database vaccine cross

2022-10-25

# Libraries

#packages  
library("pacman")  
p\_load(tidyverse, here, sjlabelled, stringr, glue, janitor, haven, stargazer,   
 ltm, skimr, readxl)  
  
#remove scientific notation  
options(scipen=999)

# Input

#Load database  
response\_original = read\_rds(here("Input", "v1.0\_ResPOnsE\_COVID\_19\_W1-W4-2.rds")) %>%   
 clean\_names()

# Processing

## W3

# 10 November to 22 December 2021  
#Select and rename variables  
W3 = response\_original %>%   
 filter(info\_wave==3) %>%   
 dplyr::select(c(v2,v3,j5\_01,v1,f10,f3,f9,h1\_04,j4bis\_b,b2\_03,b2\_06,b2\_07,d5,d1,  
 e4\_bis,k3\_03,k3\_04,k3\_05,e2\_01,e2\_04,j5\_02,g7,c5,s1,s2,  
 s9,s8,c1)) %>%   
 mutate(across(v2:c1, ~replace(., .>97 , NA))) %>%   
 na.omit()  
  
#colnames  
colnames(W3) = c("vac","vac\_int","vac\_bad","vac\_ob","worry","risk","conspiracy",  
 "nat","h\_locus","comp\_dist", "comp\_mask","comp\_hand","he\_eco",  
 "judg\_gov\_covid","judg\_gov","PTV\_L","PTV\_5SM","PTV\_BOI","tr\_par",  
 "tr\_EU","tr\_sci","pray","media","sex","age","educ",  
 "reg","eco\_insec")  
  
#combine vac\_int and vac in INT\_VAC  
W3 = W3 %>%  
 mutate(hesitancy = case\_when(  
 (vac\_int < 3 | vac == 1)~ 0,  
 (vac\_int > 2 & vac == 2)~ 1,)) %>%   
 dplyr::select(-c(vac\_int, vac))   
  
#invert polarity and recode  
W3 = W3 %>%   
 mutate((across(vac\_bad, ~ 6 - .)),   
 worry = ifelse(worry<=2, 1, 0),   
 conspiracy = ifelse(conspiracy<3, 1, 0),  
 pray = ifelse(pray<=4, 0, 1),  
 media = ifelse(media>=4 & media<=7, 1, 0),   
 sex = ifelse(sex==2, 1, 0),  
 educ = ifelse(educ<=5, 1, 0),  
 (across(age, ~ 89 - .)),  
 reg = ifelse(reg<=3, 1, 0),  
 (across(comp\_dist:comp\_hand, ~ 10 - .)),  
 (across(judg\_gov\_covid:judg\_gov, ~ 10 - .)),  
 (across(tr\_par:tr\_EU, ~ 10- .)))

## Inspect true number of missing

#Load valid cases spreadshit (avaiable at: https://dataverse.unimi.it/dataset.xhtml?persistentId=doi:10.13130/RD\_UNIMI/FF0ABQ)  
valid = read\_excel(here("Input", "Valid\_cases.xlsx"))  
  
#Select V I used in the analysis to retrieve true number of N  
valid = valid %>%   
 dplyr::select(c(v2,v3,j5\_01,v1,f10,f3,f9,h1\_04,j4bis\_b,b2\_03,b2\_06,b2\_07,d5,d1,  
 e4\_bis,k3\_03,k3\_04,k3\_05,e2\_01,e2\_04,j5\_02,g7,c5,s1,s2,  
 s9,s8,c1))  
  
#Extract miniumu number of obs per question  
min(valid)

## [1] 3767

# TRUE N = 3767. Thus I had 3767 before listwise, 1535 after. It means I m working with 40.7% of   
# the original sample.

## PCA

#PCA  
#compliance  
compliance = data.frame(W3$comp\_mask, W3$comp\_hand, W3$comp\_dist)  
compliance = na.omit(compliance)  
pr\_compliance <- princomp(na.omit(compliance), cor = TRUE)  
  
W3$low\_comp = apply(W3[8:10], 1, mean, na.rm = TRUE)  
W3 = W3 %>%   
 dplyr::select(-c(comp\_hand, comp\_mask, comp\_dist))  
  
#Gov  
gov = data.frame(W3$judg\_gov\_covid, W3$judg\_gov)  
gov = na.omit(gov)  
pr\_gov <- princomp(na.omit(gov), cor = TRUE)  
  
W3$distrust\_gov = apply(W3[9:10], 1, mean, na.rm = TRUE)  
W3 = W3 %>%   
 dplyr::select(-c(judg\_gov\_covid, judg\_gov))  
  
  
#inst\_tr  
inst\_tr = data.frame(W3$tr\_par, W3$tr\_EU)  
inst\_tr = na.omit(inst\_tr)  
pr\_inst\_tr <- princomp(na.omit(inst\_tr), cor = TRUE)  
  
W3$distrust\_inst = apply(W3[12:13], 1, mean, na.rm = TRUE)  
W3 = W3 %>%   
 dplyr::select(-c(tr\_par, tr\_EU))

# Output

## descriptives (Article)

summary\_vaccine = as.data.frame(skim(W3))  
stargazer(as.data.frame(W3), type = "text", nobs = TRUE)

##   
## ================================================  
## Statistic N Mean St. Dev. Min Max   
## ------------------------------------------------  
## vac\_bad 1,540 2.148 1.244 1 5   
## vac\_ob 1,540 2.401 1.325 1 5   
## worry 1,540 0.673 0.469 0 1   
## risk 1,540 3.173 0.756 1 5   
## conspiracy 1,540 0.456 0.498 0 1   
## nat 1,540 4.142 3.187 0 10   
## h\_locus 1,540 6.453 2.586 0 10   
## he\_eco 1,540 4.911 2.642 0 10   
## PTV\_L 1,540 2.562 3.504 0 10   
## PTV\_5SM 1,540 2.692 3.274 0 10   
## PTV\_BOI 1,540 2.525 3.460 0 10   
## tr\_sci 1,540 2.252 1.163 1 5   
## pray 1,540 0.550 0.498 0 1   
## media 1,540 0.465 0.499 0 1   
## sex 1,540 0.454 0.498 0 1   
## age 1,540 41.744 14.895 1 71   
## educ 1,540 0.584 0.493 0 1   
## reg 1,540 0.433 0.496 0 1   
## eco\_insec 1,540 2.506 0.640 1 4   
## hesitancy 1,540 0.171 0.377 0 1   
## low\_comp 1,540 1.281 1.806 0.000 10.000  
## distrust\_gov 1,540 4.403 2.264 0.000 10.000  
## distrust\_inst 1,540 5.521 2.465 0.000 10.000  
## ------------------------------------------------

## Summary of PCAs (Supplement S1 section 2)

#dimensionality  
pr\_compliance #unidimensional

## Call:  
## princomp(x = na.omit(compliance), cor = TRUE)  
##   
## Standard deviations:  
## Comp.1 Comp.2 Comp.3   
## 1.5009023 0.6149379 0.6075720   
##   
## 3 variables and 1540 observations.

pr\_gov #unidimensional

## Call:  
## princomp(x = na.omit(gov), cor = TRUE)  
##   
## Standard deviations:  
## Comp.1 Comp.2   
## 1.3608960 0.3846584   
##   
## 2 variables and 1540 observations.

pr\_inst\_tr #unidimensional

## Call:  
## princomp(x = na.omit(inst\_tr), cor = TRUE)  
##   
## Standard deviations:  
## Comp.1 Comp.2   
## 1.2820125 0.5970292   
##   
## 2 variables and 1540 observations.

#alphas  
cronbach.alpha(compliance, CI=TRUE, standardized=TRUE)

##   
## Standardized Cronbach's alpha for the 'compliance' data-set  
##   
## Items: 3  
## Sample units: 1540  
## alpha: 0.834  
##   
## Bootstrap 95% CI based on 1000 samples  
## 2.5% 97.5%   
## 0.805 0.859

cronbach.alpha(gov, CI=TRUE, standardized=TRUE)

##   
## Standardized Cronbach's alpha for the 'gov' data-set  
##   
## Items: 2  
## Sample units: 1540  
## alpha: 0.92  
##   
## Bootstrap 95% CI based on 1000 samples  
## 2.5% 97.5%   
## 0.908 0.932

cronbach.alpha(inst\_tr, CI=TRUE, standardized=TRUE)

##   
## Standardized Cronbach's alpha for the 'inst\_tr' data-set  
##   
## Items: 2  
## Sample units: 1540  
## alpha: 0.783  
##   
## Bootstrap 95% CI based on 1000 samples  
## 2.5% 97.5%   
## 0.753 0.808

## save

#export data  
saveRDS(W3, here("Input", "W3.rds"))