

Comentarios de Bioestadística

La distribución Gamma para comparar valores hormonales

Gabriel Cavada Ch.^{1,2}

¹Facultad de Medicina, Universidad de los Andes.

²División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

Gamma distribution to compare hormone levels

Como es sabido, la distribución de los valores hormonales, medidos en población normal o enferma, por lo general presentan distribuciones sesgadas hacia valores grandes, es decir, dichas distribuciones carecen de las propiedades morfológicas de la distribución normal o gaussiana que es un supuesto básico para comparar datos mediante el test t-Student, el test de análisis de la varianza (ANOVA) paramétrico, análisis de covarianza (ANCOVA) o simplemente establecer un modelo de regresión lineal donde la respuesta a explicar sea un valor hormonal. Tradicionalmente el problema de comparar grupos se resuelve mediante los test no paramétricos análogos a los paramétricos: el test de Wilcoxon, Mann-Whitney o Kruskal-Wallis. Cuando se toma el camino de los test no paramétricos además es necesario contar con un mayor tamaño de muestra, pues en general dichos test son menos potentes que los paramétricos. Ahora si el problema es explicar un valor hormonal a través de otras variables numéricas, el abanico de soluciones se restringe mucho más, pues la práctica de usos de regresiones percentilares, basadas en técnicas de remuestreo u otras que no asuman una distribución de probabilidades para la respuesta, son muy poco difundidas o usadas. Por ello se presenta la distribución Gamma como una alternativa paramétrica que debería considerarse en el análisis inferencial o de la evidencia cuando la respuesta clínica es un valor hormonal, esta alternativa era poco usada pues los cálculos involucrados son muy costosos, sin embargo, con las herramientas computacionales disponibles este argumento ya no se sustenta.

La distribución Gamma

La distribución Gamma está definida para variables aleatorias positivas (mayores que cero) y de naturaleza continua. La función densidad de probabilidades está dada por la siguiente expresión:

$$f(x) = \frac{\alpha}{\Gamma(r)} (\alpha \cdot x)^{r-1} e^{-\alpha x}$$

Donde α y r son parámetros positivos y Γ es la función matemática Gamma (homónima a la distribución) que está definida por la relación integral.

La Figura 1 muestra algunas formas distribucionales de la densidad de probabilidades Gamma.

Como se observa, las formas que ofrece la función de densidad de la Gamma son bastante similares a las distribuciones hormonales. Por ejemplo, la Figura 2 muestra los histogramas de AMH en dos grupos de sujetos (controles y casos).

La caracterización de la distribución Gamma respecto a su esperanza (valor promedio) y su varianza es:

$$E[X] = \frac{r}{\alpha} \quad \text{y} \quad Var[X] = \frac{r}{\alpha^2}$$

Es decir, su valor promedio es inversamente proporcional al parámetro α , y la varianza inversamente proporcional al cuadrado de este parámetro.

Tal como la regresión lineal clásica propone la relación:

$$E[Y|X] = \beta_0 + \beta_1 X$$

Donde Y sigue una distribución Normal con media dada por la esperanza anterior y varianza σ^2 . Ahora proponemos la relación:

$$E[Y|X] = \beta_0 + \beta_1 X$$

Pero asociada a la distribución Gamma. Esto origina un Modelo Lineal Generalizado, alternativa que los programas estadísticos potentes ofrecen.

Las estadísticas descriptivas de las AMH por casos y controles se muestran en la Tabla 1.

Comentarios de Bioestadística

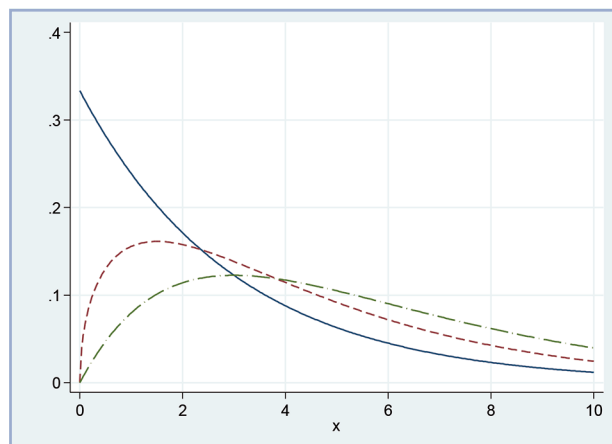


Figura 1. Funciones de densidad Gamma.

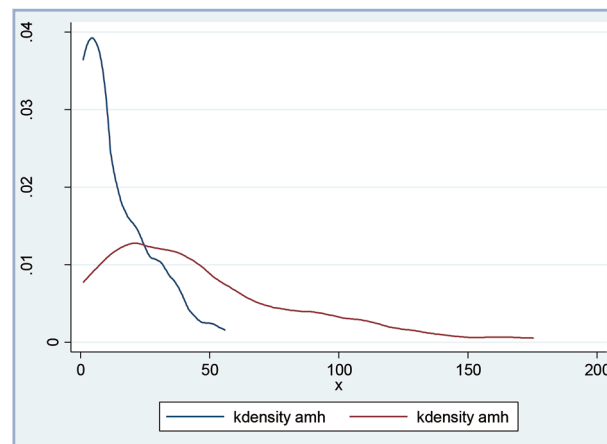


Figura 2. Distribución de AMH según casos y controles.

Tabla 1.

Grupo	n	min	p25	p50	p75	max	mean	sd
Controles	89	1	1,7	7,42	19	56,06	12,86303	13,41851
Casos	146	1	16,96	36,455	62,06	175,42	45,91517	38,41261
Total	235	1	6	24,02	47,23	175,42	33,39755	35,21617

Basta observar la diferencia entre mediana y media por grupo, para concluir que el test t-Student no es lo adecuado para comparar estos grupos, en subsidio deberíamos usar la alternativa no paramétrica o bien comparar los parámetros de la distribución Gamma respectiva, la ventaja de este último procedimiento es que podemos cuantificar la magnitud de la diferencia de la respuesta o agregar una covariable continua. Por ejemplo:

Generalized linear models	No. of obs	=	235
Optimization : ML	Residual df	=	233
	Scale parameter	=	.8465652
Deviance = 272.1458264	(1/df) Deviance	=	1.168008
Pearson = 197.2496918	(1/df) Pearson	=	.8465652
Variance function : $V(u) = u^2$	[Gamma]		
Link function : $g(u) = 1/u$	[Reciprocal]		
	AIC	=	8.706808
Log likelihood = -1021.049982	BIC	=	-999.9376

amh	Coef.	OIM Std. Err.	z	P > z	[95% Conf. Interval]
* caso	-.0559629	.00776	-7.21	0.000	-.0711721 - .0407536
**_cons	.0777422	.0075807	10.26	0.000	.0628842 .0926001

Interpretación de estos resultados

El reglón ennegrecido dice que se está modelando el inverso del promedio, que corresponde al parámetro α de la distribución Gamma.

En el reglón ** se expresa que el promedio en los controles es $1/0,0777422 = 12,863$ (ver en la Tabla 1 el promedio en los controles) y en el reglón * se muestra la diferencia o efecto de los casos respecto de los controles, o dicho de otra forma, el promedio en los casos es $1/(0,0777422-0,0559629) = 45,915$

Comentarios de Bioestadística

(ver en Tabla 1 el promedio para los casos); esta diferencia es significativa como lo muestra el p-value en negritas y cursiva que se muestra en el reglón *. Además es posible saber la magnitud del error estándar de la diferencia y con ello su intervalo de confianza, estimaciones que nos están vedadas si es que

hubiéramos preferido una alternativa no paramétrica.

En conclusión, usar el modelo lineal generalizado basado en la distribución Gamma, es una alternativa que se debiese considerar para analizar asociaciones de niveles hormonales con otras variables.