#------------------------------------------------------------------

# 昭和大学　第6回実践臨床統計学セミナー　Rハンズオン

# ロジスティック回帰分析

# 2021年11月22日

#------------------------------------------------------------------

# Rのプログラム作成には、高機能なテキストエディタを使用することをお勧めします。

# 初心者の方向けのソフトとしては、Notepad++をお勧めします（フリーソフトです）。以下のURLからダウンロードすることができます。

#　https://github.com/notepad-plus-plus/notepad-plus-plus/releases/download/v7.9.5/npp.7.9.5.Installer.exe

# 拡張子を「.r」として保存したプログラムのファイルを開くと、基本関数やコメント箇所などが色分けされて表示され、効率的なプログラミングを行うことができます。

# Macユーザーの方は、Rのパッケージのインストールができないことがあるようです。

# RStudioを使えば、特に問題なく、インストールができることが多いようですので、もし支障があるようでしたら、そちらをお試しください（以下のURLからインストールパッケージをDLできます）。

# https://download1.rstudio.org/desktop/macos/RStudio-1.4.1717.dmg

#------------------------------------------------------------------

# 1. データセットの読み込み

#------------------------------------------------------------------

setwd("C:\\R2021")

# setwdで、カレントのディレクトリを指定します。

# 他のソフトウェアと違う点は、フォルダのパスの指定において、「\」を二重にしなくてはいけないという点です。

# 通常は、「C:\」で済みますが、「C:\\」という風に「\」を「\\」と表記しなくてはいけません。

# メニューバーの中から、「ファイル」→「ディレクトリの変更」としても、同様の操作を行うことができます。

breast <- read.csv(file="breastcancer.csv")

# CSV形式のデータセットを読み込むためには、read.csvという関数を使います。

# fileという引数で、ファイル名を指定します。カレントのディレクトリにファイル名が一致するCSVファイルがあれば、読み込まれます。

head(breast)

# データセットの中身を表示します。

#------------------------------------------------------------------

# 2. ロジスティック回帰分析：単変量解析

#------------------------------------------------------------------

install.packages("epiDisplay")

library("epiDisplay")

gm1 <- glm(y ~ x, data=breast, family=binomial(link="logit"))

summary(gm1)

# glm関数は、一般化線形モデルという枠組みのモデルでの解析を行うための汎用的な関数です。ロジスティック回帰モデルは、一般化線形モデルの中のひとつのモデルです。

# lm関数と同様、「y ~ x」 というような式で、回帰式を指定することができます。

# dataという引数は、解析対象となるデータセットを指定する引数です。

# 「family=binomial(link="logit")」というオプションは、ロジスティック回帰モデルを行う上では必須の決まり文句のようなもの、としてつけてください。

# 一般化線形モデルの中でも、２項分布（binomial）によるモデルで、ロジットリンク（link="logit"）を使ったモデルが、ロジスティック回帰モデルなので、このような引数が入ります。

# ちなみに、familyの引数を変えることで、ポアソン回帰モデルなどの別な一般化線形モデルでの解析を行うことができます。

logistic.display(gm1)

# glm関数の出力では、オッズ比の推定値や信頼区間が出力されません。

# logistic.display関数で、これらの出力を得ることができます。論文化の際には、こちらの出力を使いましょう。

#------------------------------------------------------------------

# 3. ロジスティック回帰分析：多変量解析

#------------------------------------------------------------------

gm2 <- glm(y ~ x + z, data=breast, family=binomial(link="logit"))

summary(gm2)

# glm関数は、一般化線形モデルという枠組みのモデルでの解析を行うための汎用的な関数です。ロジスティック回帰モデルは、一般化線形モデルの中のひとつのモデルです。

# lm関数と同様、「y ~ x」 というような式で、回帰式を指定することができます。

# dataという引数は、解析対象となるデータセットを指定する引数です。

# 「family=binomial(link="logit")」というオプションは、ロジスティック回帰モデルを行う上では必須の決まり文句のようなもの、としてつけてください。

logistic.display(gm2)

# glm関数の出力では、オッズ比の推定値や信頼区間が出力されません。

# logistic.display関数で、これらの出力を得ることができます。論文化の際には、こちらの出力を使いましょう。

#------------------------------------------------------------------

# 4. データセットの読み込み

#------------------------------------------------------------------

setwd("C:\\R2021")

# setwdで、カレントのディレクトリを指定します。

# 他のソフトウェアと違う点は、フォルダのパスの指定において、「\」を二重にしなくてはいけないという点です。

# 通常は、「C:\」で済みますが、「C:\\」という風に「\」を「\\」と表記しなくてはいけません。

# メニューバーの中から、「ファイル」→「ディレクトリの変更」としても、同様の操作を行うことができます。

wcgs <- read.csv(file="wcgs.csv")

# CSV形式のデータセットを読み込むためには、read.csvという関数を使います。

# fileという引数で、ファイル名を指定します。カレントのディレクトリにファイル名が一致するCSVファイルがあれば、読み込まれます。

head(wcgs)

# データセットの中身を表示します。

#------------------------------------------------------------------

# 5. ロジスティック回帰分析：多変量解析

#------------------------------------------------------------------

gm3 <-glm(chd69 ~ dibpat0 + age0 + sbp0 + chol0 + ncigs0 + weight0, data=wcgs, family=binomial(link="logit"))

summary(gm3)

# glm関数は、一般化線形モデルという枠組みのモデルでの解析を行うための汎用的な関数です。ロジスティック回帰モデルは、一般化線形モデルの中のひとつのモデルです。

# lm関数と同様、「y ~ x」 というような式で、回帰式を指定することができます。

# dataという引数は、解析対象となるデータセットを指定する引数です。

# 「family=binomial(link="logit")」というオプションは、ロジスティック回帰モデルを行う上では必須の決まり文句のようなもの、としてつけてください。

logistic.display(gm3)

# glm関数の出力では、オッズ比の推定値や信頼区間が出力されません。

# logistic.display関数で、これらの出力を得ることができます。論文化の際には、こちらの出力を使いましょう。