

Comentarios de Bioestadística

Regresión de Prais-Winsten: Evaluación de tendencias lineales

Gabriel Cavada Ch.^{1,2}
¹Facultad de Medicina, Universidad de los Andes.
²Escuela de Salud Pública, Universidad de Chile.

Prais-Winsten regression: lineal trends assessment

En toda ocasión que se describe un problema de salud, un problema clásico es medirlo a través de prevalencias, incidencias o tasas, sino también cómo estas medidas evolucionan en el tiempo. En general estos indicadores están disponibles anualmente, por lo tanto disponer de una larga serie de datos, que permita usar metodologías para analizar series de tiempo, resulta prohibitivo. Aún más si nuestra atención se fija en enfermedades como la diabetes I en pacientes pediátricos, cuyo estudio de tendencias tiene un horizonte de unos 20 años. Por ello es útil conocer un método de evaluación de tendencias lineales (cambios temporales).

Por ejemplo, la Tabla 1 muestra las tasas de mortalidad por cáncer de páncreas en Chile desde el año 1997 al 2012 (Fuente: DEIS, Ministerio de Salud de Chile).

La Figura 1 muestra la tendencia de dichas tasas. Es claro que la evolución de esta tasa es lineal, por lo tanto un incauto analista, encontraría la significación de esta tendencia usando los algoritmos clásicos de la regresión lineal, además agregaría el coeficiente de correlación de Pearson y otras delicadezas. Análisis que estaría incorrecto, pues las tasas no son observaciones independientes en el tiempo, sino que existe una autocorrelación, es decir, que la tasa de un año cualquiera, depende de lo que ocurrió el año anterior. Si este supuesto se obvia, se introduce una heterocedasticidad artificial que impediría conocer en forma certera la velocidad de cambio de esta tasa, pues las varianzas de esta velocidad estarían afectadas por estos residuos cuya variabilidad cambia en el tiempo. Esta situación la corrige la estimación de Prais-Winsten, que es

Tabla 1. Tasa de mortalidad por cáncer de páncreas en Chile

Año	Tasa*
1997	4,1
1998	4,4
1999	4,8
2000	4,5
2001	4,6
2002	5,2
2003	5,6
2004	5,6
2005	5,4
2006	5,6
2007	5,8
2008	6,3
2009	6,3
2010	6,5
2011	6,4
2012	6,9

*Por 100 mil habitantes.



Figura 1.

Comentarios de Bioestadística

un modelo autoregresivo de orden 1, es decir, lo que pasa “hoy” correlaciona con lo que pasó “ayer”, en símbolos:

$$Y_t = \rho Y_{t-1}$$

Donde ρ denota la correlación y el sub índice el tiempo.

Si la tendencia lineal se quisiera estimar mediante la regresión lineal simple, el error cuadrático estaría dado por la expresión:

$$\varepsilon^2 = (Y_t - \alpha - \beta t)^2$$

Estimando sólo los parámetros α y β , y este error no se hace cargo de la dependencia “auto regresiva”, mientras que al considerar este efecto la expresión del error cuadrático es:

$$\varepsilon^2 = (\rho Y_{t-1} - \alpha - \beta t)^2$$

Lo que permite estimar esta dependencia temporal.

En nuestro ejemplo:

Al estimar la tendencia por regresión lineal simple se encuentra:

tasa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
t	.1729412	.0109531	15.79	0.000	.1494491 .1964332
_cons	4.202941	.0964246	43.59	0.000	3.996131 4.409751

Es decir, que por cada año se observa un aumento de la tasa de 0,173, pero no se sabe en cuánto influye la tasa anterior.

La regresión de Prais-Winsten entrega los siguientes resultados:

tasa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
t	.1734462	.0117831	14.72	0.000	.1481738 .1987185
_cons	4.199154	.1040909	40.34	0.000	3.975901 4.422407
rho	.0914577				

La velocidad de cambio es prácticamente la misma, sin embargo, los intervalos de confianza difieren a partir del tercer decimal, esto es demasiado, si se considera que las tasas están por 100 mil habitantes, es decir, nos debería interesar una precisión hasta el quinto decimal. Sin considerar la información extra de que la tasa de un año cualesquiera depende un 9,14% de lo que ocurrió en el año inmediatamente anterior (ρ).

La conclusión es que a veces muy poco es demasiado.