XS-2110. Métodos Estadísticos.

**Análisis de Variancia**

Cuando se quiere comparar las medias de más de 2 grupos, en el análisis paramétrico no se puede utilizar una prueba t de Student, porque necesitamos comparar las tres medias simultáneamente. Aquí es cuando se utiliza un Análisis de Variancia ANDEVA también conocido por sus siglas en inglés, ANOVA. Decimos que si tenemos k grupos, realizamos un ANOVA de una vía con k tratamientos.

Cuando k=2, el ANOVA es equivalente a una prueba t de student para muestras independientes suponiendo variancias iguales.

Supongamos que tenemos k muestras aleatorias de tamaño n1, …, nk, cada una seleccionada de sendas distribuciones con medias μ1, …, μk,

y con variancias σ21 = …=σ2k = σ2

Las hipótesis nula y alternativa se pueden plantear como:

H0: μ1 = μ2 = … = μk

H1: Al menos una μi es diferente a las demás.

La variable aleatoria es yij,

donde:

i= denota el grupo (o la población de la que se extrae)

j= la identificación de la unidad estadística dentro de cada muestra.

Entonces, llamaremos Suma de Cuadrados Total SCTotal a:



Nótese que: SCTotal/(n-1) es igual a la variancia de la variable yij

Pues la fórmula se puede expresar:



Esta fórmula la podemos descomponer en:



|  |  |  |
| --- | --- | --- |
| SCTotal = | SCT | + SCE |

Esto se sabe porque:











|  |  |  |
| --- | --- | --- |
| SCTotal = | SCE | + SCT |

En una prueba de hipótesis de un ANOVA, lo que buscamos es decir que “la SCT es más grande que la SCE”. O lo que es lo mismo, la Suma de Cuadrados entre grupos es mayor a la Suma de cuadrados Dentro de grupos.

La fórmula del CME, que es nuestra estimación de la variancia general es:



El cuadrado medio de tratamientos CMT sería:



Además, para CME, nótese que es igual a:

Para k muestras:



Para dos muestras:



 = CME

El CMT se distribuye como una χ2 con k-1 grados de libertad, y

El CME se distribuye como una χ2 con n-k grados de libertad, entonces

el estadístico

, se distribuye como una F con k-1 y n-k grados de libertad.

Si este estadístico F es grande, estamos diciendo que el CMT es considerablemente mayor al CME, y por consiguiente una proporción grande de la SC Total se explica por las diferencias de medias entre grupos, o sea por los tratamientos.

En un ANOVA, generalmente construiremos la típica tabla:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fuente de variación |  | grados de libertad | Cuadrado Medio |  |
| Entre tratamientos |  | k-1 | CMT |  |
| Error o residual (“dentro de tratamientos”) | SCE = SCTotal – SCT | n-k | CME |  |
|  |  |  |  |  |
| Total |  | n-1 |  |  |

Ejemplo.

Se realizó un experimento para investigar el efecto del color del papel (azul, verde, anaranjado) en la tasa de respuesta de volantes que se colocan en los parabrisas de parqueos de supermercados. Se escogieron 15 supermercados, en cada uno de cuyos parqueos se pusieron 100 volantes sobre un nuevo producto.

Se evaluó cuántas fueron las ventas (en miles de colones) de ese nuevo producto en el día en que se pusieron los volantes. En 5 supermercados se usaron volantes azules, en otros cinco volantes verdes, y en otros cinco volantes anaranjados. Se quiere ver si hay diferencias en las Ventas según color del volante.

Muestra de 15 supermercados. Ventas en miles de colones diarias de un nuevo producto X.

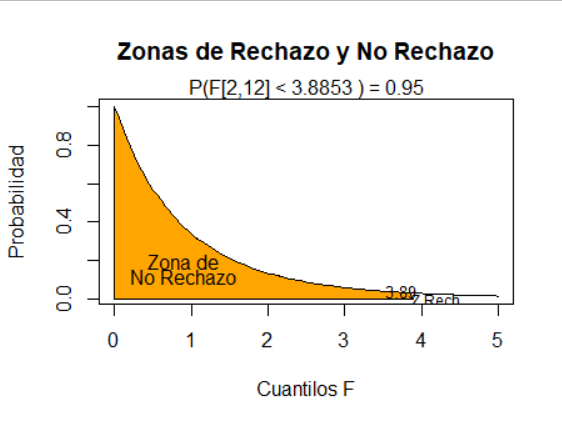
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Ventas (en miles de colones)** | | |  |
|  | **Azul** | **Verde** | **Anaranjado** |  |
|  | **28** | **34** | **41** |  |
|  | **27** | **29** | **35** |  |
|  | **33** | **24** | **37** |  |
|  | **28** | **30** | **39** |  |
|  | **34** | **28** | **33** | **Total** |
| **Promedios** | **30.0** | **29.0** | **37.0** | **32** |
| **Desv Est** |  |  |  | **4.811** |
| **Variancia** |  |  |  | **23.143** |

H0: μ1 = μ2 = μ3

H1: Al menos una μi ≠ μj para todo i≠j..

Primero se define el Ftabular=

Ft=F[1-alfa, k-1,n-k gl]= F[1-0.05, 3-1, 15-3]=F[1-0.05,2,12]=3.8853



Después se calcula las Sumas de Cuadrados para completar la tabla del ANDEVA. No es necesario usar la fórmula de la Suma de Cuadrados del Error directamente, sino que se calcula por diferencia: SCE= SCTotal-SCT

SCE=324-190 = 134

Se completa la Tabla del ANDEVA:

Los Cuadrados Medios se estiman como el cociente entre las Sumas de Cuadrados divididos por sus grados de libertad.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tabla de Análisis de Variancia | |  |  |  |  |
|  |  |  |  |  |  |
| Fuente de Variación | SC | grados de lib. | CM | Fc | Ftabular |
| Tratamiento | 190 | k-1=3-1= 2 | 95 | 8.51 | 3.89 |
| Error | 134 | n-k=15-3= 12 | 11.17 |  |  |
| **Total** | **324** | **14** |  |  |  |
|  |  |  |  |  |  |

Evidentemente, el Fc cae en la Zona de Rechazo y entonces hay suficiente evidencia estadística para rechazar la hipótesis nula de igualdad de medias, con una significancia del 5%.

**Comparaciones Múltiples**

Si la prueba de hipótesis del ANDEVA resultó significativa, es común comparar cada media contra todas las demás pues no se sabe de antemano cuál par es significativamente diferente entre sí.

El problema de las comparaciones múltiples es que se infla artificialmente el error Tipo I debido a que estamos realizando varias pruebas de hipótesis a la vez. Entonces, la probabilidad de que se cometa el error Tipo I al menos una vez es más alta.

El error Tipo I efectivo en comparaciones múltiples es: αefectivo = 1-(1-α)#.comparaciones

En todos estos métodos se calcula una diferencia mínima que debe tener la resta entre una media y otra para que sea significativa. Por eso, cada método tiene la forma de:

diferencia = (cuantil) \* (error estándar de la diferencia)

Nótese que en un intervalo de confianza común y corriente para una diferencia de medias con variancias iguales, se aplica esta fórmula:



Existen diferentes métodos para corregir el error experimental total: los métodos de Scheffé, Duncan, Newman-Keuls, LSD (ó DMS), Tukey, Bonferroni, Sidak, Dunnett, etc. En este curso, por falta de tiempo, sólo veremos el método de Tukey por ser el más conocido.

Método de Tukey:

La diferencia de medias observada se compara contra:



qα(k,f) se busca en la tabla de valores críticos de Tukey.

Es una de las pruebas más conservadoras (menos sensibles a rechazar la hipótesis nula). Con diseños balanceados, el nivel de confianza simultáneo es de 100\*(1-α)%, y en diseños desbalanceados es al menos 100\*(1-α)%.

El resultado de R de las comparaciones múltiples sería:

> TukeyHSD(anova1)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = ventas ~ color)

$color

diff lwr upr p adj

naranja-azul 7 1.361606 12.638394 0.0158322

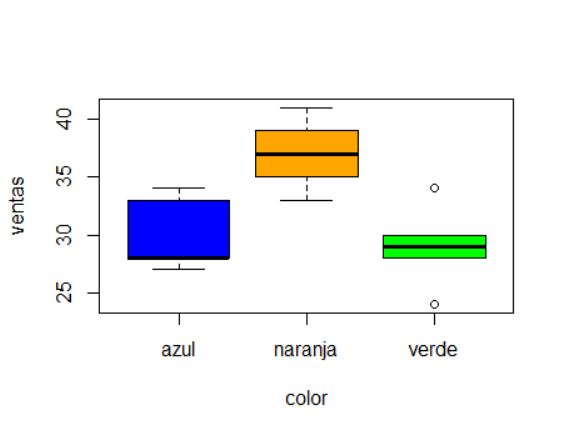
verde-azul -1 -6.638394 4.638394 0.8850473

verde-naranja -8 -13.638394 -2.361606 0.0067822

H0: Mu[naranja]=Mu[azul]

H0: Mu[verde]=Mu[azul]

H0: Mu[verde]=Mu[naranja]



ANDEVA no paramétrico ó prueba de Kruskal-Wallis:

Si no se puede mantener el supuesto de normalidad condicional de *y* (o normalidad de los residuos), o si no se cumple el supuesto de homoscedasticidad debido a valores extremos, se recomienda utilizar la prueba de Kruskal y Wallis.

Este ANDEVA también se conoce como ANDEVA de rangos. Como con el r de Spearman y el τ de Kendall, se calculan los rangos de todos los datos, indistintamente de los grupos o tratamientos. Los datos “rankeados” se denominan como Ri. Después se clasifica cada rango dentro de su grupo i, y el estadístico H se calcula como:

Si existen muchos empates, se calcula el factor de corrección FC:



Y por último, se calcula el nuevo H que va a ser igual a:



Cuando n es grande, H se distribuye como una χ2 con k-1 grados de libertad.

Se llega a un resultado similar si se utilizan las fórmulas del ANDEVA paramétrico con los rangos.

No se puede utilizar la prueba de Kruskal-Wallis si la heteroscedasticidad se mantiene, pese a la transformación en rangos. La prueba de Dunn (Dunn’s test) se recomienda para las pruebas de comparaciones múltiples.

El mismo ejemplo anterior, con Kruskal Wallis.

Muestra de 15 supermercados. Ventas en miles de colones diarias de un nuevo producto X.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Ventas (en miles de colones)** | | |  |
|  | **Azul** | **Verde** | **Anaranjado** |  |
|  | **28** | **34** | **41** |  |
|  | **27** | **29** | **35** |  |
|  | **33** | **24** | **37** |  |
|  | **28** | **30** | **39** |  |
|  | **34** | **28** | **33** | **Total** |
| **Promedios** | **30.0** | **29.0** | **37.0** | **32** |
| **Desv Est** |  |  |  | **4.811** |
| **Variancia** |  |  |  | **23.143** |

Se calculan los rangos. Para ello, se ordenan de menor a mayor los datos señalando con colores cada uno de los grupos. En negrita se resaltan los empates:

|  |  |  |
| --- | --- | --- |
| datos | rango | rango\* |
| 24 | 1 | 1 |
| 27 | 2 | 2 |
| **28** | 3 | 4 |
| **28** | 4 | 4 |
| **28** | 5 | 4 |
| 29 | 6 | 6 |
| 30 | 7 | 7 |
| **33** | 8 | 8.5 |
| **33** | 9 | 8.5 |
| **34** | 10 | 10.5 |
| **34** | 11 | 10.5 |
| 35 | 12 | 12 |
| 37 | 13 | 13 |
| 39 | 14 | 14 |
| 41 | 15 | 15 |

Ya con el rango\*, se vuelve a replantear la tabla, pero con los rangos:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Ventas (en miles de colones)** | | |  |  | **Rangos** |  | |
|  | **Azul** | **Verde** | **Ana-ran-jado** |  | **Azul** | **Verde** | **Ana-ran-jado** |
|  | **28** | **34** | **41** |  | **4** | **10.5** | **15** |
|  | **27** | **29** | **35** |  | **2** | **6** | **12** |
|  | **33** | **24** | **37** |  | **8.5** | **1** | **13** |
|  | **28** | **30** | **39** |  | **4** | **7** | **14** |
|  | **34** | **28** | **33** |  | **10.5** | **4** | **8.5** |
| **Totales (Ri)** |  |  |  |  | **29** | **28.5** | **62.5** |
| **Totales cuad (Ri2)** |  |  |  |  | **841** | **812.25** | **3906.25** |
| **Ri2/ni** |  |  |  |  | **168.2** | **162.45** | **781.25** |

Ahora hay que analizar si más del 25% de los casos son empates. Noten que 7 de 15 valores son empates (el 28 se repite 3 veces, el 33 se repite 2 veces, y el 34 se repite 2 veces). Entonces hay que aplicar el factor de corrección.

Y este H se compara con un χ2 con k-1.

χ2 [1-alfa,k-1]= χ2 [1-0.05,2]=5.99

Evidentemente, el H cae en la Zona de Rechazo y entonces hay suficiente evidencia estadística para rechazar la hipótesis nula de igualdad de medias, con una significancia del 5%.

Para analizar cuáles pares son diferentes, se analiza el resultado de R de la función dunn.test:

> dunn.test(ventas,color, method="bonferroni")

Kruskal-Wallis rank sum test

data: ventas and color

Kruskal-Wallis chi-squared = 7.6773, df = 2, p-value = 0.02

Comparison of ventas by color

(Bonferroni)

Col Mean-|

Row Mean | azul naranja

---------+----------------------

naranja | -2.381600

| 0.0259

|

verde | 0.035546 2.417146

| 1.0000 0.0235\*

alpha = 0.05

Reject Ho if p <= alpha/2