

エクセルで統計分析 統計プログラムHADについて

清水裕士

関西学院大学社会学部

HADとは

HADを作った人

- 清水裕士
 - 専門: 社会心理学 グループダイナミクス
 - 所属: 関西学院大学社会学部 准教授
 - 連絡先: simizu706(あつと)norimune.net
 - (あつと)を@に変えてください。
 - ブログ: <http://norimune.net>
 - Twitter: @simizu706



HADとは

- 清水が作ったExcelのVBAで動くプログラム
 - Excelのバージョンは2007以降に対応
 - Macにも対応 (Excel for Mac2011以降)
 - WinとMacを同じファイルで使いまわせる
- 主に心理統計分析ができる
 - 大抵の心理統計分析は可能
 - データハンドリング, 統計的検定, 多変量解析など
 - マルチレベル分析もできる
 - 階層線形モデルや, マルチレベルSEMなど

HADとは

- 無償のソフトウェアです
 - 利用は無償です
 - 清水のブログからダウンロードできます
 - <http://norimune.net/had>
 - 何度でもダウンロードできます
- 自由なソフトウェアです
 - ソースコードを自由に閲覧・変更することができます
 - 第三者への配布も自由です
 - ライセンス
 - GNU General Public License (GPL) に則ってます
 - ライセンスについては「HADとは」(<http://norimune.net/696>)のページを参照してください。

HADとは

- 「使いやすく・わかりやすく」というコンセプト
 - Excelなので、多くの環境で利用できます
 - 学生が家でも卒論の分析ができます
 - あるいは教員が指導しやすい
 - それなりの統計分析を、わかりやすい出力で
 - 卒論・修論で使うレベルの分析は一通りできる
 - 出力は、わかりやすさを重視しています
 - あまりマニアックな出力はしません
 - ただし心理学で求められる出力は、一通り出せる

HADの見た目

データシート

	A	B	C	D	E	F	G	H
1	変数名	ID	Species	Sepal.Le	Sepal.Wi	Petal.Le	Petal.Width	
2	データ 読み込み	1	1	5	3.3	1.4	0.2	
3		2	1	4.6	3.4	1.4	0.3	
4		3	1	4.6	3.6	1	0.2	
5	モデリング シート	4	1	5.1	3.3	1.7	0.5	
6		5	1	5.5	3.5	1.3	0.2	
7		6	1	4.8	3.1	1.6	0.2	
8		7	1	5.2	3.4	1.4	0.2	
9		8	1	4.9	3.6	1.4	0.1	
10		9	1	4.4	3.2	1.3	0.2	
11		10	1	5	3.5	1.6	0.6	
12	列幅の 調整	11	1	4.4	3	1.3	0.2	
13		12	1	4.7	3.2	1.6	0.2	
14		13	1	4.8	3	1.4	0.3	

HADの基礎統計分析

統計分析マクロ HAD (簡易モード)

データの要約

- 要約統計量
- ヒストグラム 箱ひげ図
- 散布図 直線回帰
- クロス集計表

変数間の関連性

- 相関分析
 - 順位相関分析
- 項目分析 (α 係数)
- 独立性の検定 (χ^2 乗検定)

差の検定

- 一標本の検定 検定値 =
- 平均値の差の検定 (t検定)
 - 対応なし 対応あり
- 順位の差の検定 (ノンパラ検定)
 - 対応なし 対応あり

マルチレベル分析

- 級内相関係数
- マルチレベル相関分析
- グループごとの回帰直線
- ペアワイズ共分散行列

詳細モード フィルタをオフにする 出力を上書きしない

OK キャンセル

GUIで変数の設定

変数情報変更

フィルター 値 ラベル

変数名 Species

値

ラベル

追加

1=セトサ
2=バーシカラー
3=バージニカ

適用

OK キャンセル

変数の編集・作成

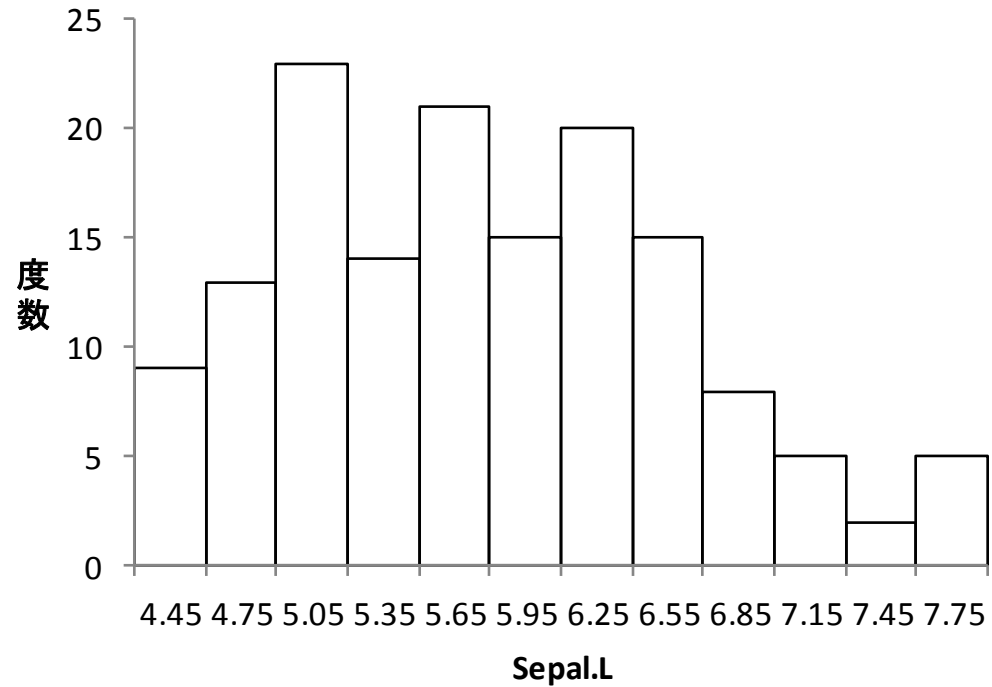
変数の作成： HAD

変数の合成 | 尺度変換 | 数値変換 | 値の再割り当て | ダミー変数

- ・使用変数から合成変数を作ります
 - 平均得点を算出
 - 合計得点を算出
 - 最初の変数から残りの変数を引く
- ・交互作用項を作ります
 - 中心化して交互作用項を作成
 - 変数の合成に、標準化した得点を使用する
- フィルタをオフにする
- 出力を上書きしない

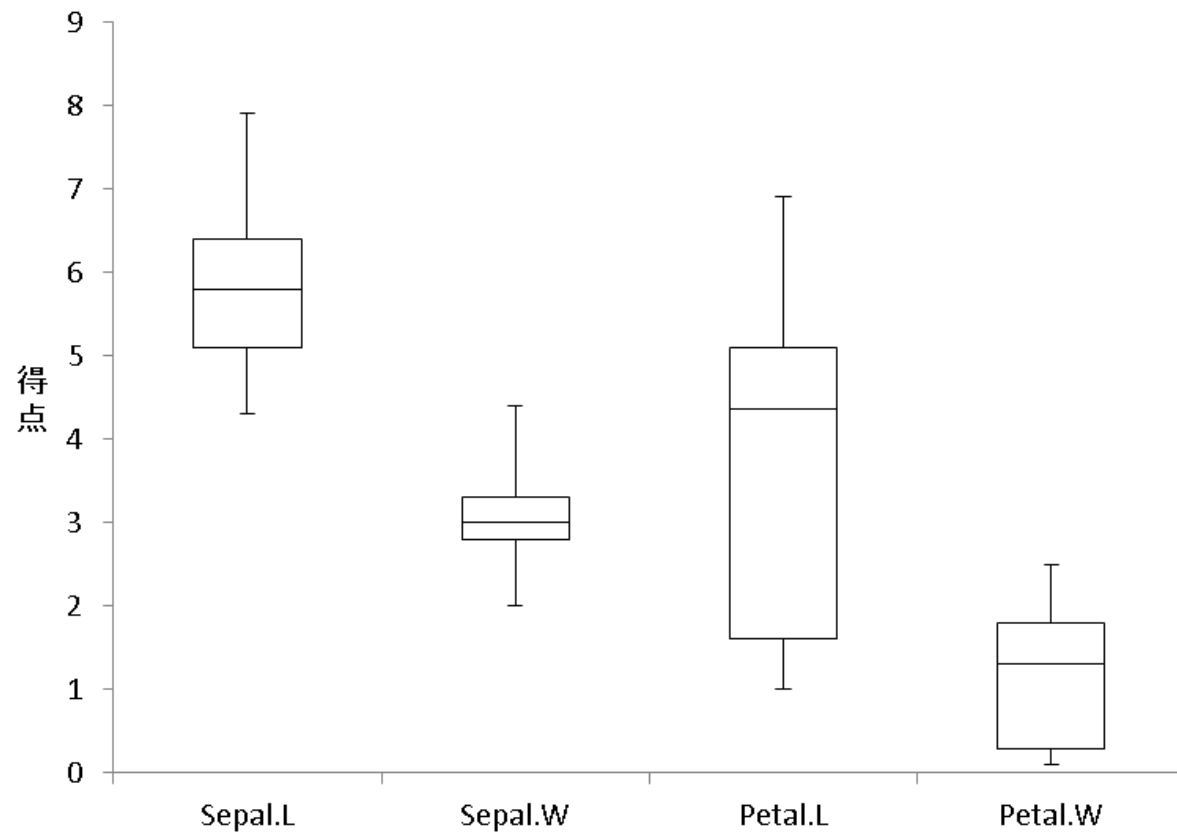
OK キャンセル

ヒストグラム

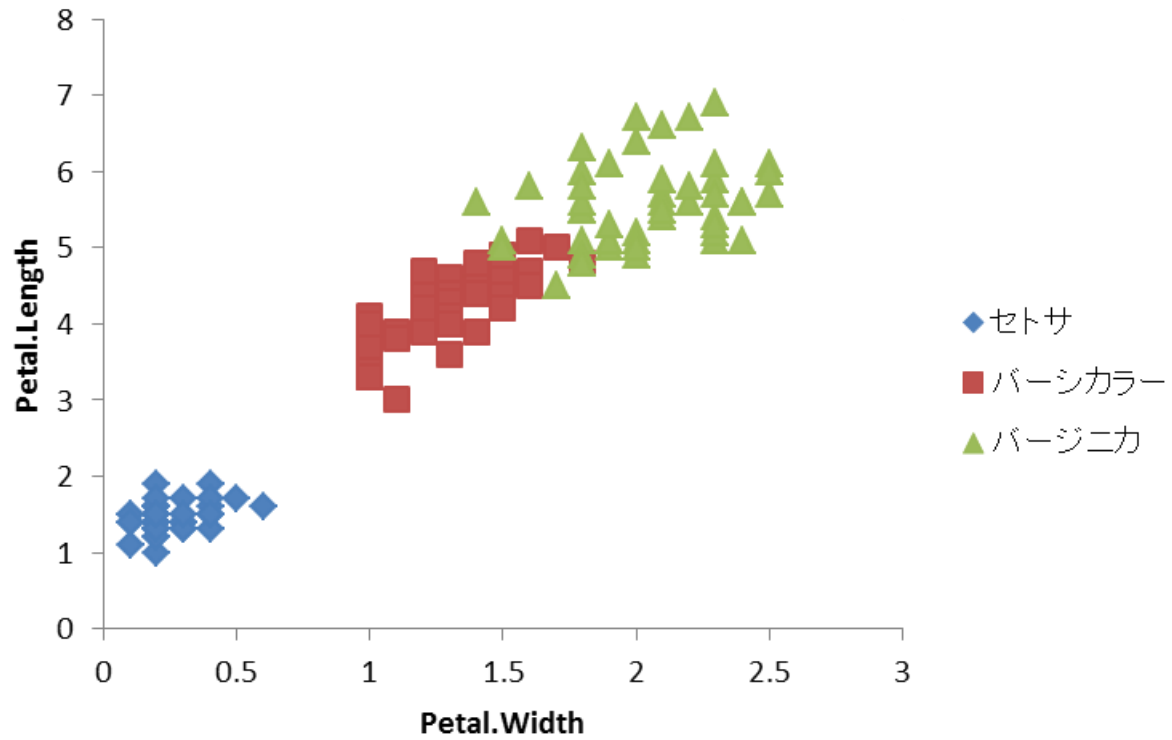


平均値	5.843	歪度	0.315	正規性	0.089
標準偏差	0.828	尖度	-0.552	補正p値	.006

箱ひげ図



散布図



基本統計分析の出力

要約統計量

	変数名	有効N数	平均値	中央値	標準偏差	分散
Sepal.Length	萼片の長さ	150	5.843	5.800	0.828	0.686
Sepal.Width	萼片の幅	150	3.057	3.000	0.436	0.190
Petal.Length	花弁の長さ	150	3.758	4.350	1.765	3.116
Petal.Width	花弁の幅	150	1.199	1.300	0.762	0.581

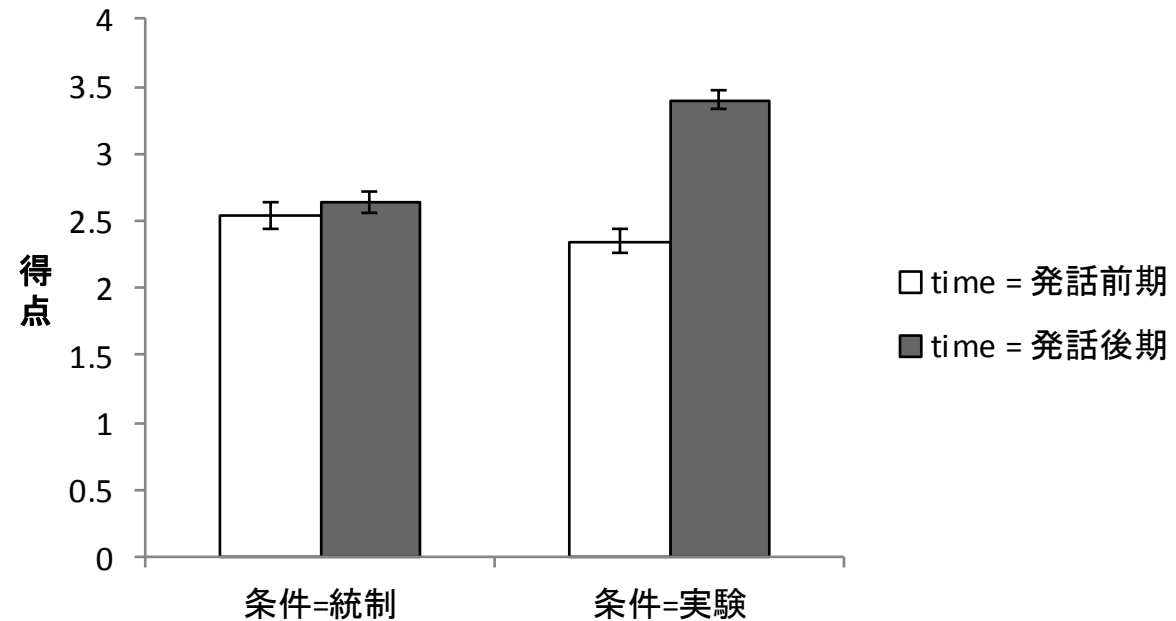
相関係数

		Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	萼片の長さ	1.000			
Sepal.Width	萼片の幅	-.118	1.000		
Petal.Length	花弁の長さ	.872 **	-.428 **	1.000	
Petal.Width	花弁の幅	.818 **	-.366 **	.963 **	1.000

分散分析の出力例

交互作用の結果

	time = 発話前期	time = 発話後期
条件=統制	2.540	2.640
条件=実験	2.347	3.400 **

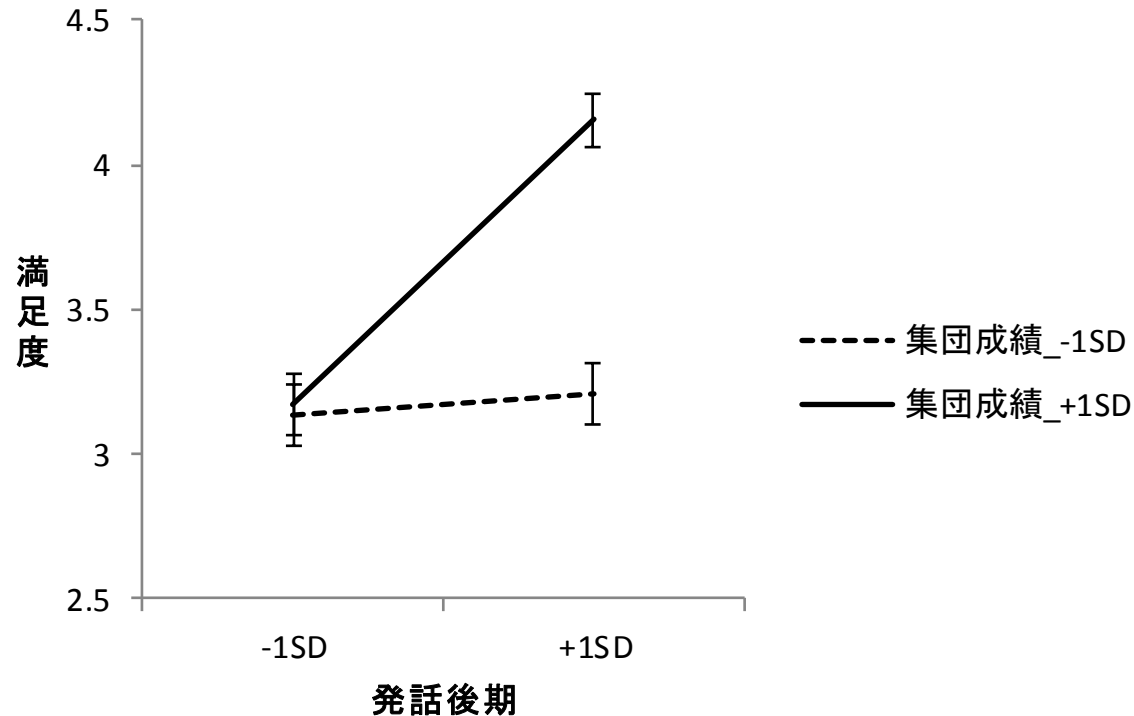


※エラーバーは標準誤差

回帰分析の出力例

交互作用効果のグラフを出力

	発話後期 -1SD	発話後期 +1SD	
集団成績_	3.133	3.206	
集団成績_	3.173	4.153	**



因子分析の出力例

因子パターン

反復回数 = 4

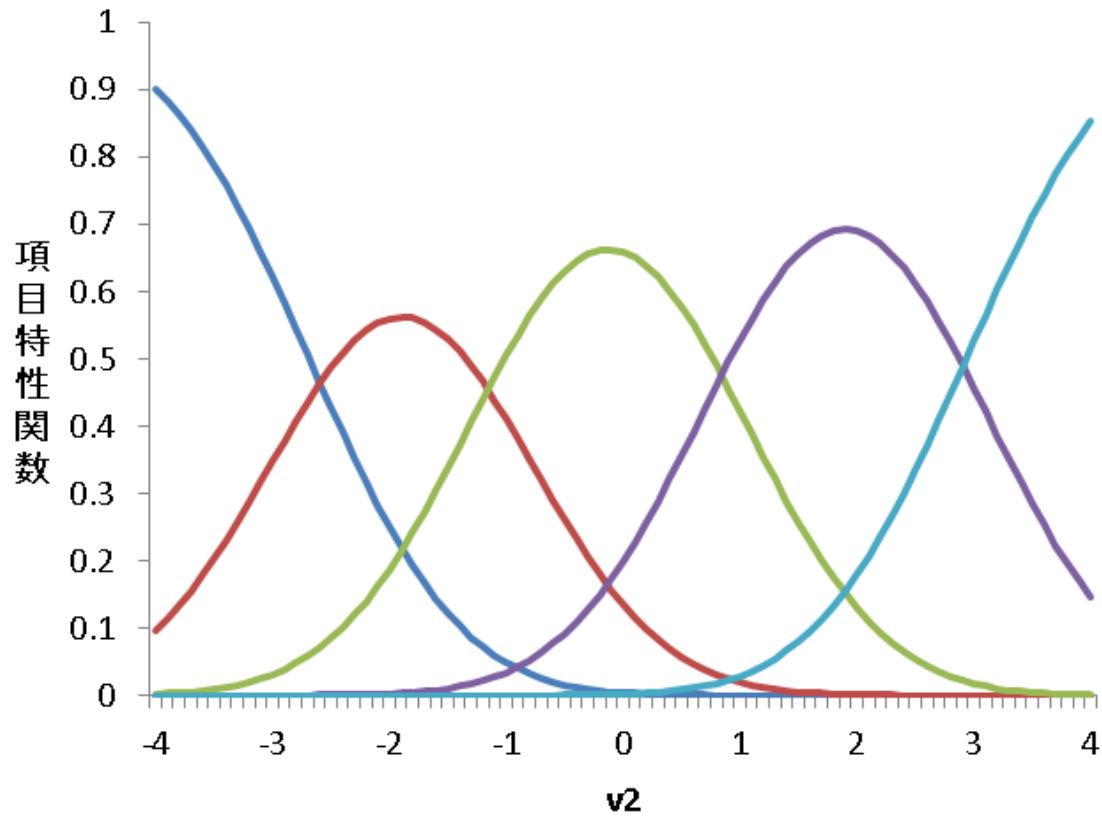
収束基準 = 0.0003

項目	Factor1	Factor2	共通性
v6	.835	-.047	.647
v8	.751	.105	.679
v10	.750	-.045	.521
v9	.744	.015	.570
v7	.716	.077	.591
v5	-.077	.810	.579
v4	-.022	.753	.545
v2	.022	.706	.519
v1	.103	.666	.544
v3	.107	.595	.450

因子寄与 4.210 3.995

適合度	乖離度 =	0.110	CFI =	1.000
	χ^2 値 =	21.279	RMSEA =	.000
	DF =	26	AIC =	59.884
	ρ =	.727	BIC =	122.552

項目反応理論の出力例



階層線形モデルの出力例

固定効果

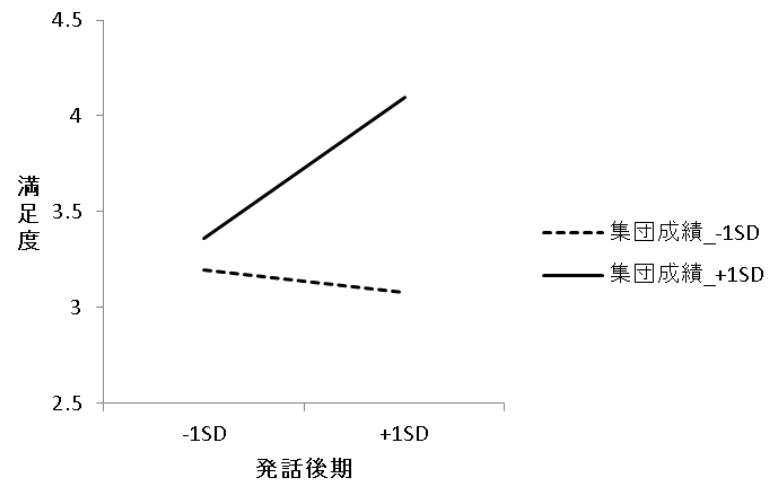
従属変数 =idt

変数名	係数	標準誤差	95%下限	95%上限	df	t値	p値
切片	3.434	0.064	3.308	3.560	97	53.994	.000 **
talk_g (wc)	0.215	0.073	0.071	0.359	97	2.957	.004 **
talk_m (gm)	0.358	0.111	0.137	0.578	97	3.215	.002 **
per (gm)	0.162	0.046	0.070	0.253	97	3.510	.001 **
talk_g*per	0.179	0.051	0.078	0.279	97	3.527	.001 **

※レベル1の説明変数はすべて集団平均で中心化しています(wc)。
※レベル2の説明変数はすべて全体平均で中心化しています(gm)。
※頑健標準誤差を用いています。

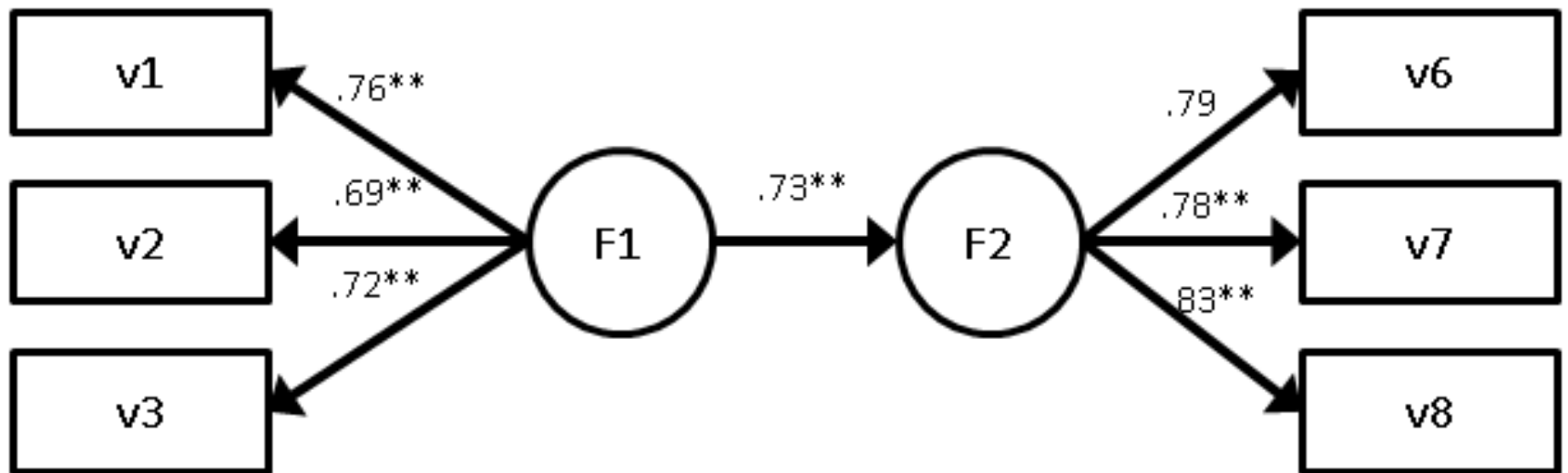
単純主効果の検定 (階層線形モデル)

交互作用の変数(⇔) talk_g*per
群分けする変数 per
検定する変数(→) talk_g

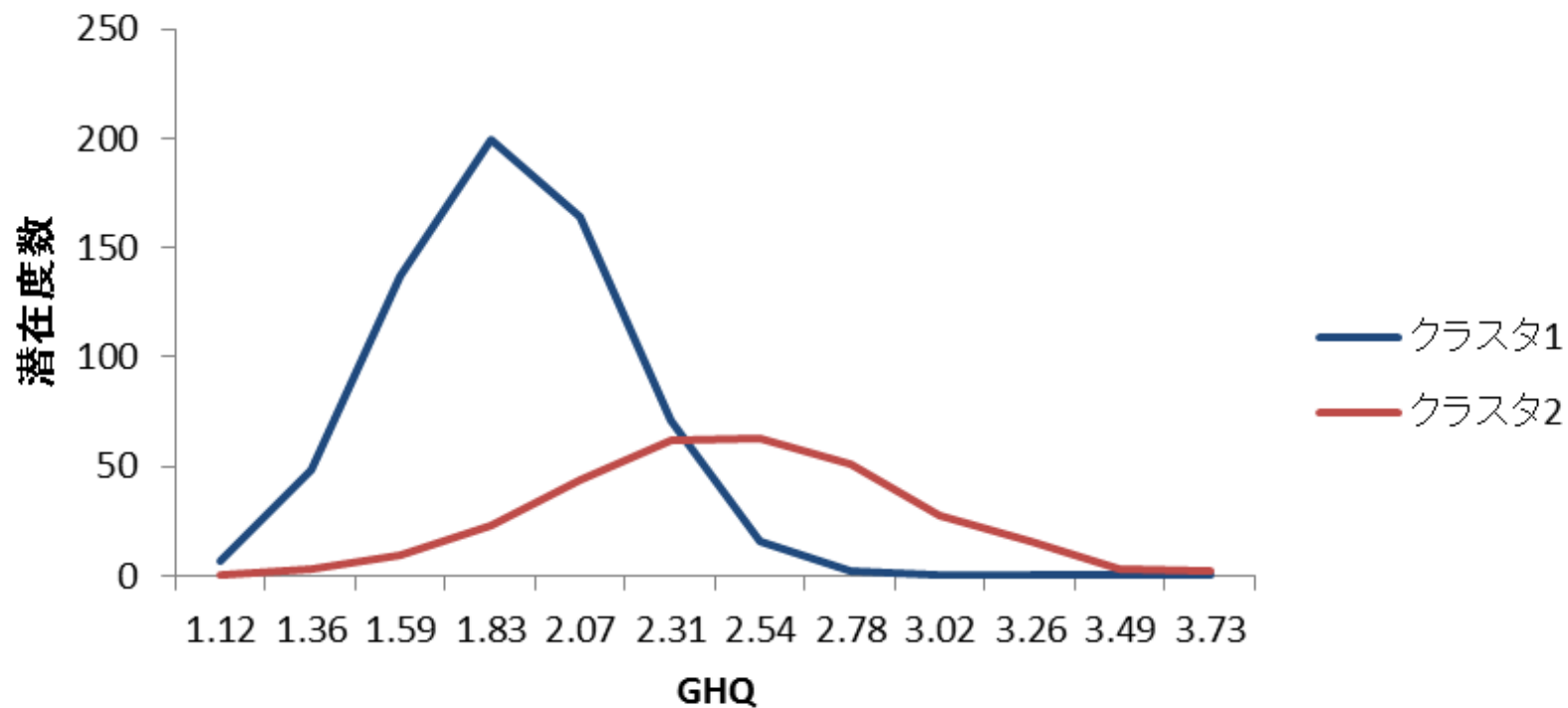


構造方程式モデルの例

※表記のパスは標準化係数



混合分布モデル



分析履歴で分析を再現

分析履歴		履歴を非表示	履歴を全消去
番号	日付	コード	要約
228	2014/7/25 7:53:46	-----	因子分析 非反復推定法 プロマックス
227	2014/7/24 16:25:45	-----	構造方程式モデル 使用変数:v1 v2 v3
226	2014/7/24 16:25:36	-----	構造方程式モデル 使用変数:v1 v2 v3
225	2014/7/24 14:28:05	-----	構造方程式モデル 使用変数:v1 v2 v3
224	2014/7/24 14:24:46	-----	構造方程式モデル 使用変数:v1 v2 v3
223	2014/7/24 14:24:17	-----	構造方程式モデル 使用変数:v1 v2 v3
222	2014/7/24 14:23:58	-----	構造方程式モデル 使用変数:v1 v2 v3
221	2014/7/24 14:23:16	-----	構造方程式モデル 使用変数:v1 v2 v3
220	2014/7/24 14:22:53	-----	構造方程式モデル 使用変数:v1 v2 v3
219	2014/7/24 14:21:55	-----	構造方程式モデル 使用変数:v1 v2 v3
218	2014/7/24 14:21:14	-----	構造方程式モデル 使用変数:v1 v2 v3
217	2014/7/24 14:21:05	-----	構造方程式モデル 使用変数:v1 v2 v3
216	2014/7/22 8:04:30	-----	重回帰分析 使用変数:v1 v2 v3
215	2014/7/22 8:02:36	1024000000001 Prtl	ポリコリック相関分析 使用変数:v1 Prir

HADの機能

HADの機能の紹介

- 基本はExcel
 - Excelで使える関数は全部使えます。
 - データの管理などはいつもの要領で可能
- VBAによるGUIをいくつか搭載
 - SPSSのようなインターフェースで操作ができる
 - GUIを使わなくてもたいのことはできる

変数の編集・作成

- 変数ラベルや値ラベルをつける
 - 変数名以外に, ラベルを付けられる
 - 数値データに, 値ラベル(1=統制条件など)をつけられる
- 変数の作成
 - 変数の合成: 合計値, 平均値, 主成分得点
 - 変数の変換: 非線形変換, 標準化, 中心化, 2値化
 - ダミー変数の作成: カテゴリデータを2値ダミー変数に変換
- データセットの再構成
 - 個人単位のデータをグループ単位に変換
 - 横並びデータを縦並びデータに変換(逆も可)

基礎統計分析

- データの要約
 - 要約統計量, 度数分布, ヒストグラム, 箱ひげ図
 - 散布図, クロス表, バブルチャート
- 基礎統計分析
 - 相関係数, 順序相関, ポリコリック相関
 - 独立性の検定 (χ^2 乗検定), 残差分析
 - 尺度分析, 信頼性分析
- 統計的推定・検定
 - 区間推定, 一標本の検定
 - 正規性の検定, 一様性の検定
 - 平均値の差の検定 (t 検定), 順位の差の検定 (ノンパラ)

多変量解析

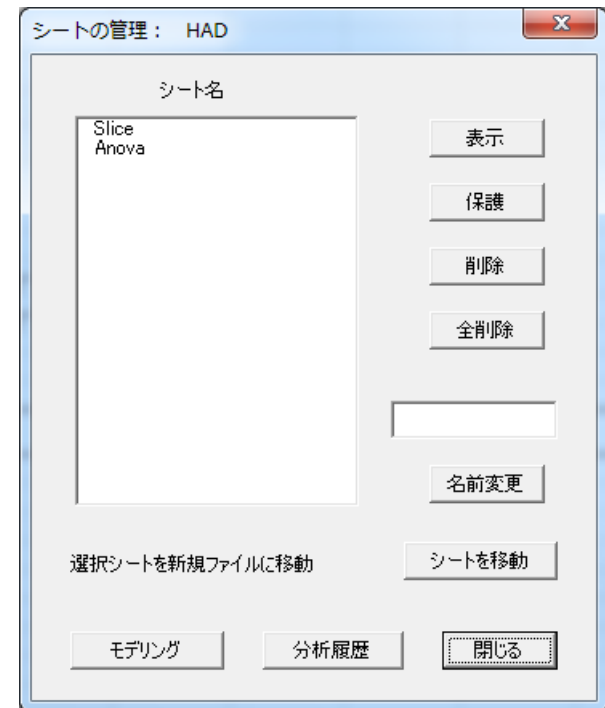
- 線形モデル
 - 分散分析, 多重比較, 単純効果分析
 - 重回帰分析, 判別分析, 多変量回帰分析, ロバスト回帰分析
 - 一般化線形モデル(順序, 名義, カウント, 打ち切り, 割合, 混合分布)
- 潜在因子の推定
 - 因子分析, 項目反応理論, 構造方程式モデル
- データの分類・圧縮
 - クラスタ分析, 混合正規分布モデル, 潜在クラス分析, 潜在ランク分析
 - 数量化分析(等質性分析・数量化Ⅲ類)
- マルチレベル分析
 - 級内相関係数, マルチレベル相関分析
 - 階層線形モデル, マルチレベル構造方程式モデル

その他分析の機能

- ラベルで結果を表示
 - 値ラベルや変数ラベルで出力
 - 結果が見やすくなる
- フィルタとグループ分け
 - 特定の値のサブジェクトだけ使用(または除外)
 - カテゴリカル変数で群分けして, それぞれに分析
- モデルの保存と再現
 - 複雑な多変量解析のモデルをHAD内部に保存
 - HADの分析履歴を保存
 - 得点化や分析結果を再現できる

シートの出力と管理

- シートはデフォルトでは上書きされる
 - 同じ分析の場合に上書きされる
 - 「出力を上書きしない」をチェックすると別の名前で出力される
- 出力シートが多い・・・
 - 「シート管理」のボタンを押す
 - 選択シートの表示
 - 選択シートの削除
 - シートの名前の変更
 - 新規ブックへシートを移動などができる



HADのコンセプト

なんでこんなもん作ったのか

- 清水が統計の勉強をするため
 - とても勉強になりました
- 統計が嫌いでも、心理学は嫌いにならないで
 - 分析のストレスを0に近づければ、純粹に学問を楽しめるのではないか
 - ただの研究の道具を、無駄に壁と思ってほしくない
- 統計そのものも、好きになってほしい
 - 家でも自分のデータを好きなだけ分析してほしい

HAD制作のコンセプト

- 使う人にとって, できるだけ使いやすく
 - 自動化できるところはできるだけ自動で
 - 一度覚えたら, 同じルールで他の分析も可能
 - 必要最低限の出力で, わかりやすさ重視
- 教える人にとって, できるだけ教えやすく
 - 学習者のレベルによって, 扱う範囲を限定できる
 - 応用的な分析も, 基礎分析の延長で実行できる
 - 結果を簡単に再現できる

こういう人にオススメ

- (研究職を志望しない) 大学生・社会人院生
 - 家でも統計分析がしたい人に
 - 大学にあまりいけないので, 自分のPCで分析したい
- 大学教員(教育用)
 - SPSSなどの商用ソフトのアカウントをそろえられない
 - 分析指導の負担を減らしたい
- (年配の) 研究者
 - すぐ分析をして結果を出したい
 - SPSSでできない分析をしたいけど, いまからRを覚えるのはちょっと...

こういう人にはオススメしない

- 研究職を志望する大学生
 - Rにチャレンジしよう！
- 大学院生
 - Rを使おう！
- ポスドク, 若手研究者
 - Rを使おう！

利用における注意点とお願い

HADの利用における注意点

- 計算は速くありません。
 - 清水のプログラミングの下手さとVBAの制約上、計算は遅いです。
 - 大規模データの分析や、シミュレーション計算には向きません。
- 結果は必ずしも信頼できるとは限りません。
 - 一応、SPSS, SAS, MplusやRと結果が一致することを確認しています。
 - しかし、バグや不具合が完全にはいえません。最終的な研究報告の際には自己責任でお願いします。

HADの利用における注意点

- 無保証です。
 - HADを利用することによって生じたいかなる問題についても、清水は保証しません
- 頻繁に更新されます。
 - 思いつきで機能が増えたり、報告された不具合が修正されます
 - できれば最新版を利用して下さい
 - Twitterで逐一報告しています (@simizu706)

お願い

- 不具合を見つけたら報告してください
 - Twitterで不具合をリプライで報告してもらえると、できるだけ早めに対応します
 - @simizu706まで
- 不特定多数に配布する場合
 - 基本自由ですが・・・
 - 大学の授業で使いたい，研究会で共有したい
 - という場合は，ご一報いただけると清水が喜びます

最後に

- Excelがあればタダで統計分析ができる
 - 卒論・修論ぐらいならこれでなんとかなります
 - 教育ツールとしても活用できます
- HADについてより詳細に知りたい人は・・・
 - <http://norimune.net/had> をご覧ください