CI0131 Diseño de experimentos I Ciclo 2025
Ignacio Díaz Oreiro

Laboratorio. ANOVA de una vía

Con base en el material estudiado en clase, vamos a realizar un laboratorio sobre ANOVA de una vía.

Este laboratorio consta de 2 partes.

En una primera parte realizaremos un ejercicio guiado, y adicionalmente se explicarán algunos conceptos que complementan lo visto en clase.

En la segunda parte, usted deberá construir un experimento utilizando de base tanto el conocimiento adquirido en clase, como la experiencia de la primera parte del laboratorio.

La primera parte no es necesario entregarla. Solamente se entrega la segunda parte.

Primera parte – Ejercicio guiado y conceptos

En este primer experimento se quiere determinar si **tres** programas de ejercicio diferentes tienen un impacto diferente en la pérdida de peso.

La variable independiente que se estamos estudiando es el programa de ejercicios y la variable de respuesta es la pérdida de peso (weight_loss), medida en libras.

Se va a realizar un ANOVA de una vía para determinar si existe una diferencia estadísticamente significativa entre la pérdida de peso resultante de los tres programas.

Se reclutaron 90 personas para que participen en un experimento en el que se asignan aleatoriamente a 30 personas para que sigan el programa A, 30 el programa B y 30 el programa C durante un mes.

1. Cargue el archivo **weight**.csv en RStudio

```
weight= (read.csv(file.choose(), header=T, encoding = "UTF-8"))
attach(weight)
```

La conversión de la columna a factor (as.factor) facilita las operaciones posteriores, asegurando que la variable es considerada como un factor (variable categórica).

```
weight$program <- as.factor(weight$program)
```

2. Revise las primeras 6 líneas del dataframe

```
head(weight)
```

```
> head(weight)
  obs program weight_loss
1   1       A    1.887342
2   2       C    2.548101
3   3       B    3.618555
4   4       A    2.061069
5   5       A    1.493098
6   6       B    3.342334
```

La primera columna del dataframe muestra el programa (A, B , C) en el que participó la persona durante un mes y la segunda columna muestra la pérdida de peso total que experimentó esa persona al final del programa, medida en libras.

3. Calculamos la media y la desviación estándar para cada tratamiento

Explorando los datos

Antes de que se ajuste (fit) el modelo ANOVA unidireccional, podemos obtener una mejor comprensión de los datos al encontrar la media y la desviación estándar de la pérdida de peso para cada uno de los tres programas que utilizan el paquete dplyr:

```
#load dplyr package
library(dplyr)
```

```
weight %>%
 group_by(program) %>%
 summarise(mean = mean(weight_loss),
           sd = sd(weight_loss))
```

```
# A tibble: 3 × 3
  program  mean    sd
  <fct>   <dbl> <dbl>
1 A        1.58 0.905
2 B        2.56 1.24
3 C        4.13 1.57
```

Puede verse que la media menor es para el grupo A, mientras que la media mayor corresponde al grupo C. Este patrón se repoite también para las desviaciones estándar: la menor es del grupo A y la mayor es la del grupo C.

4. También creamos un diagrama de cajas para cada uno de los tres programas para visualizar la distribución de la pérdida de peso para cada programa:

```
#create boxplots
boxplot(weight_loss ~ program,
      data = weight, # data2,
      main = "Pérdida de peso por programa",
      xlab = "Programa",
      ylab = "Pérdida de peso",
      col = "steelblue",
      border = "black")
```

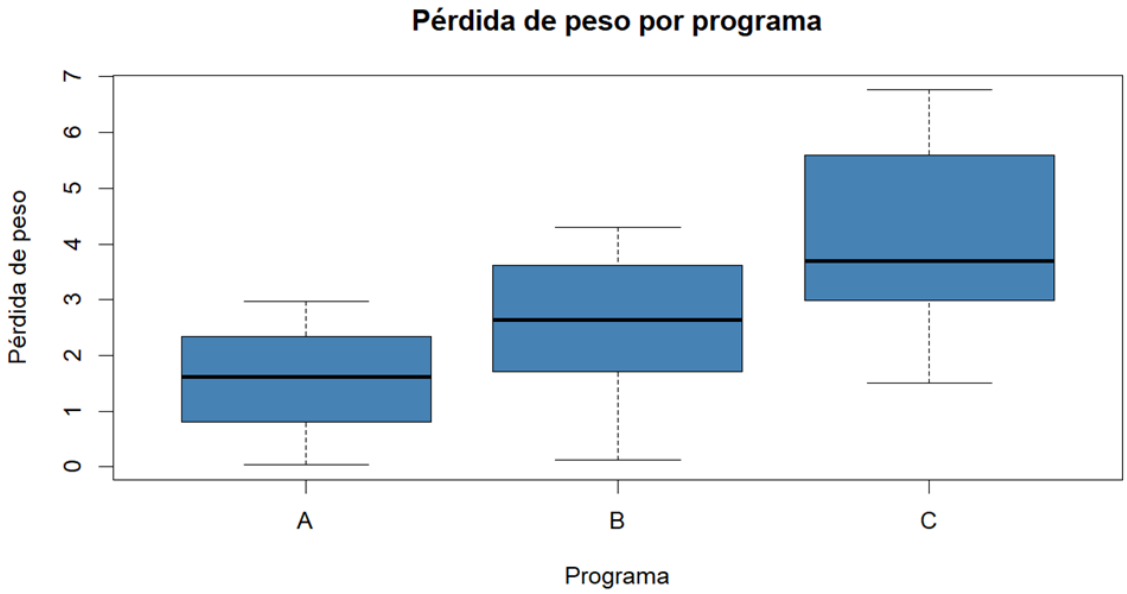

Figura 1. Boxplots de pérdida de peso para los programas A, B y C

A partir de estos diagramas de caja podemos ver que la mediana de la pérdida de peso es más alta para los participantes en el Programa C y más baja para los participantes en el Programa A.

También podemos ver que el rango intercuartil (la "longitud" del diagrama de caja) para la pérdida de peso es más alta en el Programa C en comparación con los otros dos programas.

A continuación, ajustaremos el modelo ANOVA unidireccional a nuestros datos para ver si estas diferencias visuales son estadísticamente significativas.

Definiremos un **nivel de significancia de 0.05** para este ANOVA y en general para todas las pruebas del experimento.

5. Modelo de datos e hipótesis nula y alternativa

Primero definiremos el modelo de datos para este ANOVA. Se usará el modelo de los efectos:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{Bmatrix} i = 1,2,3 \\ j = 1,2,\dots,30 \end{Bmatrix}$$

Donde μ es un parámetro común a todos los tratamientos (la media global) y τ_i es el efecto del tratamiento *i*-ésimo, donde i = 1 corresponde al programa de ejercicios A, i = 2 corresponde al programa B e i = 3 corresponde al programa C. Además, *j* = 1, 2, …, 30 corresponde a cada una de

las repeticiones que se realizan. Adicionalmente, los errores del modelo (ε_{ij}) representan una variable aleatoria que sigue una distribución normal e independiente con media 0 y varianza σ^2.

Seguidamente definimos la hipótesis alternativa y la hipótesis nula:

$$H_0: \tau_1 = \tau_2 = \tau_3 = 0$$

$$H_1: \tau_i \neq 0 \; para \; algún \; i \in \{1, 2, 3\}$$

Es decir, la hipótesis nula indica que el efecto de los programas de ejercicio A, B y C son 0 para los tres tratamientos, y la hipótesis alternativa indica que alguno de los 3 programas de ejercicio sí tiene un efecto distinto de 0 en la variable de respuesta.

Dicho de otra manera (con el modelo de las medias), sería que las medias de los 3 programas de ejercicios son iguales, mientras que la hipótesis alternativa indica que al menos dos medias son distintas.

6. Ajuste del modelo ANOVA unidireccional

Ahora ajustaremos el modelo ANOVA unidireccional en R usando la sintaxis siguiente:

```
aov(variable_respuesta ~ variable_independiente, data = dataset)
```

En este ejercicio, creamos el siguiente código para ajustar el modelo ANOVA unidireccional, usando la pérdida de peso como variable de respuesta y el programa como nuestra variable de predicción.

```
#fit the one-way ANOVA model

model <- aov(weight_loss ~ program, data = weight)
```

Luego podemos usar la función summary() para ver el resultado de nuestro modelo:

```
summary(model)
```

```
> summary(model)
            Df Sum Sq Mean Sq F value   Pr(>F)
program      2  98.93   49.46   30.83 7.55e-11 ***
Residuals   87 139.57    1.60
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A partir de la salida del modelo, podemos ver que el programa de la variable predictora es estadísticamente significativo en el nivel de significancia de 0.05, dado que 7.55e-11 es menor que 0.05, por lo que se rechaza la hipótesis nula.

En otras palabras, los efectos de los programas de ejercicio no son iguales a 0 en los tres casos, por lo que sí existe una diferencia estadísticamente significativa entre la pérdida de peso promedio que resulta de los tres programas.

7. Comprobación de los supuestos del modelo

Antes de continuar, debemos verificar que se cumplan los supuestos de nuestro modelo para que los resultados del modelo sean confiables. En particular, un ANOVA unidireccional asume:

1. **Independencia**: las observaciones de cada grupo deben ser independientes entre sí. Dado que utilizamos un diseño aleatorio (es decir, asignamos a los participantes a los programas de ejercicios al azar), esta suposición debe cumplirse para que no tengamos que preocuparnos demasiado por esto. Esto puede verificar también con los residuales.

2. **Normalidad**: los residuales del modelo deben tener una distribución aproximadamente normal para cada nivel de la variable predictora.

3. **Igualdad de varianzas**: las varianzas de los residuales son iguales o aproximadamente iguales.

Todos estos supuestos los evaluaremos usando los **residuales del modelo** y no las observaciones que son parte de la muestra.

Una forma de diagnosticar los supuestos de normalidad e igualdad de varianza es usar la función **plot**(), que produce **cuatro** gráficos sobre los **residuales del modelo**.

```
plot(model)
```

NOTA: Debe dar ENTER en la Consola de RStudio para que se vayan desplegando los 4 gráficos.

En particular, estamos interesados en los siguientes dos:

Q-Q plot: esta gráfica muestra los residuos estandarizados frente a los cuantiles teóricos. Podemos usar esta gráfica para medir aproximadamente si se cumple o no la suposición de normalidad.

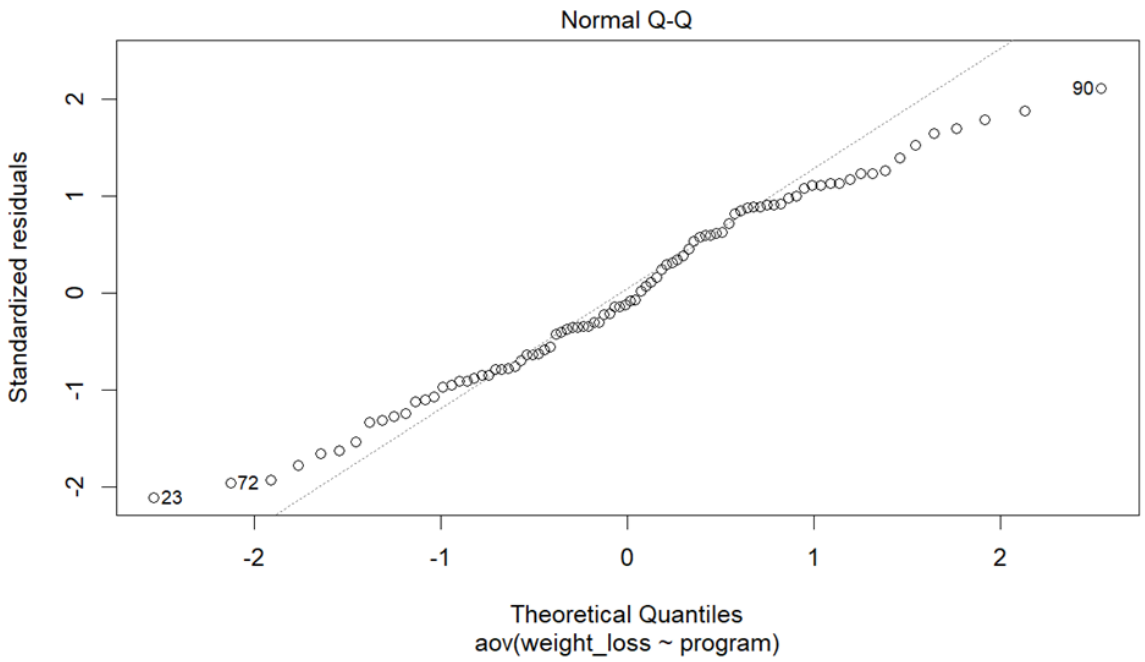

Figura 2. QQ plot de residuales para el modelo de pérdida de peso para los programas A, B y C

El gráfico Q-Q anterior nos permite verificar la suposición de **normalidad**. Idealmente, los residuos estandarizados caerían a lo largo de la línea diagonal recta en la gráfica.

Sin embargo, en el gráfico anterior podemos ver que los residuos se desvían bastante de la línea hacia el principio y el final. Esta es una indicación de que nuestra suposición de normalidad podría no cumplirse.

Residuals vs fitted: este gráfico muestra la relación entre los residuales y los valores ajustados (predichos). Podemos usar esta gráfica para medir aproximadamente si la varianza entre los grupos es igual o no.

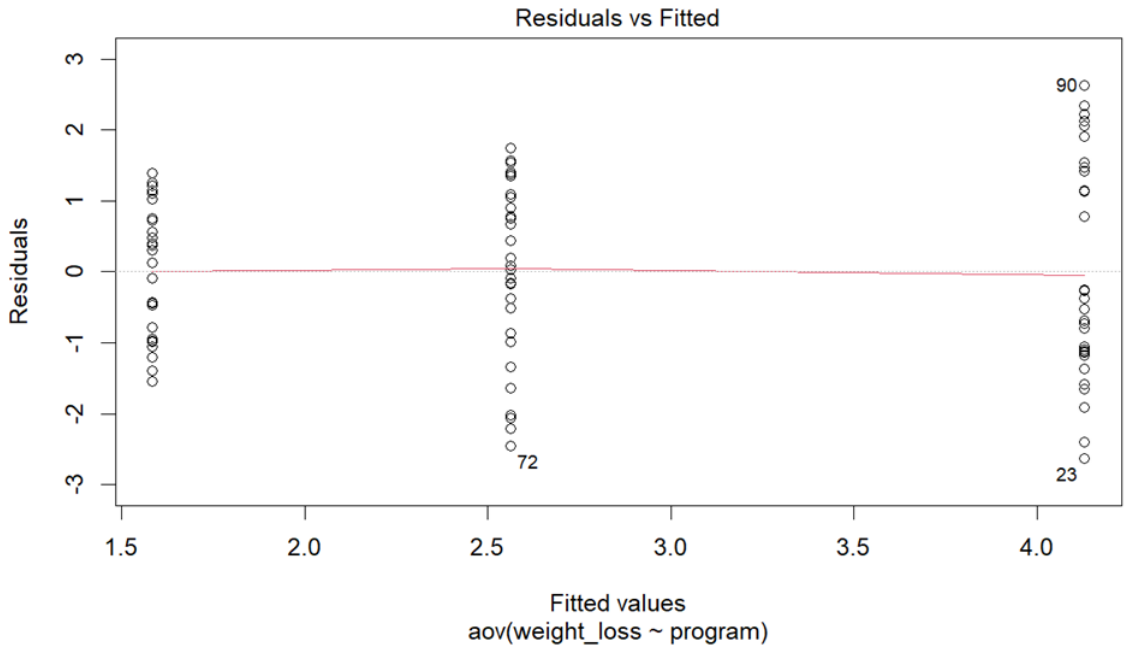

Figura 3. Gráfico de residuales versus valores ajustados para el modelo de pérdida de peso para los programas A, B y C

La gráfica anterior de Residuals vs Fitted nos permite verificar nuestra suposición de **igualdad de varianzas**. Idealmente, nos gustaría ver que los residuales se distribuyan por igual para cada nivel de los valores ajustados.

La línea roja corresponde a una línea de suavizado (loess: locally estimated scatterplot smoothing). Esta línea es una estimación suavizada que te ayuda a identificar patrones en la relación entre los residuos y los valores ajustados.

Si la línea roja muestra una curva o una tendencia notable (por ejemplo, en forma de U o S), esto puede indicar problemas con el supuesto de igualdad de varianzas. Si la línea es aproximadamente horizontal y cercana a cero, significa que no hay indicios de problemas graves.

Podemos ver que los residuos están más dispersos para los valores ajustados del tercer grupo, lo que es una indicación de que la igualdad de varianzas podría no darse. Sin embargo, la línea de suavizado es bastante plana y cercana a cero, lo que hace pensar que no habría problemas graves con este supuesto.

Aunque este gráfico no está diseñado para verificar la **independencia**, si se observa un patrón sistemático (por ejemplo, una tendencia clara, como una curva o una secuencia en los residuos), esto podría ser una señal de que los residuos no son independientes.

Otra opción sería crear un gráfico de residuales en función de un componente temporal que haya sido utilizado. Por ejemplo, el orden de las observaciones en la tabla podría responder a la aleatorización de la captura de las observaciones. En ese caso se podría crear un gráfico con los residuales en función del tiempo para revisar la independencia temporal. La **independencia** implica que los residuos no deben mostrar correlaciones en función del tiempo, representado por el orden en que se realizó la observación.

Esto también tiene sentido cuando el experimento tiene algún componente temporal.

Usaremos la columna "obs" que guarda el orden en que se hizo la observación, y bajo el entendido de que los tratamientos A, B y C fueron completamente aleatorizados.

```
# Se toman los residuales del modelo
residuos <- residuals(model)

# Se crea un gráfico de Residuales vs Tiempo (variable obs)
plot(obs, residuos,
    main = "Residuales vs Tiempo",
    xlab = "Tiempo",
    ylab = "Residuales",
    pch = 20,      # Tipo de punto
    col = "blue")  # Color de los puntos

# Se añade una línea horizontal en y = 0 para facilitar la interpretación
abline(h = 0, col = "red", lty = 2)
```

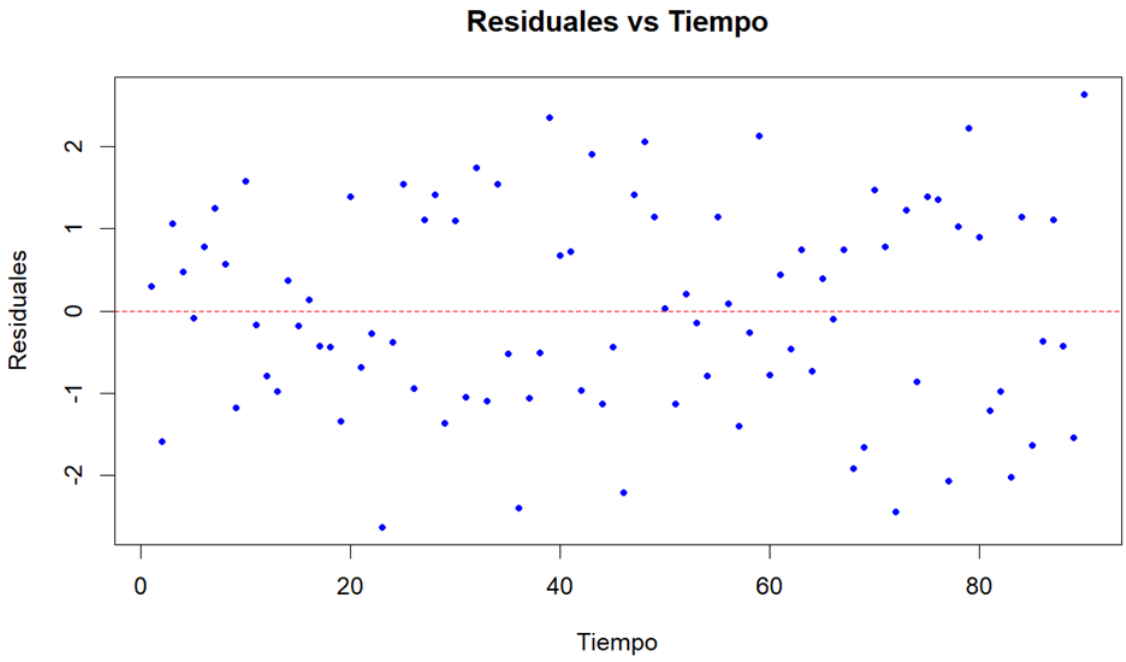

Figura 4. Gráfico de residuales en el orden en que fueron tomadas las observaciones aleatorizadas, para el modelo de pérdida de peso para los programas A, B y C

Puede verse que no aparecen patrones específicos en los datos (como por ejemplo despliegues en forma de embudo), por lo que la independencia no parece estar comprometida.

Si las filas del data frame coinciden con el orden de la variable **obs**, es decir, el orden en el dataframe corresponde con la manera aleatorizada en que se tomó la muestra, se puede hacer este mismo gráfico sin la variable **obs**:

```
# Se crea un gráfico de Residuales vs Tiempo (variable obs)
plot(residuos,
    main = "Residuales vs Tiempo",
    xlab = "Tiempo",
    ylab = "Residuales",
    pch = 20,      # Tipo de punto
    col = "blue")  # Color de los puntos
```

Los otros dos gráficos del comando **plot**(model) son:

El gráfico de Escala-Local (Scale-Location), que es similar al de residuales versus residuales ajustados, por lo que se puede usar para verificar homocedasticidad, pero es más sensible a las variaciones en la dispersión.

El gráfico de residuales versus la influencia (leverage), que sirve para identificar puntos influyentes.

8. Para probar formalmente el supuesto de normalidad debemos ejecutar la prueba Shapiro Wilks sobre **los residuales**. Usamos el mismo razonamiento de hipótesis nula e hipótesis alternativa. En este caso la hipótesis nula es que los residuales siguen una distribución normal, mientr que la alternativa sería que los residuales no siguen una distribución normal.

```
shapiro.test(model$residuals)
```

```
> shapiro.test(model$residuals)

        Shapiro-Wilk normality test

data:  model$residuals
W = 0.98024, p-value = 0.1876
```

No se puede rechazar la hipótesis nula de que los residuales siguen una distribución normal, para un nivel de significancia de 0.05.

También debemos probar formalmente la igualdad de varianzas. Podemos usar las pruebas de **Levene** o de **Bartlett**. La prueba de Bartlett es más poderosa si los datos son normalmente distribuidos, como en este caso.

```
# Bartlett

bartlett.test(weight_loss ~ program, data = weight)
```

```
> bartlett.test(weight_loss ~ program, data = weight)

        Bartlett test of homogeneity of variances

data:  weight_loss by program
Bartlett's K-squared = 8.2713, df = 2, p-value = 0.01599
```

También se puede usar la Prueba de Levene usando el paquete car:

```
#load car package

library(car)

leveneTest(weight_loss ~ program, data = weight)
```

```
> leveneTest(weight_loss ~ program, data = weight)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  4.1716 0.01862 *
      87
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

En ambos casos el valor p de la prueba es menor a 0.05, por lo que rechazaríamos la hipótesis nula de que las varianzas son iguales en los tres programas.

Aunque podríamos intentar transformar los datos para asegurarnos de que se cumplen nuestros supuestos de normalidad e igualdad de varianzas, por el momento no nos preocuparemos por esto para efectos de este ejercicio.

En cuanto a **independencia**, tomaremos como bueno el supuesto de que los datos se tomaron de forma aleatorizada y que el gráfico de residuales contra el tiempo no mostró patrones particulares. Existen algunas pruebas estadísticas como Durbin-Watson para probar independencia, pero se usan en contextos particulares como regresión lineal o series de tiempo.

9. Análisis de las diferencias de tratamiento

Una vez que hayamos verificado que los supuestos del modelo se cumplen (o se cumplen razonablemente), volvemos a la conclusión del modelo ANOVA, que indicó que la variable predictora es estadísticamente significativo en el nivel de significancia de 0.05, dado que 7.55e-11 es menor que 0.05.

Es decir, se rechazó la hipótesis nula de que los efectos son todos 0, o lo que es equivalente, que las medias de los 3 tratamientos no son iguales dos a dos.

Para determinar cuáles medias son distintas realizamos una prueba post hoc para determinar exactamente qué grupos de tratamiento difieren entre sí.

Para nuestra prueba post hoc, usaremos la función TukeyHSD() para realizar la prueba de Tukey para comparaciones múltiples:

TukeyHSD(model, conf.level=.95)

```
> TukeyHSD(model, conf.level=.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = weight_loss ~ program, data = weight)

$program
         diff       lwr      upr     p adj
B-A 0.9777414 0.1979466 1.757536 0.0100545
C-A 2.5454024 1.7656076 3.325197 0.0000000
C-B 1.5676610 0.7878662 2.347456 0.0000199
```

El valor p indica si existe o no una diferencia estadísticamente significativa entre cada par de programas.

Podemos ver en el resultado que los 3 valores son menores a 0.05, por lo que sí hay una diferencia estadísticamente significativa entre la pérdida de peso media si comparamos los programas dos a dos (B y A, C y A, C y B), en el nivel de significancia de 0.05.

10. También podemos visualizar los intervalos de confianza del 95 % que resultan de la prueba de Tukey usando la función plot(TukeyHSD()) en R:

```
plot(TukeyHSD(model, conf.level=.95))
```

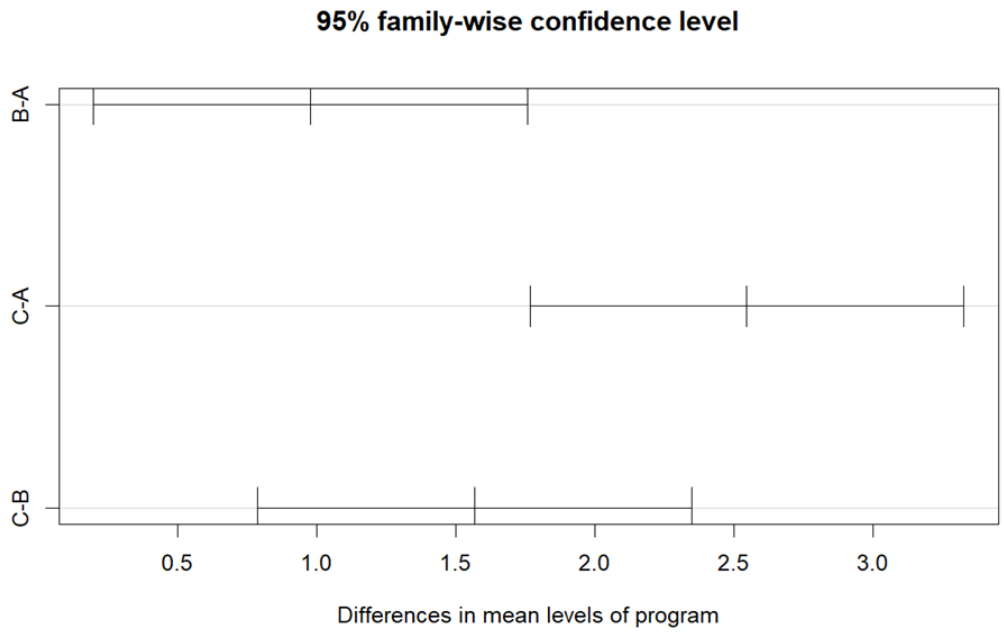

Figura 5. Intervalos de confianza para las comparaciones múltiples del modelo de pérdida de peso para los programas A, B y C

Los resultados de los intervalos de confianza son consistentes con los resultados de las pruebas de hipótesis.

En particular, podemos ver que ninguno de los intervalos de confianza para la pérdida media de peso entre programas contiene el valor cero, lo que indica que existe una diferencia estadísticamente significativa en la pérdida media entre los tres programas.

Esto es consistente con el hecho de que todos los valores p de nuestras pruebas de hipótesis están por debajo de 0.05.

11. Reportando los resultados de una ANOVA de una vía

Se deben reportar los resultados del ANOVA unidireccional de tal manera que resuma los hallazgos. En el ejemplo que estamos realizando se podría decir:

Se realizó un ANOVA de una vía para examinar los efectos del programa de ejercicios sobre la pérdida de peso (medido en libras). Hubo una diferencia estadísticamente significativa entre los efectos de los tres programas sobre la pérdida de peso (F(2, 87) = 30.83, p = 7.55e-11). Se realizaron pruebas post hoc HSD de Tukey.

La pérdida de peso media de los participantes en el programa C es significativamente mayor que la pérdida de peso media de los participantes en el programa B (p = 0.0000199), IC del 95 % = [0.7878662, 2.347456].

La pérdida de peso media de los participantes en el programa C es significativamente mayor que la pérdida de peso media de los participantes en el programa A (p < 0.0000001), IC del 95 % = [1.7656076, 3.325197].

Además, la pérdida de peso media de los participantes del programa B es significativamente mayor que la pérdida de peso media de los participantes del programa A (p = 0.0100545), IC del 95 % = [0.1979466, 1.757536].

Nótese que el Tukey de C y A reportó un valor p de 0.0000000, dado que la precisión máxima del Tukey no alcanzó para mostrar un valor. Por esta razón en los resultados reportamos que el valor p es menor que un umbral determinado.

También es importante indicar el tamaño del efecto (η2) y la potencia de la prueba. Estos dos valores se explican a continuación.

Tamaño del efecto *η*2 (eta al cuadrado)

El tamaño del efecto de un ANOVA, es el valor que permite medir cuanta varianza en la variable dependiente cuantitativa es resultado de la influencia de la variable cualitativa independiente, o lo que es lo mismo, cuanto afecta la variable independiente (factor) a la variable dependiente.

En el ANOVA la medida del tamaño del efecto más empleada es η^2 que se define como:

$$\eta^2 = SS_{entre\ grupos} / SS_{total}$$

Los niveles de clasificación más empleados para el tamaño del efecto son:

0.01 = pequeño

0.06 = mediano

0.14 > grande

Los valores necesarios para calcular η2 se obtienen del summary del ANOVA y se pueden calcular manualmente.

12. También, en R puede obtenerse mediante la función etaSquared() de paquete lsr.

```
#install.packages(lsr)
library(lsr)
etaSquared(model, anova=TRUE)
```

```
> etaSquared(model, anova=TRUE)
             eta.sq eta.sq.part        SS df        MS        F            p
program   0.4147959   0.4147959  98.92613  2 49.463064 30.83304 7.548695e-11
Residuals 0.5852041          NA 139.56740 87  1.604223       NA           NA
```

En este caso, puede verse que el efecto de la variable *program* es grande, reflejado en el eta.sq = 0.4147959.

Potencia (power) de un ANOVA de una vía

Los test de potencia permiten determinar la probabilidad de encontrar diferencias significativas entre las medias para un determinado α indicando el tamaño de los grupos, o bien calcular el tamaño que deben de tener los grupos para ser capaces de detectar con una determinada probabilidad una diferencia en las medias si esta existe.

En aquellos casos que se quiere conocer el tamaño que han de tener las muestras, es necesario conocer (bien por experimentos previos o por muestras piloto) una estimación de la varianza de la población.

La función power.anova.test() del paquete stats realiza el cálculo de potencia para modelos de ANOVA equilibrados (mismo n para cada tratamiento).

13. Para determiner la potencia:

power.anova.test(groups = 3, n = 30, between.var = 98.92, within.var = 139.56, sig.level = 0.05)

```
> power.anova.test(groups = 3, n = 30, between.var = 98.92, within.var = 139.56, sig.level
= 0.05)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 30
    between.var = 98.92
     within.var = 139.56
      sig.level = 0.05
          power = 0.9999778

NOTE: n is number in each group
```

En este caso, con n = 30 para cada grupo, y las varianzas identificadas entre grupos y dentro del grupo, la potencia de la prueba es 0.9999778, lo que es muy bueno.

14. Para determinar el tamaño de n necesario (o suficiente) para una potencia de prueba específica, no se incluye el n = 30 y se agrega power = al valor deseado. Por ejemplo, si quiero encontrar el tamaño de n para una potencia de 0.80:

power.anova.test(groups = 3, between.var = 98.92, within.var = 139.56, power = 0.8, sig.level = 0.05)

```
> power.anova.test(groups = 3, between.var = 98.92, within.var = 139.56, power = 0.8,
sig.level = 0.05)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 7.886207
    between.var = 98.92
     within.var = 139.56
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

En este caso, hubiera bastado una muestra de n = 8 en cada grupo para identificar las diferencias significativas en una población con las varianzas identificadas entre grupos y dentro del grupo.

En resumen

Se utiliza un ANOVA de una vía para determinar si existe o no una diferencia estadísticamente significativa entre las medias de tres o más grupos independientes.

Cuando informamos los resultados de un ANOVA unidireccional, siempre usamos la siguiente estructura general:

- Una breve descripción de la variable independiente y dependiente.
- El valor F general del ANOVA, los grados de libertad y el valor p correspondiente.
- Los resultados de las comparaciones post-hoc (si el valor p fue estadísticamente significativo).

Aquí está la redacción exacta que podemos usar:

- Se realizó un ANOVA de una vía para comparar el efecto de [variable independiente] sobre [variable dependiente].

- Un ANOVA de una vía reveló que [había o no] una diferencia estadísticamente significativa en [variable dependiente] entre al menos dos grupos (F(grados de libertad entre grupos, grados de libertad dentro de los grupos) = [valor F], p = [p -valor]).

- La prueba HSD de Tukey para comparaciones múltiples encontró que el valor medio de [variable dependiente] fue significativamente diferente entre [nombre del grupo] y [nombre del grupo] (p = [valor p], IC del 95 % = [inferior, superior]).

- Además, en la prueba HSD de Tukey se encontró que no hubo una diferencia estadísticamente significativa entre [nombre del grupo] y [nombre del grupo] (p=[valor de p]), IC del 95 % = [inferior, superior]).

Es conveniente también reportar la potencia de la prueba y el tamaño del efecto.

Segunda parte. Trabajo a entregar

Realice un análisis similar al presentado en la primera parte, pero esta vez sobre la **eficiencia de cuatro motores diferentes de bases de datos** (PostgreSQL, MySQL, MongoDB y Redis) en el manejo de consultas complejas (JOINs en SQL, agregaciones en NoSQL) bajo carga controlada.

Se realizan **48** pruebas diferentes, 12 en cada motor de bases de datos.

El archivo con la información es **bd_server.csv**.

La variable de respuesta es la columna **respuesta**, que guarda el tiempo de ejecución en milisegundos de cada solicitud.

Asegúrese de que cuando cargue el archivo, la columna **motor** sea interpretada como un factor por R.

Puede ejecutar el siguiente comando para estar seguros:

```
bd_server$motor <- as.factor(bd_server$motor)
```

Luego, usando summary(bd_server) debe aparecerle lo siguiente:

```
> summary(bd_server)
    motor              respuesta
 Length:48          Min.   : 8200
 Class :character   1st Qu.: 9840
 Mode  :character   Median :10622
                    Mean   :10870
                    3rd Qu.:12048
                    Max.   :14698
```

El ejercicio debe abarcar los 14 puntos estudiados en la primera parte del Laboratorio.

En cada caso presente el código utilizado, los resultados y gráficos obtenidos. Debe ir poniendo el cófigo R con su salida respectiva o su gráfico respectivo.

También debe ir incluyendo una breve descripción de los pasos que va a realizar y un breve análisis de lo que indican los resultados y gráficos.

Finalmente, reporte los resultados del ANOVA de una vía, incluyendo el tamaño del efecto y la potencia de la prueba.

También determine el tamaño de n necesario (o suficiente) para una potencia de 0.85.

Indicaciones adicionales

- El trabajo es para entregarse en grupos de 2 personas.
- Debe crear un reporte con la evidencia de la segunda parte del laboratorio. La primera parte, el ejercicio guiado, no debe incluirlo.
- El reporte es un documento con un formato uniforme y coherente, en el que se incluye una introducción breve en la que se indica de qué se trata el reporte.
- El reporte debe poder leerlo y entenderlo una persona que no cuenta con el enunciado de la tarea y no realizó la primera parte del laboratorio.
- También se debe incluir texto explicativo del trabajo que se va realizando previo a que se presente el código R (por ejemplo indicar que se va a probar la normalidad de los residuales con el gráfico tal o el estadístico tal), luego aparece el segmento de código R específico para ese punto, los resultados obtenidos en R, los gráficos que se van aportando como parte del experimento y un breve análisis de lo que indican los resultados y gráficos.
- Las entregas se realizarán en formato PDF. La letra debe ser Times New Roman, Arial o Cambria. Los tamaños permitidos son 10-12.
- Debe incluir una portada de página completa en la que incluya el nombre del curso, del laboratorio, así como los nombres completos de los estudiantes. Esta página será utilizada por el profesor o el asistente para apuntar la nota y también para incluir comentarios en caso de tener que incluir generales del informe.
- El reporte se debe entregar en Mediación Virtual, en un archivo .pdf que pueda ser leído en programas comerciales de uso habitual. Debe verificar que el .pdf que subió a Mediación Virtual contiene los ejercicios resueltos y que el archivo puede abrirse correctamente. En caso de problemas con el archivo .pdf (no abre correctamente, está corrupto, etc.) se considerará que no entregó la tarea.
- Las entregas tardías se penalizarán con un 10% de la nota luego de vencida la fecha y hora de entrega, más un 10% adicional por cada hora de retraso.