R packages and Descriptive Statistics

Contents

Installing Packages

You might want to get descriptive stats or frequencies for specific variables. There are `base R` functions, but I like to use the package `mosaic`. You can download a cheat sheet for `mosaic` by clicking here.

First we need to install the `mosaic` package using the `install.packages()` function. The package name goes inside of the parentheses in double quotes: "mosaic". This is something we do only once in the console, you wouldn't want to save it in your .Rmd file, but if you do, be sure to comment it out with a `#` like I have here.

```
#install.packages("mosaic")
```

Once a package is installed, any time we start a new R session and we want to use functions inside of that package, we will need to load the package with the `library()` function.

```
library(mosaic)

acitelli <- read.csv("acitelli.csv")
```

Basic Descriptive Statistics with `mosaic`

The function `favstats()` will give descriptive statistics for a numerical variable, and the function `tally()` will give you frequencies for a categorical variable (or a numerical variable...if you want it). Functions in `mosaic` use the formula syntax, where `y ~ x`, or for a single variable, `~x`. The `~` key can be found just below your `esc` key. The first argument is the formula, and the second argument is the data frame, e.g., `data = acitelli`.

```
favstats(~satisfaction, data = acitelli)
```

```
##       min       Q1   median Q3 max    mean        sd   n missing
##  1.166667 3.333333 3.833333  4   4 3.60473 0.4964205 296       0
```

```
tally(~gender, data = acitelli)
```

```
## gender
##  -1   1
## 148 148
```

```
#tally() can also give you percentages with the format argument
tally(~gender, data = acitelli, format = "percent")
```

```
## gender
## -1  1
## 50 50
```

You can look up more information about a function with the `?`.

```
?favstats
```

Descriptives split by gender.

```
favstats(satisfaction ~ gender, data = acitelli)
```

```
##   gender      min       Q1   median Q3 max     mean        sd   n missing
## 1     -1 1.500000 3.333333 3.833333  4   4 3.591216 0.5300260 148       0
## 2      1 1.166667 3.500000 3.833333  4   4 3.618243 0.4617875 148       0
```

What are the standard deviations of perceived `tension` split by gender?

What is the minimum of the `self_pos` variable?

The `mosaic` package also has a function for getting the correlation coefficient, it's called `cor()`. Using the same format (i.e., formula then data), how do you think would you get the correlation of satisfaction and tension?

```
cor(satisfaction~tension, data = acitelli)
```

```
## [1] -0.5971907
```

For bivariate correlation matrices, I use the `corr.test()` function from the `psych` package.

```
library(psych)

corr.test(acitelli)
```

```
## Call:corr.test(x = acitelli)
## Correlation matrix
##              cuplid Yearsmar gender self_pos other_pos satisfaction
## cuplid         1.00    -0.13   0.00     0.03     -0.13        -0.16
## Yearsmar      -0.13     1.00   0.00     0.07      0.13        -0.01
## gender         0.00     0.00   1.00    -0.25      0.04         0.03
## self_pos       0.03     0.07  -0.25     1.00      0.24         0.18
## other_pos     -0.13     0.13   0.04     0.24      1.00         0.47
```

```
## satisfaction  -0.16    -0.01   0.03     0.18      0.47         1.00
## tension        0.17    -0.11  -0.13    -0.10     -0.37        -0.60
## simhob        -0.02    -0.09  -0.17     0.06      0.18         0.29
##              tension simhob
## cuplid          0.17  -0.02
## Yearsmar       -0.11  -0.09
## gender         -0.13  -0.17
## self_pos       -0.10   0.06
## other_pos      -0.37   0.18
## satisfaction   -0.60   0.29
## tension         1.00  -0.12
## simhob         -0.12   1.00
## Sample Size
## [1] 296
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##              cuplid Yearsmar gender self_pos other_pos satisfaction
## cuplid         0.00     0.40   1.00     1.00      0.40         0.08
## Yearsmar       0.03     0.00   1.00     1.00      0.39         1.00
## gender         1.00     1.00   0.00     0.00      1.00         1.00
## self_pos       0.56     0.21   0.00     0.00      0.00         0.04
## other_pos      0.02     0.02   0.55     0.00      0.00         0.00
## satisfaction   0.00     0.92   0.64     0.00      0.00         0.00
## tension        0.00     0.06   0.02     0.07      0.00         0.00
## simhob         0.79     0.11   0.00     0.30      0.00         0.00
##              tension simhob
## cuplid          0.07   1.00
## Yearsmar        0.67   1.00
## gender          0.40   0.06
## self_pos        0.82   1.00
## other_pos       0.00   0.04
## satisfaction    0.00   0.00
## tension         0.00   0.48
## simhob          0.04   0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

We can also subset the `acitelli` dataset to get the correlation matrix split by `gender` using the brackets, but this looks like garbage!

```
#correlations for men
corr.test(acitelli[which(acitelli$gender==1),c(-3)])$r
```

```
##                 cuplid    Yearsmar     self_pos  other_pos satisfaction
## cuplid       1.00000000 -0.12821170  0.064763248 -0.1120233  -0.15068767
## Yearsmar    -0.12821170  1.00000000  0.157866700  0.1697970  -0.01488433
## self_pos     0.06476325  0.15786670  1.000000000  0.1886611   0.08142680
```

```
## other_pos    -0.11202325  0.16979696  0.188661064  1.0000000   0.50496072
## satisfaction -0.15068767 -0.01488433  0.081426804  0.5049607   1.00000000
## tension       0.15613894 -0.16045792 -0.009658491 -0.3418260  -0.56745581
## simhob       -0.04381344 -0.10357773  0.005403892  0.1595127   0.26747767
##                    tension       simhob
## cuplid        0.156138943 -0.043813443
## Yearsmar     -0.160457922 -0.103577727
## self_pos     -0.009658491  0.005403892
## other_pos    -0.341825957  0.159512724
## satisfaction -0.567455810  0.267477666
## tension       1.000000000 -0.065196943
## simhob       -0.065196943  1.000000000
```

```
#correlations for women
corr.test(acitelli[which(acitelli$gender==-1),c(-3)])$r
```

```
##                    cuplid     Yearsmar     self_pos   other_pos
## cuplid        1.000000000 -0.128211698  0.007610632 -0.14778490
## Yearsmar     -0.128211698  1.000000000 -0.004238569  0.09844142
## self_pos      0.007610632 -0.004238569  1.000000000  0.32661940
## other_pos    -0.147784898  0.098441418  0.326619400  1.00000000
## satisfaction -0.177837708  0.001917215  0.291161501  0.43982125
## tension       0.184257376 -0.067582842 -0.259670893 -0.38786934
## simhob        0.016215251 -0.085762710  0.030159790  0.21967179
##              satisfaction     tension      simhob
## cuplid       -0.177837708  0.18425738  0.01621525
## Yearsmar      0.001917215 -0.06758284 -0.08576271
## self_pos      0.291161501 -0.25967089  0.03015979
## other_pos     0.439821249 -0.38786934  0.21967179
## satisfaction  1.000000000 -0.62477474  0.33026923
## tension      -0.624774745  1.00000000 -0.23792175
## simhob        0.330269230 -0.23792175  1.00000000
```

Don't worry, the `tidyverse` to the rescue.