

SPSS Statistics 29/29.0.1

新機能のご紹介

IBM SPSS Statistics

日本アイ・ビー・エム株式会社

テクノロジー事業本部 データ・
AI・オートメーション事業部

Data & AI第一テクニカルセールス



本日のアジェンダ



1. SPSSとは

- SPSS
- 『IBM SPSS Statistics』って何？
- 特長と構成
- デモンストレーション
- SPSS Statisticsの優位性
- リリースヒストリー

2. SPSS Statistics 29/29.0.1 新機能

- データエディターの概要タブ
- ROC曲線のYouden Index
- Lasso, Ridge, Elastic Net回帰
- バイオリンプロット
- デモンストレーション

3. SPSSインフォメーション

- 関連書籍
- 紹介動画
- ブログ・コミュニティ・サポート

1.SPSSとは

SPSS

Statistical **P**ackage for the **S**ocial **S**cienceは、
1968年 アメリカ合衆国イリノイ州シカゴで
社会科学のための**統計パッケージ**として誕生し、
その後、医療やマーケティングなど**さまざまな分野**で活用され
世界中の教育・研究機関、医療機関や企業で
利用されています。



IBM SPSS Statistics

『IBM SPSS Statistics』って何？



統計解析ツールです!!

統計解析の主な役割

比較

新ワクチンと現ワクチンを接種したグループで抗体量を**比べる**
⇒ **新ワクチンには効き目があった!**

分類

アンケートの回答が似ている人を同じクラスタに**分ける**
⇒ **特徴が異なる3つのクラスタが見つかった!**

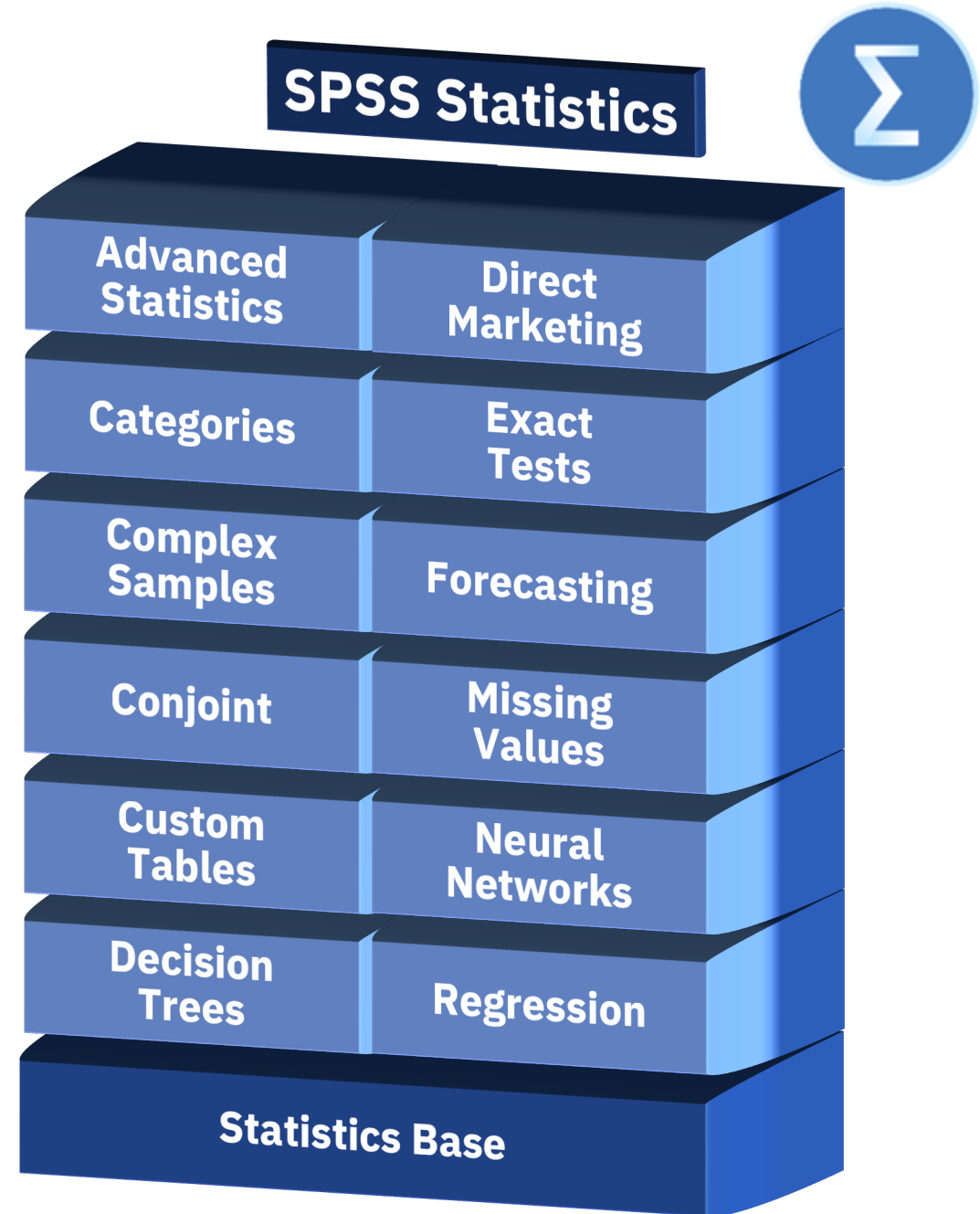
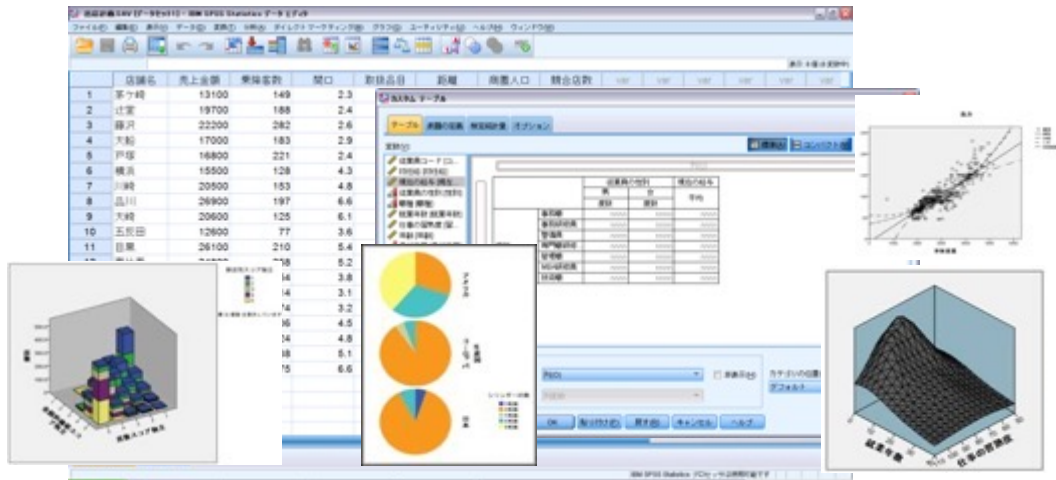
予測

賃貸物件の家賃を駅からの距離、間取り、築年数から**当てる**
⇒ **家賃に何が影響しているかわかった!**

特長と構成

統計解析のスタンダードソフトウェア

- Excel、テキスト形式、データベースなど**複数データ形式**の読み込み
- **データ加工機能・集計機能・視覚化機能**を搭載
- 仮説検証に必要な様々な統計の**検定手法**を搭載
- 高度な**多変量解析手法**を多数搭載
- **使いやすい**ユーザーインターフェース（高い操作性）
- **クライアント版**、大量データ処理用の**サーバー版**を用意
- **個人ユーザー**から、**ビジネスユーザー**まで対応
- 企業、官公庁、教育機関で**世界的に普及**



a Statistics データ読み込み・入力



a
読み込み

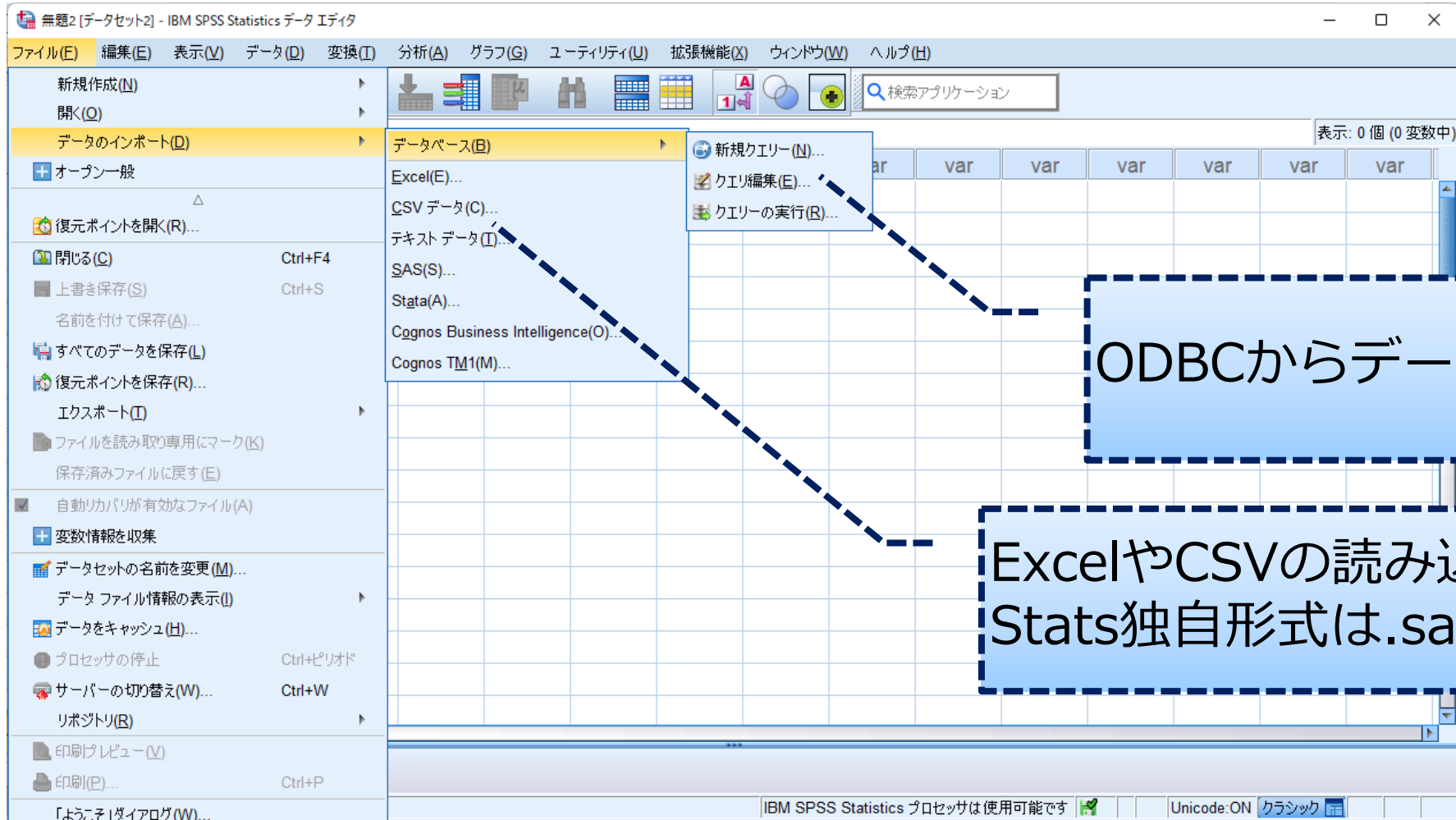
b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析



ODBCからデータベース

ExcelやCSVの読み込み
Stats独自形式は.sav

b Statistics データ表示



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

demo.sav [データセット1] - IBM SPSS Statistics データエディタ

ファイル(E) 編集(E) 表示(V) データ(D) 変換(I) 分析(A) グラフ(G) ユーティリティ(U) 拡張機能(X) ウィンドウ(W) ヘルプ(H)

検索アプリケーション

表示: 29 個 (29 変数中)

	年齢	婚姻状況	住居年数	収入	収入カテゴリー	車	車カテゴリー	教育	雇用	退職	雇用カテゴリー	仕事満足	
1	55	既婚	12	72.00	50 - 74	36.20	高い	中学	23	在職	15年以上	非常に満足	女性
2	56	未婚	29	153.00	75以上	76.90	高い	中学	35	在職	15年以上	やや満足	男性
3	28	既婚	9	28.00	25 - 49	13.70	安い	専門学校	4	在職	5年未満	普通	女性
4	24	既婚	4	26.00	25 - 49	12.50	安い	大学	0	在職	5年未満	非常に不満	男性
5	25	未婚	2	23.00	25未満	11.30	安い	高校	5	在職	5年以上 ...	やや不満	男性
6	45	既婚	9	76.00	75以上	37.20	高い	専門学校	13	在職	5年以上 ...	やや不満	男性
7	42	未婚	19	40.00	25 - 49	19.80	普通	専門学校	10	在職	5年以上 ...	やや不満	男性
8	35	未婚	15	57.00	50 - 74	28.20	普通	高校	1	在職	5年未満	非常に不満	女性
9	46	未婚	26	24.00	25未満	12.20	安い	中学	11	在職	5年以上	非常に満足	女性
10	34	既婚	0	89.00	75以上	46.10	高い	専門学校	12				
11	55	既婚	17	72.00	50 - 74	35.50	高い	専門学校	2				
12	28	未婚	3	24.00	25未満	11.80	安い	大学	4				
13	31	既婚	9	40.00	25 - 49	21.30	普通	大学	0				
14	42	未婚	8	137.00	75以上	68.90	高い	専門学校	3				
15	35	未婚	8	70.00	50 - 74	34.10	高い	専門学校	9	在職	5年以上 ...	やや満足	男性
16	52	既婚	24	159.00	75以上	78.90	高い	大学	16	在職	15年以上	非常に満足	男性

データ ビュー 変数 ビュー

IBM SPSS Statistics プロセッサは使用可能です

Unicode:ON クラシック

スプレッドシート形式

c Statistics 基礎統計



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

The screenshot shows the IBM SPSS Statistics interface. On the left, a data grid displays variables: 年齢 (Age), 婚姻状況 (Marital Status), and 住居年数 (Residence Years). The '分析' (Analyze) menu is open, with '記述統計' (Descriptive Statistics) selected. The '度数分布表' (Frequency Distribution) sub-menu is also open. On the right, the '出力' (Output) window shows a tree view with '度数分布表' selected. The main window displays the '度数分布表' (Frequency Distribution) output for '世帯全体の収入カテゴリ (千ドル)' (Household Income Category in thousands of dollars).

統計量
世帯全体の収入カテゴリ (千ドル)

度数	有効	6400
	欠損値	0

世帯全体の収入カテゴリ (千ドル)

	度数	パーセント	有効パーセント	累積パーセント
有効	25未満	1174	18.3	18.3
	25 - 49	2388	37.3	55.7
	50 - 74	1120	17.5	73.2
	75以上	1718	26.8	100.0
合計	6400	100.0	100.0	

世帯全体の収入カテゴリ (千ドル)

IBM SPSS Statistics プロセッサは

度数分布表と記述統計

d Statistics グラフ作成



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

単純箱ひげ図の定義: グループごとの集計

変数(V): 勤続年数 [雇用]

カテゴリ軸(X): 仕事に対する満足度 [仕事満足]

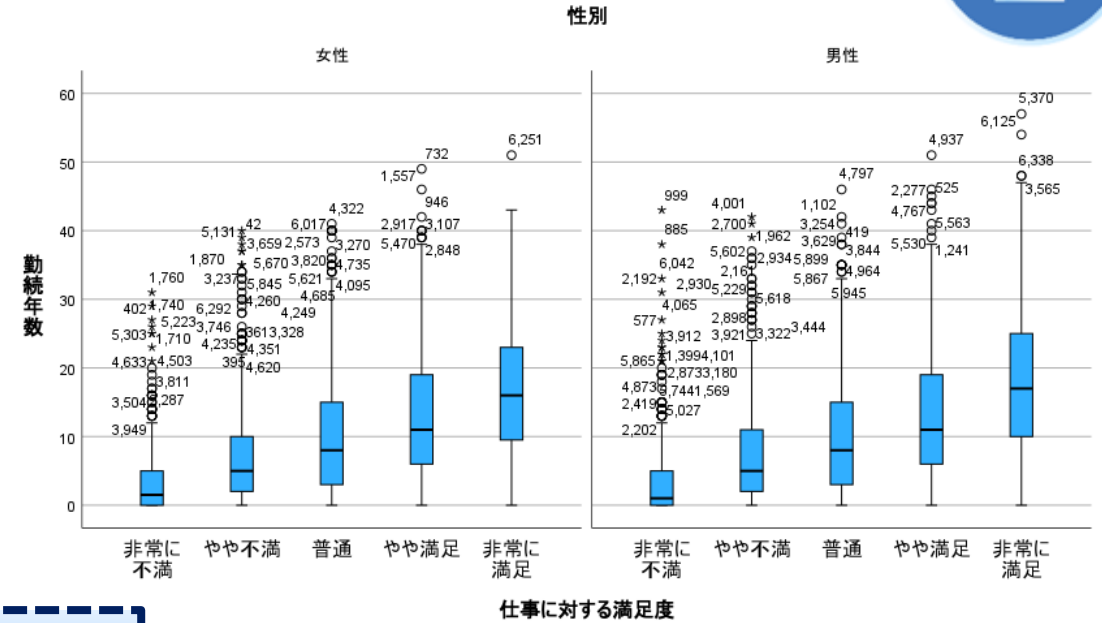
ケースのラベル(E):

行(W):

列(L): 性別 [性別]

OK 貼り付け(B)

箱ひげ図

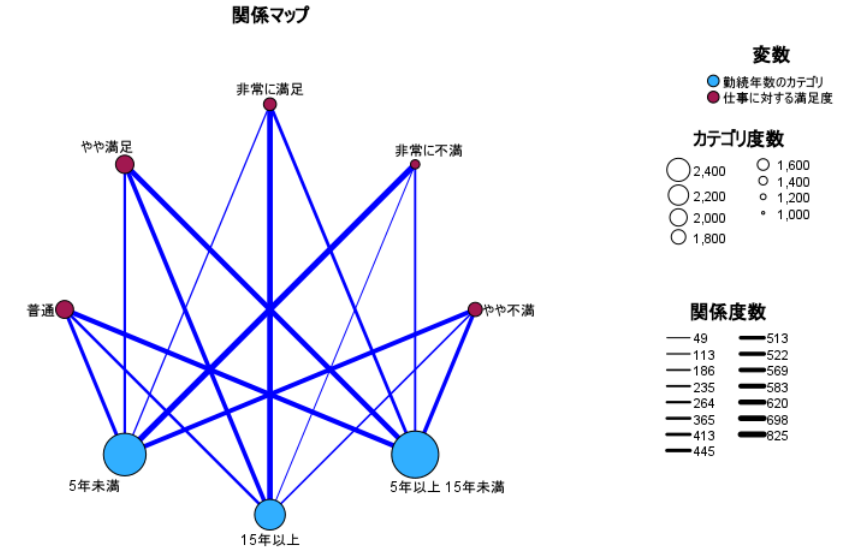


関係マップ

変数(V): 仕事に対する満足... 勤続年数のカテゴリ...

OK 貼り付け(B) 戻す(R) キャンセル ヘルプ

関連マップ



e Statistics データ加工



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

変数の計算

変数の計算(C)...

変数プロパティの定義(V)...

不明の測定の尺度を設定(L)...

データプロパティのコピー(C)

新しいユーザー指定の属性(B)...

日付と時刻を定義(E)...

多重回答グループの定義(M)...

検証(L)

重複ケースの特定(U)...

例外ケースの特定(I)...

データセットの比較(P)...

ケースの並べ替え(Q)...

変数の並べ替え(B)...

行と列の入れ換え(N)...

ファイル間での文字列幅の調整

ファイルの結合(G)

再構成(R)...

Propensity Score Matching...

傾斜重み付け...

傾向スコアによる一致...

ケースコントロールの一致...

グループ集計(A)...

直交計画(H)

複数のファイルに分割

データセットをコピー

ファイルの分割(E)

ケースの選択(S)...

ケースの重み付け(W)...

シフト値(E)...

同一の変数への値の再割り当

他の変数への値の再割り当て

連続数への再割り当て(A)...

ゲーム変数を作成

ゲーム変数を作成

連続変数のカテゴリ化(B)...

最適カテゴリ化(I)...

モデル作成のデータ準備(P)

ケースのランク付け(K)...

日付と時刻ウィザード(Q)...

時系列の作成(M)...

欠損値の置き換え(V)...

乱数ジェネレータ(G)...

保留された変換の実行

型とラベル(L)...

年齢 [年齢]

婚姻状況 [婚姻状況]

住居年数 [住居年数]

世帯全体の収入 (千ドル) [...]

世帯全体の収入カテゴリ (千...)

車の価格 [車]

車の価格カテゴリ [車カテゴリ]

教育のレベル [教育]

勤続年数 [雇用]

退職区分 [退職]

勤続年数のカテゴリ [雇用カテ...

仕事に対する満足度 [仕事...]

性別 [性別]

世帯の人数 [世帯人数]

ワイヤレスサービス [ワイヤレ]

複数回線 [複数回線]

ボイスメール [ボイス]

ポケベルサービス [ポケベル]

インターネット [インターネット]

IF()... (任)

数式(E):

入社年齢 = 年齢 - 雇用

関数グループ(G):

すべて

算術

CDFと非心度 CDF

変換

現在の日付と時刻

算術日

日付作成

関数と特殊変数(E):

\$Casenum

\$S

\$S

\$S

\$S

\$Time

Abs

Any

Applymodel

Arsin

フィルター: 説明を含める

新変数作成

変数作成ビルダ

メニュー

年齢	雇用	入社年齢
55	23	32.00
56	35	21.00
28	4	24.00
24	0	24.00
25	5	20.00
45	15	32.00
42	10	32.00
35	1	34.00
46	11	35.00
34	12	22.00
55	2	53.00
28	4	24.00

f-1 Statistics 統計解析



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

グループ統計量

	性別	度数	平均値	標準偏差	平均値の標準誤差
就学年数	男性	1225	13.36	2.966	.085
	女性	1595	13.17	2.896	.073

t検定

有意確率が0.05より高い場合は「等分散性を仮定する」をそでない場合は「等分散性を仮定しない」行の値を参照する

独立サンプルの検定

	等分散性のための Levene の検定	2つの母平均の差の検定									
		F 値	有意確率	t 値	自由度	有意確率		平均値の差	差の標準誤差	差の95%信頼区間	
						片側 p 値	両側 p 値			下限	上限
就学年数	等分散を仮定する	2.064	.151	1.738	2818	.041	.082	.193	.111	-.025	.411
	等分散を仮定しない			1.733	2601.434	.042	.083	.193	.112	-.025	.412

独立サンプルの効果サイズ

	Standardizer ^a	ポイント推定	95% 信頼区間		
			下限	上限	
就学年数	Cohen の d	2.926	.066	-.008	.141
	Hedges の補正	2.927	.066	-.008	.140
	Glass のデルタ	2.896	.067	-.008	.141

有意確率が0.05未満で2群間平均値の間に有意差があることを示す

比較

- a. 効果サイズの推定に使用する分母。
Cohen の d は、プールされた標準偏差を使用します。
Hedges の補正は、プールされた標準偏差と補正係数を使用します。
Glass's のデルタは、制御グループのサンプル標準偏差を使用します。

f-2 Statistics 統計解析



a
読み込み

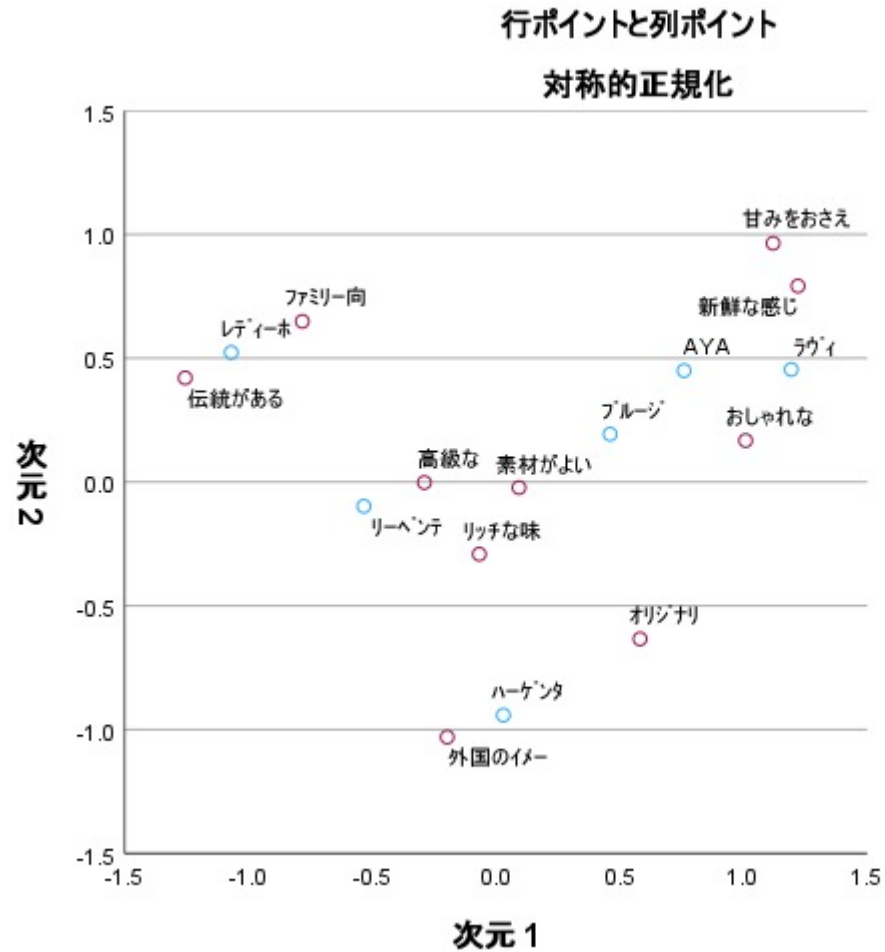
b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析



ブランド名
評価項目

コレスポネンス分析

分類

f-3 Statistics 統計解析



a
読み込み

b
データ表示

c
基礎統計

d
グラフ作成

e
データ加工

f
統計解析

モデルの要約

モデル	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.984 ^a	.967	.959	9.59838

a. 予測値: (定数)、店の間口, 取扱品目, 乗降客数。

回帰分析

分散分析^a

モデル		平方和	自由度	平均平方	F 値	有意確率
1	回帰	30124.982	3	10041.661	108.996	<.001 ^b
	残差	1013.418	11	92.129		
	合計	31138.400	14			

a. 従属変数 販売金額

b. 予測値: (定数)、店の間口, 取扱品目, 乗降客数。

係数^a

モデル		非標準化係数		標準化係数	t 値	有意確率
		B	標準誤差	ベータ		
1	(定数)	5.169	11.210		.461	.654
	取扱品目	.521	.046	.642	11.353	<.001
	乗降客数	12.646	2.635	.379	4.800	<.001
	店の間口	.284	.059	.385	4.821	<.001

a. 従属変数 販売金額



SPSS Statisticsの優位性



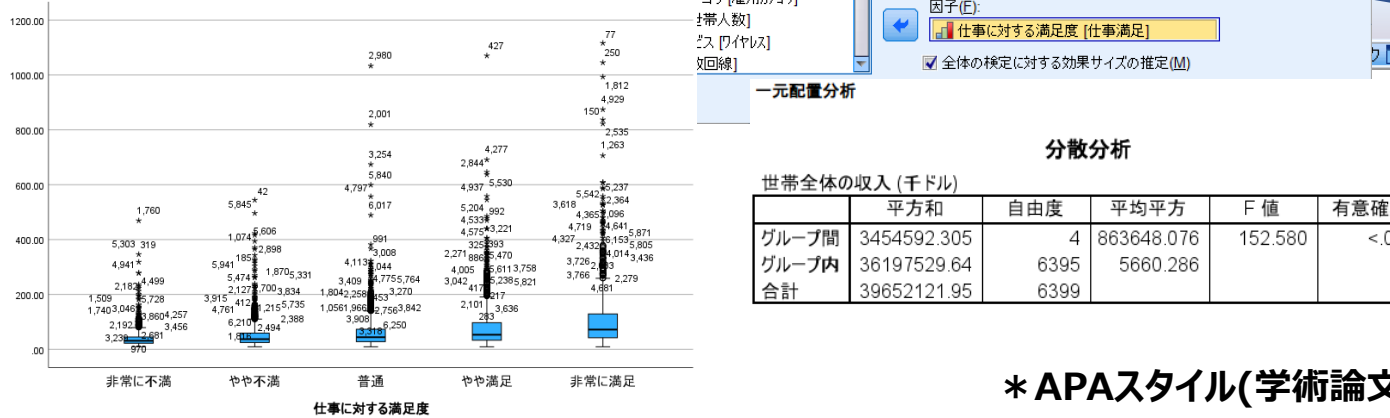
IBM SPSS Statistics データエディタ

年齢	婚姻状況	住居年数	車	車カテゴリ	教育	雇用	退職	雇用カテゴリ	仕事に対する満足度
1	55	1	12						
2	56	0	29						
3	28	1	9						
4	24	1	4						
5	25	0	2						
6	45	1	9						
7	42	0	19						
8	35	0	15						
9	46	0	26						
10	34	1	0						
11	55	1	17						
12	28	0	3						
13	31	1	9						
14	42	0	8						
15	35	0	8						
16	52	1	24						

① **コマンド不要**でわかりやすい
ユーザーインターフェース

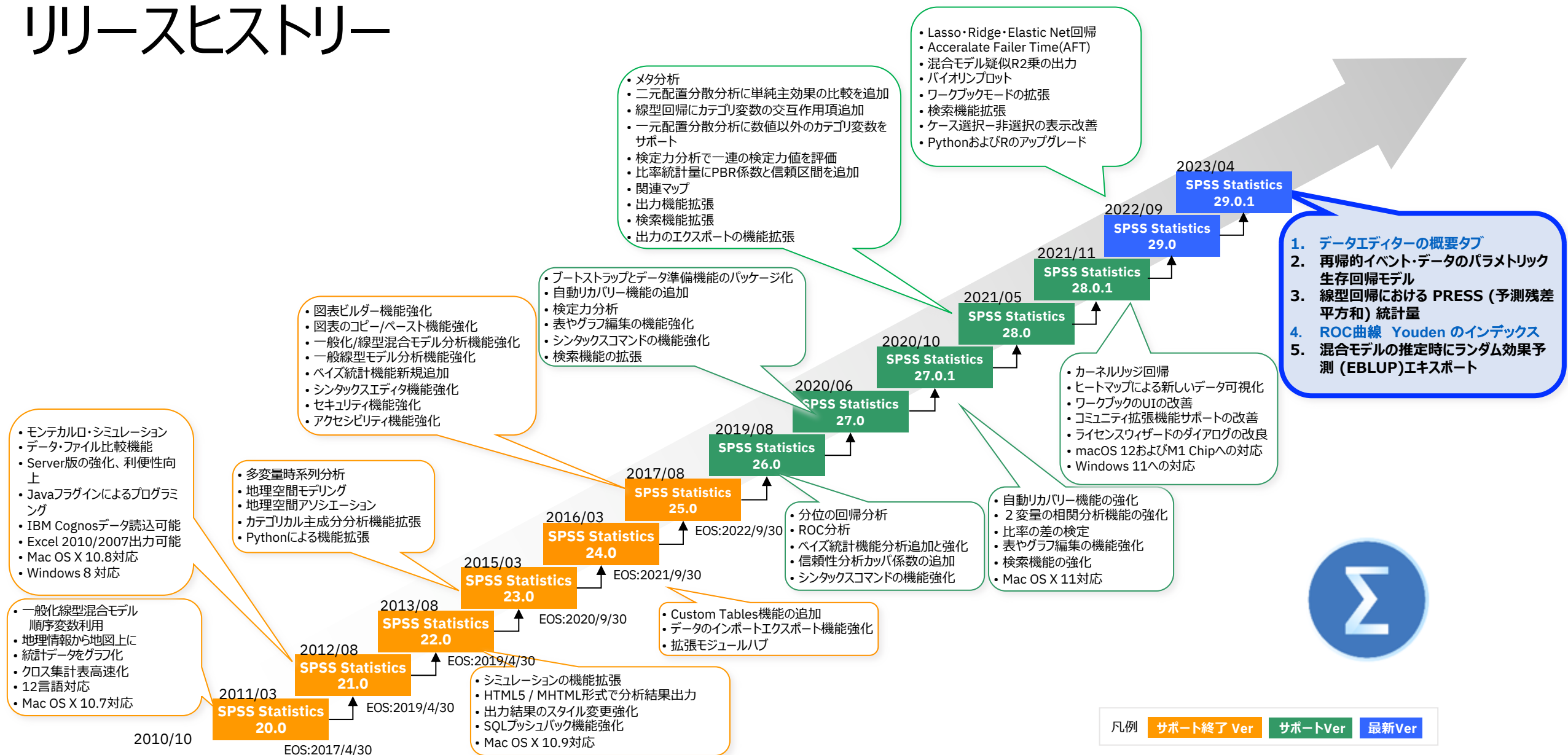
② **豊富で高度な統計解析と**
多様なデータ加工メニュー

③ **論文*やプレゼン**に適した
洗練されたアウトプット



* APAスタイル(学術論文などのアカデミックな分野の文書に用いられる書式)の選択が可能

リリースストーリー



2.SPSS Statistics 29/29.0.1

新機能

SPSS Statistics 29.0.1 | 新機能

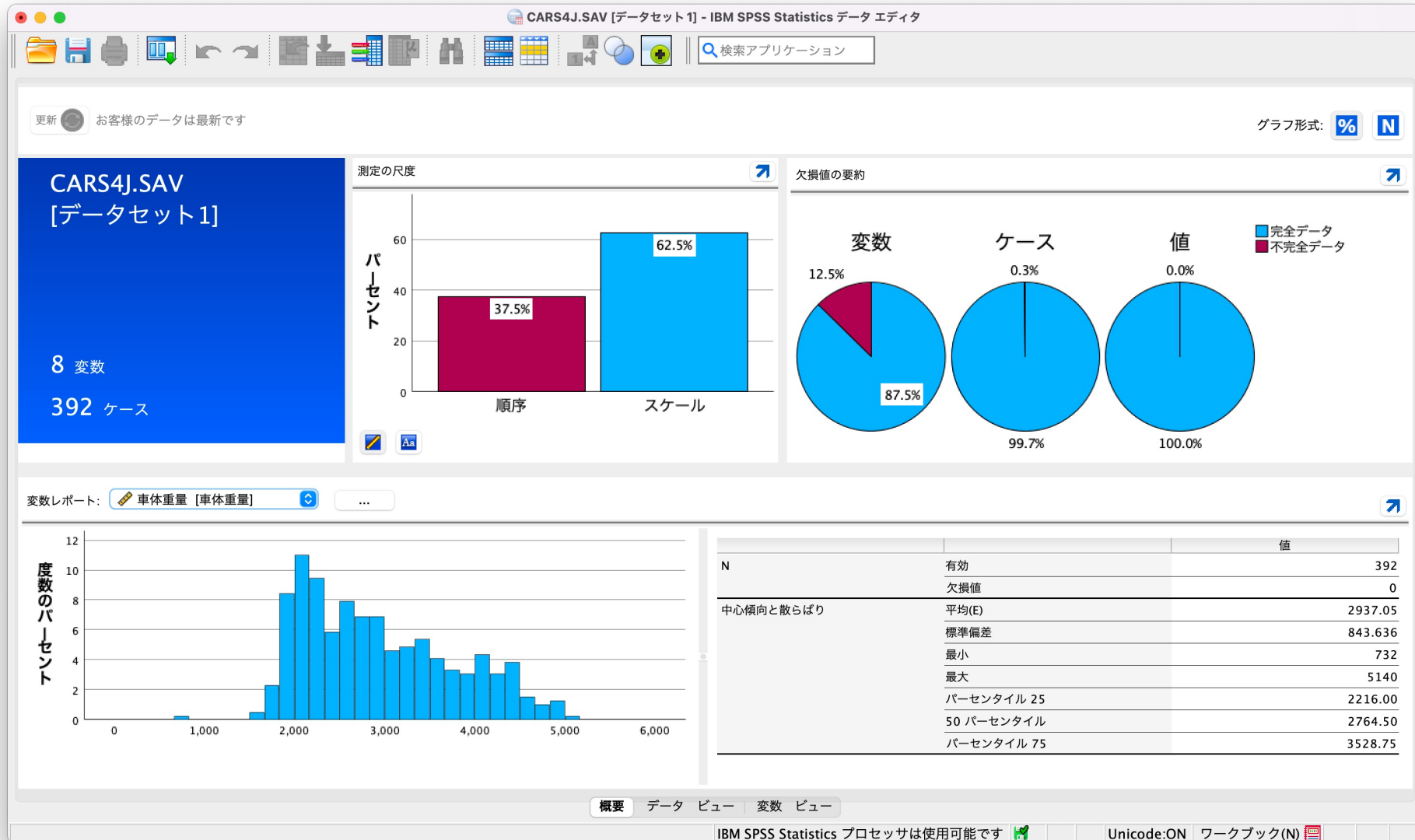
<https://www.ibm.com/docs/ja/spss-statistics/29.0.0?topic=overview-whats-new-in-version-29>

1. データエディターの概要タブ
2. 再帰的イベント・データのパラメトリック生存回帰モデル
3. 線型回帰における PRESS (予測残差平方和) 統計量
4. ROC曲線 Youden のインデックス
5. 混合モデルの推定時にランダム効果予測 (EBLUP)エクスポート



SPSS Statistics 29.0.1 | データエディタの概要タブ

- 読み込んだデータの概要を把握することができます



SPSS Statistics 29.0.1 | Youden インデックス

- ROC曲線にYouden indexが加わりました。カットオフポイントを特定に利用します。

<https://qiita.com/416nishimaki/items/be93f71354b6cd26f2ba>

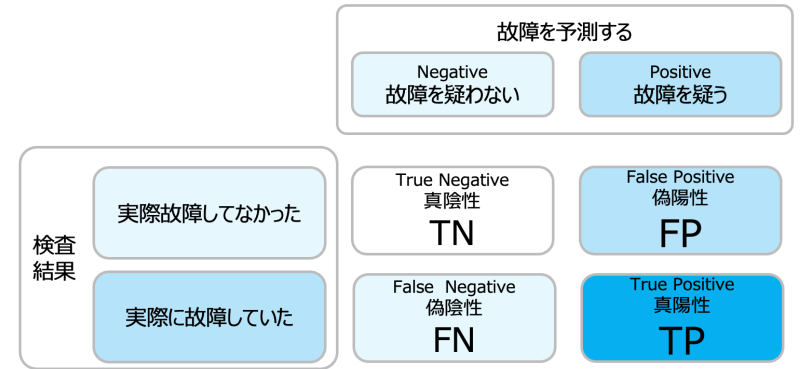
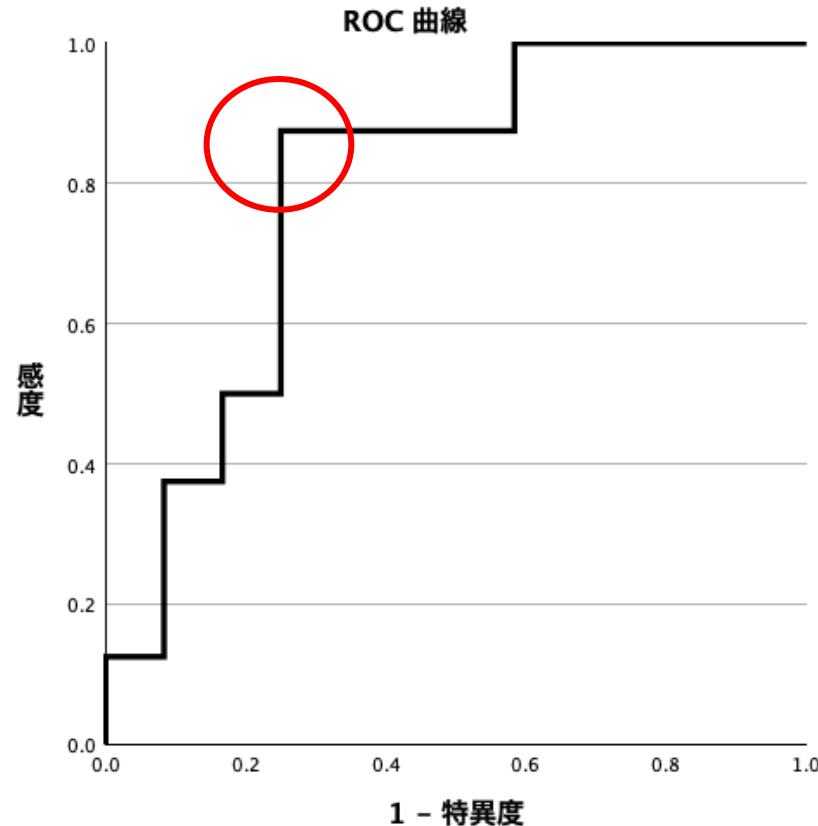
曲線の座標

検定結果変数: スコア

より大きいか、
等しい場合に正

a	感度	1 - 特異度	ヨーデン指標
-1.0000	1.000	1.000	.000
.0250	1.000	.917	.083
.0750	1.000	.833	.167
.1250	1.000	.750	.250
.1750	1.000	.667	.333
.2250	1.000	.583	.417
.2750	.875	.583	.292
.3250	.875	.500	.375
.3750	.875	.417	.458
.4250	.875	.333	.542
.4750	.875	.250	.625
.5250	.750	.250	.500
.5750	.625	.250	.375
.6250	.500	.250	.250
.6750	.500	.167	.333
.7250	.375	.167	.208
.7750	.375	.083	.292
.8250	.250	.083	.167
.8750	.125	.083	.042
.9250	.125	.000	.125
1.9500	.000	.000	.000

a. 最小の分割値は最小観測検定値から 1 を引いた値で、最大分割値は最大観測検定値に 1 を足した値です。他のすべての分割値は 2 つの連続する順序観測検定値の平均値です。



感度
Sensitivity
= 再現率

故障を見逃さない力

$$= \frac{TP}{FN + TP}$$

特異度
Specificity

故障を過度に疑わない

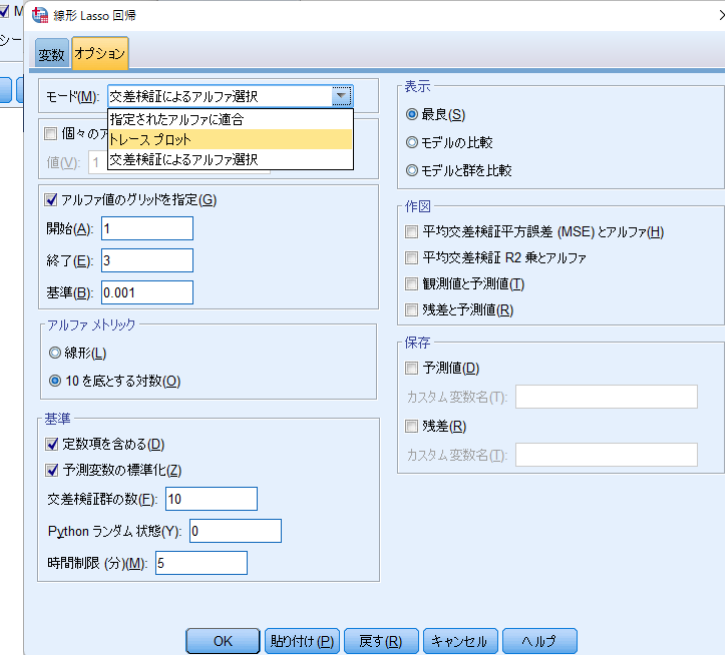
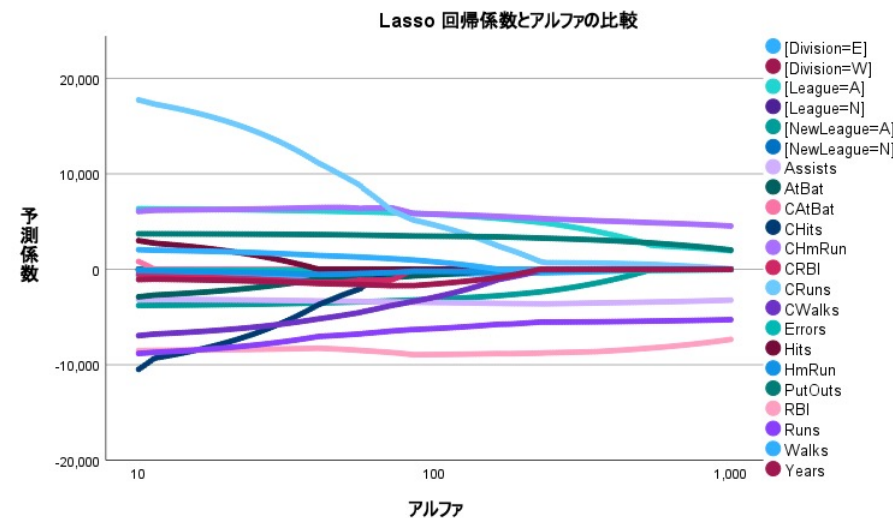
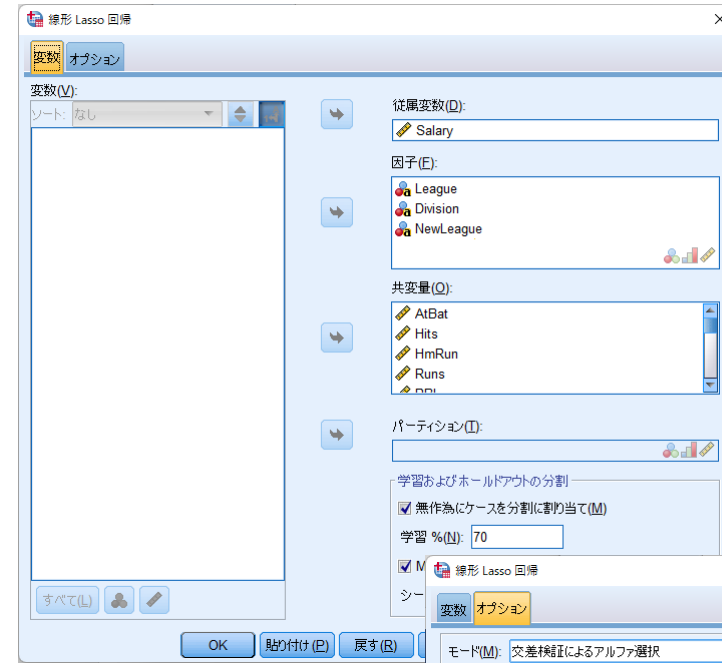
$$= \frac{TN}{TN + FP}$$

- 1. 回帰メニュー：線形回帰OLS(Ordinary Least Squares:最小二乗法)代替**
 - Lasso(ラッソ)回帰
 - Ridge(リッジ)回帰
 - Elastic Net(エラスティック ネット)回帰
- 2. 生存分析メニュー：Accelerated Failure Time (AFT) モデル**
- 3. 混合モデル(線型・一般化線型)メニュー：疑似R²乗指標**
- 4. グラフボードテンプレート選択メニュー：バイオリンプロット**
- 5. ユーザビリティの向上**
 - ワークブック機能の強化
 - 検索機能の強化
 - ケース選択の非選択ケースの改善
 - PythonとRのアップグレード



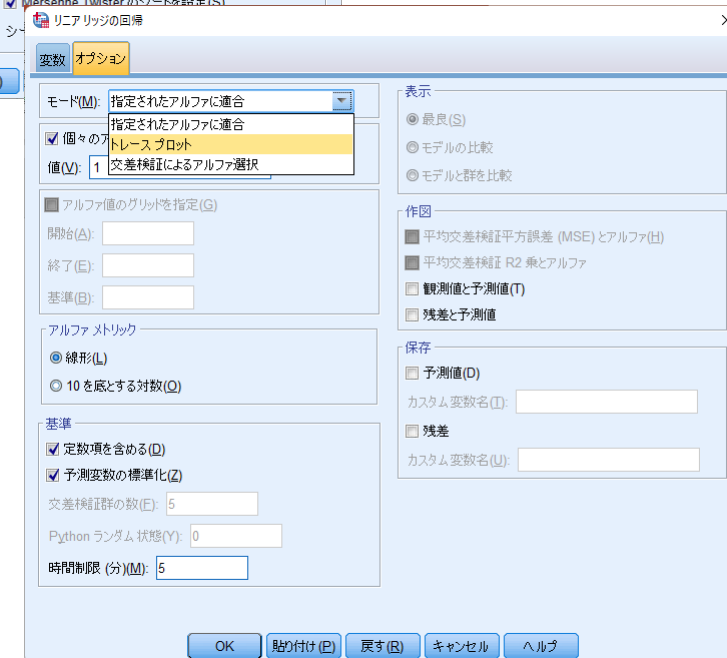
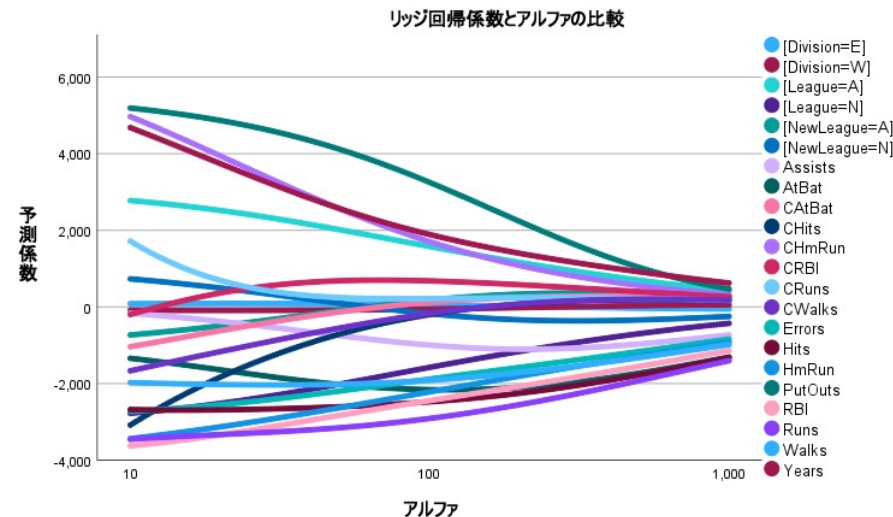
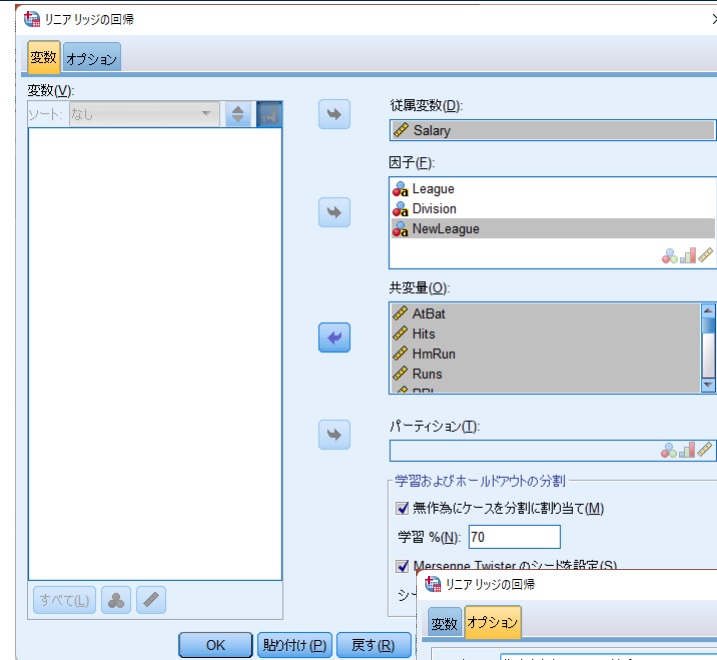
SPSS Statistics 29 | Lasso回帰

- 新しい線形回帰Lasso(ラッソ)の機能は、従属変数に対する複数の独立変数のL1正則化により線形回帰モデルを推定します。
 - トレース・プロットを表示するモードとクロスバリデーションに基づいてアルファ・ハイパーパラメータ値を選択するモードがあります。
 - 単一のモデルを適合させた場合、またはアルファ値を選択するためにクロスバリデーションを使用した場合、ホールドアウトデータのパーティションを使用して、サンプル外のパフォーマンスを推定することができます。



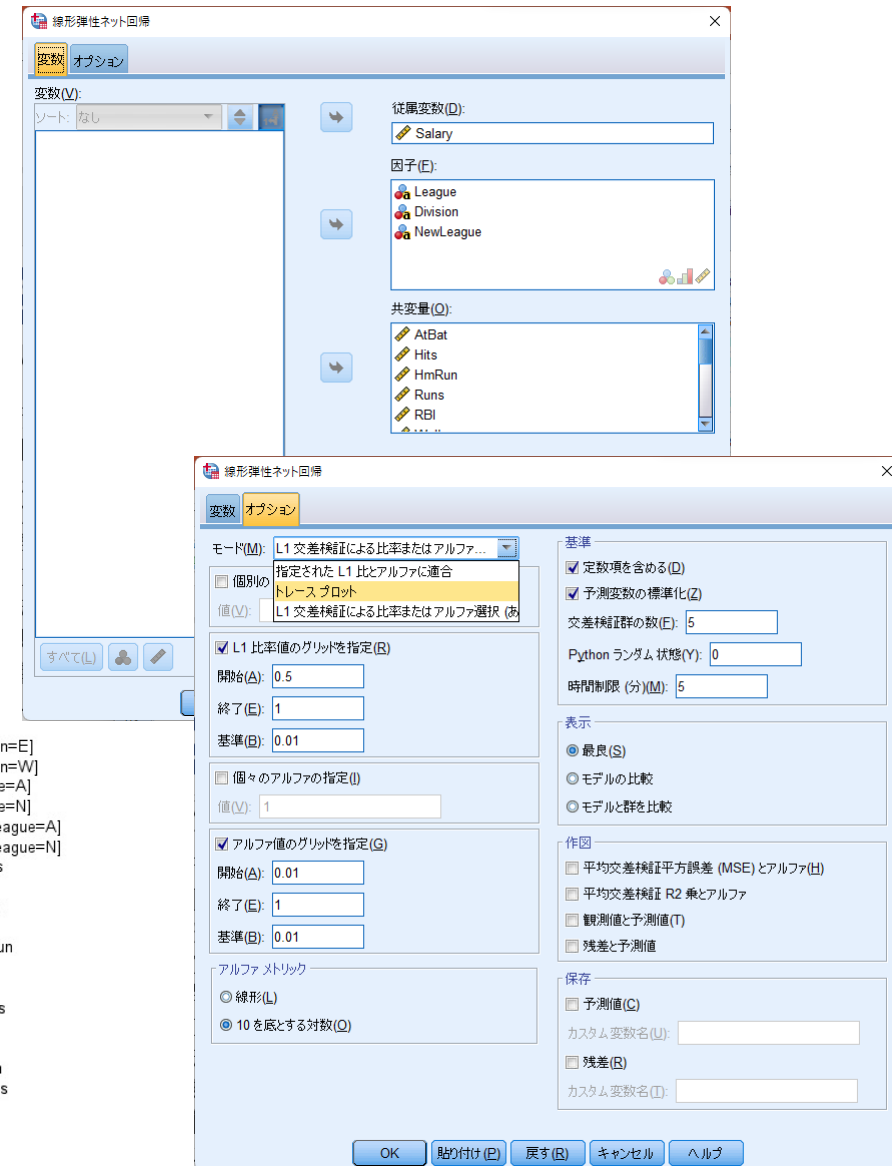
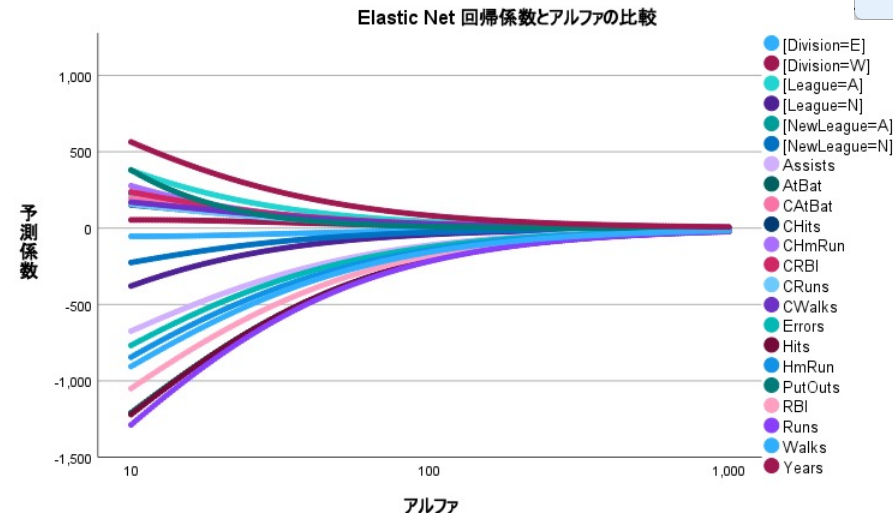
SPSS Statistics 29 | Ridge回帰

- 新しい線形回帰Ridge(リッジ)の機能は、従属変数に対する複数の独立変数のL2正則化により線形回帰モデルを推定します。
 - トレース・プロットを表示するモードとクロスバリデーションに基づいてアルファ・ハイパーパラメータ値を選択するモードがあります。
 - 単一のモデルを適合させた場合、またはアルファ値を選択するためにクロスバリデーションを使用した場合、ホールドアウトデータのパーティションを使用して、サンプル外のパフォーマンスを推定することができます。



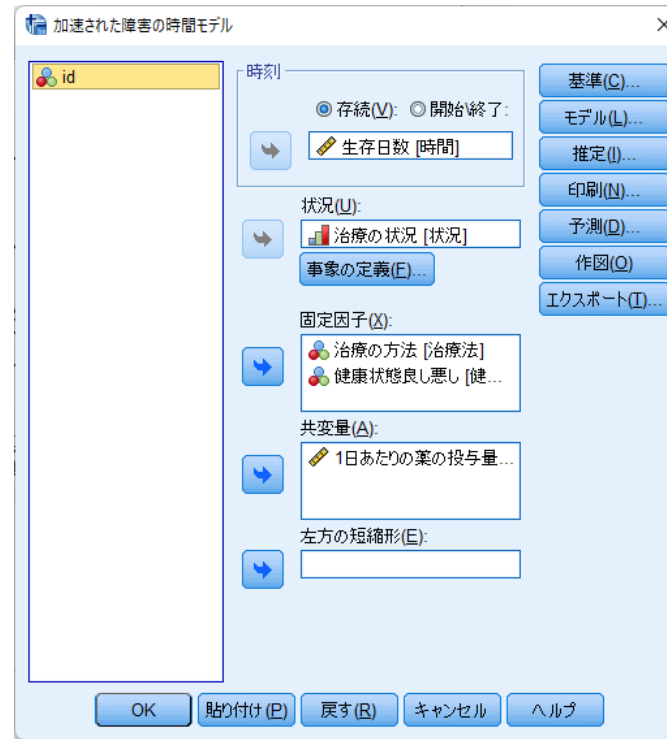
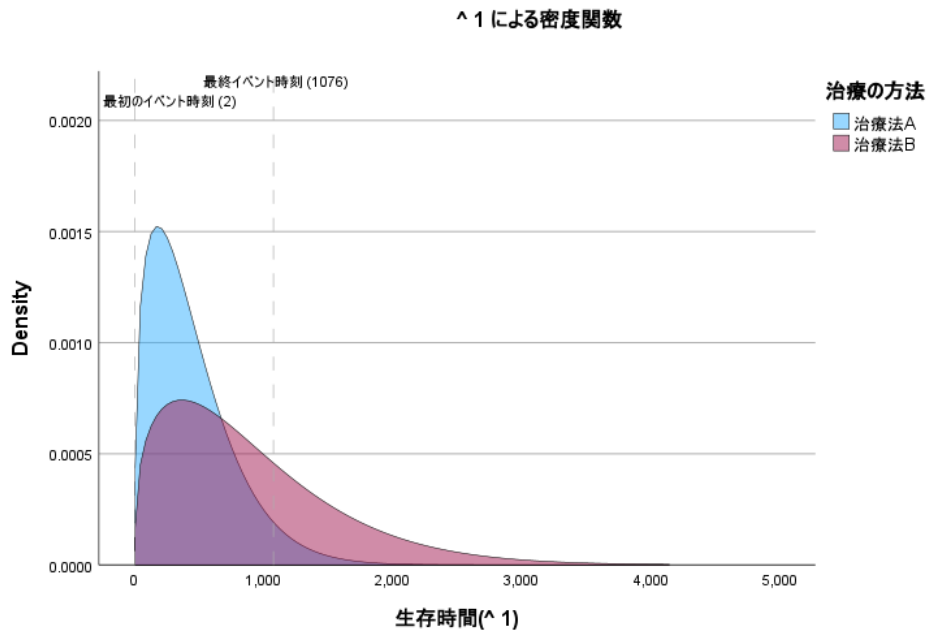
SPSS Statistics 29 | Elastic Net回帰

- 新しい線形回帰Elastic Net(エラスティック ネット)の機能は、従属変数に対する複数の独立変数の正則化を行った線形回帰モデルを推定します。正則化は、L1(Lasso)とL2(Ridge)のペナルティーを組み合わせます。
 - この拡張機能には、与えられたL1比に対するアルファの異なる値に対するトレース・プロットを表示するモードとクロスバリデーションに基づいてL1比とアルファのハイパーパラメータ値を選択するモードがオプションで用意されています。
 - 単一のモデルを適合させた場合、またはクロスバリデーションによってペナルティ比やアルファ値を選択した場合、ホールドアウトデータのパーティションを使用して、サンプル外のパフォーマンスを推定することができます。



SPSS Statistics 29 | Accelerated Failer Time(AFT)モデル

- この新しい機能は、打ち切りのある生存データを用いてパラメトリックの生存モデルを作成します。パラメトリック生存モデルは、生存時間が既知の分布に従うことを仮定し、そのモデル効果が生存時間に関して比例する加速故障(死亡)時間モデルを適合させます。



SPSS Statistics 29 | 疑似R²乗指標

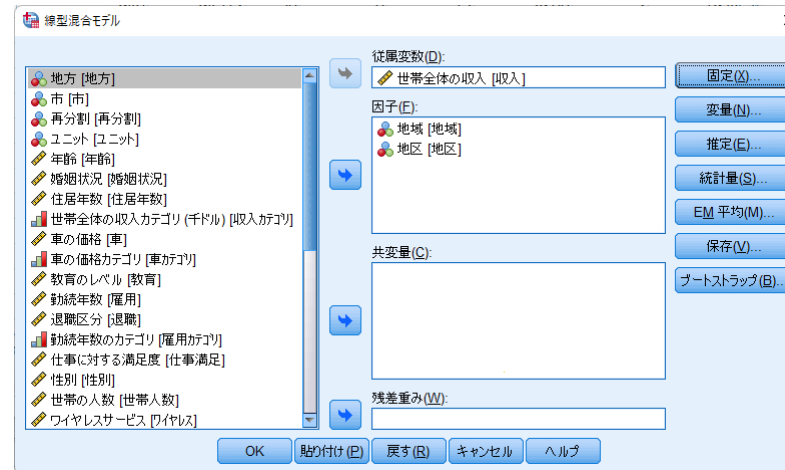
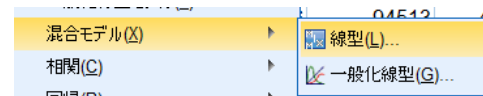
- 擬似R²指標と級内相関係数が、線形混合モデルおよび一般化線形混合モデルの出力に含まれるようになりました。
 - 決定係数R²は、線形モデルによって説明される分散の比率を表すため、よく報告される統計量です。
 - 級内相関係数(ICC: Intraclass Correlation Coefficients)は、マルチレベル/階層データでのグループ化(変量)因子によって説明される分散の比率を定量化する関連統計量です。

決定係数

疑似 R ² 乗法	境界	
	境界	.060
	条件式 (If-Then)	.084

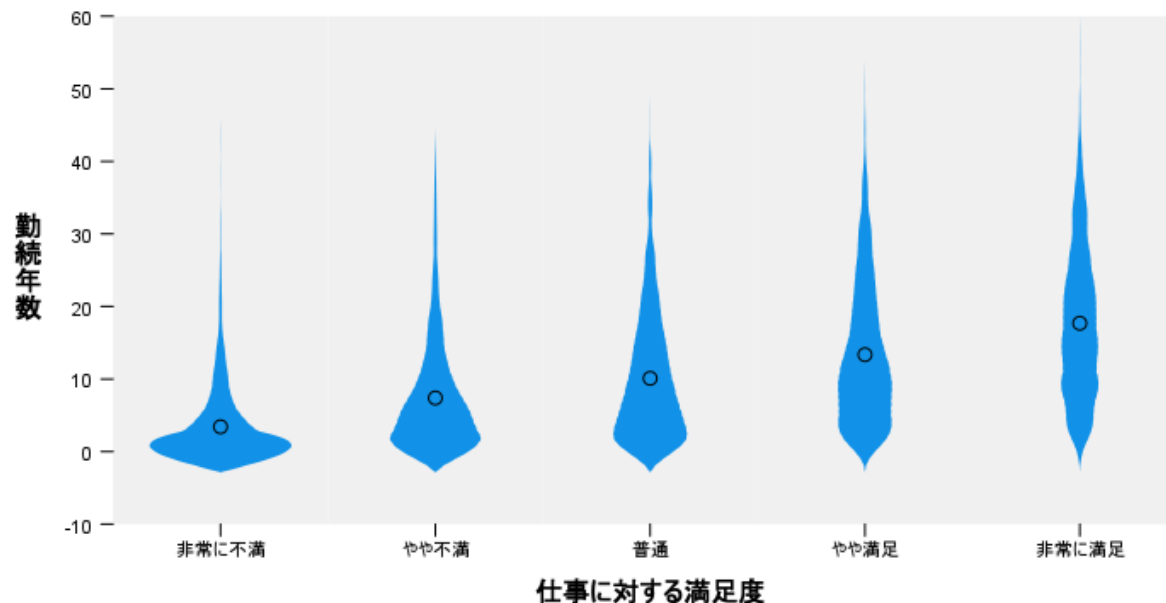
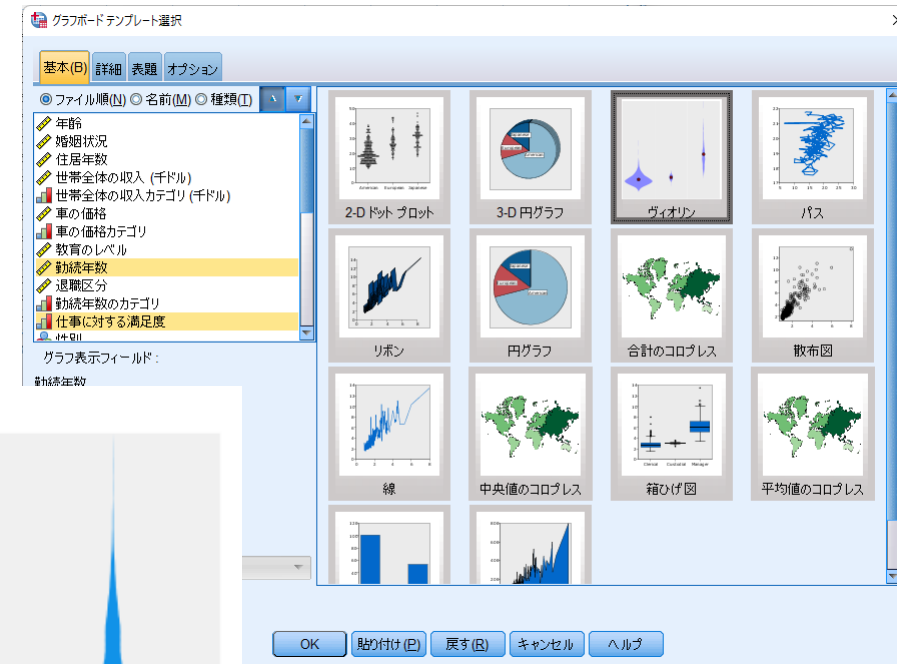
級内相関係数

全体的な ICC	調整済み	
	調整済み	.025
	条件式 (If-Then)	.023

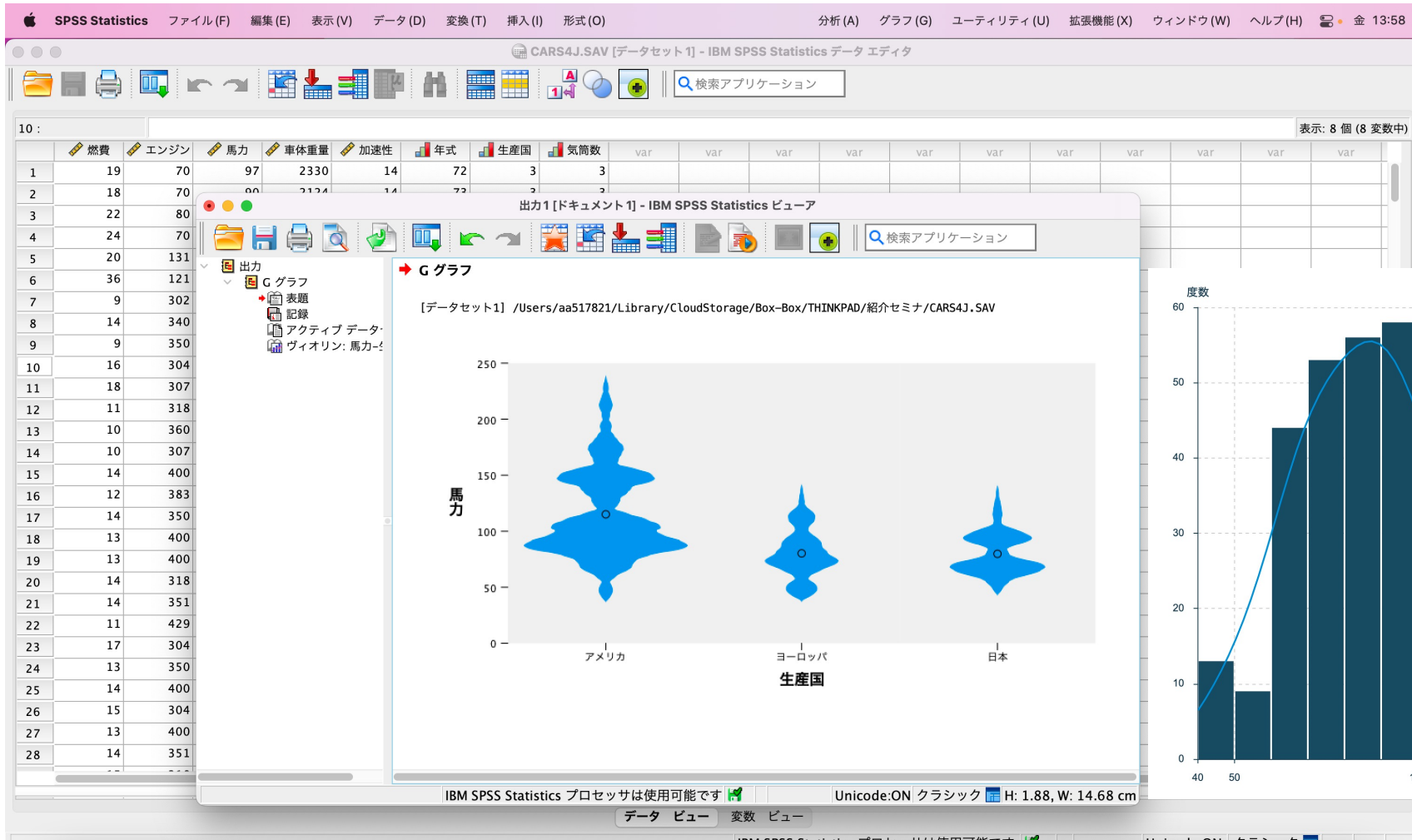


SPSS Statistics 29 | バイオリンプロット

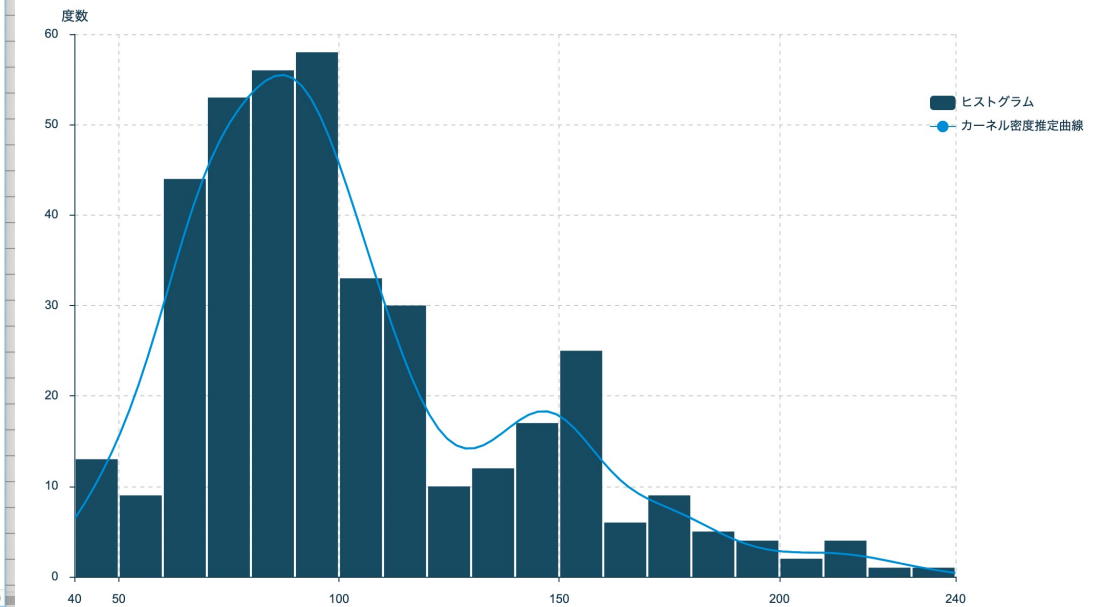
- グラフボードテンプレート選択に、箱ひげ図とカーネル密度図のハイブリッドであるバイオリンプロットが新たに追加されました。
 - バイオリンプロットは、データのピークを表示し、数値データの分布を視覚化するために使用されます。
 - 箱ひげ図は、要約統計量のみを表示するのにに対しバイオリンプロットは、要約統計量と各カテゴリの密度を表示します。



SPSS Statistics 29 | バイオリンプロット



KDE (カーネル密度推定曲線)
SPSS Modelerのチャート



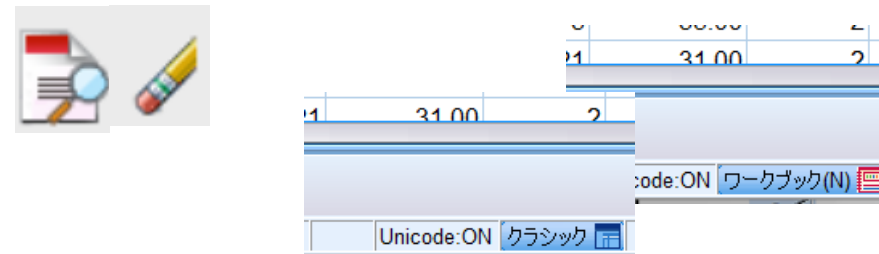
仕事に対する満足度

非常に不満 やや不満 普通 やや満足 非常に満足

SPSS Statistics 29 | ユーザビリティの向上

• ワークブックモードの機能強化

ワークブックツールバーに、「すべてのシンタックスの表示/非表示」と「すべての出力の消去」の2つの項目が追加されました。また、ステータスバーにクラシック（出力とシンタックス）モードとワークブックモードを切り替えるボタンが追加されました。



• 検索機能の強化

検索機能では、ツールバーのフィールドに直接用語を入力するオプションと、ドロップダウン・ペインで結果を表示するオプションが追加されました。



• 非選択ケースの表示機能を復元

ケースの選択機能で、「選択されなかったケースを分析から除外」を選んだ際に、選択条件に該当しないケースも印(ケース番号に斜線)を付けてデータエディターに表示します。Statistics 27.0.1 以前のバージョンの仕様に戻りました。

	年齢	婚姻状況	住居年数
1	55	既婚	12
2	56	未婚	29
3	28	既婚	9
4	24	既婚	4
5	25	未婚	2
6	45	既婚	9
7	42	未婚	19
8	35	未婚	15
9	46	未婚	26
10	34	既婚	0

	年齢	婚姻状況	住居年数
1	55	1	12
2	56	0	29
11	55	1	17
16	52	1	24
26	51	1	10
29	52	0	20
30	53	1	29
33	58	0	2
35	57	1	28
39	56	0	7

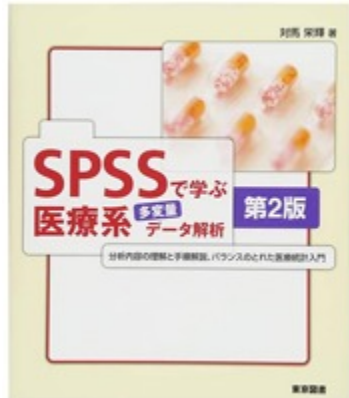
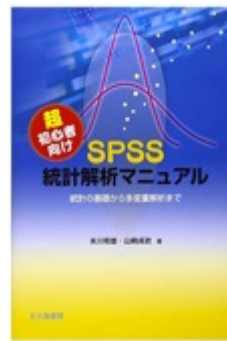
V28.x

• PythonとRのアップグレード

Python 3.10.4 と R 4.2.0 は IBM® SPSS® Statistics 29 と共にインストールされます。

3.SPSSインフォメーション

関連書籍

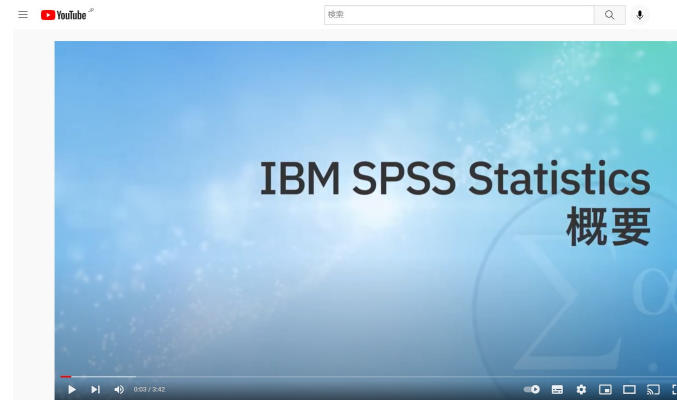


* 関連書籍の一部を掲載

紹介動画

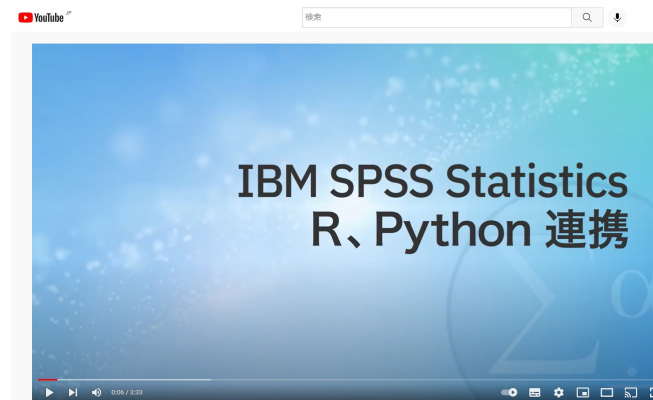
- **IBM SPSS Statistics 概要**

<https://www.youtube.com/watch?v=pbVs3cr2mTk>



- **IBM SPSS Statistics - R、Python連携**

<https://www.youtube.com/watch?v=zkWJU-edfJE&t=25s>



ブログ コミュニティ サポート

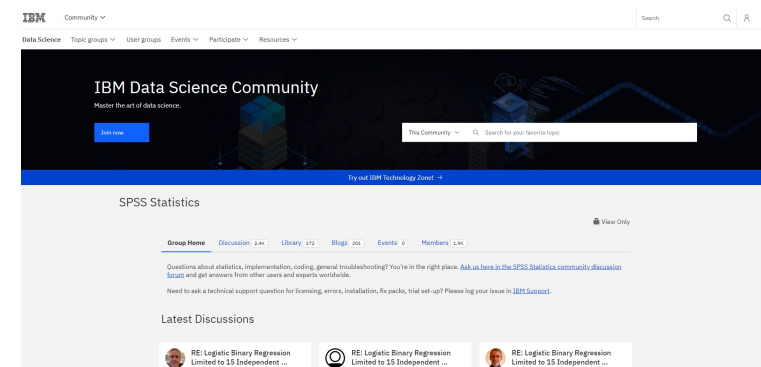
• Small Tips ブログ

<https://www.ibm.com/blogs/solutions/jp-ja/category/bigdata-analytics/spss-statistics/>



• SPSS Statistics コミュニティ

<https://community.ibm.com/community/user/datascience/communities/community-home?CommunityKey=886b6874-0fb1-402c-8243-c70ef8179a99>



• SPSSサポート

<https://www.ibm.com/jp-ja/products/spss-statistics/support>



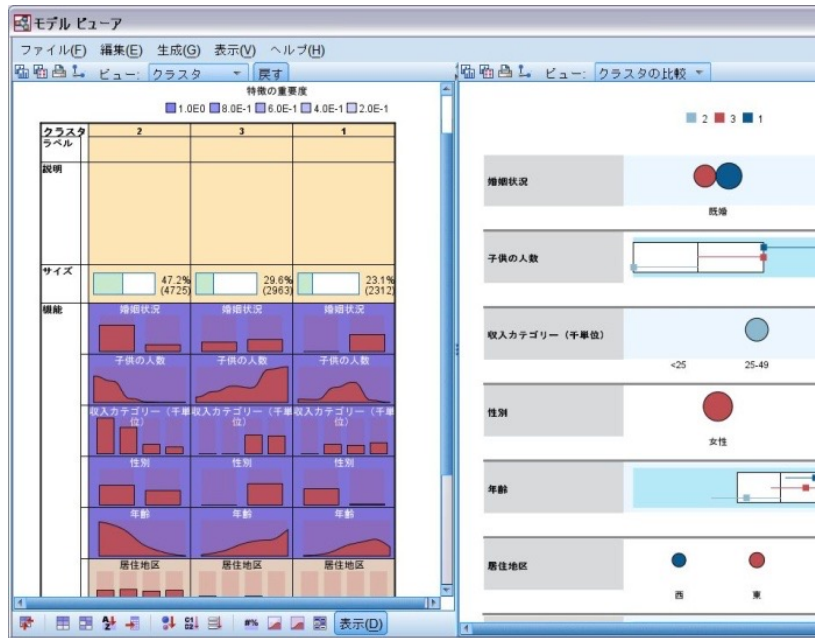
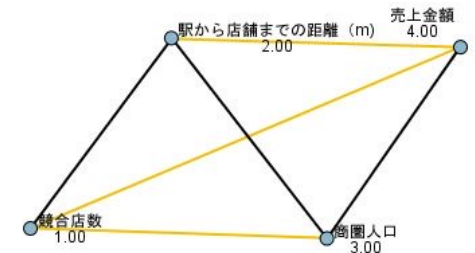
ご清聴ありがとうございました。

IBM SPSS Statistics Base

基礎統計処理と、統計解析のための基本機能を網羅する

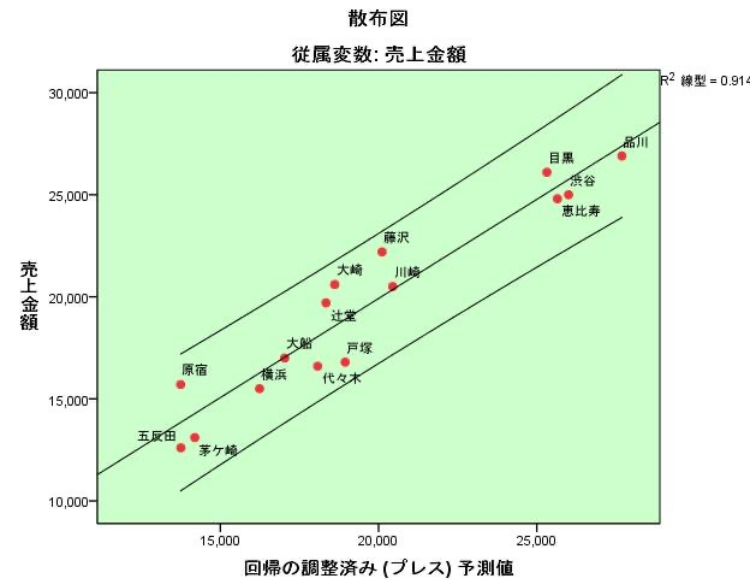
代表的な機能

- 集計 : 度数集計、クロス集計、OLAPキューブ
- 検定 : t検定、カイ二乗検定、ノンパラメトリック検定 (多重比較検定つき)
- 多変量解析 :
 相関分析、線型回帰分析、分散分析とその多重比較 (※一変量)
 因子分析、主成分分析、多次元尺度法、クラスター分析、判別分析
 最近隣法など



OLAP キューブ

	年代			
	20代	30代	40代	50代
東店店	15.98	15.25	15.16	15.16
西店	30.23	31.56	34.02	32.91
経通日				
大分類_アクセサリ_Sum	8670.12	9026.66	10232.41	11521.33
大分類_インナーウェア_Sum	30350.31	30083.83	30501.59	28241.72
大分類_バッグ_Sum	19166.94	15298.50	17545.44	17521.04
大分類_化粧品_Sum	38677.30	37494.63	38194.39	39980.18
大分類_婦人服_Sum	37268.72	37343.79	38004.15	38452.68
大分類_靴_Sum	18923.44	18192.77	19995.27	19514.46
大分類_食品_Sum	20290.71	20493.02	17862.50	21089.20



IBM SPSS Data Preparation

面倒なデータ検証作業を手軽に実現する

- データ検証のプロセスを効率化する
 - デフォルトで搭載しているルールを使い、データ内の異常値を検出する
 - ルールはカスタマイズ可能
 - 複数の変数にまたがる異常値も検出可能
 - 異常値のあるケースにフラグを立てレポートを作成するほか、データセットの先頭へ移動させることも可能
 - 自動データ準備機能(生データをそのまま使うのではなく変換した方が精度が良くなることもある)

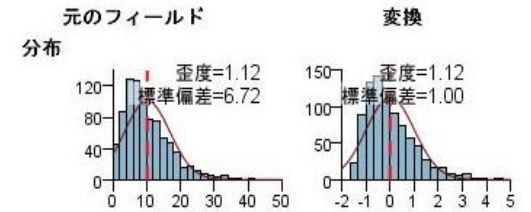
ケース報告書

ケース	検証規則違反		識別子
	単一変数 ^a	変数間	
4	1,2 dichotomy (1)		4
8		CrossVarRule 1	8

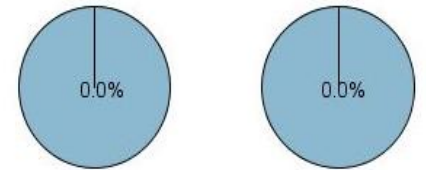
a. その規則に違反した変数の数はそれぞれの規則を満たしています。

使用するバージョン	名前	種類	予測べき集
変換	負債比		0.26
変換	雇用		0.15
変換	クレジット負債		0.14
元のフ...	教育		0.12
元のフィールド 使用しない	居住年数		0.09
変換	年齢		0.09
変換	その他負債		0.08
変換	収入		0.02

所得に対する負債の比率 (x100) の詳細



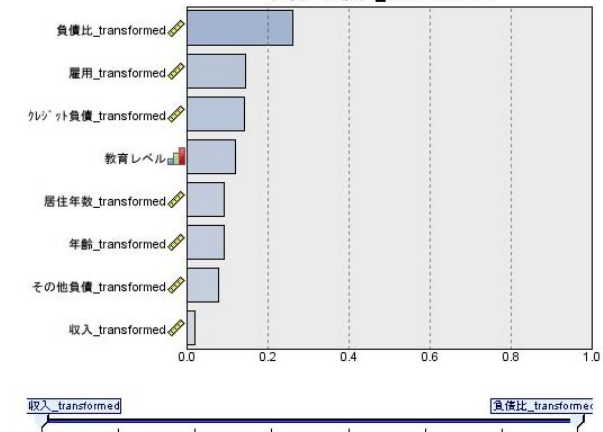
欠損値



分析での使用が推奨される特徴

予測精度

目標: 不履行_transformed



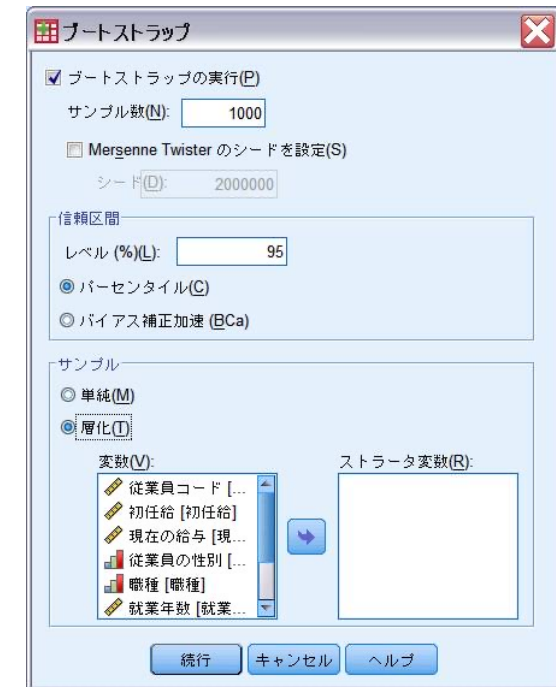
IBM SPSS Bootstrapping

ブートストラップ法の実行用アドオン オプション

- Statistics Base, Advanced Statistics, Regressionで実行可能な分析手法に対して、ブートストラップ法を適応する
 - Statistics Base:記述統計量、度数分布表、探索的分析、平均値、クロス集計表、t 検定、相関分析/ノンパラメトリック相関分析、偏相関分析、一元配置分散分析・一変量の分散分析、判別分析、順序回帰
 - Advanced Statistics:GLM、GENLIN、線型混合モデル、Cox回帰
 - Regression:回帰、名義回帰、ロジスティック回帰

記述統計			ブートストラップ ^a					
			統計量	標準誤差	バイアス	標準誤差	95% 信頼区間	
							下限	上限
小計_Sum	平均値		1188.05	107.976	-.55	106.83	987.36	1407.14
	平均値の 95% 信頼区間	下限						
		上限						
	5% trimmed 平均		1039.36		2.96	97.07	859.77	1242.01
	中央値		725.20		15.22	82.95	595.50	945.05
	分散		1550624.400		-13584.112	422166.555	861134.062	2484280.354
	標準偏差		1245.241		-16.946	168.328	927.973	1576.160
	最小値		77					
	最大値		8327					
	範囲		8250					
	4分位範囲		1371		-65	165	929	1551
	歪度		2.393	.210	-.207	.561	1.246	3.186
	尖度		8.355	.417	-1.735	4.026	.753	14.362

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

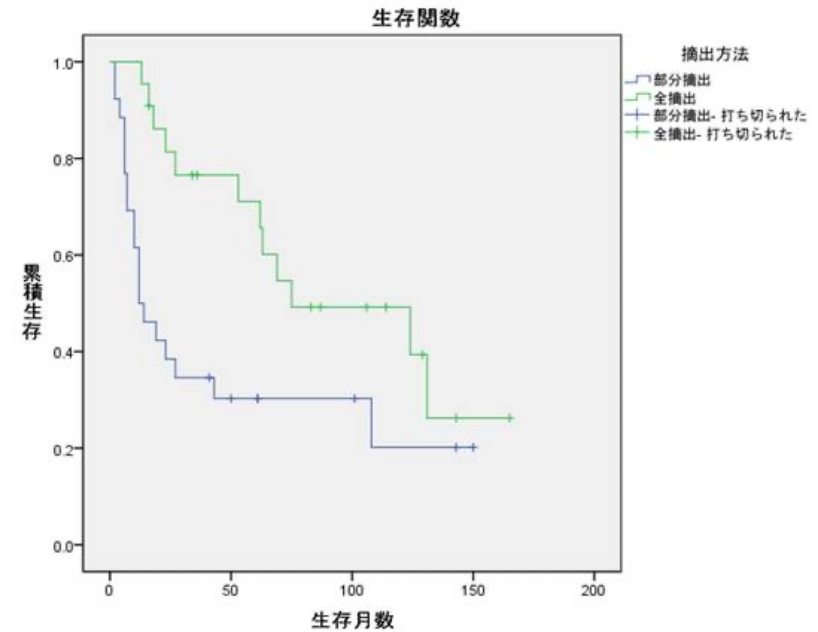
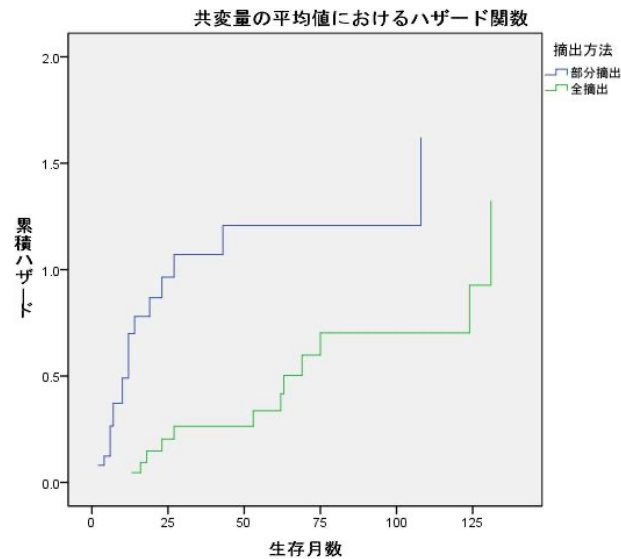


IBM SPSS Advanced Statistics

一般線型モデルや順序回帰、生命表分析、Cox回帰などを実現する

代表的な機能

- 分散分析とその多重比較（※多変量）
- 繰り返しのある分散分析（反復測定）
- 一般線型分析、Mixed Model、対数線型分析
- 一般化線形モデル（Generalized Linear Models: GZLM）
- 一般化推定方程式（Generalized Estimating Equations: GEE）
- 生存分析（COX回帰、カプラン・マイヤー）
- ベイズ統計



IBM SPSS Regression

多項・二項ロジスティック回帰、プロビット回帰など、高度な回帰分析を実行

- 回帰分析機能を強化
- 代表的な機能
 - ロジスティック回帰分析（二項 [0 / 1]、多項 [A / B / C]）
 - プロビット分析
 - 非線型回帰分析

方程式中の変数

	B	標準誤差	Wald	自由度	有意確率	Exp(B)	EXP(B) の 95% 信頼区間	
							下限	上限
ステップ 1 ^a								
雇用	-.244	.032	58.235	1	.000	.784	.736	.834
居住年数	-.082	.020	16.981	1	.000	.921	.886	.958
収入	-.007	.008	.661	1	.416	.993	.978	1.009
負債比	.071	.031	5.364	1	.021	1.073	1.011	1.140
クレジット負債	.612	.112	29.669	1	.000	1.845	1.480	2.300
その他負債	.056	.078	.509	1	.476	1.057	.907	1.232
教育			2.390	4	.664			
教育(1)	-1.036	1.253	.684	1	.408	.355	.030	4.137
教育(2)	-.786	1.250	.396	1	.529	.455	.039	5.274
教育(3)	-.715	1.259	.322	1	.570	.489	.042	5.769
教育(4)	-1.163	1.288	.815	1	.367	.313	.025	3.904
定数	.355	1.331	.071	1	.790	1.427		

a. ステップ 1: 投入された変数 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債, 教育

分類テーブル^{a, b}

観測			予測		
			不履行経歴		正解の割合
			なし	あり	
ステップ 0	不履行経歴	なし	517	0	100.0
		あり	183	0	.0
全体のパーセント					73.9

a. 定数がモデルに含まれています。

b. 遮断値は .500 です

IBM SPSS Custom Tables

カスタムテーブルを即座に作成し、多重回答データも扱って集計する

- 集計機能を強化する（テーブルプレビュービルダー）
 - 複雑な集計表を、視覚的に確認しながらの作成可能
- Statistics Baseで作成する集計表との違い
 - 多次元でのクロスや入れ子式の設定が容易に
 - 多重回答（マルチアンサー、MA）グループの変数設定を自動保存
 - 度数やパーセントだけでなく、分散や標準偏差など統計量も表示可能
 - 既存のカテゴリを編集・結合し、クロス集計へ反映する

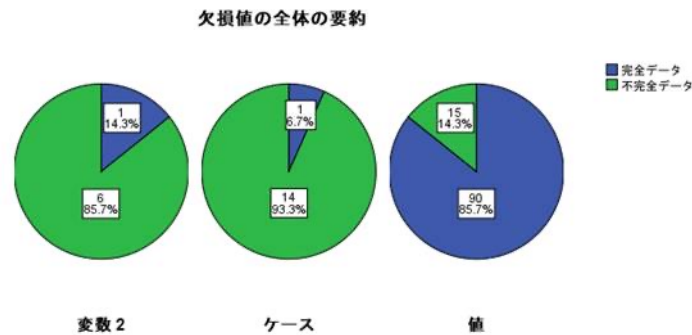
				最終学歴					
				中卒未満	高卒	短大卒	4大卒	大学院卒	
年齢がコリ	25歳未満	性別	男性	度数	13	79	5	11	0
			行の N %	12.0%	73.1%	4.6%	10.2%	.0%	
			女性	度数	26	88	8	10	0
			行の N %	19.7%	66.7%	6.1%	7.6%	.0%	
	25 - 34	性別	男性	度数	23	151	25	63	14
			行の N %	8.3%	54.7%	9.1%	22.8%	5.1%	
			女性	度数	35	198	29	72	16
			行の N %	10.0%	56.6%	8.3%	20.6%	4.6%	
	35 - 44	性別	男性	度数	34	157	29	61	27
			行の N %	11.0%	51.0%	9.4%	19.8%	8.8%	
			女性	度数	23	195	47	75	29
			行の N %	6.2%	52.8%	12.7%	20.3%	7.9%	
45 - 54	性別	男性	度数	25	109	17	53	16	
		行の N %	11.4%	49.5%	7.7%	24.1%	7.3%		
		女性	度数	24	126	23	49	38	
		行の N %	9.2%	48.5%	8.8%	18.8%	14.6%		
55 - 64	性別	男性	度数	34	55	6	26	15	
		行の N %	25.0%	40.4%	4.4%	19.1%	11.0%		
		女性	度数	34	101	7	16	25	
		行の N %	18.6%	55.2%	3.8%	8.7%	13.7%		



IBM SPSS Missing Values

欠損値を含むデータの分析を行う

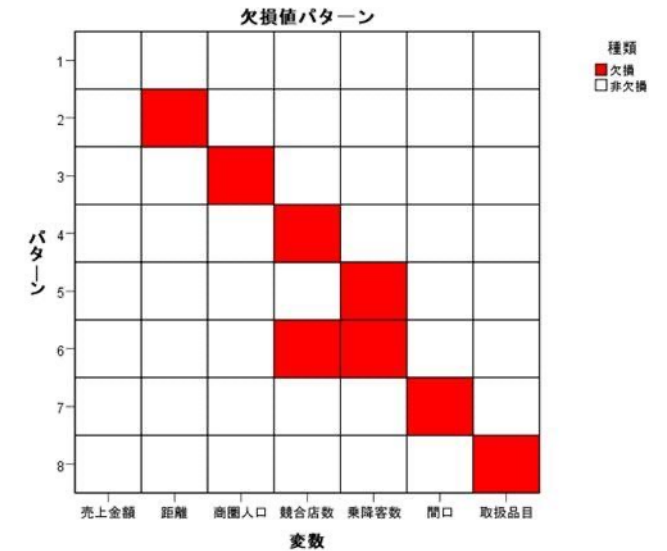
- 欠損値のパターン認識と推定し、適切な値で補完する
- Statistics Baseで行う欠損値処理の違い
 - 単一代入法(EM推定、回帰推定など)、多重代入法(MI: Multiple Imputation)に対応
 - 系列平均などを使った欠損値処理はStatistics Baseで可能



変数の要約^{a,b}

	欠損		有効な N	平均値 (ラン検定)	標準偏差
	度数	パーセント			
取扱品目数	3	20.0%	12	170.50	68.679
店の間口	3	20.0%	12	3.908	1.3767
乗降客数	3	20.0%	12	172.83	60.963
競合店数	2	13.3%	13	2.62	1.758
商圏人口	2	13.3%	13	11,301.23	3,323.185
駅から店舗までの距離…	2	13.3%	13	174.62	28.756

a. 表示する変数の最大値: 25
 b. 含める変数の欠損値の最小パーセント: 10.0%

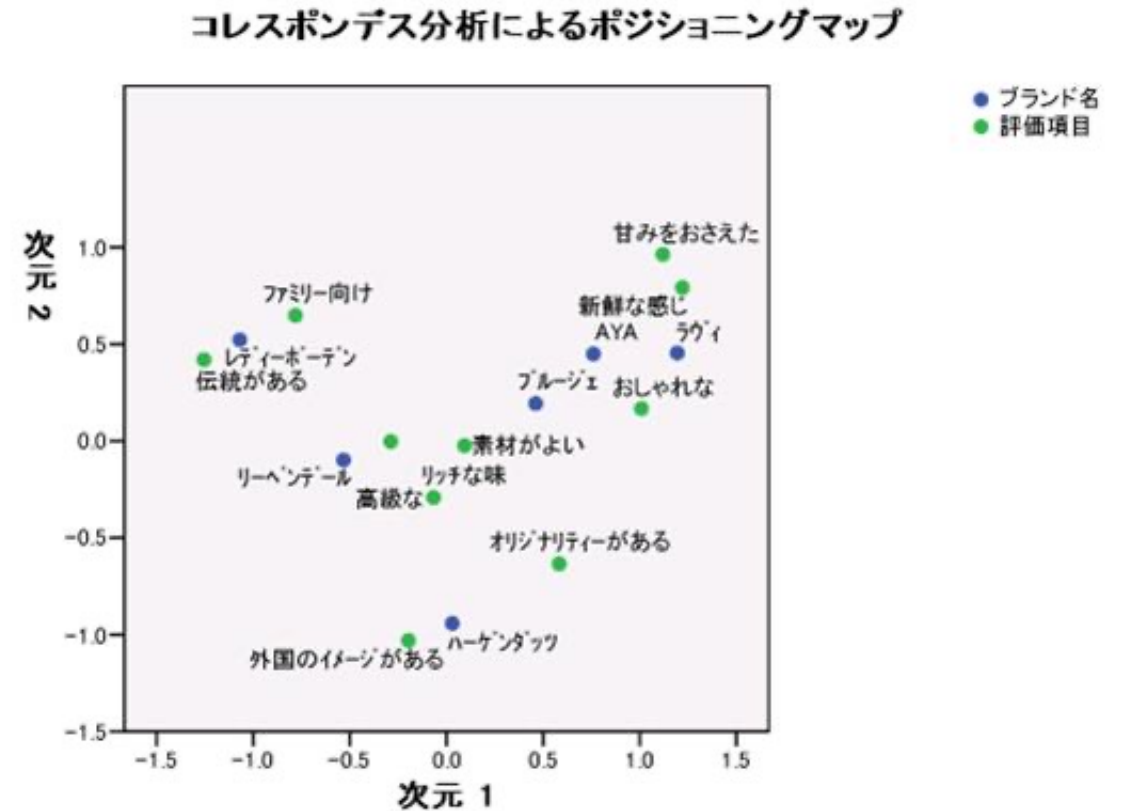
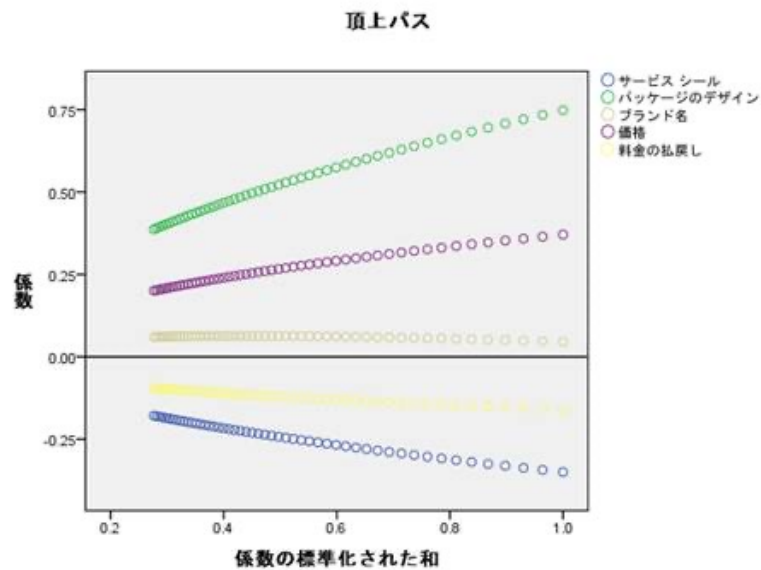


IBM SPSS Categories

コレスポンデンス分析などを実行し、カテゴリカルデータ分析をする

– カテゴリデータ分析機能を強化するオプション

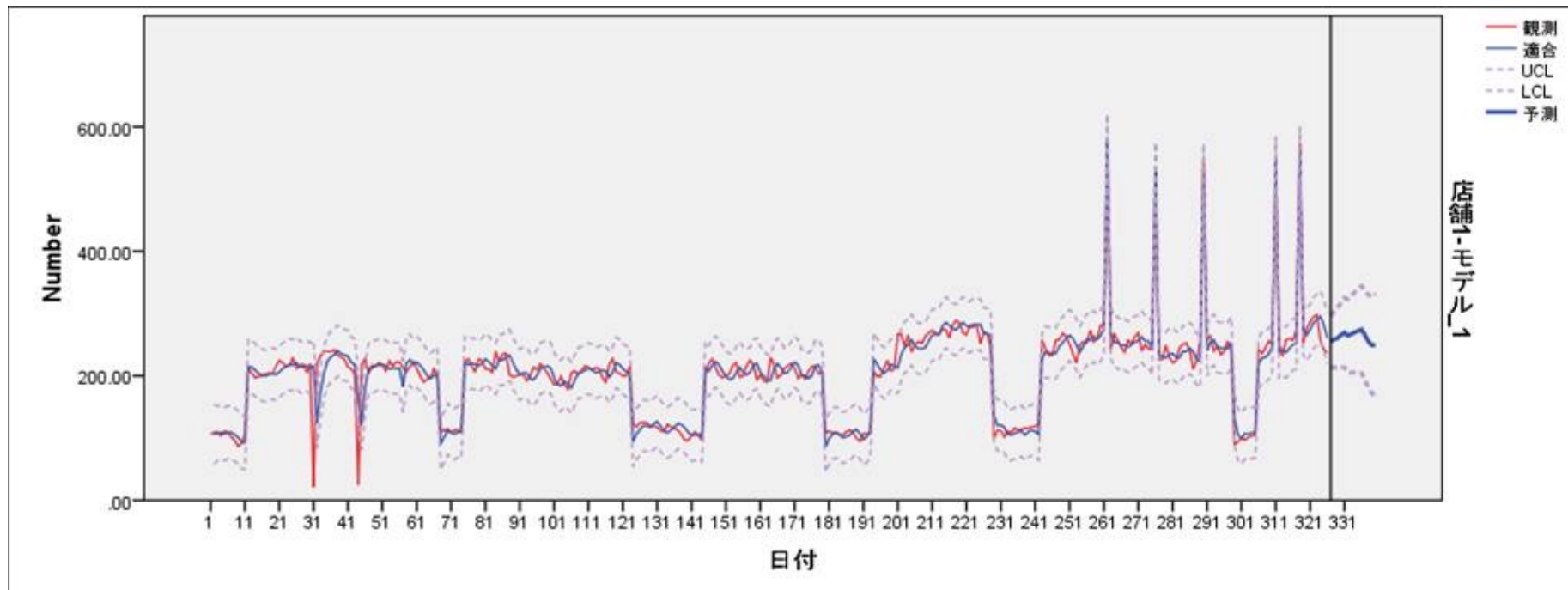
- コレスポンデンス分析（数量化Ⅲ類に相当）
- カテゴリデータ回帰分析（CATREG）
- カテゴリデータ主成分分析（CATCPA）
- 最適尺度法



IBM SPSS Forecasting

ARIMAや指数平滑モデルを使用し、時系列分析を行う

- Expert Modelerを利用することで、パラメータ決定からモデルの選択までを実行
- 分析で得られたモデルの再予測 / 再評価
- 代表的な分析手法
 - 指数平滑法、自己相関誤差の回帰、ARIMA、季節性の分解、スペクトル解析



IBM SPSS Decision Trees

CHAID, CRT, QUESTなどを搭載し、決定木分析を行う

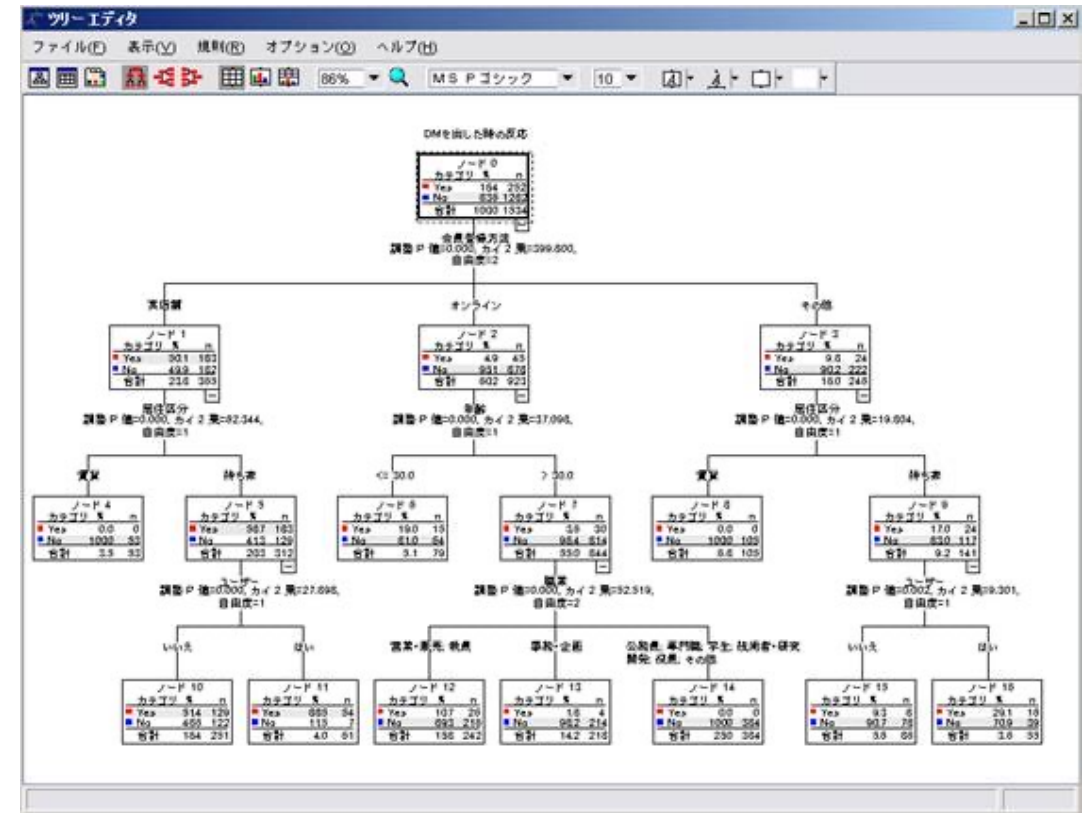
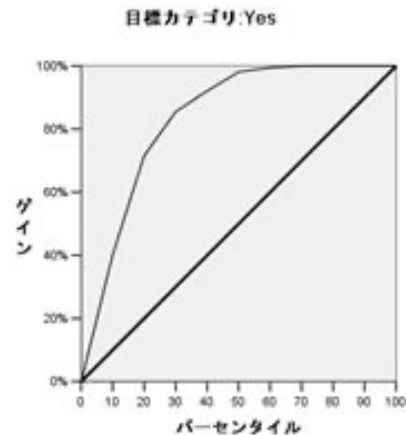
– データを構造的に理解することで特徴的なグループ(セグメント)を発見し、そこに至る背景(プロフィール、法則性)を探り出すことができる

– 数値データ / カテゴリデータ問わず分析に使用することが可能

- CHAID、Exhaustive CHAID、C&RT、QUESTを搭載

– 活用例 :

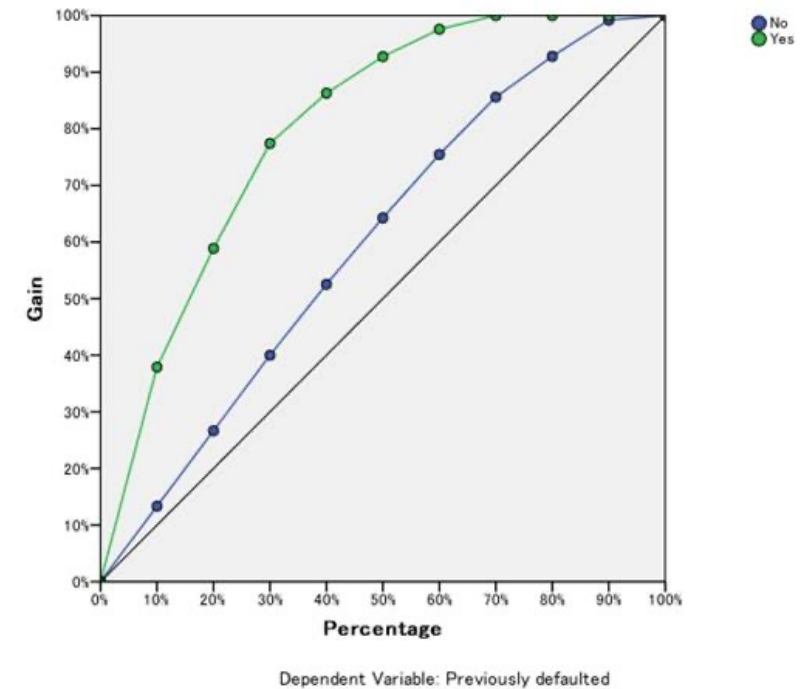
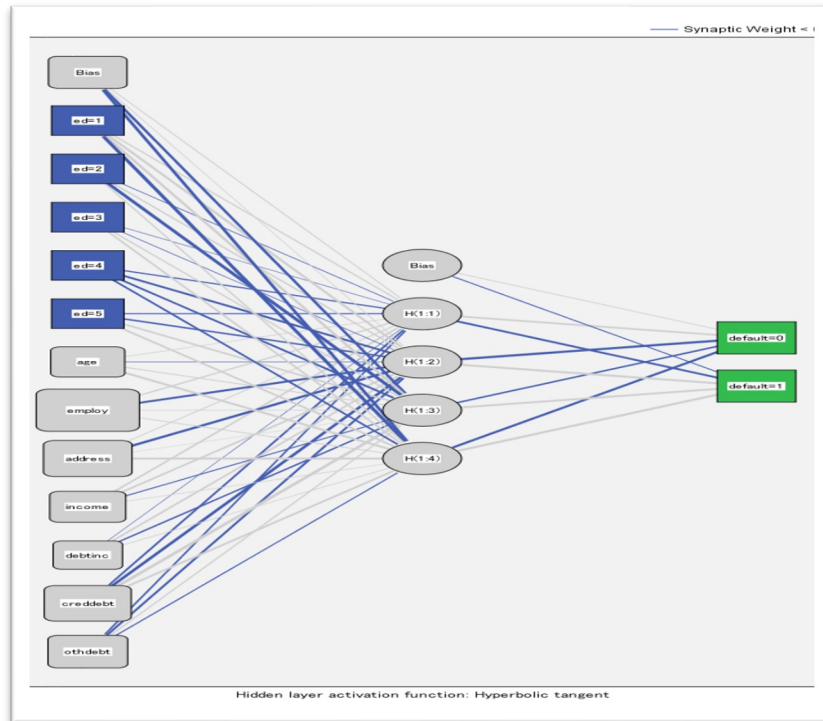
- マーケティング = DMレスポンス
- 金融 = 与信スコアリング
- 医療 = 疾病要因の特定
- 製造 = 不良品発生要因の特定、他



IBM SPSS Neural Networks

ニューラルネットワークを使用したモデリングを行う

- 従来の統計分析手法の仮定から外れるような、より複雑な構造をもつデータに対して予測・分類が可能。
- ゲインチャート、リフトチャートなどモデル評価のため多数のグラフを搭載。



IBM SPSS Direct Marketing

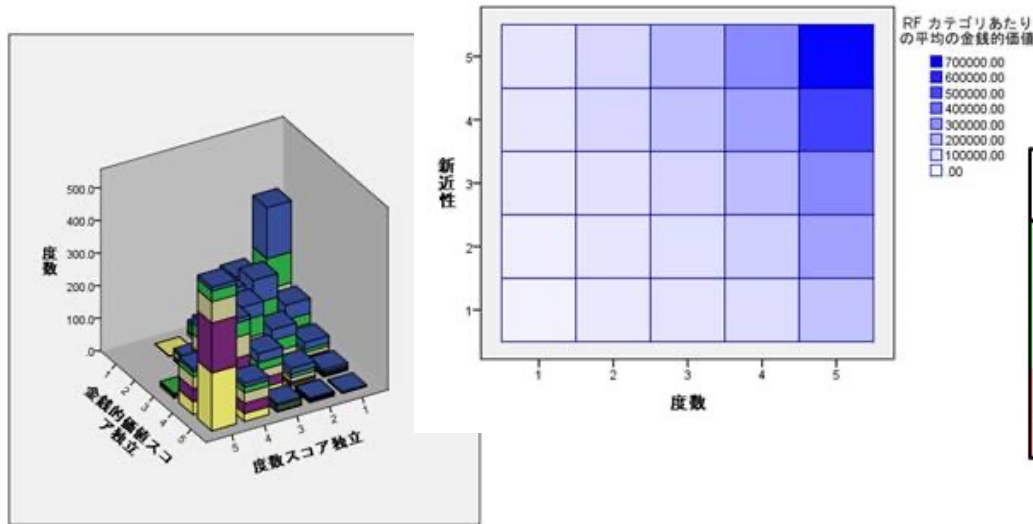
RFM分析やクラスタ分析などにより、顧客をより深く理解し施策に活用する

– ダイレクトマーケティングを行うための分析手法を6つ搭載

- RFM分析（顧客価値の算出）、クラスタ分析（顧客のグループ分け）、予測分析、グループ別回収率の測定、テストマーケティングの効果測定、複数キャンペーンの比較

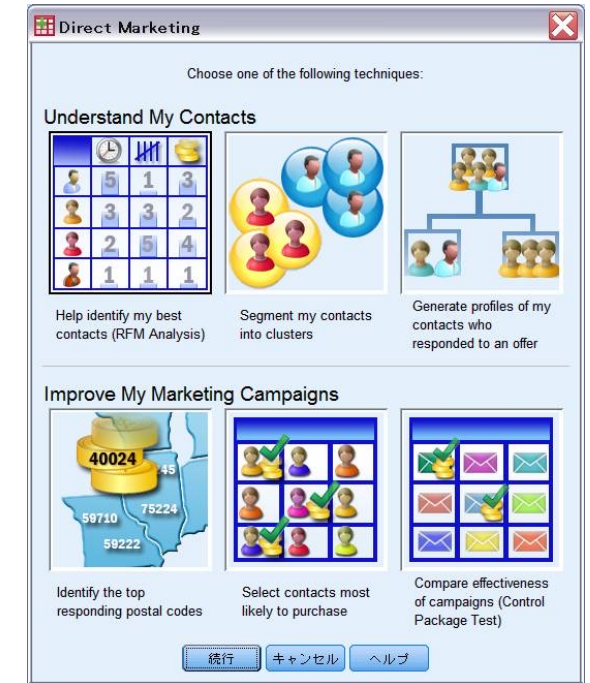
– 分析をこれから始めるビジネスユーザー向けに、統計機能名ではなく分析目的から操作できるよう設計

- 入力、出力を簡素化し、視覚的に分かりやすく
- 手間のかかるデータ加工や、分類ロジックを自動で処理が可能に



数	回答率			
	説明	グループサイズ	回答率	累積回答率
1	居住地区 = "西","南","東" 性別 = "女性" 婚姻状況 = "未婚"	379	9.2%	9.2%
2	居住地区 = "西","南","東" 性別 = "女性" 婚姻状況 = "既婚"	299	5.0%	7.4%
3	居住地区 = "西","南","東" 性別 = "男性"	722	4.7%	6.0%
4	居住地区 = "北"	517	2.5%	5.1%

緑: 目標回答率を満たす。
赤: 目標回答率を満たしていない。



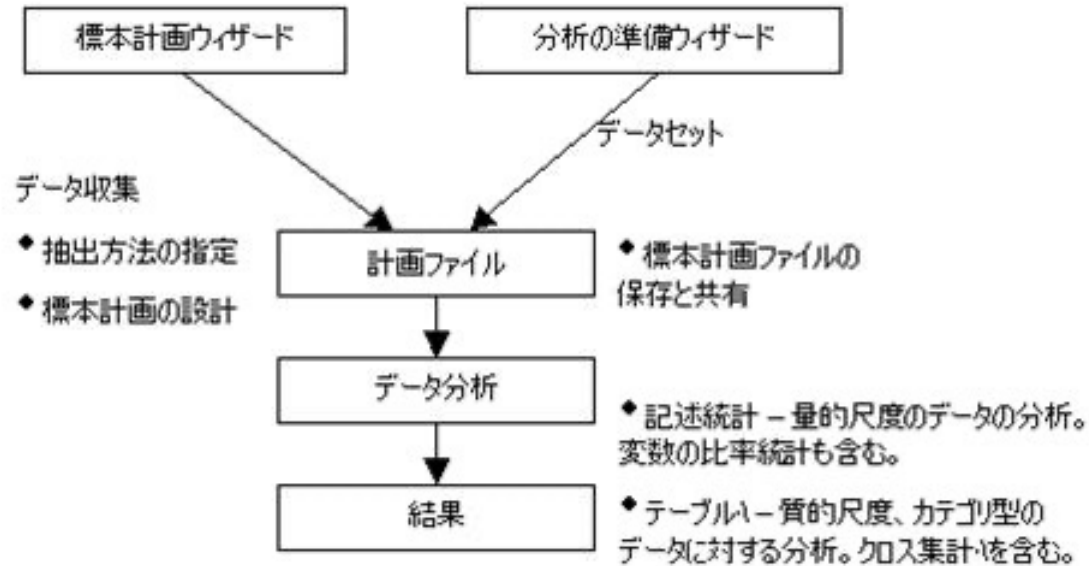
IBM SPSS Complex Samples

サンプリング手法を組み合わせた設計、および実行する

– 複雑なサンプリングの設計 / 実行をするためのオプション

– 代表的な手法

- 無作為抽出法
- 確率比例抽出法
- PPS系統抽出法



IBM SPSS Conjoint

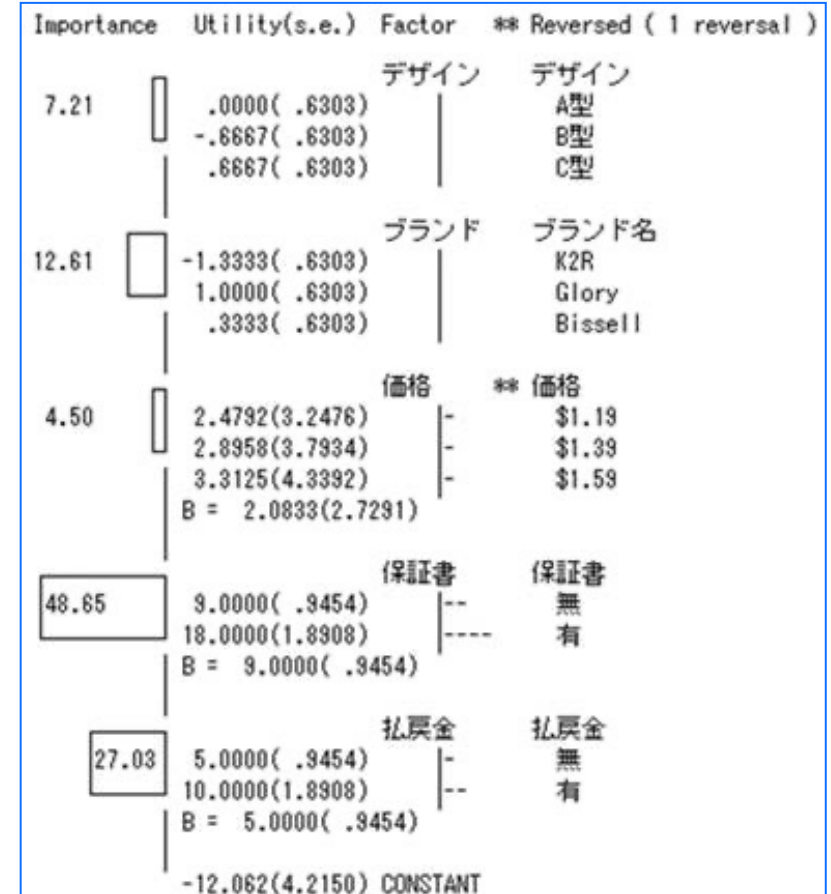
コンジョイント分析を行う

- 直交計画の作成
- コンジョイント分析の実行 (※コマンドシンタックスによる操作)

カードリスト

	カードID	デザイン	ブランド名	価格	保証書	払戻金
1	1	A型	Glory	\$1.39	有	無
2	2	B型	K2R	\$1.19	無	無
3	3	B型	Glory	\$1.39	無	有
4	4	C型	Glory	\$1.59	無	無
5	5	C型	Bissell	\$1.39	無	無
6	6	A型	Bissell	\$1.39	無	無
7	7	B型	Bissell	\$1.59	有	無
8	8	A型	K2R	\$1.59	無	有
9	9	C型	K2R	\$1.39	無	無
10	10	C型	Glory	\$1.19	無	有
11	11	C型	K2R	\$1.59	有	無
12	12	B型	Glory	\$1.59	無	無
13	13	C型	Bissell	\$1.19	有	有
14	14	A型	Glory	\$1.19	有	無
15	15	B型	K2R	\$1.39	有	有
16	16	A型	K2R	\$1.19	無	無
17	17	A型	Bissell	\$1.59	無	有
18	18	B型	Bissell	\$1.19	無	無
19 ^a	19	A型	Bissell	\$1.59	有	無
20 ^a	20	C型	K2R	\$1.19	有	無

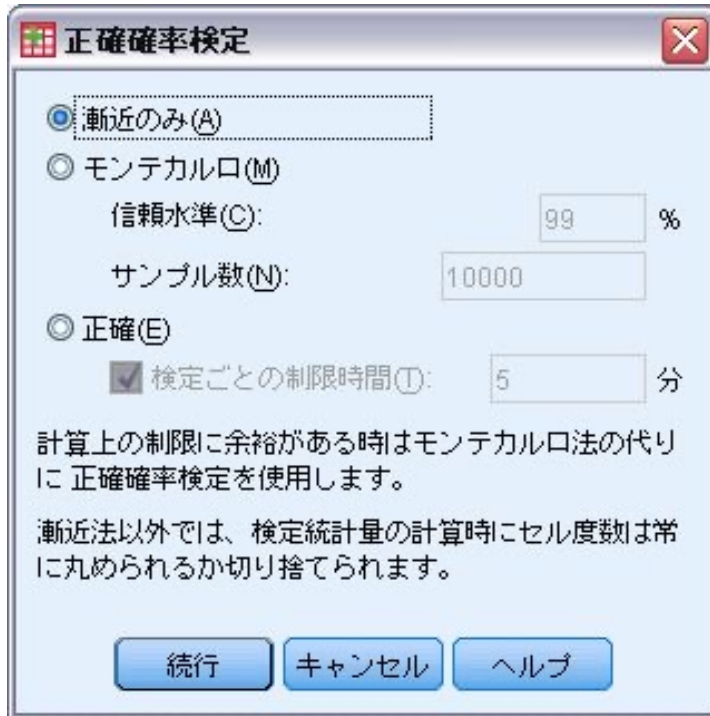
a. ホールドアウト



IBM SPSS Exact Tests

小さな標本からでも信頼性のある結果を導き出す「正確確率検定」を実行

- 小規模データの有意差検定手法を追加するオプション
- Statistics Baseの30種を超える検定機能に、正確確率検定を付加
 - フィッシャーの正確確率検定（漸近法、モンテカルロ法）、フリードマン検定、他



IBM SPSS AMOS

- 共分散構造分析に特化した分析ツールです。
- 複雑なモデルをパス図を使って簡単に表現可能です。

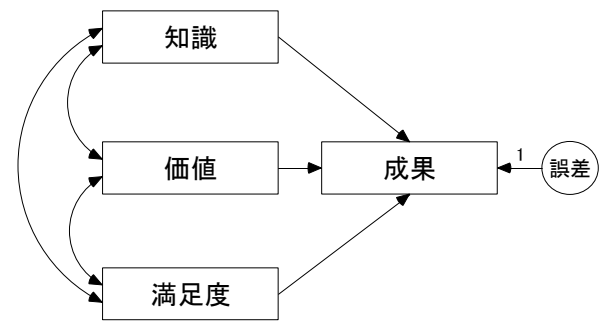
基本操作は と 、そして を組み合わせパス図を描きます

分析結果はテキスト形式でも確認

係数	推定値	標準誤差	検定統計量	確率	ラベル
関与 <-- 映画力	1.000				
ミナー <-- 映画力	1.204	.078	15.364	***	
情緒 <-- 映画力	.715	.072	9.880	***	
関与1 <-- 関与	1.000				
関与2 <-- 関与	1.044	.073	14.292	***	
関与3 <-- 関与	1.043	.089	11.667	***	
情緒1 <-- 情緒	1.000				
情緒2 <-- 情緒	.872	.123	7.080	***	
情緒3 <-- 情緒	1.380	.130	10.647	***	
ミナー1 <-- ミナー	1.000				
ミナー2 <-- ミナー	.929	.046	20.246	***	
鑑賞回数 <-- 映画力	1.241	.072	17.321	***	

ベイズ推定を使って順序カテゴリ値の分析もOK

回帰モデル



確認的因子モデル

