**Reporte exploración de datos**

1. **Prueba de Hipótesis**

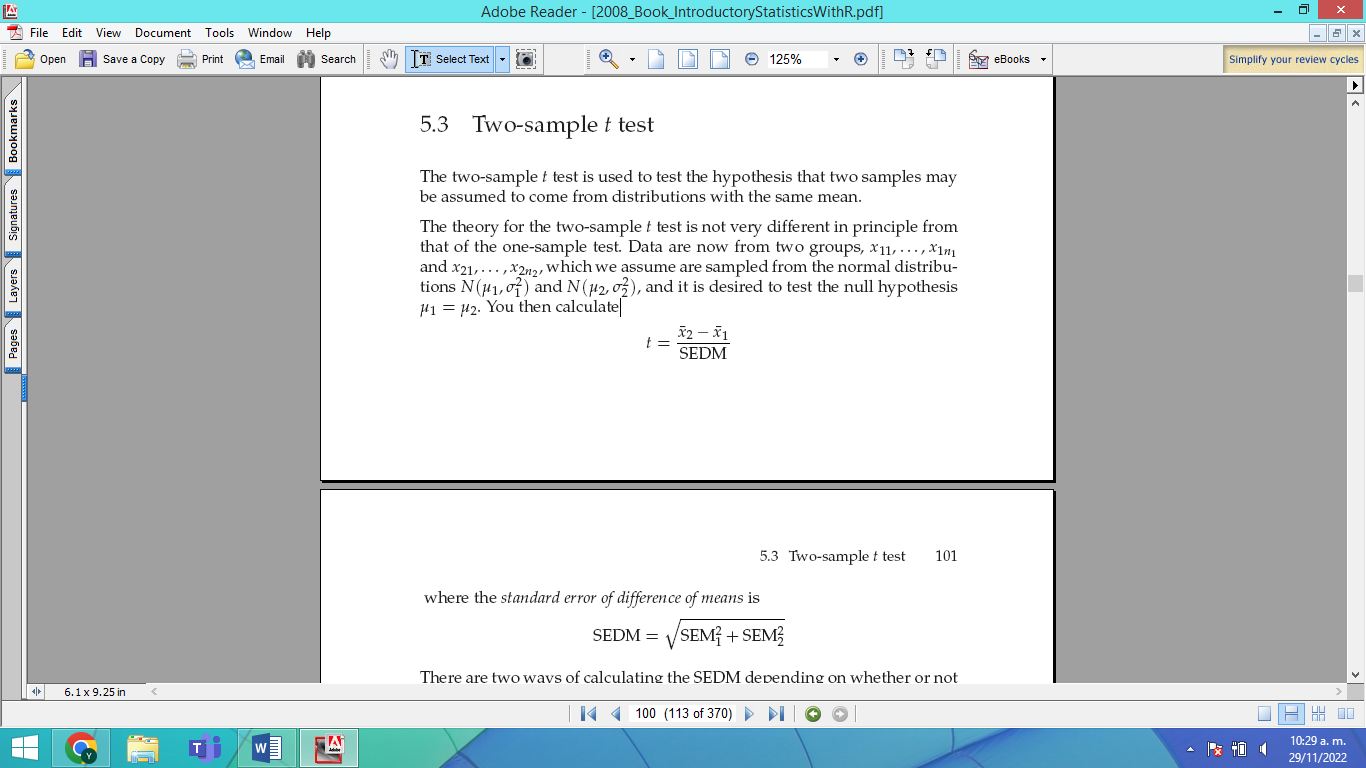
T- Student para dos muestras independientes

La prueba t de dos muestras se utiliza para probar la hipótesis de que dos muestras pueden suponer que provienen de distribuciones con la misma media.

La teoría de la prueba t de dos muestras no es muy diferente en principio de la prueba de una muestra. Los datos proceden ahora de dos grupos, x11, . . . , x1n1 y x21, . . . , x2n2 , que suponemos que se muestrean a partir de las distribuciones normales

N(µ1, σ21 ) y N(µ2, σ22 ), y se desea comprobar la hipótesis nula µ1 = µ2.

Se calcula entonces



Se muestra la tabla del proyecto la cual contiene Librería, Memoria y Tiempo que se utilizaran para esta prueba T- Student

1. **Supuestos Validar**

**Homogeneidad de varianzas**

Conocido como supuesto de homocedasticidad, considera que la varianza es constante (no varía) en los diferentes niveles de un factor, es decir, entre diferentes grupos.

**Se aplica para saber si las varianzas no varían entre los dos grupos**

**Normalidad**

Solo es aplicable si se puede asumir que las dos poblaciones que se comparan se distribuyen de forma normal. La normalidad se tiene que cumplir en las poblaciones. Sin embargo, cuando no se dispone de información sobre las poblaciones, la única forma de estimar su distribución es a partir de las muestras.

**Solo se aplica si se distribuyen igual o normal sino solo se obtiene con muestras**

**Análisis**

1. Prueba de Hipotesis para dos grupos

Un test de hipótesis para dos muestras es similar en muchos aspectos al test para una muestra. Se especifica una hipótesis nula, en la mayoría de los casos se propone que las medias de las dos poblaciones son iguales y se establece la hipótesis alternativa (uni o bilateral).

**Para saber si dos muestras son similares**

1. Analisis de covarianza

El análisis de la covarianza (ANCOVA) se trata de dos o más variantes medidas y donde cualquier variable independiente mesurable no se encuentra a niveles predeterminados, como en un experimento factorial (Badii & Castillo, 2007, Badii et al., 2007a).

Y= M + Ti + B (Xi)

**Experimento Factorial**

**Diagrama de Dispersión**

También conocido como gráfico de dispersión, gráfico de puntos, diagrama de XY, diagrama de dispersión o Scattergram.

Los diagramas de dispersión usan una colección de puntos colocados usando coordenadas cartesianas para mostrar valores de dos variables. Al mostrar una variable en cada eje, se puede detectar si existe una relación o correlación entre las dos variables.

Se pueden interpretar varios tipos de correlación a través de los patrones mostrados en los diagramas de dispersión. Estos son: positivo (los valores aumentan juntos), negativo (un valor disminuye a medida que el otro aumenta), nulo (sin correlación), lineal, exponencial y en forma de U. La fuerza de la correlación puede determinarse por la proximidad de los puntos entre sí en el gráfico. Los puntos que terminan muy lejos del conjunto general de puntos se conocen como valores atípicos.

https://datavizcatalogue.com/ES/metodos/diagrama\_de\_dispersion.html

**Coeficientes de Correlación de Pearson**

El tamaño no influye relativo al tiempo y memoria

**Mann Whitney**

La prueba U de Mann-Whitney resulta útil si tenemos dos muestras independientes y queremos si hay una diferencia en la magnitud de la variable que estamos estudiando, pero no podemos usar la prueba de t independiente o la prueba de z porque los datos no cumplen con alguno de los requisitos. Para realizar la prueba U de Mann-Whitney ponemos las observaciones de las dos muestras en orden ascendiente y asignamos un rango ordinal de manera que 1 corresponde a la observación de menor magnitud, 2 a la segunda etcétera. Luego nos fijamos en las diferencias entre las observaciones.

La prueba se basa en una comparación de cada observación de una muestra xixi con cada observación en la segunda muestra yjyj. Si las muestras tienen la misma mediana, entones cada observación tiene un 0,5 (50%) de chance de ser mayor o menor que la observación correspondiente de la otra muestra. Por tanto plantea las hipotesis:

**H0:P(xi>yj)=12H0:P(xi>yj)=12**

**H1:P(xi>yj)≠12H1:P(xi>yj)≠12**

**Tabla con mediciones de la primera tarea denominada “transformación pdf a txt”**

Ley: Cantidad de leyes las cuales son nueve

Libreria 1: PyPDF2

Libreria2: Pdfplumber

Tiempo: Preprocesamiento que tardo en milisegundos

Memoria: Preprocesamiento que consumio en bytes

Tamaño: Cantidad de hojas del documento de ley

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ley | Librería | Tiempo | Memoria | Tamaño |
| 1 | 1 | 28.67193 | 78 | 128 |
| 2 | 1 | 37.67446 | 89.6 | 112 |
| 3 | 1 | 11.1584 | 65.9 | 73 |
| 4 | 1 | 14.13138 | 63.6 | 71 |
| 5 | 1 | 15.80249 | 65 | 72 |
| 6 | 1 | 19.69258 | 92.9 | 50 |
| 7 | 1 | 14.2578 | 87.9 | 56 |
| 8 | 1 | 17.43217 | 74.7 | 49 |
| 9 | 1 | 43.59123 | 94.3 | 64 |
| 1 | 2 | 113.4356 | 89 | 128 |
| 2 | 2 | 147.1396 | 94.5 | 112 |
| 3 | 2 | 44.94758 | 76.4 | 73 |
| 4 | 2 | 57.60759 | 78.8 | 71 |
| 5 | 2 | 53.67049 | 79 | 72 |
| 6 | 2 | 61.83369 | 94.7 | 50 |
| 7 | 2 | 123.8179 | 93.2 | 56 |
| 8 | 2 | 58.25793 | 91.4 | 49 |
| 9 | 2 | 47.17654 | 84 | 64 |

A continuación se muestra una gráfica con diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría. Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente. Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico.

Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero(recordemos que el segundo cuartil coincide con la mediana). Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las lineas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

**Gráfica donde se muestra el grafico de caja y alambre con la Librería1 y Libreria2 en consumo de Tiempo**

Se muestra que la librería 1 consume un menor tiempo de preprocesamiento que la librería 2 ya que como resultados la librería 2 tiene una mayor calidad en la transformación de pdf a txt con lo cual consume mayor tiempo



**Gráfica donde se muestra el gráfico de caja y alambre con la Librería1 y Libreria2 en consumo de Memoria**

Se muestra que la librería 1 consume más memoria que la librería 2 ya que preprocesa los datos en todos los documentos de ley con mejor resultados y calidad en los datos al hacer la transformación a txt



**Gráfica donde se muestra el grafico de caja y alambre con la Librería1 y Libreria2 viendose en consumo de tiempo con relación al tamaño del documento**

Se muestra la librería 1 con el punto azul y librería 2 con un rectangulo rojo lo cual los puntos azules no tienen algun cambio y se mantiene lineal con el tamaño y tiempo de este preprocesamiento en cambio la librería 2 si se muestra diferencia muy mayor en los tamaños de documentos conforme al tiempo con lo cual no tiene una correlación



**Gráfica donde se muestra el grafico de caja y alambre con la Librería1 y Libreria2 viendose en consumo de memoria con relación al tamaño del documento**

Se muestra la librería 1 con el punto azul y librería 2 con un rectangulo rojo lo cual los puntos azules van de abajo hacia arriba y se mantiene lineal con el tamaño y memoria de este preprocesamiento en cambio la librería 2 si se muestra diferencia en los tamaños de documentos conforme a la memoria con lo cual no tiene una correlación

Aun no se verifica si es por número de imagenes, tablas, texto o alguna diferencia cualitativa que contenga la ley a analizar



**Pruebas de Hipótesis**

**Memoria y Tiempo**

T- Student para dos muestras independientes

La prueba t de dos muestras se utiliza para probar la hipótesis de que dos muestras pueden suponer que provienen de distribuciones con la misma media.

La teoría de la prueba t de dos muestras no es muy diferente en principio de la prueba de una muestra. Los datos proceden ahora de dos grupos, x11, . . . , x1n1 y x21, . . . , x2n2 , que suponemos que se muestrean a partir de las distribuciones normales

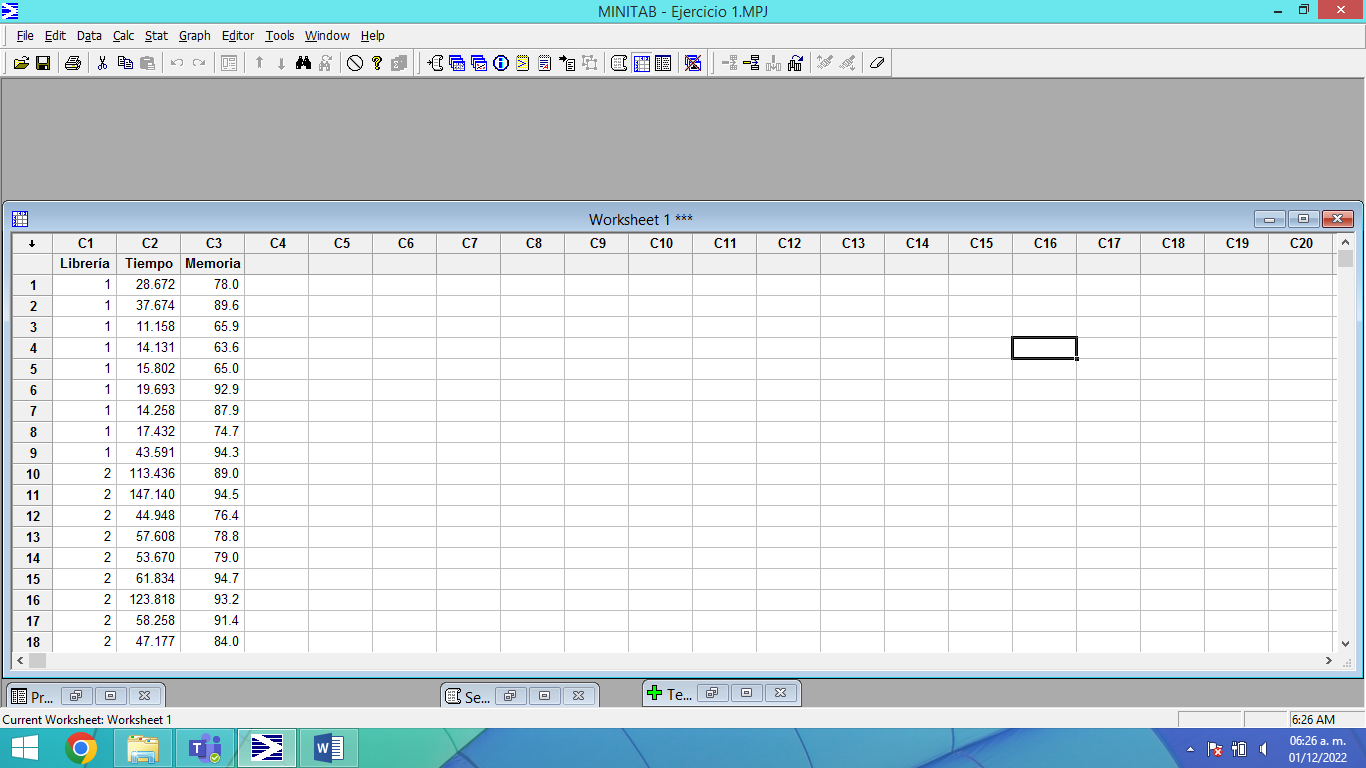
N(µ1, σ21 ) y N(µ2, σ22 ), y se desea comprobar la hipótesis nula µ1 = µ2.

Se muestra la tabla del proyecto la cual contiene Librería, Memoria y Tiempo que se utilizaran para esta prueba T- Student

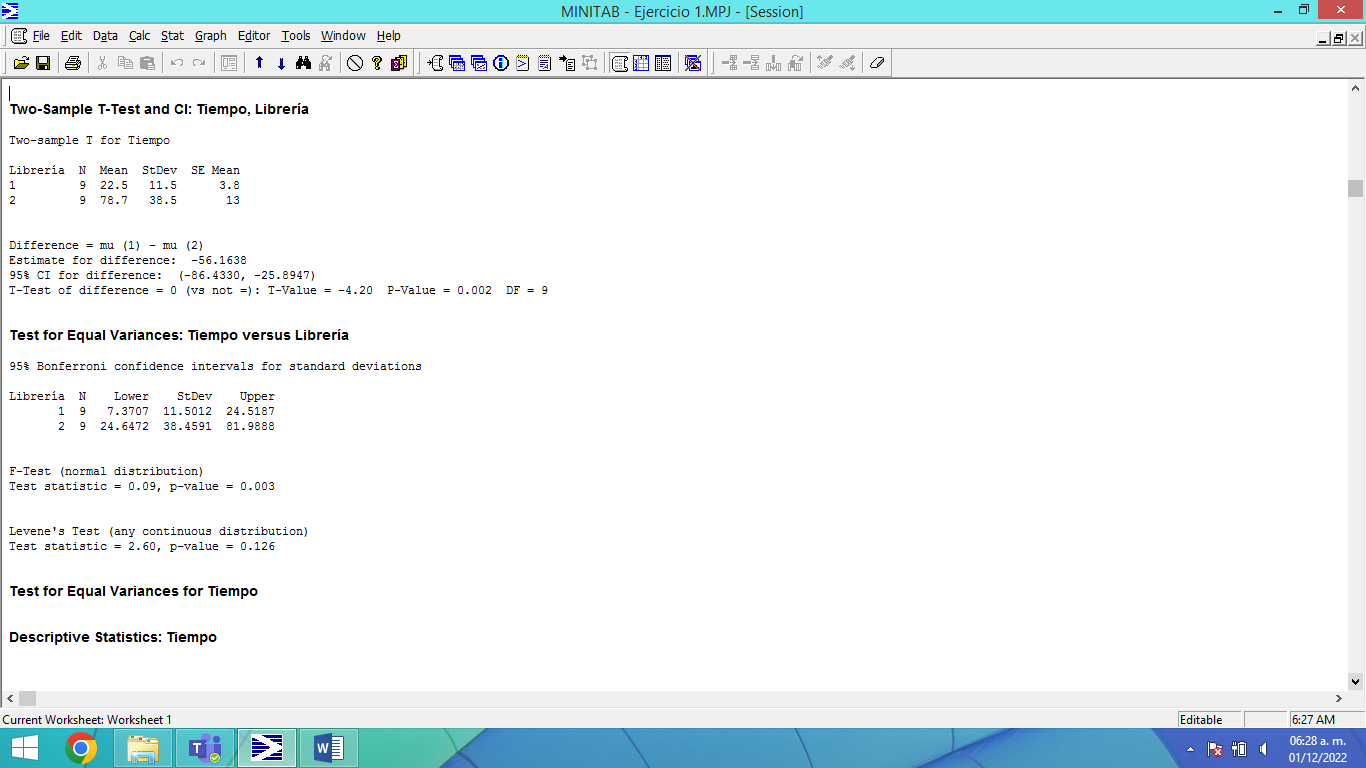
|  |  |  |
| --- | --- | --- |
| Librería | Tiempo | Memoria |
| 1 | 28.67193 | 78 |
| 1 | 37.67446 | 89.6 |
| 1 | 11.1584 | 65.9 |
| 1 | 14.13138 | 63.6 |
| 1 | 15.80249 | 65 |
| 1 | 19.69258 | 92.9 |
| 1 | 14.2578 | 87.9 |
| 1 | 17.43217 | 74.7 |
| 1 | 43.59123 | 94.3 |
| 2 | 113.4356 | 89 |
| 2 | 147.1396 | 94.5 |
| 2 | 44.94758 | 76.4 |
| 2 | 57.60759 | 78.8 |
| 2 | 53.67049 | 79 |
| 2 | 61.83369 | 94.7 |
| 2 | 123.8179 | 93.2 |
| 2 | 58.25793 | 91.4 |
| 2 | 47.17654 | 84 |

Se maneja en el software minitab 14 en el cual se introduce la tabla que se quiere probar:

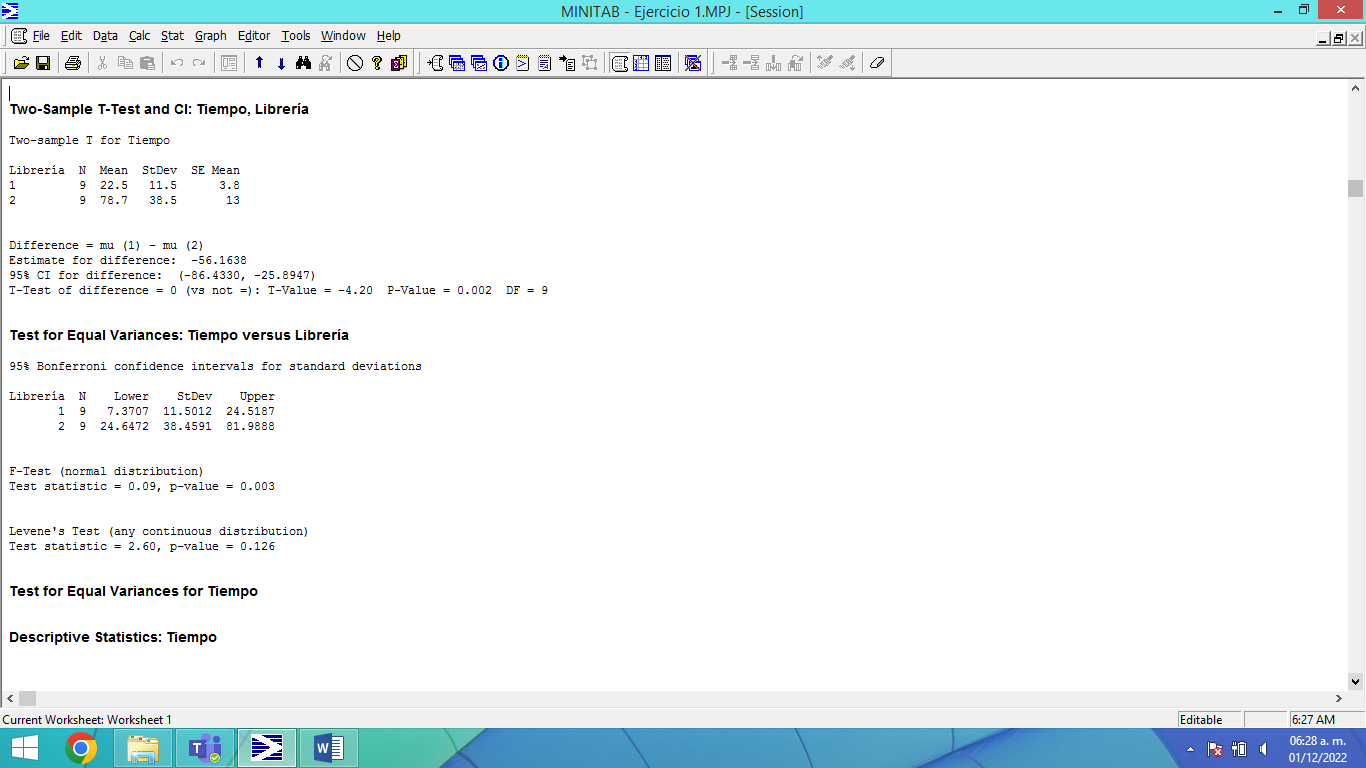
**Tiempo con Librería**

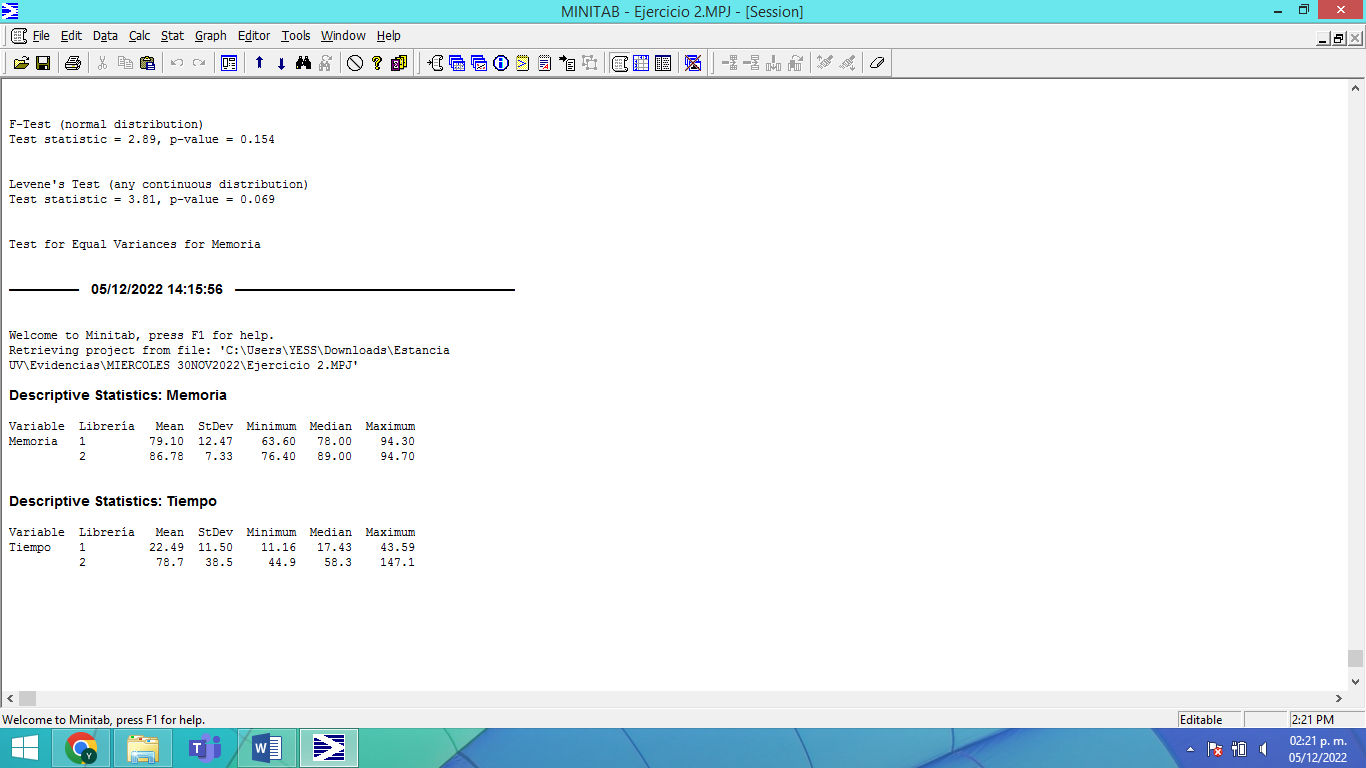


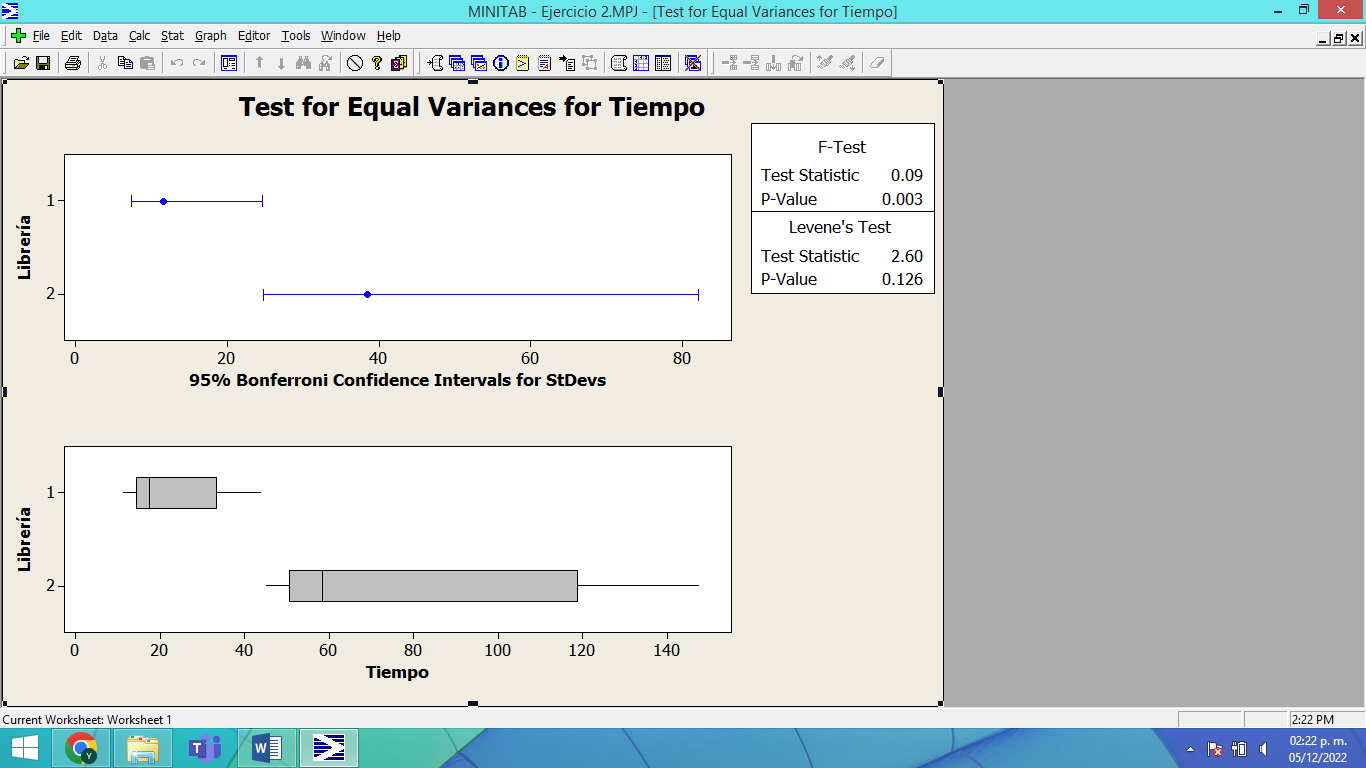
A continuación se aplica el test-t con relación al tiempo con la librería y se obtiene su distribución media y distribución estándar



Como se muestran los resultados la diferencia es del 0.002 en la prueba si fuese mayor a 0.005 se cumpliría esta prueba de varianzas para proceder a hacer las pruebas de homogeneidad de varianzas y normalidad, pruebas de hipótesis y analisis de covarianza así como la prueba de correlación de Pearson pero ya se vio que el tamaño no influye relativo al tiempo y memoria.





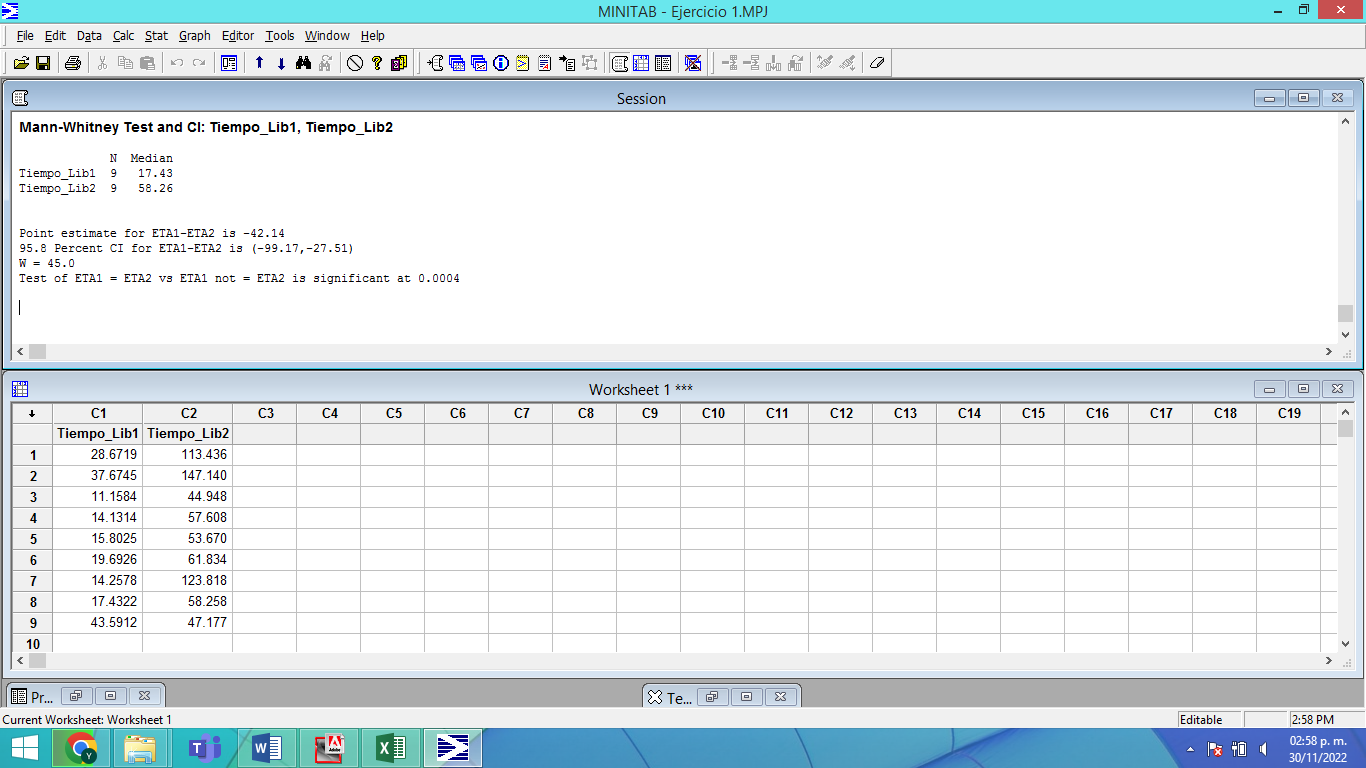


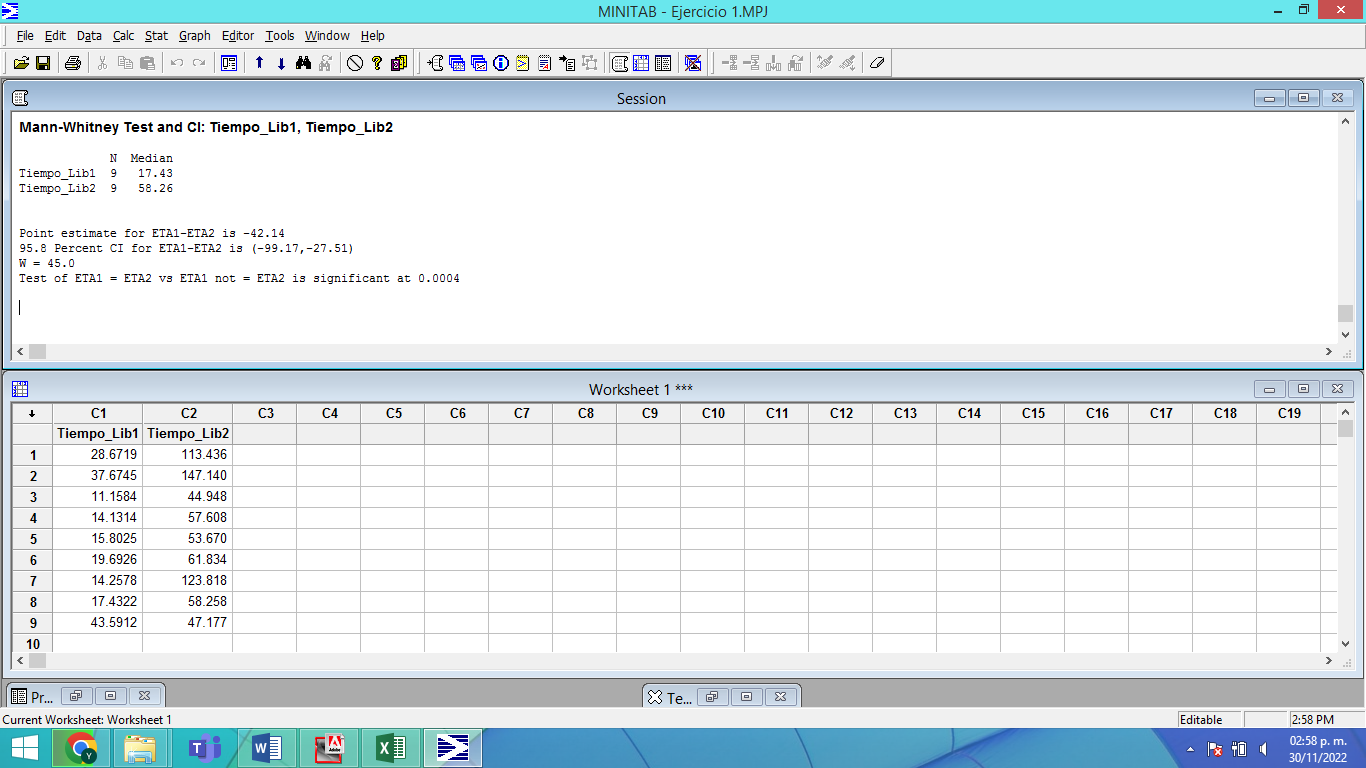
Asi que se procedio a realizar la prueba Mann Whitney la cual resulta útil si se tienen dos muestras independientes y se quiere saber si hay una diferencia en la magnitud de la variable y no no se puede realizar alguna prueba de t independiente o prueba de z porque los datos no cumplen con alguno de los requisitos.

**Mann-Whitney Tiempo Librería 1 y Tiempo Librería 2**

|  |  |  |
| --- | --- | --- |
|  | **Librería 1** | **Librería 2** |
|  | **Tiempo** | **Tiempo** |
|  | 28.67193398 | 113.4356293 |
|  | 37.67445752 | 147.1395946 |
|  | 11.15839875 | 44.94758459 |
|  | 14.13138197 | 57.6075922 |
|  | 15.80249413 | 53.6704862 |
|  | 19.69257731 | 61.83368747 |
|  | 14.25780042 | 123.8178965 |
|  | 17.43216654 | 58.25793318 |
|  | 43.59123231 | 47.17653714 |
| **Totales** | **202.4124429** | **707.8869412** |
| Media | 22.49027144 | 78.65410458 |
| Mediana | 17.43216654 | 58.25793318 |
| Moda | #N/A | #N/A |
| Desviación estandar | 11.50117843 | 38.45909963 |

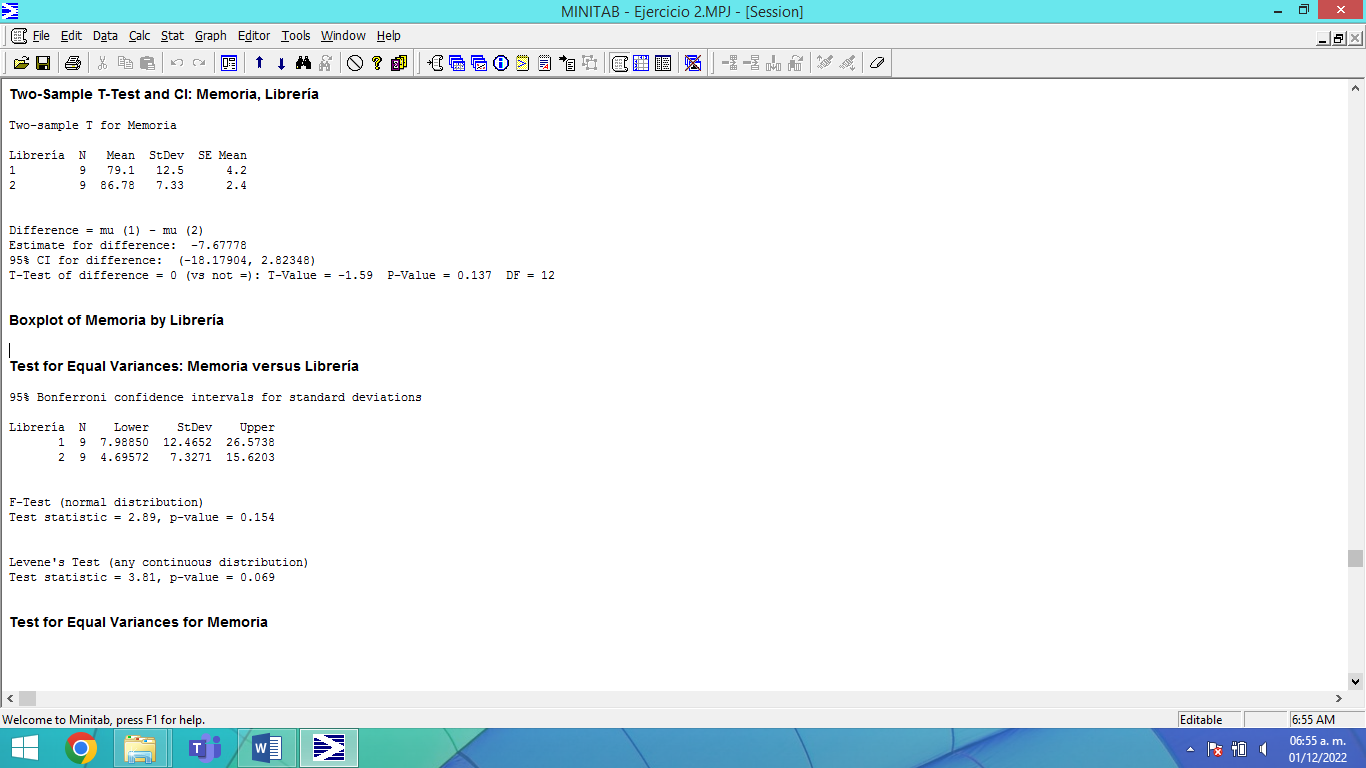
**Ho= T1 = T2**

****

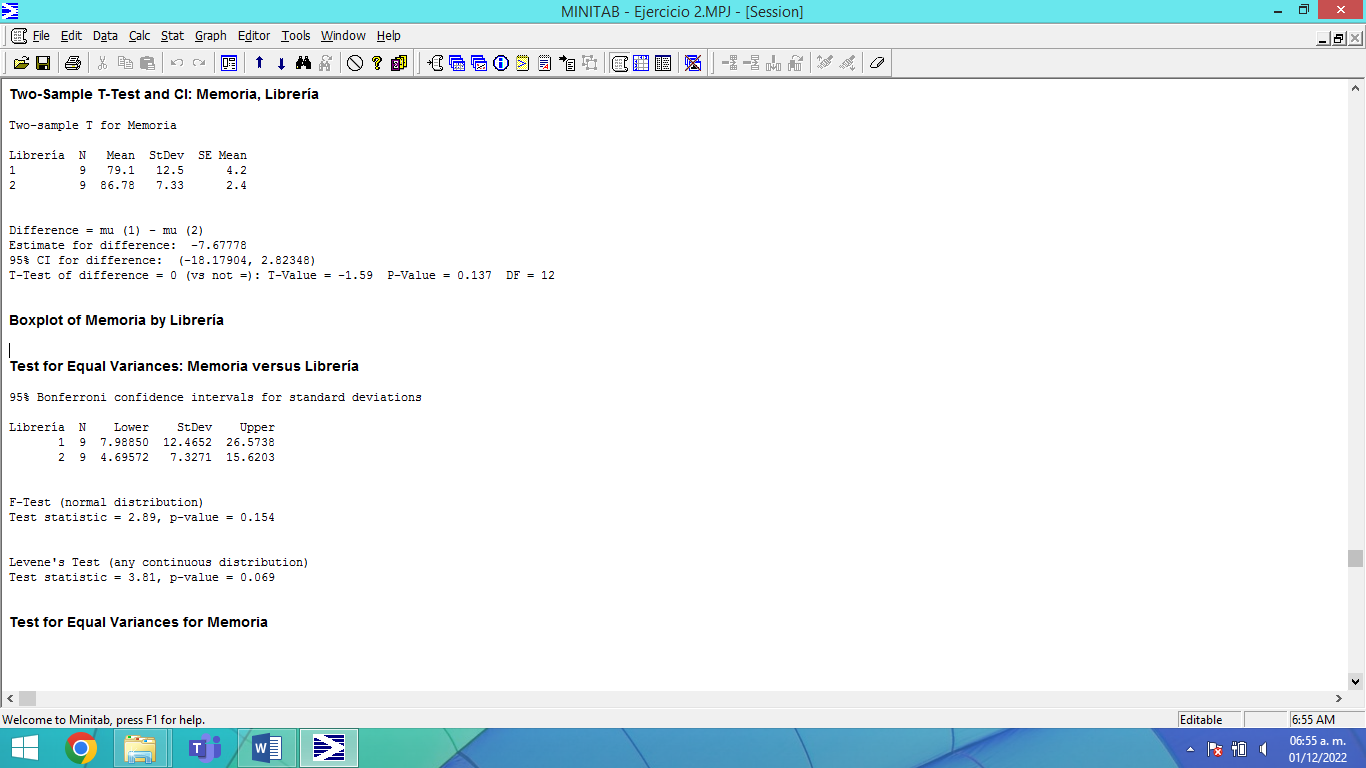
****

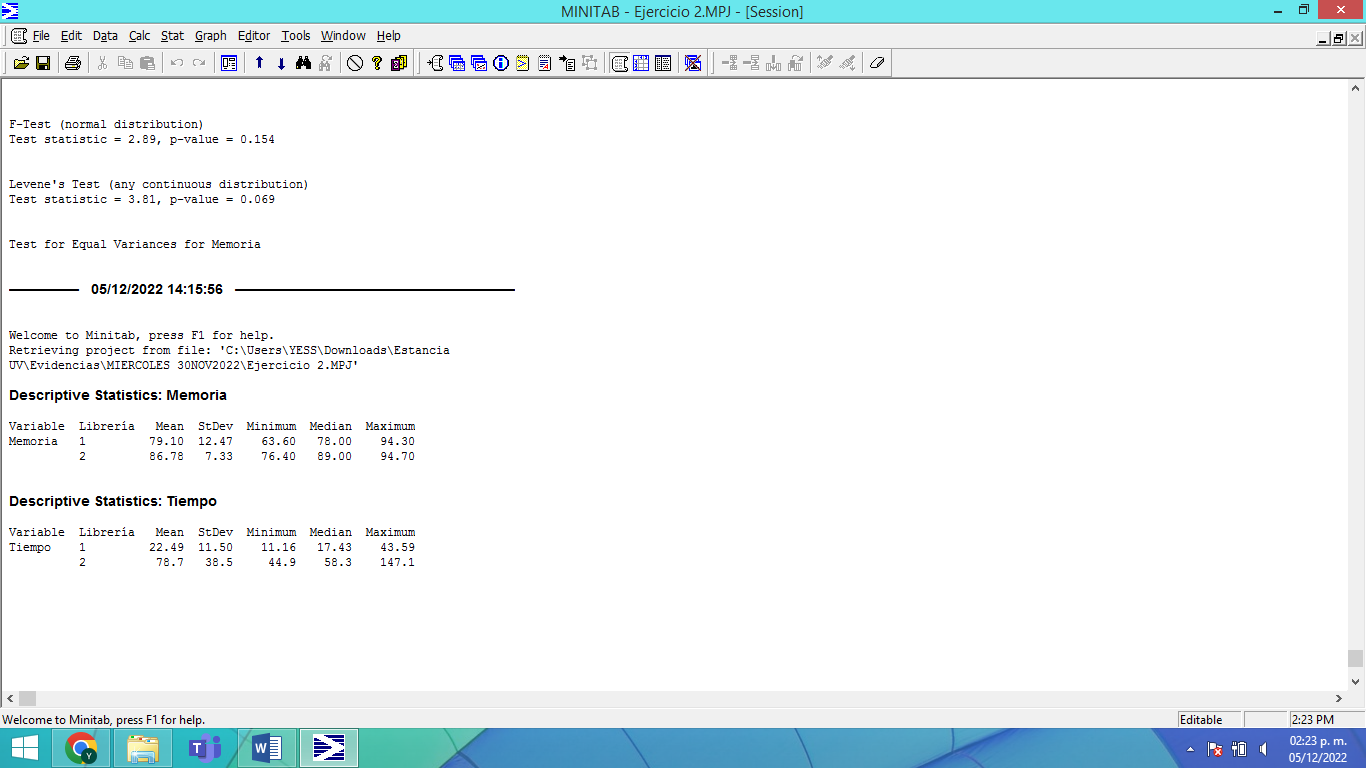
A continuación se aplica el test-t con relación a memoria con la librería y se obtiene su distribución media y distribución estándar

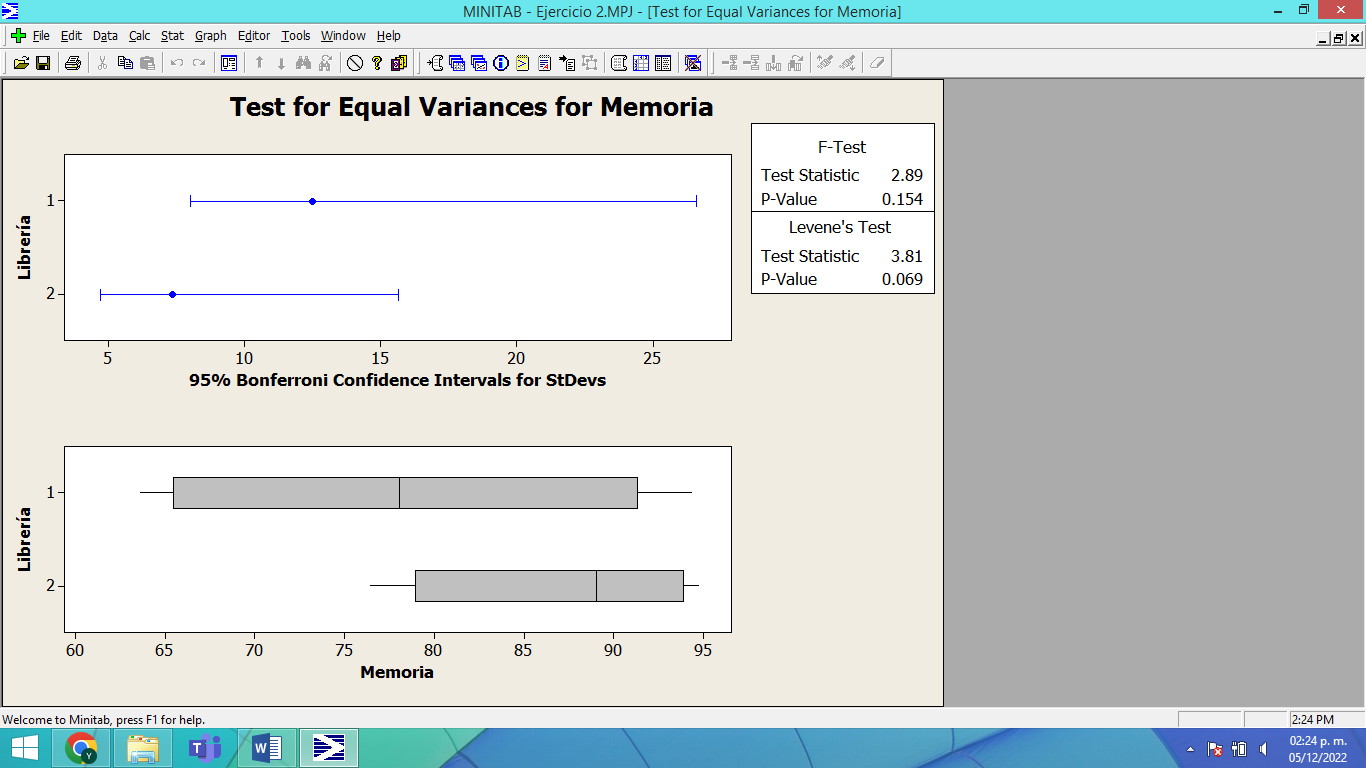
**Memoria con Librería**



Como se muestran los resultados la diferencia es del -1.59 en la prueba si fuese mayor a 0.005 se cumpliría esta prueba de varianzas para proceder a hacer las pruebas de homogeneidad de varianzas y normalidad, pruebas de hipótesis y analisis de covarianza así como la prueba de correlación de Pearson pero ya se vio que el tamaño no influye relativo al tiempo y memoria.





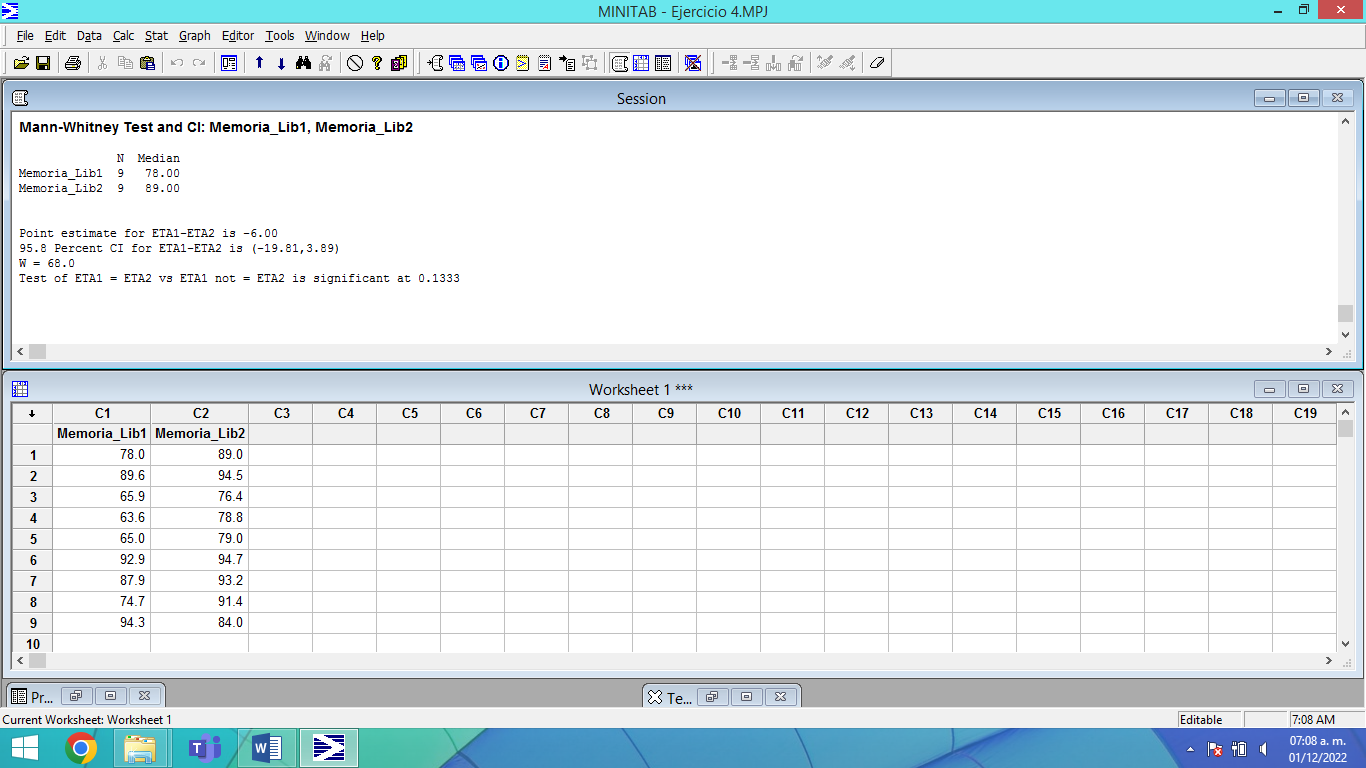


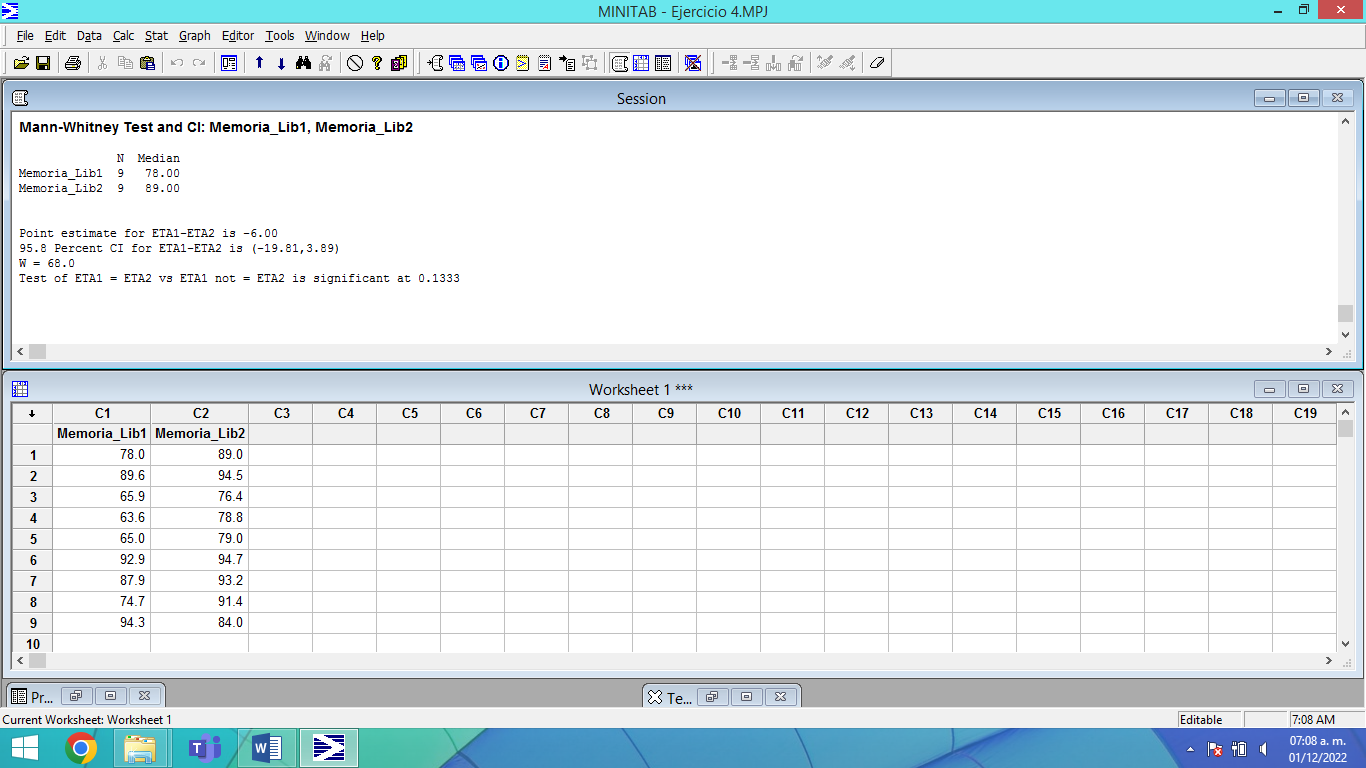
Asi que se procedio a realizar la prueba Mann Whitney la cual resulta útil si se tienen dos muestras independientes y se quiere saber si hay una diferencia en la magnitud de la variable y no no se puede realizar alguna prueba de t independiente o prueba de z porque los datos no cumplen con alguno de los requisitos.

**Mann-Whitney Memoria Libreria1 y Memoria Libreria2**

|  |  |  |
| --- | --- | --- |
|  | **Librería 1** | **Librería 2** |
|  | **Memoria** | **Memoria** |
|  | 78 | 89 |
|  | 89.6 | 94.5 |
|  | 65.9 | 76.4 |
|  | 63.6 | 78.8 |
|  | 65 | 79 |
|  | 92.9 | 94.7 |
|  | 87.9 | 93.2 |
|  | 74.7 | 91.4 |
|  | 94.3 | 84 |
| **Totales** | **711.9** | **781** |
| Media | 79.1 | 86.77777778 |
| Mediana | 78 | 89 |
| Moda | #N/A | #N/A |
| Desviación estandar | 12.46515142 | 7.327137534 |

**Ho= M1 = M2**





El lenguaje de programación que se utilizo es R Project. Las diferencias entre este lenguaje y Python son que R Project es un lenguaje orientado al análisis estadístico que se utiliza ampliamente en el campo de la ciencia de datos, mientras que Python es un lenguaje de alto nivel multipropósito utilizado además en otros campos (desarrollo web, scripting, etc.)

Los códigos para calcular el tiempo de las librerías 1 y 2 se muestran a continuación:

**Tiempo Librería1**

tiempolibreria1<-c(28.67193398,37.67445752,11.15839875,14.13138197,15.80249413,19.69257731,14.25780042,17.43216654,43.59123231)

mean(tiempolibreria1)

#Estimación del error estándar

sd(tiempolibreria1)/sqrt(length(tiempolibreria1))

#Estimación de la diferencia de medias

mean(tiempolibreria1)- mean(tiempolibreria1)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(tiempolibreria1)/length(tiempolibreria1))

#Estimación de las medianas

median(tiempolibreria1)

#Estimación de la diferencia de medianas

median(tiempolibreria1)- median(tiempolibreria1)

B<-5000

vec<-matrix(,B,2)

for(i in 1:B){

d<-sample(tiempolibreria1,9,replace=T)

vec[i,1]<-mean(d)

vec[i,2]<-vec[i,1]

}

thet<-mean(vec[,2])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,1]-thet)^2

i<-i+1

if(i>B){

break}

}

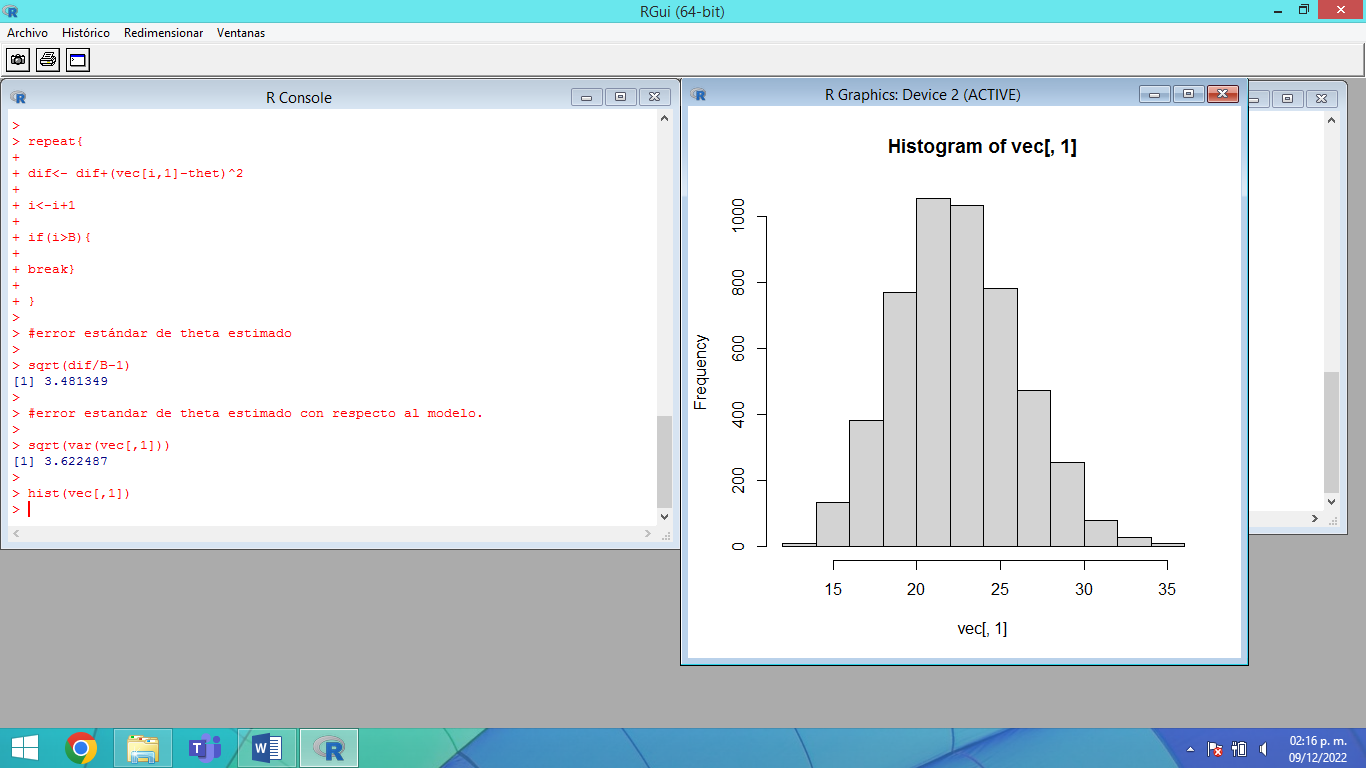
#error estándar de theta estimado

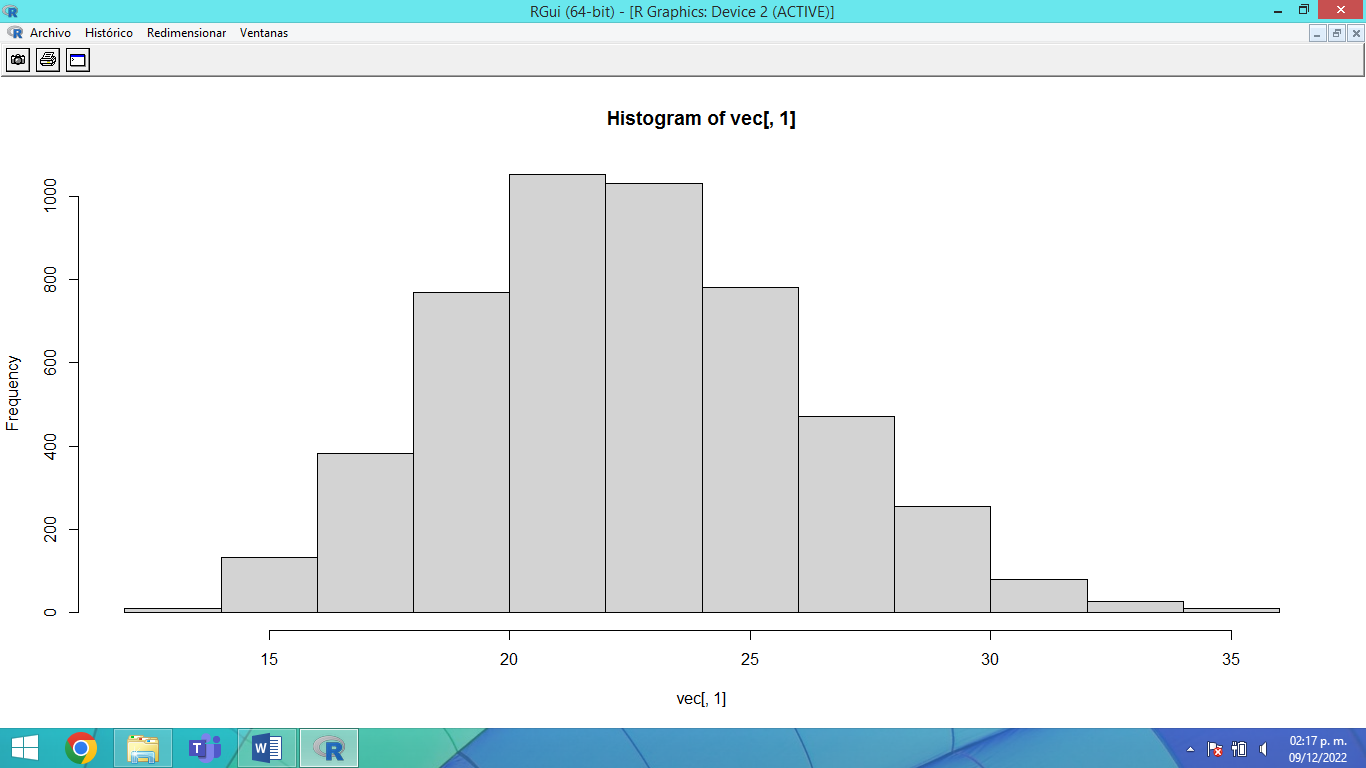
sqrt(dif/B-1)

#error estandar de theta estimado con respecto al modelo.

sqrt(var(vec[,1]))

hist(vec[,1])





**Tiempo Libreria2**

tiempolibreria2<-c(113.4356293,147.1395946,44.94758459,57.6075922,53.6704862,61.83368747,123.8178965,58.2579331,47.17653714)

mean(tiempolibreria2)

#Estimación del error estándar

sd(tiempolibreria2)/sqrt(length(tiempolibreria2))

#Estimación de la diferencia de medias

mean(tiempolibreria2)- mean(tiempolibreria2)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(tiempolibreria2)/length(tiempolibreria2))

#Estimación de las medianas

median(tiempolibreria2)

#Estimación de la diferencia de medianas

median(tiempolibreria2)- median(tiempolibreria2)

B<-5000

vec<-matrix(,B,2)

for(i in 1:B){

h<-sample(tiempolibreria2,9,replace=T)

vec[i,1]<-mean(h)

vec[i,2]<-vec[i,1]

}

thet<-mean(vec[,2])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,1]-thet)^2

i<-i+1

if(i>B){

break}

}

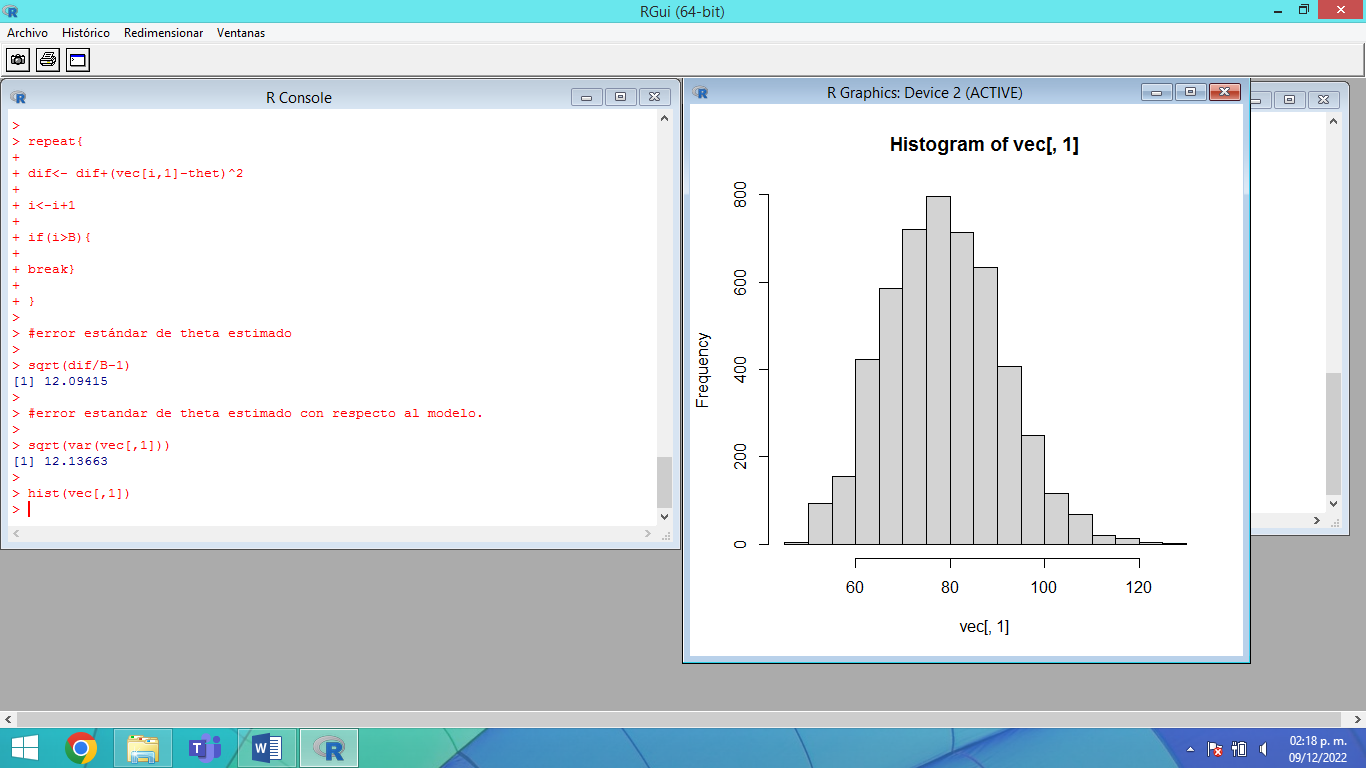
#error estándar de theta estimado

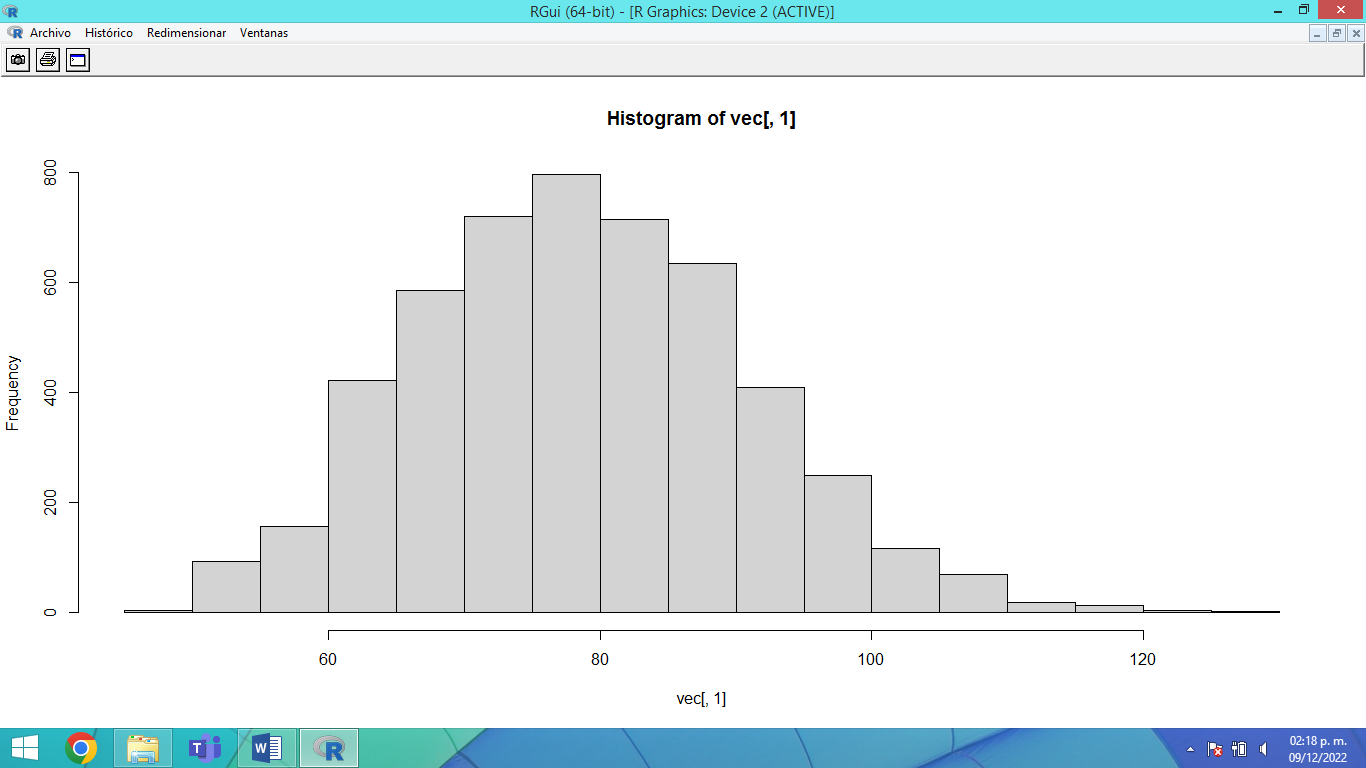
sqrt(dif/B-1)

#error estandar de theta estimado con respecto al modelo.

sqrt(var(vec[,1]))

hist(vec[,1])





**Tiempos entre Librería 1 y Librería 2**

tiempolibreria1<-c(28.67193398,37.67445752,11.15839875,14.13138197,15.80249413,19.69257731,14.25780042,17.43216654,43.59123231)

tiempolibreria2<-c(113.4356293,147.1395946,44.94758459,57.6075922,53.6704862,61.83368747,123.8178965,58.2579331,47.17653714)

mean(tiempolibreria1)

mean(tiempolibreria2)

#Estimación del error estándar

sd(tiempolibreria1)/sqrt(length(tiempolibreria1))

sd(tiempolibreria2)/sqrt(length(tiempolibreria2))

#Estimación de la diferencia de medias

mean(tiempolibreria1)- mean(tiempolibreria2)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(tiempolibreria1)/length(tiempolibreria1)+ var(tiempolibreria2)/length(tiempolibreria2))

#Estimación de las medianas

median(tiempolibreria1)

median(tiempolibreria2)

#Estimación de la diferencia de medianas

median(tiempolibreria1)- median(tiempolibreria2)

B<-5000

vec<-matrix(,B,3)

for(i in 1:B){

d<-sample(tiempolibreria1,9,replace=T)

vec[i,1]<-mean(d)

h<-sample(tiempolibreria2,9,replace=T)

vec[i,2]<-mean(h)

vec[i,3]<-vec[i,1]-vec[i,2]

}

thet<-mean(vec[,3])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,3]-thet)^2

i<-i+1

if(i>B){

break}

}

#error estándar de theta estimado

sqrt(dif/B-1)

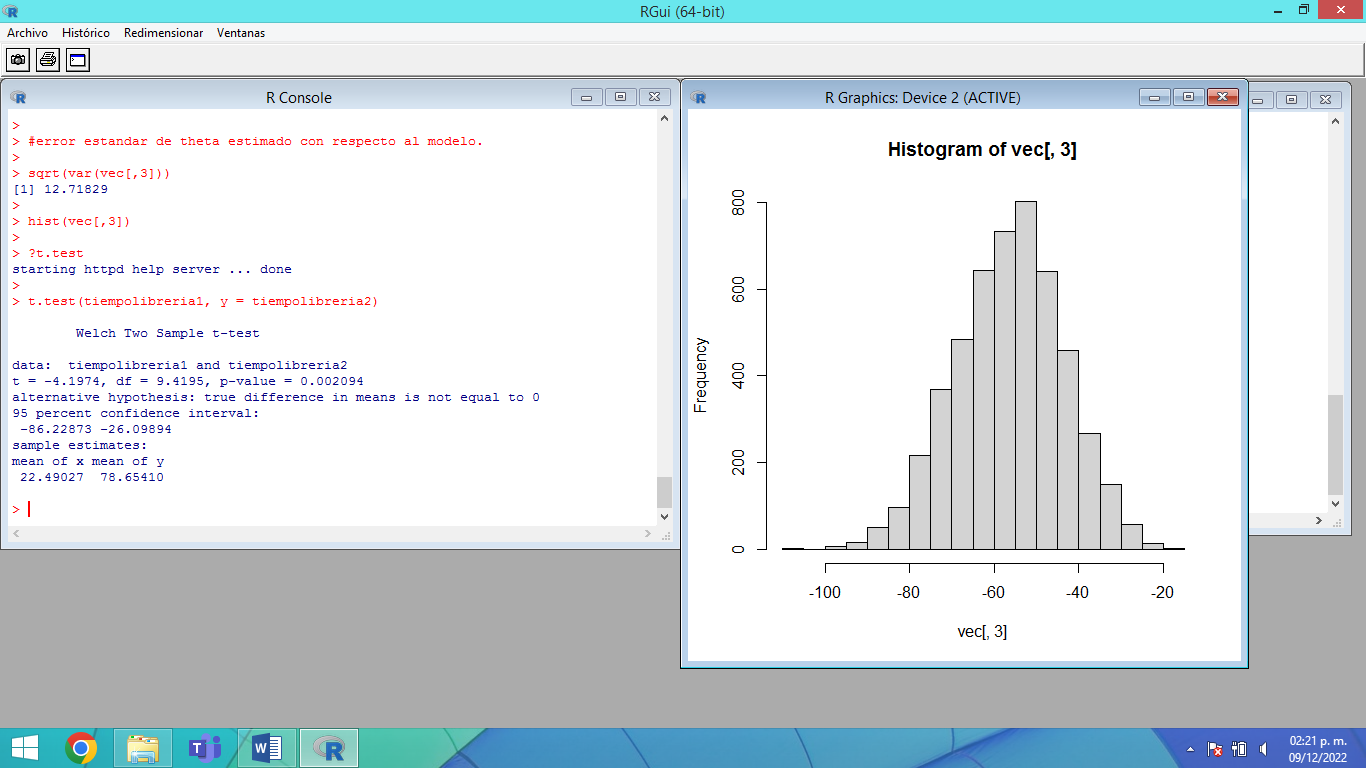
#error estandar de theta estimado con respecto al modelo.

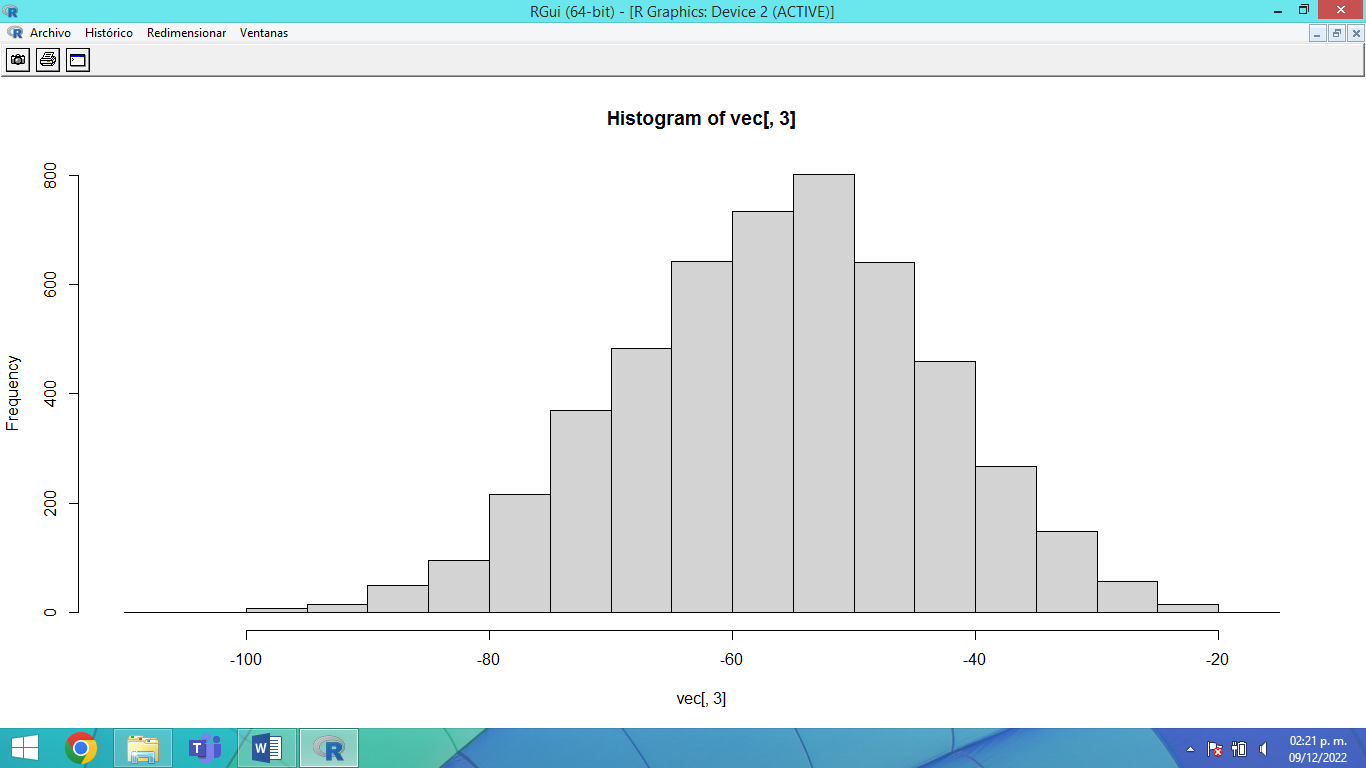
sqrt(var(vec[,3]))

hist(vec[,3])

?t.test

t.test(tiempolibreria1, y = tiempolibreria2)





Los códigos para calcular la memoria de las librerías 1 y 2 se muestran a continuación:

**Memoria Librería 1**

memorialibreria1<-c(78,89.6,65.9,63.6,65,92.9,87.9,74.7,94.3)

mean(memorialibreria1)

#Estimación del error estándar

sd(memorialibreria1)/sqrt(length(memorialibreria1))

#Estimación de la diferencia de medias

mean(memorialibreria1)- mean(memorialibreria1)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(memorialibreria1)/length(memorialibreria1))

#Estimación de las medianas

median(memorialibreria1)

#Estimación de la diferencia de medianas

median(memorialibreria1)- median(memorialibreria1)

B<-5000

vec<-matrix(,B,2)

for(i in 1:B){

d<-sample(memorialibreria1,9,replace=T)

vec[i,1]<-mean(d)

vec[i,2]<-vec[i,1]

}

thet<-mean(vec[,2])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,1]-thet)^2

i<-i+1

if(i>B){

break}

}

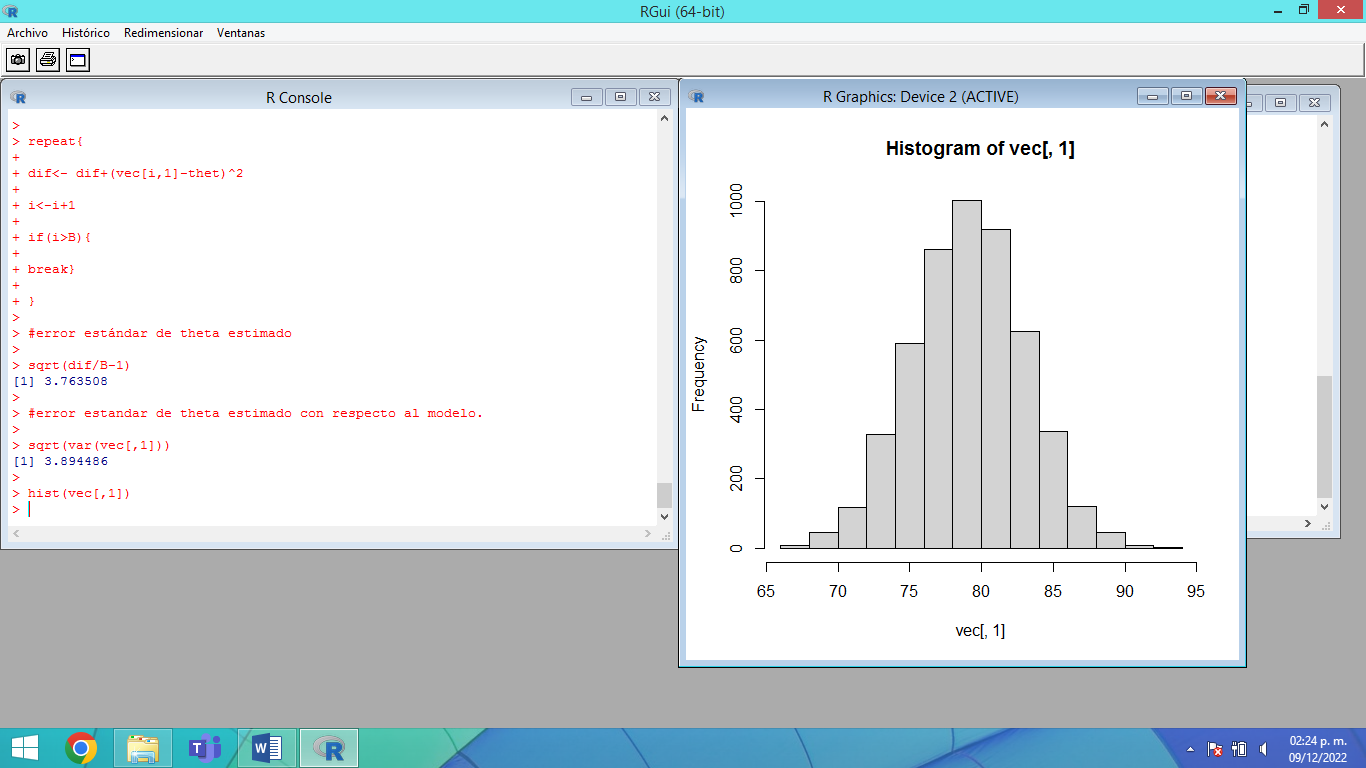
#error estándar de theta estimado

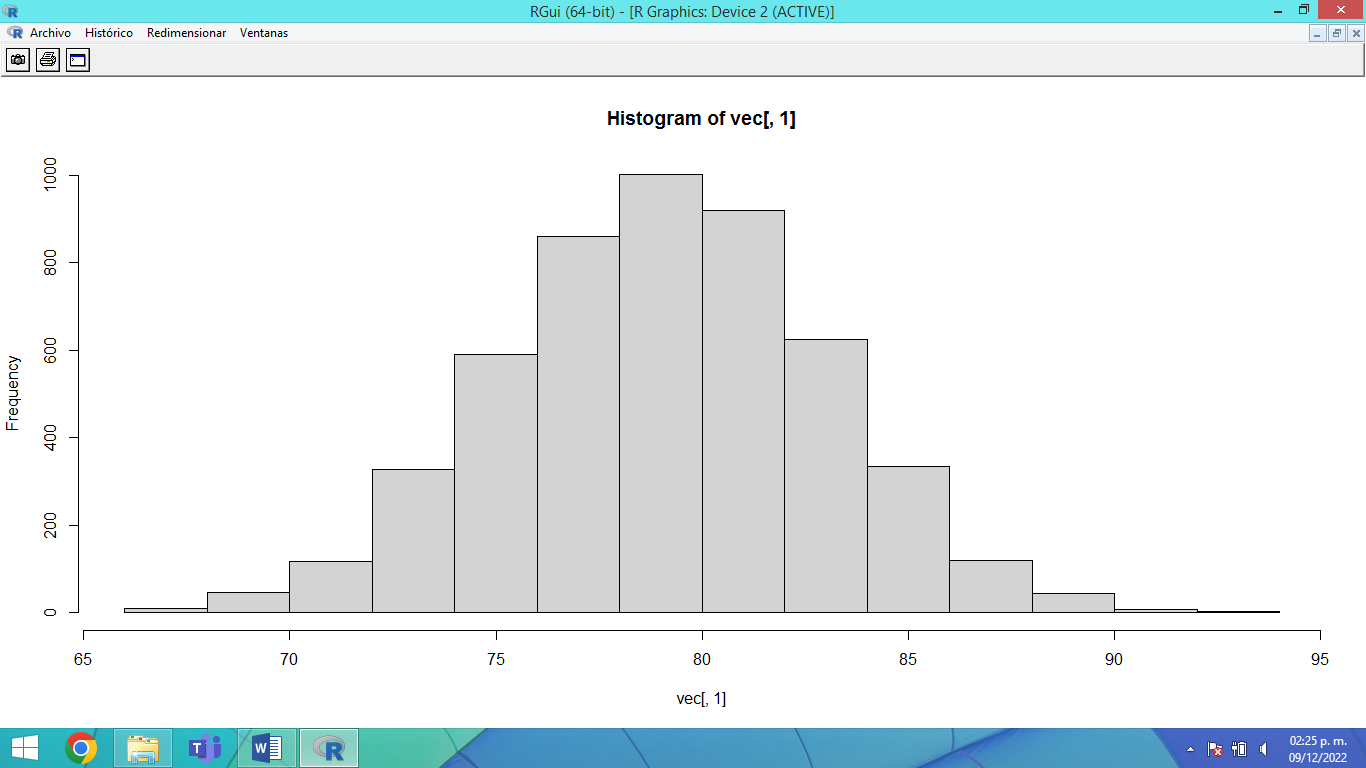
sqrt(dif/B-1)

#error estandar de theta estimado con respecto al modelo.

sqrt(var(vec[,1]))

hist(vec[,1])





**Memoria Librería 2**

memorialibreria2<-c(89,94.5,76.4,78.8,79,94.7,93.2,91.4,84)

mean(memorialibreria2)

#Estimación del error estándar

sd(memorialibreria2)/sqrt(length(memorialibreria2))

#Estimación de la diferencia de medias

mean(memorialibreria2)- mean(memorialibreria2)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(memorialibreria2)/length(memorialibreria2))

#Estimación de las medianas

median(memorialibreria2)

#Estimación de la diferencia de medianas

median(memorialibreria2)- median(memorialibreria2)

B<-5000

vec<-matrix(,B,2)

for(i in 1:B){

h<-sample(memorialibreria2,9,replace=T)

vec[i,1]<-mean(h)

vec[i,2]<-vec[i,1]

}

thet<-mean(vec[,2])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,1]-thet)^2

i<-i+1

if(i>B){

break}

}

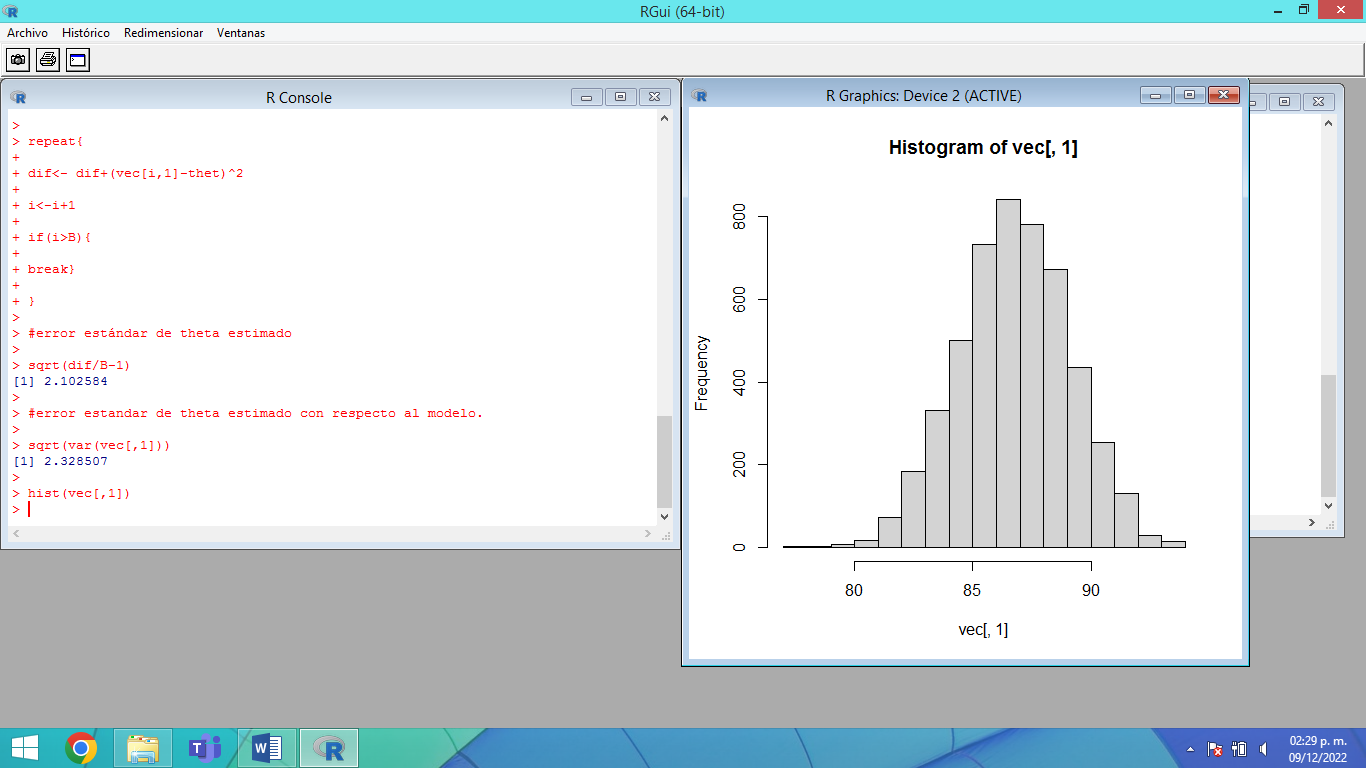
#error estándar de theta estimado

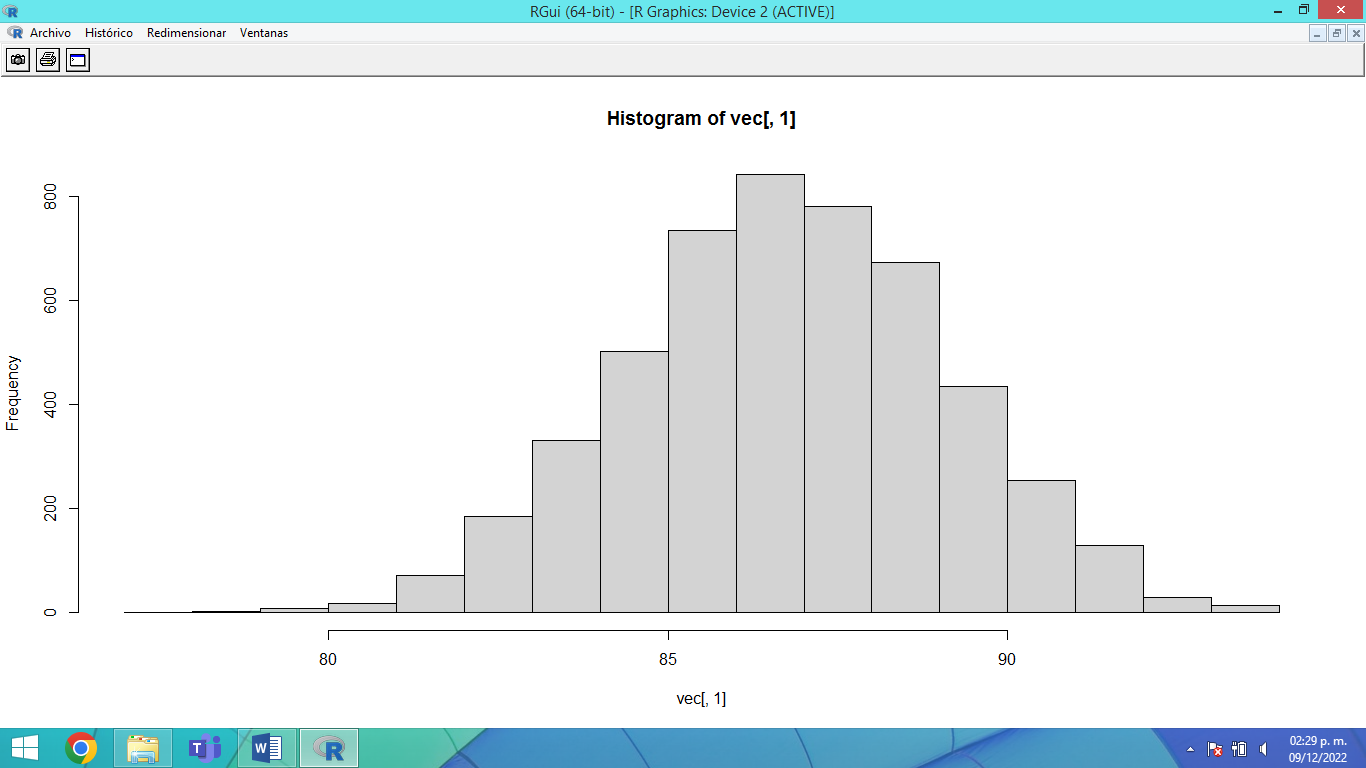
sqrt(dif/B-1)

#error estandar de theta estimado con respecto al modelo.

sqrt(var(vec[,1]))

hist(vec[,1])





**Memoria entre Librería 1 y Librería 2**

memorialibreria1<-c(78,89.6,65.9,63.6,65,92.9,87.9,74.7,94.3)

memorialibreria2<-c(89,94.5,76.4,78.8,79,94.7,93.2,91.4,84)

mean(memorialibreria1)

mean(memorialibreria2)

#Estimación del error estándar

sd(memorialibreria1)/sqrt(length(memorialibreria1))

sd(memorialibreria2)/sqrt(length(memorialibreria2))

#Estimación de la diferencia de medias

mean(memorialibreria1)- mean(memorialibreria2)

#¿Esto sugiere un efecto considerable para prolongar la vida causada por la asignación del tratamiento?

#Para dar respuesta a la pregunta se debe estimar la exactitud de los

#promedios de los tratamientos, es decir se debe obtener el estimador

#del error estándar de la diferencia de medias

#Estimación del error estándar de la diferencia de medias

sqrt(var(memorialibreria1)/length(memorialibreria1)+ var(memorialibreria2)/length(memorialibreria2))

#Estimación de las medianas

median(memorialibreria1)

median(memorialibreria2)

#Estimación de la diferencia de medianas

median(memorialibreria1)- median(memorialibreria2)

B<-5000

vec<-matrix(,B,3)

for(i in 1:B){

d<-sample(memorialibreria1,9,replace=T)

vec[i,1]<-mean(d)

h<-sample(memorialibreria2,9,replace=T)

vec[i,2]<-mean(h)

vec[i,3]<-vec[i,1]-vec[i,2]

}

thet<-mean(vec[,3])

thet

dif<-0

i<-1

repeat{

dif<- dif+(vec[i,3]-thet)^2

i<-i+1

if(i>B){

break}

}

#error estándar de theta estimado

sqrt(dif/B-1)

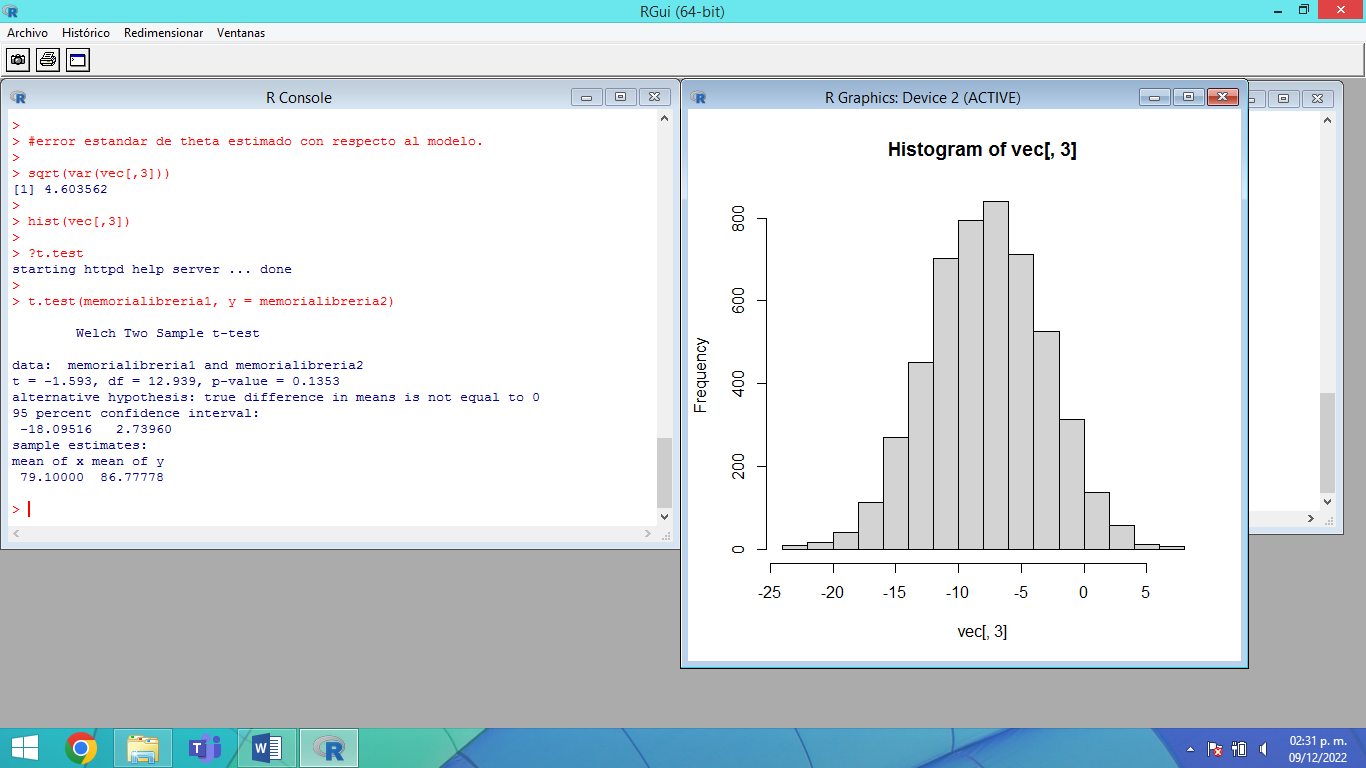
#error estandar de theta estimado con respecto al modelo.

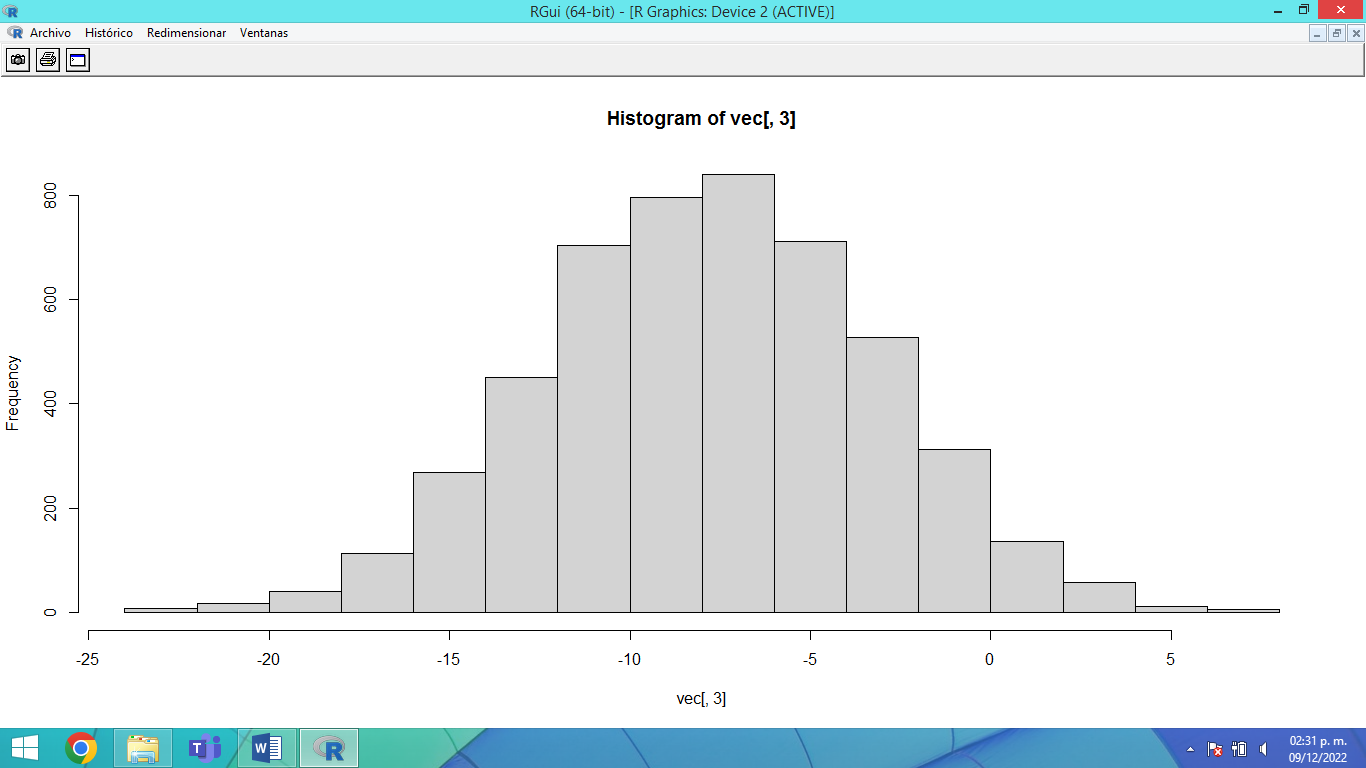
sqrt(var(vec[,3]))

hist(vec[,3])

?t.test

t.test(memorialibreria1, y = memorialibreria2)

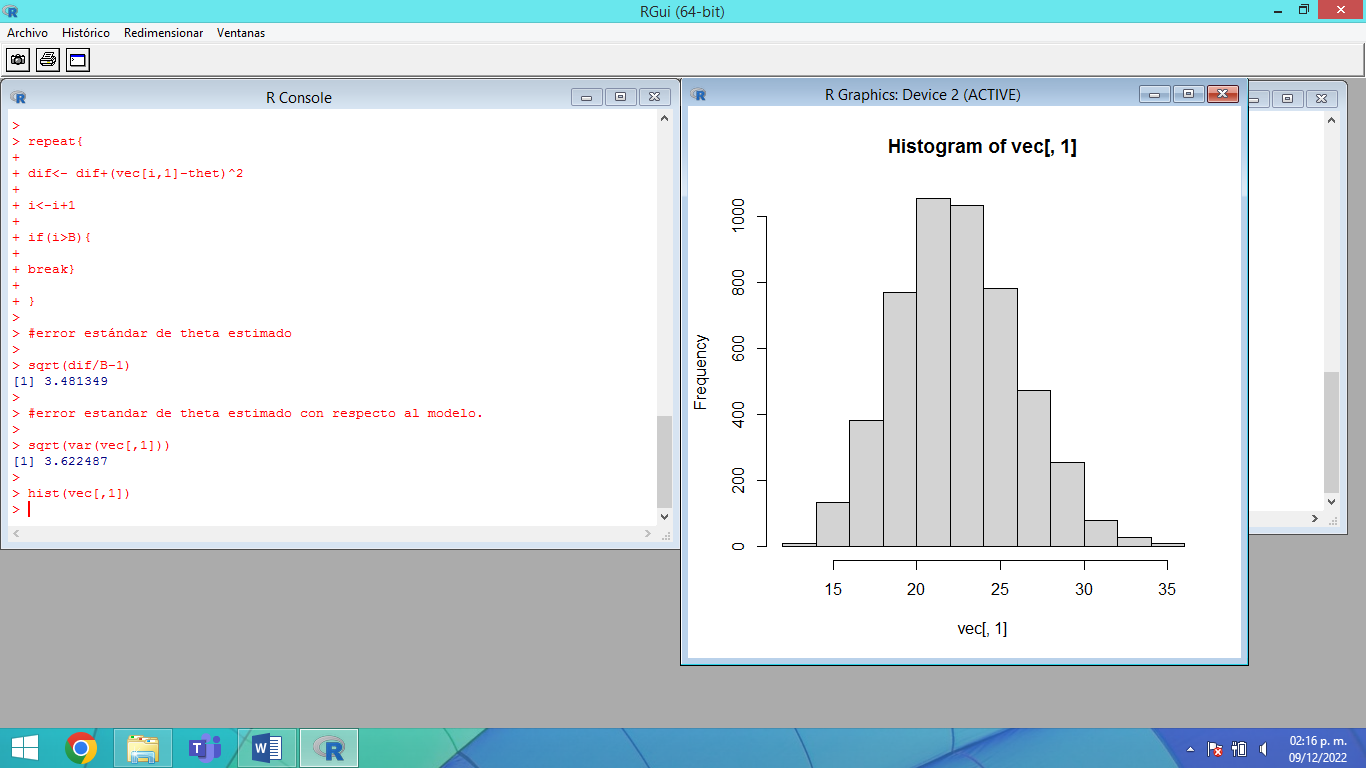




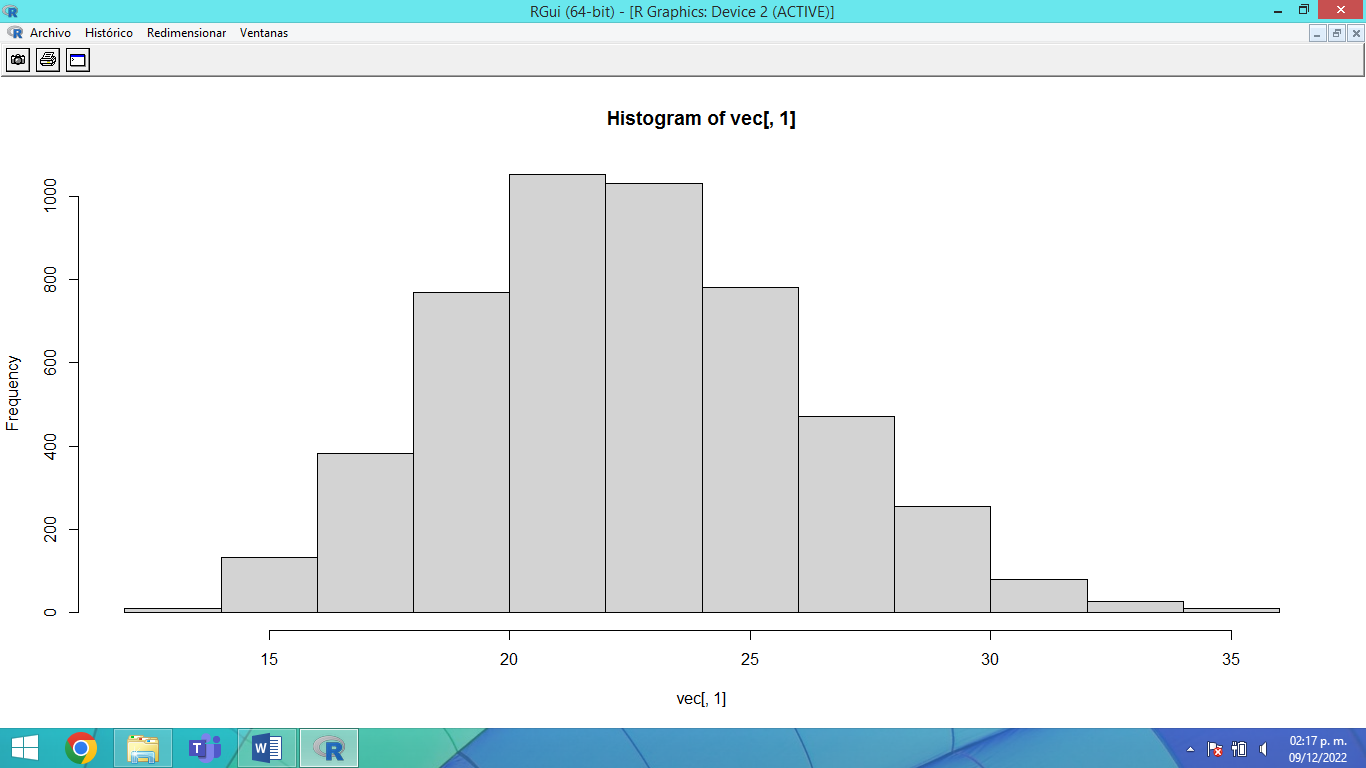
**Resultados**

Esta técnica se usa frecuentemente para aproximar el sesgo o la varianza de un análisis estadístico, así como para construir intervalos de confianza o realizar contrastes de hipótesis sobre parámetros de interés.

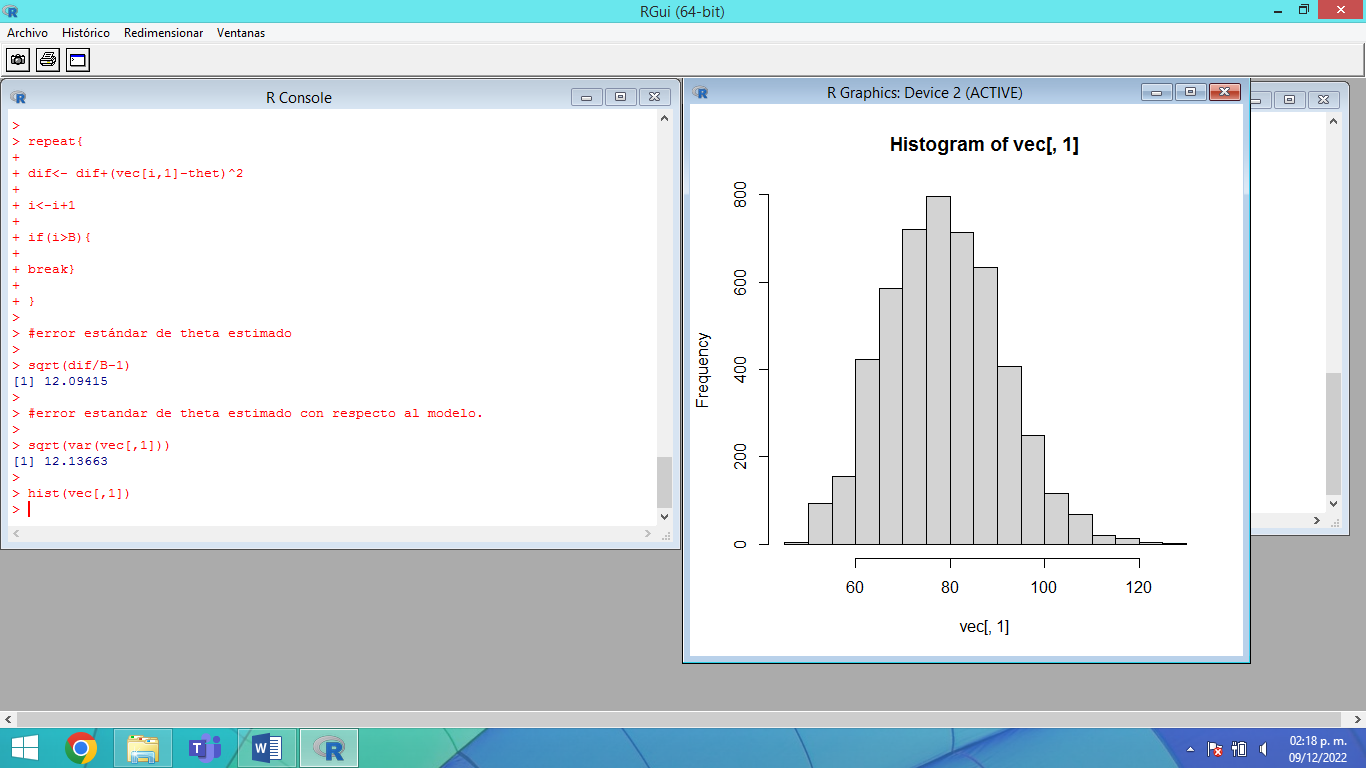
**Tiempo Librería 1**



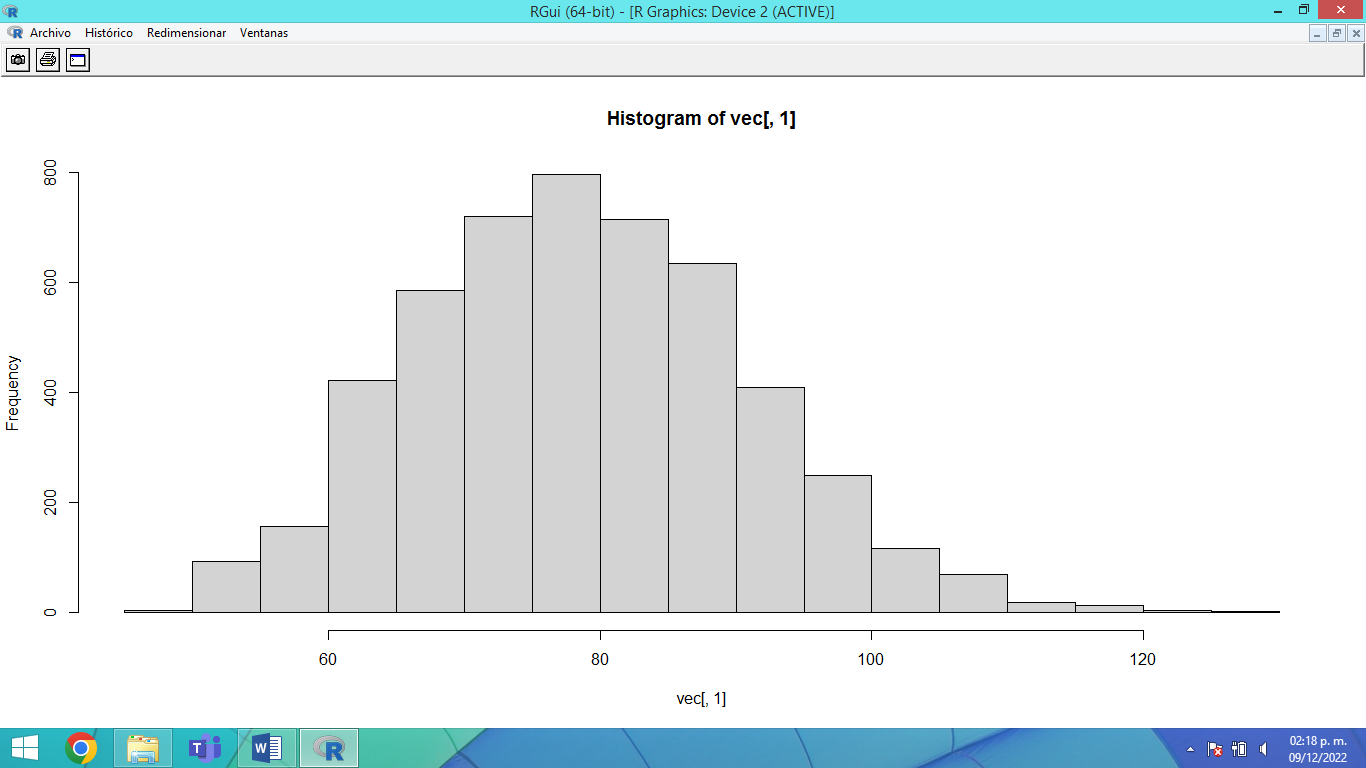
El error estándar es una estimación de la cantidad que el valor de una estadística de prueba varía de muestra a muestra en el tiempo de la librería 1 se muestra que el valor es 3.622487.



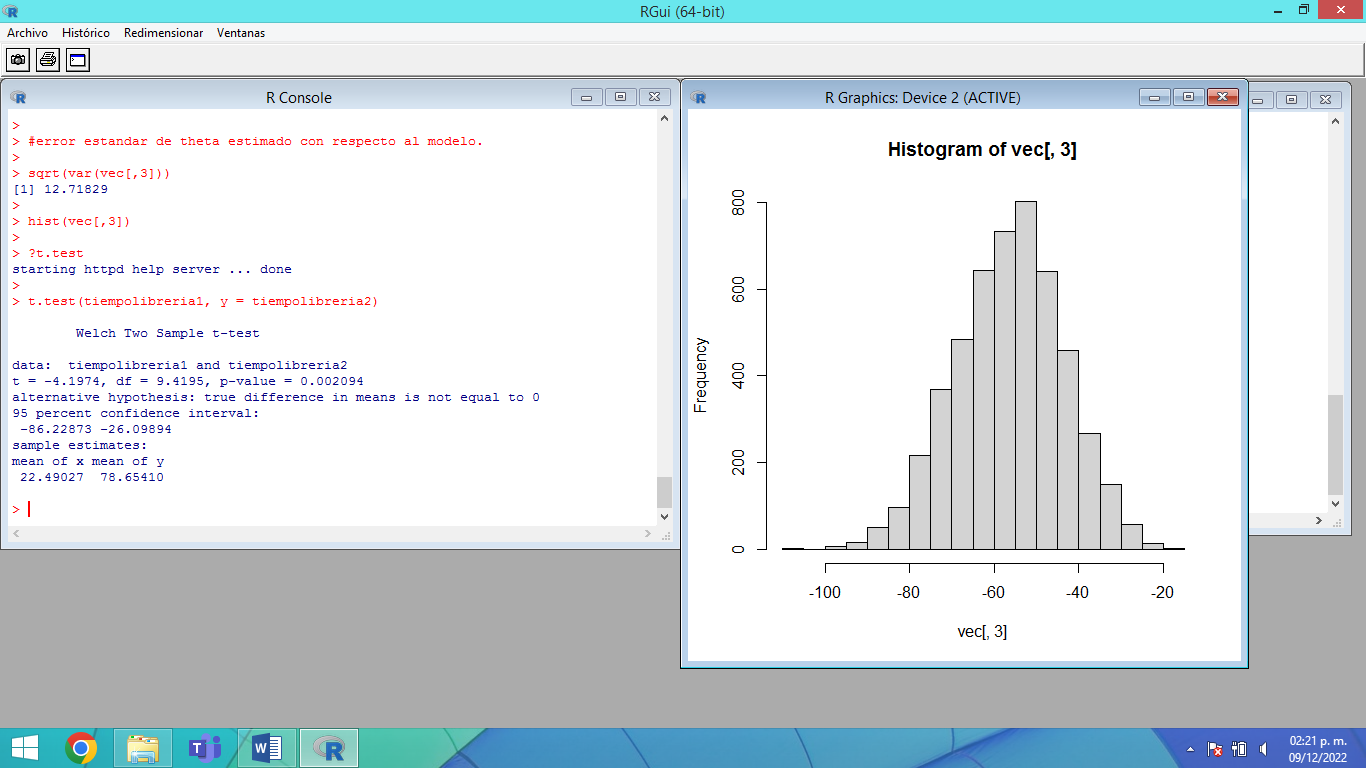
**Tiempo Librería 2**



El error estándar es una estimación de la cantidad que el valor de una estadística de prueba varía de muestra a muestra en el tiempo de la librería 2 se muestra que el valor es 12.13663

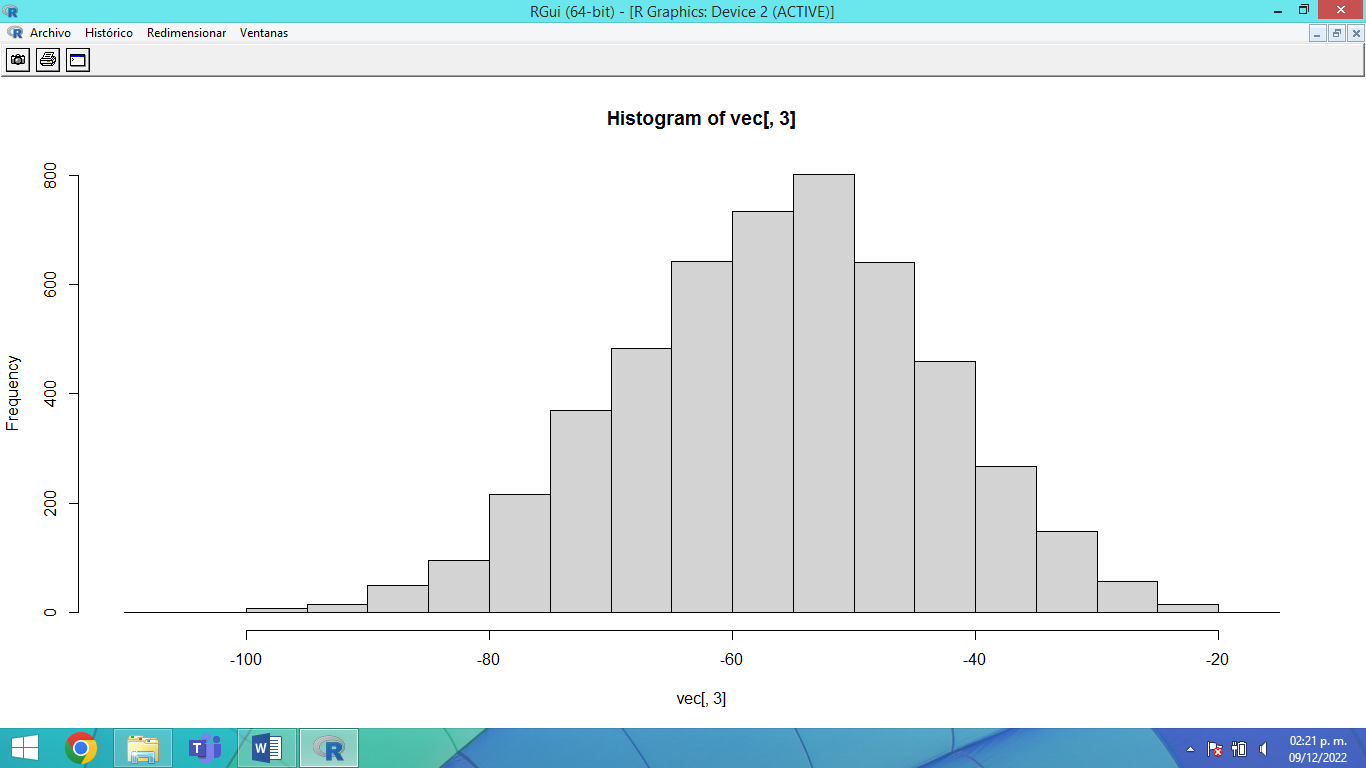


**Tiempo entre Librería 1 y Librería 2**

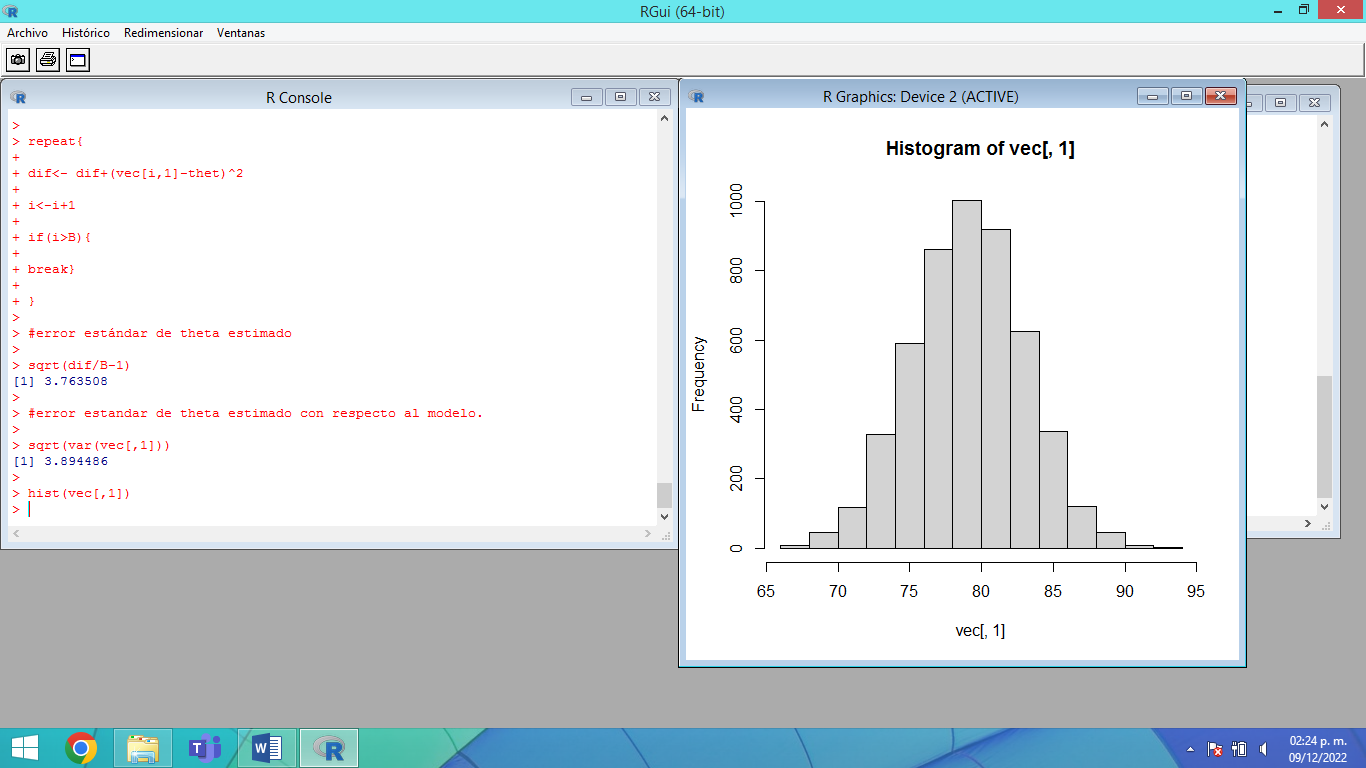


Se evaluaran los tiempos de la librería 1 y librería 2 mediante T de Student el cual es una herramienta para evaluar las medias de uno o dos grupos mediante pruebas de hipótesis el valor p es igual a 0.002094 el cual indica la probabilidad de obtener un valor tan o más extremo al observado, partiendo de la suposición de la igualdad de efecto que marca la hipótesis nula.

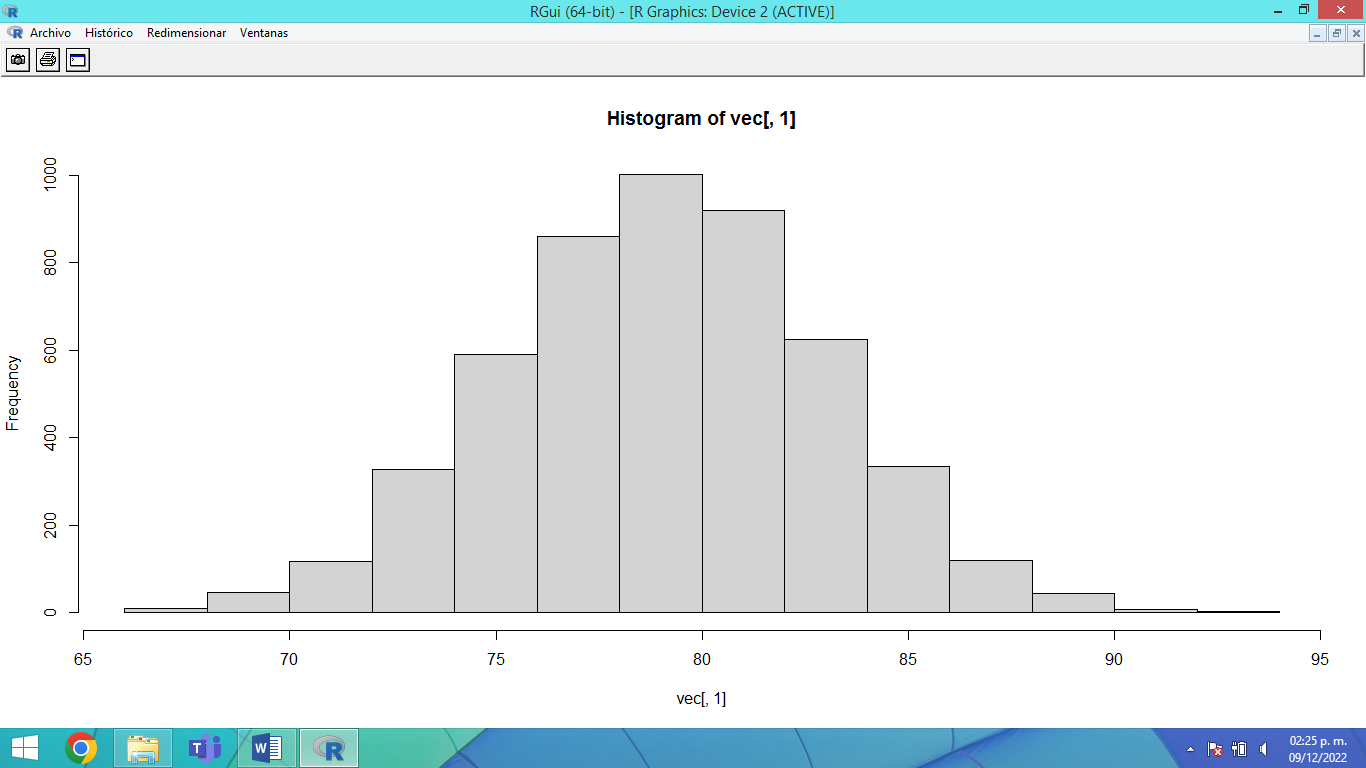
La verdadera diferencia de medias no es igual a 0. Una p < 0,05 significa que la hipótesis nula es falsa y una p > 0,05 que la hipótesis nula es verdadera: siempre nos movemos en el terreno de la probabilidad. El intervalo de confianza del 95 por ciento.



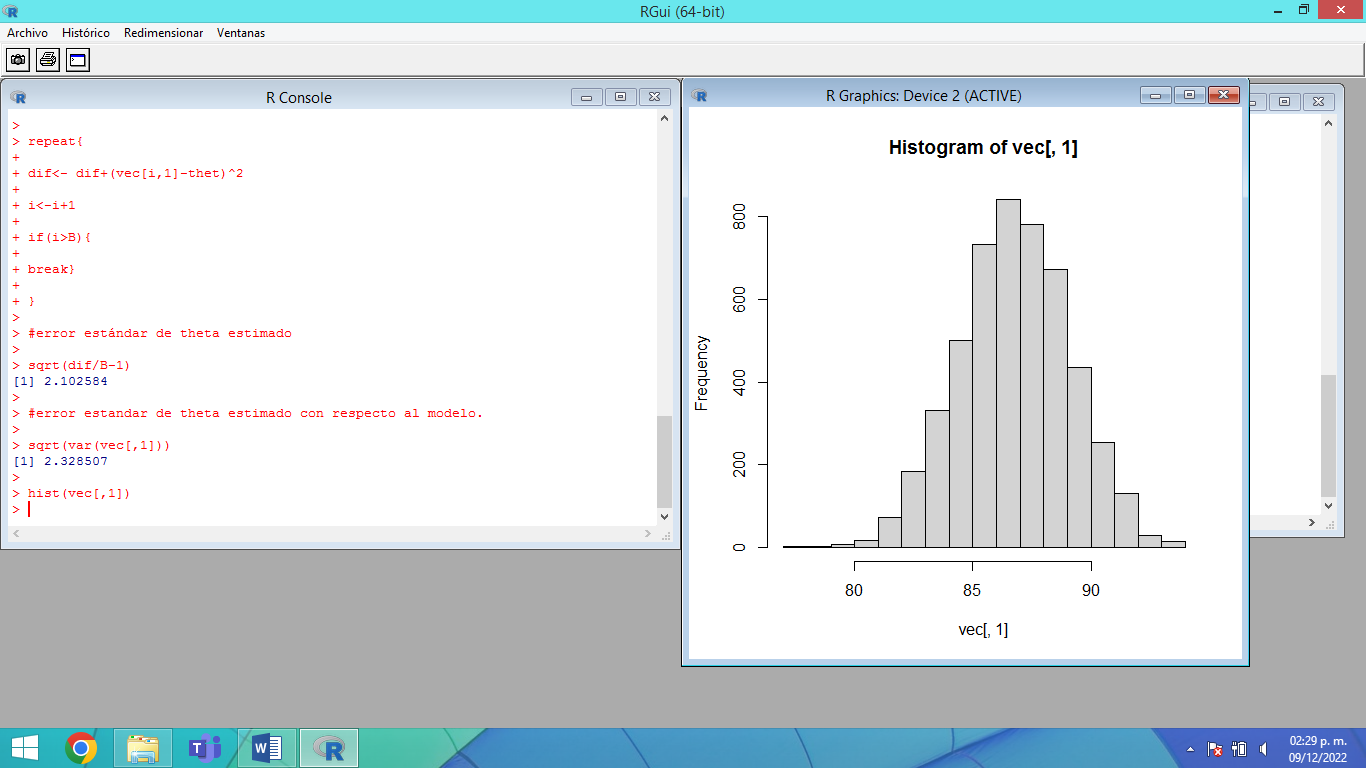
**Memoria Librería 1**



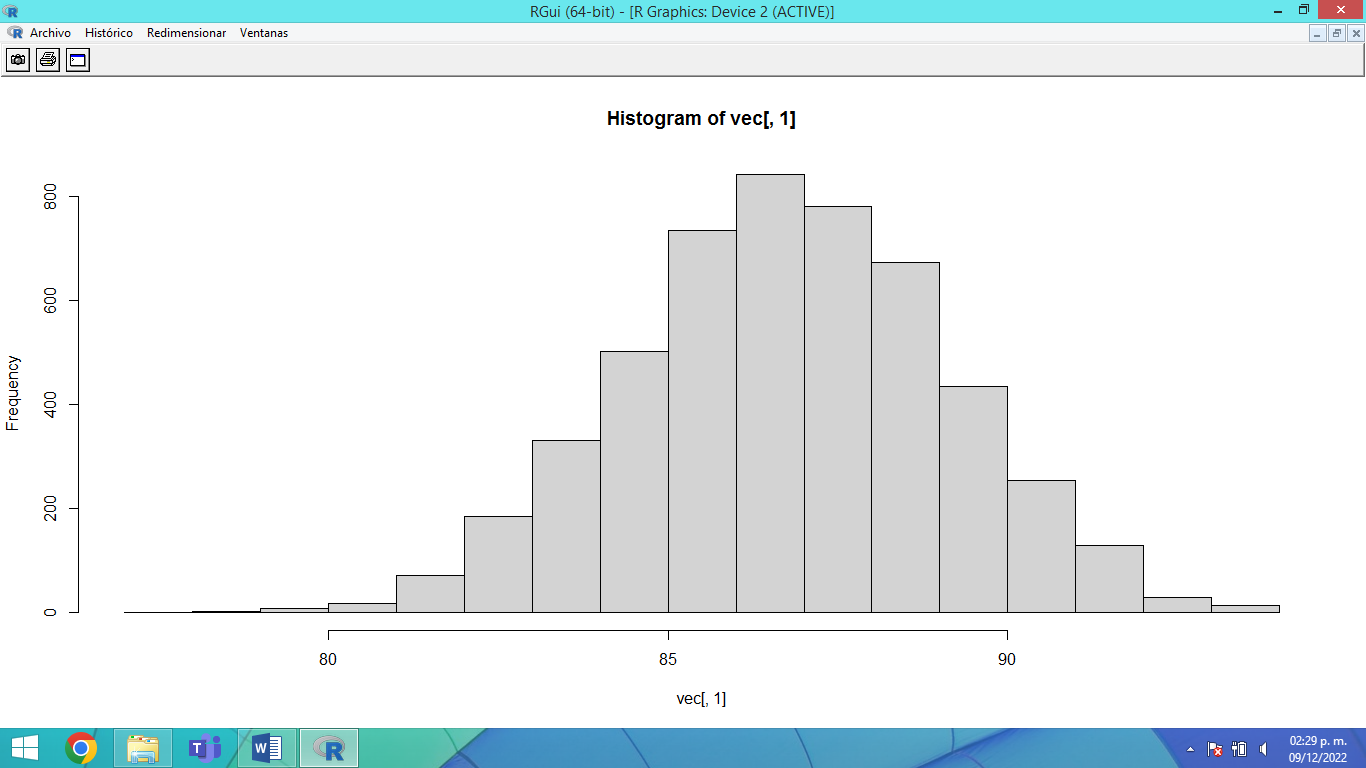
El error estándar es una estimación de la cantidad que el valor de una estadística de prueba varía de muestra a muestra en la memoria de la librería 1 se muestra que el valor es 3.894486



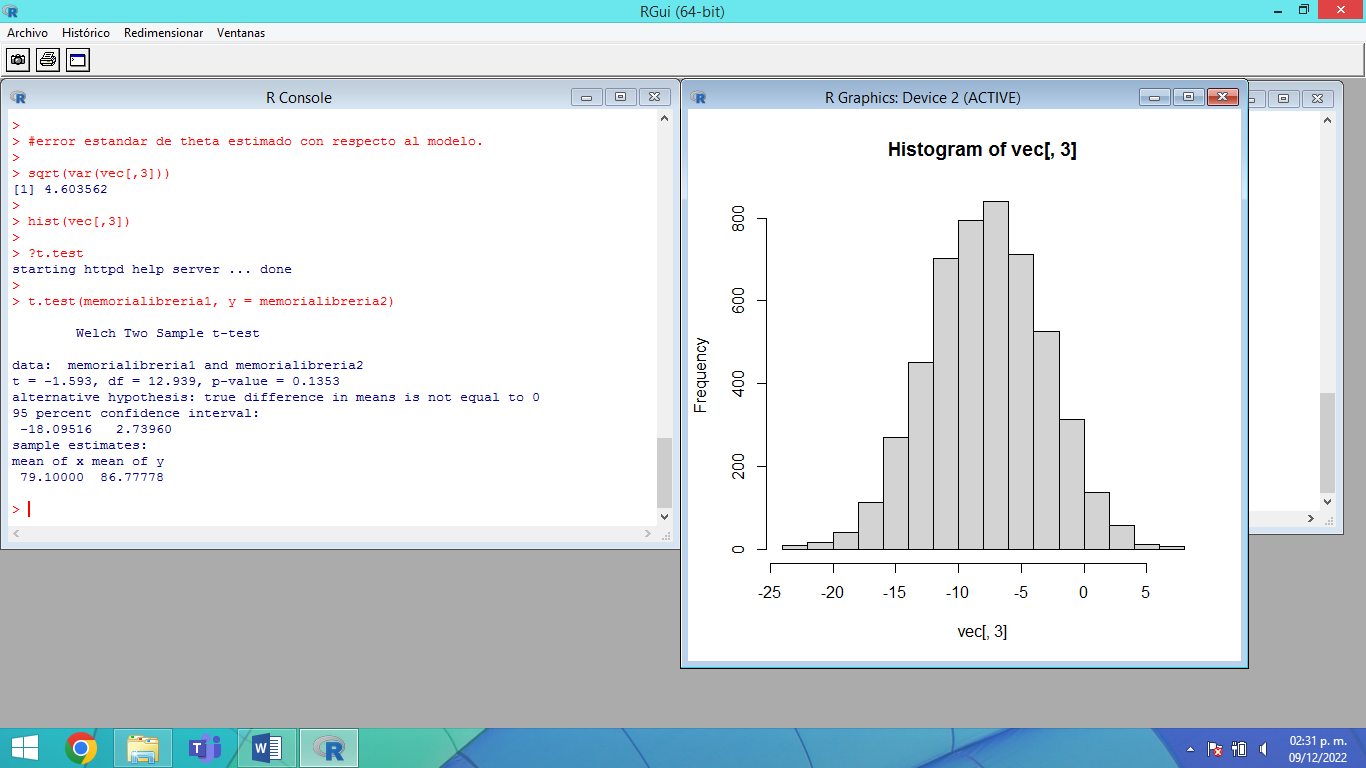
**Memoria Librería 2**



El error estándar es una estimación de la cantidad que el valor de una estadística de prueba varía de muestra a muestra en la memoria de la librería 2 se muestra que el valor es 2.328507

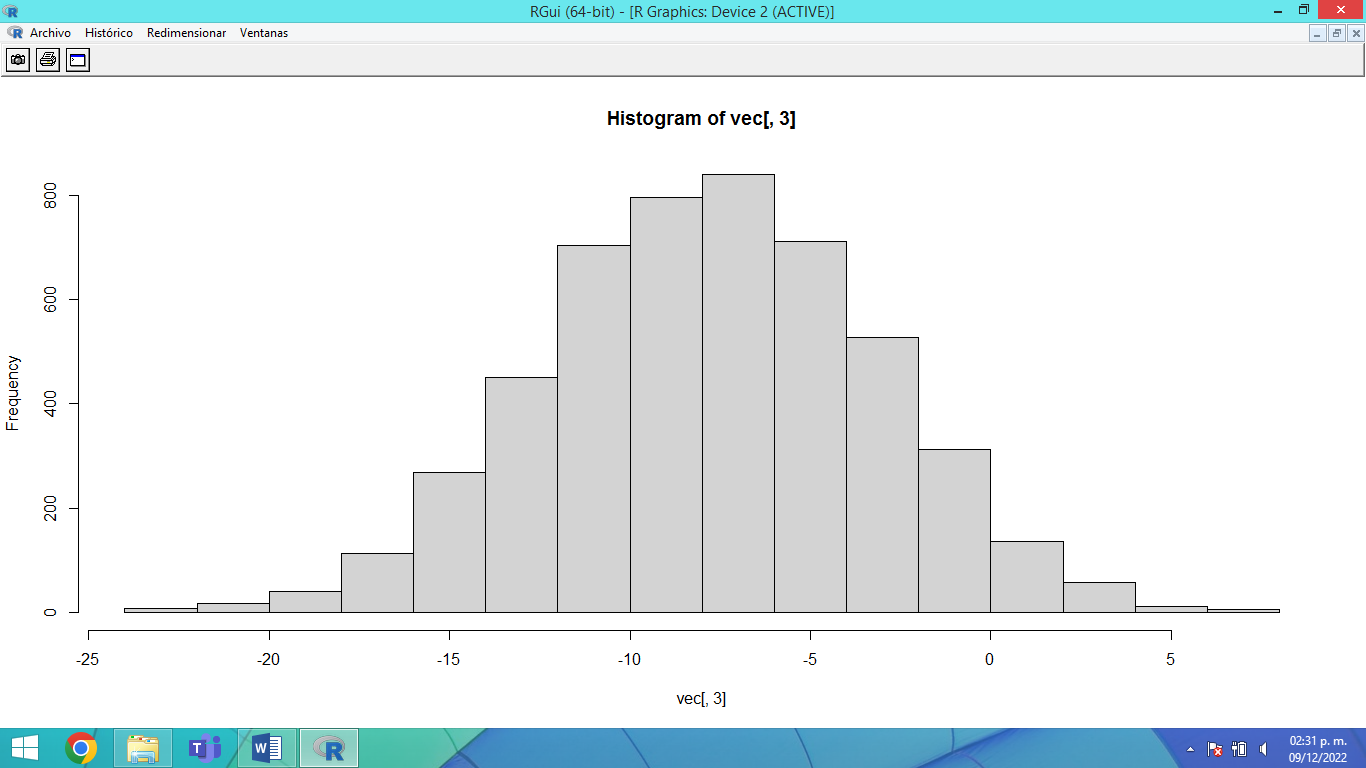


**Memoria entre Librería 1 y Librería 2**



Se evaluaran las memorias de la librería 1 y librería 2 mediante T de Student el cual es una herramienta para evaluar las medias de uno o dos grupos mediante pruebas de hipótesis el valor p es igual a 0.1353 el cual indica la probabilidad de obtener un valor tan o más extremo al observado, partiendo de la suposición de la igualdad de efecto que marca la hipótesis nula.

La verdadera diferencia de medias no es igual a 0. Una p < 0,05 significa que la hipótesis nula es falsa y una p > 0,05 que la hipótesis nula es verdadera: siempre nos movemos en el terreno de la probabilidad. El intervalo de confianza del 95 por ciento.



**Tiempo Librería 1 repetición una sola vez**

tiempolibreria1<-c(28.67193398,37.67445752,11.15839875,14.13138197,15.80249413,19.69257731,14.25780042,17.43216654,43.59123231)

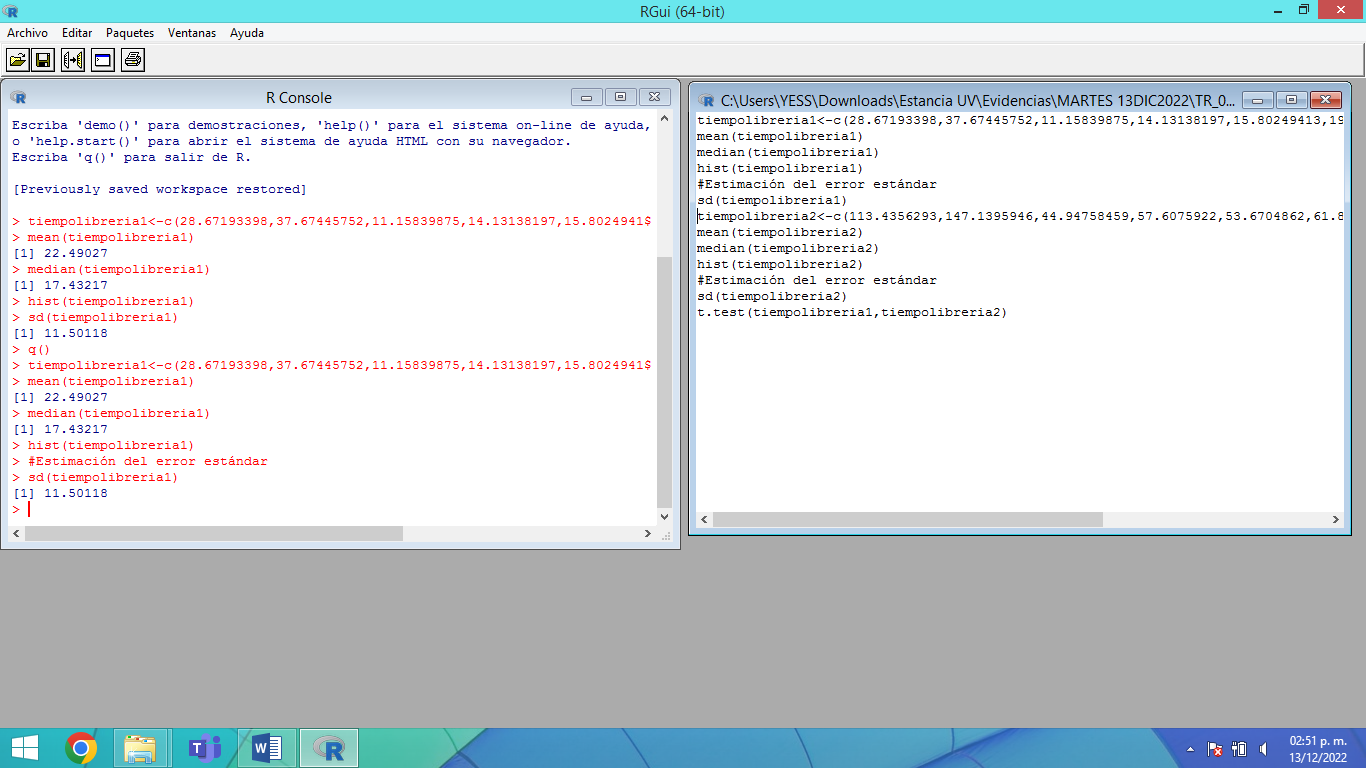
mean(tiempolibreria1)

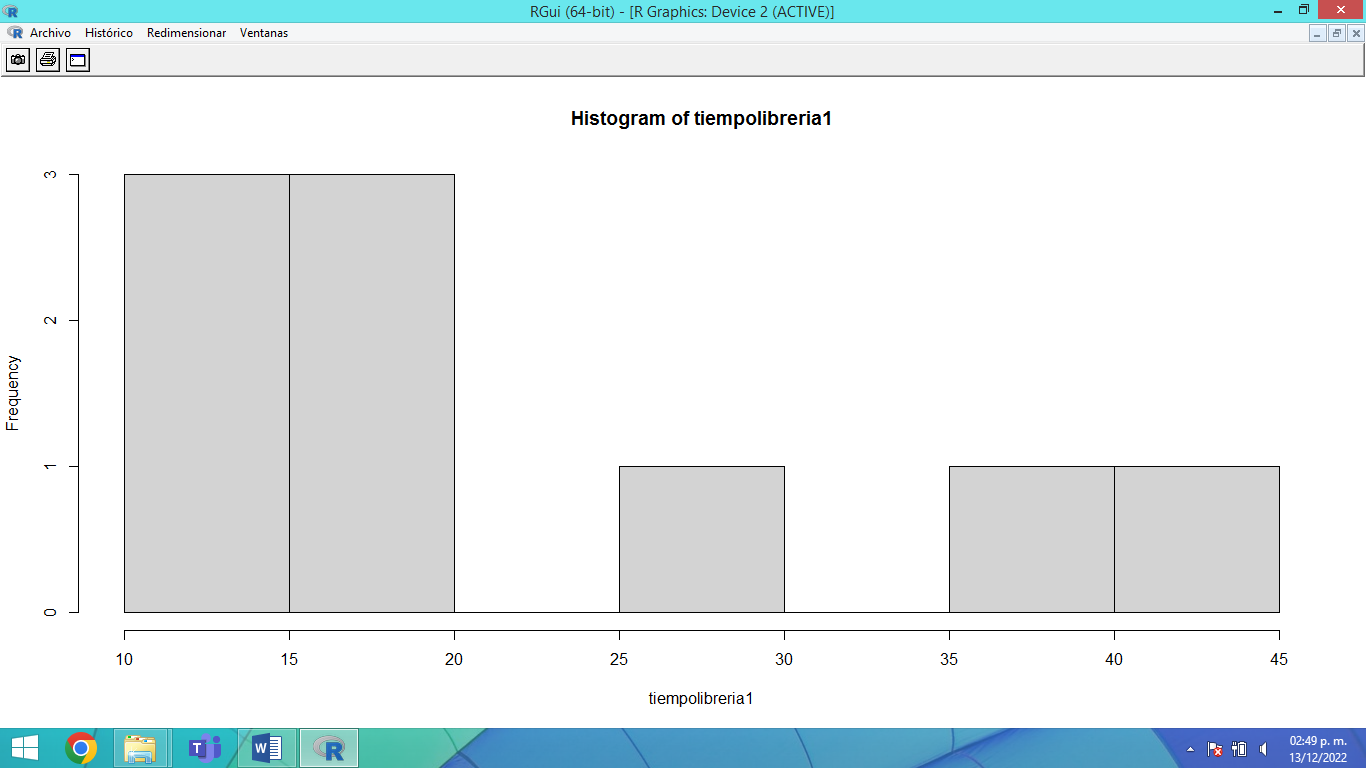
median(tiempolibreria1)

hist(tiempolibreria1)

#Estimación del error estándar

sd(tiempolibreria1)





**Tiempo Librería 2 repetición una sola vez**

tiempolibreria2<-c(113.4356293,147.1395946,44.94758459,57.6075922,53.6704862,61.83368747,123.8178965,58.2579331,47.17653714)

mean(tiempolibreria2)

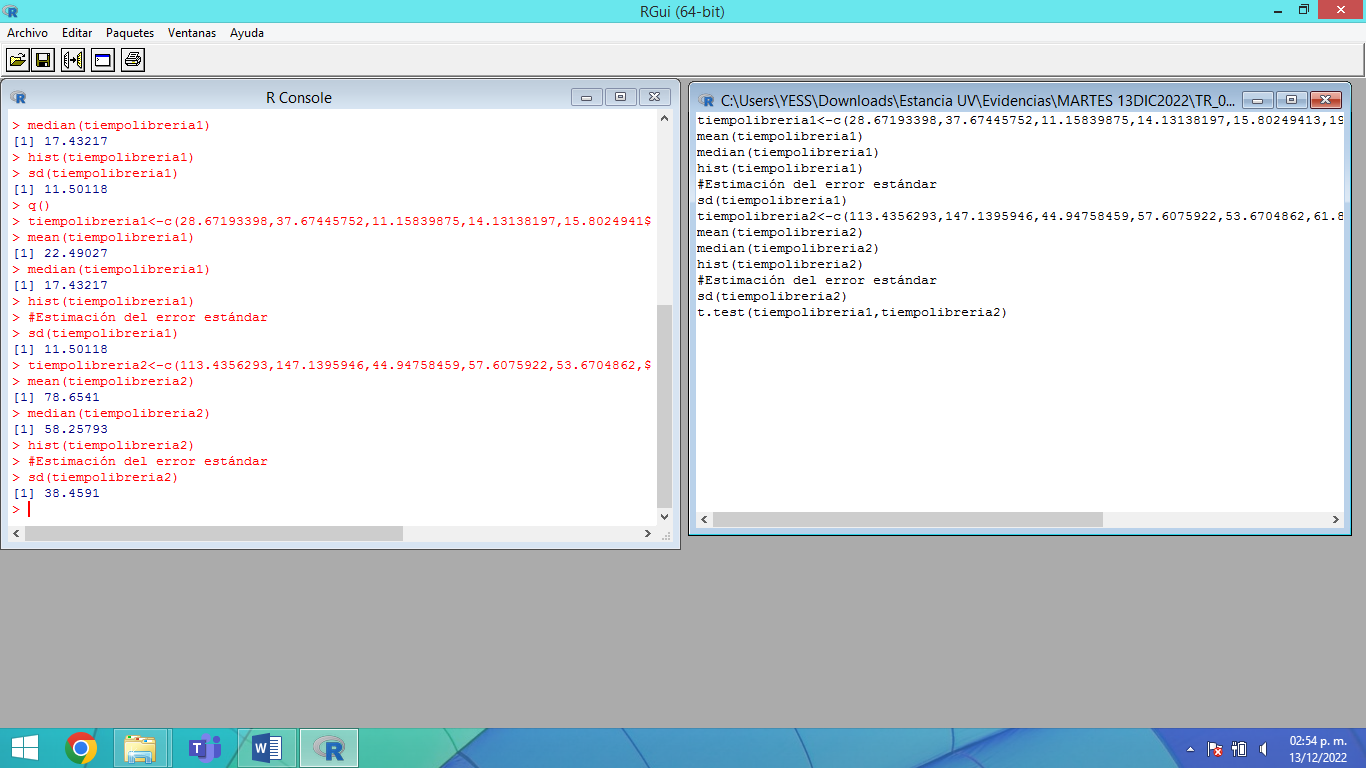
median(tiempolibreria2)

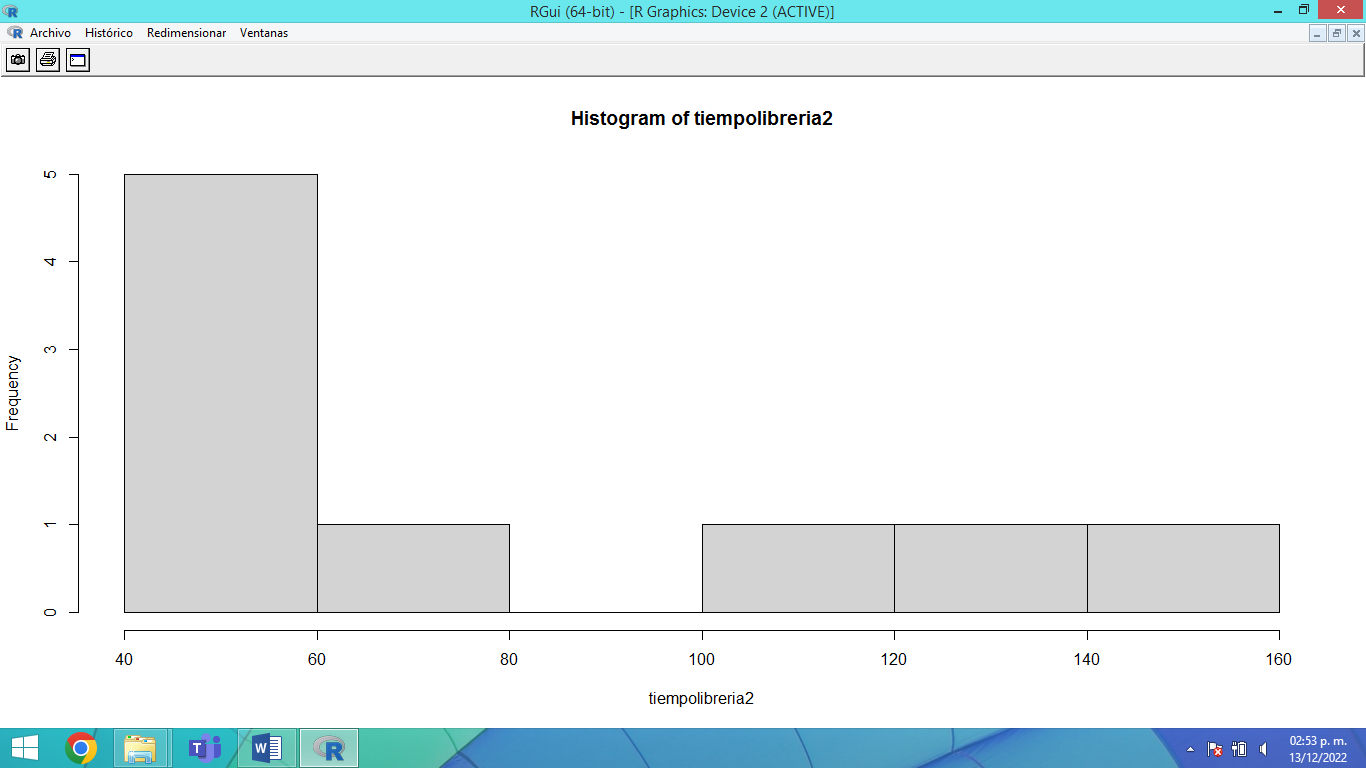
hist(tiempolibreria2)

#Estimación del error estándar

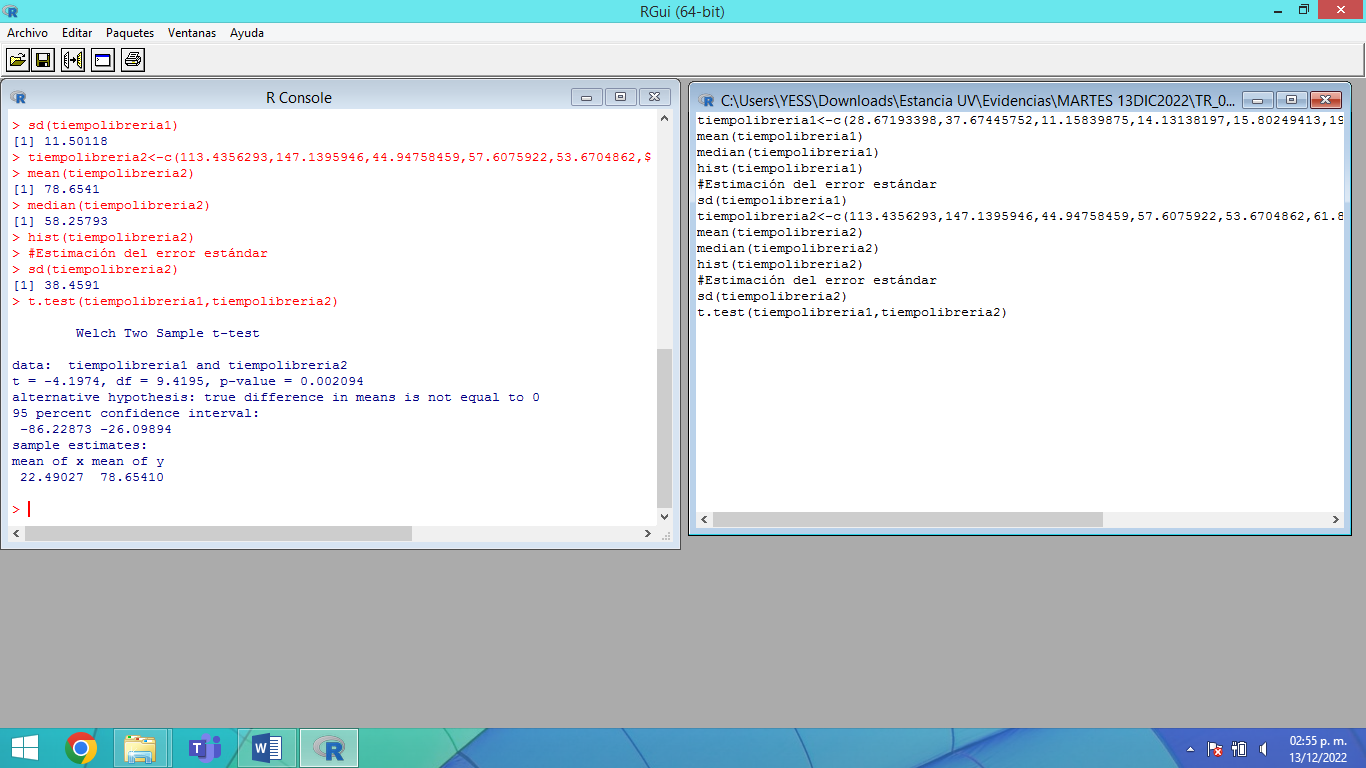
sd(tiempolibreria2)

t.test(tiempolibreria1,tiempolibreria2)





**Resultados**



**Memoria Librería 1 repetición una sola vez**

memorialibreria1<-c(78,89.6,65.9,63.6,65,92.9,87.9,74.7,94.3)

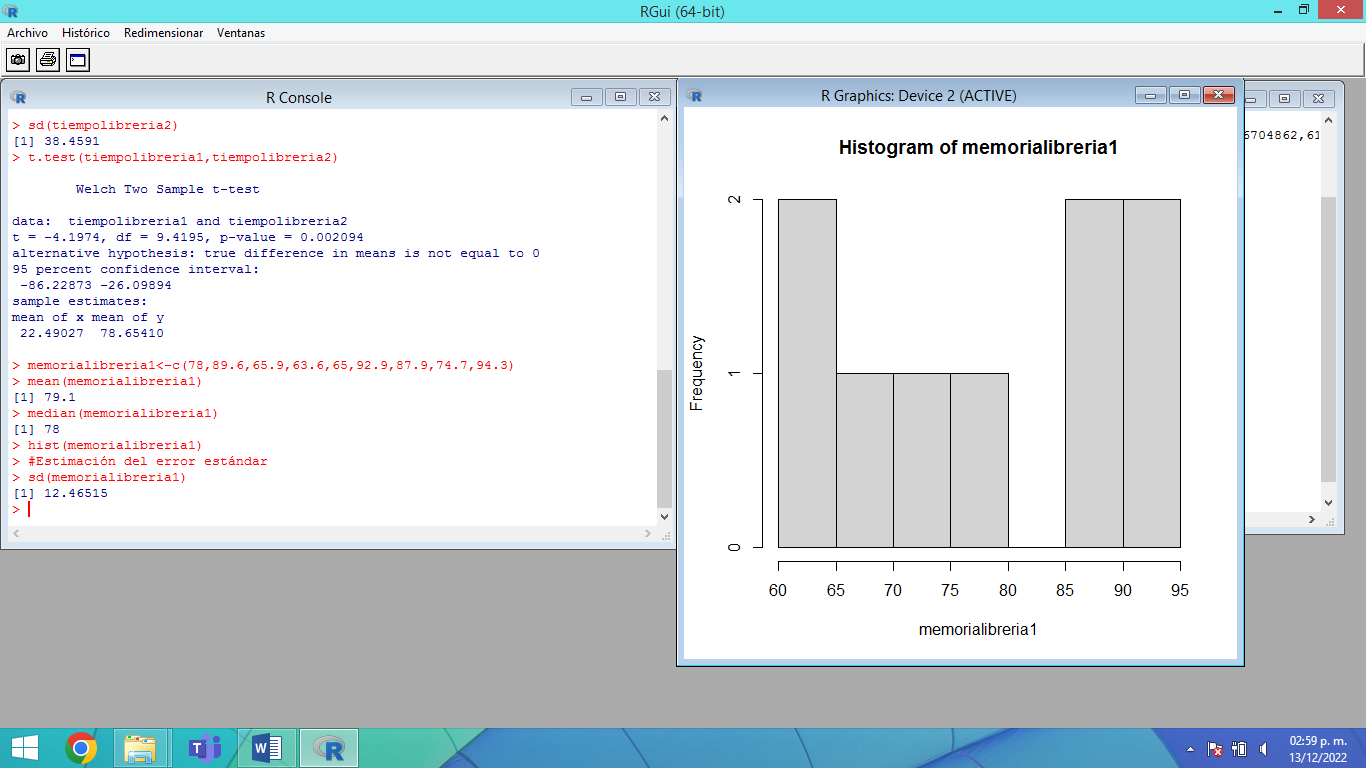
mean(memorialibreria1)

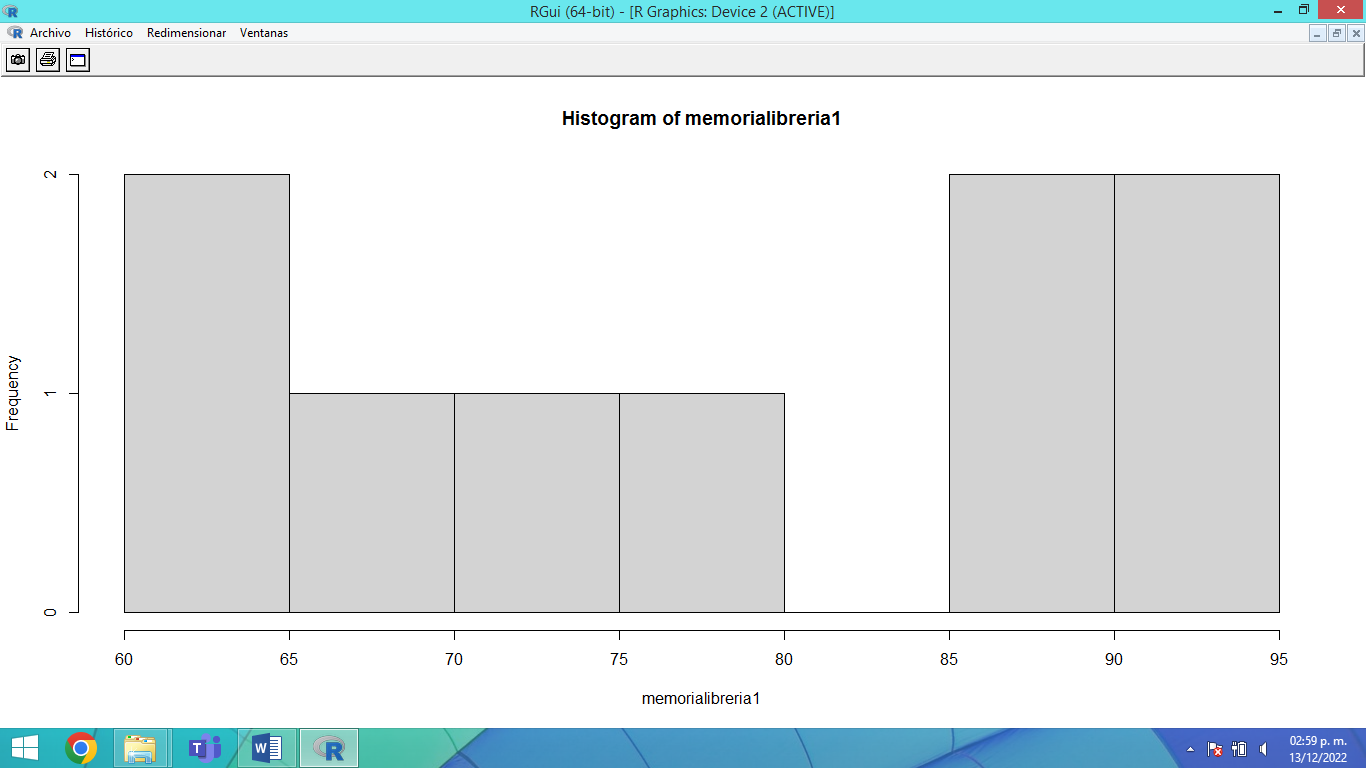
median(memorialibreria1)

hist(memorialibreria1)

#Estimación del error estándar

sd(memorialibreria1)





memorialibreria2<-c(89,94.5,76.4,78.8,79,94.7,93.2,91.4,84)

mean(memorialibreria2)

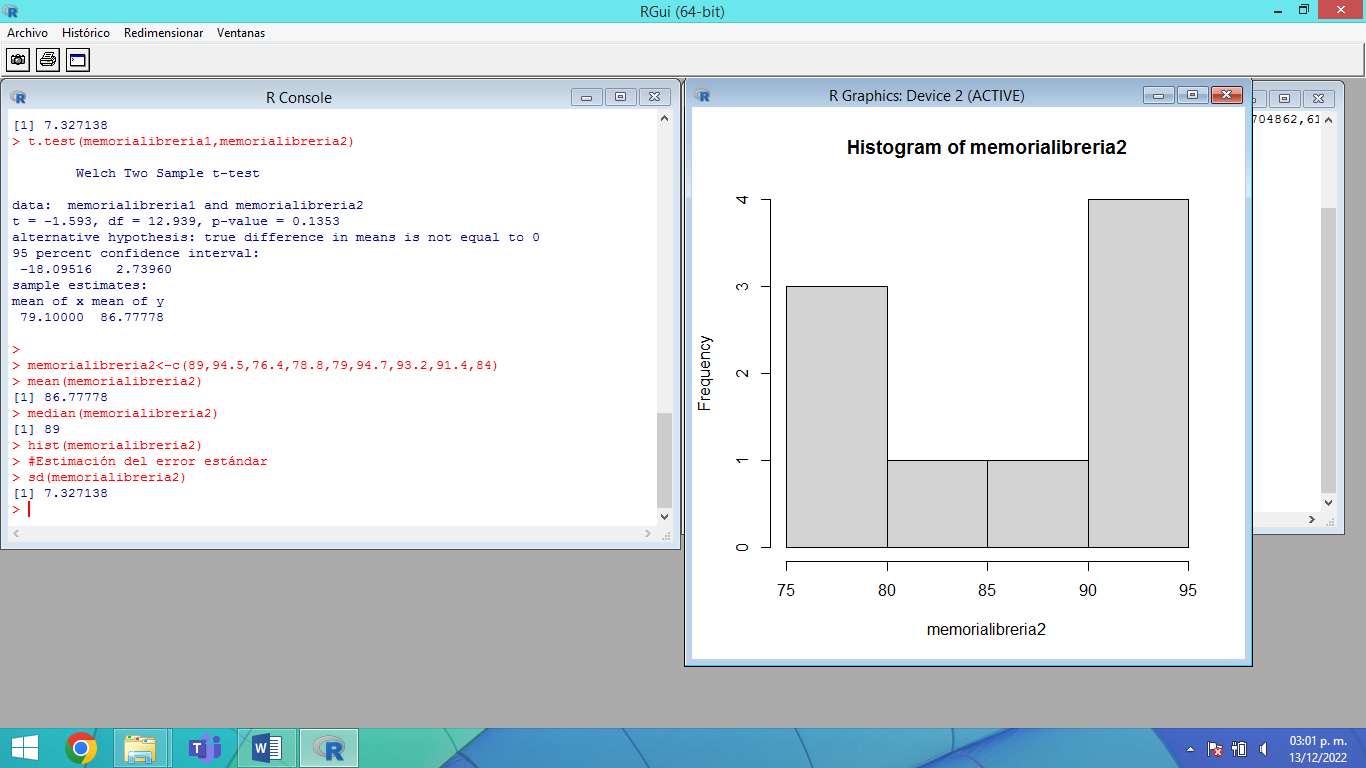
median(memorialibreria2)

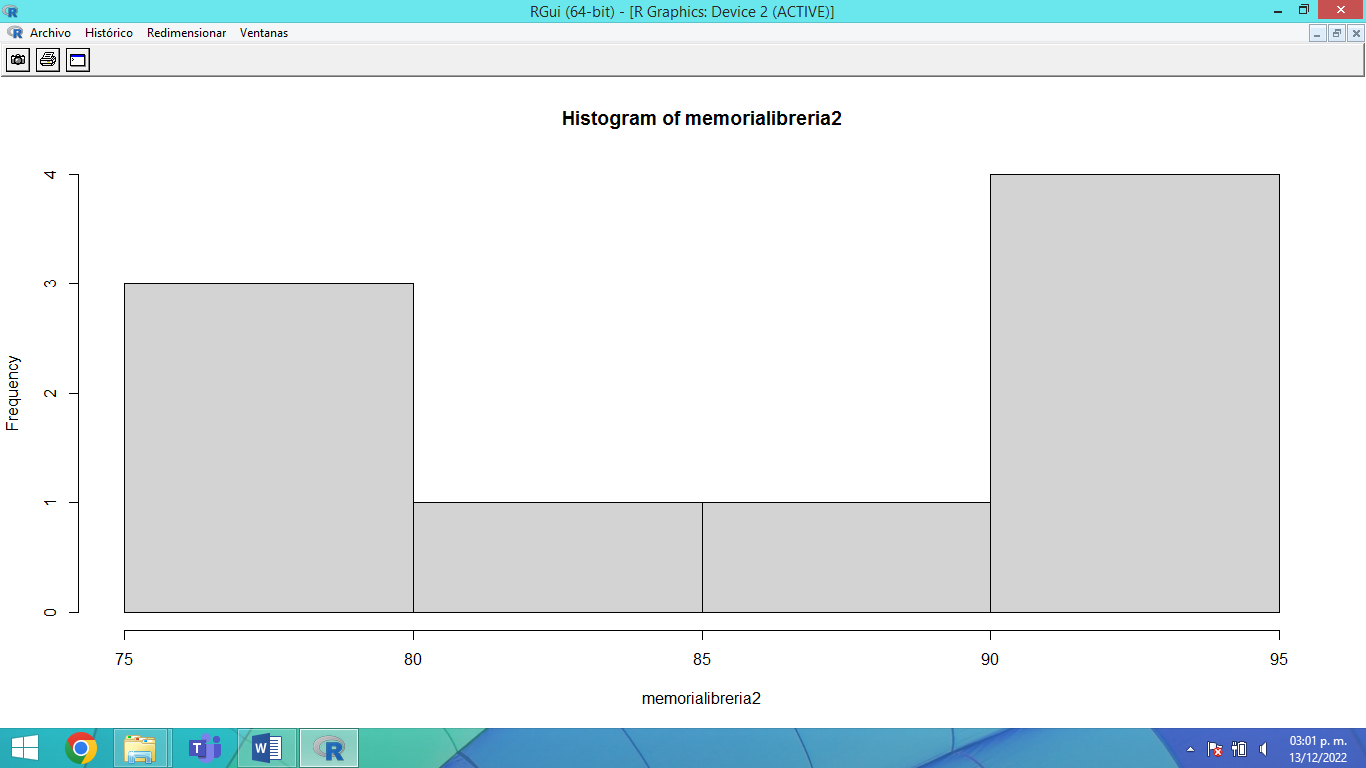
hist(memorialibreria2)

#Estimación del error estándar

sd(memorialibreria2)

t.test(memorialibreria1,memorialibreria2)





**Resultados**

