PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA

<u>EXPERIMENTACIÓN NUMÉRICA</u>
Examen 1
(Ciclo de verano 2024)

Solucionario

<u>Pregunta 1</u> (4 puntos)

En una empresa se está evaluando instituir un descanso de 15 minutos en la mañana para que los trabajadores hagan algunos ejercicios de estiramiento. La compañía cree que esto podría mejorar la satisfacción de los trabajadores. Para analizar esto, se seleccionan 20 trabajadores aleatoriamente y se registra la satisfacción de los trabajadores antes y después de la implementación de estos ejercicios. La satisfacción es medida a través de un cuestionario en donde se obtiene un índice de satisfacción. Los resultados para los veinte trabajadores antes y después de realizar los ejercicios de estiramiento están en el archivo "**estiramiento.csv**".

Realice la(s) prueba de hipótesis conveniente(s) para ayudar a decidir si los ejercicios de estiramiento son efectivos para mejorar la satisfacción del trabajador. Escriba su conclusión y recomendaciones.

Sean
$X =$ el índice de la satisfacción de los trabajadores antes de la implementación de estos ejercicios
$Y =$ el índice de la satisfacción de los trabajadores después de la implementación de estos ejercicios

De acuerdo al objetivo de la investigación, el contraste de hipótesis es:

$$\begin{cases} H_0: \mu_Y = \mu_X \\ H_1: \mu_Y > \mu_X \end{cases} \text{, o equivalentemente: } \begin{cases} H_0: \mu_Y - \mu_X = 0 \\ H_1: \mu_Y - \mu_X > 0 \end{cases} \quad \text{o} \quad \begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D > 0 \end{cases}$$

Claramente, las muestras de X e Y son pareadas, de modo que bastará trabajar con las diferencias: $D = Y - X$, observadas en cada unidad experimental.

Para comenzar, para cada trabajador, se define la variable Dif	estiramiento$Dif <- estiramiento$Despues - estiramiento$Antes
	
	head(estiramiento)
	`  Id_trabajador Antes Despues Dif`
	`1             1    24      57  33`
	`2             2    37      48  11`
	`3             3    14      43  29`
	`4             4    38      14 -24`
	`5             5    12      79  67`
	`6             6    19      63  44`

<table>
<tr>
<td>A la variable Dif se someterá a la prueba de la normalidad, para probar:

$$\begin{cases} \mathrm{H}_0: D \sim N(\mu_D\,, \sigma_D^2) \\ \mathrm{H}_1: D \nsim N(\mu_D\,, \sigma_D^2) \end{cases}$$

Aplicando la prueba Shapiro-Wilk, se obtuvo un p-valor=0.99, lo que indica que no se debe rechazar la normalidad de la diferencia de los índices de satisfacción, por lo que se acepta la normalidad.</td>
<td>

```
> normalityTest(~Dif,
test="shapiro.test",
data=estiramiento)
Shapiro-Wilk normality test
data:  Dif
W = 0.98727, p-value = 0.9922
```

</td>
</tr>
<tr>
<td>Dada la normalidad, y teniendo una muestra pequeña de la diferencia ($n = 20$), se aplica la prueba t −student, para el contraste de hipótesis con cola a la derecha:

$$\begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D > 0 \end{cases}$$

Resultando el p-valor=0.001081.
Por lo tanto, se debe rechazar la hipótesis nula a favor de la hipótesis alternativa.</td>
<td>

```
> with(estiramiento, (t.test(Dif,
alternative = "greater", mu = 0.0,
conf.level = .95)))
        One Sample t-test
data:  Dif
t = 3.5451, df = 19, p-value =
0.001081
alternative hypothesis: true mean
is greater than 0
95 percent confidence interval:
 9.29725     Inf
sample estimates:
mean of x
    18.15
```

</td>
</tr>
</table>

Conclusión: los ejercicios de estiramiento mejoran la satisfacción del trabajador.

Pregunta 2 (5 puntos)

Se realiza un experimento para analizar el desempeño de un nuevo diseño de la página web de una empresa. Para este experimento se recolectó una muestra de 40 personas, 20 de ellas interactuaron con el diseño actual y otras 20 personas interactuaron con el nuevo diseño. Los tiempos de interacción (en segundos) se registraron en el archivo "**nuevodiseno.csv**".

a) (1 punto) Indique si las muestras son relacionadas o independientes.

Como se trata de personas diferentes asignadas a cada grupo, se tiene que las muestras son independientes.

b) (4 puntos) Los diseñadores creen que el nuevo diseño incentivará a las personas a pasar mayor tiempo en la página web. Realice la prueba de hipótesis correspondiente y adecuada para contrastar la hipótesis de los diseñadores.

Sean
$X =$ el tiempo de interacción de una persona que interactuó con el diseño actual (en segundos)
$Y =$ el tiempo de interacción de una persona que interactuó con el nuevo diseño (en segundos)

De acuerdo al objetivo de la investigación, el contraste de hipótesis es:

$$\begin{cases} H_0: \mu_Y = \mu_X \\ H_1: \mu_Y > \mu_X \end{cases}$$

Verificaremos los supuestos de las pruebas paramétricas:

Cada variable se debe someter a la prueba de la normalidad: $\begin{cases} \mathrm{H}_0: X \sim N(\mu_X, \sigma_X^2) \\ \mathrm{H}_1: X \nsim N(\mu_X, \sigma_X^2) \end{cases}$ y $\begin{cases} \mathrm{H}_0: Y \sim N(\mu_Y, \sigma_Y^2) \\ \mathrm{H}_1: Y \nsim N(\mu_Y, \sigma_Y^2) \end{cases}$
Aplicando la prueba Shapiro-Wilk, se tiene que ambos p-valores son mayores que el nivel de significancia $\alpha = 0.05$, lo que indica que, para ambos grupos no se debe rechazar la normalidad del tiempo de interacción.
	`normalityTest(tiempo ~ pagina, test="shapiro.test", data=d)`
	`p-values adjusted by the Holm method:`
	`       unadjusted adjusted`
	`Actual 0.42283    0.84566`
	`Nuevo  0.90809    0.90809`

Ahora probaremos la homocedasticidad, es decir la igualdad de varianzas. Como se trata de apenas 2 grupos, se puede aplicar la prueba de razón de varianzas, para probar el contraste de hipótesis a 2 colas: $\begin{cases} \mathrm{H}_0: \dfrac{\sigma_Y^2}{\sigma_X^2} = 1 \\ \mathrm{H}_1: \dfrac{\sigma_y^2}{\sigma_X^2} \neq 1 \end{cases}$
Resultando un p-valor =0.002014, menos que el nivel de significancia $\alpha = 0.05$, por lo que se rechaza la hipótesis nula, es decir, se rechaza que las varianzas son iguales.
Conclusión: se considera que las varianzas son diferentes.
	`var.test(tiempo ~ pagina, alternative='two.sided', Rcmdr+    conf.level=.95, data=d)`
	`F test to compare two variances`
	`data:  tiempo by pagina`
	`F = 4.469, num df = 19, denom df = 19, p-value = 0.002014`
	`alternative hypothesis: true ratio of variances is not equal to 1`
	`95 percent confidence interval:`
	` 1.768889 11.290740`
	`sample estimates:`
	`ratio of variances`
	`          4.469012`

Dada la independencia de muestras, la normalidad de las muestras y la diferencia de varianzas, se está en el caso 3 de pruebas.

<table>
<tr>
<td>
Se debe aplicar la prueba t, pero considerando varianzas diferentes, para probar:

$$\begin{cases} H_0: \mu_Y = \mu_X \\ H_1: \mu_Y > \mu_X \end{cases}$$

Los resultados de la prueba muestran un $p - valor = 0.003237$, que es menor al nivel de significancia $\alpha = 0.05$, por lo que se rechaza la hipótesis nula a favor de la alternativa.
</td>
<td>

```
t.test(tiempo ~ pagina, alternative = "less",
conf.level = .95,var.equal = FALSE, data = d)
        Welch Two Sample t-test
data:  tiempo by pagina
t = -2.9503, df = 27.098, p-value = 0.003237
alternative hypothesis: true difference in
means between group Actual and group Nuevo is
less than 0
95 percent confidence interval:
     -Inf -10.8013
sample estimates:
mean in group Actual  mean in group Nuevo
              425.90               451.45
```

</td>
</tr>
</table>

Conclusión: el tiempo de interacción de los usuarios con la página del nuevo diseño es mayor que el tiempo de interacción de los usuarios con la página del diseño actual.

Pregunta 3 (6 puntos)

Se realiza un experimento para analizar el tiempo que toman tres métodos para la ubicación de contactos en un smarthphone: búsqueda por palabra, búsqueda por desplazamiento, y por reconocimiento por voz. Los datos de este experimento se encuentran en el archivo "**busqueda.csv**". La columna "Time" muestra el tiempo en segundos que toma la búsqueda de un grupo de contactos. La variable "Technique" muestra las tres técnicas que se utilizaron para buscar contactos: "Search": búsqueda por palabra, "Scroll": búsqueda por desplazamiento, y "Voice": búsqueda por voz.

a) (4 puntos) Realice una prueba adecuada para para probar si al menos una de las técnicas da un resultado diferente?

Y_1 = tiempo en segundos que lleva la búsqueda al usar la técnica Search.
Y_2 = tiempo en segundos que lleva la búsqueda al usar la técnica Scroll.
Y_3 = tiempo en segundos que lleva la búsqueda al usar la técnica Voice.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu$$
$$H_1: \exists i \mid \mu_i \neq \mu$$

Son 3 muestras independientes.

Probando la normalidad del tiempo en segundos que lleva la búsqueda, separando por grupo de personas, definidos por las técnicas usadas.

$$H_0: \forall\, i = 1,2,3: \ Y_i \sim N(\mu_i, \sigma_i^2)$$
$$H_1: \exists\, i \mid Y_i \nsim N(\mu_i, \sigma_i^2)$$

Se aplicó la prueba de Shapiro, los p-valores fueron mayores que el nivel de significancia es $\alpha = 0.05$. Por lo tanto, en ninguno de los grupos se rechaza la normalidad del tiempo que lleva la búsqueda.

```
normalityTest(Time ~ Technique,
test="shapiro.test", data=d)
p-values adjusted by the Holm method:
          unadjusted adjusted
Scroll 0.092134   0.18427
Search 0.724730   0.72473
Voice  0.060395   0.18119
```

Probando la igualdad de varianzas (homocedasticidad)

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$
$$H_1: \exists\, i \mid \sigma_i^2 \neq \sigma^2$$

Se aplicó la prueba de Bartlett, el p-valor fue menor que el nivel de significancia es $\alpha = 0.05$. Por lo tanto, se rechaza la homocedasticidad.

```
Tapply(Time ~ Technique, var,
na.action=na.omit, data=d)
Rcmdr+    # variances by group
   Scroll    Search     Voice
1282.2737  539.6421  134.4711

bartlett.test(Time ~ Technique, data=d)
        Bartlett test of homogeneity of
variances
data:  Time by Technique
Bartlett's K-squared = 20.275, df = 2, p-value
= 0.00003956
```

Dado que se cumplen los 2 supuestos: independencia y normalidad, pero no el supuesto de la homocedasticidad, se aplica la prueba de Welch, para el contraste de hipótesis:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu$$
$$H_1: \exists i \mid \mu_i \neq \mu$$

Analizando los resultados, el p-valor=0.000002291 del tiempo de búsqueda según la técnica, indica que la técnica es un factor significante para el tiempo de búsqueda. Así, que se rechaza la hipótesis de que los tiempos medios de búsqueda son iguales.

```
modelo <- aov(Time ~ Technique, data = d)
summary(modelo)
            Df Sum Sq Mean Sq F value       Pr(>F)
Technique    2  30448   15224   23.34 0.0000000393 ***
Residuals   57  37171     652

with(d, numSummary(Time, groups = Technique,
statistics=c('mean', 'sd')))
          mean       sd data:n
Scroll 137.20 35.80885     20
Search  96.80 23.23020     20
Voice   84.45 11.59617     20

oneway.test(Time ~ Technique, data = d) # Welch test
One-way analysis of means (not assuming equal variances)

data:  Time and Technique
F=20.038, num df=2.000, denom df=31.961, p-
value=0.000002291
```

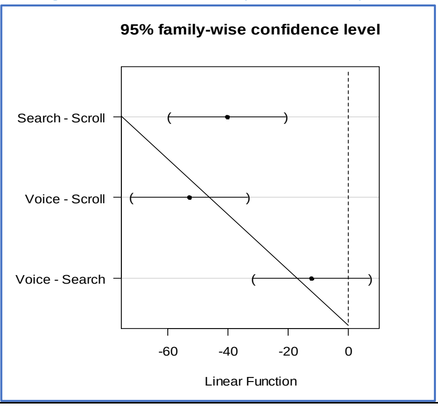
Conclusión: al menos una de las técnicas resulta con una media de tiempo de búsqueda diferente.

b) (2 puntos) Explore, mediante alguna técnica gráfica, cuál es la mejor técnica y si algunas de las técnicas se pueden considerar semejantes en relación a los resultados del tiempo de búsqueda.

Analizando las estadísticas descriptivas, se tiene que para el grupo que hizo la búsqueda por desplazamiento la media del tiempo de búsqueda mayor fue de 137.20 segundos, seguido del grupo que hizo a búsqueda por palabra, obteniendo una media de 96.80 segundos y, por último, para el grupo que hizo la búsqueda por voz, obtuvo la menor media del tiempo de búsqueda siendo 84.45.

```
with(d, numSummary(Time, groups =
Technique, statistics=c('mean',
'sd')))
         mean       sd data:n
Scroll 137.20 35.80885     20
Search  96.80 23.23020     20
Voice   84.45 11.59617     20
```

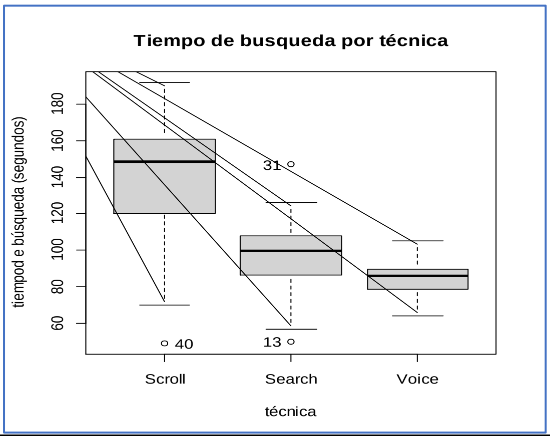
Analizando los intervalos de confianza para la diferencia de medias, se tiene la búsqueda por desplazamiento es significantemente mayor que las otras búsquedas.
Siendo mejor en menor tiempo medio la búsqueda hecha por voz y la búsqueda hecha por palabra.
Además, ya que el intervalo para media de la búsqueda por voz menos la media de la búsqueda por palabra contiene el cero, se puede decir que ambas técnicas tienen un efecto igual sobre el tiempo de búsqueda.




```
local({.Pairs  <-  glht(modelo,  linfct  =  mcp(Technique  =
"Tukey"))
print(summary(.Pairs)) # pairwise tests
print(confint(.Pairs,level=0.95))#confidence intervals
print(cld(.Pairs,level=0.05))#compact letter display
old.oma <- par(oma=c(0, 5, 0, 0))
plot(confint(.Pairs))
par(old.oma) })

Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts
Fit: aov(formula = Time ~ Technique, data = d)
Linear Hypotheses:
                     Estimate Std. Error t value Pr(>|t|)
Search - Scroll == 0  -40.400      8.075  -5.003  <0.0001 ***
Voice - Scroll == 0   -52.750      8.075  -6.532  <0.0001 ***
Voice - Search == 0   -12.350      8.075  -1.529    0.285

(Adjusted p values reported -- single-step method)
          Simultaneous Confidence Intervals
Multiple Comparisons of Means: Tukey Contrasts
Fit: aov(formula = Time ~ Technique, data = d)
Quantile = 2.4068
95% family-wise confidence level
Linear Hypotheses:
                     Estimate lwr      upr
Search - Scroll == 0 -40.4000 -59.8357 -20.9643
Voice - Scroll == 0  -52.7500 -72.1857 -33.3143
Voice - Search == 0  -12.3500 -31.7857   7.0857

Scroll Search  Voice
   "a"    "b"    "b"
```

Otra alternativa es analizando los diagramas de caja:
Definitivamente hay diferencias entre los tiempos de búsquedas. Pues hay una diferencia marcada entre los niveles de las cajas. Pero hay superposición entre las cajas de los tiempos de búsqueda para las técnicas search y voice, eso indica que la diferencia no están significante.

```
Rcmdr>  Boxplot(Time ~ Technique, data=d, id=list(method="y"),
xlab="técnica",
Rcmdr+    ylab="tiempod e búsqueda (segundos)", main="Tiempo de
busqueda por técnica")
[1] "40" "13" "31"
```



<table>
<tr>
<td>
Dado que hay presencia de datos extremos, se recomienda realizar la prueba no paramétricas.
Al realizar una prueba de Kruskal-Wallis para el contraste de hipótesis sobre la igualdad de medianas:
$$\mathrm{H_0}: Me_1 = Me_2 = Me_3 = Me$$
$$\mathrm{H_1}: \exists \mathrm{i} \mid Me_i \neq Me$$
El resultado de la prueba tiene un p-valor=0.000003679, se rechaza la hipótesis nula.
Conclusión una de las medianas del tiempo de búsqueda es diferente.
</td>
<td>

```
Tapply(Time ~ Technique, median,
na.action=na.omit, data=d)
Rcmdr+    # medians by group
Scroll Search  Voice
 148.5   99.5   86.0

Rcmdr>  kruskal.test(Time ~ Technique, data=d)
         Kruskal-Wallis rank sum test
data:  Time by Technique
Kruskal-Wallis chi-squared = 25.026, df = 2, p-
value = 0.000003679
```

</td>
</tr>
</table>

Conclusión: La técnica utilizada para la búsqueda constituye un factor que influye significantemente sobre el tiempo de búsqueda. Resultado que el menor tiempo de búsqueda es cuando se usa la voz, con una media de 84 segundos, mientras que el segundo menor tiempo medio se obtiene cuando se usa la búsqueda por palabra, que tiene una media de 97 segundos, para estos, por voz y por palabra, se puede considerar que la diferencia entre las medias no es significante. Por último, la peor técnica de búsqueda es por desplazamiento que tiene en media 137 segundos, el mayor tiempo.