

PRÁCTICA 1
DISEÑO DE EXPERIMENTOS: MODELO UNIFACTORIAL¹

1. Introducción

En esta práctica veremos cómo ajustar un modelo de análisis de la varianza unifactorial con SPSS. También veremos cómo llevar a cabo el diagnóstico del modelo mediante el análisis de los residuos.

Utilizaremos los datos del fichero `fertilizante.sav`. Este fichero (que se encuentra disponible en la página web de la asignatura) consta de dos variables llamadas `fertilizante` y `cosecha`. La primera contiene el tipo de fertilizante utilizado y la segunda el peso de la cosecha obtenida. La información de cada fila del fichero corresponde a cada una de las 30 parcelas.

2. Breve descripción de los datos

Para hacernos una idea general de cómo son los datos, siempre es conveniente hacer un breve estudio descriptivo preliminar. En el caso de un modelo unifactorial podemos representar una diagrama de cajas múltiple (cada caja corresponde a un nivel del factor o grupo) y calcular las medias y desviaciones típicas de cada grupo. En SPSS podemos ir al menú

Analizar ↔ Estadísticos descriptivos ↔ Explorar...

Pasamos la variable respuesta `cosecha` a la *lista de dependientes* y el factor `fertilizante` a la *lista de factores*. En el botón `Gráficos...` podemos eliminar el diagrama de tallo y hojas, que no suele ser muy útil. Pulsamos `Aceptar` para obtener el resultado.

Cuestiones

1. ¿Existen datos que aparecen marcados como atípicos?
2. ¿Cuál es el nivel de fertilizante para el que se obtienen mayores cosechas en la muestra?
3. ¿Cuál es el nivel de fertilizante para el que la cosecha es menos variable?

3. Ajuste del modelo (en el menú “comparar medias”)

Hay dos formas de ajustar un modelo unifactorial con SPSS. La más sencilla es utilizando el menú:

Analizar ↔ Comparar medias ↔ ANOVA de un factor...

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

En el correspondiente cuadro de diálogo debemos elegir **cosecha** como variable dependiente y **fertilizante** como **Factor**. Además:

- En el botón **Post hoc...** marcamos la opción *Bonferroni* y en la parte de abajo seleccionamos el nivel de significación global deseado (por defecto es 0,05).
- En el botón **Opciones...** marcamos la casilla *Descriptivos* si queremos que muestre las medias y desviaciones típicas para cada nivel del factor.

Una desventaja de esta manera de ajustar el modelo es que no permite calcular fácilmente ni los residuos ni los valores ajustados. En el siguiente apartado se presenta una forma alternativa de ajustar el modelo que sí lo permite, aunque es un poco más complicada.

4. Ajuste del modelo (en el menú “modelo lineal general”)

Para ajustar el modelo unifactorial vamos al menú:

Analizar ↔ Modelo lineal general ↔ Univariante...

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

- Pasamos la variable respuesta **cosecha** como *variable dependiente* y el factor **fertilizante** a la ventana de *factores fijos*.
- En el botón **Post hoc...** pasamos el factor **fertilizante** 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 del factor **fertilizante** así como la media **OVERALL**. En las opciones marcamos las *Pruebas de homogeneidad* y las *Estimaciones del tamaño del efecto*.

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

Cuestiones

1. Determina el valor de los estimadores de todos los parámetros del modelo.
2. ¿Cuánto valen las sumas de cuadrados que miden la variabilidad explicada, la residual y la total?
3. ¿Cuál es el p-valor de la hipótesis $H_0 : \mu_1 = \mu_2 = \mu_3$? ¿Se rechaza esta hipótesis a nivel $\alpha = 0,05$? ¿Y a nivel $\alpha = 0,005$?
4. ¿Por qué el error típico de lo que SPSS llama *Gran media* es diferente (y menor) que el error típico de las medias de cada uno de los niveles de fertilizante?
5. Contrasta todas las hipótesis de la forma $H_0 : \mu_i = \mu_j$ de manera que el nivel de significación global para todas las comparaciones sea $\alpha = 0,05$.

5. Análisis de los residuos

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 ↔ Estadísticos descriptivos ↔ Gráficos P-P...

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

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.

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

Gráficos ↔ Cuadro de diálogo antiguos ↔ Histograma...

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

Cuestiones

1. Determina si es razonable suponer que las varianzas de los tres grupos son iguales.
2. Determina si es razonable suponer que la variable respuesta tiene distribución normal.

6. Otro ejemplo

Las precipitaciones caídas en un país han disminuido de manera preocupante durante el último año. Antes de tomar ninguna medida se decide hacer un estudio previo para saber si el descenso de las lluvias se produjo de forma homogénea. Para ello se seleccionan aleatoriamente cinco estaciones meteorológicas en cada una de las cuatro regiones del país, obteniéndose los siguientes porcentajes de disminución de las precipitaciones en cada una de ellas:

Región Este	Región Norte	Región Oeste	Región Sur
10,4	12,8	11,2	13,9
12,8	14,2	9,8	14,2
15,6	16,3	10,7	12,8
9,2	10,1	6,3	15,0
8,7	12,0	12,4	13,7

- (a) Plantea claramente todos los elementos y las hipótesis del modelo para comparar los porcentajes de disminución de las precipitaciones en las 4 regiones.
- (b) ¿En qué zona parecen haber disminuido más las precipitaciones?
- (c) Calcula la tabla ANOVA y contrasta la hipótesis nula de que las medias de disminución del porcentaje de lluvias en el país fueron las mismas en las cuatro regiones (tomar $\alpha = 0,10$).
- (d) Compara las medias de las diferentes regiones de dos en dos, con un nivel de confianza global del 90 %.
- (e) Analiza los residuos con el fin de verificar si las hipótesis de homocedasticidad y de normalidad son razonables para estos datos.