Diseño de Experimentos

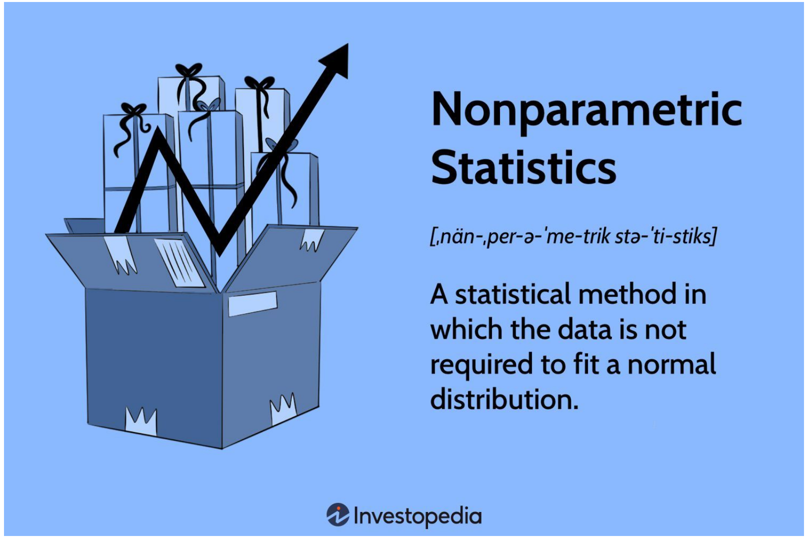
Apuntes Semana 14

Andrey Arguedas Espinoza - 2020426569

Tabla de Contenidos

Pruebas no paramétricas

Estas pruebas nos sirven para hacer análisis cuando los datos que tenemos no cumplen con normalidad, homocedasticidad. Incluso si ejecutamos transformaciones los supuestos del ANOVA no se cumplen, para esto usamos pruebas no paramétricas. Para esto vamos a tomar los datos obtenidos y los vamos a convertir en rangos, luego sacaremos muestras aleatorias y evaluaremos la probabilidad de que dos muestras pertenezcan al mismo rango. En este curso estaremos utilizando la prueba Kruskal-Wallis, sin embargo existen otras como la Mann-Whitney. También haremos análisis Post-Hoc.



Ejemplo de una prueba no paramétrica

Como este es un tipo de prueba basada en rangos utilizaremos como ejemplo la altura de los niños en un kinder. Supongamos que tomamos la altura de 4 niños, cuyos valores son (110, 125, 126 y 135 cm), los debemos ordenar de menor a mayor y asignar rangos ordinales y continuos. Con esto ya no analizaremos la variable original, sino los rangos. Esto nos va a permitir analizar cuando no se cumple la normalidad, cuando hay sesgos grandes y cuando la varianza es muy dispar. El punto negativo es las distancias, note que entre 110 cm a 125 cm hay mucha distancia en cm pero solamente 1 rango de distancia. Seguido de esto vamos a tomar dos muestras y la prueba Kruskal-Wallis nos dará la probabilidad de que ambas estén en el mismo rango. Si la probabilidad de Kruskal-Wallis nos da muy alta quiere decir que pertenecen al mismo grupo, si es baja pertenecen a distintos grupos.

Altura	110 cm	125 cm	126cm	135 cm
Rango	1	2	3	4

Creando rangos en R

```r
Altura = c(100, 132, 137, 139, 140, 142, 142, 145)
names(Altura) = letters[1:8]
Altura
rank(Altura) #Convierte en rangos
```

```
> Altura
  a   b   c   d   e   f   g   h 
100 132 137 139 140 142 142 145 
> rank(Altura)
  a   b   c   d   e   f   g   h 
1.0 2.0 3.0 4.0 5.0 6.5 6.5 8.0 
```

- Notar que f y g tienen el mismo valor por lo que el rango se convierte en el valor de rango intermedio entre ellos (originalmente hubiera sido 6 y 7)

Ventajas y desventajas de usar pruebas no paramétricas

Ventajas:

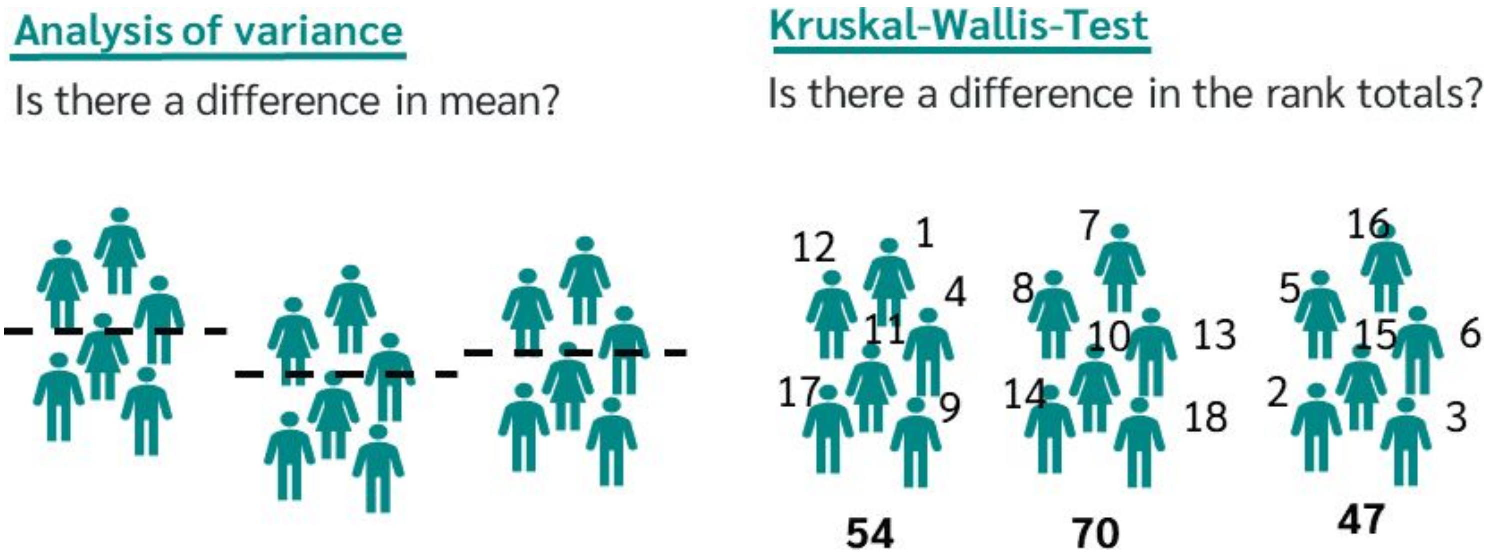
- Son pruebas muy comunes con mucha documentación.
- Son apropiadas para variables dependientes ordinales, ejemplo las calificaciones en una encuesta.
- Son apropiadas para datos que no cumplen los supuestos de pruebas paramétricas.

Desventajas:

- Es difícil recordar los nombres de las pruebas.
- Tiene limitaciones en la clase de experimentos donde se pueden usar, solo en monofactoriales.
- Hay que tener cuidado con la bibliografía consultada con respecto a hipótesis y supuestos. Hay que verificar bien los supuestos de pruebas no paramétricas.

Consideraciones con Kruskal-Wallis

Si se validan los supuestos, la hipótesis se deberá evaluar sobre las medianas. Si no se validan los supuestos, la hipótesis se evaluará sobre la probabilidad de que las dos muestras pertenezcan a un mismo rango.



Caso de uso de pruebas no paramétricas.

Supongamos el siguiente experimento donde queremos validar la arquitectura, costo, algoritmo y otro factor X. En este tenemos 144 combinaciones, si quisiéramos hacer 10 repeticiones serían 1440 experimentos. Supongamos que después de tener los resultados no hay forma de analizarlo con ANOVA debido a que no cumple los supuestos, lo que tendremos que hacer es analizar solamente un factor, y esto se convierte en un monofactorial con 4 grupos (suponiendo que analizaremos Arquitectura).

Arquitectura	Costo	Algoritmo	X
CPU	100	A	W
GPU	1000	B	Y
APU	10000	C	Z
TPU	150000		

Interpretación de pruebas no paramétricas

- Determinan si hay diferencia sistemática entre los grupos, quiere decir si la distribución de un grupo es distinta a la de otro. Si son del mismo grupo van a tener la misma forma y estarán en el mismo eje X, si son distintas pueden tener misma forma pero en distintos puntos sobre el eje X.
- Determina diferencia en la forma o dispersión de los datos.
- Vamos a tener un grupo de interés, normalmente el factor principal.

Kruskal-Wallis

Es una prueba basada en rangos, similar a Mann Whitney, se puede aplicar para un factor en varios grupos. No tiene supuestos sobre la distribución de los datos y no se comprueban hipótesis sobre las medianas de los grupos.

Vamos a poder verificar qué tan posible es que la observación de un grupo sea diferente a la observación de otro grupo (mediante los rangos).

Al igual que ANOVA aplicaremos las pruebas post-hoc para saber si hay diferencia entre los grupos.

Hipotesis

Hipótesis nula (p-value alto): Las muestras de los grupos provienen de poblaciones con distribuciones idénticas. Las poblaciones muestreadas son ***estocásticamente*** idénticas.

Hipótesis alternativa (p-value bajo): Las muestras de los grupos provienen de poblaciones con distribuciones distintas. La muestra de al menos una de las poblaciones es ***estocásticamente*** dominante.

Estocástico: Son resultados obtenidos al azar en un proceso y analizados por probabilidad, ejemplo tirar una medición 1000 veces al azar (un dado, una moneda, etc) y lo analizamos por probabilidad.

Ejemplo en R

El factor principal será Algoritmo.

Incluimos las librerías necesarias.

```
if(!require(psych)){install.packages("psych")}
if(!require(FSA)){install.packages("FSA")}
if(!require(lattice)){install.packages("lattice")}
if(!require(multcompView)){install.packages("multcompView")}
if(!require(rcompanion)){install.packages("rcompanion")}
```

Carga de datos

```
library(psych)


# Directorio donde se encuentra el archivo
setwd(this.path::here())


Data = read.table("datos_no_parametricas.txt", header=TRUE)
Data$Algoritmo = factor(Data$Algoritmo,levels=unique(Data$Algoritmo)) #Ponderacion del rango
Data$Rendimiento.f = factor(Data$Rendimiento, ordered=TRUE)
```

*** Con Rendimiento.f llevaremos el puntaje del rendimiento (ponderación de rangos)

Resumen de nuestros datos

```
        Algoritmo    Rendimiento   Rendimiento.f
 Algoritmo A:20   Min.   :1.0     1: 2
 Algoritmo B:20   1st Qu.:3.0     2:12
 Algoritmo C:20   Median :4.0     3:10
                  Mean   :3.5     4:26
                  3rd Qu.:4.0     5:10
                  Max.   :5.0
```

En Rendimiento.f podemos ver que hay 2 datos en el rango 1, 12 en el rango 2, 10 en el rango 3, 26 en el rango 4 y 10 en el rango 5.

Resumimos en una tabla

```
XT = xtabs(~Algoritmo + Rendimiento.f, data=Data)
XT
```

```
               Rendimiento.f
Algoritmo        1  2  3  4  5
  Algoritmo A    0  0  2 12  6
  Algoritmo B    2 12  4  2  0
  Algoritmo C    0  0  4 12  4
```

La tabla nos sirve para intuir si hay grupos distintos. Cada columna es el rango, por lo que se lee que hay 2 datos en el rango 1. Para cada entrada X: Y se establece que “Para el rango X existen un total de Y registros o muestras asociadas”. También podemos observar como el Algoritmo A y el Algoritmo C comparten similitudes.

Resumimos tabla de ponderación

```
prop.table(XT, margin = 1)
```

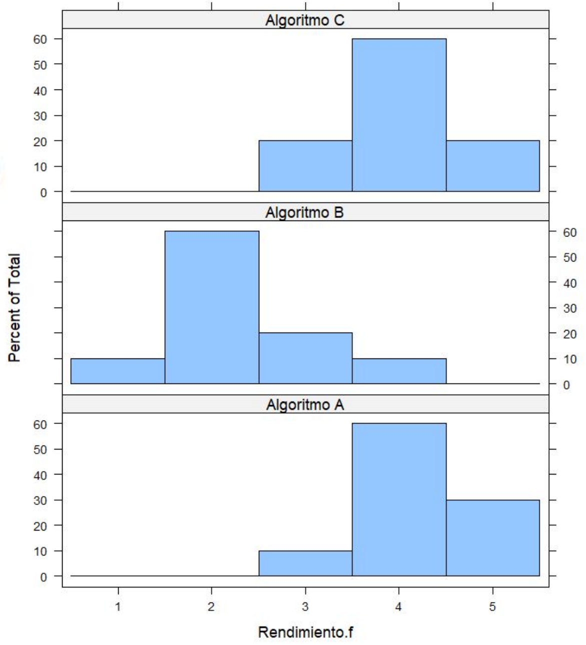
```
             Rendimiento.f
Algoritmo       1   2   3   4   5
  Algoritmo A 0.0 0.0 0.1 0.6 0.3
  Algoritmo B 0.1 0.6 0.2 0.1 0.0
  Algoritmo C 0.0 0.0 0.2 0.6 0.2
```

La cantidad de datos totales se convierten en una ponderación de los rangos mismos.

Gráfico de barras por grupo

```
library(lattice)
histogram(~Rendimiento.f | Algoritmo, data = Data, layout = c(1, 3))
```

Con el gráfico de la derecha podemos intuir que la mayor cantidad de datos del Algoritmo B no tocan a la mayoría de los de A y C. Con esto podemos entender que B es distinto a A y C.



Summarize de rendimiento en función de Algoritmo

```
library(FSA)
Summarize(Rendimiento ~ Algoritmo, data=Data, digits=3)
```

```
    Algoritmo  n mean    sd min Q1 median Q3 max
1 Algoritmo A 20  4.2 0.616   3  4      4  5   5
2 Algoritmo B 20  2.3 0.801   1  2      2  3   4
3 Algoritmo C 20  4.0 0.649   3  4      4  4   5
```

Podemos ver que la mediana de B es distinta a la mediana de A y C.

Prueba Kruskal-Wallis

```
kruskal.test(Rendimiento ~ Algoritmo, data=Data)

        Kruskal-Wallis rank sum test

data:  Rendimiento by Algoritmo
Kruskal-Wallis chi-squared = 34.265, df = 2, p-value = 3.625e-08
```

Realizamos la prueba de Kruskal Wallis y nos da un p-value pequeño por lo que sabemos que hay diferencia entre los grupos.

Interpretación de la prueba Kruskal-Wallis

La prueba de Kruskal-Wallis nos va a responder si los puntajes son significativamente distintos entre los grupos. En este caso sí lo son dado al bajo p-value que obtuvimos.

```
        Kruskal-Wallis rank sum test

data:  Rendimiento by Algoritmo
Kruskal-Wallis chi-squared = 34.265, df = 2, p-value = 3.625e-08
```

Análisis Post-Hoc con prueba de Dunn.

Usamos Post-Hoc para determinar cuál Algoritmo es distinto, para esto usamos la prueba de Dunn. Tenemos que usar el ajuste BH para minimizar el error tipo 1. Al igual que con el Anova es bueno visualizar mediante letras.

```
Data$Algoritmo = factor(Data$Algoritmo, levels = c("Algoritmo A", "Algoritmo B", "Algoritmo C"))
levels(Data$Algoritmo)
DT = dunnTest(Rendimiento ~ Algoritmo, data = Data, method = "bh")
DT
```

```
                 Comparison          Z      P.unadj        P.adj
1 Algoritmo A - Algoritmo B  5.3776947 7.544560e-08 2.263368e-07
2 Algoritmo A - Algoritmo C  0.6865142 4.923889e-01 4.923889e-01
3 Algoritmo B - Algoritmo C -4.6911805 2.716332e-06 4.074498e-06
```

Con el Dunn test obtenemos los resultados de comparaciones de los grupos, debemos observar el p-value ajustado y con este podemos concluir que A y B son distintos, A y C no lo son y B y C sí.

Despliegue compacto con letras

```
PT = DT$res
cldList(P.adj ~ Comparison, data = PT, threshold = 0.05)
```

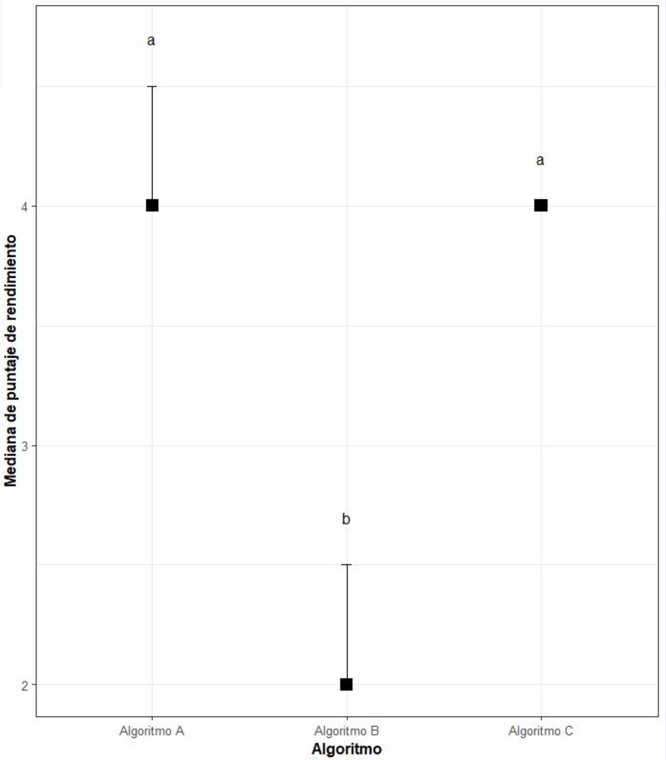
```
        Group Letter MonoLetter
1 AlgoritmoA      a          a
2 AlgoritmoB      b           b
3 AlgoritmoC      a          a
```

Gráfico final

Vamos a graficar medianas:

```
Sum = groupwiseMedian(Rendimiento ~ Algoritmo,
                      data = Data,
                      conf = 0.95,
                      R = 5000,
                      percentile = TRUE,
                      bca = FALSE,
                      digits = 3
)
Sum


X = 1:3
Y = Sum$Percentile.upper + 0.2
Label = c("a", "b", "a")
ggplot( Sum, aes(x = Algoritmo, y = Median)) +
  geom_errorbar(aes(ymin = Percentile.lower,ymax = Percentile.upper),width = 0.05,size = 0.5 ) +
  geom_point(shape = 15, size = 4) + theme_bw() +
  theme(axis.title = element_text(face="bold")) +
  ylab("Mediana de puntaje de rendimiento") +
  annotate("text", x = X, y = Y, label = Label)
```



Experimentos y control de calidad

- El control de calidad puede ser la diferencia entre buena y mala ciencia.
- Busca monitorear y mantener estándares de laboratorio.
- Detectar y corregir errores en el laboratorio (buscamos resultados consistentes).
- Ayuda a asegurar que los resultados obtenidos del experimento sean consistentes.

¿Qué es calidad?

• ISO 9000: Es el grado en el que un conjunto de características de un objeto cumplen con los requisitos.

• RAE: Propiedades o conjunto de propiedades inherentes a una cosa que permite apreciarla y compararla con los restantes de su especie

Laboratory Quality Management

Aumenta la confianza del investigador en los resultados obtenidos, consta de 6 partes:

• Manual de calidad : Describe los sistemas que se deben cumplir

• Personal y capacitación constante.

• Metodología: Debe de ser consistente, preciso y correcto.

• Manuales de referencia internos: Donde se puedan encontrar muestras de control para métodos estándar. Con la intención de verificar que el método funciona y se está aplicando correctamente antes de hacer una prueba.

• Uso de bitácoras: Escribir preparación, procedimiento y análisis del experimento.

• Costos y beneficios: El costo de su implementación debe ser bajo. Tambien ganamos en aumento de confianza de los resultados y la reducción de problemas.

Importancia del control de calidad

- Redundancia.
- Buenas prácticas al almacenar y recolectar datos.
- Estado del equipo (calibración).
- Reproducción de resultados consistentes.
- Creación de bitácoras.
- No tiene que ser cara ni complicada.
- Identificar y corregir fallas.
- Implementar un sistema de rastreo.
- Verificar utilización de controles y toma de datos.