

Putting together a simple Table 1

The data: Suppose we have an observational study with data collected on 100 people who took vitamin D and 100 people who did not take any vitamins. Here is a snippet of the data:

```
head(vitd)
```

ID	vitD	female	age	college	cat_age
1	1	1	59	0	40-64
2	1	1	69	0	65+
3	1	1	65	0	65+
4	1	1	49	0	40-64
5	1	1	60	1	40-64
6	1	1	78	0	65+

Our exposure of interest is whether someone took vitamin D or not, which is a binary factor and the exposure of interest. Our Table 1 will therefore stratify by Vitamin D intake status.

The patient characteristics we want to summarize are biological sex (male vs. female), participant age in years as well as in categories of age, and finally whether the participant has a college degree. Typically, you will have more variables than this to present but we are keeping it simple for this example.

What variable types do we want to describe? Biological sex and college degree indicator are both binary variables. Age in years is continuous in nature, so we need to decide whether we want to summarize by mean (SD) or median (IQR). In this example, we expect age to perhaps not be normally distributed, so we will present median and IQR. Age group is categorical, but it is also ordinal as the categories represent increasing ages. This doesn't matter for description, but if we want to test the association between age group and vitamin D intake, then we will want to make sure we account for the ordinal nature of age group in the analysis.

Constructing the table

```
# R packages required

library("Hmisc")
library("t.test")
library("DescTools")
library("table1")

# Each variable is made into a factor and we provide a label for each level

vitd$female <- factor(vitd$female, levels=c(1,0), labels=c("Female","Male"))
vitd$vitD <- factor(vitd$vitD, levels=c(1,0), labels=c("Yes","No"))
vitd$college <- factor(vitd$college, levels=c(1,0), labels=c("Yes","No"))

#Variable labels
label(vitd$female) <- "Biological Sex"
label(vitd$college) <- "College Degree"
label(vitd$cat_age) <- "Age group"
label(vitd$age) <- "Age (years)"

# Specifying that the summary statistics for continuous age will be median and range. For
# categorical data, it is the default n (%)

rndr <- function(x, name, ...) {
  if (!is.numeric(x)) return(render.categorical.default(x))
  what <- switch(name,
    age = "Median [Min, Max]")
  parse.abbrev.render.code(c("", what))(x)
}

# Finally, our command to produce the table, stratified by vitamin D status and using the
# summary statistics we specified above

table1(~ age + cat_age + female + college | vitD, data=vitd, render=rndr)
```

The table that is generated:

Vitamin D intake			
	Yes (N=100)	No (N=100)	Overall (N=200)
Age (years)			
Median [Min, Max]	54.5 [28.0, 85.0]	49.0 [26.0, 73.0]	51.0 [26.0, 85.0]
Age group			
<40	9 (9.0%)	17 (17.0%)	26 (13.0%)
40-64	71 (71.0%)	76 (76.0%)	147 (73.5%)
65+	20 (20.0%)	7 (7.0%)	27 (13.5%)
Biological Sex			
Female	70 (70.0%)	34 (34.0%)	104 (52.0%)
Male	30 (30.0%)	66 (66.0%)	96 (48.0%)
College Degree			
Yes	39 (39.0%)	34 (34.0%)	73 (36.5%)
No	61 (61.0%)	66 (66.0%)	127 (63.5%)

Adding p-values

Now that we understand which **tests** we need to generate p-values, we can add the p-values in a column to our Table 1. It is important to note that p-values are **not always required** for a Table 1, but with this observational data we may want to understand which variables are associated with our main exposure of interest (here, vitamin D intake).

Note: Recall, there are **specific R commands** to generate p-values and add them as a column to Table 1, however these are not part of the table1 package and require a bit more programming skill. It is easier to simply cut and paste your table 1 into a word document and add a column for the p-values (as well as any other modification to the table that you want to make).

We will use a two-group t-test for continuous age as the we have independently sampled individuals who are in two mutually exclusive groups based on vitamin D intake.

```
library("t.test")  
t.test(age ~ vitD , data = vitd, var.equal=TRUE)
```

As vitamin D status is binary (yes/no) and college degree and biological sex are also binary, we will use a chi-square test for association to test for any association of these predictors with vitamin D status.

```
chisq.test(vitd$vitD,vitd$female)  
chisq.test(vitd$vitD,vitd$college)
```

Age group is categorical with 3 levels and it is ordinal in nature so we will use a Cochran Armitage test for trend to see if age group is associated with a linear increase in the probability of vitamin D intake.

```
library("DescTools")  
table_vitDxagegrp = table(vitd$vitD,vitd$cat_age)  
CochranArmitageTest(table_vitDxagegrp)
```

We add these p-values to Table 1 along with a footnote describing what test was used with each type of predictor. Remember – *every table in a paper should stand on its own* and the reader should always be able to understand abbreviations, units, and which descriptive statistics and statistical tests were used without having to flip back to the methods section.

Vitamin D intake				
	Yes (N=100)	No (N=100)	Overall (N=200)	p-value*
Age (years)				0.0004
Median [Min, Max]	54.5 [28.0, 85.0]	49.0 [26.0, 73.0]	51.0 [26.0, 85.0]	
Age group				0.004
<40	9 (9.0%)	17 (17.0%)	26 (13.0%)	
40-64	71 (71.0%)	76 (76.0%)	147 (73.5%)	
65+	20 (20.0%)	7 (7.0%)	27 (13.5%)	
Biological Sex				<0.0001
Female	70 (70.0%)	34 (34.0%)	104 (52.0%)	
Male	30 (30.0%)	66 (66.0%)	96 (48.0%)	
College Degree				0.56
Yes	39 (39.0%)	34 (34.0%)	73 (36.5%)	
No	61 (61.0%)	66 (66.0%)	127 (63.5%)	

*Corresponds to t-test for continuous variables, chi-square test for association for nominal variables, and Cochran-Armitage test for trend for ordinal variables.

Results:

We can see that both older age ($p=0.004$) and female biological sex (<0.0001) are strongly associated with vitamin D intake. Having a college degree is not associated with vitamin D intake ($p=0.56$).