Práctica 4: Experimentos con uno y dos factores

2ª parte

Modelos Lineales

Grados en Estadística y Matemáticas

Índice

Experimentos con dos factores

Experimentos con dos factores

Toma de datos

Supongamos que estamos estudiando la posible influencia que dos factores (con r y s niveles respectivamente) pueden ejercer sobre una variable cuantitativa continua Y.

- La influencia de los dos factores puede solaparse de algún modo, es decir ambos factores pueden interactuar.
- Puede ocurrir que los niveles de uno de los dos factores estén supeditados a los del otro factor.
- Tenemos por tanto dos posibles disposiciones de los dos factores, clasificación cruzada y jerarquizado o anidado.

Ejemplo 1

Nos interesa determinar si el tiempo de coagulación (en minutos) del plasma sanguíneo presenta diferencias para 3 tratamientos y 2 concentraciones de adrenalina mezclada con el plasma.

Para cada combinación de tratamiento y concentración de adrenalina, se tomaron 3 observaciones independientes.

Los resultados se encuentran en el archivo `coagulacion.txt`.

Se trata de un diseño de clasificación cruzada

Experimentos con dos factores

Ejemplo 1

La disposición de los datos sería:

		Concentración 1	Concentración 2
Tratamiento	1	9.8 10.1 9.8	11.3 10.7 10.7
	2	9.2 8.6 9.2	10.3 10.7 10.2
	3	8.4 7.9 8.0	9.8 10.1 10.1

Ejemplo 2

Se estudia el grosor del remate de la superficie de piezas metálicas elaboradas en 4 máquinas. En este experimento cada máquina es manejada por 3 operadores diferentes, entre los cuales también estamos interesados en detectar diferencias; de cada uno de los operarios se eligen dos piezas, registrándose el grosor del remate de cada una de ellas.

A causa de la localización de las máquinas, se utilizaron distintos operarios para cada una de ellas. Los datos aparecen en el archivo `remate.dat`.

Se trata de un diseño anidado o jerárquico

Experimentos con dos factores

Ejemplo 2

La disposición de los datos sería:

Máquina	Máquina 1			Máquina 2			Máquina 3			Máquina 4		
Obrero	1	2	3	1	2	3	1	2	3	1	2	3
	79	94	46	92	85	76	88	53	46	36	40	62
	62	74	57	99	79	68	75	56	57	53	56	47

Dos factores clasificación cruzada

Dos factores clasificación cruzada

- Cada combinación de un tratamiento del primer factor con un tratamiento del segundo factor da lugar a una celda.
- En cada una de las celdas se toma una muestra con una o varias unidades experimentales en las que se mide la variable Y.
- Si el número de observaciones es el mismo en todas las celdas, se dice que el modelo es balanceado.

En el ejemplo 1 tenemos un modelo de clasificación cruzada con 3 observaciones por celda.

Dos factores clasificación cruzada

Interacción

- Puede haber una interacción de los factores en su posible influencia sobre Y.
- Diremos que existe interacción si el comportamiento de la variable Y en relación a un factor depende del tratamiento aplicado del otro factor. Podemos visualizar gráficamente la interacción con el correspondiente gráfico.
- En caso de que no exista interacción, diremos que el modelo es aditivo. En modelos aditivos, la influencia de los dos factores sobre Y se produce de forma separada.

Tenemos dos modelos matemáticos para el diseño de dos factores de clasificación cruzada: con interacción y sin interacción o aditivo.

Dos factores clasificación cruzada

Modelo de dos factores sin interacción

El modelo matemático de dos factores sin interacción es el siguiente:

$$Y_{ijk} = \mu + \tau_i + \delta_j + \varepsilon_{ijk}$$

donde

- Y_{ijk}: representa la observación k de Y en la celda (i, j).
- μ: es la media poblacional de Y sin ninguna fuente de variabilidad presente.
- τ_i: efecto que sobre la media poblacional de Y tiene el tratamiento i del primer factor.
- δ_j: efecto sobre la media poblacional de Y tiene el tratamiento j del segundo factor.
- ε_{ijk}: componente aleatoria en la medición de Y_{ijk} (no observable y con media 0).

Modelo de dos factores sin interacción

En este modelo se plantean los llamados contrastes de efectos principales:

Para el primer factor

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_r$$
$$H_1 : \text{los efectos de los tratamientos no son iguales}$$

Para el segundo factor

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_s$$
$$H_1 : \text{los efectos de los tratamientos no son iguales}$$

Tabla ANOVA

F.V.	g.l.	Suma de cuadrados	M.C.	F
Ft. A	$r-1$	$Q_A = \sum_{i=1}^{r} sn\bar{Y}_{i..}^2 - rsn\bar{Y}_{...}^2$	$\dfrac{Q_A}{r-1}$	$F_A = \dfrac{(rsn-r-s+1)Q_A}{(r-1)Q_0}$
Ft. B	$s-1$	$Q_B = \sum_{j=1}^{s} rn\bar{Y}_{.j.}^2 - rsn\bar{Y}_{...}^2$	$\dfrac{Q_B}{s-1}$	$F_B = \dfrac{(rsn-r-s+1)Q_B}{(s-1)Q_0}$
Error	$rsn-r-s+1$	$Q_0 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} Y_{ijk}^2 - \sum_{i=1}^{r} sn\bar{Y}_{i..}^2 - \sum_{j=1}^{s} rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2$	$\dfrac{Q_0}{rsn-r-s+1}$	

Dos factores clasificación cruzada

Hipótesis teóricas

Para utilizar este modelo hemos de asumir normalidad, homocedasticidad e independencia en las observaciones, que se comprobará a través del análisis de los residuos.

Comparaciones múltiples

Si aceptamos H_1, habremos detectado diferencias entre los tratamientos en cuanto al comportamiento de la variable Y. Para saber en qué tratamientos residen dichas diferencias hay que aplicar un procedimiento de comparaciones múltiples o comparaciones dos a dos, es decir realizaremos de modo simultáneo, todos los contrastes posibles

$$\begin{array}{l} H_0 : \tau_i = \tau_j \\ H_1 : \tau_i \neq \tau_j \end{array} \qquad \text{y} \qquad \begin{array}{l} H_0 : \delta_i = \delta_j \\ H_1 : \delta_i \neq \delta_j \end{array}$$

.

Dos factores clasificación cruzada

Modelo de dos factores con interacción

El modelo matemático de dos factores con interacción es el siguiente:

$$Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \varepsilon_{ijk}$$

donde

- Y_{ijk}: representa la observación k de Y en la celda (i, j).
- μ: es la media poblacional de Y sin ninguna fuente de variabilidad presente.
- τ_i: efecto que sobre la media poblacional de Y tiene el tratamiento i del primer factor.
- δ_j: efecto sobre la media poblacional de Y tiene el tratamiento j del segundo factor.
- $(\tau\delta)_{ij}$: efecto que la interacción entre el tratamiento i del primer factor y el tratamiento j del segundo factor tienen sobre la media poblacional de Y.
- ε_{ijk}: componente aleatoria en la medición de Y_{ijk} (no observable y con media 0).

Dos factores clasificación cruzada

Modelo de dos factores con interacción

En este modelo se plantea principalmente un contraste:

$$H_0 : (\tau\delta)^*_{ij} = 0 \text{ para todo } i,j$$
$$H_1 : (\tau\delta)^*_{ij} \neq 0 \text{ para algún par } i,j$$

o escrito de forma intuitiva

$$H_0 : \text{No hay interacción en el modelo}$$
$$H_1 : \text{Sí hay interacción en el modelo}$$

Dos factores clasificación cruzada

Tabla ANOVA

F.V.	g.l.	Suma de Cuadrados	M.C.	F
Inter.	$(r-1)(s-1)$	$Q_{AB}=\sum_{i,j} n\bar{Y}_{ij.}^2 - \sum_{i} sn\bar{Y}_{i..}^2 - \sum_{j} \bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2$	$\frac{Q_{AB}}{(r-1)(s-1)}$	$F_{AB}=\frac{rs(n-1)Q_{AB}}{(r-1)(s-1)Q_0}$
Error	$rs(n-1)$	$Q_0 = \sum_{i,j,k} Y_{ijk}^2 - \sum_{i,j} \bar{Y}_{ij.}^2$	$\frac{Q_0}{rs(n-1)}$	

Dos factores clasificación cruzada

Modelo de dos factores con interacción

Si en el contraste anterior se acepta H_0, entonces se pasa a trabajar con el modelo aditivo (sin interacción).

Si en el contraste anterior se acepta H_1, entonces:

- Se contrastan los efectos principales como en el modelo aditivo. En este caso la interpretación intuitiva de estos contrastes es más complicada.
- Para cada nivel de un factor, se ajusta un modelo de Y frente al otro factor (diseño completamente aleatorizado).

Dos factores clasificación cruzada

Ejemplo 3

En una fábrica se produce un determinado tipo de fibra textil. Se piensa que tanto el telar como el proveedor de la seda puede tener influencia en la resistencia de la fibra.

Se diseña un experimento en el que se elaboran dos muestras fibra en cada uno de los cuatro telares y utilizando seda de los tres proveedores de la fábrica.

Los datos de resistencia obtenidos se encuentran en el archivo `fibra.dat`.

Analizar la influencia que el telar y el proveedor pueden tener sobre sobre la resistencia de la fibra producida.

Diseño anidado o jerarquizado

Diseño anidado o jerarquizado

Estamos estudiando la posible influencia que dos factores (con r y s niveles respectivamente) pueden ejercer sobre una variable cuantitativa continua Y

- Los dos factores no están al mismo nivel y por tanto no no se pueden combinar todos los niveles de un factor con todos los niveles del otro.
- Los niveles de un factor están supeditados a los del otro.
- A éste último se le conoce como factor principal, mientras que al primer nos referiremos como factor anidado
- Si el número de observaciones es el mismo en todos los niveles del factor anidado, se dice que el modelo es balanceado.

Diseño anidado o jerarquizado

Ejemplo 2

Se estudia el grosor del remate de la superficie de piezas metálicas elaboradas en 4 máquinas. En este experimento cada máquina es manejada por 3 operadores diferentes, entre los cuales también estamos interesados en detectar diferencias; de cada uno de los operarios se eligen dos piezas, registrándose el grosor del remate de cada una de ellas.

A causa de la localización de las máquinas, se utilizaron distintos operarios para cada una de ellas. Los datos aparecen en el archivo `remate.dat`.

Se trata de un diseño anidado o jerarquizado

Diseño anidado o jerarquizado

Ejemplo 2

La disposición de los datos sería:

Máquina	Máquina 1			Máquina 2			Máquina 3			Máquina 4		
Obrero	1	2	3	1	2	3	1	2	3	1	2	3
	79	94	46	92	85	76	88	53	46	36	40	62
	62	74	57	99	79	68	75	56	57	53	56	47

Por tanto la máquina sería el factor principal y el obrero sería el factor anidado.

Diseño anidado o jerarquizado

Modelo para un diseño anidado

El modelo matemático para un diseño de dos factores anidados es el siguiente:

$$Y_{ijk} = \mu + \tau_i + (\tau\delta)_{ij} + \varepsilon_{ijk}$$

donde

- Y_{ijk}: representa la observación k de Y en la celda (i,j).
- μ: es la media poblacional de Y sin ninguna fuente de variabilidad presente.
- τ_i: efecto que sobre la media poblacional de Y tiene el tratamiento i del factor principal.
- $(\tau\delta)_{ij}$: efecto que sobre la media poblacional de Y tiene el tratamiento j del factor anidado, supeditado al tratamiento i del factor principal.
- ε_{ijk}: componente aleatoria en la medición de Y_{ijk} (no observable y con media 0).

Diseño anidado o jerarquizado

En este modelo se plantea inicialmente un contraste:

$$H_0 : (\tau\delta)_{i1} = \cdots = (\tau\delta)_{is} \;\text{ para todo }\; i$$
$$H_1 : \;\text{alguna de estas igualdades no es cierta}$$

o escrito de forma intuitiva

$$H_0 : \text{El factor anidado NO presenta influencia sobre } Y$$
$$H_1 : \text{El factor anidado SÍ presenta influencia sobre } Y$$

Diseño anidado o jerarquizado

Si en el contraste anterior se acepta H_0, entonces se ajusta un modelo con Y frente al factor principal (diseño completamente aleatorizado).

Si en el contraste anterior se acepta H_1, entonces:

- Se contrastan los efectos principales del factor principal, es decir se contrasta

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_r$$
$$H_1 : \text{los efectos de los tratamientos no son iguales}$$

- Para cada nivel del factor principal, se ajusta un modelo de Y frente al factor anidado (diseño completamente aleatorizado)

Diseño anidado o jerarquizado

Tabla ANOVA

F. V.	g.l.	Suma de Cuadrados	M.C.	F
Fct anid	$r(s-1)$	$Q_1 = \sum_{i,j} n\bar{Y}_{ij.}^2 - \sum_i sn\bar{Y}_{i..}^2$	$\frac{Q_1}{r(s-1)}$	$F = \frac{rs(n-1)Q_1}{r(s-1)Q_0}$
Error	$rs(n-1)$	$Q_0 = \sum_{i,j,k} Y_{ijk}^2 - \sum_{i,j} n\bar{Y}_{ij.}^2$	$\frac{Q_0}{rs(n-1)}$	

Diseño anidado o jerarquizado

Hipótesis teóricas

Para utilizar este modelo hemos de asumir normalidad, homocedasticidad e independencia entre las observaciones, que se comprobará a través del análisis de los residuos.

Comparaciones múltiples

Si al contrastar los efectos principales del factor principal aceptamos H_1, habremos detectado diferencias entre sus tratamientos en cuanto al comportamiento de la variable Y. Para saber en qué tratamientos residen dichas diferencias hay que aplicar un procedimiento de comparaciones múltiples o comparaciones dos a dos, es decir realizaremos de modo simultáneo, todos los contrastes posibles

$$\begin{array}{c} H_0 : \tau_i = \tau_j \\ H_1 : \tau_i \neq \tau_j \end{array}$$

.

Diseño anidado o jerarquizado

Ejemplo 4

La Confederación Hidrográfica del Guadiana realiza un estudio para determinar la concentración media de mercurio en dicho río a su paso por Extremadura. Para ello se seleccionan tres estaciones pertenecientes a la Confederación en las tres comarcas que cruza el río (Siberia, Vegas Altas y Vegas Bajas).

Cada uno de estos observatorios toma 6 mediciones (en ppm) de la sustancia en cuestión. Los datos aparecen en el archivo `conc.dat`.

Analizar las diferencias entre comarcas y entre estaciones.