

PRÁCTICA 2
DISEÑO DE EXPERIMENTOS: MODELOS BIFACTORIALES¹

1. Introducción

En esta práctica veremos cómo ajustar con SPSS los modelos de análisis de la varianza bifactoriales (aditivos o con interacciones).

Utilizaremos los datos del fichero `dientes.sav`. Este fichero contiene 60 observaciones y 3 variables. La variable respuesta es la longitud de los dientes de varios conejos de indias (`longitud`) que han recibido tres dosis diferentes de vitamina C (`dosis`, factor α) administradas de dos formas distintas (`adm`, factor β): bien tomando zumo de naranja (OJ) o tomando ácido ascórbico (VC).

2. Descripción de los datos mediante diagramas de cajas

En primer lugar podemos representar diagramas de cajas para tener una idea general de cómo son los datos. En SPSS podemos ir al menú

Gráficos ↪ Cuadros de diálogo antiguos ↪ Diagramas de caja...

Elegimos la opción `agrupado` y pulsamos `Definir`. En `Variable` situamos la variable respuesta `longitud`. Como `eje de categorías` elegimos el factor `dosis` y en `definir grupos por:` elegimos el otro factor `adm`.

Cuestiones

1. ¿Existen datos que aparecen marcados como atípicos?
2. Trata de anticipar los resultados en cuanto a si cada uno de los dos factores tiene un efecto significativo sobre la respuesta.
3. A la vista del gráfico, ¿crees que puede haber interacciones entre los dos factores?

3. Ajuste del modelo bifactorial con interacciones

Para ajustar el modelo bifactorial con interacciones:

Analizar ↪ Modelo lineal general ↪ Univariante...

En el correspondiente cuadro de diálogo debemos elegir las opciones siguientes:

¹Para escribir estas notas se ha utilizado la versión SPSS 19

- Pasamos la variable respuesta *longitud* como **variable dependiente** y los factores *dosis* y *adm* a la ventana de **factores fijos**.
- En el botón **Post hoc...** pasamos el factor *dosis* a la ventana **Pruebas post hoc para...** y marcamos la opción **Bonferroni**.
- En el botón **Guardar...** marcamos los **Valores Pronosticados** no tipificados y los **Residuos** tipificados.
- En el botón **Opciones...** indicamos que muestre las medias para los niveles de los dos factores, las interacciones, así como la media **OVERALL** o media global. En las opciones marcamos las **Pruebas de homogeneidad**. En la parte inferior se puede seleccionar el nivel de significación deseado (por defecto, $\alpha = 0,05$).

Un vez elegidas las opciones anteriores pulsamos **Aceptar** para obtener el resultado.

Cuestiones

1. Determina el valor de los estimadores de los siguientes parámetros del modelo: μ , α_1 , β_1 y $(\alpha\beta)_{11}$.
2. ¿Cuánto valen las sumas de cuadrados que miden la variabilidad explicada, la variabilidad residual y la variabilidad total?
3. ¿Para qué niveles de significación podemos afirmar que el factor *dosis* es significativo?
4. ¿Para qué niveles de significación podemos afirmar que el factor *adm* es significativo?
5. A nivel $\alpha = 0,01$, ¿podemos decir que existen interacciones significativas entre los dos factores?
6. Si α_i representa el efecto principal de la dosis *i*, contrasta todas las hipótesis de la forma $H_0 : \alpha_i = \alpha_j$ de manera que el nivel de significación global para todas las comparaciones sea 0,01.

4. Diagnóstico del modelo

Al obtener los resultados en el apartado anterior observamos que en el fichero aparecen dos nuevas variables, llamadas **PRE** y **RES**, que contienen los valores ajustados y los residuos respectivamente.

Para obtener un gráfico de probabilidad con el fin de comprobar si los residuos tienen distribución normal, utilizamos el menú:

Analizar \leftrightarrow Estadísticos descriptivos \leftrightarrow Gráficos P-P...

En la ventana de **Variables** debemos incluir la variable que contiene los residuos. Al pulsar **Aceptar** obtenemos el gráfico.

Es también útil representar un histograma de los residuos. Para ello, utilizamos:

Gráficos \leftrightarrow Cuadro de diálogo antiguos \leftrightarrow Histograma...

Pasamos la variable que contiene los residuos a la ventana **Variable** y marcamos la opción **Mostrar curva normal**.

Para representar el gráfico de los residuos frente a los valores ajustados podemos ir al menú:

Gráficos ↔ Cuadro de diálogo antiguos ↔ Dispersión/Puntos...

Elegimos el tipo **Dispersión simple** y pulsamos **Definir**. En el cuadro de diálogo, como variable *Y* elegimos la que contiene los residuos y como variable *X* la que contiene los valores pronosticados o ajustados. Al pulsar **Aceptar** obtenemos el gráfico.

El p-valor de la prueba de Levene puede utilizarse para contrastar la hipótesis nula de homogeneidad de varianzas.

Cuestiones

1. Determina si es razonable suponer homocedasticidad.
2. Determina si es razonable suponer que la variable respuesta tiene distribución normal.
3. ¿Podríamos responder a la cuestión anterior representando un histograma de la variable *longitud* en lugar de un histograma de los residuos?

5. Ajuste del modelo bifactorial aditivo

En este apartado vamos a ajustar un modelo aditivo, es decir, vamos a asumir que no hay interacciones entre los dos factores. Para ello, vamos de nuevo al menú:

Analizar ↔ Modelo lineal general ↔ Univariante...

En el botón **Modelo...** tenemos que elegir **Personalizado**. En el botón **Tipo** seleccionamos **Efectos principales** en lugar de **Interacción**. Finalmente, pasamos a la derecha los nombres de los factores que queramos incluir en el modelo.

En el botón **Opciones...** eliminamos las interacciones de la ventana de la derecha. El resto de opciones no cambian.

Cuestiones

1. En este modelo, ¿tiene el factor *dosis* una influencia significativa en la *longitud*?
2. En este modelo, ¿tiene el factor *adm* una influencia significativa en la *longitud*?
3. Mira los grados de libertad de la suma de cuadrados residual en la tabla ANOVA de este modelo. ¿Por qué no coinciden con los de la tabla que aparece en el formulario de la asignatura?

6. Problema 7 del Tema 2

Para estudiar el efecto de la iluminación (A=natural, B=muy fuerte, C=escasa) en la velocidad de lectura se realiza un experimento. Se mide el número de palabras leídas en un minuto para distintos tipos de papel y tamaño de letra. Los resultados que se obtienen son los siguientes:

	Papel satinado	Papel blanco	Papel color
Letra grande	258 A	230 C	240 B
Letra normal	235 B	270 A	240 C
Letra pequeña	220 C	225 B	260 A

¿Cuántos factores se consideran en el experimento? Construir con SPSS la tabla de análisis de la varianza y contrastar, con un nivel de significación $\alpha = 0,05$ si los factores afectan a la velocidad de lectura.

7. Problema 2 del Tema 2

En un estudio sobre el consumo de gasolina de distintos coches se realiza el siguiente experimento: se toman cuatro coches al azar de un fabricante español, cuatro de un francés, cuatro de un alemán, y cuatro de un japonés. Se prueba un coche de cada fabricante en una gran ciudad durante la hora punta, otro en ciudad fuera de la hora punta, otro se prueba en carretera de montaña y el otro en una carretera llana. El consumo en litros de gasolina por cada 100 kilómetros es:

	Hora punta	Hora normal	Montaña	Carretera llana
Español	14.7	9.4	7.2	6.8
Francés	11.6	7.7	6.8	6.0
Alemán	10.8	7.2	7.2	6.4
Japonés	16.0	10.0	9.3	7.7

- Plantear el modelo adecuado para estudiar el consumo de gasolina con dos factores.
- ¿Qué modelo de coche parece que consume más y qué modelo de coche parece que consume menos? ¿En qué condiciones parece que se consume más y en qué condiciones parece que se consume menos?
- Obtener la tabla de análisis de la varianza y decidir si el modelo de coche tiene una influencia significativa sobre el consumo (al nivel de significación 0.05).
- Comparar de dos en dos el consumo medio de los cuatro modelos de coche, con un nivel de confianza conjunto del 95%. ¿Conclusiones?
- Finalmente, analizar los datos sin tener en cuenta las condiciones en que se conducen los coches, es decir, realizando un análisis de la varianza con un solo factor. Con este modelo, ¿influye el modelo de coche en el consumo de gasolina? ¿Coincide esta conclusión con la obtenida anteriormente? ¿Cuál sería el modelo adecuado y la conclusión correcta? ¿Por qué?