

더 심층적인 통찰력을 발견하고 프레젠테이션 준비 보고서, 시각화 및 지리적 분석을 통해 분석 과제를 해결합니다.



## 2 고급사용자를 위한 고급 기능

명령어 구문 및 외부 프로그래밍 언어 (예: R, Python 등)와 같은 프로그래밍 옵션을 통해 분석 역량, 유연성 및 생산성을 극적으로 높입니다.

### ✓ 특징 / 장점

- 타 통계 툴에 비해 사용이 쉽고 기획에서 결과까지 분석 라이프사이클 전체를 지원합니다.
- 어떤 종류의 데이터셋이라도 빠르게 분석이 진행되며, 대용량 데이터의 준비 과정이 단축됩니다.
- 지리학적, 시계열 데이터를 통합하여 모델링하고 숨겨진 상관관계와 트렌드를 발견할 수 있습니다.
- 불확실한 결과에 대해서는 몬테카를로 시뮬레이션을 이용하여 모델링이 가능합니다.
- 피벗 기술을 활용하여 표, 지도, 그래프를 작성할 수 있습니다.
- 메뉴 기반의 사용자 인터페이스로 사용이 용이하며, Command Syntax와 프로그래밍도 활용 가능합니다.
- R, Python, JAVA, .Net 등 다양한 프로그래밍 옵션을 활용한 프로세스 자동화가 가능합니다.
- 통합된 R 개발환경에서 R 프로그램을 개발 및 테스트할 수 있습니다.
- IBM SPSS Statistics는 모든 데이터 형식과 다양한 프로그래밍 언어, OS를 지원합니다.
- 다른 툴과 시스템과의 연계가 용이합니다.
- 결과물을 스마트폰이나 태블릿 등 다양한 디바이스에서 확인할 수 있습니다.
- 워드, 파워포인트, 엑셀, BI, TM1 등 다양한 형태로 결과물을 내보낼 수 있습니다.

# IBM SPSS Statistics 특징

SW 설치 시 오픈소스 분석툴인 R을 만나볼 수 있습니다. 프로그래밍 기능 및 제품의 확장 허브를 통해 별도의 R프로그램 활용 없이 IBM SPSS Statistics 내에서 손쉽게 사용할 수 있도록 지원하고 있습니다. IBM SPSS Statistics 설치 시 자동으로 확장 기능들이 설치되어 활용할 수 있습니다.

## Statistics의 확장



### 01 | R 연동 기능

일반화 부스팅 회귀 예측 / 일반화 부스팅 회귀 / 이질적 상관관계 / 정준상관 / 회귀상대적중요도 / 회귀불연속 / tobit 회귀 / 로버스트 회귀 / Apriori / 문항반응모형 / 등급반응모형 / TURF 분석 / 2개 변수, 집단 Q-Q도표 / 수정 P 값 계산 / 범주순서 정의 / 요약 독립표본 T검정 / 계수 모형 / 비율 회귀 예측 / 비율 회귀 / 확장 rasch / GARCH 모형 / 모수적 회귀분석 / 분위수 회귀 / Firth 로지스틱 회귀 / 방정식 시스템 / 잠재계층분석



### 02 | 시스템 기능

R 작업 공간 가져오기 / 변수 정보 수집 / 데이터파일 검색 / 파일 전체의 문자열 너비조정 / 레이크 가중 값 / 성향 점수 매칭 / 케이스 대조 매칭 / 파일로 분할 / Programmability 변환 / 더미변수 작성



### 03 | 유틸리티

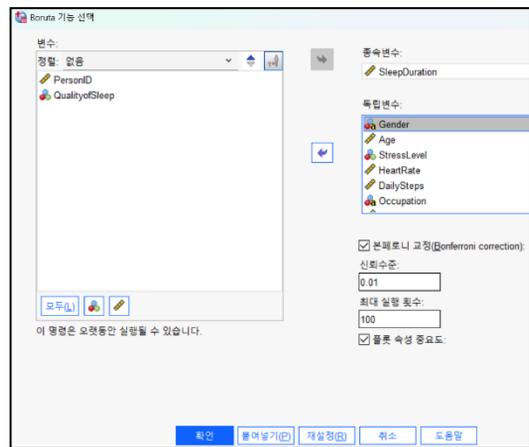
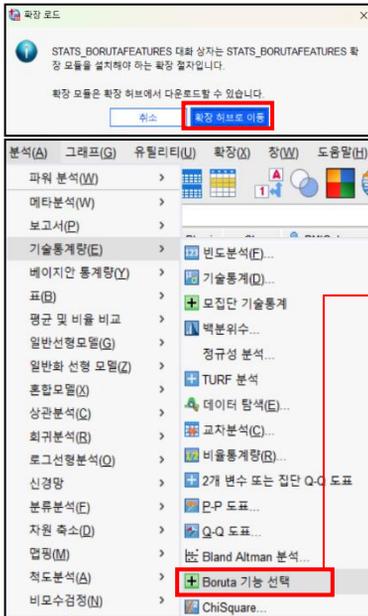
Weibull 도표 / 부집단 비교 / 회귀 변수 도표 / 피벗표를 사용한 계산 변수 매크로 정의 / 중도절단표 / 텍스트 출력 작성 / 프로세스 데이터파일

# New IBM SPSS Statistics Ver.31

Ver.31에서는 원본 변수와 데이터를 무작위로 섞어 만든 가짜 변수의 중요도를 비교하여, 중요하지 않은 변수를 제거하고 유의미한 변수를 선별하는 알고리즘이 추가되었습니다.

## ☑ Boruta Feature Selection

- Extensions > Extension hub > STATS\_BORUTAFEATURES 검색 및 설치

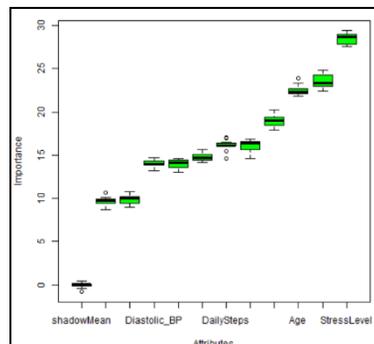


**변수 별 중요도 Box Plot ▼**  
원본 변수가 Shadow보다 높으면 Confirmed 시각적으로 변수 선택 근거 확인 가능

- 출력결과

기능	
Diastolic_BP	Confirmed
Systolic_BP	Confirmed
SleepDisorder	Confirmed
BMICategory	Confirmed
PhysicalActivityLevel	Confirmed
Occupation	Confirmed
DailySteps	Confirmed
HeartRate	Confirmed
StressLevel	Confirmed
Age	Confirmed
Gender	Confirmed
Dependent variable: SleepDuration	
Pvalue: 0.01, Bonferroni: 예, Boruta Version: 9.0.0	

◀ 선별된 변수 표시  
중요도 순으로 나열되고,  
유의미한 변수로 선별된  
변수는 Confirmed 표시

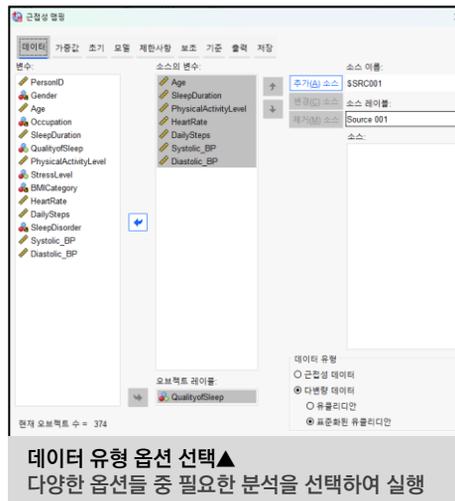


# New IBM SPSS Statistics Ver.31

Proximity mapping(근접성 맵핑)이란 다양한 형태의 근접성 데이터를 통합하여 변수들이 서로 얼마나 가까운 관계에 있는지를 공간상의 거리로 맵핑해 주는 기능입니다.

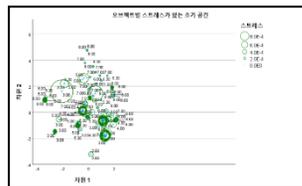
Ver.31에서는 다양한 근접성 소스 지원 및 변환/제한 옵션 제공하고 있습니다.

## Proximity mapping

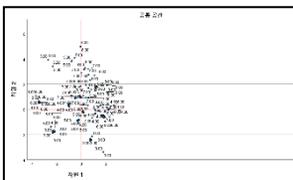


분석 요약	
오류	NO ERROR
결과	CONVERGED ( RELATIVE FUNCTION DIFFERENCE SATISFIED )
반복	219
최종 정규화된 스트레스	0.022913

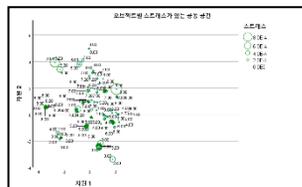
[분석 요약표]



[초기 공간 그래프]



[공통 공간]



[공통 공간 그래프]

# New IBM SPSS Statistics Ver.31

Distance Correlation(거리 상관분석)란, 선형 관계뿐만 아니라 비선형적 의존성까지 모두 포착하여, 변수 간의 진정한 연관성을 탐지하는 척도입니다. Ver.31에서 정식 프로시저로 탑재되어, 복잡한 실제 데이터의 관계를 누락 없이 정밀하게 분석할 수 있습니다.

## Distance Correlation

The screenshot illustrates the workflow for performing Distance Correlation in SPSS. It starts with navigating through the menu to 'Analyze > Distance Correlation'. The 'Distance Correlation' dialog box is shown with 'Age' and 'Diastolic\_BP' selected as input variables. The 'Standardization' sub-dialog is also visible, showing 'Min-Max' as the chosen method. Finally, the 'Distance Correlation: Plot' dialog is shown with 'Age' and 'Diastolic\_BP' selected for the X and Y axes.

가변	Label	N	평균	표준 편차	최소값(N)	최대
1	Age	374.000	42.185	8.673	27.000	59.000
2	SleepDuration	374.000	7.132	.796	5.800	8.500
3	PhysicalActivityLevel	374.000	59.171	20.831	30.000	90.000
4	HeartRate	374.000	70.166	4.136	65.000	86.000
5	DailySteps	374.000	6816.845	1617.916	3000.000	10000.000
6	Systolic_BP	374.000	128.554	7.748	115.000	142.000
7	Diastolic_BP	374.000	84.650	6.162	75.000	95.000

[분석 대상 변수의 기술 통계량]

변수 대응	dCov	dCor	95.00% 신뢰구간	
			하한(L)	상한(U)
Age 대 SleepDuration	.0080	.1893	.0096	.2852
Age 대 PhysicalActivityLevel	.0051	.1006	.0000	.2000
Age 대 HeartRate	.0028	.1216	.0205	.2203
Age 대 DailySteps	.0028	.0928	.0000	.1924
Age 대 Systolic_BP	.0159	.4048	.3164	.4863
Age 대 Diastolic_BP	.0169	.3947	.3056	.4770
SleepDuration 대 PhysicalActivityLevel	.0134	.2269	.1284	.3209
SleepDuration 대 HeartRate	.0100	.3777	.2873	.4614
SleepDuration 대 DailySteps	.0059	.1659	.0656	.2628

[거리 상관관계 추정치 표(dCov 포함)]

[산점도 이미지]

# New IBM SPSS Statistics Ver.31

Time Series Filters(시계열 필터)란, 시계열 데이터를 추세와 순환 성분으로 분해하여, 장기적 흐름과 단기적 경기 변동을 명확히 식별하는 기법입니다. Ver.31에서는 Python 기반의 확장 기능으로 고급 시계열 분해 도구가 추가되었습니다.

## Time Series Filters

### 확장 번들에서 Time Series Filters 확장 설치 Extensions > Extension hub > Time Series Filters 검색 및 설치

**확장 로드**

tsf 대화 상자는 tsf 확장 모듈을 설치해야 하는 확장 절차입니다.

확장 모듈은 확장 허브에서 다운로드할 수 있습니다.

[취소](#) [확장 허브로 이동](#)

**Time Series Filters**

Fits time series filter using Python stats model packages.

최신 버전: 1.0.0 날짜: 2025-5-26

[자세한 정보...](#)

Python 3 | 분석

필수조건  기확장 가져오기

### DATE\_변수 정의 필요 데이터 > 날짜 및 시간 정의 > 케이스 설정

**날짜 정의**

케이스의 날짜(C):

- 주, 근무일(6)
- 시, 일, 시
- 일, 근무시간(8)
- 주, 일, 시
- 주, 근무일, 근무시간
- 분
- 시, 분
- 일, 시, 분
- 초

첫번째 케이스(F):

주의 최고 수준

초: 1

현재 날짜:

초(1)

[확인](#) [재설정\(R\)](#) [취소](#) [도움말](#)

### 시계열 필터 분석 경로 분석 > 시계열 분석 > 시계열 필터

**시계열 필터**

필터:

- Hadrick-Prescott(HP)
- Baxter-King(BK)
- Christiano-Fitzgerald(CF)

변수:

필터: 없음

- time
- PUC\_STAT\_Ems
- RLY\_AC\_Ems
- TOFR\_Ems
- TQI\_ACOR\_Ems
- VS\_Ems
- ACC\_ACT\_Ems
- ACKES\_Ems
- BRAKE\_ACT\_Ems
- CONF\_ML\_FMY\_Ems
- CONF\_TCU\_Ems
- EMG\_CHL\_Ems
- EMG\_VCH\_Ems

\* HP 변수: 시계열 분해할 수치 변수

HP 변수: **N Ems**

값: 4368

BK 및 CF 변수:

제한도:

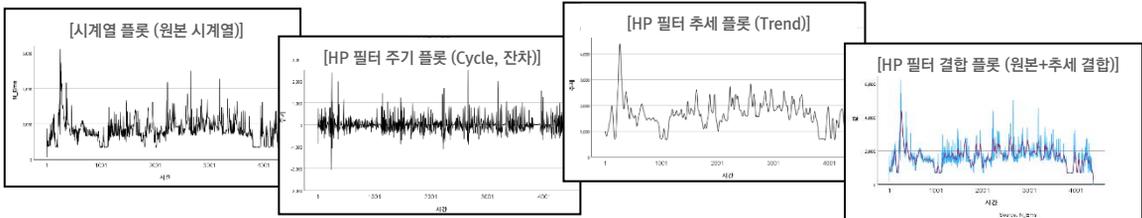
상한: 6

하한: 32

FC: 6

LC: 12

도움말

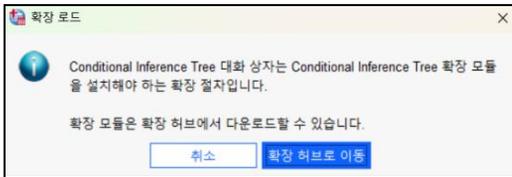


# New IBM SPSS Statistics Ver.31

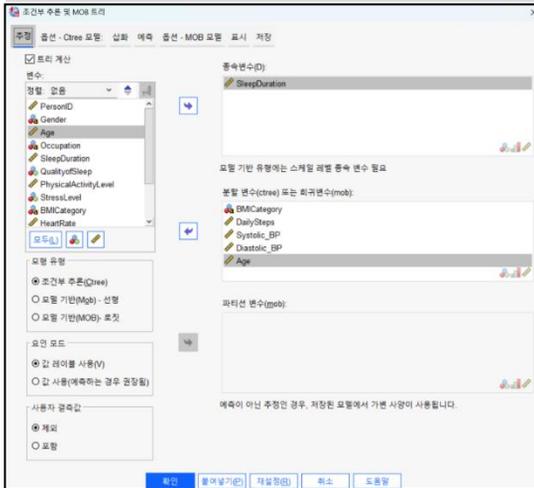
Conditional Inference Tree (조건부 추론 트리)란, 과적합을 방지하고 통계적 안정성을 높이기 위해, 조건부 추론 및 순열 검정 방식을 사용하는 분류 및 회귀 기법입니다. Ver.31에서 R 기반의 확장 기능으로 추가되었습니다.

## Conditional Inference Tree

확장 번들에서 Conditional Inference Tree 확장 설치  
Extensions > Extension hub > STATS\_CITREE 검색 및 설치



### 조건부 추론 트리 경로 분석 > 분류 분석 > CI 트리

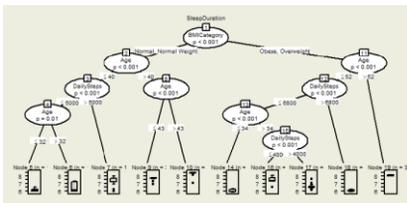


트리 매개변수		values
종속변수		SleepDuration
독립변수		BMICategory, DailySteps, Systolic_BP, Diastolic_BP, Age
요인 모드		labels
ID 변수		--
모델 유형		ctree
케이스 수		374
트리 레벨 수		5
터미널 노드 수		10
소스 파일		workspace
추정 날짜		Wed Jan 7 15:16:05 2026
저장된 모델 파일		--저장되지 않음--
사용자 결측값		제외
검정 통계량		quadratic
분할 통계		quadratic
검정 유형		Bonferroni
로그 회소 기준		-0.0512932943875506
허용되는 최소 노드 케이스		--
터미널 노드의 최소 사례		7
터미널 노드에 대한 케이스의 최소 비율		0.01
최대 깊이		99999
최대 대러언		--
R 파티클 파티클 지은 계산된 버전 1 2 24		

▼ 주요 독립 변수 기준	
변수 중요도	
가변	중요도 로그 가능성도
BMICategory	.994
Age	.923
DailySteps	.209

터미널 노드 통계: SleepDuration			
노드	케이스 수	케이스 중 %	평균
5	18.000	4.813	6.161
6	17.000	4.545	6.471
7	113.000	30.214	7.324
9	33.000	8.824	7.703
10	35.000	9.358	8.374
14	9.000	2.406	6.011
16	7.000	1.872	7.443
17	77.000	20.588	6.548
18	32.000	8.556	6.066
19	33.000	8.824	8.094

여러 노드가 있더라도 각 노드에는 하나만 표시됩니다.



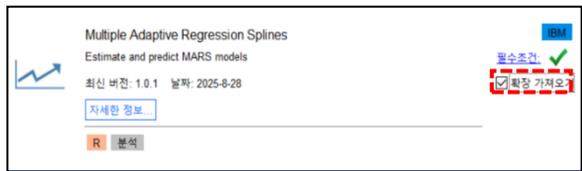
[Tree 플롯]

# New IBM SPSS Statistics Ver.31

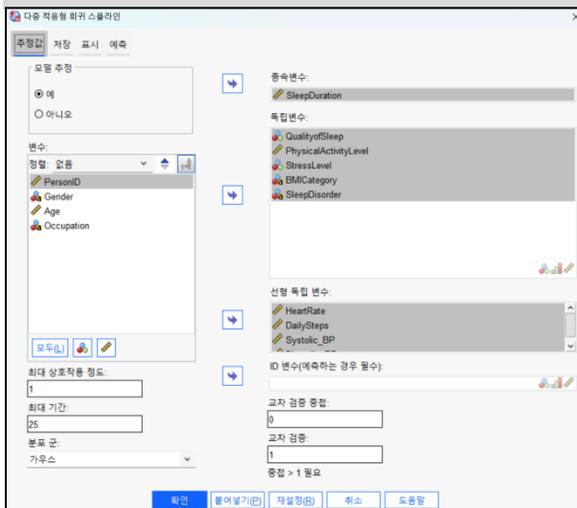
MARS(다중 적응형 회귀 스플라인) 알고리즘이란, 데이터 내의 복잡한 비선형성과 변수 간 상호작용을 정밀하게 탐지하는 회귀 분석 기법입니다. Ver.31에서 R 기반의 확장 기능으로 추가되었습니다.

## Multiple Adaptive Regression Splines

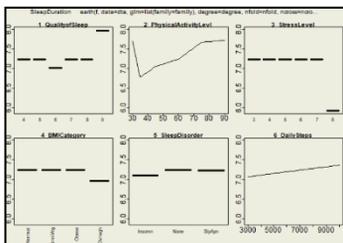
확장 번들에서 Multiple Adaptive Regression Splines 확장 설치  
Extensions > Extension hub > STATS\_EARTH 검색 및 설치



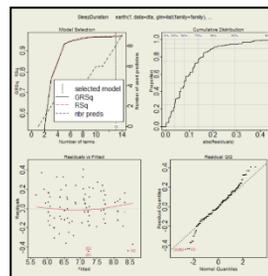
다중 적응형 회귀 스플라인 경로 분석 > 일반화 선형 모델 > 다중 적응형 회귀 스플라인



Earth 매개변수		적합성 요약				
종속 변수	values	일반화된 교차 유효성 검증				
SleepDuration		1	.020	6.635	.968	.972
독립 변수	QualityofSleep, PhysicalActivityLevel, StressLevel, BMICategory, SleepDisorder, HeartRate, DailySteps, Systolic_BP, Diastolic_BP	변수 중요도				
가용되는 최대 상호작용 경	1	QualityofSleep9	11.000	74.625	74.252	
교차 검증 중첩	0	StressLevel8	10.000	51.266	50.937	
교차 유효성 검증	1	BMICategoryOverweight	10.000	51.266	50.937	
분포 군	gaussian	PhysicalActivityLevel	9.000	39.771	39.465	
최대 기간	25	SleepDisorder	7.000	100.000	100.000	
추정 날짜	Wed Jan 7 16:44:11 2026	DailySteps	4.000	7.268	8.047	
완전한 케이스 수	374	SleepDisorderSleep_Apnea	2.000	4.231	4.922	
예측 데이터 세트	...	종속 변수 SleepDuration	1.000	3.241	3.657	
예측 임계값	0.5					
저장된 모델 파일	...					



[부분 종속 플롯]



[모델 진단 플롯]

# New IBM SPSS Statistics Ver.31

카이제곱 검정은 두 범주형 변수 간에 유의미한 연관성이 있는지 확인하는 통계적 가설 검정 방법입니다. 기존에는 교차분석 등의 하위 메뉴로 실행해야 했으나, Ver.31에서는 “분석 > 기술통계” 메뉴 아래에 ChiSquare 실행 옵션이 추가되어 접근성이 향상되었습니다.

## Chi-Square Test

**케이스 처리 요약**

	유효		카이제곱 결측		전체	
	N	퍼센트	N	퍼센트	N	퍼센트
QualityOfSleep * StressLevel	374	100.0%	0	0.0%	374	100.0%

**카이제곱 검정**

	값	자유도	근사 유의확률 (양측검정)
Pearson 카이제곱	847.685 <sup>a</sup>	25	<.001
우도비	774.693	25	<.001
선형 대 선형결합	301.293	1	<.001
유효 케이스 수	374		

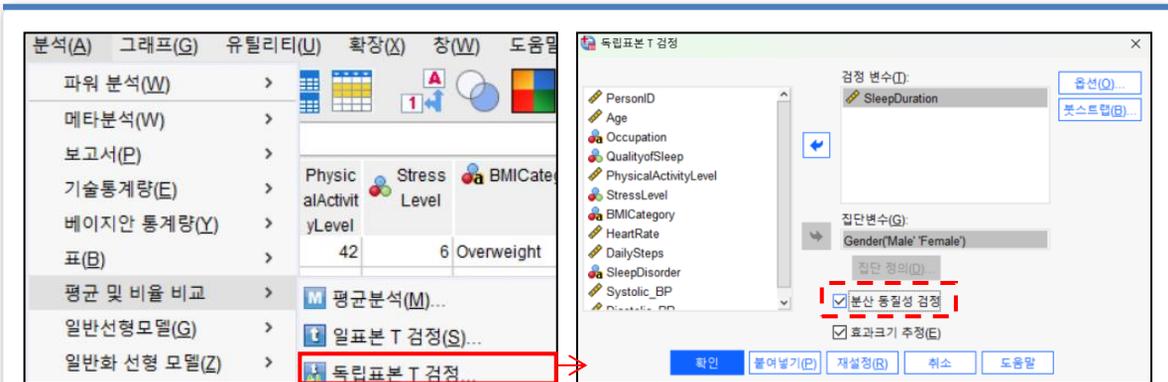
<sup>a</sup> 12 셀 (33.3%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 .61입니다.

◀ 교차분석 없이도 카이제곱 검정 결과 제공

# New IBM SPSS Statistics Ver.31

T-검정은 두 집단 간의 평균 차이가 통계적으로 유의미한지 비교하는 분석 기법입니다. Ver.31에서는 등분산 가정 결과를 선택적으로 출력할 수 있는 체크박스 옵션이 추가되었으며, 결과표에 대해 자동 해석 가이드가 추가되었습니다.

## ☑ T - Test 옵션 기능 추가



### ▼ 기본 분석

집단통계량					독립표본 효과크기					
	Gender	N	평균	표준편차	평균의 표준오차	Standardizer <sup>a</sup>	포인트 추정값	95% 신뢰구간 하한	상한	
SleepDuration	Male	189	7.037	.6932	.0504	Cohen's d	.7908	-.244	-.448	-.041
	Female	185	7.230	.8794	.0647	Hedges 수정	.7924	-.244	-.447	-.041
						Glass 델타	.8794	-.220	-.423	-.015

### ▼ 자동 해석 가이드 제공

평균의 동일성에 대한 T 검정									
	유의확률	t	자유도	유의확률		평균차이	표준오차 차이	차이의 95% 신뢰구간	
				단측 확률	양측 확률			하한	상한
SleepDuration	등분산을 가정함	-2.362	372	.009	.019	-.1932	.0818	-.3540	-.0324
	등분산을 가정하지 않음	-2.357	349.383	.009	.019	-.1932	.0820	-.3545	-.0320

독립 샘플 테스트: 단측 확률

등분산을 가정함

■ 평균 차이가 (그룹화 변수를 기준으로) 변수에 대해 유의적임: SleepDuration

등분산을 가정하지 않음

■ 평균 차이가 (그룹화 변수를 기준으로) 변수에 대해 유의적임: SleepDuration

독립 샘플 테스트: 양측 확률

등분산을 가정함

■ 평균 차이가 (그룹화 변수를 기준으로) 변수에 대해 유의적임: SleepDuration

등분산을 가정하지 않음

■ 평균 차이가 (그룹화 변수를 기준으로) 변수에 대해 유의적임: SleepDuration

참고: 류래이션된 도출값은 형식이 지정된 값이 아닌 실제 셀 값을 기준으로 계산됩니다.

# New IBM SPSS Statistics Ver.31

IBM의 AI 플랫폼 watsonx.ai를 기반으로 작동하는 LLM을 활용하여, 사용자가 선택한 분석 결과에 대해 설명을 들을 수 있는 결과 해석 지원 기능입니다.

## AI Output Assistant 기능 추가

사용 가능한 대형 언어 모델 목록에서 하나의 모델을 선택하십시오. 다국어 또는 자모 이미지 지원을 기본으로 모델 목록을 필터링하거나, 사용 가능한 모든 모델 중에서 선택할 수 있습니다.

모델 선택 필드:  
 다국어  모든 모델 표시  
 모델

모델 이름 | 설명  
 meta-llama/llama-3-2-11b-vision-instruct | The Meta Llama 3.3 multilingual large language model (LLM) is a pretrained and instruction tuned generative model in 70B (text in text out). The Llama 3.3 instruction tuned text only model is optimized for multilingual dialogue use cases and outperform many of the available open source and closed chat models on common industry benchmarks.  
 meta-llama/llama-3-70b-instruct  
 meta-llama/llama-3-405b-instruct  
 meta-llama/llama-4-maverick-17b-128e-instruct-4p8  
 mistral-large-2512  
 mistral/mistral-medium-2505  
 mistral/mistral-small-3-1-24b-instruct-2503  
 openai/gpt-oss-120b

확인 취소 도움말

출력에 대해 선택한 AI 모델에 문의:  
**전문적인 통계 해석**  
**전문적인 통계 해석**  
 비즈니스 이해 당사자를 위한 사업기획서 요약  
 학생을 위한 교육 요약  
 데이터 품질 및 모델 견고성에 대한 진단 검토  
 프롬프트 수동 입력

시 프롬프트 보기 또는 편집

시 프롬프트 텍스트  
 저는 IBM SPSS Statistics의 출력을 분석하는 통계학자입니다. 다음은 SPSS 보고서 출력(예: 회귀, 상관관계 또는 기술 통계)에서 추출한 텍스트입니다.  
 이 결과를 해석하고 다음과 같은 의미 있는 인사이트를 추출하는 데 도움을 주십시오.  
 - 계수, p값, R제곱, 상관관계 또는 검정 통계에 대한 해석.  
 - 데이터에서 발견된 추세나 주목할 만한 패턴.  
 - 추이가 필요할 수 있는 문제(예: 다중공선성, 누락된 값, 중요하지 않은 예측 변수).  
 - 추가 분석이나 개선에 대한 제안사항.

▼ '출력' 버튼을 누르면 미리 보기에 있는 결과가 출력 결과로 제공됨

	Standardized <sup>a</sup>	표준화 추정값	허한	95% 신뢰구간
SleepDuration	Cohen's d	7908	-.244	-.448 .-041
Hedges g	Hedges g	7924	-.244	-.447 .-041
Glass 델타	Glass 델타	8794	-.220	-.423 .-015

a. 효과크기를 추정하는 데 사용되는 분포입니다.  
 Cohen의 d에는 통합 표준 편차와 수정 오차가 사용됩니다.  
 Hedges g에는 통합 표준 편차와 수정 오차가 사용됩니다.  
 Glass 델타에는 대조군(즉, 두 번째 그룹)의 표준 편차가 사용됩니다.

효과크기 추정 방법  
 효과크기를 추정하는 데 사용되는 분포는 다음과 같습니다.  
 \* Cohen의 d: 통합 표준 편차  
 \* Hedges g: 통합 표준 편차와 수정 오차  
 \* Glass 델타: 대조군(즉, 두 번째 그룹)의 표준 편차

IBM SPSS Statistics 사용법  
 IBM SPSS Statistics는 사용자 지정 독립표본 효과크기 출력을 수행하는 다양한 기능을 제공합니다.  
 1. 데이터 출력: 데이터를 필터링하고, 변수를 선택합니다.  
 2. 출력에 대한 사용자 지정 프롬프트를 사용하여 결과를 생성합니다.  
 3. 프롬프트를 사용하여 결과를 생성합니다.  
 4. 프롬프트를 사용하여 결과를 생성합니다.  
 5. 프롬프트를 사용하여 결과를 생성합니다.

# New IBM SPSS Statistics Ver.31

Curated Help란 통계 분석 결과표를 자동으로 분석하여 결과 해석을 텍스트 요약으로 제공하는 기능입니다. Ver.31에서는 T-Test(일표본, 독립, 대응표본), 상관분석(이변량, 편상관, 정준상관, 거리상관), 선형 회귀 내 상관분석이 해석 지원됩니다.

## Curated Help 확장

The screenshot illustrates the 'Curated Help' feature in IBM SPSS Statistics. It shows the 'Statistical Test' dialog box with the 'Statistical Test' (통계항(S)) option selected. This option is linked to a table of statistical results. Below the table are three columns of correlation analysis results: '상관관계 0차', '상관관계 편상관', and '상관관계 부분상관'.

모델	비표준화 계수		표준화 계수		t	유의확률	상관계수		
	B	표준오차	베타	베타			0차	편상관	부분상관
1 (상수)	5.985	.245			24.404	<.001			
Age	.032	.004	.348	7.147	<.001	.345	.348	.348	
DailySteps	-2.937E-5	.000	-.060	-1.226	.221	-.040	-.064	-.060	

a. 종속변수: SleepDuration

**상관관계 0차**

- 매우 긍정적: (없음)
- 긍정적: (Age <—> 모델 1)
- 선형 상관관계 없음: (없음)
- 부정적: (DailySteps <—> 모델 1)
- 매우 부정적: (없음)

**상관관계 편상관**

- 매우 긍정적: (없음)
- 긍정적: (Age <—> 모델 1)
- 선형 상관관계 없음: (없음)
- 부정적: (DailySteps <—> 모델 1)
- 매우 부정적: (없음)

**상관관계 부분상관**

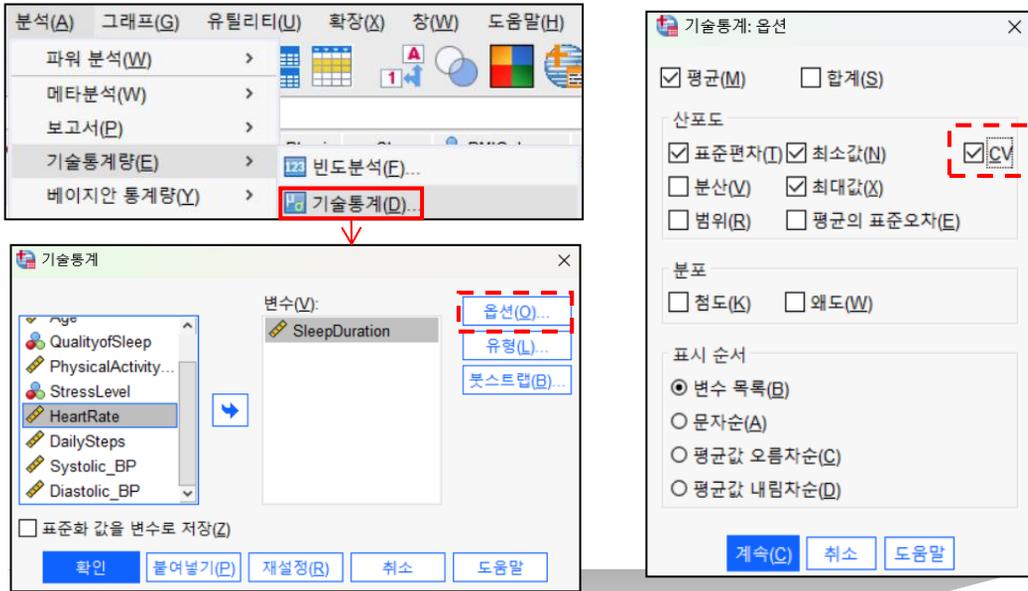
- 매우 긍정적: (없음)
- 긍정적: (Age <—> 모델 1)
- 선형 상관관계 없음: (없음)
- 부정적: (DailySteps <—> 모델 1)
- 매우 부정적: (없음)

참고: 큐레이션된 도움말은 형식이 지정된 값이 아닌, 실제 셀 값을 기준으로 계산됩니다.

# New IBM SPSS Statistics Ver.31

변동 계수는 데이터의 단위가 달라도 산포도를 비교할 수 있도록, 표준편차를 평균으로 나눈 상대적인 변동성 지표입니다. Ver.31에서는 변동 계수를 백분율(%)로 자동 계산해 주는 옵션이 추가되었습니다.

## ☑ Coefficient of Variation 기능 추가



기술통계량			
	N	최소값	최대값
SleepDuration	374	5.8	8.5
유효 N(목록별)	374		

평균	표준편차	CV
7.132	.7957	11.156

# New IBM SPSS Statistics Ver.31

Design of Experiments (실험 계획법, DOE)는 독립 변수가 종속 변수에 미치는 영향을 효율적으로 파악하기 위해 실험을 계획하고 분석하는 방법론입니다.

Ver.31에서는 검색 애플리케이션에 'DOE'를 입력하면 ANOVA, GLM 등 흩어져 있는 관련 분석 대화 상자를 자동으로 찾아 목록으로 제공하는 검색 기능이 지원됩니다.

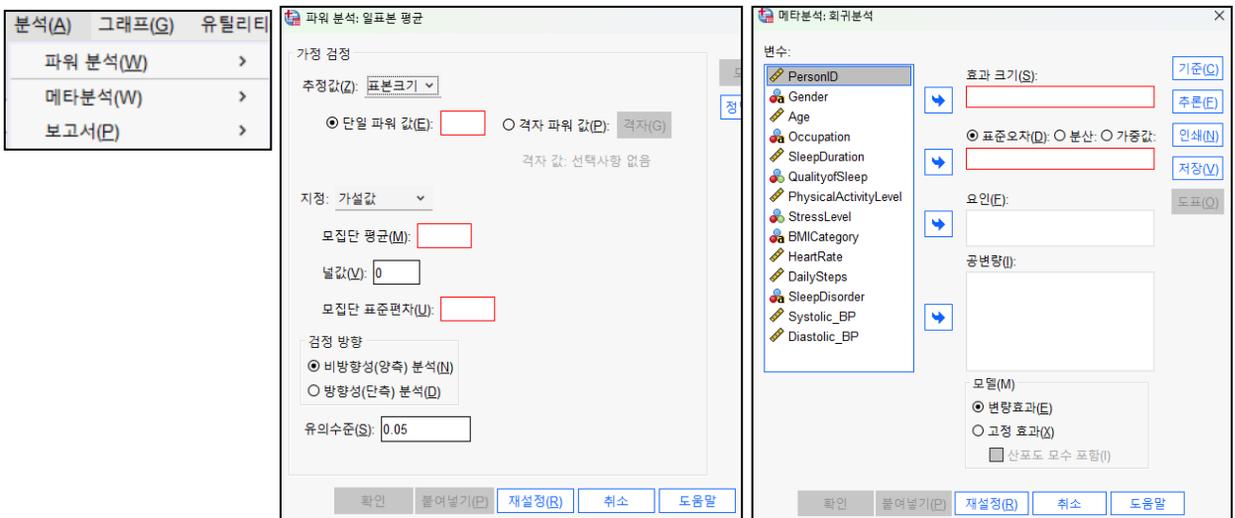
## ☑ DOE 검색 기능 추가

유형	내용
메뉴 대화 상자	분석(A) → 파워 분석(W) → 평균(M) → 일원배치 분산분석(A)
	분석(A) → 베이지안 통계량(Y) → 일원배치 분산분석(W)
	분석(A) → 베이지안 통계량(Y) → 일원배치 반복측도 분산분석(T)
	분석(A) → 평균 및 비율 비교 → 일원배치 분산분석(O)
	분석(A) → 일반선형모델(G) → 일반량(U)
	분석(A) → 일반선형모델(G) → 다변량(M)
	분석(A) → 일반선형모델(G) → 반복측도(R)
	분석(A) → 일반선형모델(G) → 분산성분(V)
	분석(A) → 일반화 선형 모델(Z) → 일반화 선형 모델(G)
	분석(A) → 일반화 선형 모델(Z) → 일반화 추정 방정식(E)
	분석(A) → 혼합모델(X) → 선형(L)
	분석(A) → 로그선형분석(O) → 일반(G)
	분석(A) → 로그선형분석(O) → 로짓(L)
	분석(A) → 로그선형분석(O) → 모델 선택(M)
확장(X)	확장(X) → 확장 허브(H)... (Propensity Score Matching)
도움말	DOE에 대한 온라인 도움말 검색
SPSS Community	IBM SPSS 커뮤니티에서 DOE 검색

# New IBM SPSS Statistics Ver.31

Ver.31에서는 파워 분석, 메타분석, 보고서 관련 모든 프로시저에서 필수 입력 항목이 빨간색으로 강조되어, 분석 실행을 위해 필요한 입력을 즉각적으로 인지할 수 있습니다.

## 필수 입력 강조 표시 추가



## Java and Python Upgrades

- Ver.31에서는 연동 프로그램인 Java와 Python이 상위 버전으로 업데이트 되었습니다.



Java 17.0.11 ➔ Java 17.0.13



Python 3.10.4 ➔ Python 3.13.1

# KoreaPlus Statistics

- Embedded on IBM SPSS Statistics Professional



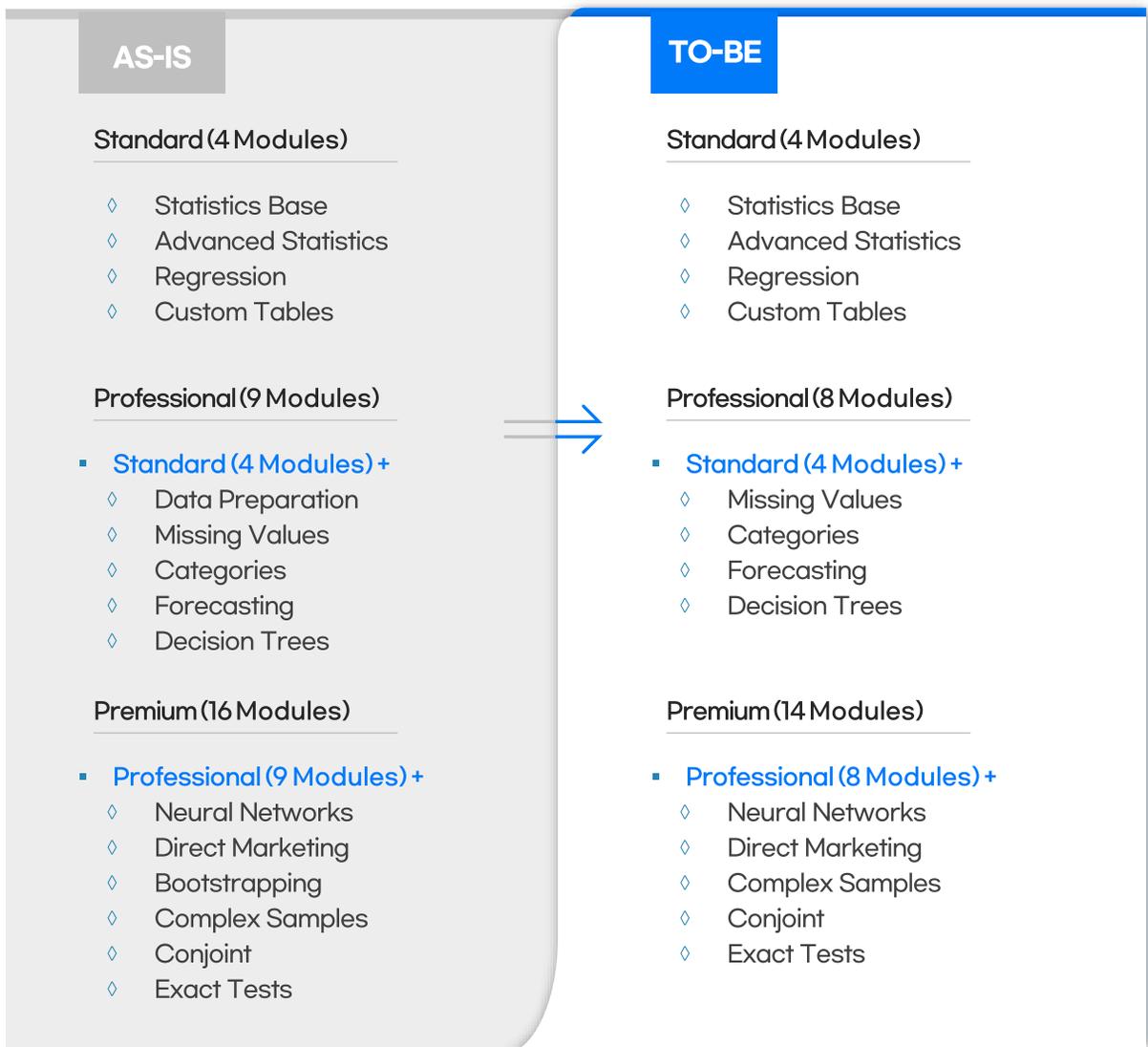
## | Packages Professional

- 패키지 구성 (Standard, Professional, Premium)
- Professional 구성 모듈
- 모듈 별 분석 기법 상세

## 패키지 구성

KoreaPlus Statistics Ver.31은 14가지 모듈의 조합으로 Packages를 구성합니다. 각각의 모듈은 전문적인 분석 기법을 포함 합니다. Statistics Base 모듈에서 과거 Data Preparation 모듈과 Bootstrapping 모듈의 분석 기능을 지원 합니다(Ver.27부터). 과거 버전의 Professional을 보유한 고객은 업그레이드를 통해 최소의 비용으로 최신버전의 기능을 모두 활용할 수 있습니다.

### ✓ KoreaPlus Statistics Packages



\* Amos : 1 Linked Program

# Professional 구성 모듈

데이터 분석을 수행하기 위해 기본적으로 필요한 기본 패키지를 포함하여 데이터 액세스에서 데이터 준비, 분석, 전개까지 분석의 전 과정을 수행할 수 있습니다. 복잡한 관계가 있는 고급 통계 분석과 전문적인 분석까지 가능하여 비즈니스 목표에 대해 최고의 결과를 얻을 수 있도록 도와줍니다. Professional 패키지는 IBM SPSS Statistics의 기능을 최대한 활용하여 더욱 풍부하고 심층적인 분석을 가능하게 합니다.

## ✓ Packages Professional 구성 모듈과 특징



### Professional 구성 모듈



Statistics Base  
Advanced Statistics  
Regression  
Custom Table  
Forecasting  
Missing Values  
Categories  
Decision Trees  
& KoreaPlus

### 특징 |

- 다양한 회귀 및 고급 통계 프로시저를 제공
- 의사결정 나무로 그룹을 쉽게 식별하고 그룹 간의 관계를 발견하여 예측
- 데이터 유효성 및 누락된 값을 통계적으로 확실한 결과를 얻을 수 있도록 지원
- 자동 모형 생성 기능을 이용하여 초보자들도 쉽게 최적의 시계열 모형 도출 가능 그룹 간의 관계를 파악하여 미래의 현상을 예측
- 지리적 분석을 사용하여 위치 및 시간 데이터를 통합, 탐색 및 모델링
- 계층 구조의 데이터, 생존 데이터 등 다양한 형태의 데이터를 이용하여 심층적으로 정교한 분석 가능

### 활용 산업군 |

- 의학, 제조, 제약 분야 : 생존 데이터 분석
- 제조 : 제품 수요, 판매량, 생산량 예측
- 공공 : 세금, 교통량 예측
- 마케팅 : 광고 시청률 예측
- 기타 : 다양한 산업 군에서 예측 업무

# 모듈 별 분석 기법 상세

## I Statistics Base

다양한 통계 기법, 데이터 관리 기능, 그리고 그래픽 옵션을 제공하여 연구자와 분석가들이 데이터를 효율적으로 분석하고 해석할 수 있도록 도와줍니다. 직관적인 사용자 인터페이스와 풍부한 기능 덕분에 초보자부터 전문가까지 쉽게 사용할 수 있습니다. 주로 사회과학, 건강 연구, 마케팅 등의 분야에서 널리 활용됩니다.

### ✓ 분석기법

- 보고서
  - OLAP 큐브
  - 케이스요약
  - 코드북
  - 행별 요약보고서
  - 열별 요약보고서
- 빈도분석
- 기술통계
- 교차분석표
- 평균
  - 합계, 중위수
  - 최소값, 최대값
  - 평균의 표준오차
- 데이터탐색
  - T검정
  - 일원분산분석
  - 일원배치 반복측도
  - 분산분석
  - GLM일변량분석
  - 이변량상관계수
  - 편상관계수
  - 거리측도(상이성측도)
  - 선형모형
  - 선형회귀분석
  - 순서회귀분석
  - 최근접이웃분석
- 판별분석
- 요인분석
- 군집화프로시저선택
- 이단계군집분석
- 계층적 군집분석
- K-평균 군집분석
- 비모수 검정
- 다중응답분석
- 신뢰성분석
- 다차원척도법
- 비율통계
- ROC곡선, ROC분석
- 시뮬레이션
- 지리공간모형화
  - 검증 규칙
  - 데이터 검증
  - 타당성 검사 (Validation Data)
  - 자동 데이터 준비 (Automated Data Preparation)
  - 특수 케이스 식별
  - 최적화 구간화
  - 케이스 대조 매칭
  - Bootstrapping
  - 기술통계, 모델링

### KEYPOINT

- 변수 이름, 변수 및 값 레이블, 결측값과 같은 사전 정보와 모든 지정된 변수의 요약통계 및 다중 응답 세트를 보고 합니다.
- 여러 종류의 변수를 설명하는데 유용한 그래픽 표시와 통계를 사용할 수 있고 이원배치표와 다원배치표를 만들어 다양한 검정과 연관 측도를 사용할 수 있습니다.
- 이외에 다중응답분석, 비모수검정, 신뢰성분석 등을 통해 전문적인 결과를 도출할 수 있습니다.

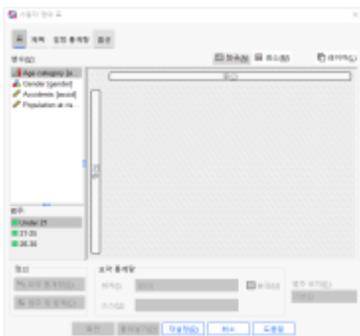
# 모듈 별 분석 기법 상세

## I Custom Tables

데이터를 요약하고 보고하는 데 최적화된 모듈입니다. 복잡한 데이터를 직관적으로 시각화할 수 있는 사용자 정의 테이블을 생성할 수 있게 해줍니다. 연구자와 분석가들은 Custom Tables을 통하여 다양한 변수와 차원을 조합해 세부적인 통계 결과를 도출하고, 보고서 작성 시 가독성과 이해도를 높일 수 있습니다. 주로 시장 조사, 여론 조사, 건강 연구 등에서 유용하게 사용됩니다.

### ✓ 분석기법

- 테이블 작성기 인터페이스
- 범주형 변수에 대한 단순 통계표
- 범주형 변수를 사용한 누적 중첩 및 레이어
- 범주형 변수에 대한 총계 및 소계
- 범주형 변수의 계산된 범주
- 공유범주가 있는 변수 테이블
- 요약 통계
- 척도변수 요약
- 신뢰구간
- 검정통계량
- 다중 응답 세트
- 결측값
- 표 형식 지정 및 사용자 정의
- 표본파일



[끌어다 놓기 테이블 빌드]



[그래픽 사용자 인터페이스]



[테이블 옵션 설정]

## KEYPOINT

- 변수의 단수 개수부터 척도변수의 평균 오차와 같은 산포 측정까지 모든 것을 포함할 수 있습니다.
- 세 종류의 유의성 검증을 사용하여 행 변수와 열 변수 간의 관계를 연구할 수 있습니다.

# 모듈 별 분석 기법 상세

## I Regression

회귀 분석을 통해 변수 간의 관계를 모델링하고 예측하는 강력한 모듈입니다. Regression은 선형 회귀, 로지스틱 회귀, 비선형 회귀 등 다양한 회귀 분석 기법을 제공하여 복잡한 데이터 패턴을 이해하고 예측할 수 있게 합니다. 연구자와 분석가는 이를 통해 중요한 예측 변수와 결과 간의 관계를 밝혀내고, 의사결정에 필요한 인사이트를 도출할 수 있습니다. 주로 경제학, 사회과학, 생물통계학 등 여러 분야에서 널리 활용됩니다.

### ✓ 분석기법

- 이분형 로지스틱 회귀모형
- 로지스틱 회귀분석
- 다항 로지스틱 회귀분석
- 프로빗 회귀분석
- 4분위수
- 비제약된 비선형 회귀분석
- 제약된 비선형 회귀분석
- 가중 추정
- 가중치 최소 제곱법
- 2-단계 최소제곱 회귀분석
- 커널리지 회귀분석
- 매개변수 가속 장애 시간 모델

## KEYPOINT

### 범주형 결과 예측

- 다수의 예측 요인의 경우 Score 및 Wald 방식을 사용하여 결과를 빠르게 얻을 수 있습니다.
- AIC(Akaike information criterion) 및 BIC(Bayesian information criterion)를 사용하여 모델 적합도를 평가합니다.

### 비선형 모델의 매개변수 추정

- CNLR을 사용하여 매개변수의 모든 조합에서 선형 및 비선형 제한조건을 사용합니다.
- 평활 손실 함수(목적 함수)를 최소화하여 매개변수를 추정하고 매개변수 표준 오류 및 상관관계(correlation)의 부트스트랩 추정을 계산합니다.

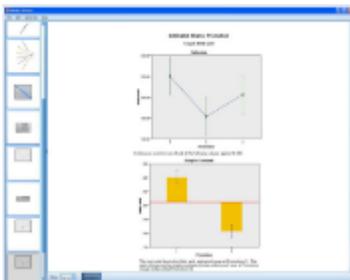
# 모듈 별 분석 기법 상세

## Advanced Statistics

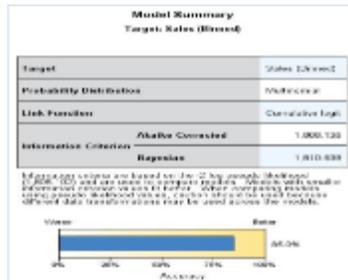
Advanced는 다변량 분석, 혼합 모델, 서바이벌 분석 등 고급 통계 기법을 제공하여 복잡한 데이터 구조를 심층적으로 분석할 수 있게 합니다. 연구자와 분석가는 이를 통해 정교한 모델을 구축하고, 데이터의 숨겨진 패턴을 발견하며, 보다 정확한 예측을 할 수 있습니다. 주로 사회과학, 생물통계학, 경제학 등 다양한 연구 분야에서 활용됩니다.

### ✓ 분석기법

- GLM 다변량분석
- GLM 반복 측도
- 분산 성분 추정 분석
- 선형 혼합 모형
- 반복 공분산 유형
- 일반화 선형 모형
- 일반화 추정 방정식
- 일반화 선형 혼합 모형
- 모형선택 로그선형 분석
- 로그선형분석
- 일반화 로그선형분석
- 계층적 로그선형분석
- 로직 로그선형 분석
- 생명테이블
- Kaplan-Meier 생존분석
- Cox 회귀분석 분석
- 시간-종속 공변량 계산
- 생존분석
- 고정효과 ANOVA
- 공분산분석(ANCOVA)
- 다변량 분산분석
- 다변량 공분산분석
- Bayesian 통계량



[선형혼합모델 평균추정치]



[일반화된 선형혼합모델]



[매개변수 추정치 테이블]

## KEYPOINT

- GLM 모형을 제공하며 GLM의 확장인 일반화 추정 방정식을 활용할 수 있습니다.
- 로그선형 분석을 통한 모형을 범주형 종속변수와 예측자 사이의 관계를 분석하기 위한 로그선형 모형, 총화 변수의 수준별로 개별 분석을 생성하는 Kaplan-Meier 생존분석, 공변량 값을 지정으로 지정 이벤트에 대한 시간을 모형화하는 Cox 회귀 분석 기법을 제공합니다.

# 모듈 별 분석 기법 상세

## I Forecasting

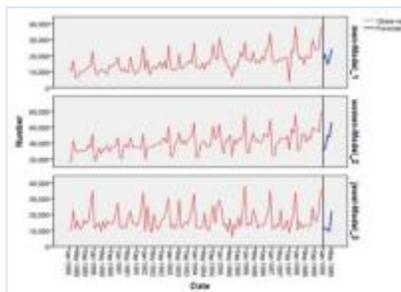
시계열 데이터 분석과 예측을 위한 모듈입니다. Forecasting은 ARIMA, 이중지수평활법 등 다양한 예측 모델을 제공하여 과거 데이터를 바탕으로 미래 동향을 예측할 수 있게 합니다. 분석가와 연구자들은 이를 통해 시장 동향, 판매 예측, 수요 계획 등을 효과적으로 수행할 수 있습니다. 경제학, 금융, 공급망 관리 등 다양한 분야에서 널리 사용됩니다.

### ✓ 분석기법

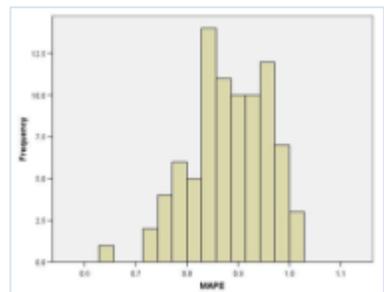
- ARIMA
- 정확한 최대우도법
- 자기회귀 이동평균
- 반적 혹은 제한적인 모형
- Intervention 모형 분석
- 단순, 승법, 가법, Holt 모형
- 시계열 모델러
- 지수평활법
- 계절분해
- 승법모형, 가법모형
- 스펙트럼 도표
- 시간 인과 모형
- 시간 인과 모형 적용
- 정확도 측도
- 이상값 유형
- ACF/PACF 도표
- GARCH 모형



[시계열 일반모델]



[예측차트]



[평균절대비율오차]

### KEYPOINT

- 지수평활, ARIMA 및 다변량 ARIMA등 모형을 추정하고 예측하며, 주기동작을 식별합니다.
- 시계열 예측의 예로는 콜 센터에 매일 필요한 직원 수 예측, 특정 제품 또는 서비스에 대한 수요 예측 등이 있습니다.

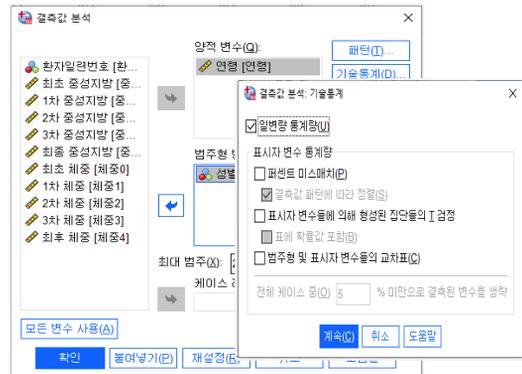
# 모듈 별 분석 기법 상세

## I Missing Values

데이터 분석에서 결측값을 처리하는 데 유용한 모듈입니다. Missing Values는 결측값 패턴을 식별하고, 다양한 방법으로 이를 대체하거나 제거할 수 있는 기능을 제공합니다. 분석가들은 이를 통해 데이터의 완전성을 유지하고, 분석 결과의 신뢰성을 높일 수 있습니다. 주로 사회과학, 건강 연구, 시장 조사 등 결측값이 자주 발생하는 분야에서 활용됩니다.

### ✓ 분석기법

- 결측값 분석
- 다중대체 (Multiple Imputation)
- 평균 추정 혹은 회귀에서 결측값 대체
- 결측 데이터에 대한 다수의 결측값 대체
- Listwise and Pairwise 통계



## KEYPOINT

### 누락 데이터 대체 문제점을 빠르게 진단

- 6개의 진단 보고서를 사용하여 다양한 각도에서 데이터를 검토합니다.
- 데이터에 대한 사례별 개요를 제공하는 데이터 패턴 보고서를 통해 누락 데이터를 진단합니다.
- 누락 데이터의 범위와 각 사례에 대해 극한 값을 판별합니다.
- 데이터 세트에서 누락 패턴을 파악하고 누락 값을 타당한 추정치로 대체합니다.
- 데이터의 특성을 기준으로 가장 적합한 방식을 선택하는 자동 대체 모델을 활용하거나 대체 모델을 사용자 정의합니다.
- 각 데이터 세트에 대해 매개변수를 생성하도록 선형 회귀 또는 예측 최대 알고리즘과 같은 기법을 사용하여 작성된 개별 데이터 세트를 모델링합니다.
- 대체 내 및 대체 사이에서 변동을 고려한 추론 통계를 계산하고 추정치를 풀링하여 최종 매개변수 추정치를 획득합니다.

# 모듈 별 분석 기법 상세

## I Categories

범주형 데이터 분석을 위한 강력한 모듈입니다. Categories는 대응 분석, 다차원 척도법, 최적 스케일링 등 다양한 기법을 제공하여 범주형 변수 간의 관계를 심층적으로 분석할 수 있습니다. 이를 통해 연구자와 분석가는 데이터의 구조를 이해하고 시각적으로 표현할 수 있으며, 의사결정에 유용한 인사이트를 얻을 수 있습니다. 주로 마케팅, 사회과학, 행동 연구 등에서 활용됩니다.

### ✓ 분석기법

- 범주형 회귀분석(CATREG)
- 범주형 주성분분석(CATPCA)
- 비선형 정준 상관 분석(OVERALS)
- 척도화 분석
- 다차원척도법(PROXSCAL)
- 다차원확장(PREFSCAL)
- 신뢰도 분석 (Fleiss Kappa)



## KEYPOINT

### 범주형 데이터에서 결과를 예측하고 관계를 발견

- 다변량 데이터와 해당 관계를 더욱 완벽하게 해석하고 쉽게 분석합니다.
- 범주형 데이터에 대한 추가 통계 조작을 수행하여 질적 변수를 양적 변수로 변환합니다.
- 시장 부문, 의료 진단, 정당 또는 생물학 종 등 연구하는 모든 유형의 범주에서 기반 관계를 그래픽으로 표시합니다.

# 모듈 별 분석 기법 상세

## I Decision Trees

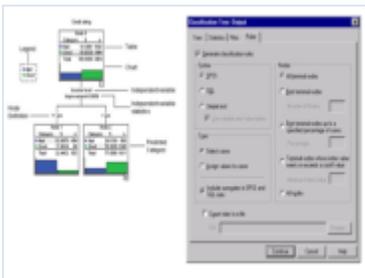
의사결정나무 분석을 통해 데이터 분류와 예측을 돕는 모듈입니다. Decision Trees는 CHAID, CART, C5.0 등 다양한 의사결정나무 알고리즘을 제공하여 복잡한 데이터의 패턴을 시각적으로 이해하고 예측 모델을 구축할 수 있습니다. 연구자와 분석가는 이를 통해 주요 예측 변수와 결과 간의 관계를 명확하게 파악하고, 데이터 기반 의사결정을 개선할 수 있습니다. 주로 마케팅, 금융, 의료 등 다양한 분야에서 활용됩니다.

### ✓ 분석기법

- 의사결정나무 작성
- 나무 편집기
- C&RT
- CHAID
- Exhaustive CHAID
- QUEST



[대화 상자 열기]



[트리 기반 분류모델]

**Risk**

Estimate	Std. Err.
277	.000

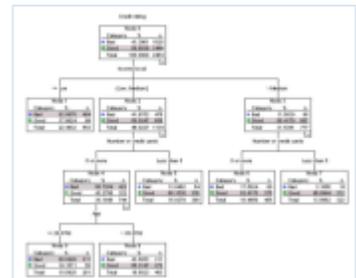
Growing Method: CHAID  
Dependent Variable: Credit rating

**Classification**

Observed	Predicted		Percent Correct
	Bad	Good	
Bad	885	365	66.2%
Good	142	1295	89.7%
Overall Percentage	30.0%	70.0%	78.0%

Growing Method: CHAID  
Dependent Variable: Credit rating

[위험 및 분류 테이블]



[CHAID 방법]

## KEYPOINT

- 독립 변수값을 기반으로 종속 변수값을 예측하거나 케이스를 집단으로 분류합니다.
- 이 모듈은 탐색 및 확인 분류분석을 위한 검증 도구를 제공합니다.

# KoreaPlus Statistics

- Embedded on IBM SPSS Statistics Professional



## | Add-On KoreaPlus Statistics Modules

- KoreaPlus의 Add-On 기능
- AI Coach
- KoreaPlus의 분석 기능 상세
- KoreaPlus Statistics Utility 1.0

# KoreaPlus의 Add-On 기능

## KoreaPlus Modules

### 기존 분석 기능

- ◆ AI RNN기반 시계열 학습/예측
- ◆ Text Analytics
  - Word2VecCluste
  - Sent2VecCluser
- ◆ Meta-Analysis
- ◆ PSM Propensity Score Matching
- ◆ Medical Analysis 생존관련 임상 시험 분석
- ◆ Scale 2모수로지스틱/일반화부분점수모형  
로지스틱 회귀/ 판별분석
- ◆ Conjoint 대화상자
- ◆ Customer Life Value 고객 생애 가치 분석
- ◆ Multivariate Analysis 주성분분석, 정준상관분석
- ◆ Sample Size Calculation 표본크기 계산
- ◆ Weibull Analysis 모수 추정

### 최신 기능

- ◆ Subgroup Analytics
  - + with Cox Regression
  - + with Logistic Regression
  - + with Poisson Regression
- ◆ PSM (기능 추가)
  - + 결측치 처리 (전처리) 기능 추가
- ◆ Meta-Analysis (기능 추가)
  - + Effect Size Data
  - + Various Effect Size Data
  - + 단일집단 사전사후 분석
- ◆ Compare Means
  - + 두 집단 이상 일 때  
독립표본 평균검정, 대응표본 평균 검정
- ◆ AI (네 가지 알고리즘 추가)
  - + RF : Random Forest
  - + GBM : Gradient Boosting Machine
  - + XGB : Extreme Gradient Boosting
  - + LGBM : Light Gradient Boosting Machine
- ◆ Text Analytics (기능 추가)
  - + 동의어 정의 사전 기능 추가
  - + 오류 개선
- ◆ AI Coach

### 편의 기능

**통계포털**  
국가통계  
포털 링크

**한국지도**  
한국지도  
제공

**Data  
Station**  
캘린더, 기상  
데이터 제공

**Tips**  
데이터  
블러오기  
Tips

**아래한글  
내보내기**  
Output  
내보내기

**Edu**  
무료 통계  
강의 제공

# AI Coach

## | KoreaPlus 생성형 AI 서비스, AI Coach

KoreaPlus의 'AI Coach'는 IBM의 생성형 AI 플랫폼인 watsonx.ai와 데이터솔루션의 KoreaPlus Statistics가 결합하여 복잡한 통계 분석 결과를 사용자들이 쉽게 이해할 수 있도록 직관적인 해석을 제공합니다. 이를 통해 비전문가도 통계 분석 결과를 효과적으로 활용할 수 있습니다.

**KoreaPlus  
Statistics**



**IBM  
watsonx.ai**



**KoreaPlus  
AI Coach**

## | AI Coach

복잡한 통계 데이터 분석결과를 자동으로 해석하여 통계 비전문가의 부담을 줄여주고, AI 해석 결과 창에 통계 용어에 대한 정의를 함께 제공해줌으로써 통계적 지식을 향상시킬 수 있습니다. 또한, 추가 질문 리스트를 통해 심도 있는 결과 해석 도출이 가능하며, 내보내기 기능을 통해 보고서 작성 시간을 단축할 수 있습니다.

AI Coach는 전 산업군(학교, 공공, 정부기관, 병원, 기업 등)에서 활용 할 수 있습니다.



강력한 해석 능력  
최신 AI 모델 기반



사용자 친화적  
인터페이스  
직관적인 결과해석



실시간 해석 지원  
빠른 분석 결과 해석



통계 지식 제공  
인사이트 도출

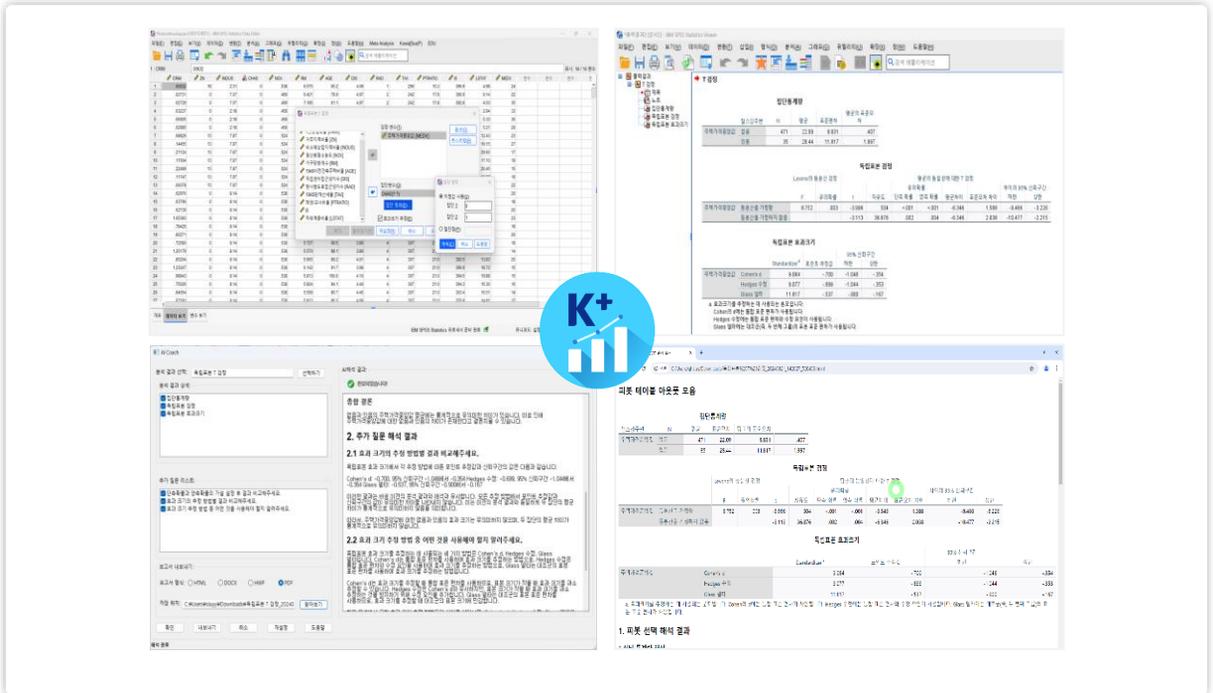


보고서 작성 시간 단축  
결과해석의 시간 절약

# AI Coach

## AI Coach 미리보기

AI Coach는 KoreaPlus Statistics의 Edu 메뉴에서 쉽고 빠르게 구동 할 수 있습니다.



## AI Coach 분석기법

### 분석기법

기술통계량

비모수검정

회귀분석

일반선형모델

기타 주요 분석기능

### 상세 분석 기법

빈도분석, 기술통계, 교차분석

카이제곱, 이항, 1-1표본 K-S, 2-독립표본, K-독립표본, 2-대응표본, K-대응표본

이분형 로지스틱, 다항 로지스틱, 곡선추정, 편최소제곱, 순서형, 기중추정, 2단계최소제곱, 프로빗, 범주형

일변량, 반복측도

추가 예정

# KoreaPlus의 분석 기능 상세

## Subgroup Analysis (1/2)

서브그룹 분석(Subgroup Analysis)은 실험군-대조군 연구에서 특정 약물 및 치료의 처치(Treatment)에 따른 질병의 발생, 사망 등의 임상적 결과를 비교하고자 할 때, 서브그룹 변수(ex. 성별, 나이 등)와 처치 변수 간의 상호작용을 분석하는 통계적 방법입니다. KoreaPlus의 서브그룹 분석은 사용자가 변수를 조정할 수 있는 세분화된 분석 대화상자를 제공하여 분석 설정과 수행이 용이하도록 도와줍니다. Forest Plot을 통해 통계량 추정치와 유의확률을 시각적으로 제시하여 결과를 명확하게 전달합니다.

### 분석 기능 상세

#### Subgroup Analysis

- with Cox Regression
- with Logistic Regression
- with Poisson Regression

### 특장점

특정 환자 서브그룹의  
처치 효과 식별

특정 서브 그룹 간의 처치 효과 차이를 식별하는데 도움을 줍니다.

추정의 정확성 향상

서브그룹 내 이질성을 감소 시키고 통계적 검정력을 높여 처치 효과 추정의 정확성을 높입니다.

새로운 가설의 발견

가설을 생성할 수 있으며, 추가 연구에서 테스트 할 새로운 가설의 발견으로 이어질 수 있습니다.

개별 대상에 대한 맞춤 처치

대상의 교유한 특성과 처치에 대한 반응을 기반으로 정보를 제공하는데 도움을 줍니다.



### 서브 그룹 분석의 활용 사례

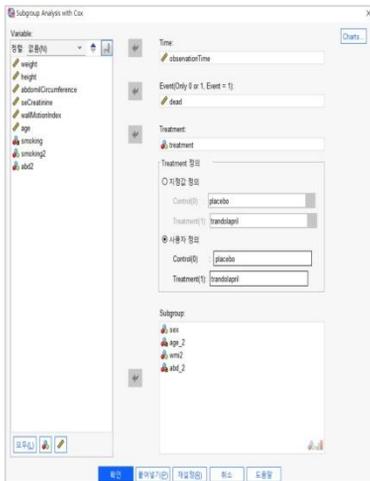
- 의학 분야의 새로운 치료 요법 및 약물에 대한 임상 실험 결과의 분류 대상 간 비교
- 사회과학 및 인문학 분야에서 추상적으로 확인된 내용에 대 통계적 검증 (범죄 재발률 요인 파악 등)

# KoreaPlus의 분석 기능 상세

## Subgroup Analysis (2/2)

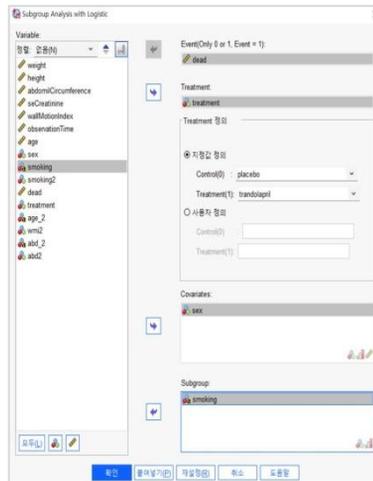
### Cox Regression

특정 하위 그룹 내에서 시간-종속 사건 발생률에 영향을 미치는 변수들의 효과를 평가하는 생존 분석 기법



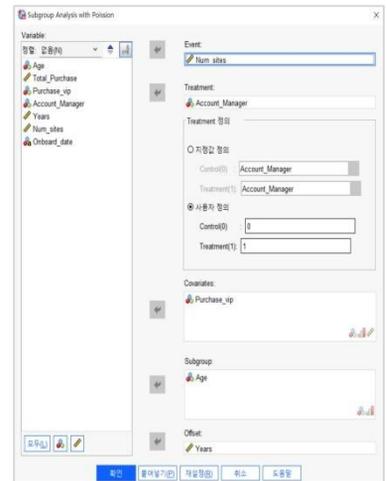
### Logistic Regression

특정 하위 그룹 내에서 이분형 결과 변수에 영향을 미치는 요인들의 효과를 평가하는 통계적 방법

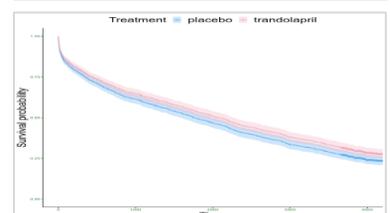
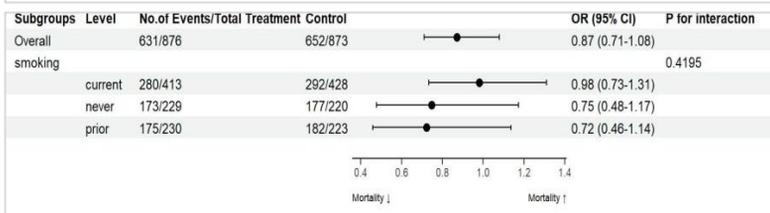
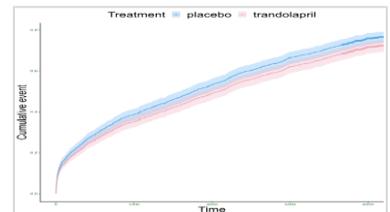
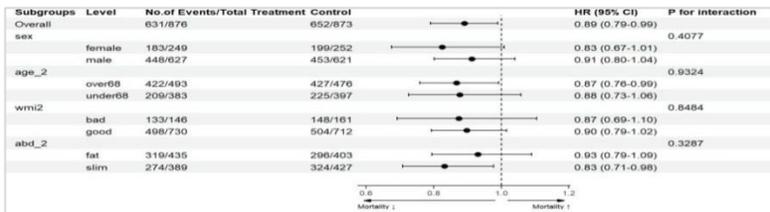


### Poisson Regression

특정 하위 그룹 내에서 사건 발생 빈도에 영향을 미치는 변수들의 효과를 평가하는 통계적 방법



## 분석 결과 (Forest Plot & Chart)



# KoreaPlus의 분석 기능 상세

## PSM (Propensity Score Matching) 기능

KoreaPlus 모듈에서 제공하는 PS Matching 기법에 결측치 데이터를 설정 할 수 있는 대화 상자를 추가 하였습니다. 데이터의 형태나 결측치가 얼마나 데이터에 포함 되어 있는지에 따라 사용자는 결측치 제거 방법을 선택하여 매칭 분석을 수행할 수 있습니다.

### 분석 기능 상세

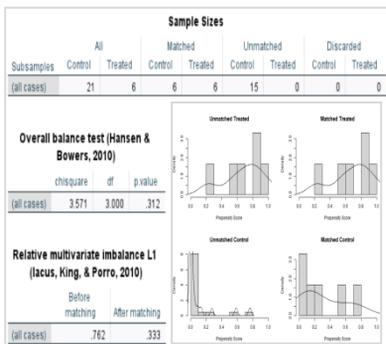
#### Missing value handling

- Row-based removal (행 제거)
- Columns-based removal (열 제거)
- Row/column removal based on missing value ratio by column (행과 열 모두 고려하여 사용자 지정 비율로 제거)

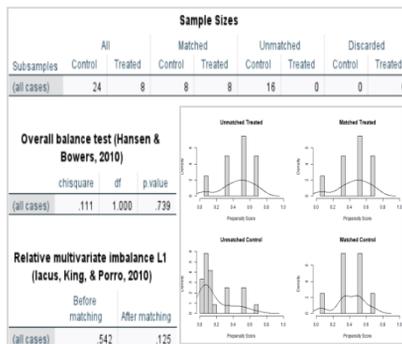
[PS Matching 대화상자]

[Missing value handling]

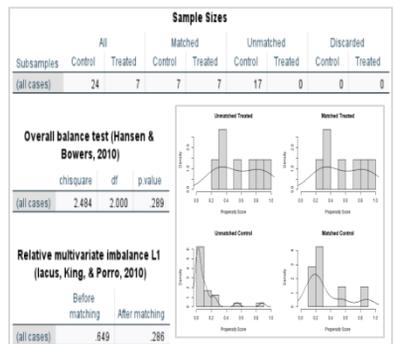
[결측치 데이터 처리 대화상자]



[Row-based removal]



[Columns-based removal]



[Row/column removal based on missing value ratio by column]

# KoreaPlus의 분석 기능 상세

## Meta Analysis 기능 (1/2)

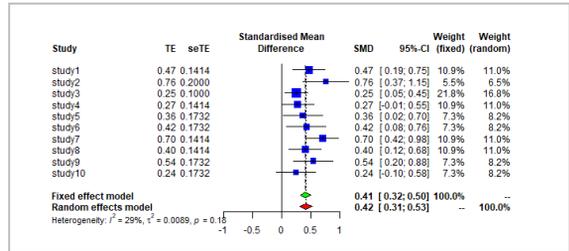
Meta Analysis는 동일한 주제하에 수행된 여러 연구들로부터 도출된 다양한 연구결과를 체계적으로 종합 분석하는 계량적 접근방법입니다. KoreaPlus Ver.31에서는 사전에 계산된 효과크기 데이터로도 메타분석을 수행 할 수 있습니다. 또한 다양한 유형의 데이터가 혼합되어 있는 파일로도 효과크기 분석이 가능하고, 두 집단이 아닌 단일 집단 연속형 데이터로도 분석 할 수 있는 기능을 추가 하였습니다.

### 분석 기능 상세

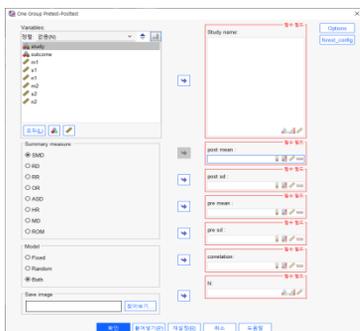
- 효과크기 데이터 (Effect Size Data)
- 다양한 효과크기 데이터 (Various Effect Size Data)
- 단일집단 사전사후 분석



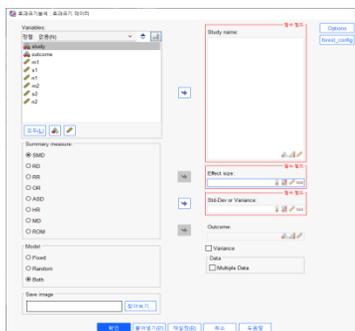
[메타분석 실행]



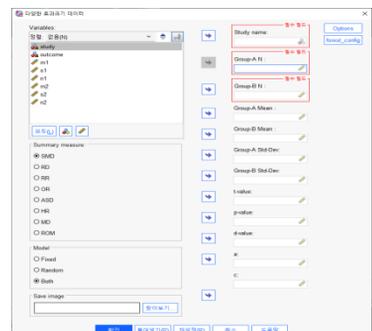
[포리스트 도표]



[단일집단 사전사후 분석 실행 창]



[효과크기 데이터 실행 창]



[다양한 효과크기 데이터 실행 창]

# KoreaPlus의 분석 기능 상세

## Meta Analysis 기능 (2/2)

IBM SPSS Statistics에서 제공되는 메타분석 기법과 KoreaPlus Statistics 모듈에 새롭게 추가 된 기법의 기능 비교표 입니다. (Ver.31 기준)

메타분석 기능 비교				
구분		방법	IBM SPSS	KoreaPlus
추정량	고정모형	Peto	불가능	가능
효과크기	이분형 효과크기	오즈비	불가능	가능
		Arcsine difference	불가능	가능
데이터 유형	연속형 데이터	단일집단 사전사후 검사	불가능	가능
분석방법	다양한 유형의 효과크기를 한가지 유형의 효과크기로 일원화 후 효과크기 분석	평균, 표준편차	불가능	가능
		t값	불가능	가능
		p값	불가능	가능
		효과크기 d	불가능	가능
		f값	불가능	가능
		log odds ratio	불가능	가능

# KoreaPlus의 분석 기능 상세

## Compare Means

KoreaPlus Ver.31에서는 2개 이상의 집단일 때 사용자가 편리하게 집단 간 평균을 비교할 수 있도록 대화상자(메뉴)를 구현하였습니다.

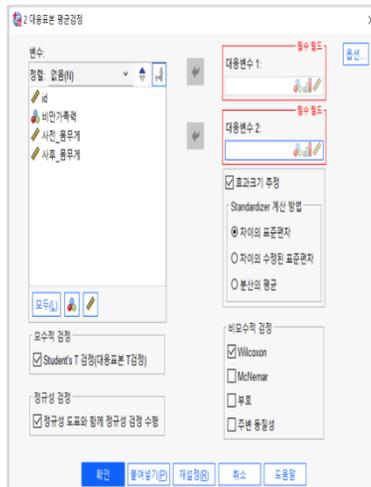
Compare Means (평균검정)는 집단 간 평균을 비교하고 집단 간에 차이가 있는지 없는지 분석하는 방법입니다.

### 분석 기능 상세

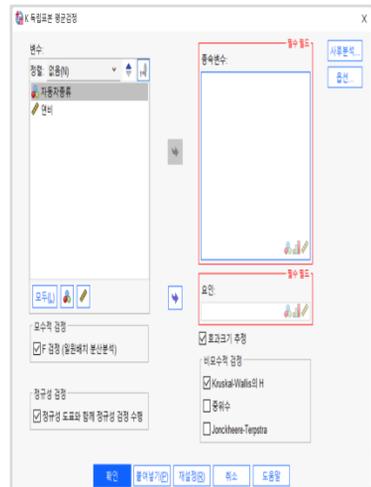
- 2 독립표본 평균검정 : 집단이 2개일 때 집단 간 서로 독립 일 때 시행
- 2 대응표본 평균 검정 : 집단이 2개일 때 독립이 아닌 영향을 받는 전후 처리 집단 일 때 시행
- K 독립표본 평균검정 : 집단이 3개 이상일 때 집단 간 독립 일 때 시행



[2 독립표본 평균검정 대화상자]



[2 대응표본 평균 검정 대화상자]



[K 독립표본 평균검정 대화상자]

		Levene의 등분산 검정			평균의 동일성에 대한 T 검정						
		F	유의확률	t	자유도	유의확률		평균차이	표준오차 차이	차이의 95% 신뢰구간	
						단측 확률	양측 확률			하한	상한
영어시험점수	등분산을 가정함	.034	.854	.534	38	.298	.597	2.188	4.100	-6.113	10.488
	등분산을 가정하지 않음			.534	37.463	.298	.597	2.188	4.100	-6.117	10.492

[2 독립표본 평균검정 - 집단 통계량]

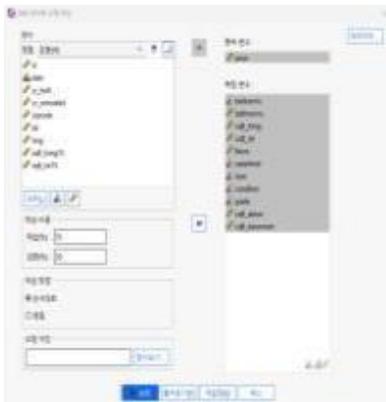
# KoreaPlus의 분석 기능 상세

## AI - 머신러닝 알고리즘

RNN 기반 예측 분석이 가능하던 AI 메뉴에 네 가지 머신러닝 알고리즘을 추가하였습니다. 머신러닝은 인공지능의 한 분야로 기계 스스로 데이터의 패턴 및 추론을 거쳐 동작하는 방법론을 의미합니다. 대량의 데이터에 적용 및 활용할 수 있고 숨겨진 패턴을 파악하는데 유용하며, 사용자가 수집한 데이터의 형태에 맞게 다양한 학습방법과 알고리즘을 유연하게 활용할 수 있습니다.

### 분석 기능 상세

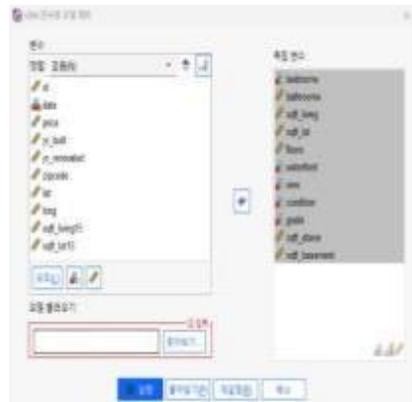
- 랜덤 포레스트(RF : Random Forest)
- 그래디언트 부스팅 머신(GBM : Gradient Boosting Machine)
- 익스트림 그래디언트 부스팅(XGB : Extreme Gradient Boosting)
- 라이트 그래디언트 부스팅 머신(LGBM : Light Gradient Boosting Machine)



[GBM 연속형 모델 학습 대화상자]



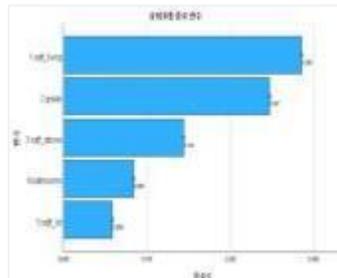
[GBM 연속형 모델 파라미터 설정]



[GBM 연속형 모델 예측 대화상자]

학습/검증 비율		입력 파라미터	
구분	비율	파라미터	값
학습	70	n_estimators	300
검증	30	max_features	sqrt
학습모델 성능검증		max_depth	5
구분	값	min_samples_leaf	2
학습데이터	79.92%	min_samples_split	5
검증데이터	78.08%	learning_rate	2

[GBM 연속형 모델 분석(학습) 결과]



[GBM 연속형 모델 분석 결과]

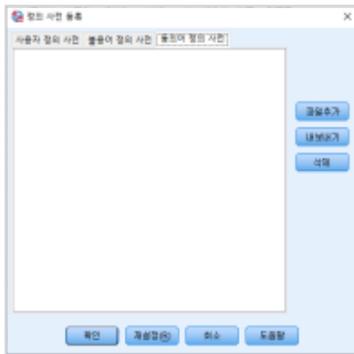
# KoreaPlus의 분석 기능 상세

## | Text Analytics 동의어 정의 사전

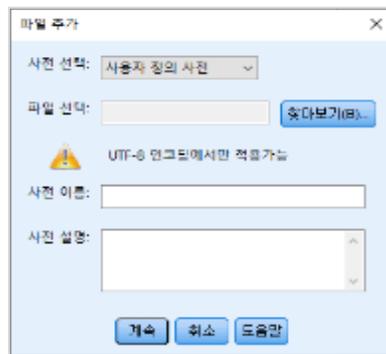
기존 Text Analytics 기능에 동의어 정의 사전을 추가 하였습니다. 사전에 등록 되지 않은 단어를 추가 할 수 있었던 사전 기능에 동의어 정의 사전을 추가하여 분석 결과의 신뢰성을 한층 더 업그레이드 하였습니다. 사용자는 불용어 정의 사전 기능을 통해 불필요한 단어들을 제거하여 모델 구축 시 학습 속도 및 성능을 개선할 수 있고, 동의어 정의 사전 기능을 통해 학습 과정에서 단어의 의미 파악을 좀 더 정확하게 실행 할 수 있습니다. 기존에 제공 된 Text Analytics의 키워드 분석, 연관성 분석과 빈도 분석 기법을 최신화 하였습니다.

### 동의어 정의 사전

- 표현 방식이 다른 2개 이상인 단어가 동일한 의미를 가질 경우 하나의 표현으로 통합 할 수 있는 사전
- 예시 : 대한민국과 한국 동의어 처리
- 대표로 사용할 단어를 첫 번째로 입력한 후 나머지 단어에 대해 쉼표 단위로 구분해서 입력



[정의 사전 등록 대화 상자]



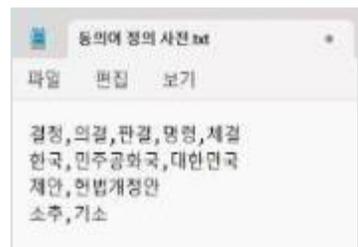
[파일추가 대화상자]



[사용자 정의 사전 형식]



[불용어 정의 사전 형식]

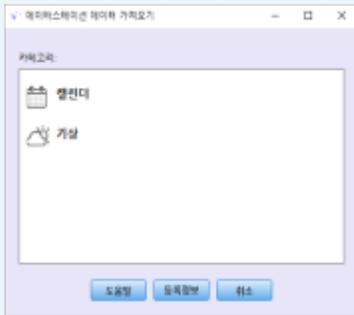


[동의어 정의 사전 형식]

# KoreaPlus의 분석 기능 상세

## Data Station

DataStation은 ㈜데이터솔루션 연구소에서 개발한 대량의 데이터 수집 프로그램입니다. 수집한 데이터들을 가공하여 데이터 프레임으로 변환해주고 요일정보, 기상정보를 수집 및 정제하여 분석 할 수 있도록 데이터 셋을 제공합니다. Data Station으로 수집 된 데이터는 CSV File(.csv)로 저장하여 IBM SPSS Statistics 실행 후 바로 분석 할 수 있습니다.



[Data Station의 대화상자]

일	연도	월	일	요일	연도
1	2020년	1월	1일	월	2020
2	2020년	1월	2일	화	2020
3	2020년	1월	3일	수	2020
4	2020년	1월	4일	목	2020
5	2020년	1월	5일	금	2020
6	2020년	1월	6일	토	2020
7	2020년	1월	7일	일	2020
8	2020년	1월	8일	월	2020
9	2020년	1월	9일	화	2020
10	2020년	1월	10일	수	2020
11	2020년	1월	11일	목	2020
12	2020년	1월	12일	금	2020
13	2020년	1월	13일	토	2020
14	2020년	1월	14일	일	2020
15	2020년	1월	15일	월	2020
16	2020년	1월	16일	화	2020
17	2020년	1월	17일	수	2020
18	2020년	1월	18일	목	2020
19	2020년	1월	19일	금	2020
20	2020년	1월	20일	토	2020
21	2020년	1월	21일	일	2020
22	2020년	1월	22일	월	2020
23	2020년	1월	23일	화	2020
24	2020년	1월	24일	수	2020
25	2020년	1월	25일	목	2020
26	2020년	1월	26일	금	2020
27	2020년	1월	27일	토	2020
28	2020년	1월	28일	일	2020
29	2020년	1월	29일	월	2020
30	2020년	1월	30일	화	2020
31	2020년	1월	31일	수	2020

[캘린더 데이터]

일	연도	월	일	날씨	기온	기온	기온
1	2020	1월	1일	맑음	10.0	1.0	4.0
2	2020	1월	2일	맑음	10.0	1.0	3.0
3	2020	1월	3일	맑음	10.0	1.0	4.0
4	2020	1월	4일	맑음	10.0	1.0	3.0
5	2020	1월	5일	맑음	10.0	1.0	3.0
6	2020	1월	6일	맑음	10.0	1.0	3.0
7	2020	1월	7일	맑음	10.0	1.0	5.0
8	2020	1월	8일	맑음	10.0	1.0	3.0
9	2020	1월	9일	맑음	10.0	1.0	3.0
10	2020	1월	10일	맑음	10.0	1.0	3.0
11	2020	1월	11일	맑음	10.0	1.0	3.0
12	2020	1월	12일	맑음	10.0	1.0	3.0
13	2020	1월	13일	맑음	10.0	1.0	3.0
14	2020	1월	14일	맑음	10.0	1.0	3.0
15	2020	1월	15일	맑음	10.0	1.0	3.0
16	2020	1월	16일	맑음	10.0	1.0	3.0
17	2020	1월	17일	맑음	10.0	1.0	3.0
18	2020	1월	18일	맑음	10.0	1.0	3.0
19	2020	1월	19일	맑음	10.0	1.0	3.0
20	2020	1월	20일	맑음	10.0	1.0	3.0
21	2020	1월	21일	맑음	10.0	1.0	3.0
22	2020	1월	22일	맑음	10.0	1.0	3.0
23	2020	1월	23일	맑음	10.0	1.0	3.0
24	2020	1월	24일	맑음	10.0	1.0	3.0
25	2020	1월	25일	맑음	10.0	1.0	3.0
26	2020	1월	26일	맑음	10.0	1.0	3.0
27	2020	1월	27일	맑음	10.0	1.0	3.0
28	2020	1월	28일	맑음	10.0	1.0	3.0
29	2020	1월	29일	맑음	10.0	1.0	3.0
30	2020	1월	30일	맑음	10.0	1.0	3.0
31	2020	1월	31일	맑음	10.0	1.0	3.0

[기상 데이터]

## Data Station

- 캘린더 옵션은 양력 기준으로 국경일/공휴일, 24절기 정보를 가진 데이터 셋 수집
- 기상 옵션은 일자별 지역별 날씨 정보와 대기질 정보를 가진 데이터 셋 수집
- 데이터 셋에서 전제 일자를 선택할 경우 2020년 1월 1일 부터 현재 날짜 수집

### Check !

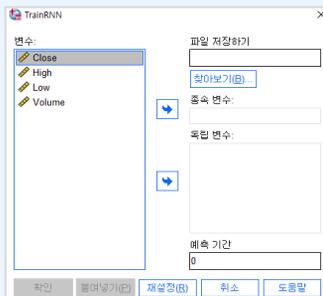
- KoreaPlus의 Data Station은 KoreaPlus Statistics 27버전 이상부터 가능합니다.
- Data Station은 Windows 10 (64bit) 이상 운영 체제에서 구동됩니다.
- 관리자 권한으로 설치 및 실행 합니다.
- Data Station은 인터넷 연결 시 실행 가능 합니다.



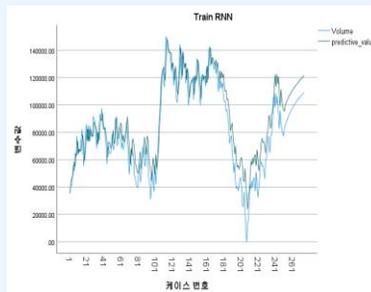
# KoreaPlus의 분석 기능 상세

## AI

RNN(Recurrent Neural Network)은 히든 노드가 방향을 가진 엷지로 연결되어 순환구조를 이루는 Neural Network의 한 종류입니다. KoreaPlus의 RNN 시계열 학습 및 예측을 활용하면 통계적 분석에 익숙하지 않은 사용자들이 손쉽게 시계열 분석을 수행할 수 있습니다.



[RNN시계열학습]



[RNN시계열학습-그래프]

	Volume	Close	Low	High	predictor_value
1	35718.00	12.00	11.00	16.00	35438.66
2	41439.00	15.00	11.00	23.00	43186.40
3	41331.00	15.00	14.00	23.00	47301.49
4	48956.00	17.00	14.00	17.00	47420.49
5	56286.00	17.00	15.00	21.00	53163.75
6	51469.00	18.00	16.00	21.00	58528.33
7	64720.00	18.00	13.00	21.00	55191.26
8	62290.00	20.00	20.00	24.00	64096.23
9	77150.00	22.00	22.00	33.00	63164.59
10	64280.00	22.00	21.00	25.00	73891.81

[결과-예측된 데이터시트]

## AI(RNN)

- LSTM(Long Short Term Memory) 사용
- 분석하고자 하는 값, 시계열 분석에 영향을 줄 수 있는 요소들과 보유한 데이터 예측
- 연장할 기간 등 설정, 예측 된 값이 추가 된 데이터 시트를 얻음
- Train loss Value 정보를 출력하여 예측한 값과 실제 값의 차이 확인

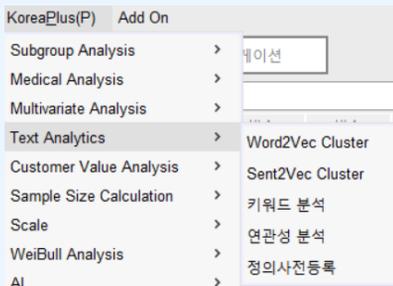
### Check !

- KoreaPlus의 AI는 KoreaPlus Statistics 25버전 이상부터 가능합니다.
- 딥러닝 알고리즘은 많은 연산이 필요하기 때문에 CPU만으로는 학습에 많은 시간이 소요 됩니다.
- 속도를 향상시키기 위해 \*GPU(Graphic Processing Unit)가 있는 환경에서 학습 하는 것을 권장합니다.
- GPU 환경확인
  - ▶ PC의 종류가 64비트이면서 그래픽 카드가 NVIDIA일 경우 사용이 가능 (속도 최소10배 차이)
- [Windows PC 종류 확인] 제어판 > 시스템 및 보안 > 시스템 > 시스템종류 : 64비트 운영체제
- [Windows 그래픽 카드 확인] 제어판 > 장치 관리처 > 디스플레이 어댑터에 NVIDIA가 존재하는지 확인

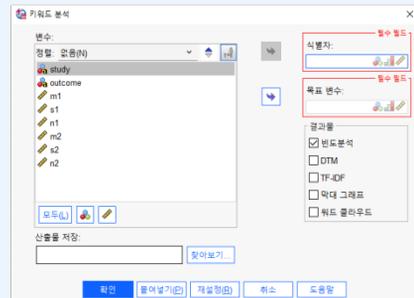
# KoreaPlus의 분석 기능 상세

## TA(Text Analytics)

고객 의견, 제품 리뷰, 피드백 등 구조화되지 않은 텍스트 데이터를 의미 있는 데이터로 변환하고 키워드 분석, 연관성 분석, 빈도수 분석 등을 통해 텍스트 데이터의 정보를 얻어 의사결정에 사용 할 수 있습니다. SNS와 뉴스 데이터 분석을 통한 사회 트렌드 분석, 고객의 피드백 분석을 통한 고객관리, 콘텐츠 강화 등 다양한 분야에 적용이 가능합니다.



[Text Analytics Menu]



[키워드 분석]

### Text Analytics

- 형태소 분석을 통해 각 단어가 현 문서에서 얼마나 출현하는지 확인 가능
- TF-IDF(Term Frequency – Inverse Document Frequency)에서 TF를 이용하여 가중치가 적용 된 중요도 분석 가능
- 형태소 단위로 문장 분류 시 기존 사전에 등록되어 있지 않은 복합명사, 신조어, 이름 등을 [사용자 정의 사전]을 통해 신규 등록 가능

### Check !

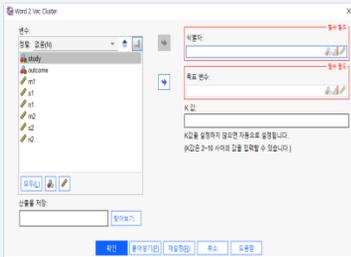
- KoreaPlus Statistics 25버전 이상부터 가능합니다.



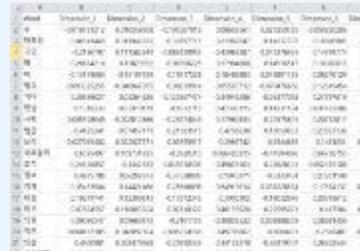
# KoreaPlus의 분석 기능 상세

## TA(Text Analytics) - Word2VecCluster

Word2Vec은 사람이 사용하는 자연어(Natural Language)를 벡터로 표현하는 임베딩 (Embedding) 기법을 적용해 단어의 의미를 수치화하여 단어 간 유사도를 반영해 주고, 머신러닝을 통해 자료 내에 비슷한 의미를 지닌 단어들을 묶어 시각화한 분석입니다. KoreaPlus의 Word2VecCluster를 활용하면 Word Embedding 분석이 익숙하지 않은 사용자들에게 손쉽게 텍스트 분석 및 원하는 정보를 제공 받을 수 있습니다.



[Word2VecCluster의 대화상자]



[분석결과 데이터셋]

## Word2VecCluster

### Word Embedding ?

- 텍스트를 구성하는 단어들을 수치화하는 방법의 일종, n차원의 벡터(vector)로 매핑(mapping)
- 벡터간 거리를 계산하여 단어간 유사도 확인
- CBOW(Continuous Bag-of-Words)와 Skip-Gram 2가지의 모델 중 Skip-Gram 활용
- 많은 양의 텍스트 데이터들을 효율적으로 표현 및 분석하여 핵심 정보를 손쉽게 찾아낼 수 있음

### Check !

- KoreaPlus Statistics 25버전 이상부터 가능합니다.



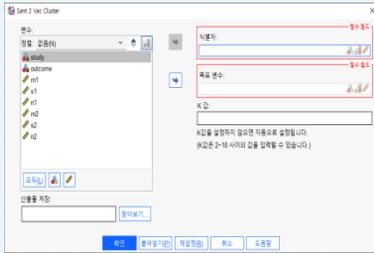
# KoreaPlus의 분석 기능 상세

## TA(Text Analytics) - Sent2VecCluser

Word2vec의 임베딩 개념을 확장한 Doc2Vec을 활용하여 문장단위로 벡터화하는 방법을 사용합니다.

단어 임베딩에 비해 더 넓은 문맥으로 이해하고 문장간의 의미를 좀 더 정확히 파악할 수 있습니다.

Sent2Vec 모델의 PV-DM(Distribute memory)방식을 채택하여 문장단위 분석과 문서단위 분석을 진행합니다.



[Sent2VecCluster대화상자]

번호	문장	클러스터	비율	중심	거리
1	대한민국은 민주 공화국이다.	1	0.500000	0.500000	0.000000
2	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
3	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
4	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
5	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
6	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
7	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
8	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
9	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
10	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
11	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
12	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
13	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
14	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
15	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
16	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
17	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
18	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
19	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000
20	대한민국의 국기는 태극 문장이다.	1	0.500000	0.500000	0.000000

[분석결과 데이터셋]

### Sent2VecCluser

#### PV-DM(Distribute memory)방식

- 네트워크 모델을 통해 학습하고 다음에 나타날 것을 예측하여 현재 문맥에서 누락된 정보나 단락의 주제를 기억함

#### Check !

- KoreaPlus Statistics 25버전 이상부터 가능합니다.



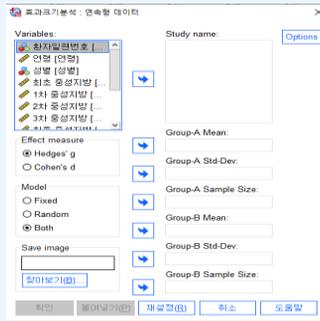
# KoreaPlus의 분석 기능 상세

## Meta Analysis

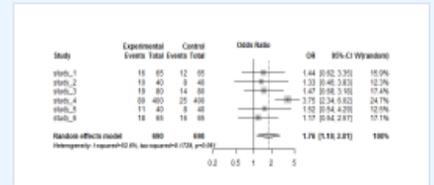
메타분석은 분석의 분석(analysis of analysis)으로, 여러 연구 결과를 종합할 목적으로 개별 연구 결과들을 모아 통계적으로 분석하는 방법입니다. 구에서 제시된 결과들의 통합된 요약 추정치를 정량적으로 산출하여 효과 및 효율성을 평가하기 위해 사용되는 통계적 방법입니다. 동일한 주제하에 수행된 여러 연구들로부터 도출된 다양한 연구결과를 체계적으로 종합 분석하는 계량적 접근 방법입니다.



[Meta Analysis 메뉴]



[연속형 데이터 대화상자]



[Forest plot]

## Meta Analysis

- Effect Size, Moderator Analysis, Publication Bias Analysis, Cumulative Meta Analysis, Sensitivity Analysis 의 연속형 데이터, 이분형 데이터, 상관관계 데이터 결과 값 제공.

### Meta Analysis의 장점

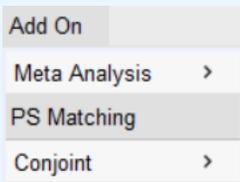
- 메타분석으로 종합된 연구 결과는 다양한 실험 조건과 방법을 포함하여 강력하고 일반화 가능한 증거를 제공합니다.
- 메타분석은 편향되지 않은 연구 표본을 선정하므로, 결과는 객관적 추론이 가능한 신뢰할 만한 연구 표본에 근거합니다.
- 표본 크기가 증가함에 따라 단일 연구에서 도출된 효과 크기보다 더 신뢰할 만한 효과 추정치를 얻을 수 있습니다.
- 상반된 연구 결과를 비교 분석하여 원인을 규명하고 논쟁을 조절할 수 있습니다.



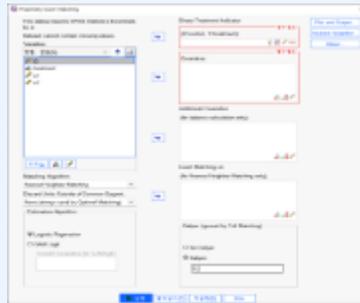
# KoreaPlus의 분석 기능 상세

## PSM (Propensity Score Matching)

Felix Thoemmes 교수가 개발한 KoreaPlus의 확장 번들로 실험군과 대조군 선정 시 나타나는 선택 편향을 줄이기 위해 Propensity Score를 기준으로 짝짓기하는 방법입니다. 선택 편향을 줄인 상태에서 실험군과 대조군을 재구성함으로써 이에 따라 처치의 효과를 보다 정확하게 비교할 수 있는 방법이며, 비슷한 성향 점수에 의해 짝짓기로 구성된 두 집단을 통해 최대한 선택편향을 줄이고, 무작위 할당을 실시한 연구와 같이 중재에 대한 보다 정확한 효과를 계산할 수 있습니다.



[대화상자 경로]



[대화상자]



[분석 결과]

### Propensity Score Matching

#### IBM SPSS 성향점수 매칭과 KoreaPlus PS Matching 비교

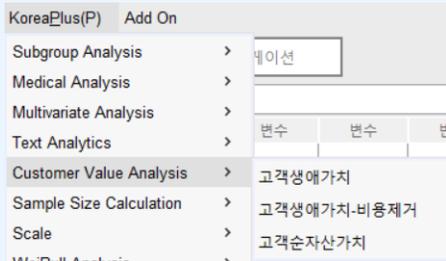
메뉴 위치	데이터 → 성향 점수 매칭	Add on → PS Matching
제품	IBM SPSS	KoreaPlus Statistics
알고리즘	Logistic Regression	<ul style="list-style-type: none"> <li>Logistic Regression</li> <li>GAM Logit</li> </ul>
매칭 알고리즘	Fuzzy	<ul style="list-style-type: none"> <li>Nearest Neighbor</li> <li>Full</li> <li>Optimal</li> </ul>
특징	<ul style="list-style-type: none"> <li>매칭되는 케이스 쉽게 확인</li> <li>1:1 매칭 만 가능</li> <li>일치 허용 오차(성향점수차)를 통해 매칭 대상 제한 가능</li> </ul>	<ul style="list-style-type: none"> <li>1:1 and 1:N 매칭</li> <li>Caliper (표준편차)를 이용한 매칭 대상 제한 가능</li> <li>추가적인 매칭 알고리즘 제공</li> <li>GAM Logit을 이용한 공변량 smoothing 제공</li> </ul>

# KoreaPlus의 분석 기능 상세

## Customer Life Value : CLTV

Customer Life Value : CLTV, 고객생애가치는 고객으로부터 미래의 일정기간 얻게 될 이익 (=수입-비용)을 할인율에 따라 현재 가치로 환산한 재무적 가치입니다.

고객 가치 분석은 고객 개인별로 전체 생애 동안의 총 기여가치를 측정하는 방법입니다.



[CLTV 메뉴]

	고객ID	공헌마진	잔존생애 기간	할인율	CLTV
1	1	15.00	5	1.0	66.86
2	2	45.00	3	.5	55.95
3	3	24.00	6	.7	73.17
4	4	75.00	10	.7	322.59
5	5	9.00	8	.9	43.21
6	6	13.50	25	.4	49.02
7	7	36.00	3	.8	71.62
8	8	19.50	8	.5	62.02
9	9	69.00	11	.3	134.45
10	10	30.00	25	.2	64.46
11	11	105.00	4	.7	232.99
12	12	96.00	20	.4	326.92
13	13	75.00	18	.5	307.55
14	14	4.50	27	.6	24.94
15	15	19.50	30	.7	128.68
16	16	135.00	20	.5	574.67
17	17	111.00	7	.1	64.04
18	18	28.50	10	.9	157.61
19	19	40.50	14	.8	238.68
20	20	45.00	23	.9	359.77

[분석 결과]

### Customer Life Value

- KoreaPlus에서의 고객 생애 가치 분석은 마케팅 비용을 고려하지 않고 단순히 관찰 기간에 나타날 공헌 마진을 현재 가치로 할인하여 측정하는 방법으로 제공
- 장점 : 측정 자체가 매우 단순하여 활용적인 측면이 강함
- 개인별 마케팅 비용의 차이가 크지 않은 산업이나 CRM 관점의 마케팅 비용을 측정하기 어려운 기업에서 활용 가능

# KoreaPlus의 분석 기능 상세

## 의학 관련 분석 기법

KoreaPlus의 의학관련 분석 기법은 의학 분야에서 주로 사용되는 고급 분석 기법들을 대화상자로 구현하여 제공합니다.

- Medical Analysis : 독성 실험 및 생존 관련 임상 시험에서 주로 활용되는 고급분석 기능 제공

KoreaPlus(P)	Add On
Subgroup Analysis	> <input type="text" value="레이션"/>
Medical Analysis	> Bartlett's Test
Multivariate Analysis	> Dixon's Q-Test
Text Analytics	> Cochran-Armitage Trend Test
Customer Value Analysis	> Poly-k Test
Sample Size Calculation	> Peto's Test
Scale	> Diagnostic & Agreement Statistics
WeiBull Analysis	> Chow's Test
AI	> Page's Trend Test
Data Station	> Rank ANCOVA
Compare Means	> Bland-Altman plot
Korea Statistics	> Bland-Altman plot

## Conjoint / Multivariate Analysis

Syntax로만 구현되었던 분석을 대화상자로 분석할 수 있도록 메뉴를 제공합니다.

- Conjoint 분석  
제품/서비스의 여러 특성 가운데 어떤 특성요인이 소비자에게 영향을 미치는지 파악하는 마케팅 리서치 기법으로 신제품 컨셉 평가, 최적 가격 설정, 시장 세분화에 활용

Add On	
Meta Analysis	> <input type="text"/>
PS Matching	> <input type="text"/>
Conjoint	> 설계
	데이터
	분석

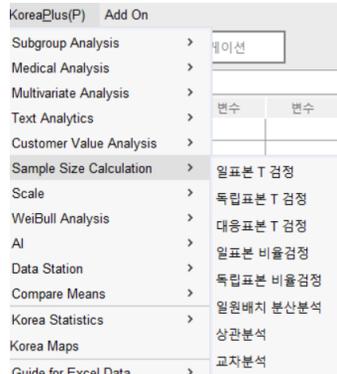
- Multivariate Analysis  
다변량 분석/주성분분석, 정준상관분석, 조사표본가중법 제공

KoreaPlus(P)	Add On
Subgroup Analysis	> <input type="text" value="레이션"/>
Medical Analysis	> <input type="text"/>
Multivariate Analysis	> 정준상관분석
Text Analytics	> 주성분분석
Customer Value Analysis	> 조사표본 가중법

# KoreaPlus의 분석 기능 상세

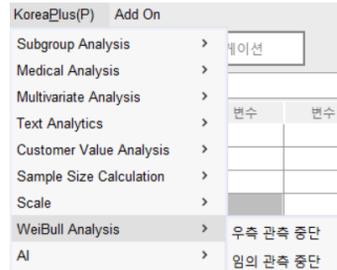
## Sample Size Calculation (표본 크기 계산)

- 최적의 표본 크기를 산출하는 기법입니다.
- 이를 통해 통계적 검정력 확보는 물론 표본 산출을 위한 시간 및 비용 절약 할 수 있습니다.



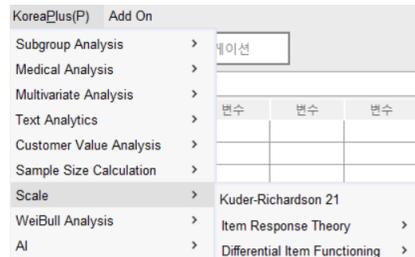
## Weibull Analysis

- 가장 적절한 수명분포를 결정하고 모수를 추정하는 접근법 입니다.
- 주로 부품의 수명 측정에 사용되며 질병으로 인한 사람의 수명이나 제품의 제조와 운반에 걸리는 시간, 날씨 예보 등에 활용됩니다.



## Scale

- Kuder-Richardson21은 문항 간의 내적 일치도 정도를 나타내어주는 지수입니다.
- Differential Item Functioning은 차별기능문항으로 각 문항이 특정 집단에 편파적으로 적용되는지 판별하는데 사용되는 기법입니다.
- Item Response Theory는 문항반응이론으로 문항에 대한 피험자의 반응과 개인의 잠재 능력 간의 관계를 수리적 함수로 나타낸 측정 이론입니다.

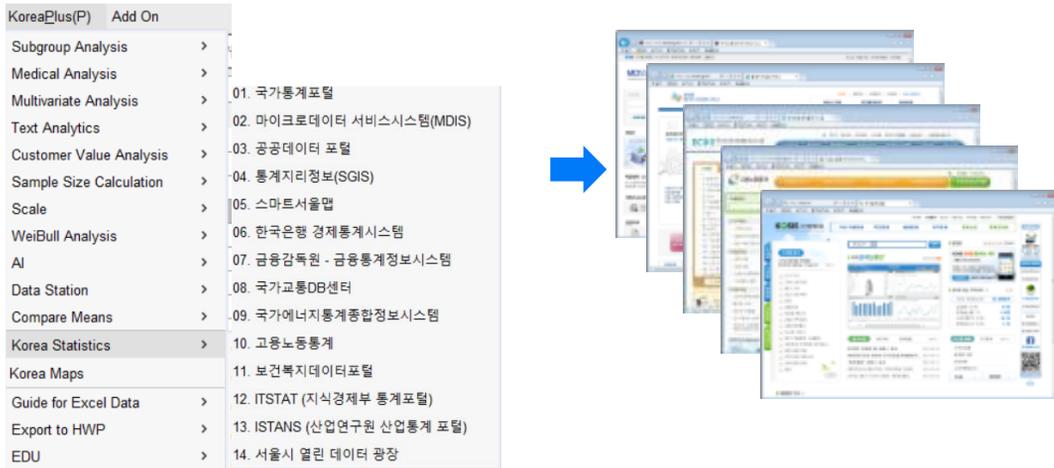


# KoreaPlus의 분석 기능 상세

## 국가통계포털 링크 추가

국가 통계 데이터를 이용하기 위한 국가 통계 제공 사이트 바로가기 기능을 제공 합니다.

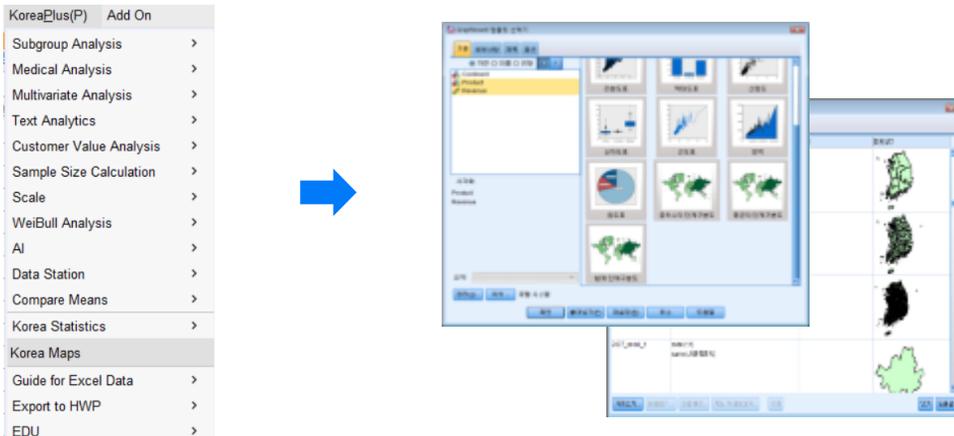
- Korea Statistics 메뉴를 통해 통계 관련 사이트(정부 · 공공)로 바로 이동이 가능합니다.



## 한국 지도 제공

KoreaPlus Statistics Maps에 한국 지도 템플릿 추가

- Korea Maps 메뉴 클릭만으로 한국지도 파일을 생성 할 수 있습니다.
- 사용자가 직접 한국 지도를 다운받지 않아도 통계지리 정보사이트에서 제공하는 다양한 한국 지도를 활용 할 수 있습니다.

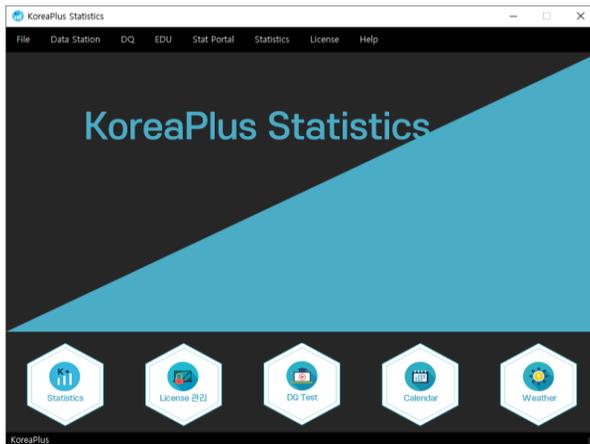




# KoreaPlus Statistics Utility 1.0

## KoreaPlus의 확장 유틸리티 솔루션

- Data Station : 캘린더, 기상 데이터 수집 기능 (데이터셋 제공)
- DQ : 데이터 활용 역량 자가진단 서비스 (보고서 PDF 형태 출력)
- Edu : 기초, 중급 통계분석 학습 자료 제공
- KoreaPlus Statistics의 License 관리



[KoreaPlus Statistics Utility 실행 창]



[주요 위젯]

## GS 인증 1등급 취득



데이터솔루션의 전문 기술팀이 개발한 KoreaPlus Statistics Utility는 KoreaPlus Statistics의 확장 유틸리티 솔루션입니다. KoreaPlus Statistics Utility는 2023년 4월 GS(Good Software)1등급을 취득하고, 데이터 처리 역량과 소프트웨어 개발 역량 및 품질 우수성을 공식적으로 인증 받았습니다.

KoreaPlus Statistics 제품 구매 시 함께 제공 되는 KoreaPlus Statistics Utility는 그래픽 기반 사용자환경(GUI)에서 위젯을 통해 KoreaPlus 제품 군의 라이선스를 손쉽게 관리할 수 있고, 캘린더 및 기상 데이터를 수집하고 분석할 수 있도록 데이터 셋을 제공하며, 분석가들이 스스로 데이터 활용역량을 점검 할 수 있는 서비스를 제공합니다.

DATASOLUTION

# KoreaPlus Statistics

- Embedded on IBM SPSS Statistics Professional



## | DATASOLUTION Service

- 제품 구성
- 기술지원 및 운영 서비스
- Community SPSS User Portal
- 교육지원 빅데이터러닝센터

# DATASOLUTION Service

## 제품 구성

### 정품 설치

- 설치파일 다운로드 링크

### 분석 매뉴얼

- 설치 책자 형태 및 pdf 파일 다운로드 링크

### 정품 사용 인증서

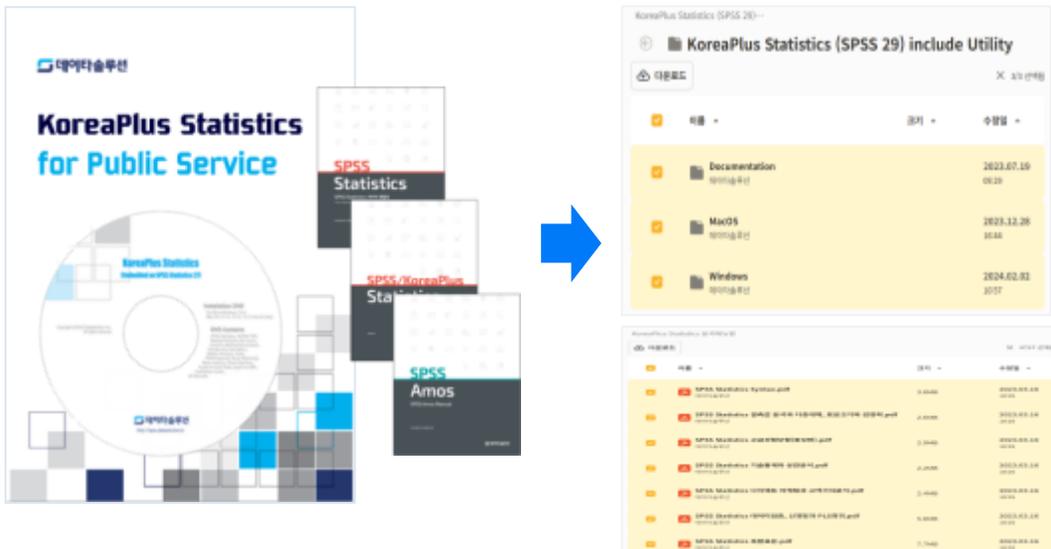
- 라이선스 인증 코드
- 기술지원 서비스 코드 (Serial Number)
- e-mail 인증서 발송

### 빅데이터러닝센터 교육 쿠폰

- 온라인(e러닝) 교육 쿠폰 e-mail
- 오프라인 실습 교육 쿠폰 e-mail

## Check

- Ver.29부터 제품 구매 시 설치 파일과 매뉴얼 도서는 e-mail로 다운로드 링크가 발송 됩니다.
- Brief Guide, KoreaPlus 매뉴얼 책자 제공



# DATASOLUTION Service

## | 기술지원 및 운영 서비스

데이터솔루션은 숙련된 전문 기술지원 팀(인력)을 보유하고 있습니다. KoreaPlus Statistics 제품 구매 시 전문 기술자의 설치 및 테스트 서비스가 제공 됩니다. 기술지원코드(Serial Number)를 통해 평생 라이선스 관리를 받고 재 인증, 재 설치 등의 기술지원 서비스를 무상으로 제공 받을 수 있습니다. 제품 구매 시점으로부터 1년 무상유지보수를 제공하며 무상 유지보수 기간 내 신 버전 출시의 경우 Major Upgrade를 제공합니다.

### 설치 및 기술지원

- 무상유지보수 1년 (SW Subscription & Support 12 Months)
- 프로그램 설치 지원
- 재 인증, 재 설치 지원
- 전담 기술 지원팀(Technical Support) 보유
- 전문 기술자 방문 / 원격 설치 지원
- License Manager 설치
- Client 설치 및 정상 운영 Test
- Log File 설정

### 운영지원 서비스

- 다양한 기술지원 채널 제공 (전화, 이메일, 사용자포털, 블로그)
- Log File 분석을 통해 사용현황 분석결과 제공
- 운영자 교육 진행 (License Manager 관리)

# DATASOLUTION Service

## SPSS User Portal 운영

데이타솔루션은 사용자 편의를 위한 SPSS User Portal을 운영하고 있습니다. User Portal에서는 라이선스 구매 및 각종 기술지원(재 인증 외), 최신버전 트라이얼을 제공하고 함께 온라인 쇼핑몰 DS몰 서비스도 함께 운영하고 있습니다. 데이타솔루션의 KoreaPlus Statistics 제품 군을 구매하신 고객 외에도 자유롭게 이용이 가능합니다.

<https://spss.datasolution.kr/main/main.asp>

### 제품

**KoreaPlus Statistics** (Embedded on IBM SPSS Statistics)

- KoreaPlus Statistics >
- AI Coach
- Embedded on IBM SPSS Statistics
- Embedded on IBM SPSS Amos
- Embedded on IBM SPSS Modeler



**KoreaPlus Statistics** (Embedded on IBM SPSS Statistics)

다양한 분야에서 활용되고 있는 SPSS 제품 고급분석과정을 대화상자(GUI)로 구현한 확장 모듈을 통해 더 편리하고 신속하게 분석을 수행할 수 있습니다.

[바로구매](#)

### TRIAL 다운로드



**KoreaPlus Statistics (Embedded on IBM SPSS Statistics 30)**

전세계에서 가장 많은 사용자를 보유하고 있는 글로벌 통계 소프트웨어로 방대한 데이터 처리가 용이하며, 편리하고 강력한 분석기능을 제공합니다.

[설치파일](#)      [설치가이드](#)



◀ DS몰

<https://dsmallstore.kr/>

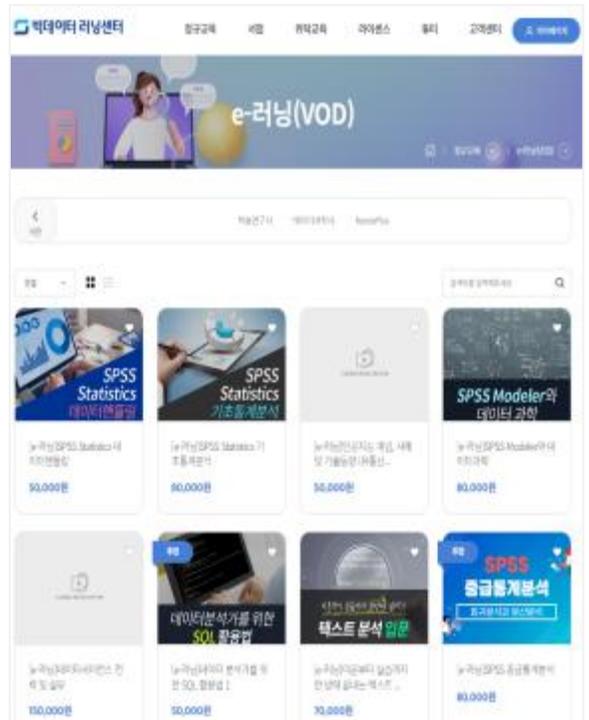
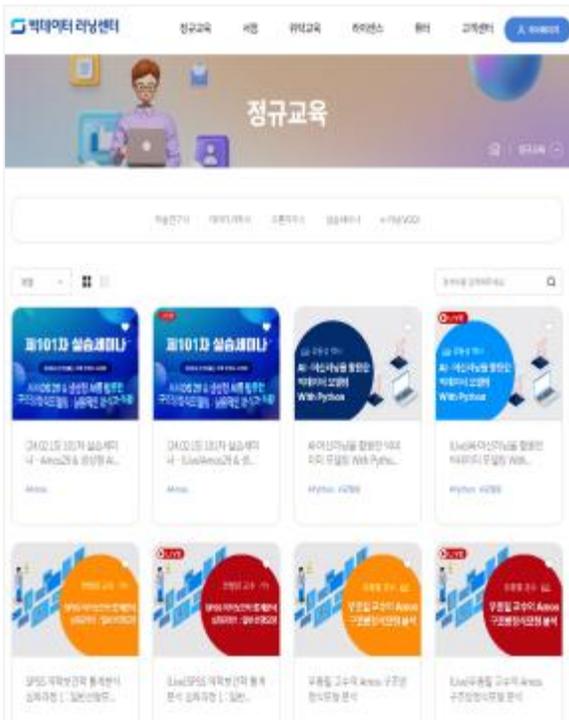
데이타솔루션이 운영하는 데이터 분석·통계·연구용 소프트웨어 및 교육상품 전문 온라인몰

# DATASOLUTION Service

## 빅데이터러닝센터 운영

빅데이터러닝센터는 데이터솔루션의 데이터 분석 전문 교육기관으로 전문강사 25명, 강의파트너 200명의 인력을 보유하고 있습니다. 러닝센터는 최대 30명이 동시에 실습할 수 있는 교육장을 제공하며, 학술연구자, 데이터과학자를 위한 다양한 커리큘럼을 제공합니다. KoreaPlus Statistics 제품 군을 구매한 고객은 온라인(e러닝), 오프라인 교육을 무상으로 제공 받을 수 있습니다.

<https://www.dataai.kr/main/index.jsp>



- 사용자 교육을 위한 자체적인 교육장, 실습 PC 보유
- 전문적이고 체계적인 위탁교육 프로그램 제공
- 계약 체결 시 정기/비정기 교육쿠폰 무상제공
- 빅데이터러닝센터 정기/비정기 교육 20% 할인
- 데이터솔루션 세미나 및 Promotion 우선 참여기회 부여



빅데이터러닝센터

DATASOLUTION

# 감사합니다.

 데이터솔루션

Contact US

DA 영업본부 [swg@datasolution.kr](mailto:swg@datasolution.kr)

[www.datasolution.kr](http://www.datasolution.kr) | 대표번호) 02-3467-7200 | 팩스) 02-563-0014

우) 06101 서울시 강남구 언주로 620 현대인텔렉스빌딩 10층