



## Un modelo bayesiano para datos cualitativos en simulación

*A bayesian model for qualitative data in simulation*

Jimmie Leppink\*

### Palabras clave:

Simulación clínica, cualitativo, cuantitativo, métodos mixtos, metaanálisis, metasíntesis.

### Keywords:

*Clinical simulation, qualitative, quantitative, mixed methods, meta-analysis, meta-synthesis.*

### RESUMEN

Tanto los datos cuantitativos como los cualitativos son importantes en la investigación en simulación. Sin embargo, se suele tratar 'cuantitativo' y 'cualitativo' como dos tipos de datos distintos con el argumento de que no se puede cuantificar información cualitativa. Esta tendencia dificulta la investigación de métodos mixtos y el metaanálisis o la metasíntesis de series de estudios. En este contexto, este artículo presenta un modelo bayesiano que ha sido utilizado para datos cuantitativos y que también puede servir para cuantificar información cualitativa. A pesar de que el ejemplo utilizado en este artículo es de un participante ( $N = 1$ ) que da 12 respuestas en cada una de dos sesiones de simulación, este modelo también se puede utilizar en contextos donde hay un interés en comparaciones entre grupos de participantes.

### ABSTRACT

*Both quantitative and qualitative data are important in simulation. However, 'quantitative' and 'qualitative' are usually treated as two different types of data with the argument that qualitative information cannot be quantified. This tendency hinders mixed-methods research and meta-analysis or meta-synthesis of series of studies. Therefore, this article presents a Bayesian model that has already been used for quantitative data and can also help to quantify qualitative information. Although the example used in this article has one participant ( $N = 1$ ) who provides twelve responses in each of two simulation sessions, this model can also be used in contexts where there is an interest in comparisons between groups of participants.*

## INTRODUCCIÓN

Las sesiones de simulación pueden darnos información cuantitativa y cualitativa sobre el comportamiento, el rendimiento o el estado emocional del participante. Aunque se suele tratar 'cuantitativo' y 'cualitativo' como dos tipos de datos distintos con el argumento de que no se puede cuantificar información cualitativa, la investigación de métodos mixtos y también el metaanálisis o la metasíntesis de series de estudios requieren algún tipo de integración de la información. Por lo tanto, este artículo presenta un modelo bayesiano para cuantificar información cualitativa. El modelo que se plantea se llama *Percentage of All Non-overlapping Data-Bayesian (PAND-B)* y ya se ha utilizado para datos cuantitativos<sup>1</sup> y datos ordinales.<sup>2</sup> Este modelo puede ayudar a educadores e investigadores a detectar transiciones en comportamiento, estado emocio-

nal u otros constructos al nivel del participante individual y a estudiar diferencias entre grupos de participantes.

## MATERIAL Y MÉTODOS

El estudiante X en departamento A tiene dos sesiones de simulación enfocadas en la habilidad de detectar tumores en radiografías pulmonares de seres humanos, con tecnología de realidad virtual. La primera sesión presenta 12 radiografías y en cada radiografía el estudiante puede elegir entre cuatro modos de estudiar: con sonido y texto (una voz explica la radiografía y hay texto en el área del tumor), con sonido (con voz, sin texto en el área del tumor), con texto en el área del tumor (sin voz), y sin nada (ni voz ni texto). La segunda sesión, que ocurre tres días después de la primera sesión, presenta 12 radiografías que en dificultad están muy parecidas a las radiografías

\* Hospital virtual Valdecilla.

Recibido: 19/08/2021  
Aceptado: 01/11/2021

doi: 10.35366/103188

**Citar como:** Leppink J. Un modelo bayesiano para datos cualitativos en simulación. Rev Latinoam Simul Clin. 2021; 3 (3): 117-119. <https://dx.doi.org/10.35366/103188>



de la primera sesión, y en cada radiografía el estudiante puede elegir entre los mismos cuatro modos de estudiar.

Entonces, el estudiante X dará un total de 24 respuestas en esta variable de cuatro categorías, porque el modo puede variar entre radiografías. Esta es una variable cualitativa del nivel de medida nominal. No es una variable cuantitativa, porque no se pueden cuantificar los cuatro modos de una manera que tenga sentido. Tampoco es una variable ordinal, porque tampoco hay un orden natural entre las categorías. Aunque se puede entender que la combinación de sonido y texto da más ayuda al estudiante que el modo que no ofrece nada, el orden entre el modo de sonido (sin texto en el área del tumor) y el modo de texto (sin sonido) no está claro, y es posible que la respuesta a la última pregunta depende de la radiografía (unas veces el sonido ayuda más, mientras que en otras el texto en el área del tumor lo hace).

**Tabla 1: Los modos elegidos (modo) por estudiante X en cada caso y en cada sesión, y codificaciones para probar las dos hipótesis (cH1 y cH2).**

Sesión	Caso	Modo	cH1	cH2
1	1	ST	1	1
1	2	ST	1	1
1	3	ST	1	1
1	4	ST	1	1
1	5	S	0	1
1	6	ST	1	1
1	7	ST	1	1
1	8	T	0	1
1	9	ST	1	1
1	10	ST	1	1
1	11	ST	1	1
1	12	ST	1	1
2	13	ST	0	0
2	14	S	1	0
2	15	N	1	1
2	16	T	1	0
2	17	N	1	1
2	18	T	1	0
2	19	N	1	1
2	20	T	1	0
2	21	N	1	1
2	22	T	1	0
2	23	N	1	1
2	24	T	1	0

ST = sonido y texto, S = sonido, T = texto, N = nada.

Los educadores del departamento A tienen los siguientes objetivos de aprendizaje para el estudiante, desde una sesión a otra:

1. Menos uso del primer modo (voz y texto) en la segunda sesión.
2. Un incremento en el uso del cuarto modo (ni voz ni texto) en la segunda sesión.

Los educadores esperan que la primera sesión sirva más para procesar nueva información y que la segunda sesión sirva más para ponerse a prueba. Para probar estas dos hipótesis, no se pueden utilizar modelos que tratan la variable de los modos como 'cuantitativo' u 'ordinal', sino que se necesita un modelo que respete el carácter nominal de los modos. Un modelo sencillo que ya se ha utilizado para datos cuantitativos<sup>1</sup> y datos ordinales,<sup>2</sup> y que también puede servir para variables nominales, es un modelo bayesiano que utiliza el porcentaje de todos los datos del residente de diferentes fases que no se solapan, en inglés: *Percentage of All Non-overlapping Data-Bayesian (PAND-B)*.<sup>1,2</sup> Es un modelo binomial bayesiano que se puede utilizar con el programa gratuito y *Open Source* JASP.<sup>3</sup>

## RESULTADOS

La *Tabla 1* presenta los modos elegidos por el estudiante y dos columnas más: 1. Una codificación para probar la primera hipótesis (menos uso del primer modo en la segunda sesión), y 2. Una codificación para probar la segunda hipótesis (más uso del cuarto modo en la segunda sesión).

Para probar la primera hipótesis, codificamos en la primera sesión cada 'ST' como '1' y los otros modos como '0'; en la segunda sesión cada 'ST' como '0' y los otros modos como '1' (cH1 en la *Tabla 1*). Para probar la segunda hipótesis, codificamos en la primera sesión cada 'N' como '0' y los otros modos como '1'; en la segunda sesión cada 'N' como '1' y los otros modos como '0' (cH2 en la *Tabla 1*).

PAND-B trata las codificaciones en cH1 y cH2 como variables binomiales y aplica una corrección para el tamaño de la muestra para evitar resultados de (casi) '0' o '100' % basados en muestras muy pequeñas,<sup>1,2</sup> y produce un intervalo creíble de 95%, que es la contraparte bayesiana del intervalo de confianza del 95% que conocemos de la escuela frecuentista. La corrección viene en la forma de una distribución binomial *a priori* de (1, 1). En la estadística bayesiana, los datos y

la distribución *a priori* se combinan para obtener la distribución *a posteriori*, que es la distribución que nos da el intervalo creíble de 95%. En el caso de cH1, los dos son (21, 3), porque hay 21 veces código '1' y 3 veces código '0'. Por lo tanto:

$$\begin{aligned} \text{Datos} + a \text{ priori} &= a \text{ posteriori}, \\ (21, 3) + (1, 1) &= (22, 4). \end{aligned}$$

La distribución *a posteriori* es una distribución binomial de (22, 4), que es una distribución con un intervalo de confianza del 95% de [0.688; 0.955] y un mediano de 0.855. Este intervalo está totalmente por encima de 0.50, lo que indica una diferencia en la dirección anticipada en la primera hipótesis.

Para probar la segunda hipótesis, los datos son (17, 7) porque hay 17 veces código '1' y 7 veces código '0' y, por lo tanto:

$$\begin{aligned} \text{Datos} + a \text{ priori} &= a \text{ posteriori}, \\ (17, 7) + (1, 1) &= (18, 8). \end{aligned}$$

La distribución *a posteriori* es una distribución binomial de (18, 8), que es una distribución con un intervalo creíble del 95% de [0.506; 0.851] y un mediano de 0.697. Este intervalo está totalmente por encima de 0.50, que indica una diferencia en la dirección anticipada en la segunda hipótesis.

## DISCUSIÓN

Aunque en el contexto de datos cuantitativos o de datos ordinales, siempre hay un rango u orden entre valores o categorías, este rango u orden no existe en el mundo de datos nominales. Sin embargo, PAND-B es útil para todo tipo de datos, porque el concepto central de este modelo –el principio de '*non-overlap*'– sí existe en todo tipo de datos.

Además, aunque el ejemplo en este artículo utiliza cuatro categorías relativamente 'fáciles' para no complicar la introducción de PAND-B en el contexto de variables nominales (variables cualitativas no ordinales), la codificación de datos cualitativos en un número de temas más reducido es una actividad clave en la investigación cualitativa, y tanto transiciones al nivel individual como diferencias entre (grupos de) individuales suelen ser de interés. PAND-B facilita estos estudios de transiciones y diferencias con un modelo sencillo que ayuda a cuantificar datos cualitativos, tanto en estudios individuales donde hay datos cualitativos o una combinación de datos cuantitativos y cualitativos como en estudios que combinan estos tipos de estudios individuales en *metaanálisis* o *metasíntesis*. Esto no es decir que codificar información cualitativa en temas cuantificables es un trabajo 'fácil' o que PAND-B, o cualquier modelo estadístico, ofrece una solución para todo tipo de información cualitativa, pero PAND-B da una vía para cerrar la brecha entre 'cualitativo' y 'cuantitativo', dos tipos de datos que desafortunadamente se suelen tratar como mundos distintos o culturas no unibles.

## REFERENCIAS

1. Leppink J. Small numbers are an opportunity, not a problem. *Sci Med*. 2021; 31(1): e40128. Available in: <https://doi.org/10.15448/1980-6108.2021.1.40128>.
2. Leppink J. Diseños de caso único en la simulación. *Rev Lat Am Sim Clin*. 2021; 3 (2): 80-82. doi: 10.35366/101432.
3. Love J, Selker R, Marsman M, et al. JASP Team (2021). JASP (Version 0.16) [Computer software]. <https://jasp-stats.org>

### Correspondencia:

**Dr. Jimmie Leppink**

Hospital virtual Valdecilla  
Avenida de Valdecilla, s/n, 39008,  
Santander, España.

**E-mail:** [jleppink@hvaldecilla.es](mailto:jleppink@hvaldecilla.es)