Using papeR with LaTeX

Benjamin Hofner

Version 1.0-0

Abstract

The main goal of the package **papeR** is to ease statistical reporting and thus to ease reproducible research. By relying on powerful tools such as the `Sweave` command kit, or the packages **knitr** and **xtable**, the package can be easily integrated in existing workflows.

The package provides an infrastructure to handle variable labels which are used in all other functions (`labels()`), allows to create (complex) summary tables of the data sets (`summarize()`) and to easily plot the data (`plot()` for labeled `data.frame`s), and enhances summary tables of statistical models by (possibly) adding confidence intervals, significance stars, odds ratios, etc. and by separating variable names and factor levels (`prettify()`).

1 Introduction

This is a short description of some of the LaTeXrelated features of `papeR`. For installation instructions and a comprehensive overview of the features of `papeR` see also the vignette on using `papeR` to (mainly) produce Markdown output (e.g. via `vignette("papeR_introduction", package = "papeR")`).

In short, we load the package, load an example data set, and set variable labels:

```
library("papeR")
data(Orthodont, package = "nlme")
labels(Orthodont) <- c("fissure distance (mm)",
                       "age (years)", "Subject", "Sex")
```

2 Pretty tables

To produce LaTeXtables, we heavily rely on the R package **xtable**. Note that all arguments to `xtable()` or `print.xtable()` can be used as usual, yet, some defaults were changed in package **papeR**. For example, we use per default the LaTeXpackage `booktabs` to print tables and tables are not set as floating objects. For further changes see the manual of `xtable.summary`.

2.1 Summary tables for numerical variables

Now we can produce summary tables for numeric variables:

```
xtable(summarize(Orthodont, type = "numeric"))
```

	N	Mean	SD	Min	Q1	Median	Q3	Max
distance	108	24.02	2.93	16.50	22.00	23.75	26.00	31.50
age	108	11.00	2.25	8.00	9.00	11.00	13.00	14.00

Grouped statistics with tests can be obtained via:

```
xtable(summarize(Orthodont, type = "numeric", group = "Sex"))
```

	Sex	N	Mean	SD	Min	Q1	Median	Q3	Max	p.value
distance	Male	64	24.97	2.90	17.00	23.00	24.75	26.50	31.50	<0.001
	Female	44	22.65	2.40	16.50	21.00	22.75	24.25	28.00	
age	Male	64	11.00	2.25	8.00	9.00	11.00	13.00	14.00	1.000
	Female	44	11.00	2.26	8.00	9.00	11.00	13.00	14.00	

Per default, t-tests are computed. To change the test, one can use:

```
xtable(summarize(Orthodont, type = "numeric", group = "Sex",
                 test = c("wilcox.test", "t.test")))
```

	Sex	N	Mean	SD	Min	Q1	Median	Q3	Max	p.value
distance	Male	64	24.97	2.90	17.00	23.00	24.75	26.50	31.50	<0.001
	Female	44	22.65	2.40	16.50	21.00	22.75	24.25	28.00	
age	Male	64	11.00	2.25	8.00	9.00	11.00	13.00	14.00	1.000
	Female	44	11.00	2.26	8.00	9.00	11.00	13.00	14.00	

To use Wicoxon tests for all variables, one could simply set `test = "wilcox.test"`, or one could switch off tests by setting `test = FALSE`.

To drop some of the statistics one can set several options to `FALSE`. E.g., if we do not want to show the five-number summaries (minimum, 25% quantile, median, 75% quantile, maximum), one can use

```
xtable(summarize(Orthodont, type = "numeric", group = "Sex",
                 quantiles = FALSE))
```

	Sex	N	Mean	SD	p.value
distance	Male	64	24.97	2.90	<0.001
	Female	44	22.65	2.40	
age	Male	64	11.00	2.25	1.000
	Female	44	11.00	2.26	

Alternatively or additionally, one could also drop N (`count = FALSE`) or mean and standard deviation (`mean_sd = FALSE`). For details see also the manual of `summary_numeric()`.

2.2 Summary tables for factor variables

In the same way, summary tables for factors can be computed. Here, we only want to print the variable `Sex`:

```
xtable(summarize(Orthodont, type = "factor", variables = "Sex"))
```

	Level	N	%
Sex	Male	64	59.3
	Female	44	40.7

If tables are longer than one page or cross a page, one can also use `"tabular.environment = "longtable""`:

```
print(xtable(summarize(Orthodont, type = "factor")),
      tabular.environment = "longtable")
```

	Level	N	%
Subject	M16	4	3.7

	Level	N	%
	M05	4	3.7
	M02	4	3.7
	M11	4	3.7
	M07	4	3.7
	M08	4	3.7
	M03	4	3.7
	M12	4	3.7
	M13	4	3.7
	M14	4	3.7
	M09	4	3.7
	M15	4	3.7
	M06	4	3.7
	M04	4	3.7
	M01	4	3.7
	M10	4	3.7
	F10	4	3.7
	F09	4	3.7
	F06	4	3.7
	F01	4	3.7
	F05	4	3.7
	F07	4	3.7
	F02	4	3.7
	F08	4	3.7
	F03	4	3.7
	F04	4	3.7
	F11	4	3.7
Sex	Male	64	59.3
	Female	44	40.7

which automatically specifies the table header such that it is repeated at the top of each new page.

To additionally obtain the cumulative frequencies, we can use:

```
xtable(summarize(Orthodont, type = "factor", variables = "Sex",
                 cumulative = TRUE))
```

	Level	N	%	∑ %
Sex	Male	64	59.3	59.3
	Female	44	40.7	100.0

As for numerical summaries, grouped statistics are tested. Per default, Fisher's

exact test is used. To speed up computations (of this non-sense test), we only use a small subset of the original data:

```
Ortho_small <- subset(Orthodont,
                      Subject %in% c("M01", "M02", "F01", "F02"))
xtable(summarize(Ortho_small, type = "factor",
                 variables = "Subject", group = "Sex"))
```

		Sex: Male		Sex: Female		
	Level	N	%	N	%	p.value
Subject	M02	4	50.0	0	0.0	< 0.001
	M01	4	50.0	0	0.0	
	F01	0	0.0	4	50.0	
	F02	0	0.0	4	50.0	

2.3 Captions

As usual, all floating tables can have captions. Per default, these are printed above the table. Note that by using the LaTeXpackage `capt-of`, one can specify table captions and labels also for non floating tables as shown in Table 2.

```
xtable(summarize(Ortho_small, type = "factor",
                 variables = "Subject", group = "Sex"),
       caption = "Example table for Fisher's exact test",
       label = "tab:Fisher")
```

Table 2: Example table for Fisher's exact test

		Sex: Male		Sex: Female		
	Level	N	%	N	%	p.value
Subject	M02	4	50.0	0	0.0	< 0.001
	M01	4	50.0	0	0.0	
	F01	0	0.0	4	50.0	
	F02	0	0.0	4	50.0	