Solución Parcial

Eddy Herrera Daza

26 de agosto de 2018

Punto 1

Se sospecha que las aguas de un lago están contaminadas por los compuestos fosforados procedentes de una industria de fertilizantes. Para tratar de verificar esta sospecha, se midieron los niveles de fósforo en distintos puntos del lago, obteniéndose los siguientes valores:

```
Lago_A=c(7.1, 8.5, 9.2, 7.3, 7.9, 8.8, 7.6, 7.8, 8.1, 6.3, 6.6, 7.3, 7.9, 6.2, 7.1)
Lago_B=c(7.1, 6.5, 5.9, 5.8, 5.9, 6.4, 6.8, 5.1, 6.2, 6.1, 5.2, 6.6, 6.3, 7.4, 6.6)
Lago_c=c(5.6, 7.1, 6.3, 5.7, 6.5, 6.5, 5.9, 6.8, 5.9, 6.1, 6.4, 6.6, 6.3, 5.6, 4.9)
```

Los valores obtenidos en el lago bajo sospecha parecen ser algo superiores a los obtenidos en los otros dos. ¿Es suficientemente importante esta diferencia como para poder concluir que el nivel de fosforo en el lago A es significativamente diferente que el que tienen los demás, y por tanto está contaminado?
b. Se cumple el principio de homocedasticidad? Justifique su respuesta

$$H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2 \tag{1}$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2; \sigma_A^2 \neq \sigma_C^2 \tag{2}$$

Para verificar si se rechaza o no H_0 se debe verificar varios supuestos. Por lo general, si no se puede alcanzar cierta seguridad de que las poblaciones que se comparan son de tipo normal, es recomendable recurrir a test que comparen la mediana de la varianza.

Test de Levene se puede aplicar con la función leveneTest() del paquete car. Se caracteriza, además de por poder comparar 2 o más poblaciones, por permitir elegir entre diferentes estadísticos de centralidad :mediana (por defecto), media, media truncada. Esto es importante a la hora de contrastar la homocedasticidad dependiendo de si los grupos se distribuyen de forma normal o no.

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.4.4
```

```
## Loading required package: carData
```

```
Lago_B=factor(Lago_B)
Lago_c=factor(Lago_c)
leveneTest(y = Lago_A, group = Lago_c, center = "median")
```

```
## Warning in anova.lm(lm(resp ~ group)): ANOVA F-tests on an essentially
## perfect fit are unreliable
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##       Df    F value    Pr(>F)
## group 10 1.1447e+29 < 2.2e-16 ***
##        4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparación entre el lago A y el lago C: ** hay diferencias significativas entre las varianzas de los dos grupos, $\rho - value < \alpha$. Por otro lado, cuando comparamos los resultados del lago A con el lago B se tiene

```r
leveneTest(y = Lago_A, group = Lago_B, center = "median")
```

```
## Warning in anova.lm(lm(resp ~ group)): ANOVA F-tests on an essentially
## perfect fit are unreliable

## Levene's Test for Homogeneity of Variance (center = "median")
##       Df    F value    Pr(>F)
## group 12 8.0209e+28 < 2.2e-16 ***
##        2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparación entre el lago A y el lago B: Hay diferencias significativas entre las varianzas de los dos grupos, $\rho - value < \alpha$
Conclusión: No se cumple el principio de homocedasticidad.
Teniendo en cuenta el resultado anterior, se realiza una prueba no paramétrica:Kruskall-Wallis
Solución:

```
    \textbf{$H_{0}$}: El nivel mediano de los niveles de fósforo del lago A es igual al de los otros dos

    \textbf{$H_{0}$}:  El nivel mediano de niveles de fósforo del lago A difiere de los niveles del lago
```

```r
kruskal.test(Lago_A ~ Lago_c)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Lago_A by Lago_c
## Kruskal-Wallis chi-squared = 12.473, df = 10, p-value = 0.2546
```

Como podemos ver, No se rechaza H_0 es decir, los niveles de fósforo asociados a la contaminación en el lago A y C son iguales.

```r
kruskal.test(Lago_A ~ Lago_B)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Lago_A by Lago_B
## Kruskal-Wallis chi-squared = 13.39, df = 12, p-value = 0.3413
```

Como podemos ver, No se rechaza H_0 es decir, los niveles de fósforo asociados a la contaminación en el lago A y B son iguales.

Conclusión Los niveles medianos de fósforo de los lagos control con el lago A No son significativamente

Punto 2

En "Experimental Design and Data Analysis for Biologists", de G.P. Quinn y M.J. Keough, (capitulo 9) , se usan los datos de un estudio de Quinn, G. P. (1988). "Ecology of the intertidal pulmonate limpet Siphonaria Diemenensis", Journal of experimental marine biology and ecology, vol. 117, no2, pp. 115-136.) sobre los efectos que tienen los factores: estación del año y la densidad de ejemplares adultos, sobre la variable respuesta dada por producción de huevos en la especie Siphonaria Diemenensis.

Utilizando el archivo "quinn" y asumiendo que se cumplen los requisitos realice un análisis de varianza y resuelva:

```r
quinn=read.table("C:/Users/lenovo/Downloads/quinn.txt",header = T)
quinn
```

```
##    DENSITY SEASON  EGGS
## 1        8 spring 2.875
## 2        8 spring 2.625
## 3        8 spring 1.750
## 4        8 summer 2.125
## 5        8 summer 1.500
## 6        8 summer 1.875
## 7       15 spring 2.600
## 8       15 spring 1.866
## 9       15 spring 2.066
## 10      15 summer 0.867
## 11      15 summer 0.933
## 12      15 summer 1.733
## 13      30 spring 2.230
## 14      30 spring 1.466
## 15      30 spring 1.000
## 16      30 summer 1.267
## 17      30 summer 0.467
## 18      30 summer 0.700
## 19      45 spring 1.400
## 20      45 spring 1.022
## 21      45 spring 1.177
## 22      45 summer 0.711
## 23      45 summer 0.356
## 24      45 summer 0.711
```

a. Determinar si las variaciones en la producción de huevos son significativamente diferentes, en las diferentes estaciones del año.\ Sea $y =$ la producción de huevos la variable respuesta y sean los factores: E:Estación del año y Densidad de ejemplares adultos

$$H_0 : \mu_y = \mu_E \tag{3}$$

$$H_1 : \mu y \neq \mu E \tag{4}$$

```r
anova_punto2=aov(EGGS~SEASON,data=quinn)
summary(anova_punto2)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## SEASON       1  3.250    3.25    8.55 0.00786 **
## Residuals   22  8.363    0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b. Determinar si las variaciones en la producción de huevos son significativamente diferentes en diferentes densidades poblacionales

```r
quinn$DENSITY=factor(quinn$DENSITY)
anova_punto2b=aov(EGGS~DENSITY,data=quinn)
summary(anova_punto2b)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## DENSITY      3  5.284  1.7614   5.566 0.00607 **
## Residuals   20  6.330  0.3165
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Realizando una **ANOVA** donde EGGS $\sim$ DENSITY o $Y \sim D$ se obtiene que el p values es de 0.00607 siendo significativo y hace que se rechace la hipótesis nula. Es decir, que hay diferencias significativas en la producción de huevos debido al efecto de la densidad poblacional
c. Si el ítem a ó b es afirmativo, donde hay mayor diferencia en la producción de huevos entre que niveles de un mismo factor

```
TukeyHSD(anova_punto2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = EGGS ~ SEASON, data = quinn)
##
## $SEASON
##                  diff       lwr        upr     p adj
## summer-spring -0.736 -1.258022 -0.2139783 0.0078605
```

Luego la diferencia en la producción de huevos entre verano y primavera es de 0.736

```
TukeyHSD(anova_punto2b)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = EGGS ~ DENSITY, data = quinn)
##
## $DENSITY
##             diff       lwr        upr     p adj
## 15-8  -0.4475000 -1.356581  0.4615808 0.5269322
## 30-8  -0.9366667 -1.845747 -0.0275859 0.0419815
## 45-8  -1.2288333 -2.137914 -0.3197526 0.0059035
## 30-15 -0.4891667 -1.398247  0.4199141 0.4525566
## 45-15 -0.7813333 -1.690414  0.1277474 0.1082534
## 45-30 -0.2921667 -1.201247  0.6169141 0.8051677
```

Los resultados de la prueba muestran que la producción de huevos tiene una diferencia significativa de 1.2288333 absoluta entre la densidad 45 y 8. Para el factor estación la diferencia entre verano y primavera es de 0.736
d. Determinar si las variaciones en la producción de huevos son significativamente diferentes, si se consideran las interacciones entre los factores nada más

```
anova_punto2d=aov(EGGS~DENSITY*SEASON,data=quinn)
summary(anova_punto2d)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## DENSITY         3  5.284   1.761   9.669 0.000704 ***
## SEASON          1  3.250   3.250  17.842 0.000645 ***
## DENSITY:SEASON  3  0.165   0.055   0.301 0.823955
## Residuals      16  2.915   0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teniendo en cuenta el modelo de iteración: EGGS $\sim$ SEASON $*$ DENSITY, la interación entre los factores no es significativa
e. Determinar si las variaciones en la producción de huevos son significativamente diferentes, cuando se

considera los efectos de los factores

```r
anova_punto2e=aov(EGGS~DENSITY+SEASON,data=quinn)
summary(anova_punto2e)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## DENSITY      3  5.284   1.761   10.87 0.000223 ***
## SEASON       1  3.250   3.250   20.05 0.000258 ***
## Residuals   19  3.079   0.162                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teniendo en cuenta el modelo de iteración: EGGS $\sim$ SEASON + DENSITY, la suma de los efectos de los factores es significativa