Ejercicio 3

¿Existen diferencias en la duración promedio de las campañas que siguen las personas casadas contra las que siguen las personas solteras?

Sugerencias: Asuma que todas las observaciones provienen de una distribución normal

Para solucionar el problema deberíamos comparar la media de la duración de las campañas que siguen las personas casadas contra las que siguen las solteras.

Pero si analizamos el problema podemos percatarnos que estamos en presencia de dos muestras de una población en la que se quiere comparar los valores de parámetros de estas muestras, en nuestro caso la duración de las campañas, y para esto en la práctica se realizan las pruebas de hipótesis con comparación como vimos en la conferencia 3.

Ahora teniendo en cuenta que no conocemos la varianza, antes necesitamos realizar una prueba de igualdad contra la diferencia en las varianzas y dependiendo del resultado de la misma realizamos la prueba de medias con la varianza desconocida correspondiente.

Entonces:

Sea y las varianzas correspondientes a la duración de las campañas que siguen las personas solteras y casadas.

Planteamos, la prueba de hipótesis para la comparación de las varianzas

Calculamos los estimadores puntuales para las varianzas:

Luego fijamos un el cual será nuestro nivel de significación.

Calculamos el estadígrafo

Elegimos la región crítica

o ,

Siendo los tamaños de la muestra de personas solteras y casadas correspondiente.

Y luego comprobamos si es posible rechazar la hipótesis nula para α=0.05

Ahora, para facilitar los cálculos, en R contamos con una función var.test() que recibe como parámetros las muestras y nos devuelve un p-value el cual es el menor nivel de significación para el cual la hipótesis nula es rechazada y si además este p-value es menor que α=0.05 entonces se cumple la región crítica.

Al calcularlo obtenemos un p-value de 4.262e-08 = 0.000000043 < 0.05 = α , por lo que se cumple que las varianzas son iguales para un nivel de significación de 0.05.

Luego pasamos a realizar el de las medias con varianzas desconocidas iguales.

Sean y los parámetros de las poblaciones normales de personas casadas y solteras correspondientes

Planteamos la prueba de hipótesis para este caso

Calculamos los estimadores puntuales para las varianzas

El estadígrafo quedaría de la siguiente manera:

Elegimos como región crítica

En R podemos encontrar una función llamada t.test() el cual recibe las muestras a analizar, el valor calculado en la varianza y una alternativa (parámetro opcional), en nuestro caso le pasamos “less”.

Comparamos el p-value obtenido 2.296602e-07 = 0.0000002297 < 0.05 = α.

El p-value es menor, por tanto, podemos decir que se rechaza la hipótesis nula. Esto quiere decir en términos de nuestro problema que el promedio de duración de las campañas que siguen las personas casadas es menor que el que siguen las personas solteras, con un nivel de significación del 5%.