

# Creació semiautomàtica de taules descriptives per grups

compareGroups

UAB

**Joan Vila & Isaac Subirana**

`jvila@imim.es, isubirana@imim.es`

RICAD Research on Inflammatory and Cardiovascular Disorders Program  
IMIM-Parc de Salut Mar, Barcelona

UAB-2014

**Table 2. Baseline Characteristics of the Participants According to Study Group.\***

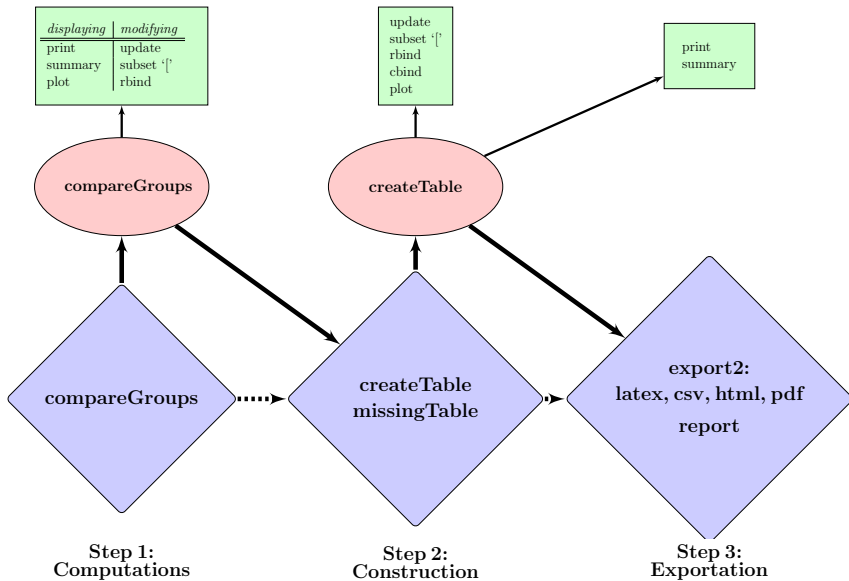
Characteristic	Mediterranean Diet with EVOO (N = 2543)	Mediterranean Diet with Nuts (N = 2454)	Control Diet (N = 2450)
Female sex — no. (%)†	1493 (58.7)	1326 (54.0)	1463 (59.7)
Age — yr†	67.0±6.2	66.7±6.1	67.3±6.3
Race or ethnic group — no. (%)			
White, from Europe	2470 (97.1)	2390 (97.4)	2375 (96.9)
Hispanic, from Central or South America	35 (1.4)	29 (1.2)	38 (1.6)
Other	38 (1.5)	35 (1.4)	37 (1.5)
Smoking status — no. (%)			
Never smoked	1572 (61.8)	1465 (59.7)	1527 (62.3)
Former smoker	618 (24.3)	634 (25.8)	584 (23.8)
Current smoker	353 (13.9)	355 (14.5)	339 (13.8)
Body-mass index‡			
Mean	29.9±3.7	29.7±3.8	30.2±4.0
<25 — no. (%)	195 (7.7)	204 (8.3)	164 (6.7)
25–30 — no. (%)	1153 (45.3)	1163 (47.4)	1085 (44.3)
>30 — no. (%)	1195 (47.0)	1087 (44.3)	1201 (49.0)
Waist circumference — cm	100±10	100±11	101±11
Waist-to-height ratio†§	0.63±0.06	0.63±0.06	0.63±0.07
Hypertension — no. (%)¶	2088 (82.1)	2024 (82.5)	2050 (83.7)
Type 2 diabetes — no. (%)†	1282 (50.4)	1143 (46.6)	1189 (48.5)
Dyslipidemia — no. (%)**	1821 (71.6)	1799 (73.3)	1763 (72.0)
Family history of premature CHD — no. (%)††	576 (22.7)	532 (21.7)	560 (22.9)

Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, Gómez-Gracia E, Ruiz-Gutiérrez V, Fiol M, Lapetra J, Lamuela-Raventos RM, Serra-Majem L, Pintó X, Basora J, Muñoz MA, Sorlí JV, Martínez JA, Martínez-González MA; PREDIMED Study Investigators (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. *New England Journal of Medicine* 368(14): 1279-1290.

methods

classes

functions



- La variable de **columna** pot ser:
  - Cadena
  - o Factors
  - o Temps fins esdeveniment
- Les variables de **fila** poden ser:
  - Com les de fila
  - i/o Numèriques

## Exemple del que no pot ser:

variable name: dat\$hp.preq

variable label:

num. of missings: 3 ( 3.19 %)

freq

- 1 (1.10%)

? 5 (5.49%)

no 28 (30.8%)

No 16 (17.6%)

NO 2 (2.20%)

si 30 (33.0%)

sI 1 (1.10%)

Si 6 (6.59%)

SI 2 (2.20%)

variable name: dat\$sup.hie

variable label:

num. of missings: 1 ( 1.06 %)

	freq
IV	3 (3.23%)
IV*	1 (1.08%)
IV. Tomo hierro oral solamente por dias	1 (1.08%)
No	7 (7.53%)
No, solo IV post Qx	1 (1.08%)
No, solo tomo 1 dia	1 (1.08%)
si	9 (9.68%)
Si	62 (66.7%)
Si	2 (2.15%)
SI	1 (1.08%)
Si* (+IV)	1 (1.08%)
Si, pero solo tomo HE post CX por un tiempo corto	1 (1.08%)
Si (solo 3 meses)	1 (1.08%)
Sup Hie	1 (1.08%)
Tomo apartir del 2012	1 (1.08%)

## “label” (library(Hmisc))

El nom/etiqueta de la variable que es veurà a la taula serà:

- el “label” si està etiquetada
- o el nom de la variable

```
> library(Hmisc)
> pes<-c(3.050, 2.840, 3.240 )
> label(pes)
```

```
[1] ""
```

```
> label(pes)<-"pes al nàixer"
> label(pes)
```

```
[1] "pes al nàixer"
```

## “factor”

Per defecte les variables de caràcter es converteixen a factor i, si no s'especifica el contrari, els “levels” són els seus valors, p.e.:

```
> smk<-c("Mai", "Ex-fumador", "Fumador", "Mai", "Fumador", "Fumador")
> table(smk)
```

```
smk
Ex-fumador   Fumador     Mai
           1         3         2
```

```
> smk2<-factor(smk, levels= c( "Mai", "Fumador", "Ex-fumador"))
> table(smk2)
```

```
smk2
    Mai   Fumador Ex-fumador
    2       3         1
```



Si s'ha d'analitzar una variable que indica temps fins esdeveniment (o censura), aquesta s'ha de definir prèviament:

```
> data(predimed)
> predimed$tmain <- with(predimed, Surv(toevent, event == 'Yes'))
> label(predimed$tmain) <- "AMI, stroke, or CV Death"
> head(predimed$tmain)
```

AMI, stroke, or CV Death

```
[1] 5.374401 6.097194+ 5.946612+ 2.907598 4.761123+ 3.148528
```

```
> (resu<-compareGroups(group ~ sex + age + p14, data = predimed))
```

```
----- Summary of results by groups of 'Intervention group'-----
```

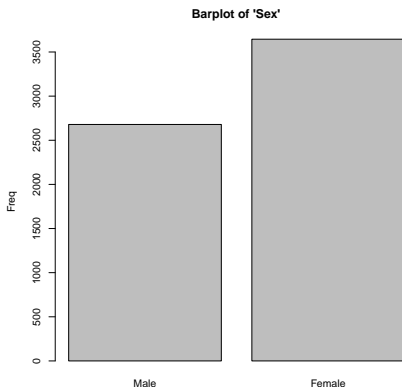
var	N	p.value	method	selection
1 Sex	6324	<0.001**	categorical	ALL
2 Age	6324	0.003**	continuous normal	ALL
3 MeDiet Adherence score	6324	<0.001**	continuous normal	ALL

```
-----
```

```
Signif. codes:  0 '**' 0.05 '*' 0.1 ' ' 1
```

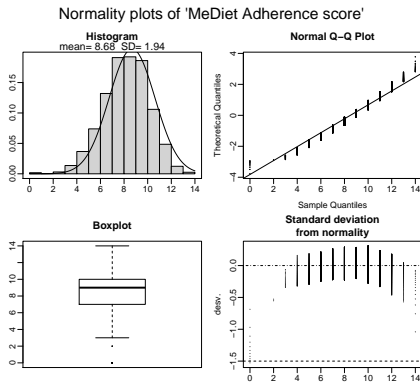
Si es fa un *plot* de l'objecte **resu** que és de classe `compareGroups` es pot veure la forma de les dades:

```
> plot(resu[1], file="./figures/")
```



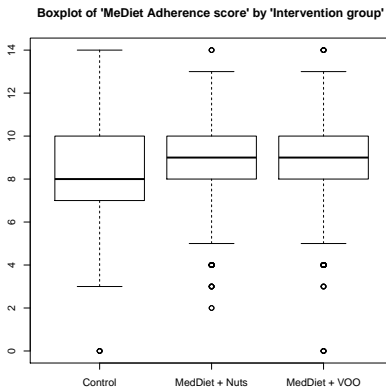
A les variables numèriques permet valorar la normalitat:

```
> plot(resu[3], file="./figures/")
```



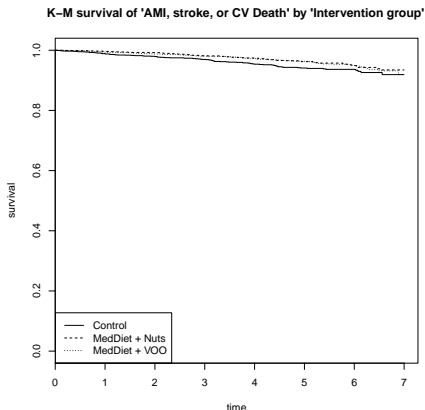
Les figures es poden fer per grups de la variable "columna":

```
> plot(resu[3], bivar=TRUE, file="./figures/bivar")
```



En el temps fins esdeveniment dibuixa corbes de KM:

```
> resu2<-compareGroups(group ~ tmain, data = predimed)  
> plot(resu2, bivar=TRUE, file="./figures/bivar")
```



Amb la funció **compareGroups** es fan tots els càlculs i amb la funció **createTable** es mostra allò que es vol d'aquest càlculs. La sortida bàsica té aquesta forma:

```
> createTable(resu)
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
MedDiet Adherence score	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001

A més del que s'ha mostrat en la anterior diapositiva, el que s'ha calculat amb la funció **compareGroups** és:

- descripció general per a tota la mostra
- la “n” de casos vàlids per a cada variable de fila
- presentar OR/HR
- p-valor: global, tendències, comparacions múltiples, OR/HR (per cada categoria)

Totes aquests càlculs es poden mostrar amb la funció **createTable**.



## Exemple de resultats que es poden obtenir:

```
> createTable(compareGroups(diab ~ smoke + sex + age + tmain,
+   data = predimed), show.p.ratio=TRUE, show.all=TRUE,
+   show.ratio=TRUE, show.p.overall=FALSE)
```

```
-----Summary descriptives table by 'Type-2 diabetes'-----
```

	[ALL] N=6324	No N=3322	Yes N=3002	OR	p
<b>Smoking:</b>					
Never	3892 (61.5%)	2053 (61.8%)	1839 (61.3%)	Ref.	
Current	858 (13.6%)	514 (15.5%)	344 (11.5%)	0.75 [0.64;0.87]	<
Former	1574 (24.9%)	755 (22.7%)	819 (27.3%)	1.21 [1.08;1.36]	<
<b>Sex:</b>					
Male	2679 (42.4%)	1267 (38.1%)	1412 (47.0%)	Ref.	
Female	3645 (57.6%)	2055 (61.9%)	1590 (53.0%)	0.69 [0.63;0.77]	<
Age	67.0 (6.17)	66.6 (6.10)	67.5 (6.22)	1.02 [1.02;1.03]	<
AMI, stroke, or CV Death	4.26%	3.20%	5.45%	. [.;.]	

Quan la variable és “temps fins..” es mostra la supervivència estimada per KM fins a un temps, que per defecte és a la mediana del seguiment.

Si la variable columna és 'temps fins..' s'obtenen HR:

```
> createTable(compareGroups(tmain ~ smoke + sex + age ,
+ data = predimed), show.ratio=TRUE, show.p.overall=FALSE)
```

-----Summary descriptives table by 'AMI, stroke, or CV Death'-----

	No event N=6072	Event N=252	HR	p.ratio
<b>Smoking:</b>				
Never	3778 (62.2%)	114 (45.2%)	Ref.	Ref.
Current	809 (13.3%)	49 (19.4%)	1.96 [1.40;2.74]	<0.001
Former	1485 (24.5%)	89 (35.3%)	2.02 [1.53;2.67]	<0.001
<b>Sex:</b>				
Male	2528 (41.6%)	151 (59.9%)	Ref.	Ref.
Female	3544 (58.4%)	101 (40.1%)	0.49 [0.38;0.63]	<0.001
Age	66.9 (6.14)	69.4 (6.65)	1.06 [1.04;1.09]	<0.001

Opcions de la funció compareGroups: **selec =**

Permet fer les anàlisis en un subset d'individus:

```
> createTable(compareGroups(group ~ diab + age ,
+   data = predimed, selec= c(sex=='Male')))
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=812	MedDiet + Nuts N=968	MedDiet + V00 N=899	p.overall
Type-2 diabetes:				0.468
No	382 (47.0%)	472 (48.8%)	413 (45.9%)	
Yes	430 (53.0%)	496 (51.2%)	486 (54.1%)	
Age	66.4 (6.62)	65.8 (6.40)	66.1 (6.61)	0.215

Opcions de la funció compareGroups: **method=**

- .- 1 - forces analysis as "normal-distributed"
- .- 2 - forces analysis as "continuous non-normal"
- .- 3 - forces analysis as "categorical"
- .- 4 - NA, which performs a Shapiro-Wilks test to decide between normal or non-normal

```
> createTable(compareGroups(group ~ diab + age ,
+   data = predimed, method= c(age=2)))
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Type-2 diabetes:				0.017
No	1072 (52.5%)	1150 (54.8%)	1100 (50.4%)	
Yes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	
Age	67.0 [62.0;72.0]	66.0 [62.0;71.0]	67.0 [62.0;72.0]	0.003

Altres opcions de la funció `compareGroups` (I):

- .- **timemax**, a quin temps s'ha de calcular la supervivència per KM
- .- **alpha**, punt de tall per decidir si és normal
- .- **z**, a quantes DEs s'ha considerar un valor com “no normal” en els *qqplot*
- .- **min.dis**, mínim diferents valors que ha de tenir una variable numèrica per a no ser considerada un factor
- .- **max.ylev**, màxim nivells que pot tenir un factor que sigui variable de fila
- .- **max.xlev**, màxim nivells que pot tenir un factor que sigui variable de columna
- .- **include.label**, mostra l'etiqueta de la variable
- .- **Q1, Q3**, per indicar els quantils

## Altres opcions de la funció compareGroups (i II):

- .- **simplify**, indica si s'ha de mostrar un *level* encara que no el tingui ningú
- .- **ref**, la categoria que s'ha de considerar de referencia quan es calcula risc
- .- **ref.no**, considera el “no” com a categoria de referencia quan es calcula risc
- .- **fact.ratio**, si p.e. es vol calcular el risc per cada 10 anys
- .- **ref.y**, categoria de referencia per la variable columna
- .- **p.corrected**, aplica la correcció per comparacions múltiples
- .- **compute.ratio**, calcula OR/HR
- .- **include.miss**, considera els *missing* com una categoria

Opcions de la funció `compareGroups` que afecten al *plot*:

- .- **file**, per salvar els *plot* en un fitxer
- .- **type**, per triar el format en que el vol guardar (p.e. `png`)
- .- **bivar**, fa figures per a cada categoria de la variable columna
- .- **n.breaks**, és el mateix argument que es fa servir a la funció *hist*

## Opcions de la funció createTable: **hide** i **hide.no**

```
> createTable(compareGroups(group ~ smoke + age + sex + htn + dial
+   data = predimed), hide=c(sex="Female", smoke="Never"),
+   hide.no = c("No"))
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Smoking:				0.444
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Sex: Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	<0.001
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249
Type-2 diabetes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017



## Opcions de la funció createTable: **digits**, **digits.ratio** i **digits.p**

```
> createTable(compareGroups(group ~ age + sex,
+   data = predimed), digits=c(2,3))
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Age	67.34 (6.28)	66.68 (6.02)	67.02 (6.21)	0.003
Sex:				<0.001
Male	812 (39.765%)	968 (46.095%)	899 (41.201%)	
Female	1230 (60.235%)	1132 (53.905%)	1283 (58.799%)	

## Opcions de la funció createTable: **sd.type** i **q.type**

```
> createTable(compareGroups(group ~ age + bmi,
+   data = predimed, method=c(bmi=2)), sd.type=2, q.type=c(2,2))
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Age	67.3±6.28	66.7±6.02	67.0±6.21	0.003
Body mass index	30.0 (27.5,32.8)	29.5 (26.9,32.2)	29.7 (27.2,32.4)	<0.001

Altres arguments de la funció createTable:

.- **type**, el 1 mostra freqüències, 2 freq i %, 3 sòls %

.- **show.p.overall**, **show.all**, **show.p.trend**, **show.p.mul** i

**show.p.ratio**, mostra els diferents p-valor

.- **show.n**, mostra els registres no *missing* per cada variable de fila

.- **show.ratio**, mostra OR/RR

.- **show.descr**, permet amagar els descriptius i posar p.e. sòls el p-valor

## update

```
> createTable(resu<-compareGroups(group ~ age + bmi,
+ data = predimed, method=c(bmi=2)))
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Body mass index	30.0 [27.5;32.8]	29.5 [26.9;32.2]	29.7 [27.2;32.4]	<0.001

```
> createTable(update(resu, . ~ . + sex))
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Body mass index	30.0 [27.5;32.8]	29.5 [26.9;32.2]	29.7 [27.2;32.4]	<0.001
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	

## varinfo/sele

```
> varinfo(resu<-compareGroups(group ~ age + bmi + bmi, data = preo
```

```
--- Analyzed variable names ----
```

```
  Orig varname Shown varname
1 group        Intervention group
2 age          Age
3 bmi          Body mass index
4 bmi.1        Body mass index
```

```
> createTable(update(resu, selec=list(bmi.1=(sex=="Female"))),
+             show.n=TRUE)
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall	N
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003	6324
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001	6324
Body mass index	30.8 (4.20)	30.2 (4.08)	30.4 (3.91)	0.002	3645

## Convinar taules per fila

```
> restab1 <- createTable(compareGroups(group ~ age + sex, data=preco)
> restab2 <- createTable(compareGroups(group ~ bmi + smoke, data=preco)
> rbind("Non-modifiable risk factors"=restab1, "Modifiable risk factors"=restab2)
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
<b>Non-modifiable risk factors:</b>				
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
<b>Modifiable risk factors:</b>				
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Smoking:				0.444
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	

## Eliminar el tabac i canviar l'ordre

```
> rbind("Modifiable"=restab1,"Non-modifiable"=restab2)[c(3,2,1)]
```

```
-----Summary descriptives table by 'Intervention group'-----
```

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
<b>Non-modifiable:</b>				
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
<b>Modifiable:</b>				
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003

## Exportar a **CSV**

```
> resu<-createTable(compareGroups(group ~ age + sex + bmi,  
+                               data = predimed))  
> export2csv(resu, file="tmp", sep=",")
```

## Exportar a **HTML**

```
> export2html(resu, file="tmp")
```



Exportar a L<sup>A</sup>T<sub>E</sub>X

&gt; export2latex(resu)

```

\begin{longtable}{lcccc}\caption{Summary descriptives table by groups of `Inte
\hline
& Control & MedDiet + Nuts & MedDiet + VOO & \multirow{2}{*}{p.overall}\
& N=2042 & N=2100 & N=2182 & \\

\hline
\hline
\endfirsthead
\multicolumn{5}{l}{\tablename\ \thetable} \textit{-- continued from previous
\hline
& Control & MedDiet + Nuts & MedDiet + VOO & \multirow{2}{*}{p.overall}\
& N=2042 & N=2100 & N=2182 & \\

\hline
\hline
\endhead
\hline
\multicolumn{5}{l}{\textit{continued on next page}} \\
\endfoot
\multicolumn{5}{l}{} \\
\endlastfoot
Age & 67.3 (6.28) & 66.7 (6.02) & 67.0 (6.21) & 0.003 \\
Sex: & & & & <$0.001 \\
\quad$Male & 812 (39.8\%) & 968 (46.1\%) & 899 (41.2\%) &

```

Amb format L<sup>A</sup>T<sub>E</sub>X:

: Summary descriptives table by groups of 'Intervention group'

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001

## Opcions **export2latex**

- **which.table**, 'descr', 'avail' or 'both'
- **size**, 'tiny', 'scriptsize', 'footnotesize', 'small'. . .
- **nmax**, per indicar la 'N' vàlida de cada grup
- **caption**, per modificar el títol de la taula
- **loc.caption**, 'top'/'bottom'
- **label**, per citar la taula en el document
- **landscape**, si la taula s'ha de fer apaisada
- **colmax**, el nombre de columnes màxim de la taula