El paquete compareGroups

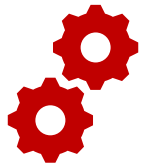
compareGroups

Esta librería incluye un conjunto de funciones extremadamente útiles para generar tablas de resultados.

Su instalación requiere un cierto tiempo, por lo que es necesario proceder con paciencia.

En este tutorial veremos las aplicaciones más interesantes analizando la base de datos ***predimed*** que se incluye en el paquete.

Instalación y carga del paquete



```
install.packages('compareGroups')
library(compareGroups)
data(predimed)
```

Base de datos *predimed*

The PREDIMED trial (Prevención con Dieta Mediterránea) is a randomized, parallel and multicentric cohort with more than 7,000 participants who were randomly assigned to three diet groups (olive oil + mediterranean diet, nuts + mediterranean diet, and low-fat diet -control group-) and followed-up during more than 7 years.

Participants of PREDIMED study were recruited in 2003 and were revisited anually until 2010. In each visit, a large variety of different information was collected: antropometric (body mass index, age, ...), blood samples (Total cholesterol, HDL cholesterol, ...), dietary pattern consumtion using validated questionaries, treatments, etc. Also, cardiovascular events were recorded during the follow-up period. At recruitment period, each participant was assigned to each of the three diet type group. Periodically (every 6-months approximately), participants were advised to follow the diet corresponding to their group. To build this example data set approx 85% of original participants were randomly selected.

http://www.predimed.es/

Base de datos *predimed*

Prediimed
Prevención con Dieta Mediterránea

group
a factor with levels Control MedDiet + Nuts MedDiet + VOO. Intervention group.

sex
a factor with levels Male Female.Sex.

age
Age.

smoke
a factor with levels Never Current Former.Smoking.

bmi
Body mass index.

waist
Waist circumference.

wth
Waist-to-height ratio.

htn
a factor with levels No Yes.Hypertension.

diab
a factor with levels No Yes .Type-2 diabetes.

hyperchol
a factor with levels No Yes.Dyslipidemia.

famhist
a factor with levels No Yes.Family history of premature CHD.

hormo
a factor with levels No Yes.Hormone-replacement therapy.

p14
MeDiet Adherence score.

toevent
follow-up to main event (years).

event
a factor with levels No Yes.AMI, stroke, or CV Death.

```
> head(predimed)
          group    sex age   smoke   bmi waist       wth htn
1       Control   Male  58  Former 33.53   122 0.7530864  No
2       Control   Male  77 Current 31.05   119 0.7300614 Yes
4  MedDiet + VOO Female  72  Former 30.86   106 0.6543210  No
5 MedDiet + Nuts   Male  71  Former 27.68   118 0.6941177 Yes
6  MedDiet + VOO Female  79   Never 35.94   129 0.8062500 Yes
8       Control   Male  63  Former 41.66   143 0.8033708 Yes
  diab hyperchol famhist hormo p14  toevent event
1   No       Yes      No    No  10 5.374401   Yes
2  Yes        No      No    No  10 6.097194    No
4  Yes        No     Yes    No   8 5.946612    No
5   No       Yes      No    No   8 2.907598   Yes
6   No       Yes      No    No   9 4.761123    No
8  Yes       Yes      No  <NA>   9 3.148528   Yes
```

Crear una tabla

```
res <- compareGroups(group ~ age + sex + smoke + waist ,
                     data = predimed)
restab <- createTable(res)
restab
```

--------Summary descriptives table by '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%)	
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%)	
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045

compareGroups proporciona tablas en formato elegante para publicación directa.

Primero definimos la comparación. En este caso, queremos comparar las variables edad, sex y smoke en los grupos de intervención.

Para las variables continuas (age y waist) obtenemos la media y su desviación estándar.

Para las qualitativa (sex, smoke) obtenemos tablas de frecuencias y proporciones.

En cada caso se obtiene un p-valor de la comparación entre grupos.

Crear una tabla

```
res <- compareGroups(group ~ . ,
                     data = predimed)
restab <- createTable(res)
restab
```

En este caso obtenemos la tabla completa que compara, de manera univariante, todas las variables entre los distintos grupos de intervención.

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

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO 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
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%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001
Hypertension:				0.249
No	331 (16.2%)	362 (17.2%)	396 (18.1%)	
Yes	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	
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%)	
Dyslipidemia:				0.423
No	563 (27.6%)	561 (26.7%)	622 (28.5%)	
Yes	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	
Family history of premature CHD:				0.581
No	1580 (77.4%)	1640 (78.1%)	1675 (76.8%)	
Yes	462 (22.6%)	460 (21.9%)	507 (23.2%)	
Hormone-replacement therapy:				0.850
No	1811 (98.3%)	1835 (98.4%)	1918 (98.2%)	
Yes	31 (1.68%)	30 (1.61%)	36 (1.84%)	
MeDiet Adherence score	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001
follow-up to main event (years)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001
AMI, stroke, or CV Death:				0.064
No	1945 (95.2%)	2030 (96.7%)	2097 (96.1%)	
Yes	97 (4.75%)	70 (3.33%)	85 (3.90%)	

Tablas estratificadas

```
res <- compareGroups(group ~ . - sex, predimed)
restab <- createTable(res, hide.no = "no")
strataTable(restab, "sex")
```

	Male				Female			
	Control N=812	MedDiet + Nuts N=968	MedDiet + VOO N=899	p.overall	Control N=1230	MedDiet + Nuts N=1132	MedDiet + VOO N=1283	p.overall
Age	66.4 (6.62)	65.8 (6.40)	66.1 (6.61)	0.215	68.0 (5.96)	67.4 (5.57)	67.7 (5.84)	0.056
Smoking:				0.851				0.907
Never	205 (25.2%)	266 (27.5%)	236 (26.3%)		1077 (87.6%)	993 (87.7%)	1115 (86.9%)	
Current	204 (25.1%)	242 (25.0%)	221 (24.6%)		66 (5.37%)	54 (4.77%)	71 (5.53%)	
Former	403 (49.6%)	460 (47.5%)	442 (49.2%)		87 (7.07%)	85 (7.51%)	97 (7.56%)	
Body mass index	29.6 (3.45)	29.1 (3.28)	29.2 (3.28)	0.018	30.8 (4.20)	30.2 (4.08)	30.4 (3.91)	0.002
Waist circumference	104 (9.82)	103 (9.36)	103 (9.65)	0.289	99.0 (11.0)	97.8 (11.0)	98.0 (10.5)	0.016
Waist-to-height ratio	0.62 (0.06)	0.62 (0.06)	0.62 (0.06)	0.191	0.64 (0.07)	0.63 (0.07)	0.63 (0.07)	0.002
Hypertension	649 (79.9%)	753 (77.8%)	682 (75.9%)	0.130	1062 (86.3%)	985 (87.0%)	1104 (86.0%)	0.780
Type-2 diabetes	430 (53.0%)	496 (51.2%)	486 (54.1%)	0.468	540 (43.9%)	454 (40.1%)	596 (46.5%)	0.007
Dyslipidemia	531 (65.4%)	653 (67.5%)	589 (65.5%)	0.575	948 (77.1%)	886 (78.3%)	971 (75.7%)	0.319
Family history of premature CHD	135 (16.6%)	171 (17.7%)	156 (17.4%)	0.841	327 (26.6%)	289 (25.5%)	351 (27.4%)	0.596
Hormone-replacement therapy	0 (0.00%)	0 (0.00%)	0 (0.00%)	.	31 (2.64%)	30 (2.81%)	36 (2.95%)	0.898
MeDiet Adherence score	8.57 (1.94)	8.86 (1.96)	8.91 (2.02)	0.001	8.36 (1.94)	8.77 (1.86)	8.68 (1.92)	<0.001
follow-up to main event (years)	4.05 (1.78)	4.38 (1.74)	4.53 (1.64)	<0.001	4.12 (1.71)	4.26 (1.67)	4.72 (1.56)	<0.001
AMI, stroke, or CV Death	58 (7.14%)	41 (4.24%)	52 (5.78%)	0.029	39 (3.17%)	29 (2.56%)	33 (2.57%)	0.576

Descriptiva total y por grupos

```
descrTable(group ~ ., predimed, hide.no = "no", show.all = TRUE)
```

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

	[ALL] N=6324	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
Sex:					<0.001
Male	2679 (42.4%)	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	3645 (57.6%)	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.0 (6.17)	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Smoking:					0.444
Never	3892 (61.5%)	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	858 (13.6%)	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	1574 (24.9%)	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.0 (3.82)	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Waist circumference	100 (10.6)	101 (10.8)	100 (10.6)	100 (10.4)	0.045
Waist-to-height ratio	0.63 (0.07)	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001
Hypertension	5235 (82.8%)	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249
Type-2 diabetes	3002 (47.5%)	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017
Dyslipidemia	4578 (72.4%)	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.423
Family history of premature CHD	1429 (22.6%)	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.581
Hormone-replacement therapy	97 (1.71%)	31 (1.68%)	30 (1.61%)	36 (1.84%)	0.850
MeDiet Adherence score	8.68 (1.94)	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001
follow-up to main event (years)	4.36 (1.69)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001
AMI, stroke, or CV Death	252 (3.98%)	97 (4.75%)	70 (3.33%)	85 (3.90%)	0.064

Tablas de porcentajes

```
res <- compareGroups(event ~ sex ,
                     data = predimed)
restab <- createTable(res,type=1)
restab
```

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

_____________________________________
              No     Yes   p.overall
            N=6072 N=252
_____________________________________
Sex:                        <0.001
    Male    41.6%  59.9%
    Female  58.4%  40.1%
_____________________________________
```

Especificando type=1 obtenemos una tabla de porcentajes. En este caso, porcentaje de hombres y mujeres según la evolución (evento).

El p-valor indica una diferencia entre hombres y mujeres.

Tablas de OR

```
res <- compareGroups(event ~ age + sex + smoke + waist ,
                     data = predimed)
restab<-createTable(update(res, subset = group != "Control"), show.ratio = TRUE)
restab
```

```
--------Summary descriptives table by 'event'---------
```

	No N=4127	Yes N=155	OR	p.ratio	p.overall
Age	66.8 (6.08)	69.5 (6.55)	1.07 [1.05;1.10]	<0.001	<0.001
Sex:					<0.001
Male	1774 (43.0%)	93 (60.0%)	Ref.	Ref.	
Female	2353 (57.0%)	62 (40.0%)	0.50 [0.36;0.70]	<0.001	
Smoking:					<0.001
Never	2540 (61.5%)	70 (45.2%)	Ref.	Ref.	
Current	553 (13.4%)	35 (22.6%)	2.30 [1.50;3.47]	<0.001	
Former	1034 (25.1%)	50 (32.3%)	1.76 [1.21;2.54]	0.004	
Waist circumference	100 (10.5)	101 (10.0)	1.01 [1.00;1.03]	0.168	0.152

Si la variable respuesta es dicotómica (en este caso prescindimos del grupo control), podemos calcular el OR.

Combinar tablas

```
createTable(compareGroups(event ~ group + age + sex, data = predimed),
            show.ratio = TRUE)

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

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

_____________________________________________________________
                          No          Yes      p.overall
                        N=6072       N=252
_____________________________________________________________
Non-modifiable risk factors:
    Age               66.9 (6.14)  69.4 (6.65)  <0.001
    Sex:                                        <0.001
        Male          2528 (41.6%) 151 (59.9%)
        Female        3544 (58.4%) 101 (40.1%)
Modifiable risk factors:
    Body mass index   30.0 (3.81)  29.8 (3.92)   0.378
    Smoking:                                    <0.001
        Never         3778 (62.2%) 114 (45.2%)
        Current       809 (13.3%)  49 (19.4%)
        Former        1485 (24.5%) 89 (35.3%)
_____________________________________________________________
```

Podemos generar distintas tablas y luego combinarlas por filas. En este caso, separamos entre variables que no podemos modificar, age y sex, y variables susceptibles de intervención (BMI y smoke).

Combinar tablas

```
res <- compareGroups(group ~ age + smoke + bmi + htn, data = predimed)
alltab <- createTable(res, show.p.overall = FALSE)
femaletab <- createTable(update(res, subset = sex == "Female"),
                                 show.p.overall = FALSE)
maletab <- createTable(update(res, subset = sex == "Male"), show.p.overall = FALSE)
cbind(ALL = alltab, FEMALE = femaletab, MALE = maletab)
```

	ALL			FEMALE			MALE		
	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	Control N=1230	MedDiet + Nuts N=1132	MedDiet + VOO N=1283	Control N=812	MedDiet + Nuts N=968	MedDiet + VOO N=899
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	68.0 (5.96)	67.4 (5.57)	67.7 (5.84)	66.4 (6.62)	65.8 (6.40)	66.1 (6.61)
Smoking:									
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	1077 (87.6%)	993 (87.7%)	1115 (86.9%)	205 (25.2%)	266 (27.5%)	236 (26.3%)
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	66 (5.37%)	54 (4.77%)	71 (5.53%)	204 (25.1%)	242 (25.0%)	221 (24.6%)
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	87 (7.07%)	85 (7.51%)	97 (7.56%)	403 (49.6%)	460 (47.5%)	442 (49.2%)
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	30.8 (4.20)	30.2 (4.08)	30.4 (3.91)	29.6 (3.45)	29.1 (3.28)	29.2 (3.28)
Hypertension:									
No	331 (16.2%)	362 (17.2%)	396 (18.1%)	168 (13.7%)	147 (13.0%)	179 (14.0%)	163 (20.1%)	215 (22.2%)	217 (24.1%)
Yes	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	1062 (86.3%)	985 (87.0%)	1104 (86.0%)	649 (79.9%)	753 (77.8%)	682 (75.9%)

OR en el diseño de casos y controles

```
res <- compareGroups(event ~ diab, predimed)
restab <- createTable(res, show.ratio=TRUE)
restab
```

Considerando la evolución final de los pacientes (evento), podemos explorar la asociación con una variable mediante el cálculo del OR.

Para los diabéticos tendríamos un IC del OR de [1.55.2.62], indicando una asociación entre ambas situaciones.

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

	No N=6072	Yes N=252	OR	p.ratio	p.overall
Type-2 diabetes:					<0.001
No	3231 (53.2%)	91 (36.1%)	Ref.	Ref.	
Yes	2841 (46.8%)	161 (63.9%)	2.01 [1.55;2.62]	<0.001	

OR en el diseño de casos y controles

```
res <- compareGroups(event ~ ., predimed)
restab <- createTable(res, show.ratio=TRUE)
restab
```

Considerando la evolución de los pacientes, podemos explorar los OR de asociación con las distintas variables (análisis univariante).

Este análisis suele incluirse en los trabajos como paso previo de selección de variables para regresión logística multivariante.

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

	No N=6072	Yes N=252	OR	p.ratio	p.overall
Intervention group:					0.064
Control	1945 (32.0%)	97 (38.5%)	Ref.	Ref.	
MedDiet + Nuts	2030 (33.4%)	70 (27.8%)	0.69 [0.50;0.95]	0.021	
MedDiet + VOO	2097 (34.5%)	85 (33.7%)	0.81 [0.60;1.09]	0.173	
Sex:					<0.001
Male	2528 (41.6%)	151 (59.9%)	Ref.	Ref.	
Female	3544 (58.4%)	101 (40.1%)	0.48 [0.37;0.62]	<0.001	
Age	66.9 (6.14)	69.4 (6.65)	1.07 [1.04;1.09]	<0.001	<0.001
Smoking:					<0.001
Never	3778 (62.2%)	114 (45.2%)	Ref.	Ref.	
Current	809 (13.3%)	49 (19.4%)	2.01 [1.41;2.82]	<0.001	
Former	1485 (24.5%)	89 (35.3%)	1.99 [1.49;2.64]	<0.001	
Body mass index	30.0 (3.81)	29.8 (3.92)	0.98 [0.95;1.02]	0.365	0.378
Waist circumference	100 (10.6)	102 (10.6)	1.01 [1.00;1.03]	0.016	0.017
Waist-to-height ratio	0.63 (0.07)	0.63 (0.07)	3.64 [0.55;23.9]	0.178	0.187
Hypertension:					0.879
No	1047 (17.2%)	42 (16.7%)	Ref.	Ref.	
Yes	5025 (82.8%)	210 (83.3%)	1.04 [0.75;1.48]	0.826	
Type-2 diabetes:					<0.001
No	3231 (53.2%)	91 (36.1%)	Ref.	Ref.	
Yes	2841 (46.8%)	161 (63.9%)	2.01 [1.55;2.62]	<0.001	
Dyslipidemia:					<0.001
No	1645 (27.1%)	101 (40.1%)	Ref.	Ref.	
Yes	4427 (72.9%)	151 (59.9%)	0.56 [0.43;0.72]	<0.001	
Family history of premature CHD:					0.403
No	4694 (77.3%)	201 (79.8%)	Ref.	Ref.	
Yes	1378 (22.7%)	51 (20.2%)	0.87 [0.63;1.18]	0.363	
Hormone-replacement therapy:					0.186
No	5341 (98.2%)	223 (99.6%)	Ref.	Ref.	
Yes	96 (1.77%)	1 (0.45%)	0.29 [0.01;1.27]	0.117	
MeDiet Adherence score	8.70 (1.94)	8.24 (1.94)	0.89 [0.84;0.95]	<0.001	<0.001
follow-up to main event (years)	4.41 (1.66)	2.96 (1.76)	0.60 [0.55;0.65]	<0.001	<0.001

Conclusión

El paquete ***compareGroups*** es una excelente opción para obtener una descriptiva rápida que compare distintas variables en grupos de interés (p.e. distintas condiciones experimentales, casos-controles, etc.)

Permite controlar distintos aspectos de las tablas de salida y proporciona resultados directamente publicables.

Como limitación, el tratamiento de las variables es univariante, por lo que es útil en una primera descriptiva general, por otra parte muy común en los trabajos médicos.