Generalized Linear Models

Flint

12/202018

|  |
| --- |
| Abstract |
| Key Words: Treatment, Onychomycosis, terbinafine, itraconazole, Logistic regression \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Group leader: Flint Chenjera |

Deendent variable:

Independent variables:

House keeping

rm(list=ls())  
chooseCRANmirror(graphics=FALSE, ind=1)

## Warning in download.file(url, destfile = f, quiet = TRUE): InternetOpenUrl  
## failed: 'The server name or address could not be resolved'

## Warning: failed to download mirrors file (cannot open URL 'https://cran.r-  
## project.org/CRAN\_mirrors.csv'); using local file 'C:/PROGRA~1/R/R-35~1.1/  
## doc/CRAN\_mirrors.csv'

knitr::opts\_chunk$set(echo = TRUE)  
library(nlme)  
library(MASS)  
library(tidyverse)

## -- Attaching packages --------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::collapse() masks nlme::collapse()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::select() masks MASS::select()

library(ggplot2)  
library(nlme)  
library(knitr)  
#library(kableExtra)  
#library(expss)  
library(pander)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(dplyr)  
#library(gmodels)  
#library(xtable)  
#library(survMisc)

# Load the data

setwd("C:/Users/user/Desktop/flint/MSc Statistics/GLM 503/Project")  
thedat <- read.table("toenail.dat",row.names=NULL)  
   
dim(thedat)

## [1] 1200 6

#The original study consisted of 1200 observations and 6 variables  
head(thedat)

## V1 V2 V3 V4 V5 V6  
## 1 1 0 1 0 7.3 1  
## 2 1 0 1 1 6.2 1  
## 3 1 0 1 2 8.0 1  
## 4 1 0 1 3 7.9 1  
## 5 1 0 1 6 7.6 1  
## 6 1 0 1 12 8.1 1

#The original variables are not properly named and the values for categorical variables are not labeled  
summary(thedat)

## V1 V2 V3 V4   
## Min. : 1.00 Min. :0.000 Min. :0.000 Min. : 0.0   
## 1st Qu.: 50.75 1st Qu.:0.000 1st Qu.:0.000 1st Qu.: 1.0   
## Median :100.50 Median :1.000 Median :1.000 Median : 2.5   
## Mean :100.50 Mean :0.665 Mean :0.505 Mean : 4.0   
## 3rd Qu.:150.25 3rd Qu.:1.000 3rd Qu.:1.000 3rd Qu.: 6.0   
## Max. :200.00 Max. :1.000 Max. :1.000 Max. :12.0   
## V5 V6   
## Min. : 0.500 Min. :1.0   
## 1st Qu.: 5.000 1st Qu.:1.0   
## Median : 6.500 Median :1.5   
## Mean : 6.776 Mean :1.5   
## 3rd Qu.: 8.025 3rd Qu.:2.0   
## Max. :19.600 Max. :2.0

#from the looks of things, V2, V3 and V6 are binary variables.

# Data manupulation

#renaming variables  
colnames(thedat) <- c("id","healthclub","gender","month","length","trt")  
dat<-thedat  
head(thedat)

## id healthclub gender month length trt  
## 1 1 0 1 0 7.3 1  
## 2 1 0 1 1 6.2 1  
## 3 1 0 1 2 8.0 1  
## 4 1 0 1 3 7.9 1  
## 5 1 0 1 6 7.6 1  
## 6 1 0 1 12 8.1 1

length(unique(thedat$id))

## [1] 200

unique(thedat$healthclub) # 0 1

## [1] 0 1

unique(thedat$gender) # 1 0

## [1] 1 0

unique(thedat$month) # 0 1 2 3 6 12

## [1] 0 1 2 3 6 12

unique(thedat$trt) # 1 2

## [1] 1 2

#adding labels to variable values  
thedat<-thedat %>%  
 mutate(Treatment=ifelse(trt==1,"terbinane","itraconazole")) %>%  
 mutate(Health\_club=ifelse(healthclub==1,"more than once a week","once a week or less")) %>%  
 mutate(Gender=ifelse(gender==1,"Male","Females")) %>%  
 mutate(Month=ifelse(month==0, "Base month",   
 ifelse(month==1,"1st month",  
 ifelse(month==2,"2nd month",   
 ifelse(month==3,"3rd month",  
 ifelse(month==6,"6th month", "12th month"))))))  
  
dat1<-   
 filter(thedat, Month=="Base month"| Month=="12th month")  
unique(dat1$Month)

## [1] "Base month" "12th month"

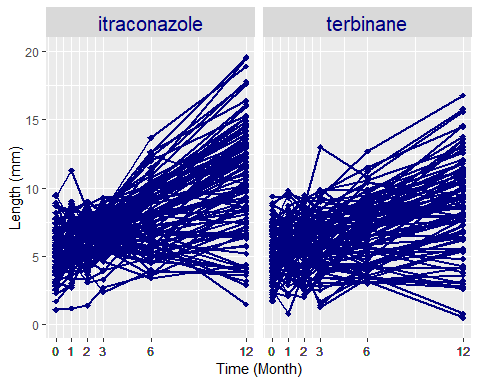
data\_wide <- reshape(dat1, idvar = "id", timevar = "Month", direction = "wide")  
summary(data\_wide)

## id healthclub.Base month gender.Base month month.Base month  
## Min. : 1.00 Min. :0.000 Min. :0.000 Min. :0   
## 1st Qu.: 50.75 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0   
## Median :100.50 Median :1.000 Median :1.000 Median :0   
## Mean :100.50 Mean :0.665 Mean :0.505 Mean :0   
## 3rd Qu.:150.25 3rd Qu.:1.000 3rd Qu.:1.000 3rd Qu.:0   
## Max. :200.00 Max. :1.000 Max. :1.000 Max. :0   
## length.Base month trt.Base month Treatment.Base month  
## Min. :1.100 Min. :1.0 Length:200   
## 1st Qu.:4.075 1st Qu.:1.0 Class :character   
## Median :5.500 Median :1.5 Mode :character   
## Mean :5.386 Mean :1.5   
## 3rd Qu.:6.800 3rd Qu.:2.0   
## Max. :9.500 Max. :2.0   
## Health\_club.Base month Gender.Base month healthclub.12th month  
## Length:200 Length:200 Min. :0.000   
## Class :character Class :character 1st Qu.:0.000   
## Mode :character Mode :character Median :1.000   
## Mean :0.665   
## 3rd Qu.:1.000   
## Max. :1.000   
## gender.12th month month.12th month length.12th month trt.12th month  
## Min. :0.000 Min. :12 Min. : 0.500 Min. :1.0   
## 1st Qu.:0.000 1st Qu.:12 1st Qu.: 6.900 1st Qu.:1.0   
## Median :1.000 Median :12 Median : 9.500 Median :1.5   
## Mean :0.505 Mean :12 Mean : 9.569 Mean :1.5   
## 3rd Qu.:1.000 3rd Qu.:12 3rd Qu.:12.300 3rd Qu.:2.0   
## Max. :1.000 Max. :12 Max. :19.600 Max. :2.0   
## Treatment.12th month Health\_club.12th month Gender.12th month   
## Length:200 Length:200 Length:200   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

data\_wide<-data\_wide %>%  
 mutate(differance=`length.12th month`- `length.Base month`) %>%  
 mutate(success=ifelse(differance>=5,"Success","Failure"))  
  
count1<-data\_wide %>%  
 count(success) %>%  
mutate("%" = n / sum(n)\*100)  
  
#count1 %>% kable("html" , caption = 'Descriptives') %>% kable\_styling()

Is there evidence to suggest that these two oral medications differ in their effects on toenail dermatophyte onychomycosis, as reflected by the proportions of successfully treated patients at 12 months?

#Graphical visualization  
p1<-ggplot(data=thedat, aes(x=month,y=length,group=id),alpha=0.90)+  
geom\_point(size=2,shape=18,color='navy blue') +  
geom\_line(size=1,color='navy blue') +  
facet\_wrap(~Treatment, nrow=1) +  
theme(strip.text.x = element\_text(size =15, colour = "navy blue")) +  
scale\_x\_continuous("Time (Month)",limits=c(0,12),breaks=thedat$month) +  
scale\_y\_continuous("Length (mm)",limits=c(0,20))  
p1



#Analytical   
#table7<- data\_wide %>% table(`Treatment.12th month`, success)   
  
table1<-   
 table(data\_wide$`Treatment.12th month`, data\_wide$success)   
  
df <-  
 data.frame(table1)  
  
  
  
#CrossTable(data\_wide$`Treatment.12th month`, data\_wide$success)  
  
mod1=glm(factor(success)~1,family = binomial(link=logit),data=data\_wide) ###---Intercept only model  
  
mod2=glm(factor(success)~factor(`Treatment.12th month`),family = binomial(link=logit),data=data\_wide) ###---Model with age as only covariate  
  
summary(mod2)

##   
## Call:  
## glm(formula = factor(success) ~ factor(`Treatment.12th month`),   
## family = binomial(link = logit), data = data\_wide)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.212 -0.776 -0.776 1.144 1.641   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) 0.08004 0.20016 0.400  
## factor(`Treatment.12th month`)terbinane -1.12601 0.30338 -3.712  
## Pr(>|z|)   
## (Intercept) 0.689235   
## factor(`Treatment.12th month`)terbinane 0.000206 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 267.50 on 199 degrees of freedom  
## Residual deviance: 253.08 on 198 degrees of freedom  
## AIC: 257.08  
##   
## Number of Fisher Scoring iterations: 4

In addition to treatment, is there an effect of the frequency with which the subject visited a gym, gender, or the length of the unafficted part of nail at baseline on the relative change of the length of the unafficted part of the nail.

mod3=glm(factor(success)~factor(`Treatment.12th month`)+ factor(`healthclub.12th month`)+  
 factor(`gender.12th month`)+ `length.Base month` ,family = binomial(link=logit),data=data\_wide)  
summary(mod3)

##   
## Call:  
## glm(formula = factor(success) ~ factor(`Treatment.12th month`) +   
## factor(`healthclub.12th month`) + factor(`gender.12th month`) +   
## `length.Base month`, family = binomial(link = logit), data = data\_wide)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7015 -0.8873 -0.5378 0.9289 2.5259   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) 3.5705 0.7548 4.730  
## factor(`Treatment.12th month`)terbinane -1.4370 0.3422 -4.199  
## factor(`healthclub.12th month`)1 -0.7670 0.3696 -2.075  
## factor(`gender.12th month`)1 -0.2114 0.3307 -0.639  
## `length.Base month` -0.5189 0.1080 -4.806  
## Pr(>|z|)   
## (Intercept) 2.24e-06 \*\*\*  
## factor(`Treatment.12th month`)terbinane 2.68e-05 \*\*\*  
## factor(`healthclub.12th month`)1 0.038 \*   
## factor(`gender.12th month`)1 0.523   
## `length.Base month` 1.54e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 267.50 on 199 degrees of freedom  
## Residual deviance: 223.14 on 195 degrees of freedom  
## AIC: 233.14  
##   
## Number of Fisher Scoring iterations: 4