Affichage d’output de régression avec Jtools

Marc Thévenin

2023-06-21

La fonction summ du package jtools (Jacob Long) permet d’obtenir des outputs de regression en format console de très bonne qualité. Il s’agit d’une très bonne alternative aux outputs par défaut. Ce package propose également des fonctionnalités d’exportation des outputs dans d’autres formats (html, docx, pdf …) et des visualisations sous forme de graphique. Ces fonctionnalité ne seront pas traitées ici, l’accent étant mis sur la qualité d’un output console avec une durée d’exécution minimale. L’utilisation des fonctionnalités d’exportation, comme le très populaire package **gtsummaty** devrait être appliqué, selon nous, pour des raisons de durée d’exécution à un résultat final.

* [Documentation sur le package **jtools** (Jacob Long)](https://jtools.jacob-long.com/index.html)
* A ce jour, la maintenance du package est assurée [version 2.2.1 à juin 2023]
* Les nombre de modèles pris en charge par le package est malheureusement assez réduit. On notera cependant la prise en charge de la fonction svyglm du package survey.
* Les fonctionnalités d’exportation ne seront pas traitées ici.

# **Installation**

install.packages("jtools")

ou

install.packages("devtools")  
devtools::install\_github("jacob-long/jtools")

# **Syntaxe de la fonction summ()**

La syntaxe est particulièrement simple, elle consiste juste à appliquée à la fonction summ() l’objet généré par la régression. Quelques options comme digits, confint, exp permettent d’améliorer et enrichir l’output.

fit = lm(y ~ x , data=df)  
summ(fit)

# **Exemples** [[1]](#footnote-23)

## Avec la fonction lm()

[Liste des options](https://jtools.jacob-long.com/reference/summ.lm.html)

library(jtools)  
library(readr)  
  
df = read.csv("https://raw.githubusercontent.com/mthevenin/intro\_logit/main/hypertension2.csv")  
  
fit = lm(bpsystol ~ age + I(sex) + I(black) + I(region), data=df)  
  
summ(fit, digits=4)

Ligne 8

digits=4 => Les résultats sont reportés avec 4 décimales

MODEL INFO:  
Observations: 10351  
Dependent Variable: bpsystol  
Type: OLS linear regression   
  
MODEL FIT:  
F(6,10344) = 552.2494, p = 0.0000  
R² = 0.2426  
Adj. R² = 0.2422   
  
Standard errors: OLS  
---------------------------------------------------------------  
 Est. S.E. t val. p  
----------------------- ---------- -------- ---------- --------  
(Intercept) 102.2001 0.8979 113.8168 0.0000  
age 0.6563 0.0116 56.5104 0.0000  
I(sex)Male 4.0350 0.3999 10.0911 0.0000  
I(black)Not Black -4.6494 0.6640 -7.0019 0.0000  
I(region)NE 0.2570 0.5892 0.4362 0.6627  
I(region)S -0.7920 0.5450 -1.4532 0.1462  
I(region)W -0.5221 0.5543 -0.9420 0.3462  
---------------------------------------------------------------

## Avec la fonction glm() [lien logit]

[Liste des options](https://jtools.jacob-long.com/reference/summ.glm.html)

fit = glm(highbp ~ age + I(sex) + I(black) + I(region), family=binomial, data=df)   
  
summ(fit, digits=4, confint=TRUE, exp=TRUE)

Ligne 3

On ajoute des intervalles de confiance (confint=TRUE) et le report des estimateurs sous forme d’Odds Ratio (exp=TRUE)

MODEL INFO:  
Observations: 10351  
Dependent Variable: highbp  
Type: Generalized linear model  
 Family: binomial   
 Link function: logit   
  
MODEL FIT:  
χ²(6) = 1623.1601, p = 0.0000  
Pseudo-R² (Cragg-Uhler) = 0.1951  
Pseudo-R² (McFadden) = 0.1151  
AIC = 12492.3709, BIC = 12543.0848   
  
Standard errors: MLE  
-------------------------------------------------------------------------  
 exp(Est.) 2.5% 97.5% z val. p  
----------------------- ----------- -------- -------- ---------- --------  
(Intercept) 0.0853 0.0701 0.1038 -24.5720 0.0000  
age 1.0496 1.0469 1.0524 36.1832 0.0000  
I(sex)Male 1.5483 1.4223 1.6856 10.0890 0.0000  
I(black)Not Black 0.5860 0.5092 0.6742 -7.4663 0.0000  
I(region)NE 1.1655 1.0287 1.3203 2.4051 0.0162  
I(region)S 1.0024 0.8930 1.1253 0.0414 0.9669  
I(region)W 1.0966 0.9746 1.2338 1.5322 0.1255  
-------------------------------------------------------------------------

## Avec la fonction svyglm() [lien logit]

[Liste des options](https://jtools.jacob-long.com/reference/summ.svyglm.html)

library(survey)  
  
w = svydesign(id=~1, weights=~w, data=df)  
  
fit = svyglm(highbp ~ age + I(sex) + I(black) + I(region), family=binomial, design=w)  
summ(fit, digits=4, confint=TRUE, exp=TRUE)

MODEL INFO:  
Observations: 10351  
Dependent Variable: highbp  
Type: Analysis of complex survey design   
 Family: binomial   
 Link function: logit   
  
MODEL FIT:  
Pseudo-R² (Cragg-Uhler) = 0.1881  
Pseudo-R² (McFadden) = 0.1125  
AIC = 12064.1922   
  
-------------------------------------------------------------------------  
 exp(Est.) 2.5% 97.5% t val. p  
----------------------- ----------- -------- -------- ---------- --------  
(Intercept) 0.0692 0.0553 0.0865 -23.3763 0.0000  
age 1.0530 1.0497 1.0562 33.0767 0.0000  
I(sex)Male 1.8298 1.6529 2.0256 11.6459 0.0000  
I(black)Not Black 0.5849 0.4913 0.6964 -6.0259 0.0000  
I(region)NE 1.1822 1.0245 1.3642 2.2906 0.0220  
I(region)S 0.9961 0.8670 1.1445 -0.0549 0.9562  
I(region)W 1.1225 0.9737 1.2940 1.5928 0.1112  
-------------------------------------------------------------------------  
  
Estimated dispersion parameter = 0.984

1. mesure de la temps artérielle (lm) et du risque d’hypertension (glm et svyglm) [↑](#footnote-ref-23)