데이타솔루션

KoreaPlus Statistics
- Embedded on SPSS Statistics 27

KoreaPlus Statistics 란 무엇입니까?

- ✓ KoreaPlus Statistics 란 SPSS Statistics에 데이타솔루션만의 Value Add Component와 서비스가 추가된 KoreaPlus 패키지 입니다.

KoreaPlus Statistics 에는 어떤 기능들이 있습니까?

- ✓ 데이타스테이션 제공(Data Station) (NEW)
- ✓ AI 분석기법 제공 (RNN 기반 시계열 학습/예측)
- ✓ TA 분석기법 제공 (Text Analytics, Tex2Vec)
- ✓ 의학 관련 분석기법 제공 (Medical Analysis, Bio-Equiv, Bio-Equiv(2x4))
- ✓ 추가 분석기법 제공 Ⅰ (Conjoint, Multivariate Analysis)
- ✓ 추가 분석기법 제공 Ⅱ (Customer Value Analysis)
- ✓ 추가 분석기법 제공 Ⅲ (Sample Size Calculation, Weibull Analysis, Scale)
- ✓ 추가 분석 기법 제공 IV (PSM(Propensity Score Matching), Meta-analysis)
- ✓ 국가통계포털 링크 (Korea Statistics)
- ✓ Maps graph를 작성하기 위한 한국지도 제공 (Korea Maps)
- ✓ 초보자를 위한 SPSS 활용하기 방법 및 샘플 데이터 제공(SPSS Training)
- ✓ 엑셀 데이터 활용을 위한 도움말 기능 (Guide for Excel Data)
- ✓ 출력결과(Output)를 한글파일(HWP)로 바로 내보내기 (Export to HWP)

KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

IBM에서 소개한 KoreaPlus Statistics

SPSS Statistics에 Value Add Component가 추가 된 KoreaPlus Statistics를 IBM Partner World에서 소개하고 있습니다.

웹사이트 주소 :

https://www-356.ibm.com/partnerworld/gsd/member/solutiondetails.do?solutionId=50941&lc=ko&stateCd=W

Datasolution KoreaPlus Statistics

Datasolution KoreaPlus Statistics는 산업영역에 따라 General Science, Public Service, Medical Service, Data Analysis로 나누어 집니다. 교육, 공공, 의료, 마케팅 등 다양한 SPSS 활용 분야에서 통계분석을 좀 더 편리하게 사용하도록 고급분석 기능 추가하여, 대한민국 실정에 맞는 현지화 기능을 제공하고, 사용자의 통계분석 역량 및 솔루션 활용 역량을 향상시키는 도움을 줍니다.

IBM PartnerWorld > Global Solutions Directory >

Datasolution KoreaPlus

솔루션 개요 | 솔루션 세부사항

ID 번호 : 50941

솔루션 이름 :	Datasolution KoreaPlus
솔루션 설명 :	DataSoluton KoreaPlus은 산업영역에 따라 General Science, Public Service, Medical Service, Data Analysis 로 나누어집니다. 교육, 공공, 의료, 마케팅 등 다양한 SPSS활용분야에서 통계분석을 좀 더 편리하게 사용 가능하도록 고급분석 기능 추가, 대한민국 실정에 맞는 현지화 기능, 사용자의 통계분석 역량 및 솔루션 활용 역량을 향상시키는 도움기능이 제공되어 있습니다.
솔루션 웹 페이지 :	http://www.datasolution.kr

이 솔루션은 IBM이 검증했습니다.

전문성과 권한 부여

IBM Gold Business Partner

Software Value Plus Capability Authorization
Business Analytics

마지막으로 수정한 날짜: 2017. 3. 6

KoreaPlus Statistics

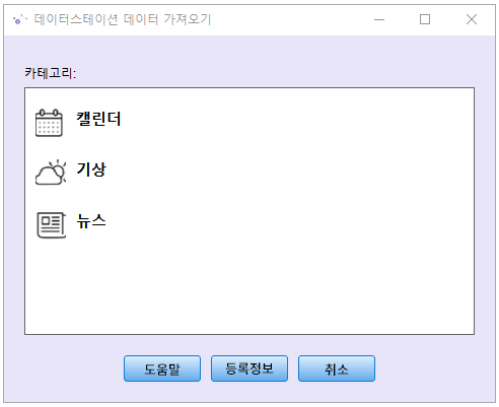
- Embedded on SPSS Statistics 27 Standard

New KoreaPlus

- Data Station 기능이 추가되었습니다.

- 데이타솔루션의 자체 솔루션인 데이터스테이션에서 분석에 필요한 데이터셋을 가져옵니다. 데이터스테이션은 날짜, 기상, 뉴스 데이터를 수집, 정제하여 제공합니다.

  - ✓ 캘린더 옵션은 양력 기준으로 음력, 국경일/공휴일, 24절기 정보를 가진 데이터셋을 가져옵니다,
  - ✓ 기상 옵션은 일자별 지역별 날씨 정보와 대기질 정보를 가진 데이터셋을 가져옵니다.
  - ✓ 뉴스 옵션은 사용자가 입력한 키워드를 갖고 네이버 뉴스를 검색하여 키워드 데이터셋을 가져옵니다.
  - ✓ KoreaPlus의 Data Station은 KoreaPlus Statistics 27버전 이상을 지원합니다.
  - ✓ Data Station 사이트에서 데이터를 가져오기 때문에 인터넷이 연결되어 있어야 합니다.



[Data Station의 대화상자]

	sol_ymd	sol_year	sol_month	sol_day	sol_nday	sol_leapyear	sol_week	lun_year
1	20200101	2020	1	1	0	0	3	2019
2	20200102	2020	1	2	0	0	4	2019
3	20200103	2020	1	3	0	0	5	2019
4	20200104	2020	1	4	0	0	6	2019
5	20200105	2020	1	5	0	0	0	2019
6	20200106	2020	1	6	0	0	1	2019
7	20200107	2020	1	7	0	0	2	2019
8	20200108	2020	1	8	0	0	3	2019
9	20200109	2020	1	9	0	0	4	2019
10	20200110	2020	1	10	0	0	5	2019
11	20200111	2020	1	11	0	0	6	2019
12	20200112	2020	1	12	0	0	0	2019
13	20200113	2020	1	13	0	0	1	2019
14	20200114	2020	1	14	0	0	2	2019
15	20200115	2020	1	15	0	0	3	2019

[캘린더 데이터]

	main_name	sub_name	city_name	w_date	area_code	actual_yn	w_description	w_avg_ta
1	강원도	남부	태백시	2020-01-01	1003	1	맑음	-4.9
2	강원도	남부	태백시	2020-01-02	1003	1	맑음	-3.4
3	강원도	남부	태백시	2020-01-03	1003	1	맑음	-.4
4	강원도	남부	태백시	2020-01-04	1003	1	맑음	-1.9
5	강원도	남부	태백시	2020-01-05	1003	1	맑음	-2.9
6	강원도	남부	태백시	2020-01-06	1003	1	흐림	.0
7	강원도	남부	태백시	2020-01-07	1003	1	비	5.2
8	강원도	남부	태백시	2020-01-08	1003	1	비	2.7
9	강원도	남부	태백시	2020-01-09	1003	1	맑음	-2.6
10	강원도	남부	태백시	2020-01-10	1003	1	구름조금	-1.4
11	강원도	남부	태백시	2020-01-11	1003	1	맑음	.0
12	강원도	남부	태백시	2020-01-12	1003	1	구름조금	-3.2
13	강원도	남부	태백시	2020-01-13	1003	1	맑음	-4.8
14	강원도	남부	태백시	2020-01-14	1003	1	맑음	-4.2
15	강원도	남부	태백시	2020-01-15	1003	1	맑음	-5.6

[기상 데이터]

	entity	entity
1	청와대	location
2	경기도	location
3	이태원	location
4	대구시	location
5	인천공항	location
6	고등학교	
7	광화문	
8	초등학교	
9	부산시	
10	후베이성	
11	대구교회	
12	인천시	
13	금융시장	
14	마시아	
15	제주도	

	subject	doccnt
1	코로나	140479
2	확진	61037
3	마스크	17922
4	대구	12316
5	지원	12174
		10862
		10805
		10491
		9646
		9541
		9094
		8455
		8060
		7744
		7555

	relword	doccnt
1	확진 양성판정	9971
2	코로나 확진판정	7098
3	코로나 양성판정	6175
4	확진 음성판정	5008
5	확진 진단검사	4395
6	코로나 진단검사	3937
7	코로나 확산방지	3649
8	코로나 음성판정	3417
9	확진 격리조치	2950
10	확진 격리해제	2695
11	코로나 마스크착용	2694
12	코로나 위기극복	2453
13	확진 검사실시	2392
14	코로나 격리조치	2323
15	확진 검사진행	2304

[뉴스 데이터]

KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

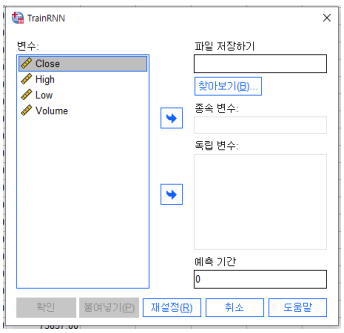
AI (RNN기반 시계열 학습/예측)

- RNN(Recurrent Neural Network)은 히든 노드가 방향을 가진 엣지로 연결되어 순환구조를 이루는 Neural Network의 한 종류.
- KoreaPlus의 RNN 시계열 학습 및 예측을 활용하면 통계적 분석에 익숙하지 않은 사용자들이 손쉽게 시계열 분석 수행 가능

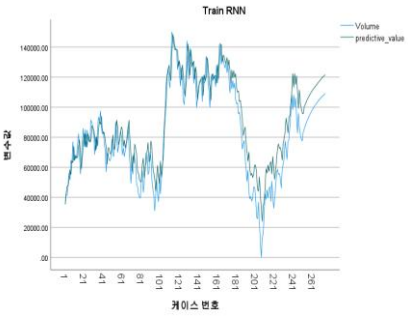
✓ LSTM(Long Short Term Memory) 사용

✓ 분석하고자 하는 값과 시계열 분석에 영향을 줄 수 있는 요소들, 갖고 있는 데이터를 예측하여 연장할 기간 등을 설정하고 예측 된 값이 추가 된 데이터 시트를 얻음

✓ Train loss Value 정보를 출력하여 예측한 값과 실제 값의 차이 확인



[RNN시계열학습]



[RNN시계열학습-그래프]

	Volume	Close	Low	High	predictive_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]

[Check Point]

* 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 **NVIDIA**일 경우 사용이 가능 (속도 최소10배 차이)

[Windows PC 종류 확인] 제어판 > 시스템 및 보안 > 시스템 > 시스템종류 : 64비트 운영체제

[Windows 그래픽 카드 확인] 제어판 > 장치 관리자 > 디스플레이 어댑터에 **NVIDIA**가 존재하는지 확인

KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

TA (Text Analytics)

- 고객 의견, 제품 리뷰, 피드백 등 구조화되지 않은 텍스트 데이터를 의미 있는 데이터로 변환하고 감성 분석, 연관성 분석, 빈도수 분석 등을 통해 텍스트 데이터의 정보를 얻어 의사결정에 사용 가능
- SNS와 뉴스 데이터 분석을 통한 사회 트랜드 분석, 고객의 피드백 분석을 통한 고객관리, 컨텐츠 강화 등 다양한 분야에서 적용 가능

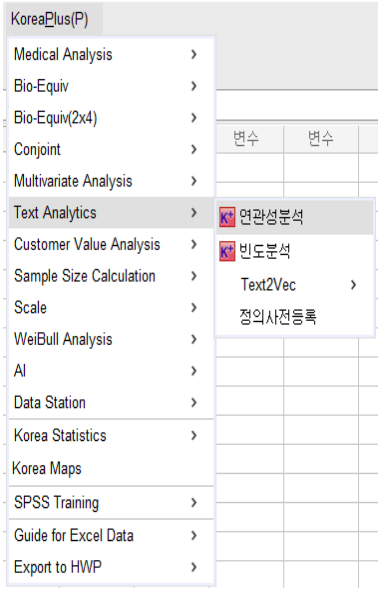
✓ 형태소 분석을 통해 각 단어가 현 문서에서 얼마나 출현하는지 확인 가능

✓ TF-IDF(Term Frequency – Inverse Document Frequency)에서 TF를 이용하여 가중치가 적용 된 중요도 분석 가능

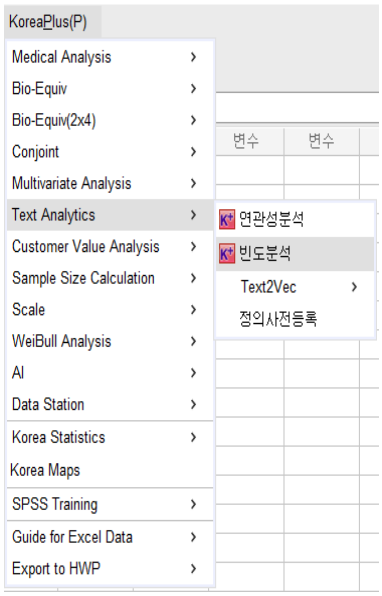
✓ Text2Vec(KoreaPlus Statistics 26버전 이상 가능)

① Word2VecCluster : 머신러닝을 통해 자료 내에 비슷한 의미를 지닌 단어들을 묶어서 시각화한 분석으로 Word Embedding 분석이 익숙하지 않은 사용자들에게 손쉽게 텍스트 분석 및 원하는 정보 제공

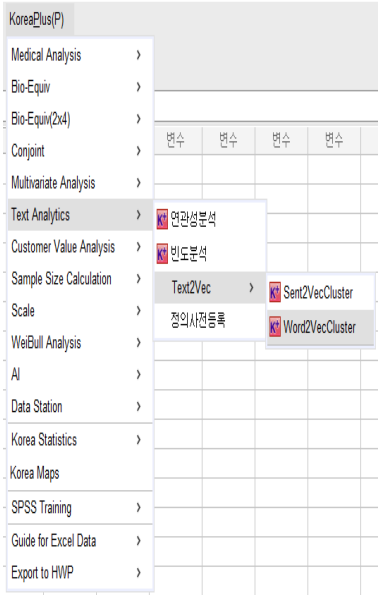
② Sent2VecCluster : word2vec에서 확장된 분석으로 비슷한 의미의 문장끼리, 비슷한 의미의 문서끼리 묶어서 시각화한 분석으로 PV-DM(Distribute memory)방식을 채택하여 문장단위 분석과 문서단위 분석을 진행



[연관성 분석]



[빈도 분석]



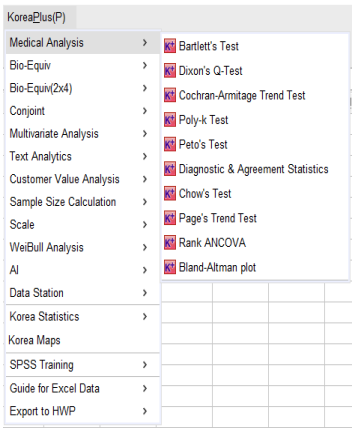
[Text2Vec]

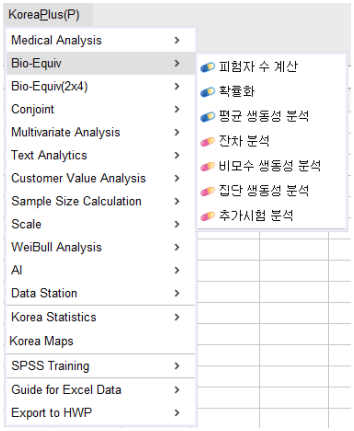
KoreaPlus Statistics

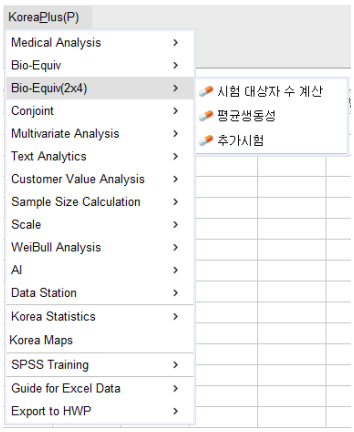
- Embedded on SPSS Statistics 27 Standard

의학 관련 분석 기법 제공

- **의학에서 주로 사용되는 고급 분석기법을 제공.**
  - ✓ SPSS 제품을 좀 더 편리하게 사용 가능하도록 대화상자 제공.
  - ✓ Medical Analysis : 생존 관련 임상 시험에서 많이 사용되는 고급분석 기능 제공.
  - ✓ Bio-Equiv : 생동성 시험의 설계와 분석을 보다 쉽고 정확하게 수행하도록 지원.
  - ✓ Bio-Equiv(2x4): 생동성 시험의 2x4 고차 교차설계 제공.

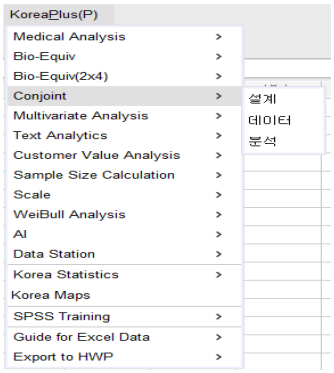


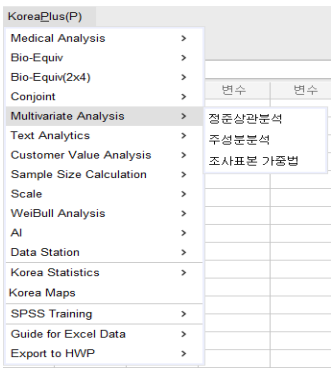




추가 분석기법 제공 I (※별도 Conjoint 모듈 구매 시, 이용 가능)

- **Conjoint, Multivariate Analysis 분석 제공.**
  - ✓ 기존에 Syntax로만 구현되었던 분석을 대화상자로 분석할 수 있도록 메뉴 제공
  - ✓ 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
  - ✓ Multivariate Analysis : 주성분분석, 정준상관분석, 조사표본가중법과 같은 다변량 분석.





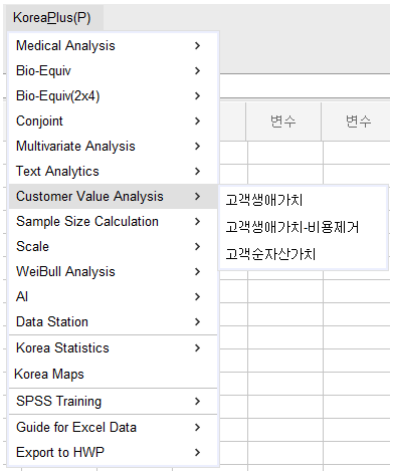
KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

추가 분석기법 제공 II

- **고객 생에 가치 분석 (Customer Life Value : CLTV) 제공.**
  - ✓ 고객생애가치 : 고객으로부터 미래의 일정기간 얻게 될 이익(=수입-비용)을 할인율에 따라 현재 가치로 환산한 재무적 가치
  - ✓ 고객 가치 분석 : 고객 개인별로 전체 생애 동안의 총 기여가치를 측정하는 방법
  - ✓ KoreaPlus에서의 고객 생애 가치 분석은 마케팅 비용을 고려하지 않고 단순히 관찰기간에 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

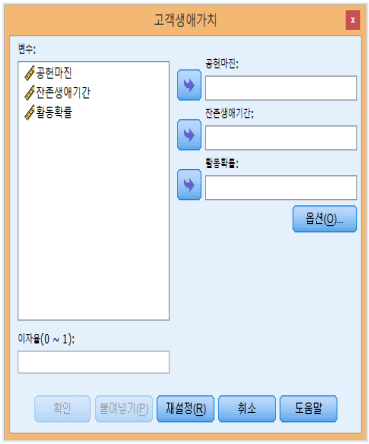
	고객ID	공헌마진	잔존생애기간	활동확률	CLTV
1	1	15.00	5	1.0	56.86
2	2	45.00	3	.5	55.95
3	3	24.00	6	.7	73.17
4	4	75.00	10	.7	322.59
5	5	9.00	8	.9	43.21
6	6	13.50	25	.4	49.02
7	7	36.00	3	.8	71.62
8	8	19.50	8	.5	52.02
9	9	69.00	11	.3	134.45
10	10	30.00	25	.2	54.46
11	11	105.00	4	.7	232.99
12	12	96.00	20	.4	326.92
13	13	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	54.04
18	18	28.50	10	.9	157.61
19	19	40.50	14	.8	238.68
20	20	45.00	23	.9	359.77

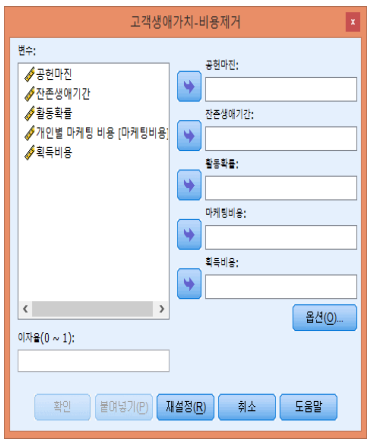


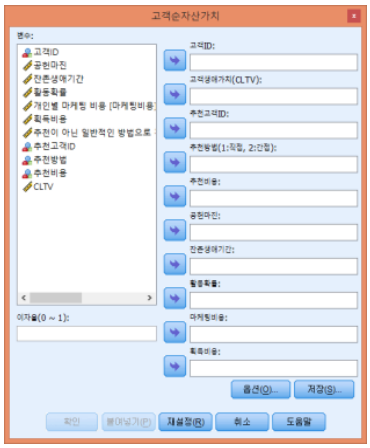
[고객생애가치]

[고객생애가치-비용제거]

[고객순자산가치]





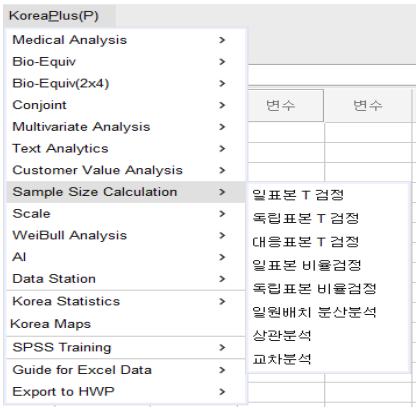


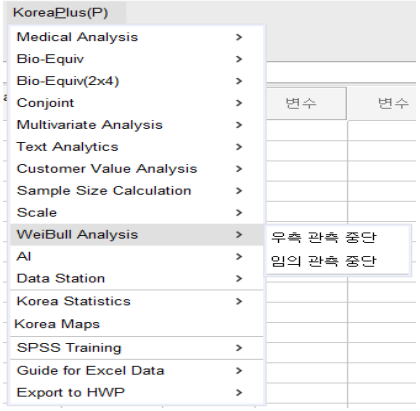
KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

추가 분석기법 제공 Ⅲ

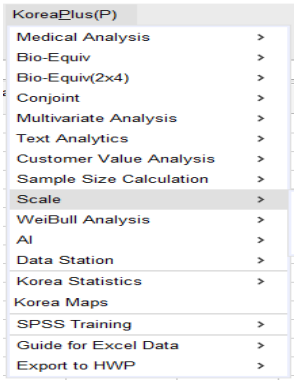
- **Sample Size Calculation과 Weibull Analysis 제공**
  - ✓ Sample Size Calculation(표본 크기 계산) : 최적의 표본 크기를 산출하는 기법. 이를 통해 통계적 검정력 확보는 물론 표본 산출을 위한 시간 및 비용 절약
  - ✓ Weibull Analysis : 가장 적절한 수명분포를 결정하고 모수를 추정하는 접근법. 주로 부품의 수명 측정에 사용되며 질병으로 인한 사람의 수명이나 제품의 제조와 운반에 걸리는 시간, 날씨 예보 등에 활용

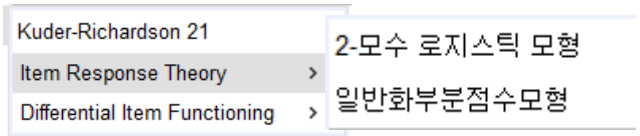


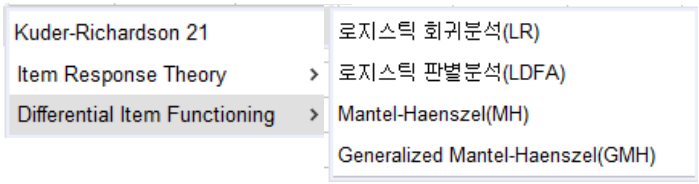


추가 분석기법 제공 Ⅳ

- **Scale 분석 기능 제공**
  - ✓ Kuder-Richardson 21 : 문항 간의 내적 일치도 정도를 나타내어주는 지수.
  - ✓ Differential Item Functioning (차별기능문항) : 각 문항이 특정 집단에 편파적으로 적용되는지 판별하는 데 사용되는 기법
  - ✓ Item Response Theory (문항반응이론) : 문항에 대한 피험자의 반응과 개인의 잠재 능력 간의 관계를 수리적 함수로 나타낸 측정 이론





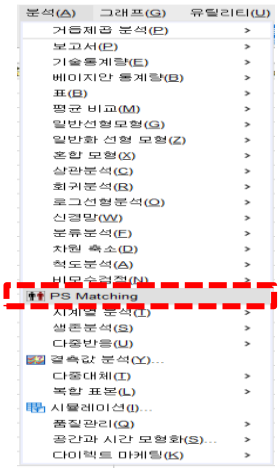


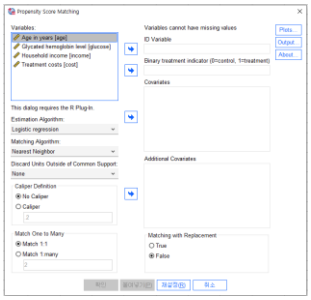
KoreaPlus Statistics

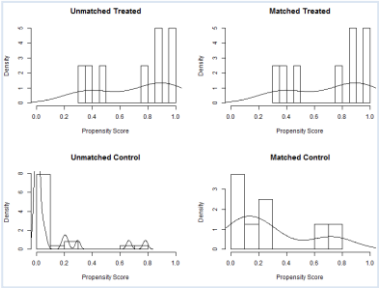
- Embedded on SPSS Statistics 27 Standard

추가 분석기법 제공 V

- **PSM(Propensity Score Matching) 기능 제공**
  - ✓ Felix Thoemmes교수가 개발한 SPSS 확장 번들로 다양한 성향점수 계산 및 매칭 기법을 제공하며, 매칭 결과를 판단할 수 있는 다양한 지표를 제공
  - ✓ 별도의 기술지원 없이 사용자가 손쉽게 설치 할 수 있는 설치 파일 제공



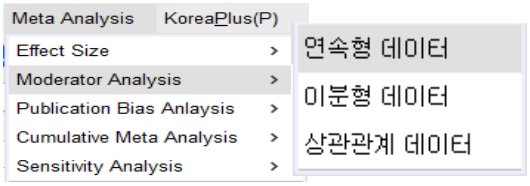


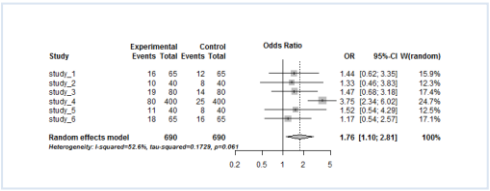


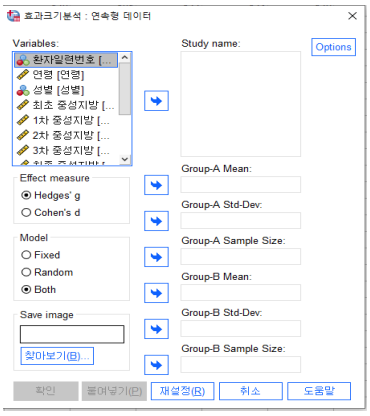
	ID_t	Treatment_t	age_t	distance_t	income_t	ps_t	psweight_t	ID_c	Treatment_c	age_c	distance_c	income_c	ps_c	psweight_c
1	1.00	1	24.00	6.00	5.50	.350	1.000	24.00	0	22.00	6.00	4.30	.047	1.000
2	2.00	1	23.00	3.00	6.10	.854	1.000	28.00	0	22.00	2.00	3.00	.216	1.000
3	3.00	1	20.00	1.00	4.50	.484	1.000	27.00	0	22.00	3.00	2.40	.064	1.000
4	4.00	1	22.00	1.00	5.30	.870	1.000	19.00	0	22.00	5.00	5.60	.297	1.000
5	5.00	1	24.00	2.00	2.40	.311	1.000	30.00	0	20.00	1.00	1.50	.038	1.000
6	6.00	1	27.00	3.00	6.30	.986	1.000	32.00	0	27.00	2.00	2.20	.663	1.000
7	7.00	1	26.00	2.00	3.30	.781	1.000	31.00	0	24.00	2.00	1.80	.193	1.000
8	8.00	1	30.00	2.00	6.50	.999	1.000	26.00	0	21.00	3.00	6.70	.781	1.000

추가 분석기법 제공 VI

- **Meta-analysis 기능 제공**
  - ✓ 동일한 주제하에 수행된 여러 연구들로부터 도출된 다양한 연구결과를 체계적으로 종합 분석하는 계량적 접근방법.
  - ✓ Effect Size, Moderator Analysis, Publication Bias Analysis, Cumulative Meta Analysis, Sensitivity Analysis 의 연속형 데이터, 이분형 데이터, 상관관계 데이터 결과 값 제공





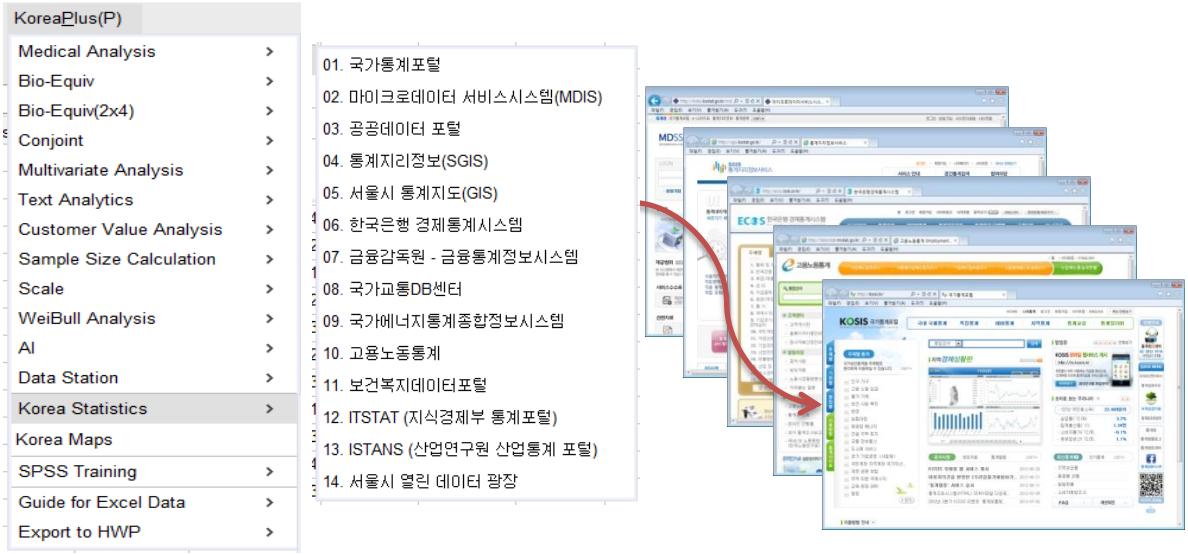


KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

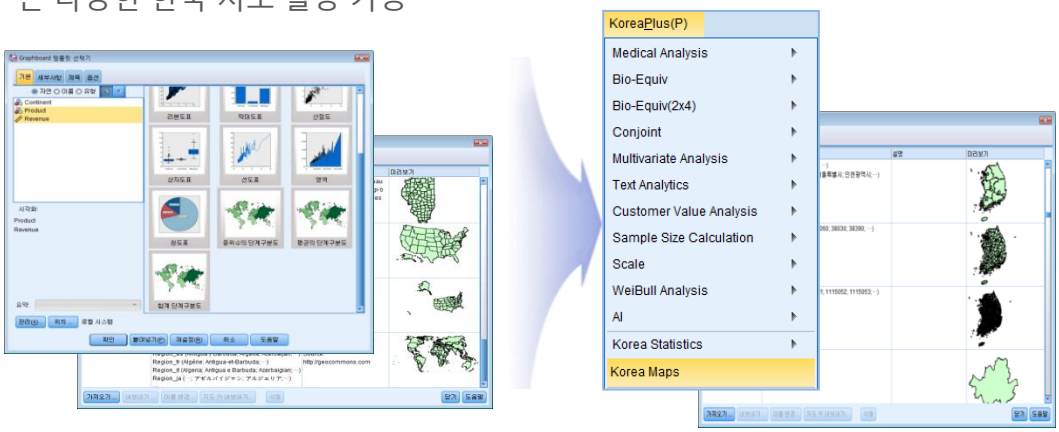
국가통계포털 링크

- **국가통계포털 링크가 추가**
  - ✓ 국가 통계 데이터를 이용하기 위한 국가 통계 제공 사이트 바로가기
  - ✓ Korea Statistics 메뉴를 통해 통계 관련 사이트(정부·공공)로 바로 이동 가능.



한국지도 제공

- **KoreaPlus Statistics Maps에 한국 지도 템플릿 추가**
  - ✓ Korea Maps 메뉴 클릭만으로 한국지도 파일생성
  - ✓ 사용자가 직접 한국 지도를 다운받지 않아도 통계지리정보사이트에서 제공하는 다양한 한국 지도 활용 가능

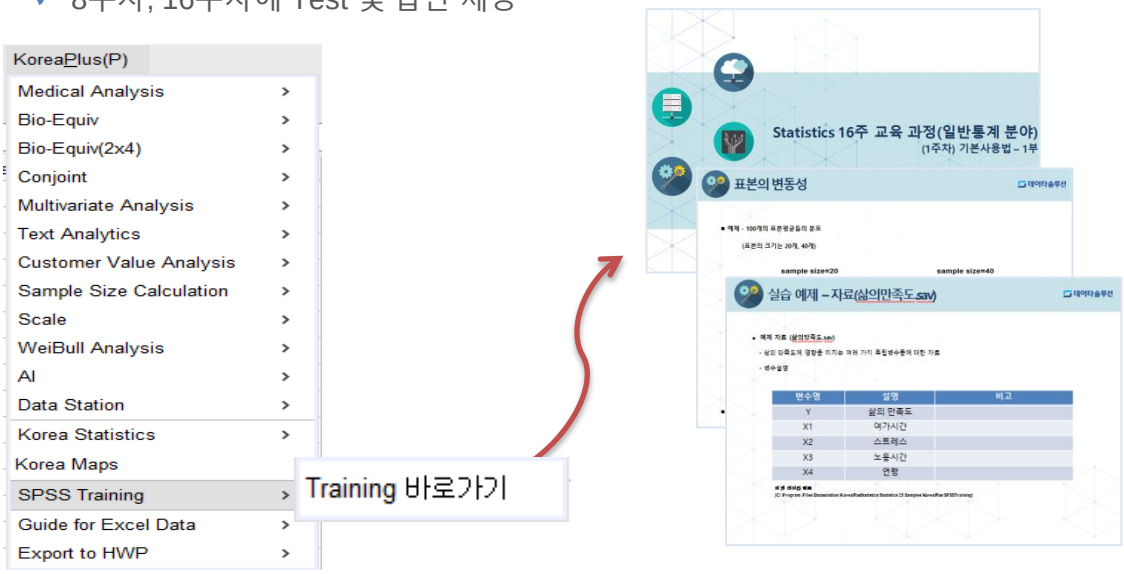


KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

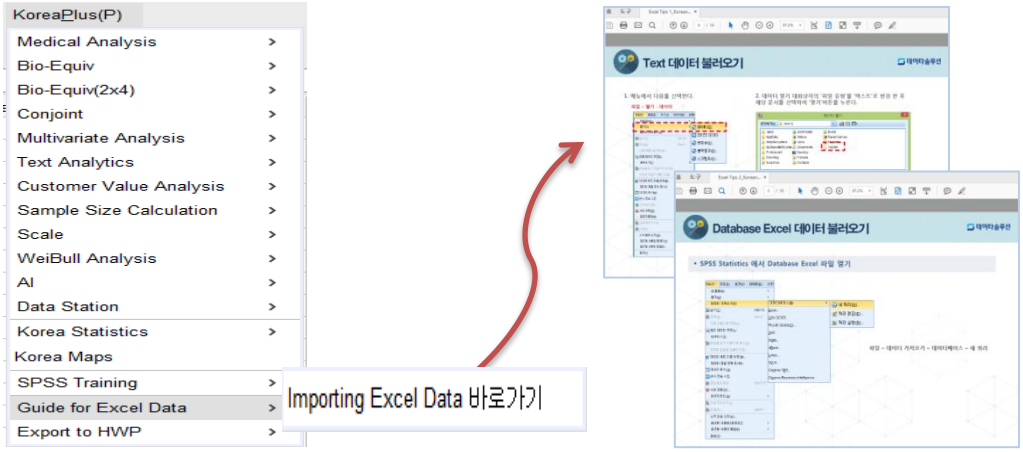
SPSS Training

- **실습 자료 및 샘플데이터 제공**
  - ✓ 초보자를 위한 16주 과정 SPSS 활용 실습 파일, 샘플 데이터 제공
  - ✓ 8주차, 16주차에 Test 및 답안 제공



엑셀 데이터 불러오기 Tips

- **엑셀 데이터 불러오기 도움말 추가**
  - ✓ 사용자가 쉽게 엑셀파일을 활용할 수 있는 Tip 제공
  - ✓ 분석 대상 데이터로 많이 쓰여지는 엑셀 데이터를 편리하고 빠르게 활용할 수 있도록 [엑셀 파일 불러오기]와 [데이터베이스에 연결된 엑셀 파일 불러오기]에 대한 도움말 제공

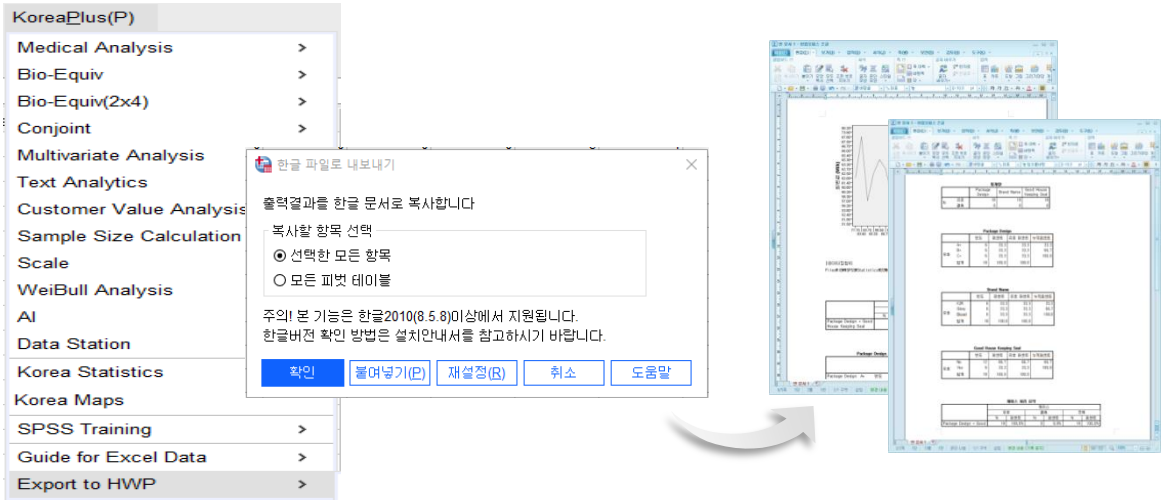


KoreaPlus Statistics

- Embedded on SPSS Statistics 27 Standard

아래한글 내보내기

- **출력결과(Output)를 아래한글(hwp)로 바로 내보내기 가능**
  - ✓ [한글 파일 내보내기] 기능으로 엑셀 및 워드 파일을 포함한 강력한 보고서 작성 가능
  - ✓ 결과값을 바로 한글 파일로 내보내어 보고서 작성 시간을 단축



기타혜택

- **SW material 제공 (DVD)**
- **제품 별 한글 매뉴얼 도서 무상 제공**