

T5. Taller encuestas. Comparaciones*

VIII Jornadas R Albacete 2016**

Álvaro Hernández Vicente, Elvira Ferre Jaén, Antonio José Perán Orcajada, Ana Belén
Marín Valverde, Antonio Maurandi López***

17 de noviembre de 2016

Índice

1. Lectura de datos	1
2. Comparación	1
2.1. Sexo	1
2.2. Preguntas Q01 a Q23	2
Referencias y bibliografía	3

1. Lectura de datos

```
load( "saeraq.RData" )
```

Consideramos que casi siempre es más adecuado analizar un cuestionario con estadística multivariante (Wasserstein and Lazar 2016, Krzywinski and Altman (2013)) que únicamente emplando contrastes, no obstante los p-valores siguen estando *muuy demandados*. Así pues solo haremos unas comparaciones en la forma en las que solemos presentarlas para que valgan de ejemplo.

2. Comparación

2.1. Sexo

```
library( stats )
wt <- wilcox.test( df$ingresos ~ df$sexo, alternative = "greater", data = df )

res1 <- data.frame( wt$statistic, wt$p.value, sig = "No" )
names( res1 ) <- c("U Mann-Whitney", "p-valor", "Sig.")
rownames( res1 ) <- "Ingresos vs. Sexo"
# Significacion
if( res1$p-valor < 0.05 ) res1$Sig.<- "Sí"

kable( res1, caption = "U de Mann-Whitney. Ingresos vs. Sexo", digits = 2 )
```

* doc:T5_comparacion.Rmd
** <http://r-es.org/8jornadasR/>
*** Servicio de Apoyo Estadístico; alvarohv@um.es, elvira@um.es, antoniojose.peran@um.es, anabelen.marin4@um.es, amaurandi@um.es



Tabla 1: U de Mann-Whitney. Ingresos vs. Sexo

	U Mann-Whitney	p-valor	Sig.
Ingresos vs. Sexo	619428.5	0.53	No

2.2. Preguntas Q01 a Q23

En este caso queremos hacer una tabla con las comparaciones de todos los ítems del cuestionario por algún factor.

Para ello definimos una función que aplicaremos sobre todos los ítems mediante la función `lapply()`.

```
# Convertimos las variables en numéricas
dfcomp <- lapply( df[, 7:ncol( df ) ], as.numeric )

compU <- function( varnum, vfactor){
  wt <- wilcox.test( varnum ~ vfactor, alternative = "greater" )
  res <- data.frame( wt$statistic, wt$p.value, sig = "No" )
  names( res ) <- c("U Mann-Whitney", "p-valor", "Sig.")
  # Significación
  if( res$p-valor < 0.05) res$Sig.<- "Sí"
  return( res )
}
```

Aplicamos la función que hemos creado sobre los ítems

```
tab <- lapply( dfcomp, compU, vfactor = df$sexo )
# str( tab )
```

Observamos que el objeto `tab` es una lista de `data.frames` con los resultados de cada comparación. Aplicamos la función `ldply()` del paquete `plyr` para juntar las comparaciones en un único `data.frame`.

```
tabla <- ldply( tab )
colnames( tabla )[1] <- "Pregunta"

kable( tabla , digits = 2, caption = "Comparaciones de los ítems por Sexo" )
```

Tabla 2: Comparaciones de los ítems por Sexo

Pregunta	U Mann-Whitney	p-valor	Sig.
Q01	892964.5	0.00	Sí
Q02	588591.5	0.99	No
Q03	375815.5	1.00	No
Q04	940498.5	0.00	Sí
Q05	895747.0	0.00	Sí
Q06	1009009.5	0.00	Sí
Q07	1051862.0	0.00	Sí
Q08	800970.0	0.00	Sí
Q09	625444.5	0.38	No
Q10	842734.5	0.00	Sí
Q11	840301.0	0.00	Sí
Q12	976111.0	0.00	Sí
Q13	1009828.0	0.00	Sí
Q14	998926.0	0.00	Sí
Q15	907249.0	0.00	Sí



Pregunta	U Mann-Whitney	p-valor	Sig.
Q16	928713.5	0.00	Sí
Q17	864043.0	0.00	Sí
Q18	1053246.0	0.00	Sí
Q19	502095.5	1.00	No
Q20	753866.5	0.00	Sí
Q21	941697.5	0.00	Sí
Q22	607627.0	0.79	No
Q23	703725.0	0.00	Sí

Referencias y bibliografía

Krzywinski, Martin, and Naomi Altman. 2013. "Points of Significance: Importance of Being Uncertain" 10 (9): 809–10. <http://dx.doi.org/10.1038/nmeth.2613>.

Wasserstein, Ronald L, and Nicole A Lazar. 2016. "The Asa's Statement on P-Values: Context, Process, and Purpose." *The American Statistician*. Taylor & Francis.