PRINCIPIOS DE DISEÑO EXPERIMENTAL

ISADORE NABI

I. INTRODUCCIÓN ______ *2*

II. CASO DE APLICACIÓN: ANÁLISIS DE RESISTENCIA PROMEDIO DE BURBUJAS ______ *2*

II.I. NIVELES (60 y 120 ml) DE GLICERINA (FACTOR) Y TIEMPO PROMEDIO DE DURACIÓN (RESPUESTA) ______ **3**

II.I. I. Análisis descriptivo de variabilidad del tiempo de resistencia promedio de las burbujas según nivel de glicerina ______ 3

II.I. II. Análisis descriptivo del tiempo promedio de resistencia de las burbujas según el nivel de glicerina ______ 4

II.I. III. Datos Centrados ______ 5

II.I. IV. Determinando la existencia (o no) de efectos de interacción ______ 8

II.I. V. Modelos de Ajuste ______ 9

II.I. VI. Sumas de Cuadrados ______ 9

II.I. VII. Estimación Adecuada de la Varianza Residual (Cuadrados Medios Residuales) ______ 10

II.I. VIII. Verificación del supuesto de homocedasticidad ______ 11

II.I VIX. Conclusiones Inferenciales ______ 11

II.II. TIPO DE AGUA (FACTOR CON NIVELES: DESTILADA. GRIFO, AÑEJADA) Y TIEMPO PROMEDIO DE DURACIÓN (RESPUESTA) ______ **12**

II.II. I. Análisis Descriptivo ______ 12

II.II. II. Prueba de Barlett y Análisis de Varianza (ANOVA) ______ 14

II.II. III. Comparación Múltiple de Promedios con Prueba de Tukey ______ 15

II.III. NIVELES (60 y 120 ml) DE GLICERINA Y TIPO DE AGUA (FACTOR CON NIVELES: DESTILADA. GRIFO, AÑEJADA) COMO FACTORES Y TIEMPO PROMEDIO DE DURACIÓN COMO RESPUESTA ______ **17**

II.III. I. Preparación de Base de Datos ______ 17

II.III. II. Analizando Efectos de Interacción entre Cantidad de Glicerina y Tipo de Agua ______ 21

II.III. III. Modelo Sin Interacciones entre Cantidad de Glicerina y Tipo de Agua ______ 22

II.III. IV. Cuantificación de Diferencias ______ 23

II.III. V. Verificando la existencia de efecto entre los niveles del tipo de agua 24

II.III. VI. Analizando diferencias significativas entre pares de medias ______ 24

III. REFERENCIAS ______ *26*

I. INTRODUCCIÓN

Cuando las unidades experimentales no son muy parecidas el diseño al azar no es óptimo. En tal escenario, se pueden formar bloques de unidades similares para aumentar la sensibilidad del experimento. El agrupamiento de unidades no homogéneas en grupos más homogéneos reduce el error e incrementa el rango de validez para las inferencias acerca de los efectos de los tratamientos.

En muchas situaciones los bloques se construyen de forma natural. Se utilizan tantas unidades dentro de un bloque como tratamientos se estén analizando. Se aplican los tratamientos aleatoriamente a las unidades dentro de cada bloque.

II. CASO DE APLICACIÓN: ANÁLISIS DE RESISTENCIA PROMEDIO DE BURBUJAS

Tres investigadoras quieren analizar el efecto que tienen las diferentes proporciones de glicerina y tipo de agua utilizados en la mezcla para la confección de burbujas sobre el tiempo promedio de resistencia de estas. Se obtiene una mezcla simple y posteriormente se ajusta al propósito de la investigación. En la mezcla se va variando el tipo de agua (destilada, grifo, añejada) y la cantidad de glicerina (60ml, 120ml). A partir de estos factores se obtienen 6 combinaciones.

La unidad experimental es la burbuja de la cual se obtienen 180 observaciones en total entre las 6 personas que funcionan como bloques, de tal forma que cada persona hizo 30 burbujas.

El orden en que cada persona tiene que utilizar un tipo de mezcla se aleatoriza de la siguiente manera: por cada persona se aleatorizan los 6 tratamientos con repeticiones hasta completar 5 burbujas por tratamiento.

La variable respuesta es el tiempo (minutos) promedio de resistencia de la burbuja. Para calcular este tiempo se procede a obtener primero 5 mediciones (burbujas) por tratamiento para cada persona (bloque) y luego se promedian, es decir, la observación final por persona para cada tratamiento es el promedio de estas 5 mediciones.

El archivo “burbujas.Rdata” contiene tres bases de datos:

1. En ‘base1’ se tienen los datos para solamente con el factor ‘glicerina’.

2. En ‘base2’ se tiene solamente el factor ‘agua’ y en base se tienen ambos.

```
load("burbujas.Rdata")
```

II.I. NIVELES (60 y 120 ml) DE GLICERINA (FACTOR) Y TIEMPO PROMEDIO DE DURACIÓN (RESPUESTA)

II.I. I. Análisis descriptivo de variabilidad del tiempo de resistencia promedio de las burbujas según nivel de glicerina

La tabla contenida en ‘base1’, con seis observaciones por bloque, contiene en su primera columna el número de observación dentro del bloque, en su segunda columna la cantidad de glicerina utilizada en la elaboración de las burbujas y en su tercera columna el tiempo de duración promedio de las burbujas. Por tanto, utilizando los datos de ‘base1’, puede construirse un gráfico para verificar la variabilidad original en los dos tratamientos[1].

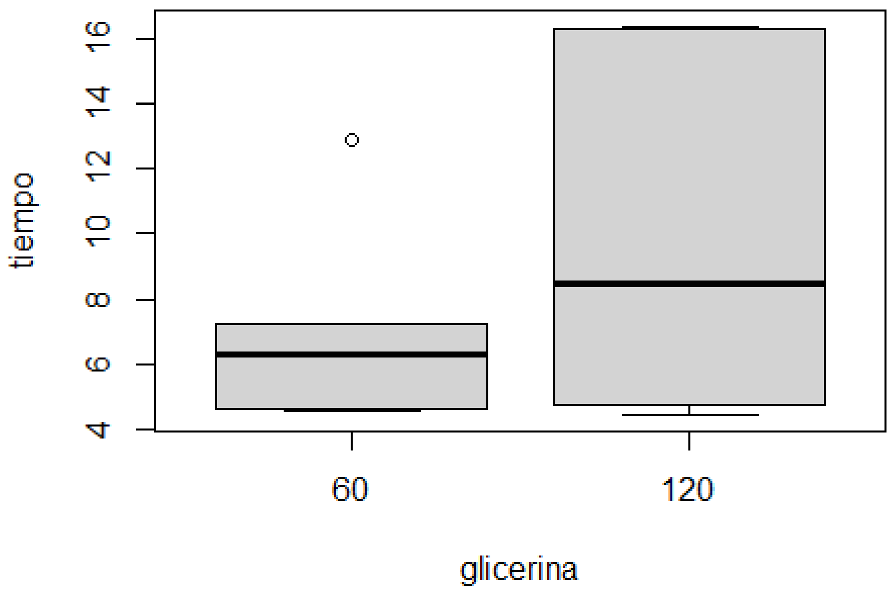
```
str(base1)

## 'data.frame':    12 obs. of  3 variables:
##  $ bloque   : int  1 2 3 4 5 6 1 2 3 4 ...
##  $ glicerina: int  60 60 60 60 60 60 120 120 120 120 ...
##  $ tiempo   : num  7.25 7.15 4.66 5.46 12.85 ...
```

[1] Con las limitaciones relativas a usar un boxplot para estudiar variabilidad, puesto que no es recomendable, salvo que se establezcan una serie de supuestos adicionales (Cross Validated, 2014).

```
base1$bloque=factor(base1$bloque)
base1$glicerina=factor(base1$glicerina)
str(base1)
```

```
## 'data.frame':    12 obs. of  3 variables:
##  $ bloque   : Factor w/ 6 levels "1","2","3","4",..: 1 2 3 4 5 6 1 2 3 4 ...
##  $ glicerina: Factor w/ 2 levels "60","120": 1 1 1 1 1 1 2 2 2 2 ...
##  $ tiempo   : num  7.25 7.15 4.66 5.46 12.85 ...
```

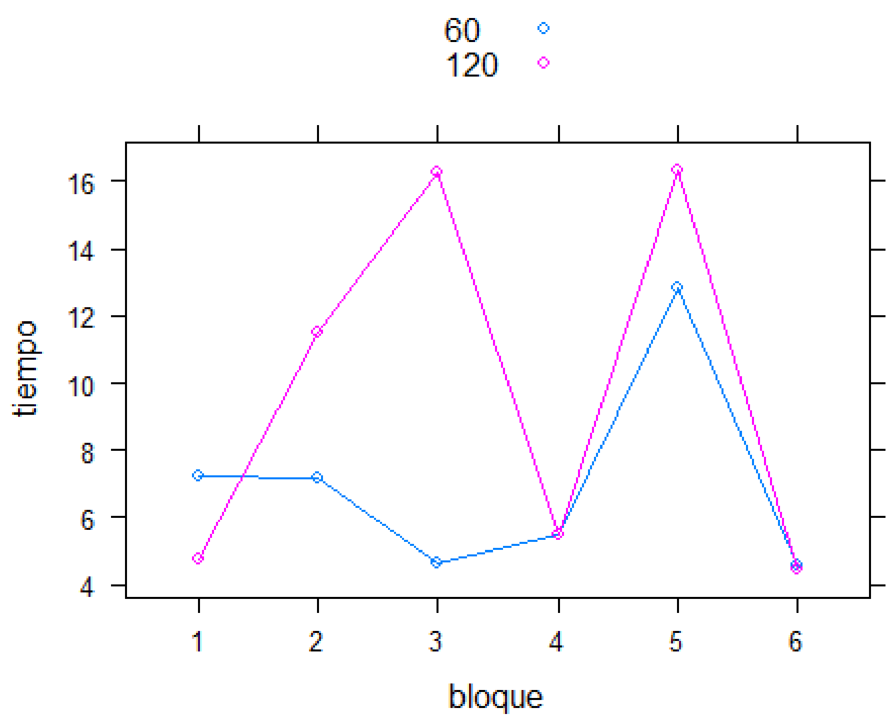
```
attach(base1)
boxplot(tiempo~glicerina)
```



II.I. II. Análisis descriptivo del tiempo promedio de resistencia de las burbujas según el nivel de glicerina

Es recomendable elaborar un gráfico de líneas para estudiar descriptivamente el comportamiento temporal de los tratamientos dentro en cada bloque. Puede usarse la sintaxis 'xyplot', con el bloque en el eje X y escribiendo "groups=glicerina" dentro de la primera sintaxis para que R dibuje una línea para cada tratamiento.

```
library(lattice)
xyplot(tiempo~bloque,groups=glicerina,type=c("p","a"),auto.key = T)
```



Preliminarmente puede afirmarse que tiende a ser mayor el tiempo cuando el nivel de glicerina es 120 que cuando es de 60.

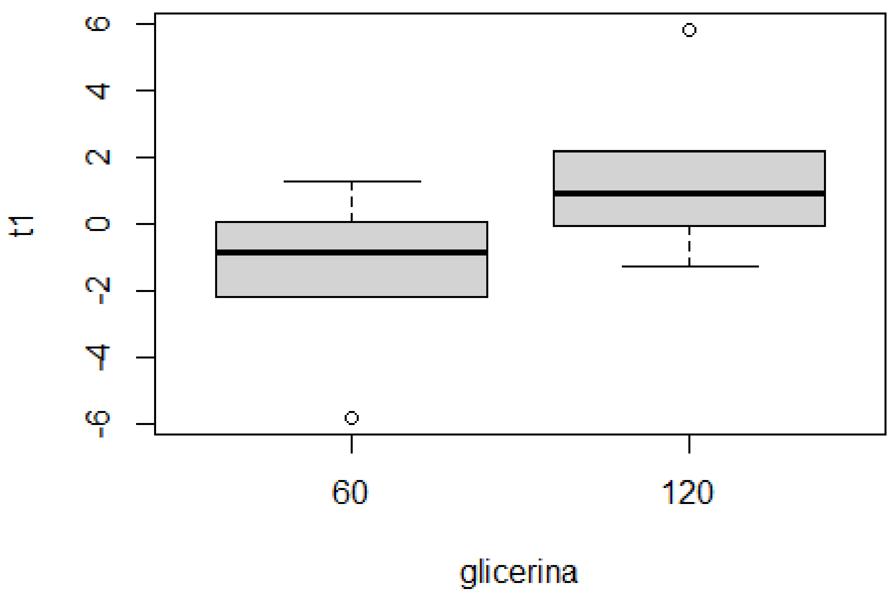
II.I. III. Datos Centrados

A continuación, se procederá a centrar los datos dentro de cada bloque, que significa eliminar el efecto del bloque restando la media del bloque al que una observación pertenece. Para centrar los datos puede ajustarse un modelo donde utilice solamente el bloque como factor y luego obtener los valores ajustados usando 'mod$fit'. Así, al restar a cada respuesta estos valores ajustados se están centrando los datos dentro de cada bloque.

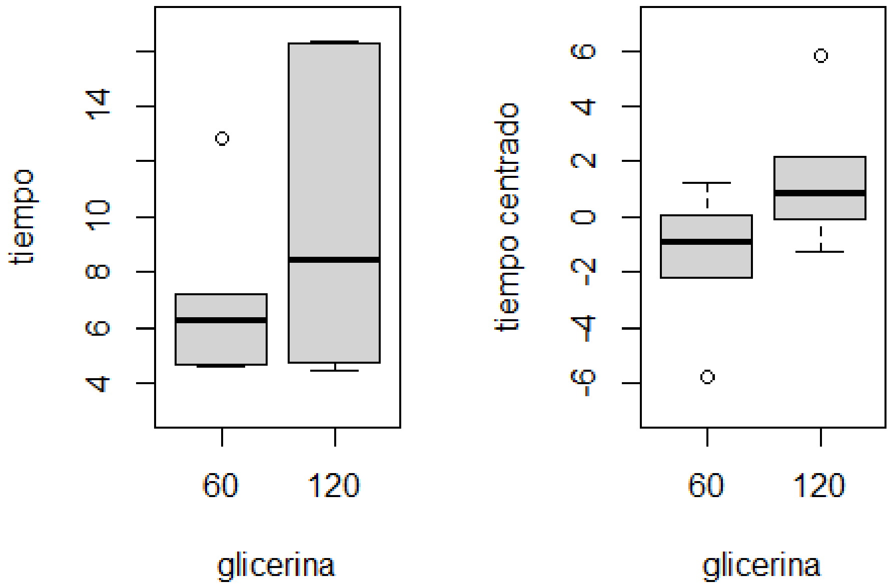
```
mod1=lm(tiempo~bloque)
pre=predict(mod1)
t1=tiempo-pre
```

Puede reelaborarse el gráfico de la sección 1 con la variable centrada. Es recomendable localizar los dos gráficos de tipo boxplot uno al lado del otro y asegurarse que ambos gráficos tengan el mismo rango en Y.

```
# Gráfico con tiempo no-centrado
boxplot(t1~glicerina)
```




```
# Gráfico con tiempo centrado
par(mfrow=c(1,2))
boxplot(tiempo~glicerina,ylim=c(3,17),ylab="tiempo")
boxplot(t1~glicerina,ylim=c(-7,7),ylab="tiempo centrado")
```

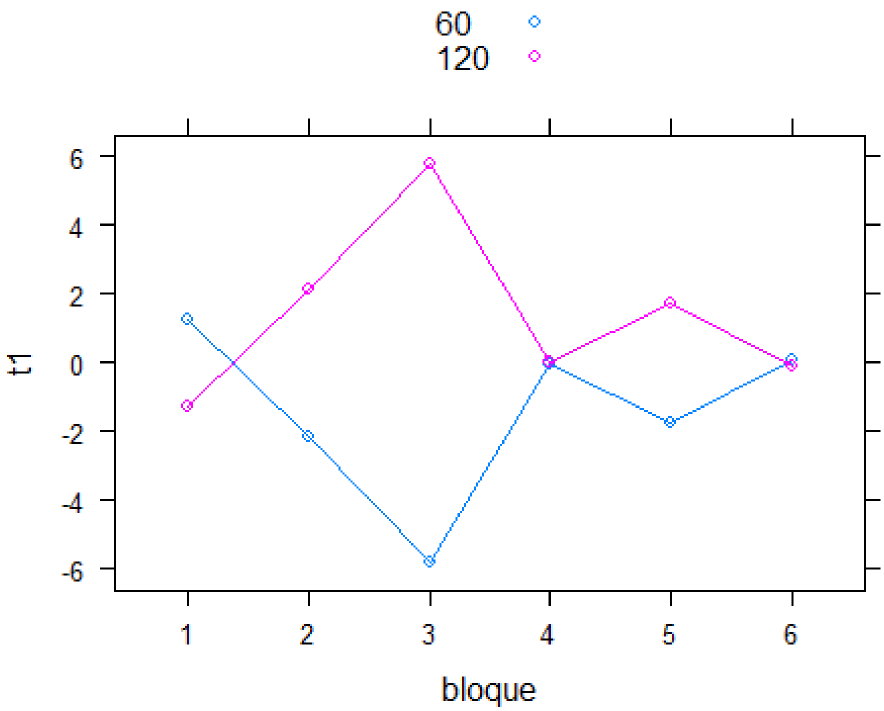



```
par(mfrow=c(1,1))
```

Se observa una gran variabilidad en el nivel 120 de glicerina en el primer gráfico, lo cual puede pensarse preliminarmente que obedece esencialmente a las diferencias entre bloques, es decir, las diferencias entre personas. Al eliminar esas diferencias (centrar los datos) se observa en el gráfico de la derecha que la variabilidad se ha reducido. Ahora la variabilidad es más parecida en los dos tratamientos, por lo que se podría pensar que el supuesto de homoscedasticidad sí se cumple, cosa que al principio podría ponerse en duda (siempre dentro del contexto descriptivo).

Adicionalmente, también puede reelaborarse el gráfico de líneas usando los datos centrados.

```
library(lattice)
xyplot(t1~bloque,groups=glicerina,type=c("p","a"),auto.key = T)
```



Al igual que en el gráfico de líneas original, se observa que los tiempos para el nivel 120 de glicerina tienden a ser mayores.

Con relación a la interacción entre glicerina y bloque, parece haber una interacción entre el factor glicerina y el bloque ya que el efecto que tiene el nivel de glicerina no es el mismo en todos los bloques, de hecho, el efecto es mucho más grande en el bloque 3. Esto podría implicar que existen diferencias en las características del proceso de creación de burbujas (capacidad pulmonar de cada persona, velocidad a la que las crea, amplitud del soplido, lugar de impacto del soplido u otros factores).

II.I. IV. Determinando la existencia (o no) de efectos de interacción

Debido a la cantidad de datos que se tienen (hay una sola repetición por tratamiento en cada bloque) no es posible verificar la hipótesis de no-interacción a pesar de que hay sospecha de una interacción. Entonces, se tiene que asumir que no hay interacción entre bloque y glicerina.

II.I. V. Modelos de Ajuste

Se utilizarán dos modelos:

1) 'mod2', que incluirá el factor de diseño y el bloque con el tiempo original como respuesta.

2) 'mod3', que incluirá sólo el factor de diseño con el tiempo centrado como respuesta.

```
mod2=lm(tiempo~glicerina+bloque)
mod3=lm(t1~glicerina)
```

II.I. VI. Sumas de Cuadrados

Es de especial interés, al comparar las tablas ANOVA de ambos modelos, analizar la suma de cuadrados total (STC, SCTot o SCT), la suma de cuadrados residual (SCE) y los grados de libertad (gl).

```
anova(mod2)

## Analysis of Variance Table
##
## Response: tiempo
##           Df  Sum Sq Mean Sq F value Pr(>F)
## glicerina  1  23.587  23.587  1.8892 0.2277
## bloque     5 145.744  29.149  2.3346 0.1868
## Residuals  5  62.427  12.485

anova(mod3)

## Analysis of Variance Table
##
## Response: t1
##           Df Sum Sq Mean Sq F value  Pr(>F)
```

```
## glicerina  1 23.587 23.5872  3.7784 0.08057 .
## Residuals 10 62.427  6.2427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sct2=sum(anova(mod2)[,2])
sct3=sum(anova(mod3)[,2])
round(c(sct2,sct3),2)
```

```
## [1] 231.76  86.01
```

```
scres2=anova(mod2)[3,2]
scres3=anova(mod3)[2,2]
round(c(scres2,scres3),2)
```

```
## [1] 62.43 62.43
```

La SCTot es diferente porque la respuesta cambió. Cuando se centran los datos se está reduciendo la variabilidad total, por lo que se tiene una menor SCTot. La SCTot original incluye la varibilidad debida a las personas (bloques), mientras que esta SCTot en el caso de datos centrados teóricamente ya no tiene esa fuente de variabilidad. Sin embargo, se mantiene la SCRes.

Así, para los dos modelos se obtiene un mismo valor de la SCRes, pero con diferentes grados de libertad. En el caso de los datos centrados, se olvida que el modelo incluyó los bloques por lo que estaría dando más grados de libertad a los residuales de los que realmente tienen.

II.I. VII. Estimación Adecuada de la Varianza Residual (Cuadrados Medios Residuales)

A pesar del hecho que los dos modelos arrojan la misma SCRes, el modelo con los datos centrados no toma en cuenta los grados de libertad que se deben atribuir a los bloques en el diseño, por lo tanto, la estimación correcta es la que da el modelo donde se incluyeron los bloques con la variable original.

```
anova(mod2)[3,3]

## [1] 12.48531
```

La estimación correcta de la variancia del error es a través de CMRes (Cuadrados Medios Residuales) del modelo, que incluye el bloque con la respuesta original, la cual es 12.49. Esta estimación de la varianza residual es una medida de la variabilidad de la respuesta dentro de cada tratamiento una vez que se ha eliminado el efecto del bloque.

II.I. VIII. Verificación del supuesto de homocedasticidad

Para hacer esta prueba no se pueden usar los tiempos originales, puesto que debe considerarse la presencia de los bloques. Hay que recordar que la homoscedasticidad opera tanto para la respuesta como para los errores, por eso en este caso lo que se recomienda es obtener los residuales del modelo con bloques y luego hacer la prueba sobre estos residuales con el factor, por ejemplo, mediante la prueba de Barlett: bartlett.test(mod2$res~glicerina).

```
bartlett.test(mod2$res~glicerina)

##
##  Bartlett test of homogeneity of variances
##
## data:  mod2$res by glicerina
## Bartlett's K-squared = 3.2297e-15, df = 1, p-value = 1
```

Se falla en rechazar H_0, que afirma que existe homocedasticidad.

II.I VIX. Conclusiones Inferenciales

```
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: tiempo
##           Df  Sum Sq Mean Sq F value Pr(>F)
## glicerina  1  23.587  23.587  1.8892 0.2277
## bloque     5 145.744  29.149  2.3346 0.1868
## Residuals  5  62.427  12.485
```

El interés es ver si el nivel de glicerina tiene un efecto sobre la resistencia promedio de las burbujas. Puesto que la probabilidad asociada es alta (0.23), no se rechaza la hipótesis de igualdad de medias entre los dos niveles de glicerina. Se concluye que no se ha encontrado un efecto del nivel de glicerina sobre la resistencia promedio de las burbujas. Debe analizarse si este experimento tiene la suficiente potencia para detectar diferencias cuando éstas existen.

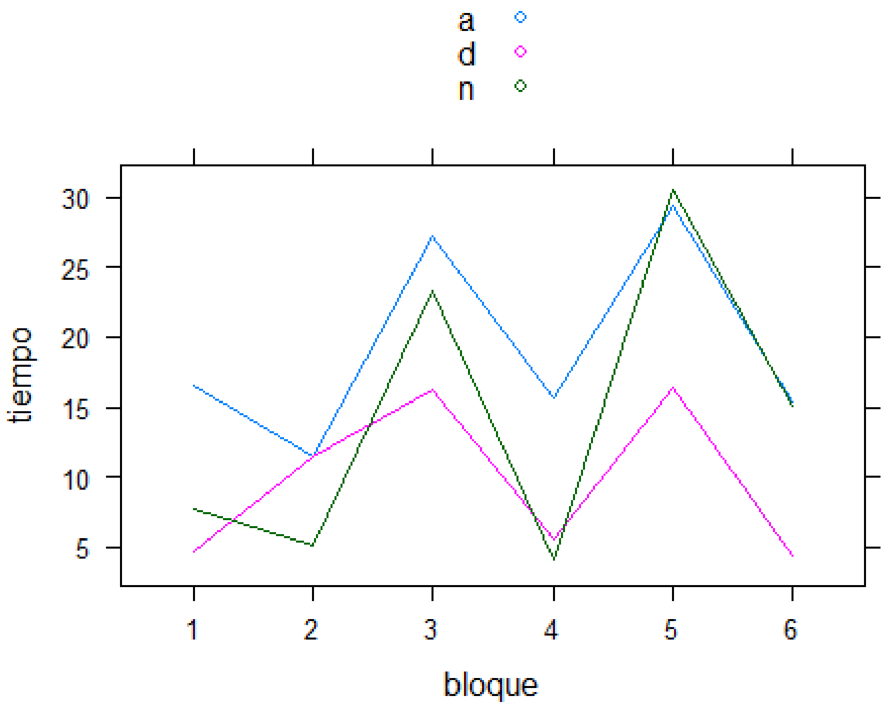
II.II. TIPO DE AGUA (FACTOR CON NIVELES: DESTILADA. GRIFO, AÑEJADA) Y TIEMPO PROMEDIO DE DURACIÓN (RESPUESTA)

II.II. I. Análisis Descriptivo

```
detach(base1)
str(base2)
```

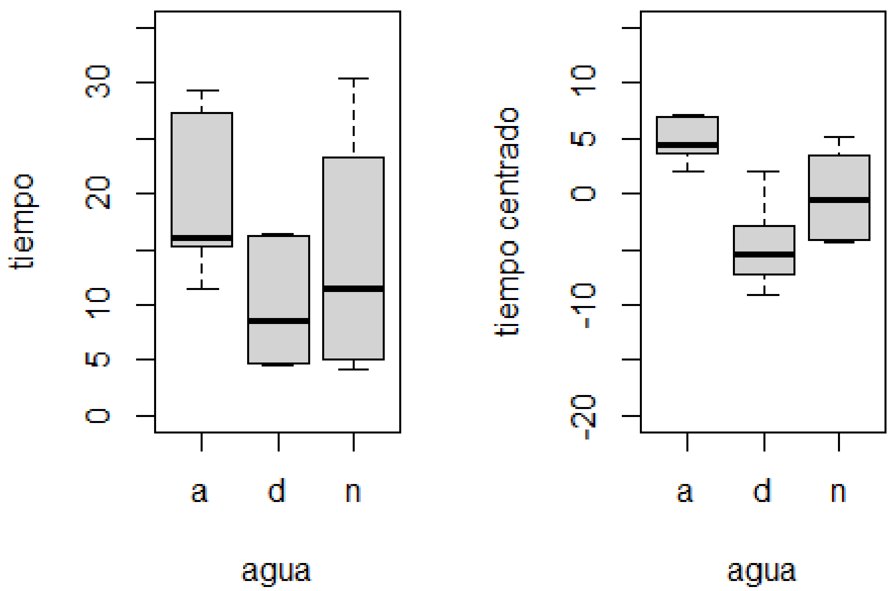
```
## 'data.frame':    18 obs. of  3 variables:
##  $ bloque: int  1 1 1 2 2 2 3 3 3 4 ...
##  $ agua  : Factor w/ 3 levels "a","d","n": 1 2 3 1 2 3 1 2 3 1 ...
##  $ tiempo: num  16.6 4.75 7.73 11.49 11.48 ...
```

```
base2$bloque=factor(base2$bloque)
attach(base2)
xyplot(tiempo~bloque,groups=agua,type="a",auto.key = T)
```



Se observa que con agua añejada (a) el tiempo promedio tiende a ser claramente mayor en relación con el agua destilada (d), aunque no parecería estar tan claro que sea diferente del agua del grifo (n).

```
# Centrando el Tiempo Promedio de Duración de Burbujas
mod4=lm(tiempo~bloque)
pre=predict(mod4)
t2=tiempo-pre
# Contruyendo Boxplot
par(mfrow=c(1,2))
boxplot(tiempo~agua,ylim=c(0,35),ylab="tiempo")
boxplot(t2~agua,ylim=c(-20,15),ylab="tiempo centrado")
```




```
par(mfrow=c(1,1))
```

Al centrar los datos se observa un promedio de tiempo mucho mayor para los datos con agua añejada que con agua destilada. Para el tratamiento con agua del grifo se observa un comportamiento intermedio.

II.II. II. Prueba de Barlett y Análisis de Varianza (ANOVA)

Puede realizarse la prueba de Bartlett se utiliza para probar si k muestras provienen de poblaciones con la misma varianza (con el fin de probar la homogeneidad de varianzas). Además, es necesario realizar el análisis de varianza.

```
mod5=lm(tiempo~agua+bloque)
bartlett.test(mod5$res~agua)

##
##  Bartlett test of homogeneity of variances
##
```

```
## data:  mod5$res by agua
## Bartlett's K-squared = 2.435, df = 2, p-value = 0.296
```

```
anova(mod5)
```

```
## Analysis of Variance Table
##
## Response: tiempo
##           Df Sum Sq Mean Sq F value   Pr(>F)
## agua       2 270.42 135.211  7.7292 0.009351 **
## bloque     5 823.60 164.720  9.4161 0.001528 **
## Residuals 10 174.93  17.493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se verifica que no se rechaza el supuesto de homoscedasticidad. Además, se rechaza la hipótesis de igualdad de promedios para los tres tipos de agua (debido a que el valor de probabilidad F del factor 'agua' indica que es significativo), con lo cual se concluye que sí existe un efecto estadísticamente significativo del tipo de agua sobre el tiempo promedio de duración de las burbujas.

II.II. III. Comparación Múltiple de Promedios con Prueba de Tukey

Puesto que el interés de la investigación es determinar cuál de los tres tipos de agua produce una resistencia mayor, conviene usar comparaciones múltiples con Tukey.

```
table(agua)
```

```
## agua
## a d n
## 6 6 6
```

```r
m=tapply(tiempo,agua,mean)
round(m,2)
```

```
##     a     d     n
## 19.29  9.80 14.33
```

```r
a.d=m[1]-m[2]
a.n=m[1]-m[3]
n.d=m[3]-m[2]
d=c(a.d,a.n,n.d)
cmres=anova(mod5)[3,3]
ee=sqrt(cmres*((1/6)+(1/6)))
q=d/ee
p=ptukey(q*sqrt(2),3,10,lower.tail = F)
names(p)=c("a-d","a-n","n-d")
round(p,3)
```

```
##   a-d   a-n   n-d
## 0.007 0.150 0.196
```

Se ha encontrado que el promedio del tiempo con agua añejada es mayor que el de agua destilada, pero el promedio con agua del grifo no se diferencia de ninguno de los otros dos (valores de probabilidad mayores que un nivel de significancia de $\alpha = 0.05$. Finalmente, se va a estimar la cota inferior de la diferencia que se puede encontrar entre los promedios donde se encontraron diferencias estadísticamente significativas, esto a través del valor t y del error estándar.

```r
t=qt(0.95,10)
lim=d[1]-t*ee[1]
round(lim,2)
```

```
##    a
## 5.11
```

Se puede esperar con un 0.95 de confianza que el tiempo promedio cuando se hacen las burbujas con agua añejada sea al menos 5.11 minutos mayor que con agua destilada.

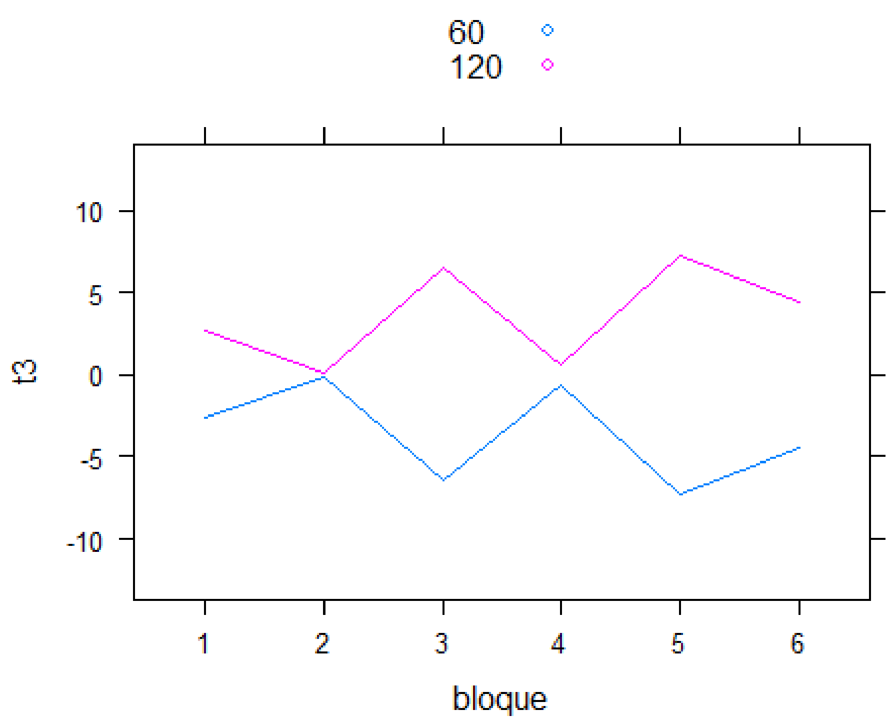
II.III. NIVELES (60 y 120 ml) DE GLICERINA Y TIPO DE AGUA (FACTOR CON NIVELES: DESTILADA. GRIFO, AÑEJADA) COMO FACTORES Y TIEMPO PROMEDIO DE DURACIÓN COMO RESPUESTA

II.III. I. Preparación de Base de Datos

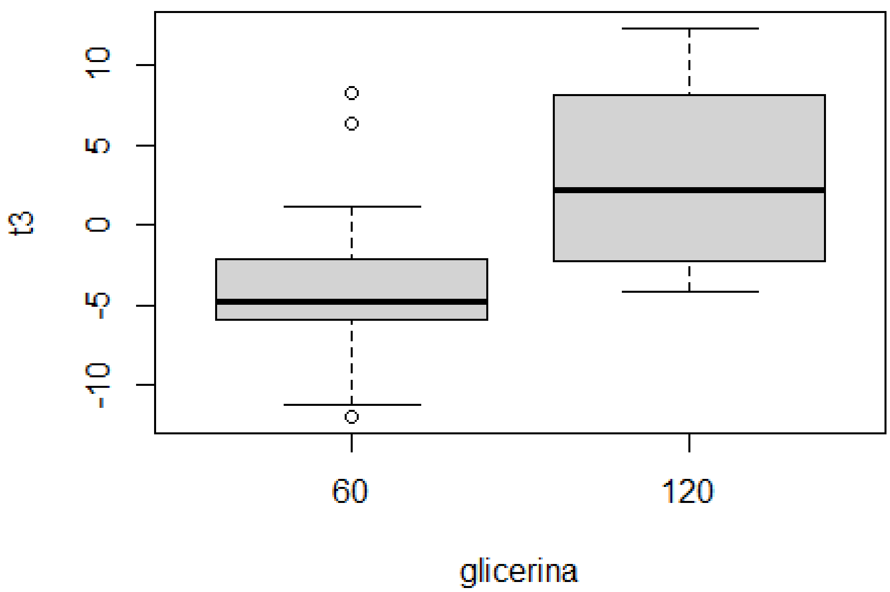
Debe verificarse que los factores están bien definidos, así como visualizar los datos. Como se trata de dos factores, se pueden centrar los datos por bloque y hacer un boxplot con cada factor. También se pueden hacer gráficos de líneas para observar el comportamiento según niveles de la cantidad de glicerina y según niveles de agua.

```
str(base)

## 'data.frame':    36 obs. of  5 variables:
##  $ bloque     : int  1 1 1 1 1 1 2 2 2 2 ...
##  $ glicerina  : int  60 120 60 120 60 120 60 120 60 120 ...
##  $ agua       : Factor w/ 3 levels "a","d","n": 1 1 2 2 3 3 1 1 2 2 ...
##  $ tratamiento: int  1 2 3 4 5 6 1 2 3 4 ...
##  $ tiempo     : num  4.54 16.6 7.25 4.75 1.28 ...

base$bloque=factor(base$bloque)
base$glicerina=factor(base$glicerina)
str(base)

## 'data.frame':    36 obs. of  5 variables:
##  $ bloque     : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 1 1 1 2 2 2 2 ...
```

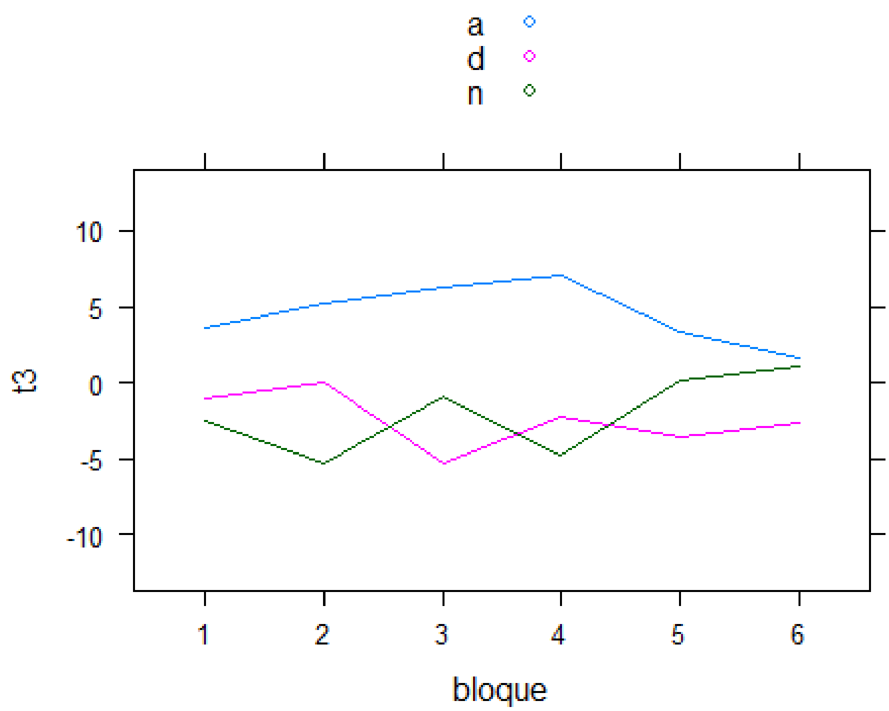
```
## $ glicerina : Factor w/ 2 levels "60","120": 1 2 1 2 1 2 1 2 1 2 ...
## $ agua      : Factor w/ 3 levels "a","d","n": 1 1 2 2 3 3 1 1 2 2 ...
## $ tratamiento: int 1 2 3 4 5 6 1 2 3 4 ...
## $ tiempo     : num 4.54 16.6 7.25 4.75 1.28 ...

detach(base2)
attach(base)

#Modelo de Ajuste
mod6=lm(tiempo~bloque)

# Centrando el tiempo promedio de duración de las burbujas
t3=tiempo-predict(mod6)

# Gráfico con Factor Cantidad de Glicerina
xyplot(t3~bloque,groups=glicerina,type=c("a"),auto.key = T)
```



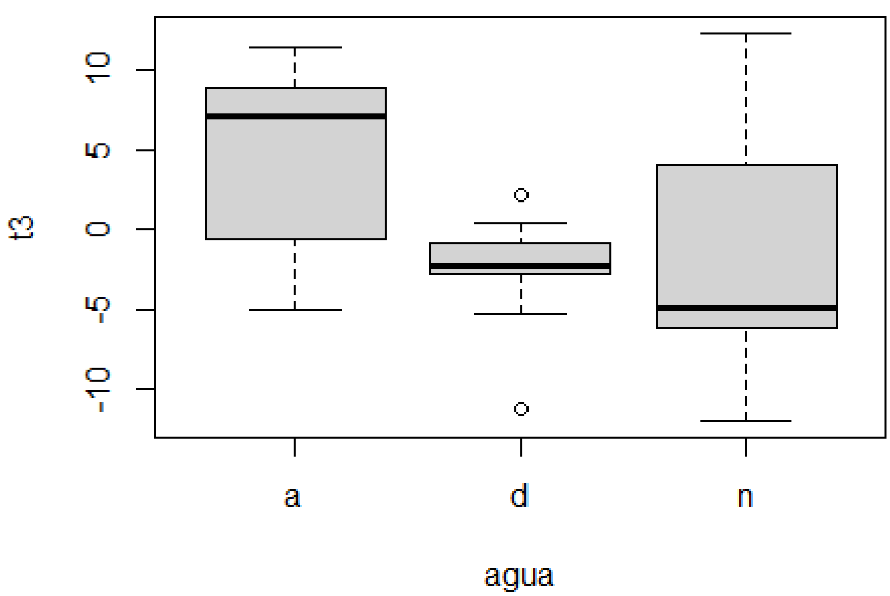
```
boxplot(t3~glicerina)
```




```
# Gráfico con Factor Tipo de Agua
xyplot(t3~bloque,groups=agua,type=c("a"),auto.key = T)
```



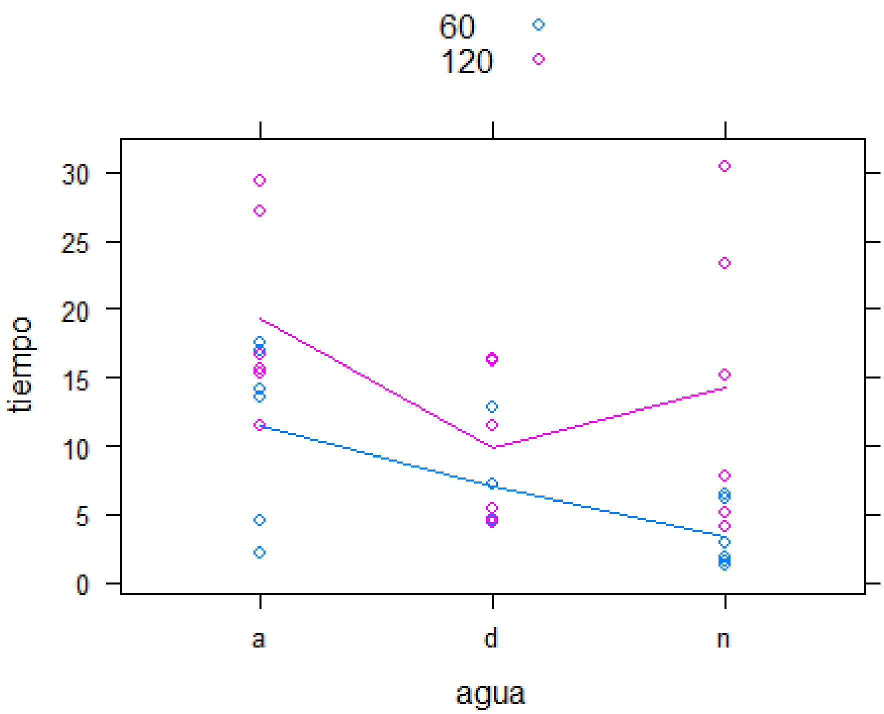

```
boxplot(t3~agua)
```



Se observa que el promedio de tiempo parece ser mayor cuando el nivel de glicerina es 120 y cuando el agua es añejada.

II.III. II. Analizando Efectos de Interacción entre Cantidad de Glicerina y Tipo de Agua

```
xyplot(tiempo~agua,groups=glicerina,auto.key = T,type=c("a","p"))
```



Parece haber interacción entre glicerina y agua, ya que cuando se tiene agua destilada la diferencia entre los promedios para los dos niveles de glicerina es un poco menor que en los otros dos casos (porque las curvas morada y azul se acercan más en la sección d, que corresponde al agua destilada -para agua de grifo las diferencias son mayores según el nivel de glicerina, mientras que son menores para agua añejada-), sin embargo, puede ser que haya mucha variabilidad en el error y esa interacción no sea tan clara. Esto podría deberse a su composición molecular u otra variable que no se está tomando en consideración.

```
mod7=lm(tiempo~glicerina*agua+bloque)
anova(mod7)
```

```
## Analysis of Variance Table
##
## Response: tiempo
##                Df Sum Sq Mean Sq F value    Pr(>F)
## glicerina       1 464.35  464.35 19.9916 0.0001467 ***
## agua            2 367.73  183.87  7.9161 0.0021713 **
## bloque          5 706.48  141.30  6.0833 0.0008139 ***
## glicerina:agua  2 101.26   50.63  2.1797 0.1340892
## Residuals      25 580.68   23.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Al hacer la prueba formal de la hipótesis de no-interacción entre glicerina y agua se obtiene una fiabilidad en la probabilidad de error tipo I (0.13) más alta que el nivel de significancia $\alpha = 0.05$, por lo que no se puede afirmar que haya interacción. En adelante, se puede asumir que estos dos factores no tienen interacción entre sí.

II.III. III. Modelo Sin Interacciones entre Cantidad de Glicerina y Tipo de Agua

Debe verificarse si existe efecto de la glicerina en este nuevo diseño.

```
mod8=lm(tiempo~glicerina+agua+bloque)
anova(mod8)
```

```
## Analysis of Variance Table
##
## Response: tiempo
##           Df Sum Sq Mean Sq F value    Pr(>F)
## glicerina  1 464.35  464.35 18.3849 0.0002059 ***
## agua       2 367.73  183.87  7.2799 0.0029602 **
## bloque     5 706.48  141.30  5.5944 0.0011569 **
## Residuals 27 681.93   25.26
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En este caso, se observa que con mayor cantidad de glicerina se obtiene un tiempo promedio de las burbujas independiente del tipo de agua (esto debido a que los coeficientes son significativos al $\alpha = 0.01$ (y aún a valores más bajos). La probabilidad asociada al factor glicerina es pequeña ($p < 0.001$) por lo que se rechaza la hipótesis de igualdad de las medias según niveles de glicerina.

II.III. IV. Cuantificación de Diferencias

Deben cuantificarse las diferencias, por ejemplo, con un nivel de confianza de 0.95.

```
table(glicerina)

## glicerina
##  60 120
##  18  18

m=tapply(tiempo,glicerina,mean)
round(m,2)

##    60   120
##  7.29 14.47

#d es la diferencia entre las medias en el tiempo de duración con una cantidad de glicerina de 60 ml y de 120 ml
d=m[2]-m[1]
cmres=anova(mod8)[4,3]
ee=sqrt(2*cmres/18)
t=qt(0.95,27)
lim=d-t*ee
round(lim,2)
```

```
##  120
## 4.33
```

Con 0.95 de nivel de confianza se espera que el tiempo promedio de las burbujas sea al menos 4.33 minutos mayor cuando se usa un nivel de glicerina de 120 ml con respecto a cuando se usa un nivel de 60, con independencia del tipo de agua que se utilice.

II.III. V. Verificando la existencia de efecto entre los niveles del tipo de agua

La probabilidad asociada al factor agua es pequeña ($p = 0.003$), por lo que se rechaza la hipótesis de igualdad de las medias según tipos de agua.

II.III. VI. Analizando diferencias significativas entre pares de medias

Es recomendable estudiar entre cuáles pares de medias existen diferencias y cuantificar aquellas donde las diferencias son estadísticamente significativas. Puede usarse un nivel de signficancia de 0.05 y un nivel de confianza de 0.95.

```
table(agua)
```

```
## agua
##  a  d  n
## 12 12 12
```

```
m=tapply(tiempo,agua,mean)
round(m,2)
```

```
##     a     d     n
## 15.39  8.40  8.85
```

```
a.d=m[1]-m[2]
a.n=m[1]-m[3]
n.d=m[3]-m[2]
d=c(a.d,a.n,n.d)
```

```
ee=sqrt(2*cmres/12)
q=d/ee
p=ptukey(q*sqrt(2),3,27,lower.tail = F)
names(p)=c("a-d","a-n","n-d")
round(p,3)

##   a-d   a-n   n-d
## 0.006 0.010 0.973

t=qt(0.95,27)
lim=d[1:2]-t*ee
names(lim)=names(p)[1:2]
round(lim,2)

##  a-d  a-n
## 3.50 3.05

detach(base)
```

Se concluye con un nivel de significancia de 0.05 que el promedio de tiempo con agua añejada es mayor que el promedio con cualquiera de los otros dos tipos de agua. Además, se espera con 0.95 de nivel de confianza que el promedio de tiempo con agua añejada sea al menos 3.5 minutos más que con agua destilada y al menos 3.05 minutos más que con agua del grifo, esto independientemente del nivel de glicerina que se use.

En particular, puesto que el nivel de glicerina que produce el mayor tiempo es 120ml, se tendrá una mayor resistencia promedio con agua añejada que sobrepasa los 3 minutos (en promedio) el tiempo que se obtiene con los otros tipos de agua. Finalmente, usar 120ml de glicerina sobrepasa (en promedio) en más de 4 minutos de duración promedio de las burbujas con relación a si se usa un nivel de glicerina de 60ml.

III. REFERENCIAS

Cross Validated. (27 de Marzo de 2014). *Deduce variance from boxplot*. Obtenido de Questions: https://stats.stackexchange.com/questions/91536/deduce-variance-from-boxplot