

2020/5/11 JMP実習③

# ロジスティック回帰



北海道大学 医学統計学  
横田 勲

## 今回の内容

2

- ▶ 計算式の習得
- ▶ Cochran-Armitageの傾向検定
- ▶ Mantel傾向検定
- ▶ 除外の習得
- ▶ ROC曲線の描画

## サンプルデータを利用

3

- ▶ ヘルプ> サンプルデータ
- ▶ 医学研究のLipid data

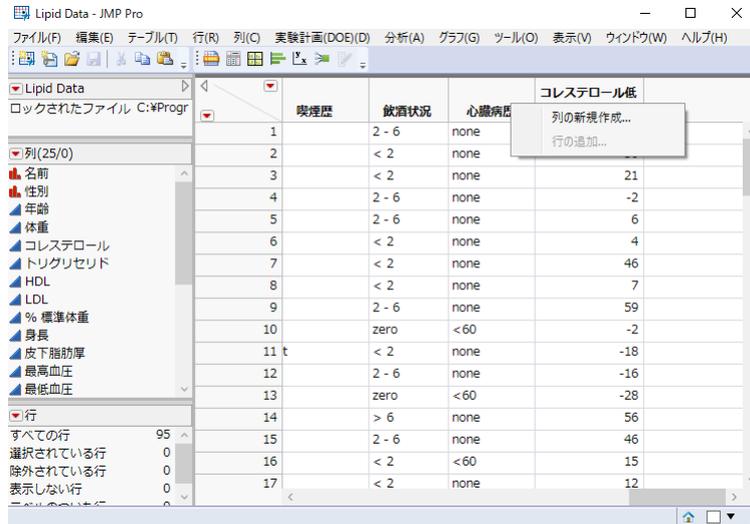
## 計算式の習得

4

- ▶ 目標（Lipidデータを使用）
  - ▶ アウトカム：3年後のLDLの改善
    - ▶ わずかでも減っていれば改善
  - ▶ 曝露：BMI
    - ▶ やせ：18.5未満
    - ▶ 正常：18.5以上、25未満
    - ▶ 肥満：25以上

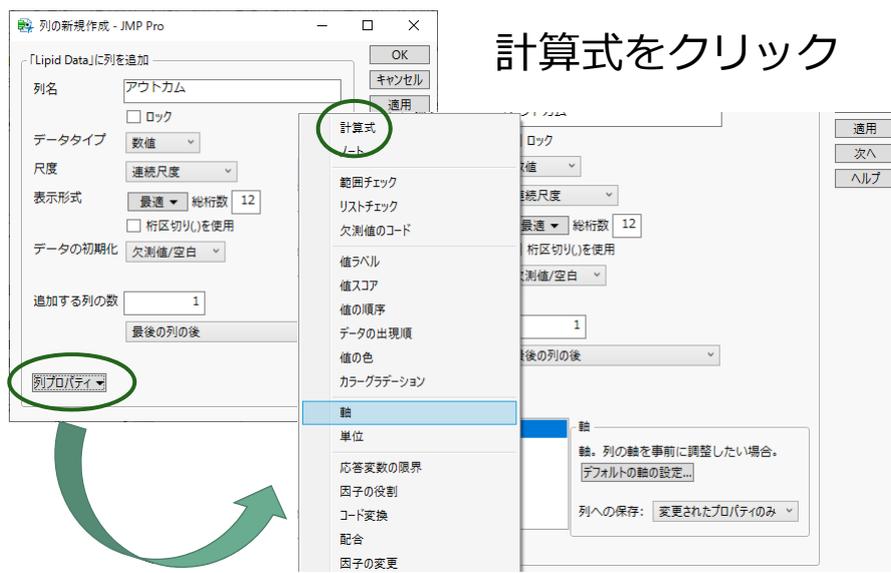
## 変数がない位置で右クリック

5



## 列名を適宜変更

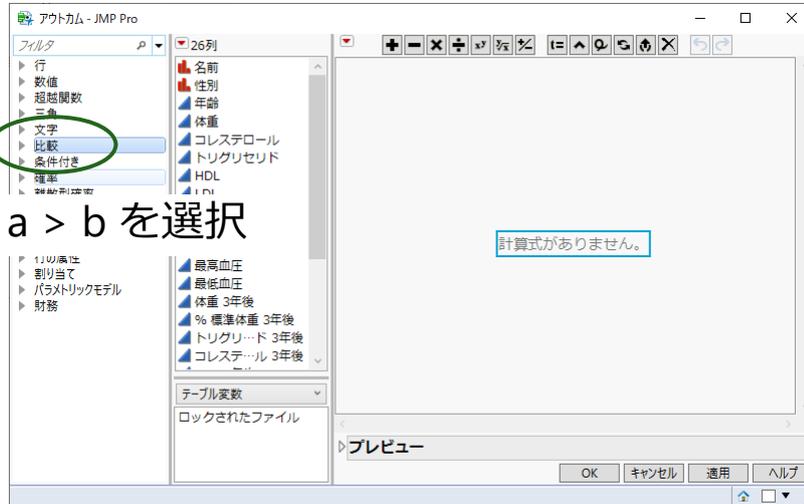
6



## 計算式をクリック

## 計算式を作成①

7

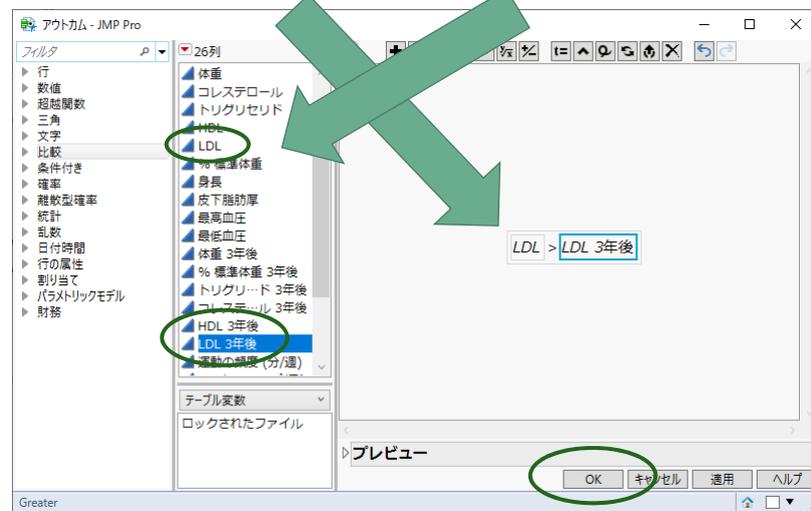


a > b を選択

## 計算式を作成②

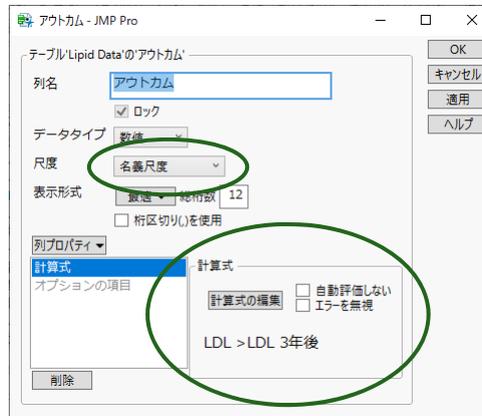
8

枠を選んで、列を選択



## 適宜名義尺度に変更

9



計算式が間違いないか確認

## 改善した人では1

10

Lipid Data - JMP Pro

ファイル(F) 編集(E) テーブル(T) 行(R) 列(C) 実験計画(DOE)(O) 分析(A) グラフ(G) ツール(S) 表示(V) ウィンドウ(W) ヘルプ(H)

行	飲酒状況	心臓病歴	コレステロール値 下	アウトカム
1	2 - 6	none	15	1
2	< 2	none	30	1
3	< 2	none	21	1
4	2 - 6	none	-2	1
5	2 - 6	none	6	0
6	< 2	none	4	0
7	< 2	none	46	1
8	< 2	none	7	1
9	2 - 6	none	59	1
10	zero	< 60	-2	0
11	< 2	none	-18	0
12	2 - 6	none	-16	0
13	zero	< 60	-28	0
14	> 6	none	56	1
15	2 - 6	none	46	1
16	< 2	< 60	15	1
17	< 2	none	12	1

## 練習① 以下の列を作成

11

列の新規作成 - JMP Pro

「Lipid Data」に列を追加

列名: BMI

ロック

データタイプ: 数値

尺度: 連続尺度

表示形式: 最速 | 桁数: 12

桁区切り(,)を使用

データの初期化: 欠測値/空白

追加する列の数: 1

選択した列の後

列プロパティ

計算式

オプションの項目

計算式の編集  自動評価しない  エラを無視

$$\left( \frac{\text{体重}}{(\text{身長} \cdot \text{身長})} \right) \cdot 703$$

削除

OK  
キャンセル  
適用  
次へ  
ヘルプ

## 練習② 以下の列を作成

12

列の新規作成 - JMP Pro

「Lipid Data」に列を追加

列名: BMIグループ

ロック

データタイプ: 数値

尺度: 順序尺度

表示形式: 最速 | 桁数: 12

桁区切り(,)を使用

データの初期化: 欠測値/空白

追加する列の数: 1

選択した列の後

列プロパティ

計算式

オプションの項目

計算式の編集  自動評価しない  エラを無視

$$\left( \begin{array}{l} \text{BMI} < 18.5 \Rightarrow 1 \\ \text{If} \\ \text{else} \Rightarrow \left( \begin{array}{l} 18.5 \leq \text{BMI} < 25 \Rightarrow 2 \\ \text{else} \Rightarrow \left( \begin{array}{l} \text{BMI} \geq 25 \Rightarrow 3 \\ \text{else} \Rightarrow . \end{array} \right) \end{array} \right) \end{array} \right)$$

削除

OK  
キャンセル  
適用  
次へ  
ヘルプ

## Cochran-Armitage傾向検定①

13

### ▶ 分析 > 二変量の関係

カム	BMIグループ	BMI
1	2	21.53107997
1	2	20.369110607
1	3	25.765817901
1	1	16.980676329
0	2	21.107149559
0	2	23.144032922
1	2	20.315631451
1	3	25.439874811
1	2	23.163311583
0	2	21.278289293
0	2	24.800900732
0	3	25.54467138
0	2	20.946630532
1	2	22.961893491
1	3	28.879345016

## Cochran-Armitage傾向検定②

14

### ▶ Yにアウトカム、XにBMIグループ

各Xに対するYの分布。いろいろな分析の種類がある。

列の選択

- 28列
  - 名前
  - 性別
  - 年齢
  - 体重
  - コレステロール
  - トリグリセリド
  - HDL
  - LDL
  - % 標準体重
  - 身長
  - 皮下脂肪厚
  - 最高血圧
  - 最低血圧
  - 体重 3年後
  - % 標準体重 3年後
  - トリグリセリド 3年後
  - コレステロール 3年後

選択した列に役割を割り当てる

Y 目的変数: アウトカム  
オプション

X 説明変数: BMIグループ  
オプション

アクション

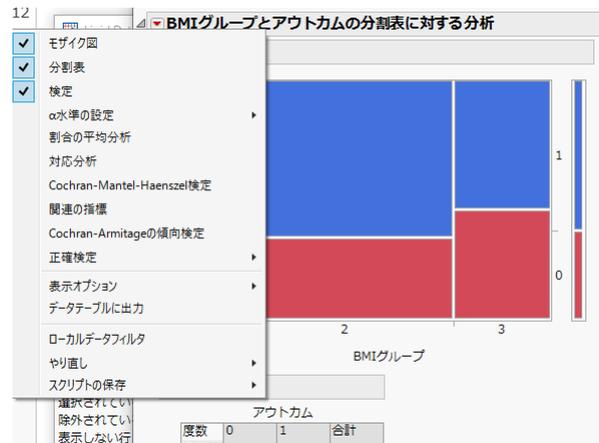
OK  
キャンセル  
削除  
前回の設定  
ヘルプ

ブロック オプション  
重み オプション(数値)  
度数 オプション(数値)  
By オプション

## Cochran-Armitage傾向検定③

15

### ▶ ▼よりCochran-Armitageの傾向検定



## 練習③ まだまだ計算式

16

- ▶ 以下のような変数を作成
- ▶ 3年後のLDLの改善を3段階で定義
  - ▶ 改善：10mg/dL以上減少
  - ▶ 悪化：10mg/dL以上増加
  - ▶ 不変：改善、悪化のいずれでもない
- ▶ BMIを「やせ・正常」と「肥満」の2水準
- ▶ ダミー層：すべて1と入力された列

## LDL3水準

17

列の新規作成 - JMP Pro

「Lipid Data」に列を追加

列名: LDL3水準

ロック

データタイプ: 数値

尺度: 順序尺度

表示形式: 最速 総桁数 12

桁区切り(,)を使用

データの初期化: 欠測値/空白

追加する列の数: 1

選択した列の後

列プロパティ

計算式

オプションの項目

計算式の編集  自動評価しない  エラーを無視

$$\text{If} \left( \begin{array}{l} \left( \text{LDL 3年後} - \text{LDL} \right) < -10 \Rightarrow 1 \\ \text{else} \Rightarrow \text{If} \left( \begin{array}{l} \left( \text{LDL 3年後} - \text{LDL} \right) > 10 \Rightarrow 3 \\ \text{else} \Rightarrow 2 \end{array} \right) \end{array} \right)$$

削除

OK  
キャンセル  
適用  
次へ  
ヘルプ

## BMI2水準

18

列の新規作成 - JMP Pro

「Lipid Data」に列を追加

列名: BMI2グループ

ロック

データタイプ: 数値

尺度: 名義尺度

表示形式: 最速 総桁数 12

桁区切り(,)を使用

データの初期化: 欠測値/空白

追加する列の数: 1

選択した列の後

列プロパティ

計算式

オプションの項目

計算式の編集  自動評価しない  エラーを無視

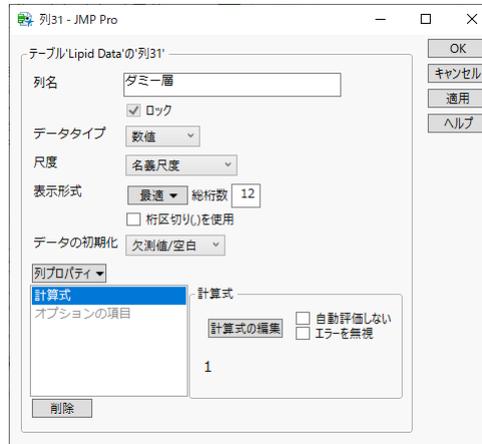
$$\left( \text{BMI2グループ} == 3 \right) + 1$$

削除

OK  
キャンセル  
適用  
次へ  
ヘルプ

## ダミー層

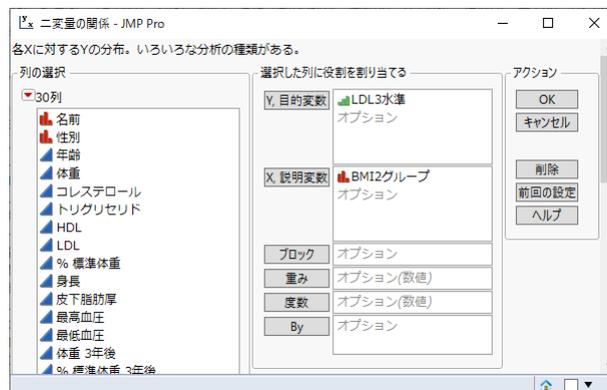
19



## Mantel傾向検定①

20

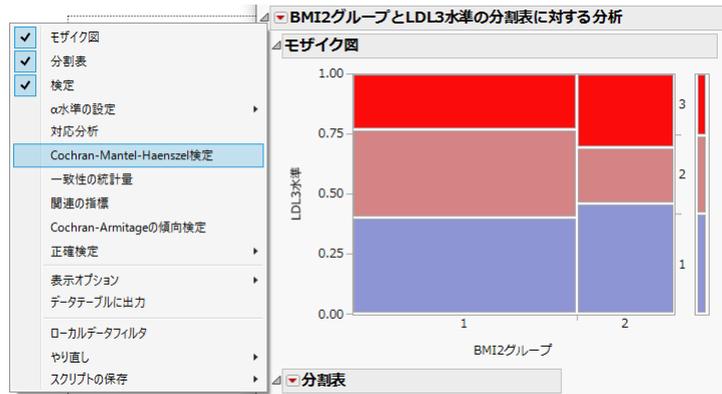
- ▶ 分析 > 二変量の関係
- ▶ YにLDL3水準、XにBMI2グループ



## Mantel傾向検定②

21

### ▶ ▼ よりCochran-Mantel-Haenszel検定



## Mantel傾向検定③

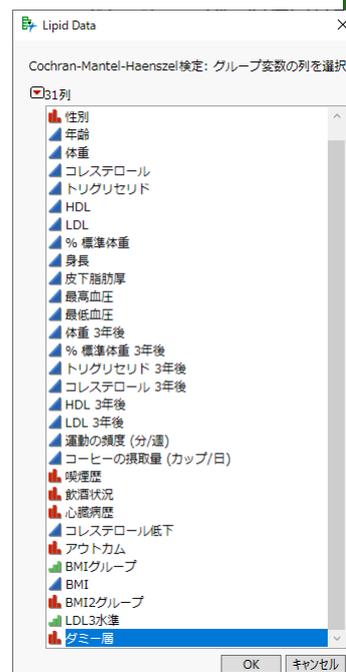
22

### ▶ グループ変数の列

に

ダミー層

を指定



## Mantel傾向検定④

23

- ▶ X間のスコアの比較 に注目

層別変数: ダミー層			
Cochran-Mantel-Haenszel検定	カイ2乗	自由度	p値(Prob>ChiSq)
スコアの相関	0.0022	1	0.9622
X間でのスコア比較	0.0022	1	0.9622
Y間でのスコア比較	0.7741	2	0.6790
カテゴリの一般連関	0.7741	2	0.6790

▶ 度数

## 除外の習得

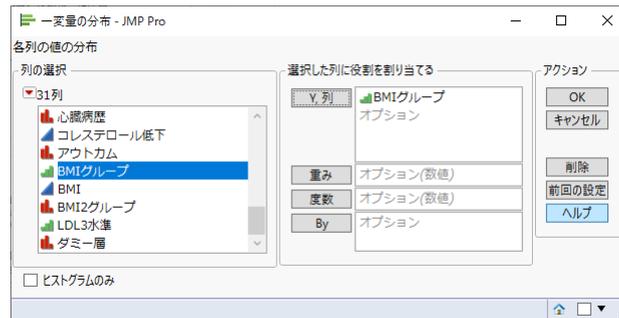
24

- ▶ 目標 (Lipidデータを使用)
  - ▶ BMIグループでのやせを除外し、性別のサブグループで正常と肥満を比較
- ▶ やせの除外には2通り
  - ▶ データフィルタを用いたやり方
  - ▶ 1変量の関係を用いたやり方

## やり方其の壱①

25

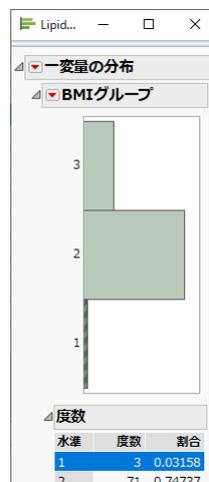
- ▶ 分析> 一変量の関係
- ▶ YにBMIグループ



## やり方其の壱②

26

- ▶ ヒストグラムの1の部分をクリック



## やり方其の壱③

27

- ▶ データに戻ってみる
- ▶ 除外する/しないをクリック

BMIグループ	BMI
1	21.53107997
2	20.369110607
3	25.765817901
4	16.980676329
5	21.107149559
6	23.144032922
7	20.315631451
8	25.439874811
9	23.163311583
10	21.278289293
11	24.800900732
12	25.54467138
13	20.946630532

## やり方其の壱④

28

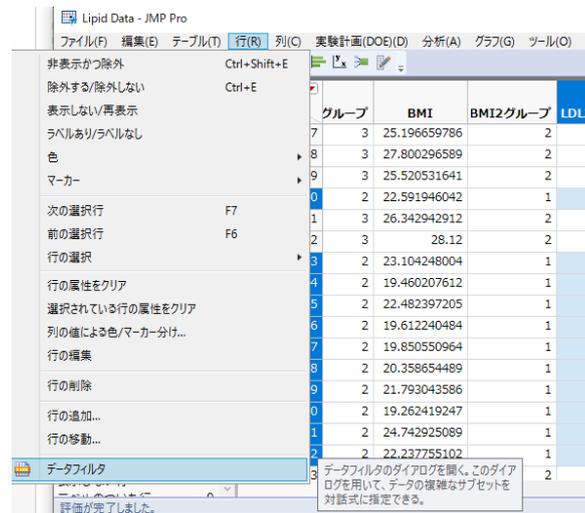
- ▶ 除外データに 

BMIグループ	BMI	BMI2グループ	LDL3水準	タミー層
1	21.53107997	1	2	1
2	20.369110607	1	1	1
3	25.765817901	2	1	1
4	16.980676329	1	2	1
5	21.107149559	1	2	1
6	23.144032922	1	2	1
7	20.315631451	1	1	1
8	25.439874811	2	1	1
9	23.163311583	1	1	1
10	21.278289293	1	3	1
11	24.800900732	1	3	1
12	25.54467138	2	3	1
13	20.946630532	1	3	1
14	22.961893491	1	1	1
15	28.879345016	2	1	1
16	21.446803841	1	2	1
17	21.452002543	1	1	1

## やり方其の式①

29

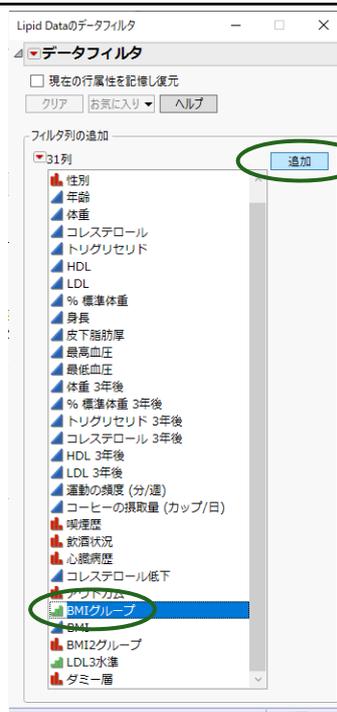
### ▶ 行> データフィルタ



## やり方其の式②

30

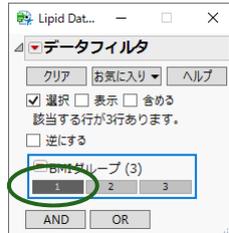
- ▶ フィルタ列の追加
  - ▶ BMIグループを選択
- ▶ 追加をクリック



## やり方其の式③

31

- ▶ 1をクリック ▶ 除外する/しないをクリック



BMIグループ	BMI
1	21.53107997
2	20.369110607
3	25.765817901
4	16.980676325
5	21.107149555
6	23.144032922
7	20.315631451
8	25.439874811
9	23.163311583
10	21.278289293
11	24.800900732
12	25.54467138
13	20.946630532

## やり方其の式④

32

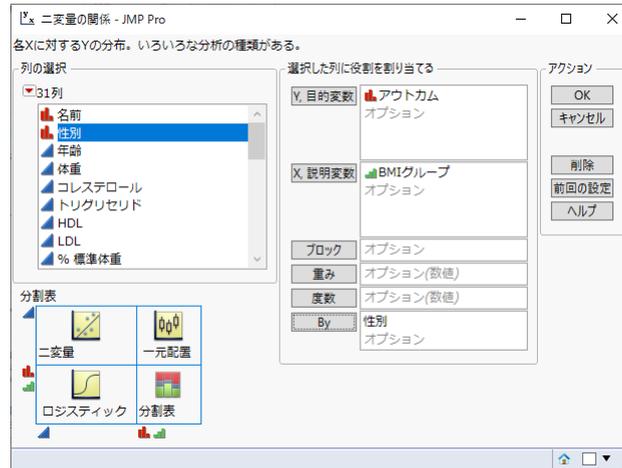
- ▶ 除外データに 

BMIグループ	BMI	BMI2グループ	LDL3水準	タミー層
1	21.53107997	1	2	1
2	20.369110607	1	1	1
3	25.765817901	2	1	1
4	16.980676325	1	2	1
5	21.107149559	1	2	1
6	23.144032922	1	2	1
7	20.315631451	1	1	1
8	25.439874811	2	1	1
9	23.163311583	1	1	1
10	21.278289293	1	3	1
11	24.800900732	1	3	1
12	25.54467138	2	3	1
13	20.946630532	1	3	1
14	22.961893491	1	1	1
15	28.879345016	2	1	1
16	21.446803841	1	2	1
17	21.452002543	1	1	1

## サブグループ解析

33

### ▶ サブグループ因子をByに



## ROC曲線

34

### ▶ アウトカム（3年後のLDLの改善有無）を

- ▶ LDLで予測した場合
- ▶ BMIで予測した場合

を求めてみよう

### ▶ 2つのROC曲線を比較しよう

## ROC曲線を描く①

35

### ▶ 分析> モデルのあてはめ



## ROC曲線を描く②

36

### ▶ ▼ よりROC曲線

名義ロジスティックのあてはめ アウトカム

効果の要約

ロジスティックプロット

勾配で収束しました, 3回の反復

反復履歴

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
Model 1	6.214450	1	1.422220	0.2311

ロジスティックプロットのプロットオプション

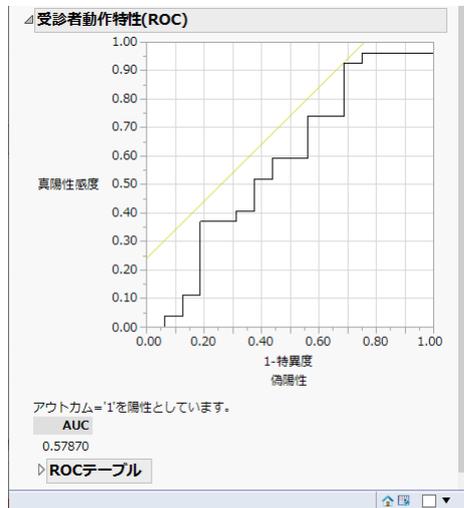
- ロジスティックプロット
- プロットオプション
- 尤度比検定
- Wald検定
- 信頼区間
- オッズ比
- 逆推定...
- 確率の計算式の保存
- 確率の計算式を発行
- ROC曲線
- リフトチャート
- 混同行列
- 利益行列の指定
- プロファイル
- モデルダイアログ
- 効果の要約
- ローカルデータフィルタ
- やり直し
- スクリプトの保存

高陽性(感度)の割合に対する偽陽性(1-特異度)の割合をプロットする。陽性(Positive Level)メッセージを使えば、陽性水準を設定できる。

飲酒状況	心臓病歴
2 - 6	none
< 2	none
< 2	none
2 - 6	none
< 2	none
2 - 6	none
zero	< 60
< 2	none
2 - 6	none
zero	< 60
> 6	none

## ROC曲線を描く③

37



## 予測確率を保存①

38

## ▶ ▼ より確率の計算式の保存

歴	飲酒状況	心臓病歴
	2 - 6	none
	< 2	none
	< 2	none
	2 - 6	none
	2 - 6	none

現在のデータテーブルに新しい列を作成し、  
応答変数の水準ごとに、説明変数とパラ  
メータの線形結合式、各水準が生じる確  
率の予測式、そして、最も生じる確率が高  
い応答変数の値を計算する予測式を保  
存する。

歴	飲酒状況	心臓病歴
	< 2	none
	2 - 6	none
	zero	<60
	> 6	none

ロジスティックプロット  
プロットのオプション  
尤度比検定  
Wald検定  
信頼区間  
オッズ比  
逆推定...

確率の計算式の保存  
確率の計算式を発行  
ROC曲線  
リフトチャート  
混同行列  
利益行列の指定  
プロフィール  
モデルダイアログ  
効果の要約  
ローカルデータフィルタ  
やり直し  
スクリプトの保存

名義ロジスティックのあてはめ アウトカム  
反復推定

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob)
差	0.244468	1	0.488936	
完全	28.138122			
縮小	28.382590			

R2乗(U) 0.0086  
AICc 60.5762  
BIC 63.7986  
オブザベーション(または重みの合計) 43

あてはめの詳細

あてはまりの悪さ(LOF)

要因	自由度	(-1)*対数尤度	カイ2乗	p値(Prob)
あてはまりの悪さ(LOF)	41	28.138122	56.2	
飽和モデル	42	0.000000		
あてはめたモデル	1	28.138122		

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>Cl)
切片	-0.4271838	1.3999971	0.09	0.7
LDL	0.00658453	0.0095125	0.48	0.4

推定値は次の対数オッズに対するものです: 1/0

## 予測確率を保存②

39

### ▶ データセットに予測確率が追加

タミ-層	線形[1]	確率[1]	確率[0]	最尤アウトカム
1	0.5708201697	0.6389524035	0.3610475965	1
2	0.3633284788	0.5898459254	0.4101540746	1
3	0.5415848591	0.6321810193	0.3678189807	1
4	0.0477978291	0.5119471828	0.4880528172	1
5	0.3729155536	0.5921632907	0.4078367093	1
6	0.8119192989	0.6925183444	0.3074816556	1
7	0.5192501353	0.6269724062	0.3730275938	1
8	0.287738081	0.5714422878	0.4285577122	1
9	0.8968860665	0.7103091682	0.2896908318	1
10	0.5578091396	0.6359454661	0.3640545339	1
11	1.0425885334	0.7393491565	0.2606508435	1
12	1.05089255932	0.6245545747	0.3754454253	1
13	0.2672997017	0.566429866	0.433570134	1
14	0.984750027	0.7280497081	0.2719502919	1

## 練習④ ROC曲線

40

- ▶ アウトカム（3年後のLDLの改善有無）をBMIで予測した場合のROC曲線を求めてみよう
- ▶ 予測確率を同様に保存しよう



## ROC曲線を比べる③

43

▶ ▼よりROC曲線、AUCの比較をクリック

The screenshot shows the JMP Pro interface. A menu is open over the 'Model Comparison' window, with options like 'ROC Curves', 'Compare AUCs', 'Lift Charts', etc. The 'Model Comparison' window displays a table of model results.

Model	AUC	Standard Error	Lower 95%	Upper 95%
1	0.6096292033	0.047856		
1	0.8290452874	0.09615		
1	0.4190142699	0.603247		
1	0.3324395679	0.582352		
1	0.8676502773	0.704256		
1	0.6330742147	0.653186		

## ROCを比べる④

44

