

Técnicas de Análisis Cuantitativas y Cualitativas

Resolución del ejercicio de evaluación 4

Marcos Rial Docampo

12 de enero de 2016

Disponemos de datos de los tiempos obtenidos por los participantes de una carrera popular. Se trata de 80 muestras de un total de 1140 que están estratificadas por sexo y categoría de edad.

Mediante un test de Shapiro-Wilk comprobamos si la variable *total.minutos* tiene una distribución normal. Para ello podemos aplicar el test a todas las observaciones de la variable o dividirla en grupos. En este caso dividiremos la variable en los cuatro grupos correspondientes a la categoría de la prueba (infantil-cadete, junior, senior y veteranos), pero podríamos haberlo hecho por sexo o por subcategoría. Obtenemos los resultados mostrados en el cuadro 1 y los gráficos de la figura 1 donde se indica que la variable *total.minutos* tiene una distribución normal en cada uno de los grupos de edad. En el caso del cuadro vemos que el p-valor supera el 5 % en todos los casos, lo que nos permite admitir la hipótesis nula H_0 de que existe una distribución normal de los datos. Lo mismo que se observa en las gráficas donde los valores de las observaciones se adaptan a la recta. En las gráficas no se observa desviaciones de linealidad respecto a la recta, sobre todo en el centro, lo que demuestra la normalidad de las observaciones para cada grupo.

Test Shapiro-Wilk		
	W	p-valor
Infantil-Cadete	0,93455	0,1888
Junior	0,93301	0,1765
Senior	0,96890	0,7315
Veterano	0,96757	0,7029

Cuadro 1: Resultado del test de Shapiro-Wilk para la variable “total.minutos”.

Es interesante saber si la variable sexo tiene influencia sobre los tiempos obtenidos por los participantes de la prueba. Al ser el sexo de los participantes un factor de dos niveles, podemos realizar un contraste de hipótesis mediante un test de Welch. Obtenemos los resultados del cuadro 2 y la figura 2 donde vemos que la media del tiempo empleado por las mujeres es casi 9 minutos superior al de los hombres. Además, un p-valor de

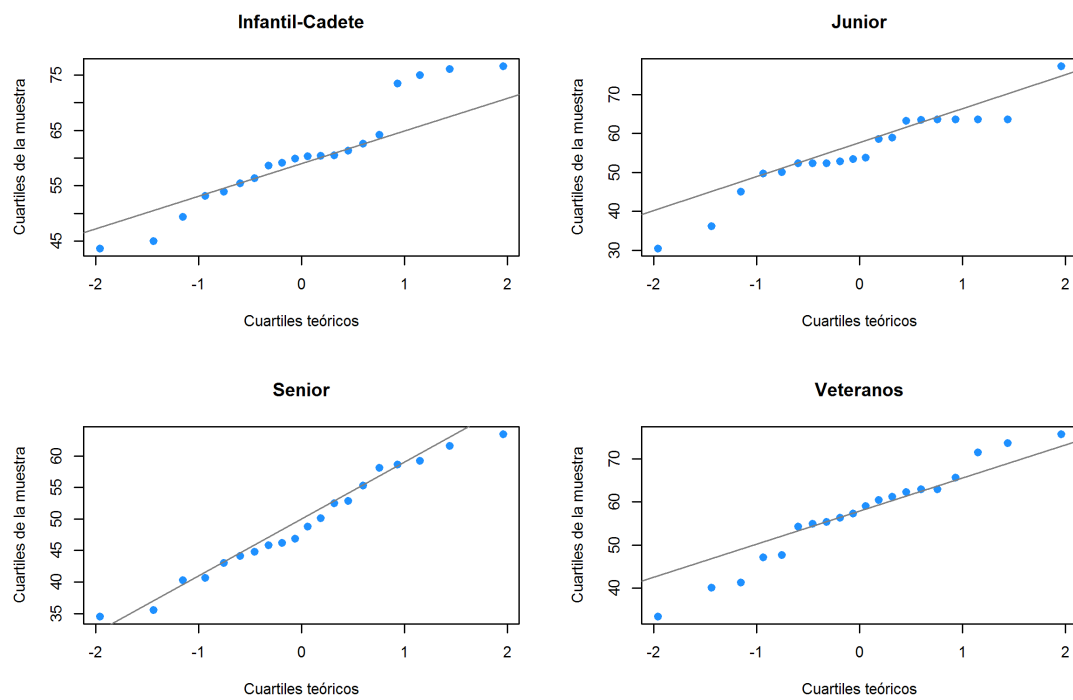


Figura 1: Diagramas de cuantiles para las categorías de la muestra.

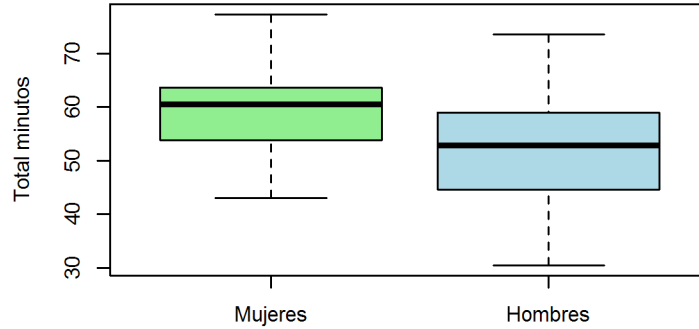


Figura 2: Representación gráfica de la diferencia en tiempos respecto al sexo de los participantes.

$7,405e^{-05}$ sugiere que la interpretación anterior de la diferencia de las medias es correcta. Realizando un análisis de varianza de un factor habríamos obtenido un resultado similar.

Sexo	Tiempo empleado (minutos)
Femenino	59,93
Masculino	50,96

Cuadro 2: Resultados de las medias de tiempo por sexo de los participantes en la prueba.

Analizamos ahora la influencia que tienen los factores sexo y categoría sobre la variable tiempo en minutos. Primero analizamos la influencia de los dos factores por separado con una hipótesis nula (H_0) de que no afectan estos factores a los tiempos obtenidos por los participantes. Obtenemos los resultados del cuadro 3 donde vemos que ambos factores influyen por separado en los tiempos de la prueba deportiva.

Factor	$\Pr(>F)$
Sexo	$2,09e^{-05}$
Categoría	0,0016

Cuadro 3: Resultado del análisis de varianza para los factores sexo y categoría.

Una vez analizados por separado procedemos a hacerlo conjuntamente teniendo en cuenta la interacción de ambos factores. En este caso el resultado que nos da es que conjuntamente no afectan al tiempo total de los participantes. Con un $\Pr(>F)$ obtenido de 0,2478 aceptamos la H_0 . Observamos en la gráfica de interacción de la figura 3 que efectivamente los tiempos obtenidos por participantes femeninos están por encima de los

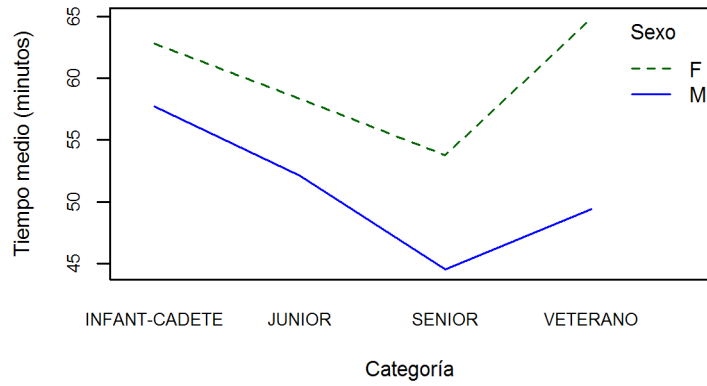


Figura 3: Gráfica de interacción entre los factores sexo y categoría con el tiempo de los participantes.

masculinos en todas las categorías. En cambio en las categorías los tiempos se mezclan en cuanto al sexo como se muestra en la figura 4 en la categoría de veteranos.

Mediante un test de varianzas comprobamos la igualdad de estas para cada factor. Es decir, que gracias a este test y con una H_0 igual a que los factores estudiados presentan varianzas iguales comprobaremos la condición de homocedasticidad de estos con la variable *total.minutos*.

Para el caso del factor sexo esta condición de homocedasticidad se cumple ya que el test devuelve un p-valor de 0,8756. En el caso de las categorías y aplicando un test de Bartlett (debido a que este factor tiene más de dos niveles) el resultado también nos dice que existe homocedasticidad con un p-valor de 0,6638.

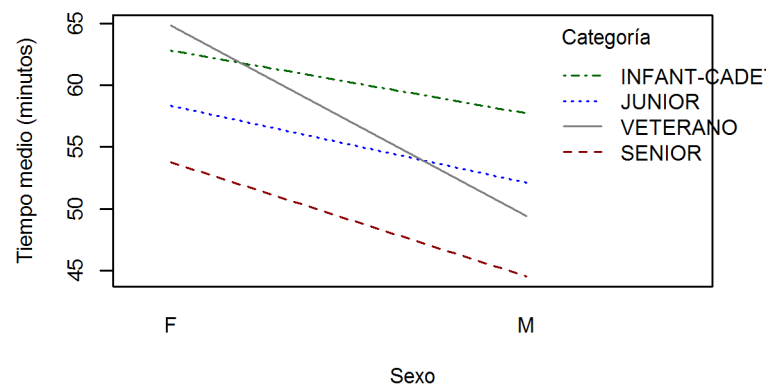


Figura 4: Gráfica de interacción entre los factores categoría y sexo con el tiempo de los participantes.