**Ejercicio propuesto 1**

**Se realiza un estudio sobre el efecto que produce la descarga de aguas residuales de una planta sobre la ecología del agua natural de un río. En el estudio se utilizaron dos lugares de muestreo. Un lugar está aguas arriba del punto en el que la planta introduce aguas residuales en la corriente; el otro está aguas abajo. Se tomaron muestras durante un periodo de cuatro semanas y se obtuvieron los datos sobre el número de diatomeas halladas. Los datos se muestran en la tabla adjunta (disponibles en el fichero** [**propuesto 1.txt):**](file:///C%3A%5CUsers%5CUsuario%5CDesktop%5Cdatos%5Cpropuesto1.txt)

|  |  |
| --- | --- |
|  | semana |
| Lugar | Semana 1 | Semana 2 | Semana 3 | Semana 4 |
| Aguas arriba | 78 9443 58 | 620 760420 913 | 204 33398 89 | 890 6555763 562 |
| Aguas abajo | 79 87145 522 | 546 65276 94 | 45 6959 62 | 254 86789 267 |

**Responder a las siguientes cuestiones:**

1. **Identificar el diseño adecuado a este experimento, escribir el modelo matemático y explicar los distintos elementos que intervienen.**
2. **Estudiar si la semana y el lugar son factores determinantes en el número de diatomeas halladas en el agua del río. ¿Hay posibilidad que una semana sea más recomendable en un lugar del río en concreto y no lo sea en el otro lugar?**
3. **Estudiar en qué semana se producen menos contaminación en el río, utilizando el método HSD de Tukey.**
4. **Estudiar en qué lugar del río se producen menos diatomeas, utilizando el método de HSD de Tukey**

.

SOLUCIÓN

1. **Identificar el diseño adecuado a este experimento, escribir el modelo matemático y explicar los distintos elementos que intervienen.**

El modelo estadístico para este diseño es:

$$y\_{ijk}=μ+τ\_{i}+β\_{j}+\left(τβ\right)\_{ij}+u\_{ijk}; i:1,2,3,4; j:1,2;k=1,2,3,4$$

donde:

$y\_{ijk}$: número de diatomeas halladas en la observación k en la semana i y lugar j;

$μ$: efecto constante, común a todos los niveles de los factores, denominado media global;

$τ\_{i}$: efecto producido por la semana i, $\left(\sum\_{i=1}^{4}τ\_{i}=0\right)$;

$β\_{j}$: efecto producido por el lugar j, $\left(\sum\_{j=1}^{2}β\_{j}=0\right)$;

$\left(τβ\right)\_{ij}$: efecto medio producido por la interacción entre la semana i y el lugar j, $\left(\sum\_{i}^{}\left(τβ\right)\_{ij}=\sum\_{j}^{}\left(τβ\right)\_{ij}=0\right)$;

$u\_{ijk}$: errores experimentales $u\_{ij}:v.a. independientes \rightarrow N(0;σ)$;

Factor: semana (4 niveles);

Factor: lugar (2 niveles);

Ambos factores de efector fijo;

Tamaño del experimento: 32 observaciones.

1. **Estudiar si la semana y el lugar son factores determinantes en el número de diatomeas halladas en el agua del río. ¿Hay posibilidad que una semana sea más recomendable en un lugar del río en concreto y no lo sea en el otro lugar?**

Para realizar el análisis de la varianza del modelo anterior se utiliza la función **TwoFactors**

En primer lugar, situarse en el directorio donde este el fichero de datos

> setwd("C:/Users/Usuario/Desktop/Datos")

Realizar el ejercicio con BrailleR

> library("BrailleR")

> datos= read.table("propuesto1.txt",header=TRUE)

> datos$lugar=as.factor(datos$lugar)

> datos$semana=as.factor(datos$semana)

> TwoFactors(Response='efecto',Factor1='lugar',Factor2='semana',Data=datos, Inter='TRUE', HSD=TRUE)

A continuación, se comentan brevemente los resultados obtenidos.

Lo primero que nos muestra es una tabla con la media, desviación típica, tamaño de muestra y error estándar de cada combinación posible entre los factores y un diagrama de puntos comparativos para cada uno de los factores.



**Comparative boxplots**

When boxplots are not included, it is because at least one group size is too small.

**Comparative dotplots**



**Interaction Plot**



A continuación, aparece la tabla de análisis de la varianza.



El único efecto que no es significativo es la interacción de los dos factores, por lo que se debe modificar el modelo suprimiendo la interacción entre ambos factores y realizar un estudio del modelo modificado que se muestra a continuación:

$$y\_{ijk}=μ+τ\_{i}+β\_{j}+u\_{ijk}; i:1,2,3,4; j:1,2;k=1,2,3,4$$

> TwoFactors(Response='efecto',Factor1='lugar',Factor2='semana',Data=datos, Inter='FALSE', HSD=TRUE)



En la nueva tabla vemos que tanto el primer factor como el segundo resultan significativos al 5% (p-valor < 0,05).

Para la validación del modelo, se muestran 4 gráficos para contrastar visualmente la independencia, homocedasticidad y normalidad de los residuos.



Y a continuación los resultados de los test para contratar la homogeneidad de las varianzas de los niveles de cada uno de los factores. Centrándonos en el test de Bartlett vemos que para el factor “lugar” el p-valor es 0.2532 por lo que se acepta que la varianza en ambos lugares en la misma, sin embargo, se rechaza para las distintas semanas (p-valor=0.01863).



1. **Estudiar en qué semana se producen menos contaminación en el río, utilizando el método HSD de Tukey.**
2. **Estudiar en qué lugar del río se producen menos diatomeas, utilizando el método de HSD de Tukey**

Finalmente, mediante los resultados del test HSD de Tukey vemos que las medias son diferentes para los dos lugares. Con respecto a las semanas, no existe diferencia entre la media de las semanas 1 y 3 (p-valor=0.9960692) y las semanas 4 y 2 (p-valor 0.9961944), considerándose el resto de medias significativamente diferentes



 