Proyecto de Metodos no parametricos

Key Hirosi Richster Quinto

2024-02-24

Índice

Proyecto de Metodos no parametricos

Key Hirosi Richster Quinto

2024-02-24

Punto 1

Una compañía de taxis está tratando de decidir si utiliza neumáticos radiales en lugar de los regulares con cinturón para mejorar la economía del combustible. Los neumáticos radiales y los neumáticos normales con cinturón se manejarán a lo largo de una pista prescrita para pruebas. Sin cambiar conductores, los mismos vehículos se equiparon con neumáticos normales de cinturón y se condujeron por la misma pista de pruebas. El consumo de gasolina, en kilómetros por litro, se registró de la siguiente manera.

```
automovil<- c(seq(1,16,1))
#Neumaticos_radiales
S1<-c(4.2,4.7,6.6,7,6.7,4.5,5.7,6,7.4,4.9,6.1,5.2,5.7,6.9,6.8,4.9)
#Neumatico_con_cinturon
S2<-c(4.2,4.9,6.2,6.9,6.8,4.4,5.7,5.8,6.9,4.9,6.0,4.9,5.3,6.5,7.1,4.8)
```

¿Se puede concluir al nivel de significancia de 0.01, que los vehículos equipados con neumáticos radiales dan en promedio una economia de combustible que aquellos equipados con neumaticos regulares con cinturón?

deseamos probar que:

H_0: Los vehículos equipados con neumáticos radiales en promedio dan una economia de combustible que aquellos equipados con neumaticos regulares con cinturón

H_0: Los vehículos equipados con neumáticos radiales en promedio NO dan una economia de combustible que aquellos equipados con neumaticos regulares con cinturón

```
n_1<-length(S1)
n_2<-length(S2)
```

Calculamos los rangos para ambas varables:

```
Ri<-rank(c(S1,S2));Ri
```

```
##  [1]  1.5  5.0 23.0 30.0 24.0  4.0 15.0 18.5 32.0  9.0 20.0 12.0 15.0 28.0 25.5
## [16]  9.0  1.5  9.0 21.0 28.0 25.5  3.0 15.0 17.0 28.0  9.0 18.5  9.0 13.0 22.0
## [31] 31.0  6.0
```

```
sort(c(S1,S2))
```

```
##  [1] 4.2 4.2 4.4 4.5 4.7 4.8 4.9 4.9 4.9 4.9 4.9 5.2 5.3 5.7 5.7 5.7 5.8 6.0 6.0
## [20] 6.1 6.2 6.5 6.6 6.7 6.8 6.8 6.9 6.9 6.9 7.0 7.1 7.4
```

Ahora calculamos los U y los W

```
U_1<- sum(Ri[1:length(S1)]);U_1
```

```
## [1] 271.5
```

```
U_2<- sum(Ri[(length(S1) + 1):length(Ri)]);U_2
```

```
## [1] 256.5
```

```
W_1<-U_1-((n_1*(n_1+1))/2);W_1
```

```
## [1] 135.5
```

```
W_2<-U_2-((n_2*(n_2+1))/2);W_2
```

```
## [1] 120.5
```

```
alpha <-0.01
tab<- qnorm(1-alpha/2);tab
```

```
## [1] 2.575829
```

```r
E_W<- (n_1*n_2)/2;E_W
```

```
## [1] 128
```

```r
var_W<-  (n_1*n_2*(n_1+n_2+1))/12;var_W
```

```
## [1] 704
```

```r
Z1<- (W_1-E_W)/sqrt(var_W);Z1
```

```
## [1] 0.2826669
```

```r
Z2<- (W_2-E_W)/sqrt(var_W);Z2
```

```
## [1] -0.2826669
```

```r
if (Z1>tab ){
  print("Se rechaza h0")
}else{
  print("No se rechaza H0")
}
```

```
## [1] "No se rechaza H0"
```

Como $Z_1 = 0.2826669 < 2.575829 = Z_{\alpha/2}$, No se rechaza H_0, es decir, que los vehículos equipados con neumáticos radiales en promedio Sí dan una economia de combustible que aquellos equipados con neumaticos regulares con cinturón.

punto 2

Se utilizan cuatro diferentes métodos para cultivar trigo, sobre 4 parcelas de tierra, y se midió la producción por acre en cada parcela. Se desea comparar las medianas de cada una de las parcelas. Los resultados fueron:

```r
metodo1<-c(83,91,94,89,89,96,91,92,90)
metodo2<-c(91,90,81,83,84,83,88,91,89,84)
metodo3<-c(101,100,91,93,93,95,94)
metodo4<-c(78,82,81,77,79,81,80,81)
```

Hay evidencia muestral sufiente para sugerir que algunos métodos para cultivar trigo tienden a dar mayor producción que otros?

lo que se desea probar es que:

$H_0 : \mu_1 = \mu_2 = ... = \mu_t \quad vs \quad H_1 :$ Al menos una media es distinta

Entonces para iniciar calculamos los rangos

```r
ri<-rank(c(metodo1,metodo2,metodo3,metodo4));ri
```

```
##  [1] 11.0 23.0 29.5 17.0 17.0 32.0 23.0 26.0 19.5 23.0 19.5  6.5 11.0 13.5 11.0
## [16] 15.0 23.0 17.0 13.5 34.0 33.0 23.0 27.5 27.5 31.0 29.5  2.0  9.0  6.5  1.0
## [31]  3.0  6.5  4.0  6.5
```

```r
R.j1<- sum(ri[1:length(metodo1)]);R.j1
```

```
## [1] 198
```

```r
R.j2 <- sum(ri[10:19]);R.j2
```

```
## [1] 153
```

```
R.j3<- sum(ri[20:26]);R.j3
```

```
## [1] 205.5
```

```
R.j4<- sum(ri[27:34]);R.j4
```

```
## [1] 38.5
```

```
R.j_Bar1<- mean(ri[1:length(metodo1)]);R.j_Bar1
```

```
## [1] 22
```

```
R.j_Bar2 <- mean(ri[10:19]);R.j_Bar2
```

```
## [1] 15.3
```

```
R.j_Bar3<- mean(ri[20:26]);R.j_Bar3
```

```
## [1] 29.35714
```

```
R.j_Bar4<- mean(ri[27:34]);R.j_Bar4
```

```
## [1] 4.8125
```

```
n<-length(ri);n
```

```
## [1] 34
```

```
H<-(12/(n*(n+1)))*sum((R.j1^2)/length(metodo1),(R.j2^2)/length(metodo2),(R.j3^2)/length(metodo3), (R.j4
```

```
## [1] 25.23604
```

Como tenemos observaciones pareadas procedemos a hacer los siguientes cálculos.

```
dat1<-c(metodo1,metodo2,metodo3,metodo4)
freqx<-table(dat1)
freqx<-freqx[which(freqx>1)]

freqx
```

81	83	84	89	90	91	93	94
4	3	2	3	2	5	2	2

```
C<-1-(sum(freqx^3-freqx)/(n^3-n));C
```

```
## [1] 0.9935829
```

```
H1<- H/C;H1
```

```
## [1] 25.39903
```

```
k<-4

chi<-qchisq(1-0.05,k-1);chi
```

```
## [1] 7.814728
```

```
if (H1>=chi){
  print("Se rechaza h0")
}else{
```

```r
  print("No se rechaza H0")
}
```

```
## [1] "Se rechaza h0"
```

En conclución se rechaza $ H_0$, ya que $H_1 > X_4^2 = 7.814728$, entonces existe diferencias es decir que hay evidencia para sugerir que algunos métodos de producción tienden a dar mayor producción que otros.

Teniendo en cuenta que se rechaza H_0 procedemos a hacer las comparaciones de medias para ver cual de los tratamientos que difieren.

```r
alpha<-0.05
zk<-qnorm(1-alpha/(4*(4-1)));zk
```

```
## [1] 2.638257
```

```r
estadistica1<-zk*sqrt((n*(n-1)/12)*(1/9+1/10));estadistica1 #1 con 2
```

```
## [1] 11.72137
```

```r
estadistica2<-zk*sqrt((n*(n-1)/12)*(1/9+1/7));estadistica2  #1 con 3
```

```
## [1] 12.8562
```

```r
estadistica3<-zk*sqrt((n*(n-1)/12)*(1/9+1/8));estadistica3  #1 con 4
```

```
## [1] 12.39599
```

```r
estadistica4<-zk*sqrt((n*(n-1)/12)*(1/10+1/7));estadistica4  #2 con 3
```

```
## [1] 12.57183
```

```r
estadistica5<-zk*sqrt((n*(n-1)/12)*(1/10+1/8));estadistica5  #2 con 4
```

```
## [1] 12.1008
```

```r
estadistica6<-zk*sqrt((n*(n-1)/12)*(1/7+1/8));estadistica6  #3 con 4
```

```
## [1] 13.20306
```

Cuantas combinaciones puedo hacer?

```r
#posibles combinaciones
choose(4,2)
```

```
## [1] 6
```

```r
#diferencias
dif1<-abs(R.j_Bar1-R.j_Bar2);dif1
```

```
## [1] 6.7
```

```r
dif2<-abs(R.j_Bar1-R.j_Bar3);dif2
```

```
## [1] 7.357143
```

```r
dif3<-abs(R.j_Bar1-R.j_Bar4);dif3
```

```
## [1] 17.1875
```

```r
dif4<-abs(R.j_Bar2-R.j_Bar3);dif4
```

```
## [1] 14.05714
```

```r
dif5<-abs(R.j_Bar2-R.j_Bar4);dif5
```

```
## [1] 10.4875
```

```r
dif6<-abs(R.j_Bar3-R.j_Bar4);dif6
```

```
## [1] 24.54464
```

En este caso como tenemos diferentes observaciones vamos a tener una estadistica para cada diferencia de media

Estadistica 1 con difernecia de media 1 ($6.7 < 11.72137$ ns

Estadistica 2 con difernecia de media 2 ($7.357143 < 12.8562$) ns

Estadistica 3 con difernecia de media 3 ($17.1875 > 12.39599$) $*$

Estadistica 4 con difernecia de media 4 ($14.05714 > 12.57183$) $*$

Estadistica 5con difernecia de media 5 ($10.4875 < 12.1008$) ns

Estadistica 6 con difernecia de media 6 ($24.54464 > 13.20306$) $*$

Teniendo en cuenta que las estadísticas obtenidas tenemos que los tratamientos que difieren son el 1 con el 4, el 2 con el 3 y la 3 con la 4 lo que me indica que estos son diferentes viendo que las diferencias de medias son mayores que la estadística obtenida para esa relación.

Punto 3

Se diseña un experimento de degustación de modo que cuatro marcas de café colombiano sean clasificadas por 9 expertos. Para evitar cualquier efecto acumulado, la sucesión de pruebas para las 4 infusiones se determina aleatoriamente para cada uno de los 9 probadores expertos hasta que se dé una clasificación en una escala de 7 puntos (1 = extremo desagradable, 7 = extremo agradable) para cada una de las siguientes 4 categorías: sabor, aroma, cuerpo y acidez. La suma de los puntajes de las cuatro características se convierte en rangos.

H_0 : La mediana de los resultados sumados son iguales. H_1 : Por lo menos dos marcas tengan resultados diferentes.

```r
marcas <- data.frame(
  experto= c(seq(1,9,1)),
  A = c(24,27,19,24,22,26,27,25,22),
  B = c(26,27,22,27,25,27,26,27,23),
  C = c(25,26,20,25,22,24,22,24,20),
  D = c(22,24,16,23,21,24,23,21,19)
);marcas
```

experto	A	B	C	D
1	24	26	25	22
2	27	27	26	24
3	19	22	20	16
4	24	27	25	23
5	22	25	22	21
6	26	27	24	24
7	27	26	22	23
8	25	27	24	21
9	22	23	20	19

Calculamos los rangos para las columnas B, C y D organizando por filas

```r
ranks1 <- apply(marcas[, -1], 1, rank)


rangos1<-t(ranks1);rangos1
```

A	B	C	D
2.0	4.0	3.0	1.0
3.5	3.5	2.0	1.0
2.0	4.0	3.0	1.0
2.0	4.0	3.0	1.0
2.5	4.0	2.5	1.0
3.0	4.0	1.5	1.5
4.0	3.0	1.0	2.0
3.0	4.0	2.0	1.0
3.0	4.0	2.0	1.0

calculamos las sumas de los rangos por columnas

```r
suma1<- colSums(rangos1);suma1
```

```
##    A    B    C    D
## 25.0 34.5 20.0 10.5
```

Vemos el número de filas y el nmero de columnas

```r
n1<- nrow(marcas);n1
```

```
## [1] 9
```

```r
k1<- ncol(marcas[-1]);k1
```

```
## [1] 4
```

Ahora contamos las observaciones repetidas por filas

```r
contar_repetidas <- function(fila) {
  freq <- table(fila)  # Calcular la frecuencia de cada observación en la fila
  freq <- freq[freq > 1]  # Filtrar solo las observaciones repetidas
  return(freq)
}

# Aplicar la función a cada fila del DataFrame
obs_rep_por_fila <- apply(rangos1, 1, contar_repetidas)

# Unificar las frecuencias de observaciones repetidas en un vector
Tk <- unlist(obs_rep_por_fila);Tk
```

```
## 3.5 2.5 1.5
##   2   2   2
```

Obtenemos el valor F

```r
fk1<-(12*sum(suma1^2)-(3*n1^2*k1*(k1+1)^2))/(n1*k1*(k1+1)+((n1*k1-sum(Tk^2))/(k1-1)));fk1
```

```
## [1] 19.18085
```

```r
#sale el mismo fk
friedman.test(rangos1)
```

```
##
##  Friedman rank sum test
##
## data:  rangos1
## Friedman chi-squared = 20.724, df = 3, p-value = 0.0001201
```

Como el tamaño de la muestra es 9 y $F = 19.18085$, se encuentra que el p_valor es $p = 0.0001201$, por lo tanto $0.0001201 < 0.05$, entonces rechazamos H_0. Por lo que se concluye que por lo menos dos marcas de café tengan resultados diferentes.