

保健統計学実習

第3日目

- 第7回 重回帰分析
- 第8回 ロジスティック回帰、検査データの解析
- 第9回 調査データ解析(2):(web調査ツールを使用した)調査票作成、データ入力

滋賀医科大学NCD疫学研究センター
医療統計学部門

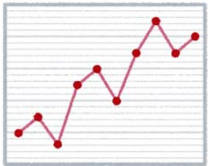
原田 亜紀子
(aharada@belle.shiga-med.ac.jp)

講義・演習スケジュール

1 R, EZRの使い方、データセットの読み込み、頻度集計、記述統計, 相関	8/31(木)
2 EZRのコード保存, R-studio commander	
3 エクセルの基礎(1)	
4 仮説検定の基礎, 2群の比較(t検定, Wilcoxon検定)	9/1(金)
5 カイニ乗検定、マクネマー検定	
6 調査データ解析(1): 調査票作成、データ入力	
7 重回帰分析	9/5(火)
8 ロジスティック回帰、検査データの解析	
9 調査データ解析(2):(web調査ツールを使用した)調査票作成、データ入力	
10 分散分析	9/6(水)
11 サンプルサイズ	
12 調査データ解析(3): 解析用データの作成	
13 主成分分析、因子分析、クラスター分析	9/8(金)
14 解析実習・まとめ(復習・課題の時間)	

2

第7回 重回帰分析



3

解析の手順 線形モデルのあてはめとモデル診断

なぜ線形モデルを使うのか？

- データ削減
- 予測式を算出したい
- パラメータ値の解釈(従属変数に対する独立変数の影響)

決定係数、残差の検討を丁寧に行う

仮定は正しいか？

- $y = x + \text{誤差}$
- 誤差の変動: 平均は0, 分散は等しい, 独立, ほぼ正規分布

偏重回帰係数の有意性を重視してモデルを作る

回帰診断

- 残差の検討と影響度分析



4

線形回帰

- xに対するyの線形回帰
- y: 応答, 従属変数
- x: 説明変数, 独立変数

予測

$\hat{y} = a + bx$

残差

$y - \hat{y}$

確率分布は等分散正規分布

Fig.1 : (General) liner model

yという結果に対して、原因と考えられるxがどのように影響しているかを検討する手法

(例) 握力 = a + b₁×(年齢) + b₂×(性別) + b₃×(体重)

5

回帰と最小二乗法

残差が最小となるよう、係数を推定する

$y = b_0 + b_1x_{1n} + b_px_{pn}$

実測値(y)	予測値(\hat{y})	誤差 e_j
y_1	$b_0 + b_1x_{11} + \dots + b_px_{p1}$	$y_1 - \hat{y}_1$
y_2	$b_0 + b_1x_{12} + \dots + b_px_{p2}$	$y_2 - \hat{y}_2$
y_n	$b_0 + b_1x_{1n} + \dots + b_px_{pn}$	$y_n - \hat{y}_n$

残差が最も小さくなるように回帰直線を引く
(そのように、データから b_0 と b_1 を推定する)

6

1. モデルのあてはまり

2. 変数選択によるモデル構築

3. モデルの解釈・評価

7

線形モデルフィッティング

Anscombe's quartet

データセット
Ans1.csv~Ans4.csv

- 単純な記述統計量はほぼ同じだが、分布が大きく異なり、グラフにすると大きく異なって見える4つのデータセット。
- これらは、統計学者Francis Anscombeによって、データを分析する前にグラフ化することの重要性と、外れ値やその他の影響力のある観測値が統計的性質に及ぼす影響の両方を示すために、1973年に作成されたものです。

I

II

III

IV

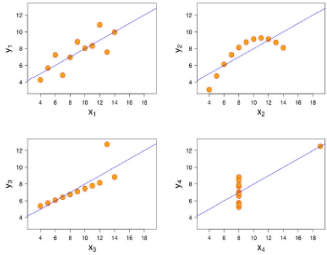
	I	II	III	IV				
X ₁	Y ₁	X ₂	Y ₂	X ₃	Y ₃	X ₄	Y ₄	
10	8.04	10	9.14	10	7.46	8	6.58	
8	6.95	8	8.14	8	6.77	8	5.76	
13	7.58	13	8.74	13	12.74	8	7.71	
9	8.81	9	8.77	9	7.11	8	8.84	
11	8.33	11	9.26	11	7.81	8	8.47	
14	9.96	14	8.10	14	8.84	8	7.04	
6	7.24	6	6.13	6	6.08	8	5.25	
4	4.26	4	3.10	4	5.39	19	12.5	
12	10.84	12	9.13	12	8.15	8	5.56	
7	4.82	7	7.26	7	6.42	8	7.91	
5	5.68	5	4.74	5	5.73	8	6.09	
AVG	9	7.50	9	7.50	9	7.50	7.50	
VAR	11	4.127	11	4.128	11	4.123	11	4.123
CORREL	0.816	0.816	0.816	0.817				

8

線形モデルフィッティング
Anscombe's quartet

データセット
Ans1.csv~Ans4.csv

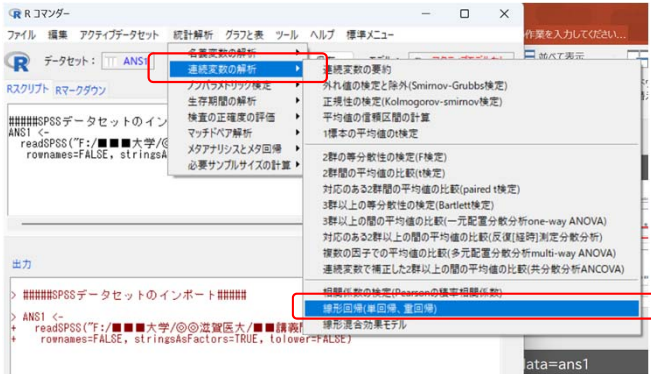
形状は全く異なるが、要約統計では類似している4つのデータセット



	I	II	III	IV
	X ₁	X ₂	X ₃	X ₄
Y ₁	10	8.04	10	9.14
Y ₂	8	6.95	8	8.14
Y ₃	13	7.58	13	8.74
Y ₄	9	8.81	9	8.77
Y ₅	11	8.33	11	9.26
Y ₆	14	9.96	14	8.10
Y ₇	6	7.24	6	6.13
Y ₈	4	4.26	4	3.10
Y ₉	12	10.84	12	9.13
Y ₁₀	7	4.82	7	7.26
Y ₁₁	5	5.68	5	4.74
Y ₁₂	9	7.50	9	7.50
Y ₁₃	11	4.127	11	4.128
Y ₁₄	10	4.127	10	4.123
Y ₁₅	11	4.127	11	4.123
Y ₁₆	11	4.127	11	4.123
Y ₁₇	11	4.127	11	4.123
Y ₁₈	11	4.127	11	4.123
Y ₁₉	11	4.127	11	4.123
Y ₂₀	11	4.127	11	4.123
Y ₂₁	11	4.127	11	4.123
Y ₂₂	11	4.127	11	4.123
Y ₂₃	11	4.127	11	4.123
Y ₂₄	11	4.127	11	4.123
Y ₂₅	11	4.127	11	4.123
Y ₂₆	11	4.127	11	4.123
Y ₂₇	11	4.127	11	4.123
Y ₂₈	11	4.127	11	4.123
Y ₂₉	11	4.127	11	4.123
Y ₃₀	11	4.127	11	4.123
Y ₃₁	11	4.127	11	4.123
Y ₃₂	11	4.127	11	4.123
Y ₃₃	11	4.127	11	4.123
Y ₃₄	11	4.127	11	4.123
Y ₃₅	11	4.127	11	4.123
Y ₃₆	11	4.127	11	4.123
Y ₃₇	11	4.127	11	4.123
Y ₃₈	11	4.127	11	4.123
Y ₃₉	11	4.127	11	4.123
Y ₄₀	11	4.127	11	4.123
Y ₄₁	11	4.127	11	4.123
Y ₄₂	11	4.127	11	4.123
Y ₄₃	11	4.127	11	4.123
Y ₄₄	11	4.127	11	4.123
Y ₄₅	11	4.127	11	4.123
Y ₄₆	11	4.127	11	4.123
Y ₄₇	11	4.127	11	4.123
Y ₄₈	11	4.127	11	4.123
Y ₄₉	11	4.127	11	4.123
Y ₅₀	11	4.127	11	4.123
Y ₅₁	11	4.127	11	4.123
Y ₅₂	11	4.127	11	4.123
Y ₅₃	11	4.127	11	4.123
Y ₅₄	11	4.127	11	4.123
Y ₅₅	11	4.127	11	4.123
Y ₅₆	11	4.127	11	4.123
Y ₅₇	11	4.127	11	4.123
Y ₅₈	11	4.127	11	4.123
Y ₅₉	11	4.127	11	4.123
Y ₆₀	11	4.127	11	4.123
Y ₆₁	11	4.127	11	4.123
Y ₆₂	11	4.127	11	4.123
Y ₆₃	11	4.127	11	4.123
Y ₆₄	11	4.127	11	4.123
Y ₆₅	11	4.127	11	4.123
Y ₆₆	11	4.127	11	4.123
Y ₆₇	11	4.127	11	4.123
Y ₆₈	11	4.127	11	4.123
Y ₆₉	11	4.127	11	4.123
Y ₇₀	11	4.127	11	4.123
Y ₇₁	11	4.127	11	4.123
Y ₇₂	11	4.127	11	4.123
Y ₇₃	11	4.127	11	4.123
Y ₇₄	11	4.127	11	4.123
Y ₇₅	11	4.127	11	4.123
Y ₇₆	11	4.127	11	4.123
Y ₇₇	11	4.127	11	4.123
Y ₇₈	11	4.127	11	4.123
Y ₇₉	11	4.127	11	4.123
Y ₈₀	11	4.127	11	4.123
Y ₈₁	11	4.127	11	4.123
Y ₈₂	11	4.127	11	4.123
Y ₈₃	11	4.127	11	4.123
Y ₈₄	11	4.127	11	4.123
Y ₈₅	11	4.127	11	4.123
Y ₈₆	11	4.127	11	4.123
Y ₈₇	11	4.127	11	4.123
Y ₈₈	11	4.127	11	4.123
Y ₈₉	11	4.127	11	4.123
Y ₉₀	11	4.127	11	4.123
Y ₉₁	11	4.127	11	4.123
Y ₉₂	11	4.127	11	4.123
Y ₉₃	11	4.127	11	4.123
Y ₉₄	11	4.127	11	4.123
Y ₉₅	11	4.127	11	4.123
Y ₉₆	11	4.127	11	4.123
Y ₉₇	11	4.127	11	4.123
Y ₉₈	11	4.127	11	4.123
Y ₉₉	11	4.127	11	4.123
Y ₁₀₀	11	4.127	11	4.123

モデル診断をやってみよう

回帰分析を行う



線形回帰(単回帰, 重回帰)

モデル名を入力: RegModel.1
複数の変数を選択する場合はCtrlキーを押しながらクリック。
目的変数 (1つ選択)
説明変数 (1つ以上選択)

☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)
☐ モデル解析結果をグラフとして表示
☒ 図表診断プロットを表示する
☐ AICを用いたステップワイズの実数選択を行う。
☐ BICを用いたステップワイズの実数選択を行う。
☐ P値を用いたステップワイズの実数選択(減少)を行う。
一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 Page<50
<全ての有効なケース>

Call:
lm(formula = y ~ x, data = ans1)
Residuals:
Min 1Q Median 3Q Max
-1.92127 -0.45577 -0.04136 0.70941 1.83882
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.0001 1.1247 2.667 0.02573 *
x 0.5001 0.1179 4.241 0.00217 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.237 on 9 degrees of freedom
Multiple R-squared: 0.6665, Adjusted R-squared: 0.6295
F-statistic: 17.99 on 1 and 9 DF, p-value: 0.00217
分散分析
F値 17.99 P値 0.00217
> multreg.table
回帰係数推定値 95%信頼区間下限 95%信頼区間上限 標準誤差 t統計量
(Intercept) 3.0000909 0.4557949 5.4444488 1.1247468 2.6673948
x 0.5000909 0.2393701 0.7668117 0.1179055 4.241455
P値
(Intercept) 0.025734051
x 0.002169629

偏回帰係数、標準偏回帰係数

偏回帰係数

偏回帰係数は重回帰モデルにおける独立変数の係数である
他の独立変数の影響を除外した回帰係数となる

標準偏回帰係数(β)

平均0、分散1に標準化した単位に依存しない係数(変数間の比較が行えるようにする)
各独立変数が従属変数にどのくらい影響しているかを評価できる

標準偏回帰係数

- 標準偏回帰係数(A_i) a_i $Sx:x$ の標準偏差, $SDy:y$ の標準偏差

$$A_i = a_i \times \frac{SD_x}{SD_y}$$
$$= 0.50 \times (3.31 / 2.03) = 0.815$$

- 標準化したデータセットで重回帰
 - `z <- scale(Ans1)`
 - `z <- data.frame(z)`

データセットZで重回帰分析を実施

	回帰係数推定値	95%信頼区間下限	95%信頼区間上限	標準誤差	t統計量	P値
(Intercept)	-1.673724e-17	-0.4151695	0.4151695	0.1835281	-9.119714e-17	1.000000000
x	8.164205e-01	0.3809871	1.2518540	0.1924859	4.241455e+00	0.002169629

13

data=ans1

モデルの要約^b

モデル	R	R ² 乗	調整済み R ² 乗	推定値の標準誤差	Durbin-Watson
1	.816 ^a	.667	.629	1.23660	3.212

分散分析^a

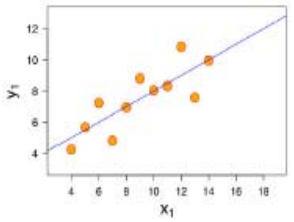
モデル		平方和	自由度	平均平方	F 値	有意確率
1	回帰	27.510	1	27.510	17.990	.002 ^b
	残差	13.763	9	1.529		
	合計	41.273	10			

a. 従属変数 y
b. 予測値: (定数), x_0

係数^a

モデル		非標準化係数 B	標準化係数 ベータ	t 値	有意確率	B の 95.0% 信頼区間 下限	上限	ゼロ次	相関 偏	部分	共線性の統計量 許容度	VIF
1	(定数)	3.000	1.125	2.667	.026	.456	5.544					
	x	.500	.118	.816	.002	.233	.767	.816	.816	.816	1.000	1.000

a. 従属変数 y



4

R, R², 調整済みR²

重相関係数(R)

- 重回帰式から得られる予測値と実測値の相関係数
- 1に近いほどあてはまりがよい
- 変数の数が多いと1に近づく(変数の数の影響をうける)

決定係数(R²)

- 重回帰モデルの適合性を評価する指標
- 変数の数が多いと1に近づく(変数の数の影響をうける)

自由度調整済み重相関係数・決定係数

- 独立変数の数、nを補正した指標

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$
$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$

15

残差

- 予測値の大小によって残差の分布に違いがないかをみる
- 0を中心に上下に均等に分布しているか確認する

Q-Qプロット

- 残差が正規分布であれば直線になる

RSTUDENT

- 残差をその標準偏差推定値で割ったもの。
- 値が±3を超えている場合は異常値である。

LEVERAGE

- 説明変数空間における外れ値

Cook's D

- 問題のオブザベーションを除外したときのリグレッションの変化
- 4/nを超える場合は注意

Cook' Dでは、ID3, 9の影響度が大きい

Residuals vs Fitted

Normal Q-Q

Scale-Location

Residuals vs Leverage

Variance inflation factor (VIF)

従属変数y, 独立変数x₁, x₂, x₃, x₄で構築された重回帰モデルがあったとして、
x₁を従属変数、残りのx₂, x₃, x₄を独立変数とした
重回帰モデルを構築し、重相関係数を求める

VIF = $\frac{1}{(1-R^2)}$ (分母は“許容度”とよばれる)

VIF ≧ 10 となるような変数は除いた方がよい

17

参考: 影響度・残差関連の指標

MODEL Option or Statistic	Formula	
PRED (Y _i)	X _i b	予測値
RES (r _i)	Y _i - Ŷ _i	残差
H (h _i)	x _i (X'X) ⁻¹ x _i '	デコ比: 説明変数空間での外れ度、何個の 回帰係数を決定しているか<1
STDP	$\sqrt{h_i \sigma^2}$	
STDI	$\sqrt{(1+h_i) \sigma^2}$	
STDR	$\sqrt{(1-h_i) \sigma^2}$	
LCL	Ŷ _i - t _{α/2} STDI	
LCLM	Ŷ _i - t _{α/2} STDP	
UCL	Ŷ _i + t _{α/2} STDI	
UCLM	Ŷ _i + t _{α/2} STDP	
STUDENT	$\frac{r_i}{f_i}$	基準化した残差
RSTUDENT	$\frac{r_i}{\hat{\sigma}_{(i)} \sqrt{1-h_i}}$	基準化した残差 (分散を当該obsを除いて計算)
COOKD	$\frac{1}{p} \frac{STUDENT^2}{STDR^2}$	Cook's D (当該obsを除いたときの あてはめ結果の全体としての変化)
COVRATIO	$\frac{\det(\hat{\sigma}_{(i)}^2 (X_{(i)} X_{(i)}')^{-1})}{\det(\hat{\sigma}^2 (X'X)^{-1})}$	
DFBETIS	$\frac{(\hat{Y}_i - \hat{Y}_{(i)})}{(\hat{\sigma}_{(i)} \sqrt{h_i})}$	これを2乗しpで割り分散を置き換えたものがCook's D
DFBETAS _j	$\frac{b_j - b_{(i,j)}}{\hat{\sigma}_{(i)} \sqrt{(X'X)^{-1}_{jj}}}$	当該obsを除いた時の個々の回帰係数の変化
PRESS(pred _r)	$\frac{r_i}{1-h_i}$	

1718

data=ans2

モデルの要約^b

モデル	R	R ² 乗	調整済み R ² 乗	推定値の標準誤差	Durbin-Watson
1	.816 ^a	.666	.629	1.23721	2.188

a. 予測値: (定数)、x₁.
b. 従属変数 y

分散分析^a

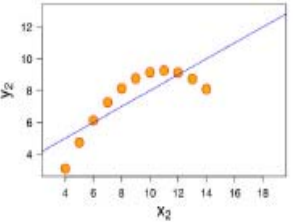
モデル	平方和	自由度	平均平方	F 値	有意確率
1 回帰	27.500	1	27.500	17.966	.002 ^b
残差	13.776	9	1.531		
合計	41.276	10			

a. 従属変数 y
b. 予測値: (定数)、x₁.

係数^a

モデル	非標準化係数 B	標準化係数 ベータ	標準誤差	t 値	有意確率	B の 95.0% 信頼区間	
						下限	上限
1 (定数)	3.001	1.125		2.667	.026	.455	5.547
x ₁	.500	.118	.816	4.239	.002	.233	.767

a. 従属変数 y



19

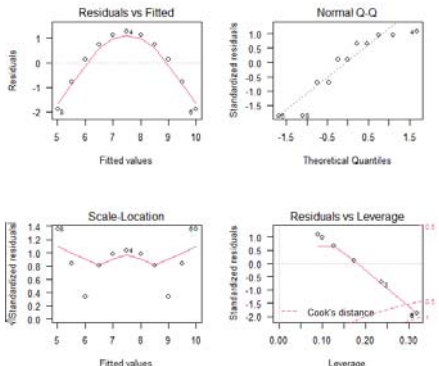
lm(y ~ x)

Call:
lm(formula = y ~ x, data = ANS2)

Residuals:
Min 1Q Median 3Q Max
-1.9009 -0.7608 0.1291 0.9491 1.2691

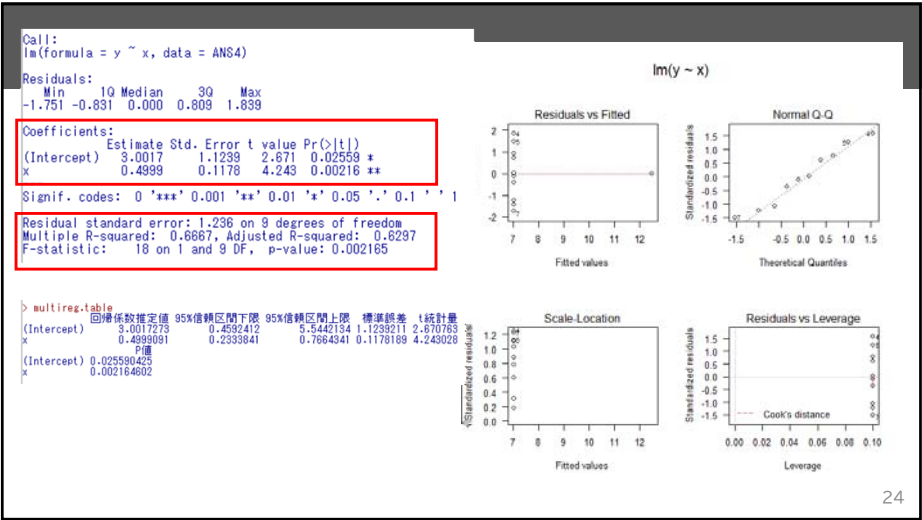
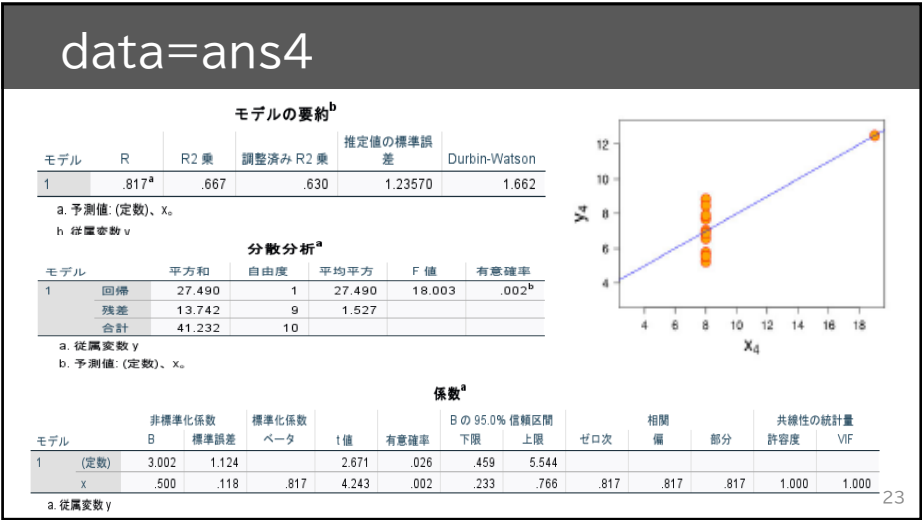
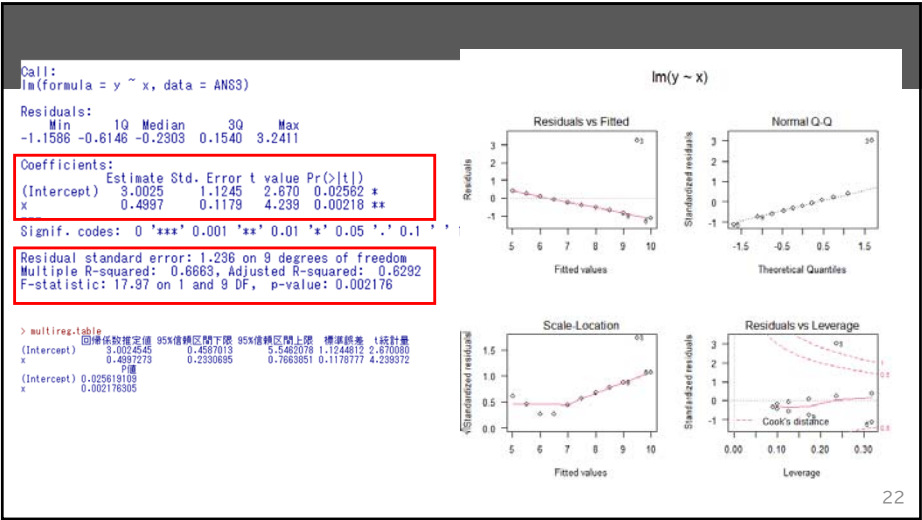
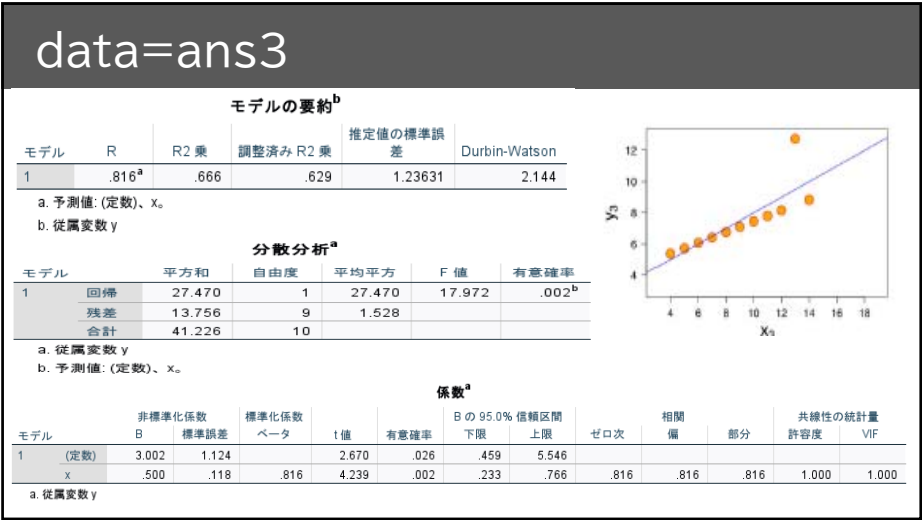
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.001 1.125 2.667 0.02576 *
x 0.500 0.118 4.239 0.00218 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.237 on 9 degrees of freedom
Multiple R-squared: 0.6662, Adjusted R-squared: 0.6292
F-statistic: 17.97 on 1 and 9 DF, p-value: 0.002179



```
> multreg.table  
回帰係数推定値 95%信頼区間下限 95%信頼区間上限 標準誤差 t統計量  
(Intercept) 3.000909 0.4552982 5.5465200 1.1253024 2.666758  
x 0.500000 0.2331475 0.7666525 0.1178637 4.239590  
F値  
(Intercept) 0.025768841  
x 0.002178816
```

20



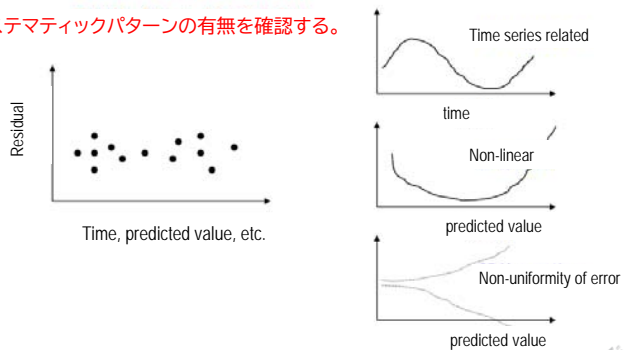
ans1~ans4

- 4つの回帰について、これだけ異なるデータの集まりだが、同じような直線が引けてしまう
 - グラフを見ずに、平均、回帰係数などの数値だけみるとほぼ同じ
- 残差に系統的なパターンが残されているなら、さらにモデル化すべき

25

残差分析

システムティックパターンの有無を確認する。



26

1. モデルのあてはまり

2. 変数選択によるモデル構築

3. モデルの解釈・評価

27

変数選択

相関の高い変数をモデルに含めると係数が不安定になる(多重共線性)。

<対策>

- ① 変数の選択
 - ALL(全変数を一度に入力)
 - FORWARD(前方、増加)
 - BACKWARD(後方、減少)
 - STEPWISE(増加/減少)を使用する。
- ② 合成変数の作成
- ③ 事前情報の利用(ベイズ型)

28

多重共線性

■ 通常の回帰分析
(x_1 と x_2 の相関が低い)

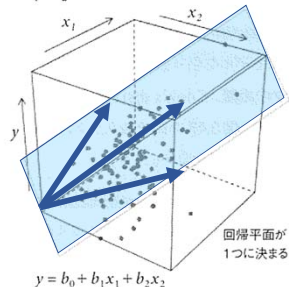
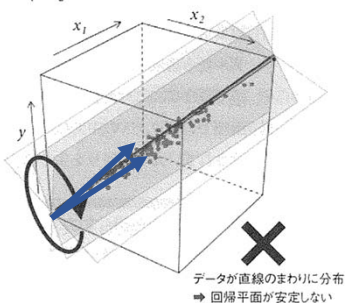


図3.46 多重共線性

■ 多重共線性がある状態
(x_1 と x_2 の相関が高い)



演習 変数選択

データセット
select.csv

“select.csv”データは、 $y=x_1+x_2+x_3$ +誤差(x_1, x_2, x_3 は独立)というモデルで作られています。

- x_4 は x_1 と高い相関がある
- x_5 は x_2 と高い相関がある
- x_6 - x_{15} は互いに独立であり、 y とも独立である。

変数選択を適用して、どのような結果が得られるか確認してください。

x_1 - x_{15} を用いてモデルを作成する。相関の高い変数を含む場合、どのような変数が選択されるのか

• AICの値をもとに

Call:
lm(formula = y ~ x1 + x2 + x3, data = TempDF)

Residuals:
Min 1Q Median 3Q Max
-2.4259 -0.7411 0.1107 0.6243 2.4522

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0003582	0.1077994	0.003	0.997
x1	1.0127592	0.1134334	8.928	2.86e-14 ***
x2	1.0043676	0.1057964	9.493	1.82e-15 ***
x3	0.8505119	0.1024468	8.302	6.50e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.053 on 96 degrees of freedom
Multiple R-squared: 0.7171, Adjusted R-squared: 0.7083
F-statistic: 81.13 on 3 and 96 DF, p-value: < 2.2e-16

• P値をもとに

-----最終モデル

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0003582008	0.1077994	0.00322845	9.973557e-01
x1	1.0127591938	0.1134334	8.928231348	2.982083e-14
x2	1.0043675576	0.1057964	9.493397871	1.823653e-15
x3	0.8505119339	0.1024468	8.301979590	6.496778e-13

重回帰分析 注意点

1. 目的の再確認

1) 予測式を算出したい
→ 決定係数や残差の検討を詳細に

2) パラメータ値の解釈(従属変数に対する独立変数の影響)
→ 偏回帰係数の有意性を重視してモデルを構築する
→ 仮説に合わせた変数の投入(強制投入)

2. 標本数と独立変数の数

独立変数1つに対し $n \geq 20 \sim 30$ 程度が目安

3. 外れ値のチェック

4. 独立変数間の相関

5. 正規分布からのズレが大きい場合は変数変換(対数変換など)

33

第8回
ロジスティック回帰

34

一般線形モデル
一般化線形モデル

Development of linear models

階層ベイズモデル

より柔軟な統計モデリング

一般化線形混合モデル

個人差や部位差などのランダム効果への対応

一般化線形モデル

正規分布以外の確率分布を扱う

一般線形モデル

Statistical Inference

MCMC

最尤法

Logistic Poisson

Linear regression Analysis of Variance

最小二乗法

35

一般線形モデル→一般化線形モデル

GLMIは正規分布だけでなく、指数族(二項、ガンマ、ポアソンなど)の分布も扱うことができます！

確率分布は正規分布

Response variable

Explanatory variables

Link Function : identity

Fig.1 : 一般線形モデル

確率分布はポアソン分布

Response variable

Explanatory variables

Link function: log

正規分布を超えて拡張することを確認

Fig.2 : 一般化線形モデル(例:ポアソン)

36

GLMは以下の3つの要素で構成される

1. 確率分布の指数族
指数族(二項、ガンマ、ポアソンなど)

2. 線形予測量
 $\eta = X\beta$
線形予測量(η)は独立変数の情報をモデルに取り込む量である。
 η は未知パラメータ β の線形結合(したがって「線形」)で表され、線形結合の係数は独立変数の行列Xとして表される。

3. リンク関数
ある式を線形に変換する関数。

37

Link function

式を線形に変換する関数

Distribution	Support	Discrete variable	Continuous variable	Probability Distribution	family	Frequently used link functions	Variance	Mean function
Normal	real: $(-\infty, +\infty)$							
Exponential	real: $(0, +\infty)$							
Gamma	real: $(0, +\infty)$							
Inverse Gaussian	real: $(0, +\infty)$							
Poisson	integer: $0, 1, 2, \dots$	Bernoulli	Binomial	binomial	logit	$\mu(1-\mu)$	$(\mu)^{-1}(1-\mu)$	$\mu(X;\beta)$
Bernoulli	integer: $\{0, 1\}$		Poisson	poisson	log	μ		
Binomial	integer: $0, 1, \dots, K$		Gamma	gamma	log $\frac{1}{\mu}$	μ^2		
Categorical	K-vector of integer element in the set		Normal	gaussian	identity	$-\mu^2$		$\frac{\exp(X;\beta)}{1 + \exp(X;\beta)}$
Multinomial	K-vector of integer: $\{0, K\}$							

38

回帰モデルの拡張

0-1データに対するロジスティック回帰

頻度データに対するポアソン回帰

生存時間データに対するCox回帰

一般化推定方程式(Generalized Estimating Equation)

- 反復測定データに対するGEE

個人差、階層、繰り返しを扱うことができるモデル(混合モデル)

- 線型:
- 一般化線型:
- 非線型:

39

Example
データセット 神経痛(N=60)

高齢者における神経痛の鎮痛剤に関する研究

痛みの有無と薬剤の関連を検討

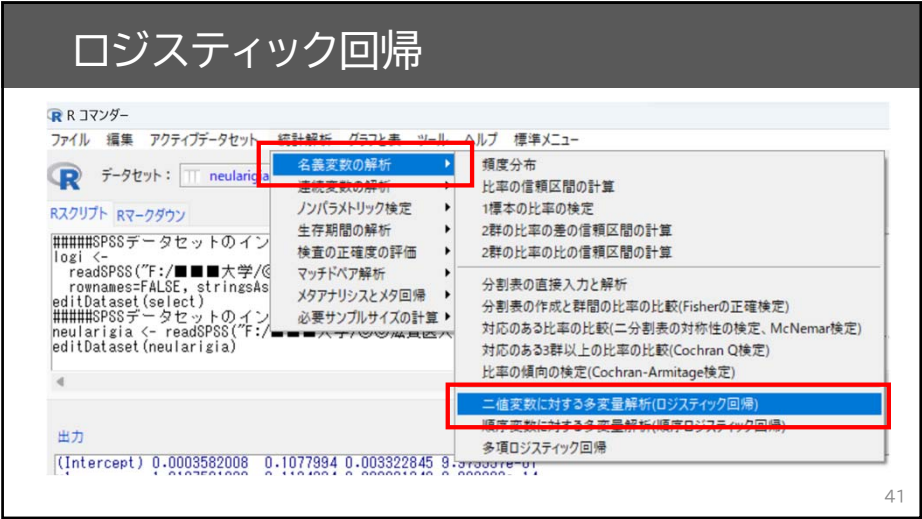
従属変数: 痛みの消失率(痛みの有無:有、無) --- 2値

独立変数: 治療法(Treatment: P(プラセボ)、A、B)

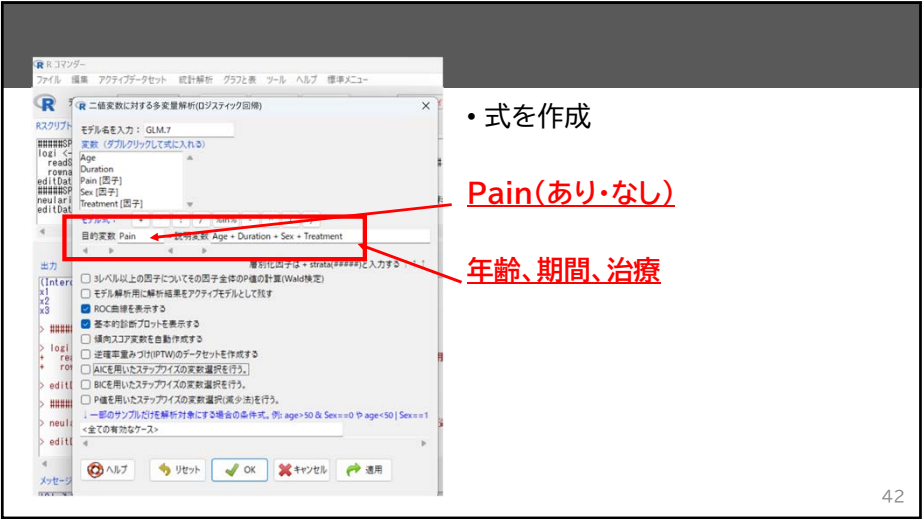
共変量: Sex(性別:F、M)
年齢(治療開始時の年齢:連続変数)
Duration(罹病期間:連続量(月))

40

ロジスティック回帰



41



• 式を作成

Pain(あり・なし)

年齢、期間、治療

42

Logistic regression

ロジスティック回帰では、従属変数はオッズの自然対数であるロジットである、

$$\log(odds) = \text{logit}(P) = \ln\left(\frac{P}{1-P}\right)$$

ロジットとはオッズの対数であり、オッズはPの関数である。ロジスティック回帰では、次のようになる。

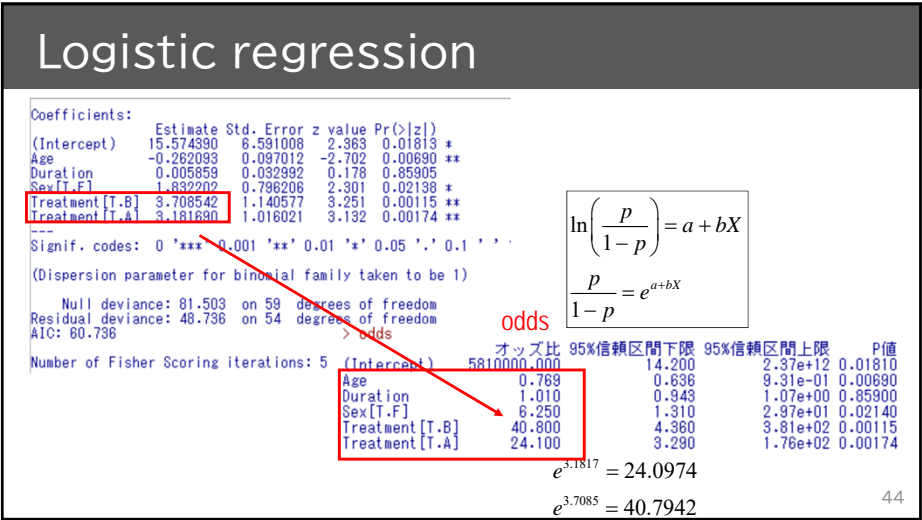
$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = a + bX$$

$$\frac{P}{1-P} = e^{a+bX}$$

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

43

Logistic regression



$$\ln\left(\frac{P}{1-P}\right) = a + bX$$

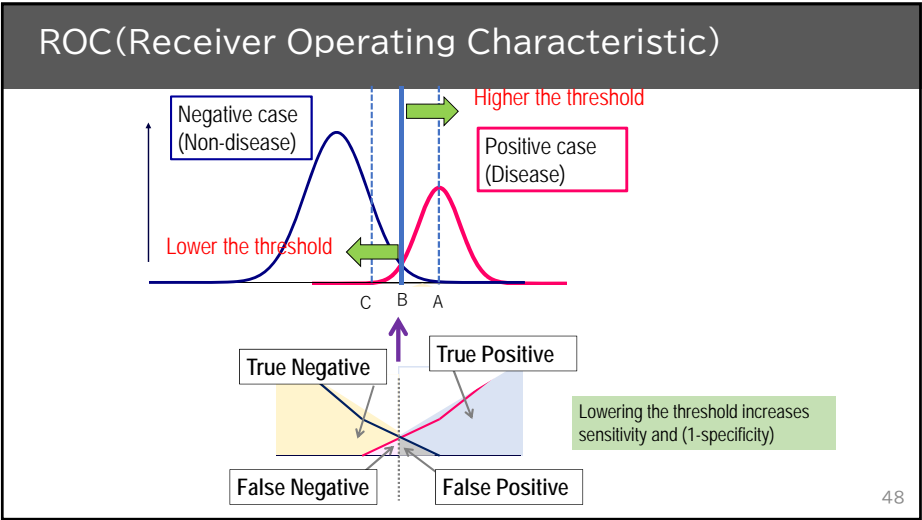
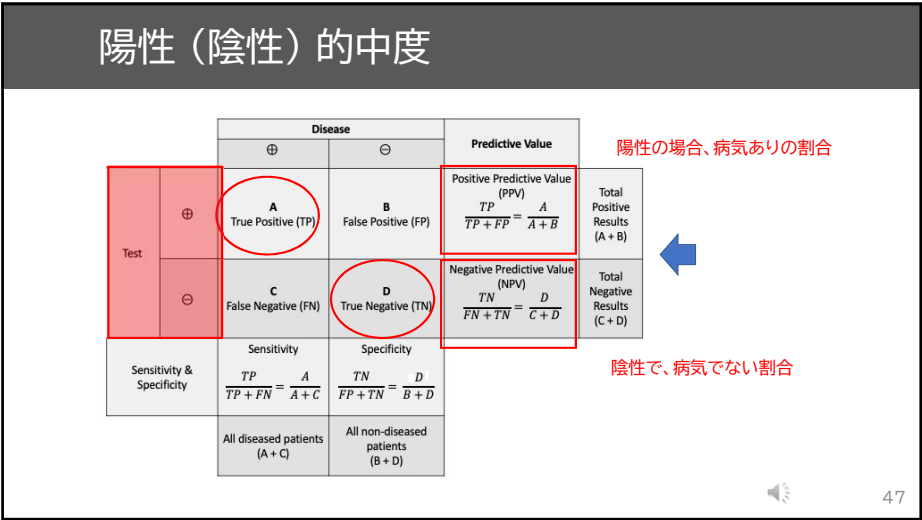
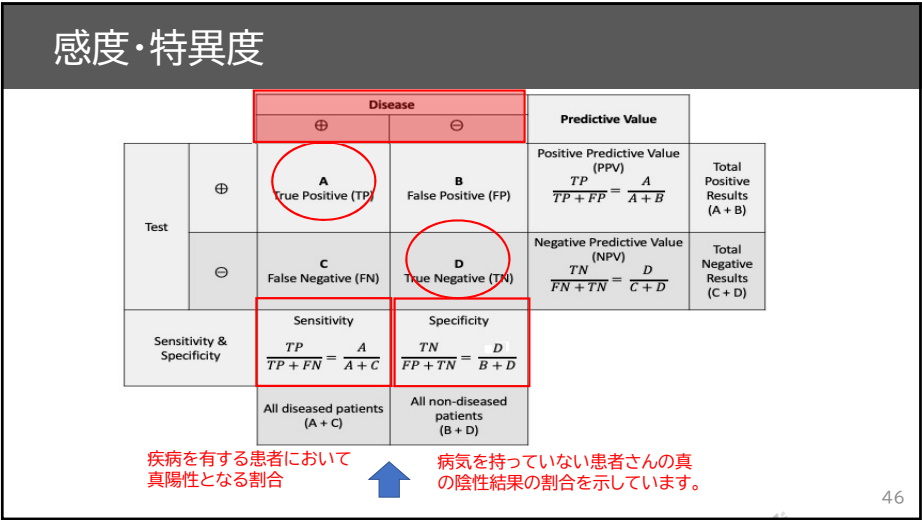
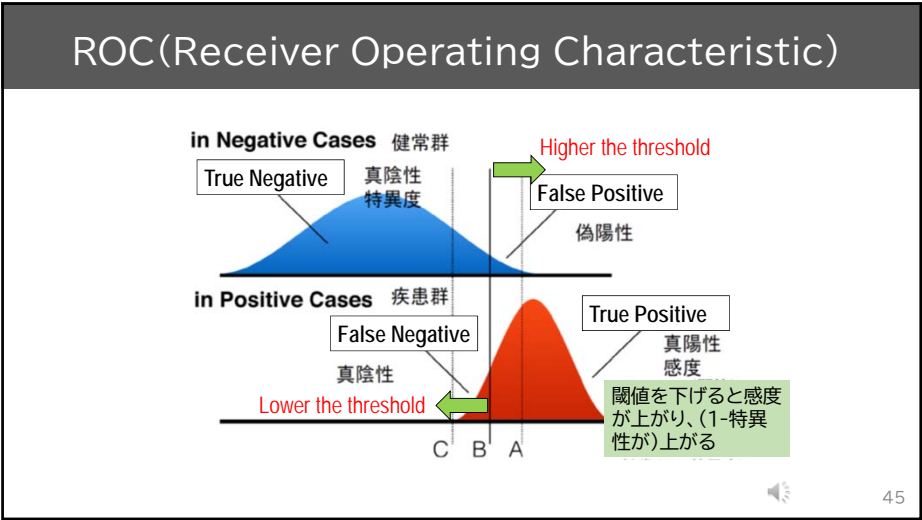
$$\frac{P}{1-P} = e^{a+bX}$$

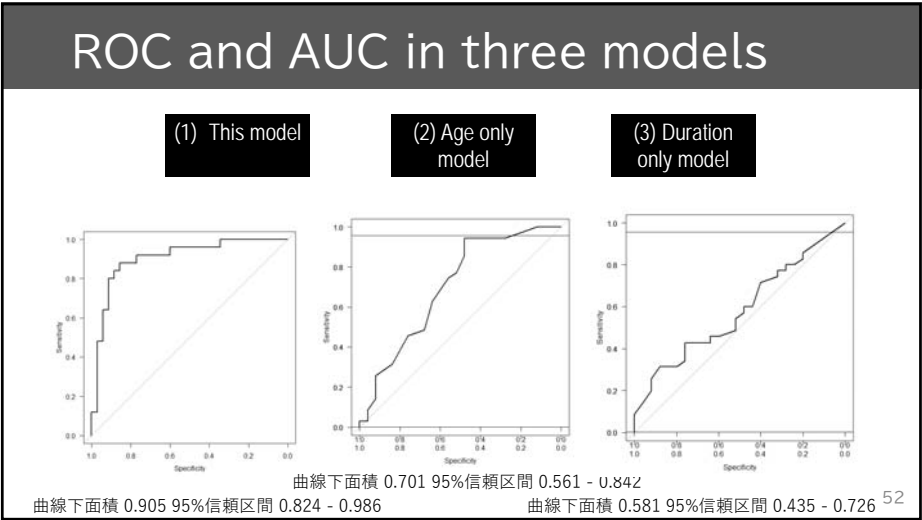
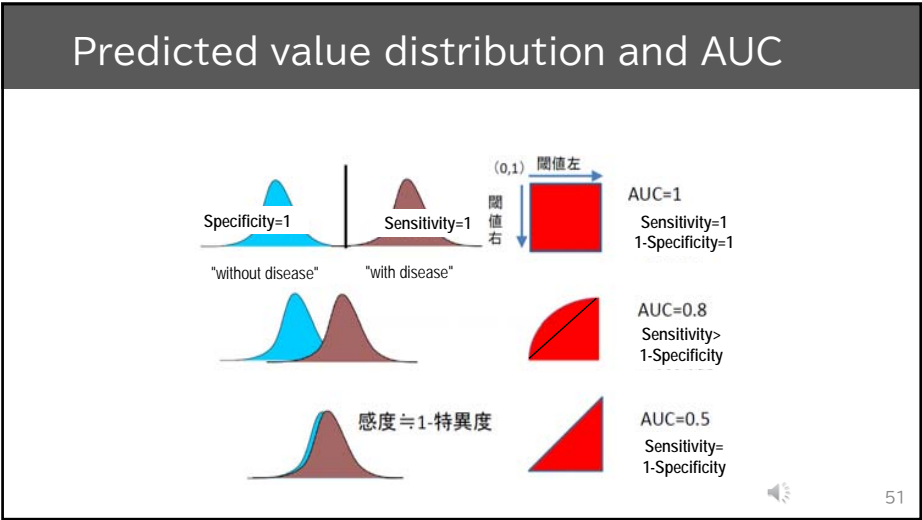
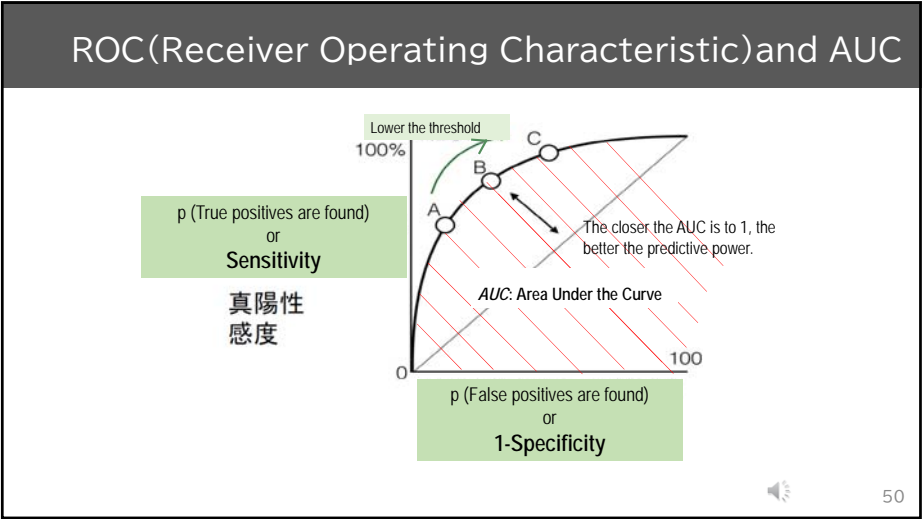
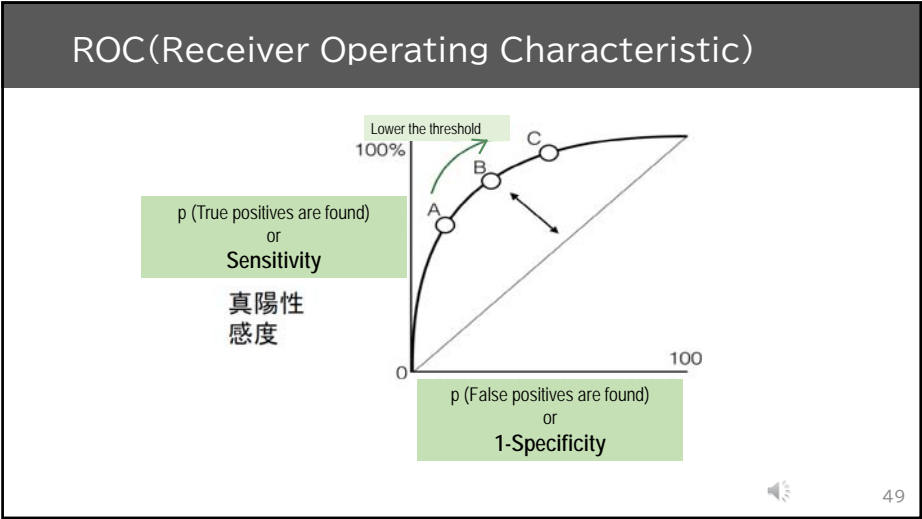
odds

$$e^{3.1817} = 24.0974$$

$$e^{3.7085} = 40.7942$$

44





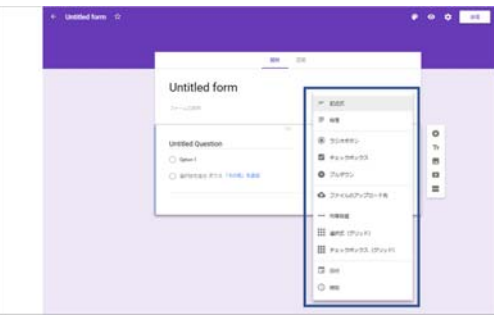
53

5

55

56


4. 設問 / 選択肢の作成




- 記述式…一文程度の短い文章を記載してもらったとき
- 段落式…段落をわけけるような長い文章を記載してもらったとき
- ラジオボタン…複数の回答から1つの回答を選択してもらったとき
- チェックボックス…複数の回答から1つ以上の回答を選択してもらったとき
- プルダウン…プルダウン形式で回答を選択してもらったとき
- 均等目盛…例えば1～5などの段階にわけて評価を知りたいとき

57

4. 設問 / 選択肢の作成



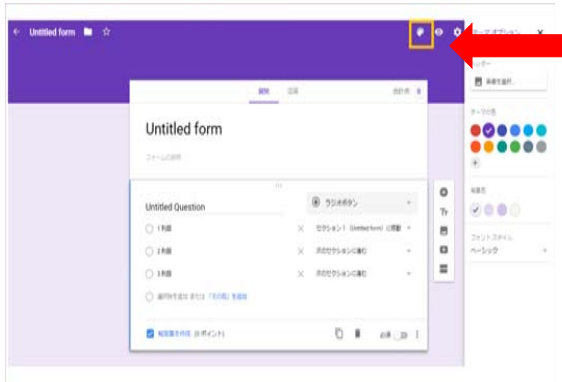
設問項目に対して説明を追加したいときや、回答完了時に次のセクションへと移動させたいときには、【必須】の横にある三点リーダーをクリックします。



自由記載の場合は、「回答の検証」で文字数などの制限を付けられる

58

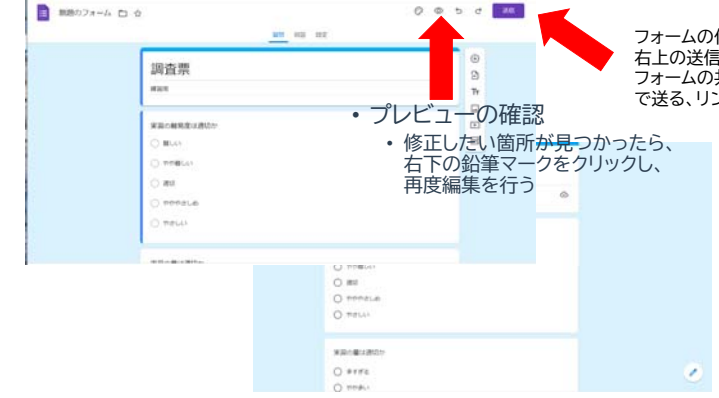
5. デザインの設定



- デザインの設定は、画面上部のパレットマークから行うことができる

59

6. フォームの確認



- プレビューの確認
- 修正したい箇所が見つかったら、右下の鉛筆マークをクリックし、再度編集を行う

フォームの作成が完了したら、右上の送信をクリックします。フォームの共有方法はメールで送る、リンクを共有する。

60

7. 回答結果の集計

9件の回答

回答 5

上部の【回答】をクリックすると、フォームへの回答結果を確認することができる

回答を受け付中

○○について教えてください

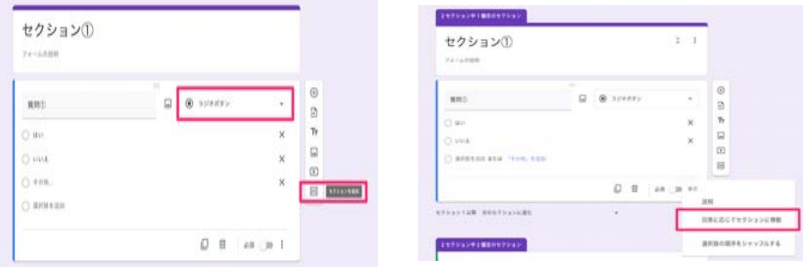
9件の回答



61

<条件分岐>

条件分岐をするには、まず「ラジオボタン」「プルダウン」のどちらかの形式で質問を作成しましょう。作成できたら、【セクションを追加】をクリック



62

<スプレッドシートへエクスポート>

スプレッドシートと連携させてグラフを調整することも可能

質問 回答

0件の回答

回答を受け付けています

回答先の選択

新しいスプレッドシートを作成 既存のフォーム (回答)

既存のスプレッドシートを選択

Googleフォームに回答がくるたびに、回答がスプレッドシートにリアルタイムで反映されていきます

63

調査項目例

人格特性: 日本語版 Ten Item Personality Inventory (TIPI-J)

- 10問7段階評価(1〜7点)
- 外向性、協調性、勤勉性、神経症傾向、開放性を評価
- 得点が高いほど各項目が強い

クロノタイプ(朝方夜型)

- 19問

【クロノタイプ判定】

【70〜86点】明らかな朝型 【59〜69点】ほぼ朝型 【42〜58点】中間型 【31〜41点】ほぼ夜型 【16〜30点】明らかな夜型

アテネ睡眠調査票

- 8問(0〜3)
- [1〜3点]…睡眠がとれています [4〜5点]…不眠症の疑いが少しあります [6点以上]…不眠症の可能性が高いです

フォームを作成し回答しあう

- 条件
 - 数値情報を調査する項目を設定(仮想でも良い)
 - 自由記載の欄を作成
 - 必ず入力してもらう



65

Google Formの調査データのデータセット化

66

睡眠調査データ

- 睡眠調査データを配布します
 - データ(1)とデータ(2)があります
 - データ(1)を原本とします。データ(2)は同じデータを入力したファイルです(注:早めにデータ回収を打ち切り、重複送信をそのままにしてあるという設定。二つはデータが一致しない可能性がある)。
- この二つのデータが一致しているか、確認してもらいます
 - いくつか入力間違いも作っています。

67

COMPARE

例題
ae.sav
ae2.sav
Sleep.q.csv
Sleep.q2.csv

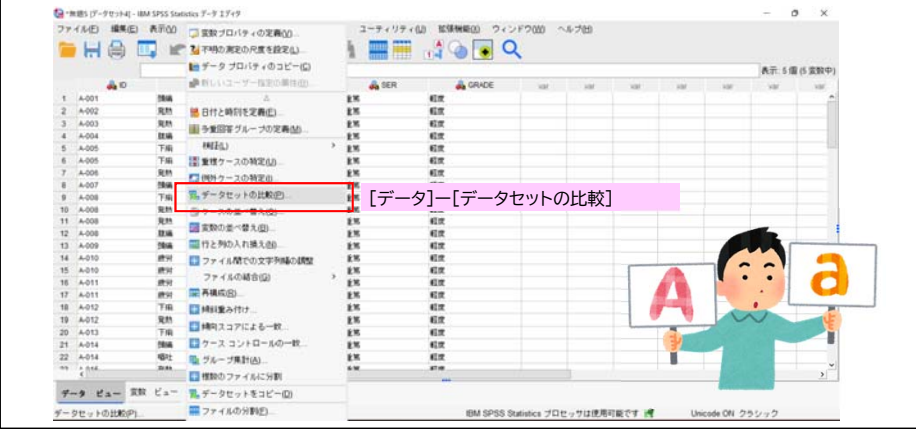
データセットを比較するには？

- データファイルを開きます。(データセットの「データエディター」ウィンドウをクリックすると、そのデータセットをアクティブなデータセットにすることができます。)
- メニューから以下を選択します。
- データ>データセットの比較
- アクティブなデータセットと比較するオープンデータセットまたはIBM SPSS Statisticsデータファイルを選択します。
- 比較する1つ以上のフィールド(変数)を選択します。

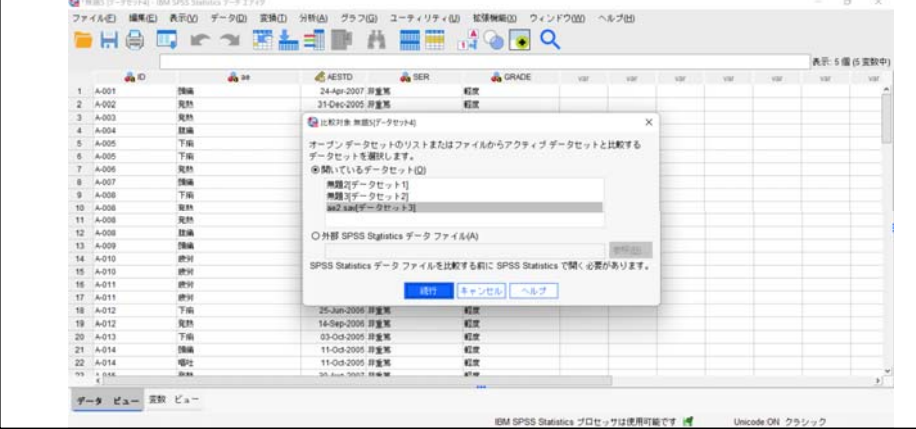


68

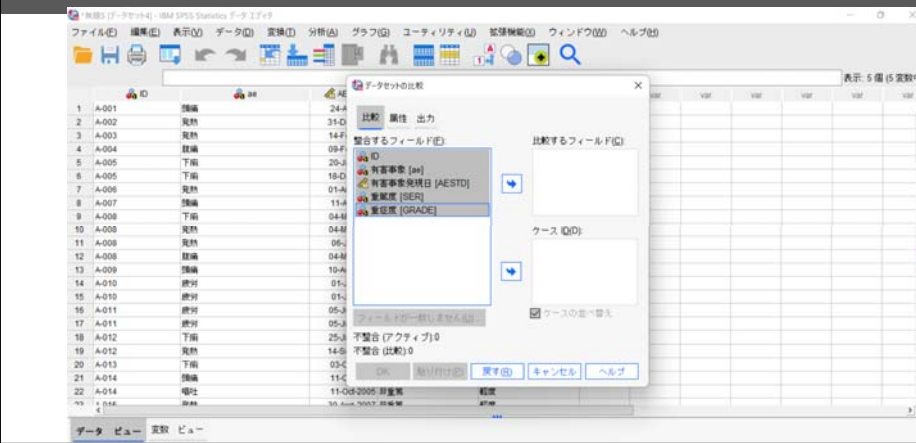
COMPARE 手順



COMPARE 手順



COMPARE 手順



COMPARE 結果

比較の概要

情報	アクティブ データセット	比較
データ ファイル	C:\Users\biosta\Local_35\Desktop\SPSSiae.sav	C:\Users\biosta\Local_35\Desktop\SPSSiae2.sav
データセット	データセット4	
フィルタ		
重み		
ファイルの分割		
ケース	86	86

ケースによる不一致

比較されるケース	度数	
不一致を含むケース	4	
パーセント	4.7%	

変数による不一致

変数 ビュー	度数	パーセント ^a
ID	0	0.0%
有害事象	4	4.7%
有害事象発現日	0	0.0%
重篤度	0	0.0%
重症度	0	0.0%

a. 比較される 86 のケースに基づく

一致した概要

結果	統計	アクティブ	比較
ケース	度数	86	86
比較されるケース	度数	86	86
	パーセント	100.0%	100.0%
比較されないケース	度数	0	0
	パーセント	0.0%	0.0%

ケースごとの比較

行	アクティブ	比較	ID	有害事象	有害事象発現日	重篤度	重症度
5	5		(1) 下痢 (2) 腹痛				
6	6		(1) 下痢 (2) 腹痛				
11	11		(1) 発熱 (2) 腹痛				
14	14		(1) 嘔吐 (2) 疲労感				

(1) はアクティブ データセットで (2) は比較データセットです

