

TEMA 3. CONCEPTUALIZACIÓN DEL DISEÑO DE SERIES TEMPORALES (II): FUNDAMENTACIÓN ANALÍTICA.

EL ESTUDIO DE LA INDEPENDENCIA O NO DE LAS OBSERVACIONES DENTRO DE LAS SERIES: IMPLICACIONES EN EL USO DE ANOVA.

1. INTRODUCCIÓN

Cuando realizamos una intervención y queremos observar si nuestro tratamiento ha tenido efectos o si, por el contrario, no ha producido cambio alguno, es decir, cuando pretendemos comparar los datos antes y después a la aplicación del tratamiento, debemos someter los datos a algún tipo de análisis estadístico. Uno de ellos es el llamado ANOVA. El cálculo del ANOVA se supedita al cumplimiento de determinados requisitos que exigen normalidad (distribución normal de los datos), homocedasticidad (homogeneidad de las varianzas de los términos de error) e independencia (entre los errores de cualquier par de puntuaciones). En el momento en que no se cumplan alguno de ellos, los datos no pueden analizarse con esta prueba estadística o si se realiza, los resultados estarán afectados por este hecho.

En los diseños transversales, puesto que los sujetos de un tratamiento son distintos a los de otro, no tiene por qué existir relación entre las puntuaciones de ambos, esto es, habrá independencia de los términos de error. Sin embargo, en los diseños longitudinales es una misma persona o grupo quien aporta puntuaciones en los distintos momentos temporales, de modo que éstas pueden presentar dependencia, ya sea por pertenecer a un mismo sujeto, o por ser emitidas en una serie temporal en la que pueden influirse unas a otras. Así, en los diseños longitudinales de series temporales se hace necesaria la comprobación del supuesto de independencia del ANOVA.

2. DEFINICIÓN

Cuando las medidas obtenidas a lo largo del tiempo (denominadas *datos secuenciales*) no son independientes entre sí, es decir, cuando están relacionadas unas con otras, se produce el fenómeno de la *dependencia serial*. Este constructo intenta explicar la relación entre una ocurrencia de conducta y otra, a lo largo de una serie temporal. Se dice que dos ocurrencias de conducta son independientes cuando la probabilidad condicional de una de ellas, dada la otra, no es diferente de la probabilidad incondicional.

El hecho de que nuestros datos muestren dependencia serial nos indica que existe cierta asociación entre las medidas, dicho de otro modo, que los datos secuenciales están correlacionados unos con otros. Así, para determinar si existe dependencia serial atendemos a si existe correlación entre las observaciones sucesivas de la serie. La cuantificación de la dependencia serial se basa en el estudio de la autocorrelación, que se define como los valores que en el tiempo t se relacionan con los del tiempo $t+k$ (siendo k un retardo 1, 2, etc). Más adelante nos extenderemos más en caracterizar la autocorrelación, pero baste decir aquí que, si ésta es significativa, podemos predecir las observaciones sucesivas a partir de las anteriores; nuestros datos presentan entonces dependencia serial. Y la dependencia serial trae, como apuntábamos antes, determinadas consecuencias al realizar el ANOVA.

3. PATRONES DE NO INDEPENDENCIA

En los datos de un diseño de series temporales podemos encontrar diferentes casos de dependencia puesto que, como hemos señalado, ésta puede darse por diversas causas. A continuación vemos algunos de estos patrones posibles.

1) Debido a los grupos

	Y11	Y12	Y13	Y21	Y22	Y23	Y31	Y32	Y33
Y11	-								
Y12	r	-							
Y13	r	r	-						
Y21	0	0	0	-					
Y22	0	0	0	r	-				
Y23	0	0	0	r	r	-			
Y31	0	0	0	0	0	0	-		
Y32	0	0	0	0	0	0	r	-	
Y33	0	0	0	0	0	0	r	r	-

En este cuadro existe relación entre las puntuaciones de sujetos del mismo grupo pues, como observamos, sólo en ese caso la correlación es distinta de cero.

Si los patrones de no independencia se deben a la configuración de los grupos es que dentro de un mismo grupo, los sujetos que lo conforman, están asociados de alguna forma, es decir, que sus medidas mantienen algún tipo de correlación. Esta dependencia puede venir de tres fuentes distintas:

Algunas causas para este patrón de dependencia podemos encontrarlas en:

- Los efectos de un muestreo no aleatorio. Si la asignación de los sujetos a cada grupo se realiza de acuerdo a un determinado criterio, dichos sujetos pueden estar relacionados en cuanto a ciertas características. Este hecho haría que sus medidas también se relacionaran en cierto grado con las medidas de los miembros de su mismo grupo.
- La historia local intragrupo. Esto es, que los sujetos de cada grupo comparten determinada historia común que influye en su conducta.
- Interacción social como liga grupal. Los miembros de un grupo actúan de forma distintiva respecto a los demás sólo por serlo.

La medida estadística empleada para cuantificar este patrón es la *correlación intraclase*, que se calcula:

$$(C_{Mentre} - C_{Mintra}) / (C_{Mentre} + C_{Mintra}(n-1))$$

Su valor puede ser +1 (indicando que la semejanza intragrupo es mayor que la entregupo) y $-1/(n-1)$.

2) Debido a la secuencia temporal

La forma en que esté configurada la secuencia temporal de un diseño de series temporales también puede influir en el estudio de la dependencia serial de los datos. En este caso la dependencia pueda derivar del hecho de que las conductas se sucedan en una secuencia temporal, en la que cada una puede tener efecto sobre las demás.

En el *modelo autorregresivo de orden 1* cada variable (cada conducta) se relaciona con la siguiente, y así sucesivamente, aunque cada vez en menor grado. A medida que avanza la serie el valor de las correlaciones va disminuyendo. Ej: Y1-Y2 están correlacionadas con un $\rho=0.3$; Y1-Y3 tendrían un valor de $\rho=(0.3)^2$; Y1-Y4 tienen un $\rho=(0.3)^3$ y así sucesivamente.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9
Y1	-								
Y2	ρ	-							
Y3	ρ^2	ρ	-						
Y4	ρ^3	ρ^2	ρ	-					
Y5	ρ^4	ρ^3	ρ^2	ρ	-				
Y6	ρ^5	ρ^4	ρ^3	ρ^2	ρ	-			
Y7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ	-		
Y8	ρ^7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ	-	
Y9	ρ^8	ρ^7	ρ^6	ρ^5	ρ^4	ρ^3	ρ^2	ρ	-

Modelo autorregresivo de orden 1

En el *modelo de media móvil de orden 1* cada variable se relaciona sólo con la que le sigue inmediatamente. Ej: Y4 sí está correlacionada con Y5 pero no con Y6.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9
Y1	-								
Y2	ρ	-							
Y3	0	ρ	-						
Y4	0	0	ρ	-					
Y5	0	0	0	ρ	-				
Y6	0	0	0	0	ρ	-			
Y7	0	0	0	0	0	ρ	-		
Y8	0	0	0	0	0	0	ρ	-	
Y9	0	0	0	0	0	0	0	ρ	-

Modelo de media móvil de orden 1

Tendríamos un modelo de media móvil de orden 1, por ejemplo, en un test en el que la respuesta a una pregunta determinara la respuesta a la siguiente, pero no tuviera relación con las demás.

Podríamos hablar de otros muchos patrones de no independencia, ya que son muchos los existentes, como aquéllos que se basan de ciclos, tendencias, etc.

En general, la variación secuencial puede deberse a causas exógenas (como sería, por ejemplo, que la conducta de beber alcohol dependiera del día de la semana) o a la propia variable de conducta (por ejemplo, si el hecho de comer mucho un día hiciera que al día siguiente se comiera menos).

La medida estadística empleada para cuantificar la dependencia serial es la *autocorrelación*, que se refiere a una correlación entre puntos de datos separados por diferentes intervalos temporales (retardos) en las series (Barlow y Hersen, 1988). Concretamente la dependencia serial se refleja en una autocorrelación de intervalo 1 (r_1), que se obtiene mediante el emparejamiento de la observación inicial con la siguiente, y así sucesivamente. La forma de calcular la autocorrelación es hallar el cociente entre covarianza a partir de ciertos retardos [orden 1 (m_1-m_2); orden 2 ($m_1-m_2-m_3$)] y la varianza de la serie. Su valor oscila entre -1 y $+1$. Si la autocorrelación es significativamente diferente de cero significa que la conducta del sujeto en cierto punto puede ser predicha a partir de la conducta en la ocasión anterior. El signo de la autocorrelación no indicará la dirección de la predicción.

3) Debido a la secuencia espacial

También encontramos patrones de dependencia debidos a la secuencia espacial, los cuales pueden ser abordados de forma similar al caso anterior. En este caso, la dependencia de las observaciones se debe al lugar donde fueron tomadas. Las medidas tomadas en un mismo espacio tienden a ser más semejantes que las que se obtienen en sitios lejanos entre sí.

Un ejemplo de esto es el *Modelo del vecino más cercano*. Este modelo afirma que sólo correlacionan las observaciones de las barriadas adyacentes y no las de las barriadas separadas, dentro de una ciudad que se compone de la siguiente manera:

1	2	3
4	5	6
7	8	9

Veamos cuál sería la matriz de datos que se obtendría siguiendo el postulado de este modelo:

	X	X2	X3	X4	X5	X6	X7	X8	X9
X1									
X2	ρ								
X3	0	ρ							
X4	ρ	ρ	0						
X5	ρ	ρ	ρ	ρ					
X6	0	ρ	ρ	0	ρ				
X7	0	0	0	ρ	ρ	0			
X8	0	0	0	ρ	ρ	ρ	ρ		
X9	0	0	0	0	ρ	ρ	0	ρ	

Como vemos, cada sólo correlacionan los datos de las barriadas conlindantes. Por ejemplo, los datos de la barriada 7 sólo tienen un índice de correlación distinto de cero con respecto a los datos de las barriadas 4, 5 y 8.

Esta dependencia puede deberse a distintos factores, como son:

- Contagio: la proximidad hace que se difundan los mismos estilos de conducta, las mismas opiniones, los mismos problemas, etc.
- Historia o destinos comunes: los sujetos que viven en un mismo espacio han compartido sucesos vitales y pueden tener propósitos en común.
- Interacción: el compartir un espacio común aumenta la probabilidad de que se de interacción y se produzcan fenómenos de influencia grupal, dándose un patrón semejante al de no independencia debida a los grupos.

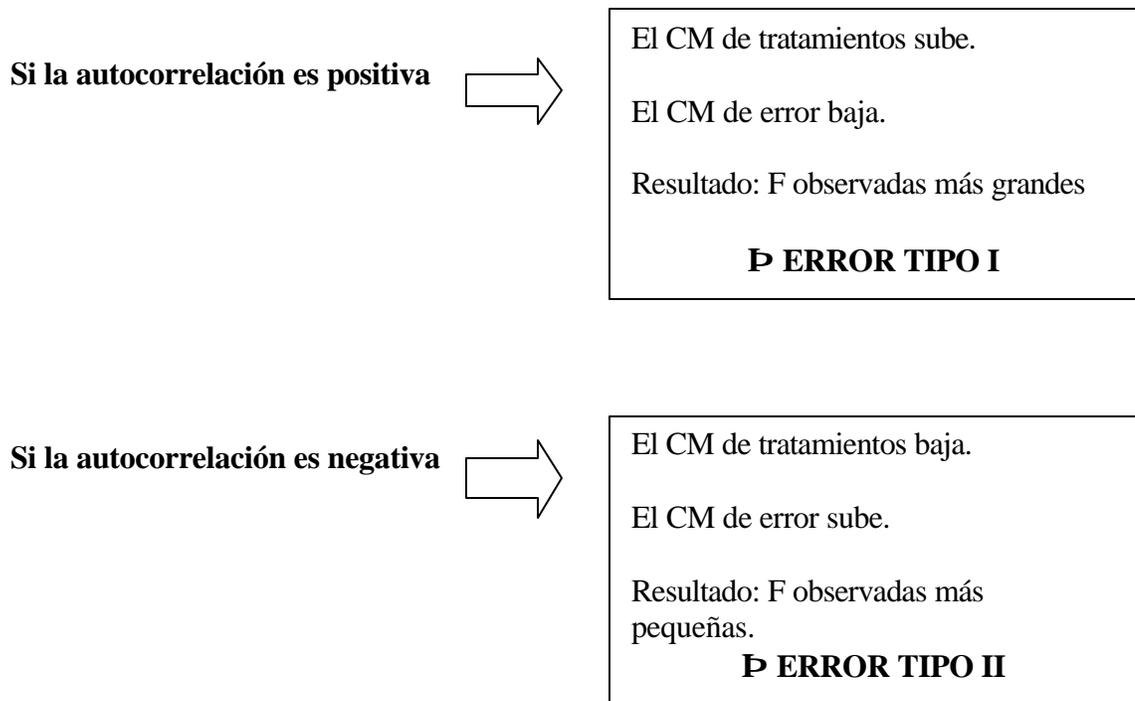
4. CONSECUENCIAS EN EL ANOVA AL VIOLAR LA INDEPENDENCIA

Cuando las autocorrelaciones no se desvían significativamente de cero podemos decir que no hay dependencia entre los datos: los errores son aleatorios. En este caso se pueden analizar con las pruebas estadísticas convencionales (t y F).

Sin embargo, cuando sí existe dependencia se pueden producir grandes sesgos en ambas pruebas. La estimación de los grados de libertad se hace de forma errónea y se cae en error tipo I (rechazar la hipótesis nula cuando es acertada) y tipo II (aceptar la hipótesis nula cuando es falsa). Concretamente, si la correlación es positiva, aumenta el cuadrado medio (CM) de los tratamientos y disminuye el CM del error. Así, obtenemos una F observada mayor de lo que debiera y cometemos error tipo I.

Si, por el contrario, la correlación es negativa, disminuye el CM de los tratamientos y aumenta el del error. Con ello infraestimamos la F observada y caemos en error tipo II.

En el siguiente gráfico se esquematiza lo que puede ocurrir en cada caso:



Vemos, por tanto, que es imprescindible realizar pruebas de autocorrelación a los datos de series temporales antes de aplicarles cualquier prueba estadística. Aunque no hay que olvidar que, a pesar de estos intentos por minimizar los errores, siempre está presente el riesgo de que los datos con que contamos estén influidos por un factor de variación espúreo, interno a la propia naturaleza de la conducta que estamos midiendo, y que puede llevar a que nuestros resultados sean inconcluyentes.