campR – summary tables

Marcus Vollmer

2019-06-06

Contents

Summary tables **1**
Libraries and data . . . 1
xtable . . . 2
stargazer . . . 5
Own summary functions . . . 10
strtable and stargazer_long . . . 10
mytable . . . 11
characteristics_table . . . 12
Rapport . . . 17

This is a campR notebook introduces summary tables with 'xtable', 'stargazer', 'rapport' and show how to use own helper functions for group comparisons.

Rapport needs 'Pandoc' which is not available in all R versions. After that run: 'install_github('rapport', 'rapporter')'.

Summary tables

Libraries and data

The following commands imports the libraries of **xtable**, **stargazer** and **rapport**. Further, we import some user defined function from GitHub. The latter package makes the data frame *movies* available a collection of motion pictures with user ratings frome the internet movie database (IMDB).

```
library(xtable)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(rapport)

# Load user created functions
library(devtools)
repo = "https://raw.githubusercontent.com/MarcusVollmer/R-Helper-Functions/master/"
source_url(paste0(repo, "strtable.R"))
```

```
## SHA-1 hash of file is 6309796f32309527278d63789fae2e3cc2229d99
```

```
source_url(paste0(repo, "stargazer_long.R"))
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

```
source_url(paste0(repo, "characteristics_table.R"))
```

```
## SHA-1 hash of file is 871b5963575aaf03f2b9965d0bd4da15543b531e
```

```
source_url(paste0(repo, "mytable.R"))
```

```
## SHA-1 hash of file is 4f787c370d850c7771c3dd5762291edd9da27f71
```

```
library(ggplot2movies)
```

xtable

Lets have short look what kind of data xtable can process:

```
methods(xtable)
```

```
##  [1] xtable.anova*            xtable.aov*
##  [3] xtable.aovlist*          xtable.coxph*
##  [5] xtable.data.frame*       xtable.glm*
##  [7] xtable.gmsar*            xtable.lagImpact*
##  [9] xtable.lm*               xtable.matrix*
## [11] xtable.prcomp*           xtable.sarlm*
## [13] xtable.sarlm.pred*       xtable.spautolm*
## [15] xtable.sphet*            xtable.splm*
## [17] xtable.stsls*            xtable.summary.aov*
## [19] xtable.summary.aovlist*  xtable.summary.glm*
## [21] xtable.summary.gmsar*    xtable.summary.lm*
## [23] xtable.summary.prcomp*   xtable.summary.sarlm*
## [25] xtable.summary.spautolm* xtable.summary.sphet*
## [27] xtable.summary.splm*     xtable.summary.stsls*
## [29] xtable.table*            xtable.ts*
## [31] xtable.zoo*
## see '?methods' for accessing help and source code
```

The xtable documentation summarizes the programming syntax and lists available functions.

We will plot the head of the data and will display the summary of a linear model in the default R:markdown fashion.

```
# Lets quickly check the movie data base
(h = head(movies))
```

```
##                       title year length budget rating votes   r1   r2  r3
## 1                         $ 1971    121     NA    6.4   348  4.5  4.5 4.5
## 2        $1000 a Touchdown 1939     71     NA    6.0    20  0.0 14.5 4.5
## 3   $21 a Day Once a Month 1941      7     NA    8.2     5  0.0  0.0 0.0
## 4                  $40,000 1996     70     NA    8.2     6 14.5  0.0 0.0
## 5 $50,000 Climax Show, The 1975     71     NA    3.4    17 24.5  4.5 0.0
## 6                     $pent 2000     91     NA    4.3    45  4.5  4.5 4.5
##     r4   r5   r6   r7   r8   r9  r10 mpaa Action Animation Comedy Drama
## 1  4.5 14.5 24.5 24.5 14.5  4.5  4.5           0         0      1     1
## 2 24.5 14.5 14.5 14.5  4.5  4.5 14.5           0         0      1     0
## 3  0.0  0.0 24.5  0.0 44.5 24.5 24.5           0         1      0     0
## 4  0.0  0.0  0.0  0.0  0.0 34.5 45.5           0         0      1     0
## 5 14.5 14.5  4.5  0.0  0.0  0.0 24.5           0         0      0     0
## 6 14.5 14.5 14.5  4.5  4.5 14.5 14.5           0         0      0     1
##   Documentary Romance Short
## 1           0       0     0
## 2           0       0     0
## 3           0       0     1
## 4           0       0     0
## 5           0       0     0
## 6           0       0     0
```

```
movies$genre = factor(with(movies, paste(Action, Comedy, Drama, Documentary,
    Romance, Short, sep = "")), c("100000", "010000", "001000", "000100", "000010",
    "000001"), c("Action", "Comedy", "Drama", "Documentary", "Romance", "Short"))
```

```
# Build a generalized linear model to explain the budget
mymodel = lm(budget ~ length + rating + year + genre, movies)
(s = summary(mymodel))
```

```
##
## Call:
## lm(formula = budget ~ length + rating + year + genre, data = movies)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -63233855 -10859681  -3929875   4962836 157885203
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -570902777   37296310 -15.307  < 2e-16 ***
## length              258081      16863  15.305  < 2e-16 ***
## rating              167936     276357   0.608    0.543
## year                288880      18693  15.454  < 2e-16 ***
## genreComedy       -14565943    1298609 -11.217  < 2e-16 ***
## genreDrama        -20369294    1240837 -16.416  < 2e-16 ***
## genreDocumentary  -27896161    2454030 -11.367  < 2e-16 ***
## genreRomance      -16120791    3200503  -5.037 5.06e-07 ***
## genreShort        -11885684    2517163  -4.722 2.46e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19630000 on 2588 degrees of freedom
##   (56191 observations deleted due to missingness)
## Multiple R-squared:  0.2816, Adjusted R-squared:  0.2793
## F-statistic: 126.8 on 8 and 2588 DF,  p-value: < 2.2e-16
```

Next, we will plot the results with **xtable** as LaTeXand HTML output.

```
# The basic LaTeX output
xtable(h)
```

% latex table generated in R 3.4.4 by xtable 1.8-3 package % Wed Jun 5 17:15:32 2019

	title	year	length	budget	rating	votes	r1	r2	r3	r4	r5	r6
1	$	1971	121		6.40	348	4.50	4.50	4.50	4.50	14.50	24.50
2	$1000 a Touchdown	1939	71		6.00	20	0.00	14.50	4.50	24.50	14.50	14.50
3	$21 a Day Once a Month	1941	7		8.20	5	0.00	0.00	0.00	0.00	0.00	24.50
4	$40,000	1996	70		8.20	6	14.50	0.00	0.00	0.00	0.00	0.00
5	$50,000 Climax Show, The	1975	71		3.40	17	24.50	4.50	0.00	14.50	14.50	4.50
6	$pent	2000	91		4.30	45	4.50	4.50	4.50	14.50	14.50	14.50

```
xtable(mymodel)
```

% latex table generated in R 3.4.4 by xtable 1.8-3 package % Wed Jun 5 17:15:33 2019

If you need to change some properties use **options**:

```
# The basic LaTeX output
xtable(h)
```

% latex table generated in R 3.4.4 by xtable 1.8-3 package %

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-570902776.9225	37296310.3349	-15.31	0.0000
length	258081.1427	16862.6383	15.30	0.0000
rating	167935.4543	276357.2885	0.61	0.5435
year	288879.5304	18692.7647	15.45	0.0000
genreComedy	-14565942.9648	1298608.8297	-11.22	0.0000
genreDrama	-20369293.4501	1240837.0406	-16.42	0.0000
genreDocumentary	-27896160.7441	2454030.4535	-11.37	0.0000
genreRomance	-16120790.9017	3200502.7152	-5.04	0.0000
genreShort	-11885684.2953	2517162.9568	-4.72	0.0000

title	year	length	budget	rating	votes	r1	r2	r3	r4	r5	r6	r7	r8
$	1971	121		6.40	348	4.50	4.50	4.50	4.50	14.50	24.50	24.50	14.50
$1000 a Touchdown	1939	71		6.00	20	0.00	14.50	4.50	24.50	14.50	14.50	14.50	4.50
$21 a Day Once a Month	1941	7		8.20	5	0.00	0.00	0.00	0.00	0.00	24.50	0.00	44.50
$40,000	1996	70		8.20	6	14.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$50,000 Climax Show, The	1975	71		3.40	17	24.50	4.50	0.00	14.50	14.50	4.50	0.00	0.00
$pent	2000	91		4.30	45	4.50	4.50	4.50	14.50	14.50	14.50	4.50	4.50

```
xtable(mymodel)
```

% latex table generated in R 3.4.4 by xtable 1.8-3 package %

Estimate	Std. Error	t value	Pr(>|t|)
-570902776.9225	37296310.3349	-15.31	0.0000
258081.1427	16862.6383	15.30	0.0000
167935.4543	276357.2885	0.61	0.5435
288879.5304	18692.7647	15.45	0.0000
-14565942.9648	1298608.8297	-11.22	0.0000
-20369293.4501	1240837.0406	-16.42	0.0000
-27896160.7441	2454030.4535	-11.37	0.0000
-16120790.9017	3200502.7152	-5.04	0.0000
-11885684.2953	2517162.9568	-4.72	0.0000

stargazer

Stargazer produces LaTeX code, HTML/CSS code and ASCII text for well-formatted tables.

```
# The basic LaTeX output
stargazer(h)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mi, Jun 05, 2019 - 17:15:33

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
year	6	1,970.333	26.090	1,939	1,948.5	1,990.8	2,000
length	6	71.833	37.397	7	70.2	86	121
rating	6	6.083	1.972	3.400	4.725	7.750	8.200
votes	6	73.500	135.252	5	8.8	38.8	348
r1	6	8.000	9.664	0.000	1.125	12.000	24.500
r2	6	4.667	5.298	0.000	1.125	4.500	14.500
r3	6	2.250	2.465	0.000	0.000	4.500	4.500
r4	6	9.667	9.801	0.000	1.125	14.500	24.500
r5	6	9.667	7.488	0.000	3.625	14.500	14.500
r6	6	13.750	10.069	0.000	7.000	22.000	24.500
r7	6	7.250	10.153	0.000	0.000	12.000	24.500
r8	6	11.333	17.090	0.000	1.125	12.000	44.500
r9	6	13.750	13.468	0.000	4.500	22.000	34.500
r10	6	21.333	14.006	4.500	14.500	24.500	45.500
Action	6	0.000	0.000	0	0	0	0
Animation	6	0.167	0.408	0	0	0	1
Comedy	6	0.500	0.548	0	0	1	1
Drama	6	0.333	0.516	0	0	0.8	1
Documentary	6	0.000	0.000	0	0	0	0
Romance	6	0.000	0.000	0	0	0	0
Short	6	0.167	0.408	0	0	0	1

```
stargazer(mymodel)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mi, Jun 05, 2019 - 17:15:37

```
#
stargazer(h, summary = FALSE)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mi, Jun 05, 2019 - 17:15:37

Lets have short look at the supported objects:

```
`?`(`stargazer models`)
```

Stargazer in action to display results of linear models:

```
# Lets model the budget of movies with and without the interaction of movie
# length and genre
m1 = lm(budget ~ length + genre + rating + year, movies)
m2 = lm(budget ~ length * genre + rating + year, movies)
```

Table 2:

	Dependent variable:
	budget
length	258,081.100***
	(16,862.640)
rating	167,935.500
	(276,357.300)
year	288,879.500***
	(18,692.760)
genreComedy	−14,565,943.000***
	(1,298,609.000)
genreDrama	−20,369,293.000***
	(1,240,837.000)
genreDocumentary	−27,896,161.000***
	(2,454,030.000)
genreRomance	−16,120,791.000***
	(3,200,503.000)
genreShort	−11,885,684.000***
	(2,517,163.000)
Constant	−570,902,777.000***
	(37,296,310.000)
Observations	2,597
R^2	0.282
Adjusted R^2	0.279
Residual Std. Error	19,625,079.000 (df = 2588)
F Statistic	126.782*** (df = 8; 2588)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3:

	title	year	length	budget	rating	votes	r1	r2	r3	r4
1	$	1,971	121		6.400	348	4.500	4.500	4.500	4.500
2	$1000 a Touchdown	1,939	71		6	20	0	14.500	4.500	24.500
3	$21 a Day Once a Month	1,941	7		8.200	5	0	0	0	0
4	$40,000	1,996	70		8.200	6	14.500	0	0	0
5	$50,000 Climax Show, The	1,975	71		3.400	17	24.500	4.500	0	14.500
6	$pent	2,000	91		4.300	45	4.500	4.500	4.500	14.500

```r
# Stargazer has a nice combined representation of multiple models
stargazer(m1, m2, title = "Regression results")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mi, Jun 05, 2019 - 17:15:37

```r
# You have various properties you can change according to your need
stargazer(m1, m2, title = "Regression results", single.row = TRUE, ci = TRUE,
    ci.level = 0.95, decimal.mark = ".", digits = 0)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mi, Jun 05, 2019 - 17:15:37

Or you can use the html output:

Table 4: Regression results

	Dependent variable:	
	budget	
	(1)	(2)
length	258,081.100***	710,893.200***
	(16,862.640)	(45,936.600)
genreComedy	−14,565,943.000***	38,161,577.000***
	(1,298,609.000)	(7,052,174.000)
genreDrama	−20,369,293.000***	32,697,800.000***
	(1,240,837.000)	(5,437,043.000)
genreDocumentary	−27,896,161.000***	41,149,598.000***
	(2,454,030.000)	(8,454,675.000)
genreRomance	−16,120,791.000***	53,521,988.000***
	(3,200,503.000)	(13,686,829.000)
genreShort	−11,885,684.000***	41,595,682.000***
	(2,517,163.000)	(5,766,692.000)
rating	167,935.500	−121,085.800
	(276,357.300)	(272,222.000)
year	288,879.500***	293,678.600***
	(18,692.760)	(18,314.040)
length:genreComedy		−500,094.700***
		(69,912.850)
length:genreDrama		−493,844.000***
		(49,261.050)
length:genreDocumentary		−695,810.900***
		(89,157.600)
length:genreRomance		−667,970.200***
		(135,037.900)
length:genreShort		−718,418.100***
		(131,061.600)
Constant	−570,902,777.000***	−627,081,945.000***
	(37,296,310.000)	(36,977,034.000)
Observations	2,597	2,597
R^2	0.282	0.314
Adjusted R^2	0.279	0.310
Residual Std. Error	19,625,079.000 (df = 2588)	19,196,315.000 (df = 2583)
F Statistic	126.782*** (df = 8; 2588)	90.921*** (df = 13; 2583)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5: Regression results

	Dependent variable:	
	budget	
	(1)	(2)
length	258,081*** (225,031, 291,131)	710,893*** (620,859, 800,927)
genreComedy	−14,565,943*** (−17,111,170, −12,020,716)	38,161,577*** (24,339,569, 51,983,585)
genreDrama	−20,369,293*** (−22,801,289, −17,937,298)	32,697,800*** (22,041,391, 43,354,209)
genreDocumentary	−27,896,161*** (−32,705,972, −23,086,349)	41,149,598*** (24,578,740, 57,720,456)
genreRomance	−16,120,791*** (−22,393,661, −9,847,921)	53,521,988*** (26,696,296, 80,347,680)
genreShort	−11,885,684*** (−16,819,233, −6,952,136)	41,595,682*** (30,293,173, 52,898,192)
rating	167,935 (−373,715, 709,586)	−121,086 (−654,631, 412,460)
year	288,880*** (252,242, 325,517)	293,679*** (257,784, 329,573)
length:genreComedy		−500,095*** (−637,121, −363,068)
length:genreDrama		−493,844*** (−590,394, −397,294)
length:genreDocumentary		−695,811*** (−870,557, −521,065)
length:genreRomance		−667,970*** (−932,640, −403,301)
length:genreShort		−718,418*** (−975,294, −461,542)
Constant	−570,902,777*** (−644,002,202, −497,803,352)	−627,081,945*** (−699,555,600, −554,608,2
Observations	2,597	2,597
R^2	0	0
Adjusted R^2	0	0
Residual Std. Error	19,625,079 (df = 2588)	19,196,315 (df = 2583)
F Statistic	127*** (df = 8; 2588)	91*** (df = 13; 2583)
Note:		*p<0.1; **p<0.05; ***p<0

Own summary functions

https://github.com/MarcusVollmer/R-Helper-Functions can be sourced directly from GitHub to fasten and to standardize the analysis and to allow reproducability.

Currently there are four useful functions provided:

1. strtable.R – a modified str implementation for data frames (originally published by Jason Bryer)
2. stargazer_long.R – a modification of stargazer to display long tables, to allow rotations and LaTeX column specifications
3. mytable.R – a cross-table for factor variables to display counts or summary statistics of a continuous variable with univariable statistical testing
4. characteristics_table.R – creates stratified tables of data along with univariable statistical testing

strtable and stargazer_long

```
# Some genre columns should be set to logicals
booleanvars = colnames(movies[, 18:24])
movies[, booleanvars] = lapply(booleanvars, function(x) as.logical(movies[,
    x]))

# Generate a summary table using strtable
s = strtable(movies, n = 10, width = 300)

# For LaTeX output as a longtable use the follwing command.  The compiling
# of the LaTeX script to PDF need the inclusion of the **longtable** in the
# header.
stargazer_long(s[, 1:4], summary = FALSE, rownames = FALSE, output = "cccp{9cm}",
    rotate = 60)
```

variable	NAs	class	stats
title	0	character	
year	0	integer	m=1976.13, sd=23.74, q=[1893.00, 1958.00, 1983.00, 1997.00, 2005.00]
length	0	integer	m=82.34, sd=44.35, q=[1.00, 74.00, 90.00, 100.00, 5220.00]
budget	53,573	integer	m=13412513.25, sd=23350084.93, q=[0.00, 250000.00, 3000000.00, 15000000.00, 200000000.00]
rating	0	numeric	m=5.93, sd=1.55, q=[1.00, 5.00, 6.10, 7.00, 10.00]
votes	0	integer	m=632.13, sd=3829.62, q=[5.00, 11.00, 30.00, 112.00, 157608.00]
r1	0	numeric	m=7.01, sd=10.94, q=[0.00, 0.00, 4.50, 4.50, 100.00]
r2	0	numeric	m=4.02, sd=5.96, q=[0.00, 0.00, 4.50, 4.50, 84.50]
r3	0	numeric	m=4.72, sd=6.45, q=[0.00, 0.00, 4.50, 4.50, 84.50]
r4	0	numeric	m=6.37, sd=7.59, q=[0.00, 0.00, 4.50, 4.50, 100.00]
r5	0	numeric	m=9.80, sd=9.73, q=[0.00, 4.50, 4.50, 14.50, 100.00]
r6	0	numeric	m=13.04, sd=10.98, q=[0.00, 4.50, 14.50, 14.50, 84.50]
r7	0	numeric	m=15.55, sd=11.59, q=[0.00, 4.50, 14.50, 24.50, 100.00]
r8	0	numeric	m=13.88, sd=11.32, q=[0.00, 4.50, 14.50, 24.50, 100.00]
r9	0	numeric	m=8.95, sd=9.44, q=[0.00, 4.50, 4.50, 14.50, 100.00]
r10	0	numeric	m=16.85, sd=15.65, q=[0.00, 4.50, 14.50, 24.50, 100.00]
mpaa	0	character	

Action	0	logical	"FALSE" (54100), "TRUE" (4688)
Animation	0	logical	"FALSE" (55098), "TRUE" (3690)
Comedy	0	logical	"FALSE" (41517), "TRUE" (17271)
Drama	0	logical	"FALSE" (36977), "TRUE" (21811)
Documentary	0	logical	"FALSE" (55316), "TRUE" (3472)
Romance	0	logical	"FALSE" (54044), "TRUE" (4744)
Short	0	logical	"FALSE" (49330), "TRUE" (9458)
genre	27,488	Factor w/ 6 levels	"Action" (2075), "Comedy" (8350), "Drama" (14270), "Documentary" (2392), "Romance" (546), "Short" (3667)

mytable

Cross table for count data with confidence intervals

Printing a cross table with p value and percentages (rows sums up to 100%).

```
library(PropCIs)
mytable(movies$Animation, movies$Action, ci = FALSE, prec = "%.2f", latex = TRUE)
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

	FALSE	TRUE
FALSE	50494 (91.64)	4604 (8.36)
TRUE	3606 (97.72)	84 (2.28)

Overall: Fisher's exact test: p-Value=2.04e-52

Printing the same table with 95% exact Clopper-Pearson confidence intervals (PropCIs package required) and less digits with % sign, a fixed column width and floating point p-value style:

```
mytable(movies$Animation, movies$Action, ci = 0.95, prec = "%.1f", prec_p = "%.4f",
    latex = "rp{4cm}p{4cm}", pct_sign = "%")
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

	FALSE	TRUE	95 CI
FALSE	50494 (91.6%)	4604 (8.4%)	8.1 to 8.6
TRUE	3606 (97.7%)	84 (2.3%)	1.8 to 2.8

Overall: Fisher's exact test: p-Value=0.0000

Cross table for a continuous variable with statistical testing

For group-wise testing of a continuous variable pass a third variable to mytable, e.g. comparing the movie rating in relation to the genres:

```
mytable(movies$Animation, movies$Action, movies$budget, latex = TRUE)
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

	FALSE	TRUE	p
FALSE	2000000.0 (200000.0,11000000.0)	16000000.0 (2500000.0,50000000.0)	1.59e-84

TRUE	18000000.0 (1750000.0,60000000.0)	14500000.0 (3000000.0,77500000.0)	8.17e-01

Line by line: med (q_{25},q_{75}), Wilcoxon rank sum test
Overall: Kruskal-Wallis rank sum test: p-Value=2.11e-89

```
# Change the precision of the numbers and the p-value by adding the
# properties and setting the scientific format with digits:
mytable(movies$Animation, movies$Action, movies$budget, prec = "%.2e", prec_p = "%.2e",
    latex = TRUE)
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

	FALSE	TRUE	p
FALSE	2.00e+06 (2.00e+05,1.10e+07)	1.60e+07 (2.50e+06,5.00e+07)	1.59e-84
TRUE	1.80e+07 (1.75e+06,6.00e+07)	1.45e+07 (3.00e+06,7.75e+07)	8.17e-01

Line by line: med (q_{25},q_{75}), Wilcoxon rank sum test
Overall: Kruskal-Wallis rank sum test: p-Value=2.11e-89

```
mytable(movies$Action, movies$length < 60, movies$budget, prec = "%.2e", prec_p = "%.2e",
    latex = TRUE)
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

	FALSE	TRUE	p
FALSE	3.00e+06 (5.00e+05,1.40e+07)	1.00e+04 (3.81e+03,3.05e+04)	1.04e-219
TRUE	1.80e+07 (3.43e+06,5.20e+07)	5.00e+03 (2.45e+03,2.50e+04)	2.13e-18

Line by line: med (q_{25},q_{75}), Wilcoxon rank sum test
Overall: Kruskal-Wallis rank sum test: p-Value=3.46e-299

characteristics_table

This is a helpful function to compute and display a characteristics table of the study population with separate columns for each cohort, e.g. treatment vs. control. In our example data set we stratify the movie database by `Documentary` which is a binary variable, indicating whether the movie is a documentary or not. characteristics_table.R will do the job of printing characteristics seperated by a binary or factorial response (genre). P-values are the results of statistical testing comparing both groups: T test and Wilcoxon ranksum test for continous variables with mean and SD or median and quartiles respectively. Statistical testing with categorical data is conducted by Fishers exact test or Chi-Squared-Test (categorical with more than 3 levels). Missing values (NA's) was omitted for this analysis. Precision of numerials can be set individually for continuous values and p values. Default is one digit after decimal place `prec="%.1f"`, `prec_continuous="%.0f"` and four digits for p values `prec_p="%.4f"`.

```
# For LaTeX output as a longtable use the following command.  The compiling
# of the LaTeX script to a PDF needs the inclusion of the **longtable** in
# the header.
characteristics_table(-2, "Documentary", movies[, -1], "col", prec = "%.1f",
    prec_continuous = "%.1f", latex = "p{1.5cm}p{2cm}rrrr", tablefootnote = FALSE)
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

Variable	Level	FALSE	TRUE	P	NAs
Action	True	4672 (8.4)	16 (0.5)	0.0000	0
	False	50644 (91.6)	3456 (99.5)		
Animation	True	3647 (6.6)	43 (1.2)	0.0000	0
	False	51669 (93.4)	3429 (98.8)		
Comedy	True	17140 (31.0)	131 (3.8)	0.0000	0
	False	38176 (69.0)	3341 (96.2)		
Drama	True	21684 (39.2)	127 (3.7)	0.0000	0
	False	33632 (60.8)	3345 (96.3)		
Romance	True	4734 (8.6)	10 (0.3)	0.0000	0
	False	50582 (91.4)	3462 (99.7)		
Short	True	8591 (15.5)	867 (25.0)	0.0000	0
	False	46725 (84.5)	2605 (75.0)		
genre	Action	2075 (7.2)	0 (0.0)	0.0000	27488
	Comedy	8350 (28.9)	0 (0.0)		
	Drama	14270 (49.4)	0 (0.0)		
	Documentary	0 (0.0)	2392 (100.0)		
	Romance	546 (1.9)	0 (0.0)		
	Short	3667 (12.7)	0 (0.0)		
year	Median (Quartiles)	1982.0 (1957.0,1997.0)	1995.0 (1976.0,2001.0)	0.0000	0
year	Mean (SD)	1975.6 (23.7)	1985.3 (22.3)	0.0000	0
length	Median (Quartiles)	90.0 (75.0,101.0)	76.0 (45.0,91.0)	0.0000	0
length	Mean (SD)	83.1 (44.3)	70.4 (43.5)	0.0000	0
budget	Median (Quartiles)	3000000.0 (300000.0,16000000.0)	150000.0 (25000.0,600000.0)	0.0000	53573
budget	Mean (SD)	13724631.4 (23552882.6)	908125.2 (1999650.9)	0.0000	53573
rating	Median (Quartiles)	6.0 (5.0,6.9)	6.9 (5.8,7.7)	0.0000	0
rating	Mean (SD)	5.9 (1.5)	6.7 (1.6)	0.0000	0
votes	Median (Quartiles)	31.0 (12.0,120.0)	16.0 (8.0,39.0)	0.0000	0
votes	Mean (SD)	665.4 (3938.6)	102.6 (935.8)	0.0000	0
r1	Median (Quartiles)	4.5 (0.0,4.5)	0.0 (0.0,4.5)	0.0000	0
r1	Mean (SD)	7.1 (10.9)	6.0 (10.9)	0.0000	0
r2	Median (Quartiles)	4.5 (0.0,4.5)	0.0 (0.0,4.5)	0.0000	0
r2	Mean (SD)	4.1 (6.0)	2.6 (5.5)	0.0000	0
r3	Median (Quartiles)	4.5 (0.0,4.5)	0.0 (0.0,4.5)	0.0000	0
r3	Mean (SD)	4.8 (6.4)	3.0 (6.5)	0.0000	0
r4	Median (Quartiles)	4.5 (0.0,14.5)	0.0 (0.0,4.5)	0.0000	0
r4	Mean (SD)	6.5 (7.6)	3.9 (6.8)	0.0000	0
r5	Median (Quartiles)	4.5 (4.5,14.5)	4.5 (0.0,14.5)	0.0000	0
r5	Mean (SD)	10.0 (9.7)	6.7 (9.6)	0.0000	0
r6	Median (Quartiles)	14.5 (4.5,14.5)	4.5 (0.0,14.5)	0.0000	0
r6	Mean (SD)	13.2 (10.9)	10.1 (11.3)	0.0000	0

r7	Median (Quartiles)	14.5 (4.5,24.5)	14.5 (4.5,24.5)	0.0000	0
r7	Mean (SD)	15.6 (11.5)	14.5 (12.4)	0.0000	0
r8	Median (Quartiles)	14.5 (4.5,24.5)	14.5 (4.5,24.5)	0.0000	0
r8	Mean (SD)	13.7 (11.2)	16.5 (12.9)	0.0000	0
r9	Median (Quartiles)	4.5 (4.5,14.5)	14.5 (4.5,14.5)	0.0000	0
r9	Mean (SD)	8.7 (9.2)	12.7 (12.4)	0.0000	0
r10	Median (Quartiles)	14.5 (4.5,24.5)	24.5 (14.5,34.5)	0.0000	0
r10	Mean (SD)	16.4 (15.3)	24.7 (19.0)	0.0000	0

It also works for categorical variables with more than two levels, such as `genre`, which has six levels specified in movies with a single/unique genre assignment. Statistical testing will change to Chi-Squared-Testing, Kruskal-Wallis rank sum test and One-way analysis of variance (ANOVA). Footnotes will tell you the statistical test behind the p values.

```
# You may change the footnote labeling as follows:
characteristics_table(-2, "genre", movies[, -c(1, 7:16)], "col", prec = "%.1f",
    prec_continuous = "%.1f", prec_p = "%.3f", tablefootnote = TRUE, latex = "p{1.2cm}p{1.5cm}p{1.4cm}p
    fn = c("'", "_", "''", "_", "'''", "_"))
```

```
## SHA-1 hash of file is fc2ebd2a87d6da2f803de4c9243eaf5954836225
```

Variable	Level	Action	Comedy	Drama	Documentary	Romance	Short	P	NAs
Action	True	2075 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000’	27488
	False	0 (0.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Animation	True	35 (1.7)	113 (1.4)	35 (0.2)	8 (0.3)	9 (1.6)	943 (25.7)	0.000’	27488
	False	2040 (98.3)	8237 (98.6)	14235 (99.8)	2384 (99.7)	537 (98.4)	2724 (74.3)		
Comedy	True	0 (0.0)	8350 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000’	27488
	False	2075 (100.0)	0 (0.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Drama	True	0 (0.0)	0 (0.0)	14270 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000’	27488
	False	2075 (100.0)	8350 (100.0)	0 (0.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Documentary	True	0 (0.0)	0 (0.0)	0 (0.0)	2392 (100.0)	0 (0.0)	0 (0.0)	0.000’	27488
	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	0 (0.0)	546 (100.0)	3667 (100.0)		
Romance	True	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	546 (100.0)	0 (0.0)	0.000’	27488
	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	0 (0.0)	3667 (100.0)		
Short	True	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3667 (100.0)	0.000’	27488

	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	0 (0.0)		
year	Median (Quartiles)	1990.0 (1979.0,1997.0)	1979.0 (1959.0,1994.0)	1979.5 (1957.0,1996.0)	1997.0 (1985.0,2002.0)	1978.0 (1951.0,1996.0)	1995.0 (1966.0,2001.0)	0.000”	0
year	Mean (SD)	1985.2 (16.9)	1974.8 (21.8)	1975.3 (22.7)	1991.4 (14.7)	1972.4 (24.6)	1980.6 (27.1)	0.000”’	0
length	Median (Quartiles)	94.0 (88.0,102.0)	91.0 (84.0,99.0)	96.0 (88.0,108.0)	85.0 (70.0,95.0)	95.0 (85.0,106.0)	11.0 (7.0,20.0)	0.000”	0
length	Mean (SD)	100.0 (28.6)	91.4 (14.1)	99.0 (23.0)	87.2 (38.1)	99.1 (24.5)	14.1 (9.7)	0.000”’	0
budget	Median (Quartiles)	20000000.0 (4050000.0,50000000.0)	4000000.0 (600000.0,17000000.0)	3000000.0 (500000.0,12000000.0)	155000.0 (50000.0,550000.0)	1853132.5 (28975.5,6000000.0)	9000.0 (3000.0,28000.0)	0.000”	53573
budget	Mean (SD)	33088683.5 (36845498.0)	13657908.8 (21083219.4)	10589139.0 (17949423.6)	728864.7 (1871151.4)	6698891.1 (11852982.5)	391038.9 (3459856.0)	0.000”’	53573
rating	Median (Quartiles)	4.7 (3.7,6.0)	5.7 (4.7,6.6)	6.3 (5.4,7.1)	7.0 (6.0,7.8)	6.2 (5.3,7.1)	6.5 (5.4,7.5)	0.000”	0
rating	Mean (SD)	4.9 (1.6)	5.6 (1.5)	6.2 (1.4)	6.8 (1.6)	6.1 (1.4)	6.4 (1.7)	0.000”’	0
votes	Median (Quartiles)	65.0 (20.0,282.5)	40.0 (14.0,166.0)	39.0 (14.0,148.0)	18.0 (9.0,45.0)	34.5 (12.0,113.0)	10.0 (7.0,19.0)	0.000”	0
votes	Mean (SD)	2152.9 (9591.6)	630.2 (2868.1)	689.5 (4311.2)	106.1 (796.0)	357.0 (1551.6)	23.4 (82.9)	0.000”’	0

’ Chi-squared test
” Kruskal-Wallis rank sum test
”’ One-way analysis of variance (ANOVA)

```
# Or firstly store the results, and use stargazer_long for the subsequent
# LaTeX output, e.g. with rotation:
movies$budget = movies$budget/1e+06
s = characteristics_table(-2, "genre", movies[, -c(1, 7:16)], "col", prec = "%.1f",
    prec_continuous = "%.1f", prec_p = "%.3f", tablefootnote = FALSE)
stargazer_long(s, summary = FALSE, rownames = FALSE, output = "p{1.2cm}p{1.5cm}p{1.4cm}p{1.4cm}p{1.4cm}p
    rotate = 60)
```

Variable	Level	Action	Comedy	Drama	Documentary	Romance	Short	P	NAs
Action	True	2075 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000	27488
	False	0 (0.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Animation	True	35 (1.7)	113 (1.4)	35 (0.2)	8 (0.3)	9 (1.6)	943 (25.7)	0.000	27488

	False	2040 (98.3)	8237 (98.6)	14235 (99.8)	2384 (99.7)	537 (98.4)	2724 (74.3)		
Comedy	True	0 (0.0)	8350 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000	27488
	False	2075 (100.0)	0 (0.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Drama	True	0 (0.0)	0 (0.0)	14270 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.000	27488
	False	2075 (100.0)	8350 (100.0)	0 (0.0)	2392 (100.0)	546 (100.0)	3667 (100.0)		
Documentary	True	0 (0.0)	0 (0.0)	0 (0.0)	2392 (100.0)	0 (0.0)	0 (0.0)	0.000	27488
	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	0 (0.0)	546 (100.0)	3667 (100.0)		
Romance	True	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	546 (100.0)	0 (0.0)	0.000	27488
	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	0 (0.0)	3667 (100.0)		
Short	True	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3667 (100.0)	0.000	27488
	False	2075 (100.0)	8350 (100.0)	14270 (100.0)	2392 (100.0)	546 (100.0)	0 (0.0)		
year	Median (Quartiles)	1990.0 (1979.0,1997.0)	1979.0 (1959.0,1994.0)	1979.5 (1957.0,1996.0)	1997.0 (1985.0,2002.0)	1978.0 (1951.0,1996.0)	1995.0 (1966.0,2001.0)	0.000	0
year	Mean (SD)	1985.2 (16.9)	1974.8 (21.8)	1975.3 (22.7)	1991.4 (14.7)	1972.4 (24.6)	1980.6 (27.1)	0.000	0
length	Median (Quartiles)	94.0 (88.0,102.0)	91.0 (84.0,99.0)	96.0 (88.0,108.0)	85.0 (70.0,95.0)	95.0 (85.0,106.0)	11.0 (7.0,20.0)	0.000	0
length	Mean (SD)	100.0 (28.6)	91.4 (14.1)	99.0 (23.0)	87.2 (38.1)	99.1 (24.5)	14.1 (9.7)	0.000	0
budget	Median (Quartiles)	20.0 (4.0,50.0)	4.0 (0.7,17.0)	3.0 (0.5,12.0)	0.2 (0.1,0.6)	1.9 (0.3,6.0)	0.0 (0.0,0.0)	0.000	53573
budget	Mean (SD)	33.1 (36.8)	13.7 (21.1)	10.6 (17.9)	0.7 (1.9)	6.7 (11.9)	0.4 (3.5)	0.000	53573
rating	Median (Quartiles)	4.7 (3.7,6.0)	5.7 (4.7,6.6)	6.3 (5.4,7.1)	7.0 (6.0,7.8)	6.2 (5.3,7.1)	6.5 (5.4,7.5)	0.000	0
rating	Mean (SD)	4.9 (1.6)	5.6 (1.5)	6.2 (1.4)	6.8 (1.6)	6.1 (1.4)	6.4 (1.7)	0.000	0
votes	Median (Quartiles)	65.0 (20.0,282.5)	40.0 (14.0,166.0)	39.0 (14.0,148.0)	18.0 (9.0,45.0)	34.5 (12.0,113.0)	10.0 (7.0,19.0)	0.000	0
votes	Mean (SD)	2152.9 (9591.6)	630.2 (2868.1)	689.5 (4311.2)	106.1 (796.0)	357.0 (1551.6)	23.4 (82.9)	0.000	0

Rapport

http://rapport-package.info/ is an R package that facilitates creation of reproducible statistical report templates. Once created, rapport templates can be exported to various external formats: HTML, LaTeX, PDF, ODT, DOCX etc.

Let's have short look at the available templates:

```
rapport.ls()
```

```
##  [1] "AnalyzeWizard.rapport"
##  [2] "ANOVA.rapport"
##  [3] "BartlettsTest.rapport"
##  [4] "BrownForsythTest.rapport"
##  [5] "Correlation.rapport"
##  [6] "Crosstable.rapport"
##  [7] "Descriptives.rapport"
##  [8] "Example.rapport"
##  [9] "FactorAnalysis.rapport"
## [10] "FTest.rapport"
## [11] "GenerateBeta.rapport"
## [12] "GLM.rapport"
## [13] "HierarchicalClusterAnalysis.rapport"
## [14] "HomogeneityCrosstable.rapport"
## [15] "KMeansCluster.rapport"
## [16] "KolmogorovSmirnovTest.rapport"
## [17] "KruskalWallisTest.rapport"
## [18] "LinearRegression.rapport"
## [19] "MDS.rapport"
## [20] "Minimal.rapport"
## [21] "NormalityTest.rapport"
## [22] "OutlierTest.rapport"
## [23] "PCA.rapport"
## [24] "TTestOneSample.rapport"
## [25] "TTestTwoSample.rapport"
## [26] "WilcoxonTest.rapport"
```

Lets try to rebuild the Wilcoxon test results and compare with mytable results:

```
# rapport('WilcoxonTest.tpl', data=movies, ...)

# mytable(movies$Animation, movies$Action, movies$budget, prec='%.2e',
# prec_p='%.2e', latex=TRUE)
```

... to be continued