

Publication-ready analytical and summary tables with R
Cheat Sheet

tbl summary()

Calculates descriptive stats for continuous, categorical, and dichotomous variables.

tbl_regression()

Turns a regression model object into a customized, formatted table.

Core Table Functions

tbl_survfit()

Turns a survfit object into a customized table with time-to-event estimates.



tbl_summary() using tidyverse syntax to summarize specific columns of a dataset with flexible customization options (See <u>vignettel</u>)

Basic code

trial %>% select(trt,age,grade,response) %>% tbl_summary()

Basic table

Characteristic	N = 200 ¹
Chemotherapy Treatment	
Drug A	98 (49%)
Drug B	102 (51%)
Age, yrs	47 (38, 57)
Unknown	11
Grade	
I	68 (34%)
II	68 (34%)
III	64 (32%)
Tumor Response	61 (32%)
Unknown	7
¹ n (%); Median (IQR)	

label Formula list of variable labels

Argument Input

missing_text

tbl_svysummary() for survey objects

Same functionality as tbl_summary(), but takes a survey object as input, and accounts for survey weights and design.

More info at:

http://www.danieldsjoberg.com/gtsum
mary/reference/tbl_svysummary.html

Customized code



Customized table

1		Characte	ristic	Drug A , N = 98	B [†] Drug B , N = 102 [†]
	→	Age (year	s)	47.01 (14.71)	47.45 (14.01)
		Missing		7	4
+	+>	Tumor gra	ade		
		I		35 (51%)	33 (49%)
		II		32 (47%)	36 (53%)
		III		31 (48%)	33 (52%)
I	Н	Missing		0	0
		Tumor Re	sponse		
I		0		67/132 (51%)	65/132 (49%)
		1		28/61 (46%)	33/61 (54%)
I		Missing		3	4
		¹ Mean (SD)); n (%); n/	N (%)	

For more info on customization arguments and options, visit http://www.danieldsjoberg.com/gtsummary/reference/tbl_summary.html

Customization options

Changes text of variable name in table

Effect on table

by Column to crosstabulate by Summary statistics will be calculated separately for each level of the variable

statistic

Formula list of summary statistic type for each variable

Changes summary statistics displayed for specified variables in table

digits Formula list of number of decimal places in places to display Changes number of rounded decimal places in table for specified continuous variables

Formula list specifying variable Changes variable type for specified variables, affecting which summary statistics are displayed

value Formula list of value to display for dichotomous variables

Changes the value displayed for dichotomous type variables

missing "no", "ifany", "always" Changes whether missing observations are reported

String to display for count of Changes the name of the missing data level for appropriate variables

Formula list of type of sorting to sort perform ("frequency" or "alphanumeric")

Changes the type of sorting for categorical variables

percent "column", "row", or "cell"

Changes how percentage statistics are calculated and displayed

Helper functions useful extensions to tbl_summary()

Extended code

add_n() %>%

add_p()

add_overall() %>%

Extended table

Characteristic	N	Overa	II, N = 200 ¹	Drug A , N = 98 ¹			
Age, yrs	189	47	(38, 57)	46 (37, 59)			
Tumor Response	193	61	(32%)	28 (29%)			
¹ Median (IQR); n (%) ² Wilcoxon rank sum		earson's	Chi-squared	d test			
+ +							
add_n() Adds a column v	vith tl	he	_	overall() olumn with			

Adds a column with the total number of non-missing observations

Adds a column with overall summary statistics

See also: add_q() bold_p() bold_labels() add_stat()

48 (39, 56)

33 (34%)

Drug B, N = 1021 p-value2

0.7

0.5

add_p() Adds column of p-values generated by testing for differences between groups. Takes arguments below.

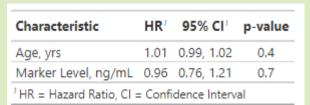
Argument	Default	Input	Effect on table
test	Continuous: "Kruskal test"; Categorical, cell counts ≥5: "chisq.test.no.correct"; Categorical, cell counts > 5: "fisher.test"	Formula list specifying statistical test to perform for each variable: "t.test", "aov", "wilcox.test", "kruskal.test", "chisq.test" and "lme4"; custom tests possible too	Changes p-value in table based on specified statistical test
pvalue_fun	style_pvalue()	Function to round and format p-values	Changes format of p-values in table

For more info, visit http://www.danieldsjoberg.com/gtsummary/reference/index.html

tbl_regression() Present regression model object in publication-ready table

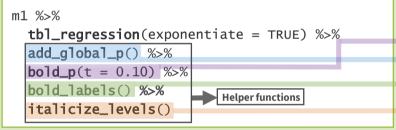
cox model: basic code

cox model: basic table



glm model: basic code

glm model: custom code using helper functions



1. Build model of interest

2. Use tbl_regression() to present results

tbl_regression() supports most commonly used regression models, and uses broom::tidy(x) to perform initial tidying.

For more info:

http://www.danieldsjoberg.com/gtsummary/reference/tbl_regression.html

glm model: basic table

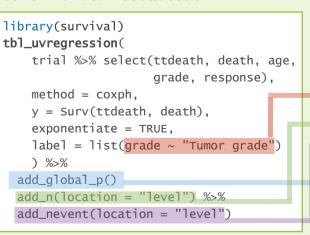
Characteristic	OR ¹	95% CI ¹	p-value		
Age, yrs	1.02	1.00, 1.04	0.091		
T Stage					
T1	_	_			
T2	0.58	0.24, 1.37	0.2		
T3	0.94	0.39, 2.28	0.9		
T4	0.79	0.33, 1.90	0.6		
OR = Odds Ratio, CI = Confidence Interval					

glm model: custom table

	Characteristic		OR ⁷	95%	Cl	p-value		
Age, yrs			1.02	1.00,	1.04	0.087		
	T Stage					4	0.6	
	T1			_	_	-		
	T2			0.58	0.24,	1.37		
	<i>T3</i>			0.94	0.39,	2.28		
	T4			0.79	0.33,	1.90		
OR = Odds Ratio, CI = Confidence Interval								

tbl_uvregression() displays multiple univariate regression models at once

cox univariate models: code



cox univariate models: table

Characteristic	N	Event N	\mathbf{HR}^{I}	95% CI ¹	p-value		
Age, yrs	189	103	1.01	0.99, 1.02	0.3		
Tumor grade					0.075		
1	68	33	_	_			
II	68	36	1.28	0.80, 2.05			
III	64	43	1.69	1.07, 2.66			
Tumor Response	193	107	0.50	0.31, 0.78	0.001		
HR = Hazard Ratio, CI = Confidence Interval							

Requires "method" parameter specifying model type. Can estimate univariate regression models holding either outcome ("y") or covariate ("x") constant, or both (see "formula" parameter). For more info about "formula" and other parameters, see:

https://www.danieldsjoberg.com/gtsummary/reference/tbl_uvregression.html

tbl_survfit() Present survfit object with custom estimates In publication-ready table

```
library(survival)
tbl_survfit(
list(
    survfit(Surv(ttdeath, death) ~ 1, trial),
    survfit(Surv(ttdeath, death) ~ trt, trial)
),
times = c(12, 24),
label_header = "**{time} Month**"
)
```

<u>\</u>		
12 Month	24 Month	
88% (84%, 93%)	44% (38%, 51%)	
91% (85%, 97%)	47% (38%, 58%)	
86% (80%, 93%)	41% (33%, 52%)	
	88% (84%, 93%) 91% (85%, 97%)	

Options for defining model:

- 1. x=explicit survfit model (or list of models) from dataframe
- 2. x=dataframe and designate y=Surv object and include=covariates in model

Options for output:

- 1. time-specific survival estimates using times=times of interest
- 2. quantile (e.g. median) survival times using probs=quantile of interest

For more info: http://www.danieldsjoberg. com/gtsummary/reference/ tbl_survfit.html

library(survival) tbl_survfit(trial, y = Surv(ttdeath, death), include = c(trt, grade), probs = 0.5, label_header = "**Median Survival**") %>% add_p()

Characteristic	Median Survival	p-value ¹
Chemotherapy Treatment		0.2
Drug A	24 (21, —)	
Drug B	21 (18, —)	
Grade		0.072
I	— (22, —)	
II	22 (18, —)	
III	20 (18, 23)	
¹ Log-rank test		

tbl_merge(), tbl_stack() combine tables by row or column

```
t1 = tbl_survfit(
  list(survfit(Surv(ttdeath, death) ~ trt +
  grade, trial),
   times = c(12, 24),
   label_header = "**{time} Month**"
)
```

```
t2 = tbl_survfit(
   trial,
   y = Surv(ttdeath, death),
   include = c(trt, grade),
   probs = 0.5,
   label_header = "**Median Survival**"
) %>% add_p()
```

tbl_merge(list(t1,t2), tab_spanner = FALSE)

Characteristic	12 Month	24 Month	Median Survival	p-value ¹
Chemotherapy Treatment				0.2
Drug A	91% (85%, 97%)	47% (38%, 58%)	24 (21, —)	
Drug B	86% (80%, 93%)	41% (33%, 52%)	21 (18, —)	
Grade				0.072
I	97% (93%, 100%)	51% (41%, 65%)	— (22, —)	
II	82% (74%, 92%)	47% (37%, 61%)	22 (18, —)	
III	86% (78%, 95%)	33% (23%, 47%)	20 (18, 23)	
¹ Log-rank test				

tbl_merge combines columns, tbl_stack() combines rows. For more info, see https://www.danieldsjoberg.com/gtsummary/reference/