

Comentarios de Bioestadística

Regresión percentílica

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

Una pregunta metodológica recurrente, es asociar una respuesta “Y”, con un conjunto de covariables, cuando la respuesta no sigue una distribución normal o gaussiana; si la respuesta es continua y con una distribución de probabilidades conocida la solución está en los modelos lineales generalizados. Pero si la respuesta “Y” no tiene una distribución de probabilidades conocida o si “Y” es de naturaleza ordinal, como por ejemplo la intensidad del dolor medida a través de la escala EVA, o, niveles de desarrollo puberal medidos a través de los tanners, los modelos lineales generalizados no dan una respuesta satisfactoria. Esta última situación justifica conocer la regresión percentílica.

La regresión percentílica intenta explicar un determinado percentil de la respuesta “Y” a través de un conjunto de covariables en forma lineal, es decir, estima los parámetros de una asociación del tipo:

$$Y_k = X\beta$$

Donde Y_k , es el k-ésimo percentil de la respuesta y la expresión $X\beta$ la combinación lineal de covariables que explican esta.

Usualmente se intenta estimar los coeficientes de la relación:

$$Y_{0,5} = X\beta$$

Es decir, explicar el percentil 50 o mediana de la respuesta por la combinación lineal de covariables.

Obviamente el método es robusto, en el sentido que no interesa qué distribución de probabilidades tiene la respuesta “Y”, como ya se dijo sólo se requiere que la respuesta esté medida a lo menos en escala ordinal.

Si se tuviera sólo una covariable y esta identificara grupos, la regresión percentilar sería una excelente alternativa al test de Kruskal-Wallis (ANOVA no paramétrico), con la ventaja que inmediatamente se tendría la evaluación de los p-valores asociados a comparaciones múltiples.

Ilustremos con un de ejemplo:

Supongamos que una Unidad de tratamiento paliativo del dolor en pacientes oncológicos, desea saber que pacientes se benefician más respecto de la disminución del EVA, ajustando por edad y género. Si la edad está en años cumplidos y género distingue entre hombres y mujeres, los datos se muestran a continuación:

edad	Sexo (0:hombres 1: mujeres)	eva
84	0	6
66	0	9
64	0	8
62	0	8
60	0	9
33	0	6
85	0	8
60	0	8
57	0	5
61	0	5
61	1	3
97	1	5
51	1	3
53	1	2
42	1	1
92	1	3
58	1	3
54	1	1
39	1	3
89	1	1

Por género, ¿Quiénes tienen más sensación de dolor? La tabla siguiente muestra los cuartiles de dolor por género:

Comentarios de Bioestadística

sexo	N	min	p25	p50	p75	max
0	10	5	6	8	8	9
1	10	1	1	3	3	5
Total	20	1	3	5	8	9

En hombres la mediana de EVA es 8, mientras que en mujeres es de 3. Esta diferencia, ¿es significativa?

```
. qreg eva sexo
Iteration 1: WLS sum of weighted deviations = 22.861757
Iteration 1: sum of abs. weighted deviations = 21
Median regression          Number of obs = 20
Raw sum of deviations      47 (about 5)
Min sum of deviations      21 Pseudo R2 = 0.5532
```

eva	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Sexo	-5	1.432818	-3.49	0.003	-8.010238 -1.989762
_cons	8	1.058206	7.56	0.000	5.776792 10.22321

Este modelo de regresión genera la siguiente ecuación para la mediana de EVA:

$$EVA_{0,5} = 8 - 5 * \text{sexo}$$

Si la persona es hombre entonces sexo = 0 y la ecuación expresa que la mediana de EVA es 8 y si la persona es mujer entonces sexo = 1 y la ecuación expresa que la mediana de EVA es 8-5=3, como el p-value asociado a sexo es p = 0,003 (en negritas), la diferencia de la mediana de EVA por sexo es significativa.

Si se ajusta por edad, la estimación es:

```
. qreg eva sexo edad, nolog
Median regression          Number of obs = 20
Raw sum of deviations      47 (about 5)
Min sum of deviations      21 Pseudo R2 = 0.5532
```

Eva	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sexo	-5	1.670378	-2.99	0.008	-8.524189 -1.475811
edad	-7.21e-19	.0526888	-0.00	1.000	-.1111637 .1111637
_cons	8	3.688383	2.17	0.045	.2181929 15.78181

Como la variable edad no es significativa, la conclusión es que la mediana de EVA solamente es explicada por el género del sujeto.