Making Your Table 1 Reproducible

Steve Ampah

May 17, 2020

1 Need these packages installed. No need to know LaTeXto do these Tables

```
install.packages("plyr")
install.packages("dplyr")
install.packages("stringr")
install.packages("rms")
install.packages("knitr")

----------------------------------------
How to specify a test in this function:
----------------------------------------
aov.t   is for  "ANOVA test"
fisher.t  is for "Fisher exact test"
chisq.t  is for "Chi-squared test"
t.test is for "T-test"
kruskal.t is for "Kruskal-Wallis test"
wilcox.t  is for "Wilcoxon ranked sum test"

----------------------------------------
 Arguements of myTable1 function:
----------------------------------------

myTable1(dat,
      contvar = NULL, contTest = NULL,
       catvar = NULL,  catTest = NULL, Test = T,
        splitvar, mydec = 1,
         pdec = 3,
          docaption = F,
           my.docaption = NULL,
            tsdec = 2,
             rowPCT = F,
              prmsd = NULL, my.loc = "./tab1.tex",
               mysize = "scriptsize",
                myinsert.bottom = NULL,
                 Trace = F, splitlabel = NULL, showtable = F,
                  mysim.p = F, myB = 200, mylongtable = F, mylandscape = F,
                   myeps = 0.001, bracket = F, mylevels = NULL,
                     chi.correct = F, exceloutput = F,
    exceloutputName = NULL, latexoutput = F, ...)
```

2 Table 1 : Summary across gender without variable labels

```
tt=myTable1(dat=dat1, splitvar="sex",splitlabel ="Gender",
       contvar=c("age","BP.sys","BP.dia","N.smokePday"), # continuous variables
       contTest=c("t.test","wilcox.t","t.test","aov.t"), # Test to be applied respectively to the contvars
       catvar=c("diabetic",  "Treatment","Race"),        # Categorical variable
       catTest=c("fisher.t","fisher.t","chisq.t"),       # Test to use for categorical variables
       docaption = T,                                    # Should code do caption for you ?
       my.docaption="xxxxxxxx",                    # If false, then write caption eg. "Summaries by sex"
       prmsd=c("mean","median","mean","mean"),   # Specify statistics for summaries
       my.loc="./tabhold/mytable1s1.tex",       # location for tex file
       Trace=F,                                 # Used for my editing
       pdec=2,                                  # Decimal place for p-values
       Test=T,                                  # Test statistic column to be included in table
       latexoutput=T,                           # Whether to spit out tex file
       exceloutput=T,exceloutputName ="ExxcelT1" , # Produce an excel file of Table 1
       showtable = F)                           # Whether to show Table on current screen
```

Table 1: Summary of patients' variables across Gender

Variables	N	**Gender**		
		Female N=40	Male N=60	Combined N=100
age	100	52.2±17.8	48.2±18	49.8±17.9
BP.sys	100	112 **119.7** 128.6	102.1 **114.9** 130.9	104.4 **117.2** 130.1
BP.dia	100	64±8.3	59.3±10.3	61.1±9.8
N.smokePday	100	2.2±1.5	1.8±1.4	2±1.5
diabetic	100			
No		85% (34)	83% (50)	84% (84)
Yes		15% (6)	17% (10)	16% (16)
Treatment	100			
Drug A		22% (9)	18% (11)	20% (20)
Drug A+B		45% (18)	45% (27)	45% (45)
Drug B		32% (13)	37% (22)	35% (35)
Race	100			
African American		57% (23)	65% (39)	62% (62)
White		42% (17)	35% (21)	38% (38)

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after percents are counts.

3 Replacing column 1 with variable labels

```
#Creating labels and units for variables
mylabel<-c("","Treatment Grps","Age (yrs)",
           "Systolic Blood Pressure (mmHg)",
           "Diastolic Blood Pressure (mmHg)",
           "Gender","Smoked Pks/day","Diabetic",
           "Cholesterol (mg/dl)","Race")

 j=1
for(i in names(dat1)){
  label(dat1[[i]])<-mylabel[j]
        j=j+1
}
```

Table 2: Summary of baseline variables across gender

Variables	N	**Gender**		
		Female N=40	Male N=60	Combined N=100
Age (yrs)	100	52(18)	48(18)	50(18)
Systolic Blood Pressure (mmHg)	100	112 **120** 129	102 **115** 131	104 **117** 130
Diastolic Blood Pressure (mmHg)	100	64(8)	59(10)	61(10)
Smoked Pks/day	100	2(2)	2(1)	2(1)
Cholesterol (mg/dl)	100	168(19)	206(27)	191(30)
Diabetic	100			
No		85% (34)	83% (50)	84% (84)
Yes		15% (6)	17% (10)	16% (16)
Treatment Grps	100			
Drug A		22% (9)	18% (11)	20% (20)
Drug A+B		45% (18)	45% (27)	45% (45)
Drug B		32% (13)	37% (22)	35% (35)
Race	100			
African American		57% (23)	65% (39)	62% (62)
White		42% (17)	35% (21)	38% (38)

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x(s)$ represents $\bar{X}(1SD)$. Numbers after percents are counts.

4 Include a column of P-values for comparison of variables across gender

```
tt3=myTable1(dat=dat1, splitvar="sex",splitlabel ="Gender",
       contvar=c("age","BP.sys","BP.dia","N.smokePday","cholesterol"),
       contTest=c("t.test","wilcox.t","t.test","t.test","t.test"),
       catvar=c("diabetic",  "Treatment"),
       catTest=c("chisq.t","fisher.t"),
       docaption = F,
       my.docaption="Summary of baseline variables across gender",
       prmsd=c("mean","median","mean","mean","mean"),
       my.loc="./tabhold/mytable1s3.tex",
       Trace=F,
       pdec=2,
       Test=T,  # This part is changed to T to report p-values
       latexoutput=T,
       exceloutput=F,exceloutputName ="testx" ,
       showtable = F)
```

Table 3: Summary of baseline variables across gender

Variables	N	**Gender**		Combined	Test Statistic
		Female	Male		
		N=40	N=60	N=100	
Age (yrs)	100	52.2±17.8	48.2±18	49.8±17.9	$t(98) = 1.09,\ P = 0.28$[1]
Systolic Blood Pressure (mmHg)	100	112 **119.7** 128.6	102.1 **114.9** 130.9	104.4 **117.2** 130.1	$W = 1352,\ P = 0.29$[2]
Diastolic Blood Pressure (mmHg)	100	64±8.3	59.3±10.3	61.1±9.8	$t(98) = 2.43,\ P = 0.02$[1]
Smoked Pks/day	100	2.2±1.5	1.8±1.4	2±1.5	$t(98) = 1.39,\ P = 0.17$[1]
Cholesterol (mg/dl)	100	168.2±19.2	205.8±26.6	190.8±30.2	$t(97) = -8.22,\ P < 0.001$[1]
Diabetic	100				$\chi^2_1 = 0.05,\ P = 0.82$[3]
No		85% (34)	83% (50)	84% (84)	
Yes		15% (6)	17% (10)	16% (16)	
Treatment Grps	100				$P = 0.8$[4]
Drug A		22% (9)	18% (11)	20% (20)	
Drug A+B		45% (18)	45% (27)	45% (45)	
Drug B		32% (13)	37% (22)	35% (35)	

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after percents are counts. Tests used: [1] T-test, [2] Wilcoxon ranked sum test, [3] Chi-squared test, [4] Fisher exact test.

5 Split variable of 2 or more levels

```
tt4=myTable1(dat=dat1, splitvar="Treatment",splitlabel =NULL,
       contvar=c("age","BP.sys","BP.dia","N.smokePday"),
       contTest=c("aov.t","kruskal.t","aov.t","aov.t"),
       catvar=c("sex",  "diabetic"),
       catTest=c("fisher.t","fisher.t"),
       docaption = F,
       my.docaption = "Summary of baseline variables across treatment",
       prmsd=c("mean","median","mean","mean"),
       my.loc="./tabhold/mytable1s4.tex",
       Trace=T,
       pdec=2,
       tsdec=0,
       Test=F,
       latexoutput = T,
       showtable = F)
```

Table 4: Summary of baseline variables across treatment

Variables	N	**Treatment**			
		Drug A N=20	Drug A+B N=45	Drug B N=35	Combined N=100
Age (yrs)	100	55.1±14.5	47.9±17.3	49.2±20.2	49.8±17.9
Systolic Blood Pressure (mmHg)	100	111 **124.4** 129.3	101.7 **113.4** 125.4	111.8 **117.6** 134.6	104.4 **117.2** 130.1
Diastolic Blood Pressure (mmHg)	100	64.9±9.2	60±10.7	60.4±8.5	61.1±9.8
Smoked Pks/day	100	2.5±1.5	1.7±1.2	2.1±1.7	2±1.5
Gender	100				
Female		45% (9)	40% (18)	37% (13)	40% (40)
Male		55% (11)	60% (27)	63% (22)	60% (60)
Diabetic	100				
No		90% (18)	87% (39)	77% (27)	84% (84)
Yes		10% (2)	13% (6)	23% (8)	16% (16)

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after percents are counts.

6 Ordering Treatment levels as : A, B and A+B

```
#table(dat1£Treatment)
tt5=myTable1(dat=dat1, splitvar="Treatment",splitlabel =NULL,
       mylevel=c("Drug A","Drug B", "Drug A+B"),  # ordering Trt was specified here
       contvar=c("age","BP.sys","BP.dia","N.smokePday"),
       contTest=c("aov.t","kruskal.t","aov.t","aov.t"),
       catvar=c("sex",  "diabetic"),
       catTest=c("fisher.t","fisher.t"),
       docaption = F,
       my.docaption = "Summary of baseline variables across treatment",
       prmsd=c("mean","median","mean","mean"),
       my.loc="./tabhold/mytable1s5.tex",
       Trace=T,
       pdec=2,
       tsdec=0,
       Test=F,
       latexoutput = T,
       showtable = F)
```

Table 5: Summary of baseline variables across treatment

Variables	N	Treatment			Combined
		Drug A	Drug B	Drug A+B	
		N=20	N=35	N=45	N=100
Age (yrs)	100	55.1±14.5	49.2±20.2	47.9±17.3	49.8±17.9
Systolic Blood Pressure (mmHg)	100	111 **124.4** 129.3	111.8 **117.6** 134.6	101.7 **113.4** 125.4	104.4 **117.2** 130.1
Diastolic Blood Pressure (mmHg)	100	64.9±9.2	60.4±8.5	60±10.7	61.1±9.8
Smoked Pks/day	100	2.5±1.5	2.1±1.7	1.7±1.2	2±1.5
Gender	100				
Female		45% (9)	37% (13)	40% (18)	40% (40)
Male		55% (11)	63% (22)	60% (27)	60% (60)
Diabetic	100				
No		90% (18)	77% (27)	87% (39)	84% (84)
Yes		10% (2)	23% (8)	13% (6)	16% (16)

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after percents are counts.

7 Introducing missing in numeric variables

```
for(i in names(dat1)){
  dd=sample(c(F,T),size=nrow(dat1), prob=c(.91,.09),replace=T)
  if( is.numeric(dat1[[i]])){
  dat1[[i]][dd]<-NA
  }

}


tt5=myTable1(dat=dat1, splitvar="Treatment",splitlabel =NULL,
       mylevel=c("Drug A","Drug B", "Drug A+B"),  # ordering Trt was specified here
       contvar=c("age","BP.sys","BP.dia","N.smokePday"),
       contTest=c("aov.t","kruskal.t","aov.t","aov.t"),
       catvar=c("sex",  "diabetic"),
       catTest=c("fisher.t","fisher.t"),
       docaption = F,
       my.docaption = "Summary of baseline variables across treatment",
       prmsd=c("mean","median","mean","mean"),
       my.loc="./tabhold/mytable1s6.tex",
       Trace=T,
       pdec=2,
       tsdec=0,
       Test=T,
       latexoutput = T,
       showtable = F)
```

Table 6: Summary of baseline variables across treatment

Variables	N	Treatment			Combined	Test Statistic
		Drug A N=20	Drug B N=35	Drug A+B N=45	N=100	
Age (yrs)	87	54.8±14.8	51.8±18.8	49.6±17.6	51.5±17.4	$F_{2,84} = 1,\ P = 0.56$[1]
Systolic Blood Pressure (mmHg)	92	108.2 **120.8** 129.3	111.1 **117** 135	101.5 **113.3** 122.2	103.3 **114.9** 129.9	$\chi^2_2 = 3,\ P = 0.21$[2]
Diastolic Blood Pressure (mmHg)	91	65.8±9.3	61.5±8.4	60±10.9	61.6±10	$F_{2,88} = 2,\ P = 0.11$[1]
Smoked Pks/day	91	2.5±1.6	2.2±1.7	1.6±1.2	2±1.5	$F_{2,88} = 2,\ P = 0.1$[1]
Gender	100					$P = 0.8$[3]
Female		45% (9)	37% (13)	40% (18)	40% (40)	
Male		55% (11)	63% (22)	60% (27)	60% (60)	
Diabetic	100					$P = 0.43$[3]
No		90% (18)	77% (27)	87% (39)	84% (84)	
Yes		10% (2)	23% (8)	13% (6)	16% (16)	

$_a\ b\ _c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after percents are counts. Tests used: [1] ANOVA test, [2] Kruskal-Wallis test, [3] Fisher exact test.