Untitled

Md Mushfiqur Rahman

5/31/2022

#install.packages("gtsummary")  
#install.packages("finalfit")  
library(gtsummary)  
library(finalfit)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.7 v stringr 1.4.0  
## v tidyr 1.2.0 v forcats 0.5.1  
## v readr 2.0.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# Summary Statistics table (Finalfit package)

# Load example dataset, modified version of survival::colon  
data(colon\_s)  
# glimpse of the colon\_s dataset  
str(colon\_s)

## 'data.frame': 929 obs. of 32 variables:  
## $ id : num 1 2 3 4 5 6 7 8 9 10 ...  
## $ rx : Factor w/ 3 levels "Obs","Lev","Lev+5FU": 3 3 1 3 1 3 2 1 2 3 ...  
## $ sex : num 1 1 0 0 1 0 1 1 1 0 ...  
## $ age : num 43 63 71 66 69 57 77 54 46 68 ...  
## ..- attr(\*, "label")= chr "Age (years)"  
## $ obstruct : num NA 0 0 1 0 0 0 0 0 0 ...  
## $ perfor : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ adhere : num 0 0 1 0 0 0 0 0 1 0 ...  
## $ nodes : num 5 1 7 6 22 9 5 1 2 1 ...  
## $ status : num 1 0 1 1 1 1 1 0 0 0 ...  
## $ differ : num 2 2 2 2 2 2 2 2 2 2 ...  
## $ extent : num 3 3 2 3 3 3 3 3 3 3 ...  
## $ surg : num 0 0 0 1 1 0 1 0 0 1 ...  
## $ node4 : num 1 0 1 1 1 1 1 0 0 0 ...  
## $ time : num 1521 3087 963 293 659 ...  
## $ sex.factor : Factor w/ 2 levels "Female","Male": 2 2 1 1 2 1 2 2 2 1 ...  
## ..- attr(\*, "label")= chr "Sex"  
## $ rx.factor : Factor w/ 3 levels "Obs","Lev","Lev+5FU": 3 3 1 3 1 3 2 1 2 3 ...  
## ..- attr(\*, "label")= chr "Treatment"  
## $ obstruct.factor: Factor w/ 2 levels "No","Yes": NA 1 1 2 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "Obstruction"  
## $ perfor.factor : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "Perforation"  
## $ adhere.factor : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 1 1 1 2 1 ...  
## ..- attr(\*, "label")= chr "Adherence"  
## $ differ.factor : Factor w/ 3 levels "Well","Moderate",..: 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Differentiation"  
## $ extent.factor : Factor w/ 4 levels "Submucosa","Muscle",..: 3 3 2 3 3 3 3 3 3 3 ...  
## ..- attr(\*, "label")= chr "Extent of spread"  
## $ surg.factor : Factor w/ 2 levels "Short","Long": 1 1 1 2 2 1 2 1 1 2 ...  
## ..- attr(\*, "label")= chr "Time from surgery"  
## $ node4.factor : Factor w/ 2 levels "No","Yes": 2 1 2 2 2 2 2 1 1 1 ...  
## ..- attr(\*, "label")= chr ">4 positive nodes"  
## $ status.factor : Factor w/ 2 levels "Alive","Died": 2 1 2 2 2 2 2 1 1 1 ...  
## ..- attr(\*, "label")= chr "Status"  
## $ age.factor : Factor w/ 3 levels "<40 years","40-59 years",..: 2 3 3 3 3 2 3 2 2 3 ...  
## ..- attr(\*, "label")= chr "Age"  
## $ loccomp : num NA 0 1 1 0 0 0 0 1 0 ...  
## $ loccomp.factor : Factor w/ 2 levels "No","Yes": NA 1 2 2 1 1 1 1 2 1 ...  
## ..- attr(\*, "label")= chr "Local complications"  
## $ time.years : num 4.167 8.458 2.638 0.803 1.805 ...  
## ..- attr(\*, "label")= chr "Time (years)"  
## $ mort\_5yr : Factor w/ 2 levels "Alive","Died": 2 1 2 2 2 2 2 1 1 1 ...  
## ..- attr(\*, "label")= chr "Mortality 5 year"  
## $ age.10 : num 4.3 6.3 7.1 6.6 6.9 5.7 7.7 5.4 4.6 6.8 ...  
## $ mort\_5yr.num : num 2 1 2 2 2 2 2 1 1 1 ...  
## $ hospital : Factor w/ 5 levels "hospital\_1","hospital\_2",..: 5 3 5 4 5 4 2 2 2 2 ...

# Table 1 - Patient demographics by variable of interest ----  
explanatory = c("age", "age.factor", "sex.factor", "obstruct.factor")  
dependent = "mort\_5yr"   
t1 <- colon\_s %>%  
 summary\_factorlist(dependent, explanatory,  
 cont = "mean", # cont is by default set to mean and corresponding hypothesis is tested. "median" gives non-parametric hypothesis with the IQR  
 p=TRUE, # adds p-value column  
 add\_dependent\_label=TRUE)

## Note: dependent includes missing data. These are dropped.

# Note: dependent(mort\_5yr) includes missing data. These are dropped.  
# To print the table as a knitr table  
knitr::kable(t1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: Mortality 5 year |  | Alive | Died | p |
| Age (years) | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 0.986 |
| Age | <40 years | 31 (6.1) | 36 (8.9) | 0.020 |
|  | 40-59 years | 208 (40.7) | 131 (32.4) |  |
|  | 60+ years | 272 (53.2) | 237 (58.7) |  |
| Sex | Female | 243 (47.6) | 194 (48.0) | 0.941 |
|  | Male | 268 (52.4) | 210 (52.0) |  |
| Obstruction | No | 408 (82.1) | 312 (78.6) | 0.219 |
|  | Yes | 89 (17.9) | 85 (21.4) |  |

# We can adjust the alignment of the table using the align argument.  
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: Mortality 5 year |  | Alive | Died | p |
| Age (years) | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 0.986 |
| Age | <40 years | 31 (6.1) | 36 (8.9) | 0.020 |
|  | 40-59 years | 208 (40.7) | 131 (32.4) |  |
|  | 60+ years | 272 (53.2) | 237 (58.7) |  |
| Sex | Female | 243 (47.6) | 194 (48.0) | 0.941 |
|  | Male | 268 (52.4) | 210 (52.0) |  |
| Obstruction | No | 408 (82.1) | 312 (78.6) | 0.219 |
|  | Yes | 89 (17.9) | 85 (21.4) |  |

# Here we can see that Chi-squared approximation may be incorrect warning is showed cause finalfit conducts Chi-square test as a default for categorical variables.  
# we can use p\_cat argument to specify which kind of test we want finalfit to perform or the categorical variables.  
# Here we are conducting fisher's exact test for the categorical variables.  
  
# summary\_factorlist only consists of "chisq" and "fisher" test for the categorical varables & for continious explanatory variables "aov"(anova test), "t.test", Kruskal wallis test and also Mann-Whitney U/ Wilcoxon rank sum test can be done.  
  
t1 <- colon\_s %>%  
 summary\_factorlist(dependent, explanatory,  
 p=TRUE, add\_dependent\_label=TRUE, p\_cat = "fisher")

## Note: dependent includes missing data. These are dropped.

# We can adjust the alignment of the table using the align argument.  
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: Mortality 5 year |  | Alive | Died | p |
| Age (years) | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 0.986 |
| Age | <40 years | 31 (6.1) | 36 (8.9) | 0.019 |
|  | 40-59 years | 208 (40.7) | 131 (32.4) |  |
|  | 60+ years | 272 (53.2) | 237 (58.7) |  |
| Sex | Female | 243 (47.6) | 194 (48.0) | 0.894 |
|  | Male | 268 (52.4) | 210 (52.0) |  |
| Obstruction | No | 408 (82.1) | 312 (78.6) | 0.203 |
|  | Yes | 89 (17.9) | 85 (21.4) |  |

# Suppose we need to show the total for each of the row of the previous table.   
# total\_col = TRUE argument does this for you.  
t1 <- colon\_s %>%  
 summary\_factorlist(dependent, explanatory,  
 p=TRUE, add\_dependent\_label=FALSE, p\_cat = "fisher",  
 total\_col = TRUE  
 )

## Note: dependent includes missing data. These are dropped.

# We can adjust the alignment of the table using the align argument.  
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| label | levels | Alive | Died | Total | p |
| Age (years) | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 59.8 (11.9) | 0.986 |
| Age | <40 years | 31 (6.1) | 36 (8.9) | 67 (7.3) | 0.019 |
|  | 40-59 years | 208 (40.7) | 131 (32.4) | 339 (37.0) |  |
|  | 60+ years | 272 (53.2) | 237 (58.7) | 509 (55.6) |  |
| Sex | Female | 243 (47.6) | 194 (48.0) | 437 (47.8) | 0.894 |
|  | Male | 268 (52.4) | 210 (52.0) | 478 (52.2) |  |
| Obstruction | No | 408 (82.1) | 312 (78.6) | 720 (80.5) | 0.203 |
|  | Yes | 89 (17.9) | 85 (21.4) | 174 (19.5) |  |

# to show the number of missing values for each of the variables.  
t1 <- colon\_s %>%  
 summary\_factorlist(dependent, explanatory,  
 p=TRUE, add\_dependent\_label=FALSE, p\_cat = "fisher",  
 total\_col = TRUE,   
 na\_include = TRUE # by default FALSE  
 )

## Note: dependent includes missing data. These are dropped.

# We can adjust the alignment of the table using the align argument.  
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| label | levels | Alive | Died | Total | p |
| Age (years) | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 59.8 (11.9) | 0.986 |
| Age | <40 years | 31 (6.1) | 36 (8.9) | 67 (7.3) | 0.019 |
|  | 40-59 years | 208 (40.7) | 131 (32.4) | 339 (37.0) |  |
|  | 60+ years | 272 (53.2) | 237 (58.7) | 509 (55.6) |  |
| Sex | Female | 243 (47.6) | 194 (48.0) | 437 (47.8) | 0.894 |
|  | Male | 268 (52.4) | 210 (52.0) | 478 (52.2) |  |
| Obstruction | No | 408 (79.8) | 312 (77.2) | 720 (78.7) | 0.203 |
|  | Yes | 89 (17.4) | 85 (21.0) | 174 (19.0) |  |
|  | (Missing) | 14 (2.7) | 7 (1.7) | 21 (2.3) |  |

# to add row total  
  
t1 <- colon\_s %>%  
 summary\_factorlist(dependent, explanatory,  
 p=TRUE, add\_dependent\_label=FALSE, p\_cat = "fisher",  
 add\_row\_totals = TRUE  
 )

## Note: dependent includes missing data. These are dropped.

# We can adjust the alignment of the table using the align argument.  
knitr::kable(t1, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| label | Total N | Missing N | levels | Alive | Died | p |
| Age (years) | 915 (100.0) | 0 | Mean (SD) | 59.8 (11.4) | 59.9 (12.5) | 0.986 |
| Age | 915 (100.0) | 0 | <40 years | 31 (6.1) | 36 (8.9) | 0.019 |
|  |  |  | 40-59 years | 208 (40.7) | 131 (32.4) |  |
|  |  |  | 60+ years | 272 (53.2) | 237 (58.7) |  |
| Sex | 915 (100.0) | 0 | Female | 243 (47.6) | 194 (48.0) | 0.894 |
|  |  |  | Male | 268 (52.4) | 210 (52.0) |  |
| Obstruction | 894 (97.7) | 21 | No | 408 (82.1) | 312 (78.6) | 0.203 |
|  |  |  | Yes | 89 (17.9) | 85 (21.4) |  |

# this is the end of the summary\_factorlist function from the finalfit packsge  
# to know more about this function type ?summary\_factorlist and hit run!!

# Summary table of gtsummary package

library(gtsummary)  
explanatory = c("age", "age.factor", "sex.factor", "obstruct.factor")  
dependent = "mort\_5yr"   
  
# Univariate tables  
t2 <- colon\_s %>%select("age","age.factor","sex.factor","obstruct.factor","mort\_5yr") %>% tbl\_summary()  
  
t2

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |
| --- | --- |
| **Characteristic** | **N = 929** |
| Age (years) | 61 (53, 69) |
| Age |  |
| <40 years | 70 (7.5%) |
| 40-59 years | 344 (37%) |
| 60+ years | 515 (55%) |
| Sex |  |
| Female | 445 (48%) |
| Male | 484 (52%) |
| Obstruction | 176 (19%) |
| Unknown | 21 |
| Mortality 5 year |  |
| Alive | 511 (56%) |
| Died | 404 (44%) |
| Unknown | 14 |

# this shows the univariate descriptive table for the selected variables along with the missing values. Note that for continious variable thei function provides Median and 1st & 3rd quantile values inside the round brackets.  
  
# for bivarite table "by" argument is used in the tbl\_summary which automatically creates bivariate table  
t2 <- colon\_s %>%select("age","age.factor","sex.factor","obstruct.factor","mort\_5yr") %>% tbl\_summary(by = "mort\_5yr")

## 14 observations missing `mort\_5yr` have been removed. To include these observations, use `forcats::fct\_explicit\_na()` on `mort\_5yr` column before passing to `tbl\_summary()`.

t2 # note that 14 missing values of the mort\_5yr variable has been removed automatically

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Alive**, N = 511 | **Died**, N = 404 |
| Age (years) | 60 (53, 68) | 62 (52, 69) |
| Age |  |  |
| <40 years | 31 (6.1%) | 36 (8.9%) |
| 40-59 years | 208 (41%) | 131 (32%) |
| 60+ years | 272 (53%) | 237 (59%) |
| Sex |  |  |
| Female | 243 (48%) | 194 (48%) |
| Male | 268 (52%) | 210 (52%) |
| Obstruction | 89 (18%) | 85 (21%) |
| Unknown | 14 | 7 |

# For continious is shows the median and 1st & 3rd quantile values inside the round brackets. the top "N=" shows the total for each pof the column  
  
# For hypothesis testing we just need to add "add\_p()" funciton with the help of pipe ( %>% )  
# then for removing the warning of the missing values of mort\_5yr we use forcats::fct\_explicit\_na()  
t2 <- colon\_s %>%select("age","age.factor","sex.factor","obstruct.factor","mort\_5yr") %>%   
 tbl\_summary(by = "mort\_5yr",  
 missing = "always", # Creates a row named "Unknown" for each of the variables  
 missing\_text = "(Missing)", # Changes the newly created "Unknown" rows to "Missing"  
 # digits = everything()~ 2 # number of digits the summary statistics will be rounded to  
 ) %>% add\_p() # Wilcoxon sum rank test was applied on continious variable, for categorical data pearcon shi-square test was applied

## 14 observations missing `mort\_5yr` have been removed. To include these observations, use `forcats::fct\_explicit\_na()` on `mort\_5yr` column before passing to `tbl\_summary()`.

t2

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Alive**, N = 511 | **Died**, N = 404 | **p-value** |
| Age (years) | 60 (53, 68) | 62 (52, 69) | 0.5 |
| (Missing) | 0 | 0 |  |
| Age |  |  | 0.020 |
| <40 years | 31 (6.1%) | 36 (8.9%) |  |
| 40-59 years | 208 (41%) | 131 (32%) |  |
| 60+ years | 272 (53%) | 237 (59%) |  |
| (Missing) | 0 | 0 |  |
| Sex |  |  | 0.9 |
| Female | 243 (48%) | 194 (48%) |  |
| Male | 268 (52%) | 210 (52%) |  |
| (Missing) | 0 | 0 |  |
| Obstruction | 89 (18%) | 85 (21%) | 0.2 |
| (Missing) | 14 | 7 |  |

# to change the statistics calculation for the variables  
# for the continuous variable we want the mean and the sd   
  
t2 <- colon\_s %>%select("age","age.factor","sex.factor","obstruct.factor","mort\_5yr") %>%   
 tbl\_summary(by = "mort\_5yr",  
 missing = "always", # Creates a row named "Unknown" for each of the variables  
 missing\_text = "(Missing)", # Changes the newly created "Unknown" rows to "Missing"  
 # digits = everything()~ 2, # number of digits the summary statistics will be rounded to  
 statistic = list(  
 all\_continuous() ~ "{mean} ({sd})" # , # calculates mean and standard deviation for each of the continuous variables   
 # all\_categorical() ~ "{n} / {N} ({p}%)" # shows the number of number of observations for each row divided by total of that column and their percentage  
 ),  
 label = list(age.factor ~ "Age Group", # changes the name of the variable age.factor to "Age Group" in the output table.  
 age ~ "Age continuos"), # Changes the name of the age variable to "Age Continuous" in the output table.  
 ) %>%   
 add\_p() %>%   
 bold\_labels() %>% # Bolds the labels  
# add\_n() # Shows the total number of observations per variable  
 add\_overall(last =TRUE) # %>% add\_p()

## 14 observations missing `mort\_5yr` have been removed. To include these observations, use `forcats::fct\_explicit\_na()` on `mort\_5yr` column before passing to `tbl\_summary()`.

# adds a total column for each row. notice that this column is at the last of thable (for last = TRUE)  
# to have p value column at the last just cut the add\_p column from before and paste is after add\_overall(last= t ) with pipe  
   
  
t2

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **Alive**, N = 511 | **Died**, N = 404 | **p-value** | **Overall**, N = 915 |
| **Age continuos** | 60 (11) | 60 (13) | 0.5 | 60 (12) |
| (Missing) | 0 | 0 |  | 0 |
| **Age Group** |  |  | 0.020 |  |
| <40 years | 31 (6.1%) | 36 (8.9%) |  | 67 (7.3%) |
| 40-59 years | 208 (41%) | 131 (32%) |  | 339 (37%) |
| 60+ years | 272 (53%) | 237 (59%) |  | 509 (56%) |
| (Missing) | 0 | 0 |  | 0 |
| **Sex** |  |  | 0.9 |  |
| Female | 243 (48%) | 194 (48%) |  | 437 (48%) |
| Male | 268 (52%) | 210 (52%) |  | 478 (52%) |
| (Missing) | 0 | 0 |  | 0 |
| **Obstruction** | 89 (18%) | 85 (21%) | 0.2 | 174 (19%) |
| (Missing) | 14 | 7 |  | 21 |

# For multiple statistics for variables we can add statistics we want to calculate.  
t2 <- colon\_s %>%select("age","age.factor","sex.factor","obstruct.factor","mort\_5yr") %>%   
 tbl\_summary(by = "mort\_5yr",  
 missing = "always", # Creates a row named "Unknown" for each of the variables  
 missing\_text = "(Missing)", # Changes the newly created "Unknown" rows to "Missing"  
 # digits = everything()~ 2, # number of digits the summary statistics will be rounded to  
 type = list(all\_continuous() ~ "continuous2"  
 ),# add this for continious variable  
 statistic = list(all\_continuous() ~ c("{N\_nonmiss}", # stats to show  
 "{median} ({p25}, {p75})",   
 "{min}, {max}")  
   
 ),  
 label = list(age.factor ~ "Age Group", # changes the name of the variable age.factor to "Age Group" in the output table.  
 age ~ "Age continuos"), # Changes the name of the age variable to "Age Continuous" in the output table.  
 ) %>%   
 bold\_labels() %>%   
 add\_overall(last =TRUE) %>% add\_p()

## 14 observations missing `mort\_5yr` have been removed. To include these observations, use `forcats::fct\_explicit\_na()` on `mort\_5yr` column before passing to `tbl\_summary()`.

t2

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **Alive**, N = 511 | **Died**, N = 404 | **Overall**, N = 915 | **p-value** |
| **Age continuos** |  |  |  | 0.5 |
| N | 511 | 404 | 915 |  |
| Median (IQR) | 60 (53, 68) | 62 (52, 69) | 61 (53, 69) |  |
| Range | 22, 83 | 18, 85 | 18, 85 |  |
| (Missing) | 0 | 0 | 0 |  |
| **Age Group** |  |  |  | 0.020 |
| <40 years | 31 (6.1%) | 36 (8.9%) | 67 (7.3%) |  |
| 40-59 years | 208 (41%) | 131 (32%) | 339 (37%) |  |
| 60+ years | 272 (53%) | 237 (59%) | 509 (56%) |  |
| (Missing) | 0 | 0 | 0 |  |
| **Sex** |  |  |  | 0.9 |
| Female | 243 (48%) | 194 (48%) | 437 (48%) |  |
| Male | 268 (52%) | 210 (52%) | 478 (52%) |  |
| (Missing) | 0 | 0 | 0 |  |
| **Obstruction** | 89 (18%) | 85 (21%) | 174 (19%) | 0.2 |
| (Missing) | 14 | 7 | 21 |  |

# N.B you can also modify the table header, footnote and everything as you wish. For more information visit https://www.danieldsjoberg.com/gtsummary/articles/tbl\_summary.html

# Different regression model table (finalfit package & gtsummary)

# Creating data for linear regression & logistics model  
reg\_data <- data.frame(  
 y = rnorm(1000, mean = 5, sd = 5),  
 x1 = rnorm(1000),  
 x2 = rpois(1000, 3),  
 x3 = sample(1:10000, 1000, replace = T),  
 x4 = as.factor(sample(c("Male","Female"), 1000, replace = TRUE)),  
 x5 = as.factor(sample(c("USA","UK","Canada","France","Spain"), 1000, replace = TRUE)),  
 Disease = as.factor(sample(c(0,1), 1000, replace = TRUE))  
)  
  
head(reg\_data)

## y x1 x2 x3 x4 x5 Disease  
## 1 -5.6525456 -1.04766098 2 4259 Female USA 1  
## 2 -0.5600106 -0.02844858 6 2550 Female Spain 0  
## 3 -5.1688401 1.26222368 3 4579 Female UK 0  
## 4 1.7344571 0.29459992 3 1301 Female France 0  
## 5 3.6312287 1.90466072 4 2574 Male Canada 1  
## 6 14.5306482 -1.60253887 3 1722 Female UK 1

ff\_glimpse(reg\_data)

## $Continuous  
## label var\_type n missing\_n missing\_percent mean sd min  
## y y <dbl> 1000 0 0.0 4.9 4.9 -13.6  
## x1 x1 <dbl> 1000 0 0.0 0.1 1.0 -3.0  
## x2 x2 <int> 1000 0 0.0 3.0 1.7 0.0  
## x3 x3 <int> 1000 0 0.0 5102.9 2881.8 7.0  
## quartile\_25 median quartile\_75 max  
## y 1.8 4.7 8.2 19.3  
## x1 -0.6 0.1 0.8 3.3  
## x2 2.0 3.0 4.0 10.0  
## x3 2559.8 5171.5 7515.5 9982.0  
##   
## $Categorical  
## label var\_type n missing\_n missing\_percent levels\_n  
## x4 x4 <fct> 1000 0 0.0 2  
## x5 x5 <fct> 1000 0 0.0 5  
## Disease Disease <fct> 1000 0 0.0 2  
## levels levels\_count  
## x4 "Female", "Male" 542, 458  
## x5 "Canada", "France", "Spain", "UK", "USA" 194, 210, 201, 217, 178  
## Disease "0", "1" 533, 467  
## levels\_percent  
## x4 54, 46  
## x5 19, 21, 20, 22, 18  
## Disease 53, 47

explanatory = c("x1","x2","x3","x4","x5")  
dependent = 'y'  
t4<- reg\_data %>%  
 finalfit(dependent, explanatory)  
  
knitr::kable(t4, row.names=FALSE, align=c("l", "l", "r", "r", "r", "r"))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent: y |  | unit | value | Coefficient (univariable) | Coefficient (multivariable) |
| x1 | [-3.0,3.3] | Mean (sd) | 4.9 (4.9) | 0.13 (-0.18 to 0.43, p=0.412) | 0.11 (-0.19 to 0.42, p=0.475) |
| x2 | [0.0,10.0] | Mean (sd) | 4.9 (4.9) | -0.01 (-0.18 to 0.17, p=0.913) | -0.02 (-0.20 to 0.16, p=0.812) |
| x3 | [7.0,9982.0] | Mean (sd) | 4.9 (4.9) | 0.00 (-0.00 to 0.00, p=0.563) | 0.00 (-0.00 to 0.00, p=0.546) |
| x4 | Female | Mean (sd) | 4.8 (5.1) | - | - |
|  | Male | Mean (sd) | 5.2 (4.7) | 0.39 (-0.22 to 1.00, p=0.211) | 0.41 (-0.20 to 1.03, p=0.186) |
| x5 | Canada | Mean (sd) | 5.3 (4.8) | - | - |
|  | France | Mean (sd) | 4.3 (5.1) | -1.07 (-2.03 to -0.11, p=0.028) | -1.07 (-2.04 to -0.11, p=0.029) |
|  | Spain | Mean (sd) | 5.0 (5.2) | -0.30 (-1.27 to 0.67, p=0.545) | -0.27 (-1.24 to 0.71, p=0.590) |
|  | UK | Mean (sd) | 5.1 (4.7) | -0.25 (-1.21 to 0.70, p=0.601) | -0.22 (-1.17 to 0.74, p=0.657) |
|  | USA | Mean (sd) | 5.1 (4.8) | -0.26 (-1.26 to 0.74, p=0.606) | -0.23 (-1.24 to 0.78, p=0.652) |

# here we have a table for the Linear Regression model. the first column has the explanatory variable names.  
# 2nd column has the range(For continuous variable) & levels(for categorical variables).  
# 4th column has the mean and sd. 5th column contains the uni-variate regression coefficients, 95% CI and P-value.  
# the last column has the multivariate regression coefficients, 95%CI & P-value  
  
  
# Select any column of interest  
knitr::kable(t4[,-c(6)], row.names=FALSE, align=c("l", "l", "r", "r", "r")) # drop the univariate regression coefficient column

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: y |  | unit | value | Coefficient (univariable) |
| x1 | [-3.0,3.3] | Mean (sd) | 4.9 (4.9) | 0.13 (-0.18 to 0.43, p=0.412) |
| x2 | [0.0,10.0] | Mean (sd) | 4.9 (4.9) | -0.01 (-0.18 to 0.17, p=0.913) |
| x3 | [7.0,9982.0] | Mean (sd) | 4.9 (4.9) | 0.00 (-0.00 to 0.00, p=0.563) |
| x4 | Female | Mean (sd) | 4.8 (5.1) | - |
|  | Male | Mean (sd) | 5.2 (4.7) | 0.39 (-0.22 to 1.00, p=0.211) |
| x5 | Canada | Mean (sd) | 5.3 (4.8) | - |
|  | France | Mean (sd) | 4.3 (5.1) | -1.07 (-2.03 to -0.11, p=0.028) |
|  | Spain | Mean (sd) | 5.0 (5.2) | -0.30 (-1.27 to 0.67, p=0.545) |
|  | UK | Mean (sd) | 5.1 (4.7) | -0.25 (-1.21 to 0.70, p=0.601) |
|  | USA | Mean (sd) | 5.1 (4.8) | -0.26 (-1.26 to 0.74, p=0.606) |

# N.B: first you'll need to fix your model in usual way.  
  
# Linear regression using gtsummary  
reg <- lm(y~x1+x2+x3+x4+x5, data = reg\_data)  
summary(reg)

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = reg\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.3460 -3.2118 -0.2577 3.3272 14.6840   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.017e+00 5.653e-01 8.876 <2e-16 \*\*\*  
## x1 1.113e-01 1.559e-01 0.714 0.4754   
## x2 -2.141e-02 8.999e-02 -0.238 0.8120   
## x3 3.275e-05 5.418e-05 0.604 0.5457   
## x4Male 4.142e-01 3.132e-01 1.322 0.1863   
## x5France -1.073e+00 4.907e-01 -2.188 0.0289 \*   
## x5Spain -2.676e-01 4.960e-01 -0.539 0.5897   
## x5UK -2.166e-01 4.872e-01 -0.444 0.6568   
## x5USA -2.319e-01 5.143e-01 -0.451 0.6522   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.914 on 991 degrees of freedom  
## Multiple R-squared: 0.008374, Adjusted R-squared: 0.0003693   
## F-statistic: 1.046 on 8 and 991 DF, p-value: 0.3992

reg %>% tbl\_regression() # Basic regression table

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| x1 | 0.11 | -0.19, 0.42 | 0.5 |
| x2 | -0.02 | -0.20, 0.16 | 0.8 |
| x3 | 0.00 | 0.00, 0.00 | 0.5 |
| x4 |  |  |  |
| Female |  |  |  |
| Male | 0.41 | -0.20, 1.0 | 0.2 |
| x5 |  |  |  |
| Canada |  |  |  |
| France | -1.1 | -2.0, -0.11 | 0.029 |
| Spain | -0.27 | -1.2, 0.71 | 0.6 |
| UK | -0.22 | -1.2, 0.74 | 0.7 |
| USA | -0.23 | -1.2, 0.78 | 0.7 |

# - Global p-values for Stage are reported - Large p-values are rounded to two decimal places  
# - P-values less than 0.10 are bold - Variable labels are bold  
# - Variable levels are italicized  
reg %>%  
 tbl\_regression(  
 pvalue\_fun = ~style\_pvalue(.x, digits = 2),  
 ) %>%   
 bold\_labels() %>%  
 italicize\_levels() %>%  
 add\_global\_p() %>% # add global p-value   
 add\_q() %>% # adjusts global p-values for multiple testing  
 bold\_p() %>% # bold p-values under a given threshold (default 0.05)  
 bold\_p(t = 0.10, q = TRUE) %>% # now bold q-values under the threshold of 0.10  
 bold\_labels()

## add\_q: Adjusting p-values with  
## `stats::p.adjust(x$table\_body$p.value, method = "fdr")`  
## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **Beta** | **95% CI** | **p-value** | **q-value** |
| **x1** | 0.11 | -0.19, 0.42 | 0.48 | 0.68 |
| **x2** | -0.02 | -0.20, 0.16 | 0.81 | 0.81 |
| **x3** | 0.00 | 0.00, 0.00 | 0.55 | 0.68 |
| **x4** |  |  | 0.19 | 0.53 |
| *Female* |  |  |  |  |
| *Male* | 0.41 | -0.20, 1.0 |  |  |
| **x5** |  |  | 0.21 | 0.53 |
| *Canada* |  |  |  |  |
| *France* | -1.1 | -2.0, -0.11 |  |  |
| *Spain* | -0.27 | -1.2, 0.71 |  |  |
| *UK* | -0.22 | -1.2, 0.74 |  |  |
| *USA* | -0.23 | -1.2, 0.78 |  |  |

# To know more about this visit: https://www.danieldsjoberg.com/gtsummary/articles/tbl\_regression.html

# Logistic regression

t4<- reg\_data %>%  
 finalfit(dependent, explanatory)  
  
knitr::kable(t4, row.names=FALSE, align=c("l", "l", "r", "r", "r", "r"))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent: y |  | unit | value | Coefficient (univariable) | Coefficient (multivariable) |
| x1 | [-3.0,3.3] | Mean (sd) | 4.9 (4.9) | 0.13 (-0.18 to 0.43, p=0.412) | 0.11 (-0.19 to 0.42, p=0.475) |
| x2 | [0.0,10.0] | Mean (sd) | 4.9 (4.9) | -0.01 (-0.18 to 0.17, p=0.913) | -0.02 (-0.20 to 0.16, p=0.812) |
| x3 | [7.0,9982.0] | Mean (sd) | 4.9 (4.9) | 0.00 (-0.00 to 0.00, p=0.563) | 0.00 (-0.00 to 0.00, p=0.546) |
| x4 | Female | Mean (sd) | 4.8 (5.1) | - | - |
|  | Male | Mean (sd) | 5.2 (4.7) | 0.39 (-0.22 to 1.00, p=0.211) | 0.41 (-0.20 to 1.03, p=0.186) |
| x5 | Canada | Mean (sd) | 5.3 (4.8) | - | - |
|  | France | Mean (sd) | 4.3 (5.1) | -1.07 (-2.03 to -0.11, p=0.028) | -1.07 (-2.04 to -0.11, p=0.029) |
|  | Spain | Mean (sd) | 5.0 (5.2) | -0.30 (-1.27 to 0.67, p=0.545) | -0.27 (-1.24 to 0.71, p=0.590) |
|  | UK | Mean (sd) | 5.1 (4.7) | -0.25 (-1.21 to 0.70, p=0.601) | -0.22 (-1.17 to 0.74, p=0.657) |
|  | USA | Mean (sd) | 5.1 (4.8) | -0.26 (-1.26 to 0.74, p=0.606) | -0.23 (-1.24 to 0.78, p=0.652) |

# here we have a table for the Linear Regression model. the first column has the explanatory variable names.  
# 2nd to 4th column is for summar table  
# 5th column contains the uni-variate regression coefficients, 95% CI and P-value.  
# the last column has the multivariate regression coefficients, 95%CI & P-value  
  
  
# Select any column of interest  
knitr::kable(t4[,-c(5)], row.names=FALSE, align=c("l", "l", "r", "r", "r")) # drop the univariate regression coefficient column

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: y |  | unit | value | Coefficient (multivariable) |
| x1 | [-3.0,3.3] | Mean (sd) | 4.9 (4.9) | 0.11 (-0.19 to 0.42, p=0.475) |
| x2 | [0.0,10.0] | Mean (sd) | 4.9 (4.9) | -0.02 (-0.20 to 0.16, p=0.812) |
| x3 | [7.0,9982.0] | Mean (sd) | 4.9 (4.9) | 0.00 (-0.00 to 0.00, p=0.546) |
| x4 | Female | Mean (sd) | 4.8 (5.1) | - |
|  | Male | Mean (sd) | 5.2 (4.7) | 0.41 (-0.20 to 1.03, p=0.186) |
| x5 | Canada | Mean (sd) | 5.3 (4.8) | - |
|  | France | Mean (sd) | 4.3 (5.1) | -1.07 (-2.04 to -0.11, p=0.029) |
|  | Spain | Mean (sd) | 5.0 (5.2) | -0.27 (-1.24 to 0.71, p=0.590) |
|  | UK | Mean (sd) | 5.1 (4.7) | -0.22 (-1.17 to 0.74, p=0.657) |
|  | USA | Mean (sd) | 5.1 (4.8) | -0.23 (-1.24 to 0.78, p=0.652) |

# N.B: first you'll need to fix your model in usual way.  
  
# Linear regression using gtsummary  
reg <- glm(Disease~x1+x2+x3+x4+x5, data = reg\_data, family = "binomial")  
summary(reg)

##   
## Call:  
## glm(formula = Disease ~ x1 + x2 + x3 + x4 + x5, family = "binomial",   
## data = reg\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2631 -1.1257 -0.9752 1.2133 1.4334   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.816e-02 2.314e-01 -0.078 0.937   
## x1 -1.390e-03 6.382e-02 -0.022 0.983   
## x2 -2.531e-02 3.694e-02 -0.685 0.493   
## x3 5.313e-06 2.220e-05 0.239 0.811   
## x4Male -2.197e-01 1.283e-01 -1.712 0.087 .  
## x5France -2.112e-01 2.023e-01 -1.044 0.296   
## x5Spain 8.042e-02 2.028e-01 0.397 0.692   
## x5UK 1.504e-01 1.991e-01 0.756 0.450   
## x5USA 1.687e-01 2.101e-01 0.803 0.422   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1381.9 on 999 degrees of freedom  
## Residual deviance: 1373.0 on 991 degrees of freedom  
## AIC: 1391  
##   
## Number of Fisher Scoring iterations: 4

reg %>% tbl\_regression(  
 exponentiate = TRUE # provides odds ratio instead of log of odds  
) # Basic regression table

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **OR** | **95% CI** | **p-value** |
| x1 | 1.00 | 0.88, 1.13 | >0.9 |
| x2 | 0.98 | 0.91, 1.05 | 0.5 |
| x3 | 1.00 | 1.00, 1.00 | 0.8 |
| x4 |  |  |  |
| Female |  |  |  |
| Male | 0.80 | 0.62, 1.03 | 0.087 |
| x5 |  |  |  |
| Canada |  |  |  |
| France | 0.81 | 0.54, 1.20 | 0.3 |
| Spain | 1.08 | 0.73, 1.61 | 0.7 |
| UK | 1.16 | 0.79, 1.72 | 0.4 |
| USA | 1.18 | 0.78, 1.79 | 0.4 |

# - Global p-values for Stage are reported - Large p-values are rounded to two decimal places  
# - P-values less than 0.10 are bold - Variable labels are bold  
# - Variable levels are italicized  
reg %>%  
 tbl\_regression(  
 exponentiate = TRUE,  
 pvalue\_fun = ~style\_pvalue(.x, digits = 2),  
 ) %>%   
 bold\_labels() %>%  
 italicize\_levels() %>%  
 add\_global\_p() %>% # add global p-value   
 add\_q() %>% # adjusts global p-values for multiple testing  
 bold\_p() %>% # bold p-values under a given threshold (default 0.05)  
 bold\_p(t = 0.10, q = TRUE) %>% # now bold q-values under the threshold of 0.10  
 bold\_labels()

## add\_q: Adjusting p-values with  
## `stats::p.adjust(x$table\_body$p.value, method = "fdr")`  
## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **OR** | **95% CI** | **p-value** | **q-value** |
| **x1** | 1.00 | 0.88, 1.13 | 0.98 | 0.98 |
| **x2** | 0.98 | 0.91, 1.05 | 0.49 | 0.82 |
| **x3** | 1.00 | 1.00, 1.00 | 0.81 | 0.98 |
| **x4** |  |  | 0.087 | 0.43 |
| *Female* |  |  |  |  |
| *Male* | 0.80 | 0.62, 1.03 |  |  |
| **x5** |  |  | 0.31 | 0.79 |
| *Canada* |  |  |  |  |
| *France* | 0.81 | 0.54, 1.20 |  |  |
| *Spain* | 1.08 | 0.73, 1.61 |  |  |
| *UK* | 1.16 | 0.79, 1.72 |  |  |
| *USA* | 1.18 | 0.78, 1.79 |  |  |