

Asunto: Re: TP Reg Log

De: Franco Assanelli <assafranco@gmail.com>

Fecha: 18/11/2023, 17:04

A: nicolas@kuzminski.com.ar

<https://github.com/Hurled3574/TP-Reg-Log.git>

al final, logré subir todo a github... nada de archivos particionados jaja

El 18/11/2023 a las 16:52, Franco Assanelli escribió:

Te adjunto, Nico, lo que pude hacer. Hay un script que se llama "Criterio de selección del modelo" (creo) en el que me puse medio caótico porque estaba buscando la función que me sirva para hacer varios AICs de un saque y no estar haciendo uno por uno.

Recién me desperté de una pseudo-siesta (una siesta en la que te tirás y no te podés dormir) que necesité porque volví cansado de la clase... y de la semana, se ve... no sé. así que lamento la tardanza

Te paso lo que hice en tres partes, para no pasarme del límite permitido por correo

El 17/11/2023 a las 16:30, Franco Assanelli escribió:

Descargué también el paper del segundo dataset sobre el sueño. Nos podría orientar...

El 17/11/2023 a las 16:21, Franco Assanelli escribió:

Nico:

descargué el paper del dataset. Quizás nos oriente en el análisis de los datos y en la redacción del informe.

Temas a resolver, todavía:

1. Las variables a mantener a priori
2. Las variables a descartar a priori
3. Darnos cuenta si esas variables que mantenemos las podemos comunicar en el informe
4. Evtl. comparar con las variables del paper-
5. Considerar que el paper usa los términos sensibility y accuracy. Esto nos lleva a una prueba ROC, si no me equivoco... o alguna de las últimas que vimos

El 16/11/2023 a las 17:39, Franco Assanelli escribió:

Hasta acá llego por hoy, Nico

Saludos
Franco

El 16/11/2023 a las 14:04, Franco Assanelli escribió:

En un Newsletter que sigo de la revista Nature [Terrifying tarantula fangs and more — October's best science images \(nature.com\)](#), encontré una foto de células cancerosas de mama, sacada por una fotógrafa en una competencia de la empresa Nikon. Adjunto foto que puede ser interesante

Saludos
Fran

El 16/11/2023 a las 12:31, Franco Assanelli escribió:

Y te adjunto el proyecto completo

El 16/11/2023 a las 12:26, Franco Assanelli escribió:

Te paso lo que fui redactando hasta ahora en el script.

```
rm(list = ls())
library(corrplot)
library(olsrr)
library(wooldridge)
library(lmtest)
datos <- read.csv("data.csv")
View(datos)
```

```
#####
```

```
# Eliminación a priori de variables.
##Nos quedamos con las variables que contienen las medias por sencillez para
comunicar los resultados, y porque expresan mejor la situación de cada caso.
```

```
datos2 <- datos[,c(2:12)] #Recorte de datos
colnames(datos2) #Lista de variables
```

```
##Yo creo que podemos omitir otras tantas variables o mediciones porque
desconocemos de qué tratan y porque son muy específicas del campo de estudio,
quizás. POr ejemplo, compactness.... Pero hay variables, como perimeter que SON
```

CLAVE, porque hablan específicamente del perímetro del tumor.

#Análisis exploratorio de datos

##Valores duplicados

```
datos[duplicated(datos$id), "id"] #NO hay
```

##Boxplots

```
# boxplot(datos$radius_mean, horizontal = T)
```

```
# boxplot(datos$texture_mean, horizontal = T)
```

```
# boxplot(datos$perimeter_mean, horizontal = T)
```

```
# boxplot(datos$area_mean, horizontal = T)
```

```
# boxplot(datos$smoothness_mean, horizontal = T)
```

```
# boxplot(datos$compactness_mean, horizontal = T)
```

##Variables

```
colnames(datos)
```

##Eliminar variable X?

##Cambio valores variable rta

```
table(datos$diagnosis)
```

```
datos$diagnosis <- factor(datos$diagnosis,
  levels = c("B", "M"),
  labels = c(0, 1))
```

```
datos$diagnosis
```

Según los scripts de las clases, hay que aplicar los WOE, agrupar variables, u otras cosas, antes de realizar el modelo de regresión

#Regresión

```
diagnostico <-
```

```
  glm(
```

```
    diagnosis ~ radius_mean + texture_mean + perimeter_mean + area_mean +
    smoothness_mean + compactness_mean + concavity_mean +
```

```
    concave.points_mean + symmetry_mean + fractal_dimension_mean,
```

```
    family = binomial(link = "logit"),
```

```
    datos
```

```
  )
```

Pruebas Ómnibus

```
ols_regress(diagnostico)
```

lrtest(diagnostico)

#####

#####

data("sleep75")

Saludos

Fran

El 16/11/2023 a las 11:34, Franco Assanelli escribió:

Estaba pensando, además, algo que te dije por wasap.

tenemos 3 bloques de variables:

1. promedios de observaciones
2. errores estándar de observaciones
3. las, ahora no recuerdo bien, peores medidas ("worst")

Entonces, a partir de las consignas (que adjunto al final del presente mensaje), tenemos que eliminar variables que nos resulten inviables para analizar esto.

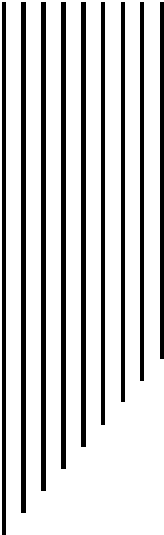
Yo pienso que los errores estándar y las medidas "worst" las podríamos descartar porque analizamos casos y no veo la utilidad de analizar el error estándar de cada caso, ni la medida "worst". Además, ¿cómo las comunicarías...?

Decime qué pensás como para ya hacer o terminar la primera consigna.

El 16/11/2023 a las 11:26, Franco Assanelli escribió:

Nico, cómo andás?

Los profes me contestaron que la prueba ómnibus es la prueba de razón de verosimilitud pero hecha según añadidura de variables una por una, en bloques y entre constante y modelo completo. Ahora te adjunto:

- 
1. fragmento del Kutner que explica sobre la prueba de razón de veros.,
 2. un ppt que está en el aula v.
 3. el mail del profe

Saludos
Fran