Analyse des correspondances multiples

Axel-Cleris Gailloty

2023-02-18

library(readxl)  
mental\_health <- read\_excel("mental\_health\_survey.xlsx")

# Analyse exploratoire des données

pretty\_present\_data(mental\_health)

|  | Column | Type | Preview |
| --- | --- | --- | --- |
| I am currently employed at least part-time | I am currently employed at least part-time | numeric | c(0, 1, 1, 0, 1) |
| I identify as having a mental illness | I identify as having a mental illness | numeric | c(0, 1, 0, 0, 1) |
| Education | Education | character | c(“High School or GED”, “Some Phd”, “Completed Undergraduate”, “Some Undergraduate”, “Completed Undergraduate”) |
| I have my own computer separate from a smart phone | I have my own computer separate from a smart phone | numeric | c(0, 1, 1, 1, 1) |
| I have been hospitalized before for my mental illness | I have been hospitalized before for my mental illness | numeric | c(0, 0, 0, 0, 1) |
| How many days were you hospitalized for your mental illness | How many days were you hospitalized for your mental illness | numeric | c(0, 0, 0, NA, 35) |
| I am legally disabled | I am legally disabled | numeric | c(0, 0, 0, 0, 1) |
| I have my regular access to the internet | I have my regular access to the internet | numeric | c(1, 1, 1, 1, 1) |
| I live with my parents | I live with my parents | numeric | c(0, 0, 0, 1, 0) |
| I have a gap in my resume | I have a gap in my resume | numeric | c(1, 0, 0, 1, 1) |
| Total length of any gaps in my resume in months. | Total length of any gaps in my resume in months. | numeric | c(24, 1, 0, 11, 33) |
| Annual income (including any social welfare programs) in USD | Annual income (including any social welfare programs) in USD | numeric | c(35, 22, 100, 0, 32) |
| I am unemployed | I am unemployed | numeric | c(1, 0, 0, 1, 0) |
| I read outside of work and school | I read outside of work and school | numeric | c(1, 1, 1, 1, 1) |
| Annual income from social welfare programs | Annual income from social welfare programs | numeric | c(0, 0, 0, 0, 30) |
| I receive food stamps | I receive food stamps | numeric | c(0, 0, 0, 0, 0) |
| I am on section 8 housing | I am on section 8 housing | numeric | c(0, 0, 0, 0, 0) |
| How many times were you hospitalized for your mental illness | How many times were you hospitalized for your mental illness | numeric | c(0, 0, 0, 0, 4) |
| Lack of concentration | Lack of concentration | numeric | c(1, 1, 0, 0, 1) |
| Anxiety | Anxiety | numeric | c(1, 1, 0, 0, 1) |
| Depression | Depression | numeric | c(1, 1, 0, 0, 1) |
| Obsessive thinking | Obsessive thinking | numeric | c(1, 0, 0, 0, 1) |
| Mood swings | Mood swings | numeric | c(0, 0, 0, 0, 1) |
| Panic attacks | Panic attacks | numeric | c(1, 1, 0, 0, 1) |
| Compulsive behavior | Compulsive behavior | numeric | c(0, 0, 0, 0, 1) |
| Tiredness | Tiredness | numeric | c(0, 1, 0, 0, 1) |
| Age | Age | character | c(“30-44”, “18-29”, “30-44”, “30-44”, “30-44”) |
| Gender | Gender | character | c(“Male”, “Male”, “Male”, “Male”, “Male”) |
| Household Income | Household Income | character | c(“$25,000-$49,999”, “$50,000-$74,999”, “$150,000-$174,999”, “$25,000-$49,999”, “$25,000-$49,999”) |
| Region | Region | character | c(“Mountain”, “East South Central”, “Pacific”, “New England”, “East North Central”) |
| Device Type | Device Type | character | c(“Android Phone / Tablet”, “MacOS Desktop / Laptop”, “MacOS Desktop / Laptop”, “Windows Desktop / Laptop”, “iOS Phone / Tablet”) |

## Renommer les colonnes

* I am currently employed at least part-time : EMPLOYED
* I identify as having a mental illness : MENTAL\_ILLNESS
* Education : EDUCATION
* I have my own computer separate from a smart phone : OWN\_COMPUTER
* I have been hospitalized before for my mental illness : HOSPITALIZED
* How many days were you hospitalized for your mental illness : DAYS\_HOSPITALIZED
* I am legally disabled : DISABLED
* I have my regular access to the internet : HAS\_INTERNET
* I live with my parents : LIVE\_WITH\_PARENTS
* I have a gap in my resume : GAP\_RESUME
* Total length of any gaps in my resume in months. : GAP\_LENGTH
* Annual income (including any social welfare programs) in USD : ANNUAL\_INCOME
* I am unemployed : UNEMPLOYED
* I read outside of work and school : READ\_OUTSIDE
* Annual income from social welfare programs : ANNUAL\_SOCIAL\_WELFARE\_INCOME
* I receive food stamps : RECEIVE\_FOOD\_STAMPS
* I am on section 8 housing : SECTION\_8\_HOUSING
* How many times were you hospitalized for your mental illness : TIMES\_HOSPITALIZED
* Lack of concentration : LACK\_CONCENTRATION
* Anxiety : ANXIETY
* Depression : DEPRESSION
* Obsessive thinking : OBSESSIVE\_THINKING
* Mood swings : MOOD\_SWINGS
* Panic attacks : PANIC\_ATTACKS
* Compulsive behavior : COMPULSIVE\_BEHAVIOR
* Tiredness: TIREDNESS
* Age : AGE
* Gender : GENDER
* Household Income : HOUSEHOLD\_INCOME
* Region : REGION
* Device Type : DEVICE\_TYPE

library(dplyr)  
  
mental\_health <- rename(mental\_health,   
 "EMPLOYED" = "I am currently employed at least part-time",  
 "MENTAL\_ILLNESS" = "I identify as having a mental illness",  
 "EDUCATION" = "Education",  
 "OWN\_COMPUTER" = "I have my own computer separate from a smart phone",  
 "HOSPITALIZED" = "I have been hospitalized before for my mental illness",  
 "DAYS\_HOSPITALIZED" = "How many days were you hospitalized for your mental illness",  
 "DISABLED" = "I am legally disabled",  
 "HAS\_INTERNET" = "I have my regular access to the internet",  
 "LIVE\_WITH\_PARENTS" = "I live with my parents",  
 "GAP\_RESUME" = "I have a gap in my resume",  
 "GAP\_LENGTH" = "Total length of any gaps in my resume in months.",  
 "ANNUAL\_INCOME" = "Annual income (including any social welfare programs) in USD",  
 "UNEMPLOYED" = "I am unemployed",  
 "READ\_OUTSIDE" = "I read outside of work and school",  
 "ANNUAL\_SOCIAL\_WELFARE\_INCOME" = "Annual income from social welfare programs",  
 "RECEIVE\_FOOD\_STAMPS" = "I receive food stamps",  
 "SECTION\_8\_HOUSING" = "I am on section 8 housing",  
 "TIMES\_HOSPITALIZED" = "How many times were you hospitalized for your mental illness",  
 "LACK\_CONCENTRATION" = "Lack of concentration",  
 "ANXIETY" = "Anxiety",  
 "DEPRESSION" = "Depression",  
 "OBSESSIVE\_THINKING" = "Obsessive thinking",  
 "MOOD\_SWINGS" = "Mood swings",  
 "PANIC\_ATTACKS" = "Panic attacks",  
 "COMPULSIVE\_BEHAVIOR" = "Compulsive behavior",  
 "TIREDNESS" = "Tiredness",  
   
 "AGE" = "Age",  
 "GENDER" = "Gender",  
 "HOUSEHOLD\_INCOME" = "Household Income",  
 "REGION" = "Region",  
 "DEVICE\_TYPE" = "Device Type")

Nous allons enlever la colonne UNEMPLOYED du jeu de donnée car elle est corrélée à la colonne EMPLOYED.

mental\_health <- select(mental\_health, -UNEMPLOYED)

## Est-ce que R a bien reconnu le type des données ?

### Convertir des colonnes en facteur

Convertir toutes les colonnes en facteur à l’exception des colonnes DAYS\_HOSPITALZED, GAP\_LENGTH,ANNUAL\_SOCIAL\_WELFARE\_INCOME, ANNUAL\_INCOME, TIMES\_HOSPITALIZED

all\_columns <- colnames(mental\_health)  
num\_cols <- c("DAYS\_HOSPITALIZED", "GAP\_LENGTH", "ANNUAL\_SOCIAL\_WELFARE\_INCOME" , "ANNUAL\_INCOME", "TIMES\_HOSPITALIZED")  
factor\_cols <- setdiff(all\_columns, num\_cols)  
mental\_health <- data.frame(mental\_health)  
mental\_health[factor\_cols] <- lapply(mental\_health[factor\_cols], factor)

summary(mental\_health)

EMPLOYED MENTAL\_ILLNESS EDUCATION OWN\_COMPUTER  
 0:107 0:254 Completed Undergraduate:100 0: 42   
 1:227 1: 80 Some Undergraduate : 82 1:292   
 High School or GED : 63   
 Completed Masters : 49   
 Some Masters : 12   
 Completed Phd : 10   
 (Other) : 18   
 HOSPITALIZED DAYS\_HOSPITALIZED DISABLED HAS\_INTERNET LIVE\_WITH\_PARENTS  
 0:308 Min. : 0.000 0:301 0: 12 0:297   
 1: 26 1st Qu.: 0.000 1: 33 1:322 1: 37   
 Median : 0.000   
 Mean : 3.276   
 3rd Qu.: 0.000   
 Max. :100.000   
 NA's :37   
 GAP\_RESUME GAP\_LENGTH ANNUAL\_INCOME READ\_OUTSIDE  
 0:252 Min. : 0.000 Min. : 0.00 0: 37   
 1: 82 1st Qu.: 0.000 1st Qu.: 12.00 1:297   
 Median : 0.000 Median : 30.00   
 Mean : 8.497 Mean : 37.43   
 3rd Qu.: 4.750 3rd Qu.: 55.00   
 Max. :100.000 Max. :100.00   
   
 ANNUAL\_SOCIAL\_WELFARE\_INCOME RECEIVE\_FOOD\_STAMPS SECTION\_8\_HOUSING  
 Min. : 0.000 0:312 0:327   
 1st Qu.: 0.000 1: 22 1: 7   
 Median : 0.000   
 Mean : 3.326   
 3rd Qu.: 0.000   
 Max. :100.000   
   
 TIMES\_HOSPITALIZED LACK\_CONCENTRATION ANXIETY DEPRESSION OBSESSIVE\_THINKING  
 Min. : 0.000 0 :282 0:234 0:248 0 :291   
 1st Qu.: 0.000 1 : 51 1:100 1: 86 1 : 42   
 Median : 0.000 NA's: 1 NA's: 1   
 Mean : 1.195   
 3rd Qu.: 0.000   
 Max. :100.000   
   
 MOOD\_SWINGS PANIC\_ATTACKS COMPULSIVE\_BEHAVIOR TIREDNESS AGE   
 0 :295 0 :284 0 :304 0 :233 > 60 : 80   
 1 : 38 1 : 49 1 : 29 1 :100 18-29: 52   
 NA's: 1 NA's: 1 NA's: 1 NA's: 1 30-44:103   
 45-60: 99   
   
   
   
 GENDER HOUSEHOLD\_INCOME REGION   
 Female:176 $25,000-$49,999 :68 South Atlantic :63   
 Male :158 $50,000-$74,999 :57 Middle Atlantic :56   
 Prefer not to answer:38 East North Central:50   
 $10,000-$24,999 :35 Pacific :46   
 $75,000-$99,999 :33 Mountain :32   
 $0-$9,999 :27 (Other) :85   
 (Other) :76 NA's : 2   
 DEVICE\_TYPE   
 Android Phone / Tablet : 93   
 iOS Phone / Tablet : 93   
 MacOS Desktop / Laptop : 24   
 Other : 2   
 Windows Desktop / Laptop:122

str(mental\_health)

'data.frame': 334 obs. of 30 variables:  
 $ EMPLOYED : Factor w/ 2 levels "0","1": 1 2 2 1 2 2 2 2 2 2 ...  
 $ MENTAL\_ILLNESS : Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 2 1 2 ...  
 $ EDUCATION : Factor w/ 8 levels "Completed Masters",..: 4 6 3 7 3 4 7 7 3 8 ...  
 $ OWN\_COMPUTER : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 2 2 2 2 ...  
 $ HOSPITALIZED : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
 $ DAYS\_HOSPITALIZED : num 0 0 0 NA 35 0 0 0 0 0 ...  
 $ DISABLED : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
 $ HAS\_INTERNET : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...  
 $ LIVE\_WITH\_PARENTS : Factor w/ 2 levels "0","1": 1 1 1 2 1 2 1 2 1 1 ...  
 $ GAP\_RESUME : Factor w/ 2 levels "0","1": 2 1 1 2 2 1 1 1 1 1 ...  
 $ GAP\_LENGTH : num 24 1 0 11 33 0 0 0 0 0 ...  
 $ ANNUAL\_INCOME : num 35 22 100 0 32 0 1 11 73 12 ...  
 $ READ\_OUTSIDE : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...  
 $ ANNUAL\_SOCIAL\_WELFARE\_INCOME: num 0 0 0 0 30 0 0 0 0 0 ...  
 $ RECEIVE\_FOOD\_STAMPS : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
 $ SECTION\_8\_HOUSING : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
 $ TIMES\_HOSPITALIZED : num 0 0 0 0 4 0 0 0 0 0 ...  
 $ LACK\_CONCENTRATION : Factor w/ 2 levels "0","1": 2 2 1 1 2 1 1 2 2 1 ...  
 $ ANXIETY : Factor w/ 2 levels "0","1": 2 2 1 1 2 1 1 2 2 2 ...  
 $ DEPRESSION : Factor w/ 2 levels "0","1": 2 2 1 1 2 1 1 2 2 2 ...  
 $ OBSESSIVE\_THINKING : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 1 1 1 ...  
 $ MOOD\_SWINGS : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
 $ PANIC\_ATTACKS : Factor w/ 2 levels "0","1": 2 2 1 1 2 1 1 2 1 1 ...  
 $ COMPULSIVE\_BEHAVIOR : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...  
 $ TIREDNESS : Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 2 2 2 ...  
 $ AGE : Factor w/ 4 levels "> 60","18-29",..: 3 2 3 3 3 3 2 2 3 3 ...  
 $ GENDER : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...  
 $ HOUSEHOLD\_INCOME : Factor w/ 11 levels "$0-$9,999","$10,000-$24,999",..: 8 9 5 8 8 1 3 3 9 4 ...  
 $ REGION : Factor w/ 9 levels "East North Central",..: 4 2 6 5 1 7 3 3 7 7 ...  
 $ DEVICE\_TYPE : Factor w/ 5 levels "Android Phone / Tablet",..: 1 3 3 5 2 1 5 5 2 1 ...

## Réaliser des visualisations

# Analyse des correspondances multiples

L’Analyse des Correspondances Multiples (ACM) est une méthode qui permet d’étudier l’association entre au moins deux variables qualitatives.

L’Analyse des Correspondances Multiples est aux variables qualitatives ce que l’Analyse en Composantes Principales est aux variables quantitatives. Elle permet en effet d’aboutir à des cartes de représentation sur lesquelles on peut visuellement observer les proximités entre les catégories des variables qualitatives et les observations.

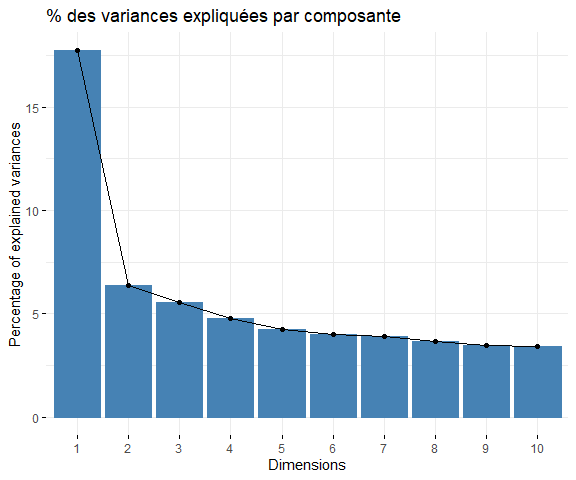
## Filtrer les colonnes

var\_quali\_illustratives <- c("HOUSEHOLD\_INCOME", "REGION", "DEVICE\_TYPE")  
idx\_quali\_illustratives <- match(var\_quali\_illustratives, all\_columns)

idx\_num\_cols <- match(num\_cols, all\_columns)

library(FactoMineR)  
library(factoextra)  
premiere\_acm <- MCA(X = mental\_health[1:333,], quanti.sup = idx\_num\_cols, quali.sup = idx\_quali\_illustratives, graph = FALSE)

fviz\_screeplot(premiere\_acm, title = "% des variances expliquées par composante")

 Nous retenons les 4 premières dimensions.

premiere\_acm$eig[1:4,]

eigenvalue percentage of variance cumulative percentage of variance  
dim 1 0.24179805 17.731857 17.73186  
dim 2 0.08717498 6.392832 24.12469  
dim 3 0.07537718 5.527660 29.65235  
dim 4 0.06489595 4.759036 34.41138

var\_actives <- setdiff(colnames(mental\_health), c(num\_cols, var\_quali\_illustratives))  
modalites\_acm <- row.names(premiere\_acm$var$contrib)

Les 4 premières dimensions expliquent 34.14% de la variance totale. En ACM le pouvoir explicatif des axes est généralement plus faible. Les variables actives que nous avons sélectionnées pour l’ACM sont **EMPLOYED, MENTAL\_ILLNESS, EDUCATION, OWN\_COMPUTER, HOSPITALIZED, DISABLED, HAS\_INTERNET, LIVE\_WITH\_PARENTS, GAP\_RESUME, READ\_OUTSIDE, RECEIVE\_FOOD\_STAMPS, SECTION\_8\_HOUSING, LACK\_CONCENTRATION, ANXIETY, DEPRESSION, OBSESSIVE\_THINKING, MOOD\_SWINGS, PANIC\_ATTACKS, COMPULSIVE\_BEHAVIOR, TIREDNESS, AGE, GENDER**. Contrairement à une ACP qui s’effectue directement sur des variables quantitatives continues, l’ACM se réalise sur les modalités que prend chaque variable. Il y a donc nécessairement une multiplication du nombre de colonnes qui entrent dans le calcul des composantes. Chaque modalité est considérée comme une colonne à part entière. Voilà la liste complète des colonnes qui servent à calculer les composantes : **EMPLOYED\_0, EMPLOYED\_1, MENTAL\_ILLNESS\_0, MENTAL\_ILLNESS\_1, Completed Masters, Completed Phd, Completed Undergraduate, High School or GED, Some highschool, Some Phd, Some Undergraduate, Some Masters, OWN\_COMPUTER\_0, OWN\_COMPUTER\_1, HOSPITALIZED\_0, HOSPITALIZED\_1, DISABLED\_0, DISABLED\_1, HAS\_INTERNET\_0, HAS\_INTERNET\_1, LIVE\_WITH\_PARENTS\_0, LIVE\_WITH\_PARENTS\_1, GAP\_RESUME\_0, GAP\_RESUME\_1, READ\_OUTSIDE\_0, READ\_OUTSIDE\_1, RECEIVE\_FOOD\_STAMPS\_0, RECEIVE\_FOOD\_STAMPS\_1, SECTION\_8\_HOUSING\_0, SECTION\_8\_HOUSING\_1, LACK\_CONCENTRATION\_0, LACK\_CONCENTRATION\_1, ANXIETY\_0, ANXIETY\_1, DEPRESSION\_0, DEPRESSION\_1, OBSESSIVE\_THINKING\_0, OBSESSIVE\_THINKING\_1, MOOD\_SWINGS\_0, MOOD\_SWINGS\_1, PANIC\_ATTACKS\_0, PANIC\_ATTACKS\_1, COMPULSIVE\_BEHAVIOR\_0, COMPULSIVE\_BEHAVIOR\_1, TIREDNESS\_0, TIREDNESS\_1, > 60, 18-29, 30-44, 45-60, Female, Male**.

Nous passons donc de *22* colonnes initiales à *52*.

dimdesc(premiere\_acm)

$`Dim 1`  
  
Link between the variable and the continuous variables (R-square)  
=================================================================================  
 correlation p.value  
GAP\_LENGTH 0.3585876 1.534298e-11  
DAYS\_HOSPITALIZED 0.3085145 8.948495e-09  
TIMES\_HOSPITALIZED 0.2627934 1.154794e-06  
ANNUAL\_INCOME -0.2087179 1.246669e-04  
  
Link between the variable and the categorical variable (1-way anova)  
=============================================  
 R2 p.value  
DEPRESSION 0.57744815 6.968417e-64  
MENTAL\_ILLNESS 0.54113907 6.055699e-58  
OBSESSIVE\_THINKING 0.52775269 7.151139e-56  
ANXIETY 0.52309816 3.641033e-55  
PANIC\_ATTACKS 0.47590876 2.308628e-48  
COMPULSIVE\_BEHAVIOR 0.46997715 1.496083e-47  
MOOD\_SWINGS 0.45210608 3.687904e-45  
LACK\_CONCENTRATION 0.33756220 1.881540e-31  
HOSPITALIZED 0.26557117 5.504913e-24  
GAP\_RESUME 0.18815069 1.040330e-16  
TIREDNESS 0.18114325 4.394317e-16  
AGE 0.18856238 7.538136e-15  
DISABLED 0.11332971 2.880684e-10  
LIVE\_WITH\_PARENTS 0.10173124 2.605386e-09  
RECEIVE\_FOOD\_STAMPS 0.08861320 3.063642e-08  
HOUSEHOLD\_INCOME 0.14312734 2.305075e-07  
OWN\_COMPUTER 0.07767471 2.347317e-07  
DEVICE\_TYPE 0.09289733 1.845499e-06  
EDUCATION 0.10589902 5.449513e-06  
EMPLOYED 0.05181425 2.748417e-05  
GENDER 0.02686652 2.698490e-03  
SECTION\_8\_HOUSING 0.01345318 3.436469e-02  
  
Link between variable abd the categories of the categorical variables  
================================================================  
 Estimate p.value  
DEPRESSION=DEPRESSION\_1 0.42851085 6.968417e-64  
MENTAL\_ILLNESS=MENTAL\_ILLNESS\_1 0.42517202 6.055699e-58  
OBSESSIVE\_THINKING=OBSESSIVE\_THINKING\_1 0.53800303 7.151139e-56  
ANXIETY=ANXIETY\_1 0.38905115 3.641033e-55  
PANIC\_ATTACKS=PANIC\_ATTACKS\_1 0.47879022 2.308628e-48  
COMPULSIVE\_BEHAVIOR=COMPULSIVE\_BEHAVIOR\_1 0.59778253 1.496083e-47  
MOOD\_SWINGS=MOOD\_SWINGS\_1 0.51994571 3.687904e-45  
LACK\_CONCENTRATION=LACK\_CONCENTRATION\_1 0.39665040 1.881540e-31  
HOSPITALIZED=HOSPITALIZED\_1 0.47225355 5.504913e-24  
GAP\_RESUME=GAP\_RESUME\_1 0.24754249 1.040330e-16  
TIREDNESS=TIREDNESS\_1 0.22828309 4.394317e-16  
DISABLED=DISABLED\_1 0.28083697 2.880684e-10  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_1 0.24952903 2.605386e-09  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_1 0.29464448 3.063642e-08  
DEVICE\_TYPE=Android Phone / Tablet 0.31231936 3.489006e-08  
AGE=30-44 0.19008073 2.228590e-07  
OWN\_COMPUTER=OWN\_COMPUTER\_0 0.20639982 2.347317e-07  
HOUSEHOLD\_INCOME=$0-$9,999 0.51934090 3.481945e-07  
EMPLOYED=EMPLOYED\_0 0.11984479 2.748417e-05  
EDUCATION=Some Undergraduate 0.23928377 3.070356e-05  
AGE=18-29 0.23695288 6.049056e-05  
GENDER=Female 0.08070468 2.698490e-03  
HOUSEHOLD\_INCOME=$10,000-$24,999 0.27498369 8.291028e-03  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_1 0.19879032 3.436469e-02  
EDUCATION=High School or GED 0.15419052 4.642777e-02  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_0 -0.19879032 3.436469e-02  
HOUSEHOLD\_INCOME=$75,000-$99,999 -0.11603508 2.642109e-02  
AGE=45-60 -0.13271520 4.424380e-03  
GENDER=Male -0.08070468 2.698490e-03  
DEVICE\_TYPE=Windows Desktop / Laptop -0.03370884 1.830527e-03  
EDUCATION=Completed Masters -0.20243639 1.385467e-04  
EMPLOYED=EMPLOYED\_1 -0.11984479 2.748417e-05  
OWN\_COMPUTER=OWN\_COMPUTER\_1 -0.20639982 2.347317e-07  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_0 -0.29464448 3.063642e-08  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_0 -0.24952903 2.605386e-09  
AGE=> 60 -0.29431841 2.490796e-09  
DISABLED=DISABLED\_0 -0.28083697 2.880684e-10  
TIREDNESS=TIREDNESS\_0 -0.22828309 4.394317e-16  
GAP\_RESUME=GAP\_RESUME\_0 -0.24754249 1.040330e-16  
HOSPITALIZED=HOSPITALIZED\_0 -0.47225355 5.504913e-24  
LACK\_CONCENTRATION=LACK\_CONCENTRATION\_0 -0.39665040 1.881540e-31  
MOOD\_SWINGS=MOOD\_SWINGS\_0 -0.51994571 3.687904e-45  
COMPULSIVE\_BEHAVIOR=COMPULSIVE\_BEHAVIOR\_0 -0.59778253 1.496083e-47  
PANIC\_ATTACKS=PANIC\_ATTACKS\_0 -0.47879022 2.308628e-48  
ANXIETY=ANXIETY\_0 -0.38905115 3.641033e-55  
OBSESSIVE\_THINKING=OBSESSIVE\_THINKING\_0 -0.53800303 7.151139e-56  
MENTAL\_ILLNESS=MENTAL\_ILLNESS\_0 -0.42517202 6.055699e-58  
DEPRESSION=DEPRESSION\_0 -0.42851085 6.968417e-64  
  
$`Dim 2`  
  
Link between the variable and the continuous variables (R-square)  
=================================================================================  
 correlation p.value  
GAP\_LENGTH 0.2056641 1.572167e-04  
DAYS\_HOSPITALIZED 0.1103563 4.417867e-02  
ANNUAL\_INCOME -0.2546445 2.511719e-06  
  
Link between the variable and the categorical variable (1-way anova)  
=============================================  
 R2 p.value  
EDUCATION 0.30937834 4.313255e-23  
RECEIVE\_FOOD\_STAMPS 0.25153708 1.296527e-22  
EMPLOYED 0.23872551 2.206811e-21  
OWN\_COMPUTER 0.21099079 8.747851e-19  
HAS\_INTERNET 0.15243038 1.432896e-13  
LIVE\_WITH\_PARENTS 0.12091402 6.739579e-11  
SECTION\_8\_HOUSING 0.09443148 1.030148e-08  
HOUSEHOLD\_INCOME 0.14957304 8.087232e-08  
TIREDNESS 0.08055286 1.375769e-07  
READ\_OUTSIDE 0.07663250 2.847616e-07  
DEVICE\_TYPE 0.09185855 2.203175e-06  
DISABLED 0.06307287 3.479373e-06  
LACK\_CONCENTRATION 0.06162686 4.539196e-06  
DEPRESSION 0.05063281 3.412900e-05  
OBSESSIVE\_THINKING 0.03707919 4.096358e-04  
GAP\_RESUME 0.03294956 8.757794e-04  
MENTAL\_ILLNESS 0.03151922 1.140175e-03  
GENDER 0.02541430 3.535808e-03  
COMPULSIVE\_BEHAVIOR 0.01857622 1.279524e-02  
PANIC\_ATTACKS 0.01753405 1.560867e-02  
AGE 0.02734707 2.753237e-02  
HOSPITALIZED 0.01166718 4.890292e-02  
  
Link between variable abd the categories of the categorical variables  
================================================================  
 Estimate p.value  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_1 0.29807094 1.296527e-22  
EMPLOYED=EMPLOYED\_0 0.15445909 2.206811e-21  
OWN\_COMPUTER=OWN\_COMPUTER\_0 0.20425409 8.747851e-19  
EDUCATION=Some highschool 0.63675089 5.102463e-14  
HAS\_INTERNET=HAS\_INTERNET\_0 0.30924507 1.432896e-13  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_1 0.16334347 6.739579e-11  
EDUCATION=High School or GED 0.17227669 6.676511e-10  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_1 0.31623559 1.030148e-08  
DEVICE\_TYPE=Android Phone / Tablet 0.14867993 9.156367e-08  
TIREDNESS=TIREDNESS\_0 0.09140557 1.375769e-07  
READ\_OUTSIDE=READ\_OUTSIDE\_0 0.13003790 2.847616e-07  
DISABLED=DISABLED\_1 0.12579785 3.479373e-06  
LACK\_CONCENTRATION=LACK\_CONCENTRATION\_0 0.10176198 4.539196e-06  
DEPRESSION=DEPRESSION\_0 0.07618875 3.412900e-05  
HOUSEHOLD\_INCOME=$0-$9,999 0.23210611 5.670388e-05  
OBSESSIVE\_THINKING=OBSESSIVE\_THINKING\_0 0.08562570 4.096358e-04  
GAP\_RESUME=GAP\_RESUME\_1 0.06220011 8.757794e-04  
MENTAL\_ILLNESS=MENTAL\_ILLNESS\_0 0.06161231 1.140175e-03  
HOUSEHOLD\_INCOME=$10,000-$24,999 0.16732472 1.400651e-03  
GENDER=Female 0.04713046 3.535808e-03  
COMPULSIVE\_BEHAVIOR=COMPULSIVE\_BEHAVIOR\_0 0.07135970 1.279524e-02  
PANIC\_ATTACKS=PANIC\_ATTACKS\_0 0.05518158 1.560867e-02  
AGE=18-29 0.07741673 2.058438e-02  
HOUSEHOLD\_INCOME=Prefer not to answer 0.10846509 3.752579e-02  
HOSPITALIZED=HOSPITALIZED\_1 0.05943430 4.890292e-02  
EDUCATION=Completed Masters -0.10611688 4.922236e-02  
HOSPITALIZED=HOSPITALIZED\_0 -0.05943430 4.890292e-02  
HOUSEHOLD\_INCOME=$125,000-$149,999 -0.14946551 2.265743e-02  
DEVICE\_TYPE=MacOS Desktop / Laptop -0.12150735 2.221795e-02  
HOUSEHOLD\_INCOME=$200,000+ -0.13548949 1.894663e-02  
HOUSEHOLD\_INCOME=$100,000-$124,999 -0.12370535 1.710716e-02  
PANIC\_ATTACKS=PANIC\_ATTACKS\_1 -0.05518158 1.560867e-02  
EDUCATION=Completed Undergraduate -0.09039736 1.351118e-02  
AGE=45-60 -0.07244946 1.290362e-02  
COMPULSIVE\_BEHAVIOR=COMPULSIVE\_BEHAVIOR\_1 -0.07135970 1.279524e-02  
HOUSEHOLD\_INCOME=$75,000-$99,999 -0.10786131 1.194269e-02  
EDUCATION=Some Undergraduate -0.10125148 1.177987e-02  
DEVICE\_TYPE=iOS Phone / Tablet -0.05533520 1.035633e-02  
GENDER=Male -0.04713046 3.535808e-03  
EDUCATION=Some Phd -0.35413978 1.575561e-03  
MENTAL\_ILLNESS=MENTAL\_ILLNESS\_1 -0.06161231 1.140175e-03  
GAP\_RESUME=GAP\_RESUME\_0 -0.06220011 8.757794e-04  
OBSESSIVE\_THINKING=OBSESSIVE\_THINKING\_1 -0.08562570 4.096358e-04  
DEPRESSION=DEPRESSION\_1 -0.07618875 3.412900e-05  
LACK\_CONCENTRATION=LACK\_CONCENTRATION\_1 -0.10176198 4.539196e-06  
DISABLED=DISABLED\_0 -0.12579785 3.479373e-06  
READ\_OUTSIDE=READ\_OUTSIDE\_1 -0.13003790 2.847616e-07  
TIREDNESS=TIREDNESS\_1 -0.09140557 1.375769e-07  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_0 -0.31623559 1.030148e-08  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_0 -0.16334347 6.739579e-11  
HAS\_INTERNET=HAS\_INTERNET\_1 -0.30924507 1.432896e-13  
OWN\_COMPUTER=OWN\_COMPUTER\_1 -0.20425409 8.747851e-19  
EMPLOYED=EMPLOYED\_1 -0.15445909 2.206811e-21  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_0 -0.29807094 1.296527e-22  
  
$`Dim 3`  
  
Link between the variable and the categorical variable (1-way anova)  
=============================================  
 R2 p.value  
AGE 0.44320277 1.424407e-41  
DISABLED 0.21971229 1.362762e-19  
READ\_OUTSIDE 0.16741213 7.156843e-15  
EMPLOYED 0.14486522 6.390747e-13  
LIVE\_WITH\_PARENTS 0.13160878 8.535718e-12  
SECTION\_8\_HOUSING 0.12875097 1.485785e-11  
EDUCATION 0.13875685 2.356652e-08  
RECEIVE\_FOOD\_STAMPS 0.07563767 3.423926e-07  
HOSPITALIZED 0.07229941 6.349712e-07  
GENDER 0.03177750 1.087099e-03  
ANXIETY 0.02163940 7.168061e-03  
OWN\_COMPUTER 0.02129860 7.643050e-03  
DEVICE\_TYPE 0.04132954 7.667317e-03  
TIREDNESS 0.01908725 1.161087e-02  
  
Link between variable abd the categories of the categorical variables  
================================================================  
 Estimate p.value  
AGE=> 60 0.27987630 1.020591e-23  
DISABLED=DISABLED\_1 0.21832482 1.362762e-19  
READ\_OUTSIDE=READ\_OUTSIDE\_1 0.17872320 7.156843e-15  
EMPLOYED=EMPLOYED\_0 0.11188455 6.390747e-13  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_0 0.15846374 8.535718e-12  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_1 0.34336151 1.485785e-11  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_1 0.15198885 3.423926e-07  
HOSPITALIZED=HOSPITALIZED\_1 0.13757695 6.349712e-07  
EDUCATION=Completed Undergraduate 0.11186320 6.463328e-07  
GENDER=Male 0.04900569 1.087099e-03  
ANXIETY=ANXIETY\_0 0.04418057 7.168061e-03  
OWN\_COMPUTER=OWN\_COMPUTER\_1 0.06034463 7.643050e-03  
TIREDNESS=TIREDNESS\_1 0.04137404 1.161087e-02  
DEVICE\_TYPE=Windows Desktop / Laptop -0.02404875 1.466297e-02  
DEVICE\_TYPE=iOS Phone / Tablet -0.13184196 1.368415e-02  
TIREDNESS=TIREDNESS\_0 -0.04137404 1.161087e-02  
OWN\_COMPUTER=OWN\_COMPUTER\_0 -0.06034463 7.643050e-03  
ANXIETY=ANXIETY\_1 -0.04418057 7.168061e-03  
GENDER=Female -0.04900569 1.087099e-03  
EDUCATION=Some Undergraduate -0.09386435 4.175555e-04  
EDUCATION=High School or GED -0.12466778 6.019193e-05  
HOSPITALIZED=HOSPITALIZED\_0 -0.13757695 6.349712e-07  
RECEIVE\_FOOD\_STAMPS=RECEIVE\_FOOD\_STAMPS\_0 -0.15198885 3.423926e-07  
SECTION\_8\_HOUSING=SECTION\_8\_HOUSING\_0 -0.34336151 1.485785e-11  
LIVE\_WITH\_PARENTS=LIVE\_WITH\_PARENTS\_1 -0.15846374 8.535718e-12  
EMPLOYED=EMPLOYED\_1 -0.11188455 6.390747e-13  
READ\_OUTSIDE=READ\_OUTSIDE\_0 -0.17872320 7.156843e-15  
DISABLED=DISABLED\_0 -0.21832482 1.362762e-19  
AGE=18-29 -0.31521741 5.994728e-26