

Omnia Año 27, No.1 (enero-junio, 2021) pp. 27-46
Universidad del Zulia. ISSN: 1315-8856
Depósito legal pp 199502ZU2628

Simulación computacional de datos y aplicaciones en estadística

*Claudio Segundo Hurtado Petit** y *Víctor Riveros***

Resumen

El objetivo del trabajo es generar datos por simulación computacional usando una estructura determinada por un modelo preestablecido. La investigación es descriptiva y explicativa, se describen las características del modelo propuesto para la simulación de datos y se explica la implementación de la simulación. El diseño de la investigación es experimental dada la naturaleza de la simulación de datos como metodología de “experimentación numérica”. Para ejecutar la simulación se seleccionó una de las dimensiones del instrumento SCL-90-R, instrumento utilizado para evaluar y describir diversos síntomas psicopatológicos. Se manejó la metodología simulación Monte Carlo, y el programa R; como resultado se obtuvo una matriz de datos correspondientes a la dimensión PAR del instrumento SCL-90-R. Finalmente se comparan histogramas de frecuencia de los datos obtenidos por la simulación, con aquellos obtenidos por investigaciones empíricas hechas usando el mencionado instrumento.

Palabras clave: Modelo, Simulación, Computación, Datos, Estadística.

* Lcdo en Educación, Mención Matemática y Física (1989); Magister Scientiarum en Matemática Aplicada (1994); Profesor Titular a Dedicación Exclusiva e Investigador activo de la Universidad del Zulia adscrito al Departamento de Matemática y Física de la Universidad del Zulia. E-mail: hurtadoclaudio6@gmail.com.

** Lcdo en Educación Mención Cs. Matemáticas. MSc. en Matemáticas Aplicadas. Dr. En Ciencias Humanas, Postdoctorado en Ciencias Humanas. Profesor adscrito al Dpto de Matemática y Física de la FHE. Universidad del Zulia. E-mail: Vriveros75@gmail.com.

Computational simulation of data and applications in statistics

Abstract

The objective of the work is to generate data by computational simulation using a structure determined by a pre-established model. The research is descriptive and explanatory, the characteristics of the proposed model for data simulation are described and the implementation of the simulation is explained. The research design is experimental given the nature of data simulation as a “numerical experimentation” methodology. To run the simulation, one of the dimensions of the SCL-90-R instrument was selected, instrument used to assess and describe various psychopathological symptoms. The Monte Carlo simulation methodology was used, and the R program; a data matrix corresponding to the PAR dimension of the SCL-90-R instrument was obtained. Finally, frequency histograms of the data obtained by the simulation are compared with those obtained by empirical investigations made using the aforementioned instrument.

Key words: Model, Simulation, Computing, Data, Statistics

Introducción

Las simulaciones computacionales en las últimas décadas han venido aplicándose no sólo en las ciencias formales sino que ya ocupa espacios importantes en ámbitos de la ciencia sociales; ésta metodología ha permitido llevar a cabo experimentación previa antes de ejecutar intervenciones directas en la realidad empíricas. Permite a los investigadores construir modelos de la realidad y trabajar sobre ellos para explorar el comportamiento de éstos a medida que se controlan los datos de entrada al modelo y se analizan los datos de salida.

Ejecutar una simulación computacional implica el diseño un modelo de la situación, fenómeno o concepto a simular; el investigador debe tener conocimiento sobre el objeto a simular con la finalidad de que el modelo sea una buena aproximación de esa realidad. En ésta investigación se conduce una simulación para generar datos por simulación computacional de un modelo preestablecido. El modelo seleccionado ha sido una de las dimen-

siones de la estructura factorial recogida en el instrumento SCL-90-R; el cual es un instrumento utilizado para evaluar y describir diversos síntomas psicopatológicos. La dimensión escogida de la estructura factorial es la dimensión PAR (que evalúa comportamientos paranoides: pensamiento proyectivo, suspicacia, temor a la pérdida de autonomía). Como referenciado del SCL-90-R se escogió la aplicación referida por un grupo de investigadores (Sánchez et al, 2020), en el marco de la pandemia por COVID-19 en el Gran Mendoza extensión de la capital de Mendoza en Argentina.

En general, la necesidad de tener datos a disposición en una investigación es fundamental ya que el investigador los necesita para validar y certificar los resultados de su investigación; en éste sentido se puede: a) generar datos desde la teoría; b) generar teoría desde los datos. En ambos casos, los cuales son enfoques epistemológicos bien diferenciados (enfoque deductivo, enfoque inductivo), es evidente la interrelación datos y teoría.

Con el desarrollo de la computación y la disponibilidad de hardware y software, la actividad de los científicos se ha reorientado debido a la inclusión de éstas herramientas informáticas y computacionales en la solución de problemas; las simulaciones computacionales ha permitido el desarrollo de nuevo conocimiento científico en diversas áreas de las ciencias formales, naturales y sociales; constituyéndose así en un nuevo método para hacer ciencia (Seoane, Arriasecq y Greca, 2015).

Objetivos de la Investigación

Objetivo general: Generar datos por simulación computacional con estructura determinada por un modelo preestablecido.

Objetivos específicos

- Describir características de un modelo computacional.
- Caracterizar las diferentes etapas de la simulación computacional de datos.
- Simular computacionalmente datos con base al modelo seleccionado.
- Explicar los alcances de la modelación y simulación computacional de datos.

La simulación de datos, permite la realización de “experimentaciones” que en la realidad empírica social en muchas ocasiones no es factible ejecutar por razones prácticas, éticas, financieras, entre otras.

Una investigación de éste tipo es factible siempre y cuando se cuente con una teoría de sustento para el modelo a simular, así como un mínimo conocimiento sobre programación, contar con software y hardware que permita ejecutar simulaciones.

La presente investigación, en términos de lo anteriormente expresado, aporta en su desarrollo elementos teóricos y prácticos de soporte teórico-metodológico a investigaciones donde la simulación de datos sea el objetivo, así como también pretende explicar el abordaje de la simulación de datos partiendo de un modelo muy sencillo en su construcción con la finalidad de presentar de manera pedagógica, métodos de modelaje y simulación computacional de datos.

Referentes teóricos

Desde mediados de los años 1.990 se ha extendido la aplicación de métodos computacionales para el estudio en ciencias sociales (Rodríguez y Roggero, 2014); por esta razón los métodos computacionales ya no son exclusivos para ciencias como la física y la matemática.

La simulación computacional de datos tiene muchas aplicaciones sobretodo en el estudio de modelos en ecuaciones estructurales, se puede hacer mención a trabajos como el realizado por Jackson (2009); en este trabajo mediante una simulación Monte Carlo se generaron datos a través de un modelo con distribución normal multivariada con uno, dos, tres, cuatro o cinco factores y con cargas factoriales poblacionales de 0,80 o 0,90. En los casos con modelos con más de un factor, los datos se generaron con correlaciones poblacionales de 0,50; los datos se obtuvieron usando el generador de números aleatorios incorporado en el programa SAS que produce una distribución normal con media cero y desviación estándar de uno; con esa información se generaron 200 replicaciones (matrices de covarianza) las cuales sirvieron para simular varios escenarios aplicados a determinar tamaños de muestras adecuados en modelos de ecuaciones estructurales para ayudar a determinar el número de parámetros a estimar en un contexto de análisis factorial confirmatorio con el método de máxima verosimilitud.

Otro antecedente sobre la temática a referenciar es la investigación llevada a cabo por (Carballo, et al. 2016); en la investigación se trata de establecer el tamaño de muestra suficiente para el ajuste de parámetros implícitos en la estructura de modelos multivariados basados en relaciones de causalidad; se implementó una simulación Monte Carlo para simular el tamaño de muestra suficiente para el modelaje de las relaciones de causalidad estructural con datos categóricos, para los indicadores de calidad de vida en la comunidad del conjunto residencial. Se construyó un modelo teórico estructural con método mínimos cuadrados generalizados (GLS) y máxima verosimilitud (ML), las simulaciones se ejecutaron bajo el supuesto de escala continua y con multinormalidad. Se hizo la comparación entre métodos no paramétricos de distribución libre asintótica, métodos robustos paramétricos como los ya mencionados GLS-ML, en ambos casos se generaron matrices de covarianzas, y la simulación bajo los supuestos ya mencionados permitió un buen procedimiento muestral indicando que 150 casos es suficiente para el estudio de los indicadores de calidad de vida en el complejo habitacional estudiado.

Modelo: Un término de importancia en ésta investigación es el concepto de modelo, debido a las variadas tipologías o conceptualizaciones presentes en la literatura y porque se quiere fijar el marco conceptual que es utilizado.

Según (Armatte, 2006), el concepto de modelo se origina en el campo de la Física y la Lógica-Matemática, en sus palabras afirma al respecto que: “Un modelo es una construcción racional siempre abierta, frágil y transitoria que se ve desafiada por lo real”.

Por otro lado, para (Rodríguez y Roggero, 2014), la conceptualización de un modelo parte de la relación sujeto/observador con un objeto; es una relación triádica compuesta por un sujeto S, el modelo M y el objeto O. En la representación M aparece como mediador epistémico de la relación S-O; es así, como S puede aprender algo de O a través de M. El modelo se presenta como un instrumento que aporta conocimiento del objeto. En ésta investigación se entiende como modelo a la representación de un objeto real o teórico mediante un sistema formal, que permite estudiarlo en sus partes, en su funcionamiento y con la posibilidad de actuar sobre él para hacerlo evolucionar de manera estacionaria.

Es así como un modelo M es un conjunto de variables relacionadas a través de ecuaciones matemáticas y sujetas a restricciones. Se identifican en la literatura dos tipos de modelos:

- Modelos deterministas: aquellos que establecen el valor de una variable cualquiera conocidas las otras.

- Modelo probabilístico (con componente aleatorio): en el modelo se incluyen variables para las cuales no es posible determinar qué valores va a tomar.

La presente investigación se fundamenta en modelos probabilísticos.

Simulación Estadístico computacional: Simular significa imitar, representar algo; en otras palabras, una simulación duplica características y comportamientos de un sistema real. Con la simulación se imita el comportamiento del sistema real (O) a través de un modelo (M) construido como representación de esa realidad.

Desde el punto de vista histórico los Métodos de Monte Carlo se declaran como el origen de la simulación computacional (Durán, 2015); son métodos que combinan conceptos estadísticos (muestreo aleatorio) con la capacidad que tienen las computadoras para generar números pseudo aleatorios y automatizar cálculos; se utilizó en sus inicios para resolver problemas matemáticos mediante la simulación de variables aleatorias.

Las simulaciones Método Monte Carlo son experimentos numéricos realizados sobre los modelos (M), utilizando un algoritmo. La simulación genera datos provenientes de reglas especificadas por el modelo (M), y los datos obtenidos se analizan de manera inductiva. El nivel de representación que tienen los modelos utilizados para la simulación estadístico computacional, permiten al investigador poner a prueba hipótesis, estimar valores de parámetros y estudiar el comportamiento del modelo M y su relación con el objeto O.

En conclusión, una simulación estadístico computacional es un programa computacional que intenta representar el mundo real o teoría, basado en un modelo probabilístico de orden estadístico-matemático; en tal sentido, la precisión de la simulación depende de la precisión con la cual se ha construido el modelo estadístico-matemático.

Simulación Monte Carlo y generación de datos: el método Monte Carlo es una familia de métodos de simulación con los cuales se puede reproducir y estudiar sistemas empíricos de forma controlada. La base de los métodos Monte Carlo está en la generación de variables aleatorias. A cada

variable aleatoria generada, el investigador le asigna las propiedades relacionadas con: a) las características de la distribución (media, varianza, asimetría, curtosis, etc); b) relaciones entre ellas (correlación) y c) cantidad de error presente en la medida de esas variables. Esas propiedades definen la población de referencia de la cual se extraen cierto número de muestras con tamaño definido (que puede ser fijo o variable de acuerdo con los objetivos de la investigación); de esas muestras se calculan los estadísticos objeto de estudio. Al finalizar la simulación el investigador dispone de las distribuciones muestrales de los estadísticos requeridos en la investigación, con esos estadísticos puede el investigador deducir el comportamiento de éstos, en función de las propiedades definidas en la población.

En el método Monte Carlo, las propiedades de las distribuciones de las variables aleatorias se investigan mediante el uso de números aleatorios (Gentle, 1985). Por lo tanto, es importante tener a disposición un buen generador de “números aleatorios”. En éste sentido queda claro que los programas que generan variables aleatorias, se basan en algoritmos generadores de tipo determinístico y por lo tanto generan los llamados números pseudo-aleatorios. En la práctica un generador produce una secuencia de números aparentemente aleatorios, pero a un cierto punto tienden a repetirse. Es así que un buen generador es aquel que produce una secuencia de valores lo más larga posible antes de volver a repetirse la secuencia. En el ambiente del programa R, que es el utilizado en ésta investigación, se utiliza por defecto el método generador de Mersenne-Twister (Matsumoto y Nishimura, 1998), que garantiza una secuencia de $2^{19937} - 1$ elementos antes volver a repetir la secuencia.

En general para llevar a cabo una simulación con método Monte Carlo se toman en cuenta los siguientes 9 pasos propuestos por Paxton, Curran, Bollen, Kirby y Chen, (2001); a continuación:

- 1) Desarrollo de la pregunta de investigación;
- 2) Creación de un modelo válido;
- 3) Diseño de condiciones experimentales específicas;
- 4) Elección de valores para los parámetros del modelo;
- 5) Seleccionar el software apropiado para ejecutar la simulación;
- 6) Ejecutar la simulación;

- 7) Almacenar archivos;
- 8) Resolución de los problemas planteados y verificación;
- 9) Presentar los resultados definitivos.

Software para ejecutar la simulación: Con la finalidad de llevar a cabo la simulación de datos en la presente investigación se maneja el software R, software de código libre. El programa R es un entorno computacional y gráfico con exclusiva orientación al análisis estadístico de datos, las versiones del programa son continuamente actualizadas y desarrolladas por colaboradores en todo el mundo mediante el grupo de trabajo R Development Core Team. En la actualidad R se ha convertido en una importantísima herramienta para en análisis científico de datos (data science). Para descargar, instalar y tener a disposición materiales e información sobre programa se recomienda la página: www.r-project.org.

Metodología

Se presenta a continuación la metodología llevada a cabo para producir los datos por simulación con estructura determinada por un modelo preestablecido. Se detallan los procedimientos para alcanzar los objetivos específicos desde la presentación del modelo a simular, las secuencias en programación R hasta la obtención de los datos.

Tipo de investigación: Según Hernández, Baptista y Fernández (2014) la investigación es cuantitativa con alcance descriptivo y explicativo. Descriptivo porque se describen las características del modelo propuesto para ilustrar el proceso de simulación de datos así con también las variables que los componen, caracterizando cada etapa de la simulación. Se describe el modelo del instrumento SCL-90-R. La investigación explicativa porque se centra en explicar el funcionamiento de las simulaciones para generar los datos.

Diseño de la investigación: Tomando en cuenta el objetivo general de la investigación “Producir datos por simulación computacional con estructura determinada por un modelo preestablecido” y según Hernández, Baptista y Fernández (2014), el diseño de la investigación es experimental debido a que la simulación de datos es considerada como un “experimento

numérico” ya que el investigador puede manipular de manera artificial variables en el modelo propuesto para analizar resultados obtenidos.

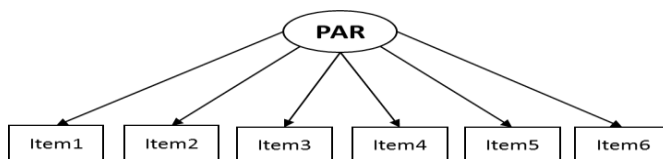
Selección del modelo a simular: Como ejemplo ilustrativo para la simulación de datos en ésta investigación se tomó como referencia el instrumento SCL-90-R, instrumento utilizado para evaluar y describir diversos síntomas psicopatológicos.

En la presente investigación se simulan los datos relacionados con la dimensión PAR del instrumento SCL-90-R referido por un grupo de investigadores (Sánchez, Vela, Neila, Ragone, Quiero, Mesa, Lucero, Rossi, Maschi, y Belamate, 2020) en el marco de la pandemia por COVID-19 en el Gran Mendoza extensión de la capital de Mendoza en Argentina. El instrumento SCL-90-R utilizado por los investigadores Argentinos está integrado por 90 reactivos (ítems) y se compone de 9 dimensiones y 3 índices generales. Los reactivos individuales están operacionalizados con una escala que mide el grado de malestar o molestia que causa el síntoma del 0 al 4, donde 0 corresponde a “nada”, 1 corresponde a “Muy poco”, 2 corresponde a “Poco”, 3 corresponde a “Bastante” y 4 a “Mucho”.

En el trabajo de (Sánchez et al, 2020), describen cada una de las dimensiones del instrumento: SOM (somatización); OBS (Obsesiones y compulsiones); SI (Sensibilidad interpersonal); DEP (Depresión): incluye la mayoría de los principales síntomas típicos de los trastornos; ANS (Ansiedad); HOS (Hospitalidad); FOB (Ansiedad fóbica); PAR (Ideación Paranoide); PSIC (Psicoticismo); IA (Ítems adicionales): 7 ítems que no forman parte de ninguna dimensión pero si de los índices generales. Se relacionan con problemas de sueño, problemas con la comida, culpa y pensamiento de muerte.

Modelo representativo de la dimensión PAR y los ítems que le componen: De acuerdo con la información reportada por (Sánchez, et al; 2020), el modelo de medida de la dimensión PAR es un unidimensional y se puede representar como pathdiagram en la figura 1.

Figura 1. Modelo para la dimensión PAR



Fuente: Hurtado (2020).

En la figura 1, se muestra cómo la dimensión PAR refleja su manifestación en los 6 ítems diseñados para medirla o evaluarla. En (1) se presenta modelo de medida relacionado con la figura 1:

$$X = \Lambda_x PAR + \varepsilon \quad (1)$$

En términos estadísticos y psicométricos, la ecuación (1) es un modelo de medida que establece la relación entre la dimensión (Variable Latente) PAR y las variables observables X (ítems) el cual es un vector de dimensiones 6×1 ; Λ_x es una matriz de dimensiones $6 \times n$, donde n es el tamaño de las muestra y contiene las cargas factoriales que relacionan los ítems con la dimensión PAR; por otro lado ε es otro vector de dimensiones 6×1 que contiene los errores de medida de cada uno de los ítems que miden la dimensión PAR.

Diseño de condiciones experimentales específicas para la simulación del modelo: Para establecer las condiciones experimentales específicas de la simulación de datos para la dimensión PAR, se asume:

- 1) Para cada ítem (variable observable) de la dimensión PAR se asume una distribución Normal.
- 2) De acuerdo con la información obtenida de la publicación hecha por Robles, Andreu y Peña, (2002), se presentan los promedios y desviaciones estándar de cada uno de los ítems:

Tabla 1. Estadísticos descriptivos de los ítems que componen la dimensión PAR

Ítems	N	Media	Desv. típ.
8	460	,35	,75
18	460	,73	,87
43	460	,52	,88
68	460	,60	,95
76	459	,49	,84
83	458	,45	,76

Fuente: Robles, Andreu y Peña (2002).

3) Arbitrariamente se selecciona como tamaño de muestra N=700. A continuación, en la tabla2, se presentan todas las propiedades de las variables que necesitan ser simuladas.

Tabla. 2. Propiedades de las variables

Dimensión	Variable	Rango	Distribución
PAR	Ítem 1	{0,1, 2, 3, 4}	N($\alpha=0,35$; $\beta=0,75$)
PAR	Ítem 2	{0,1, 2, 3, 4}	N($\alpha=0,73$; $\beta=0,87$)
PAR	Ítem 3	{0,1, 2, 3, 4}	N($\alpha=0,52$; $\beta=0,88$)
PAR	Ítem 4	{0,1, 2, 3, 4}	N($\alpha=0,60$; $\beta=0,95$)
PAR	Ítem 5	{0,1, 2, 3, 4}	N($\alpha=0,49$; $\beta=0,84$)
PAR	Ítem 6	{0,1, 2, 3, 4}	N($\alpha=0,45$; $\beta=0,76$)

Fuente: Hurtado (2020).

Elección de valores para los parámetros del modelo: Se toma el valor del parámetro coeficiente de consistencia interna entre los ítems asociados con la dimensión PAR; el coeficiente Alpha de Cronbach 0,75 obtenido de la publicación de (Sánchez, et al. 2.020).

Para realizar la simulación se necesita crear una matriz de correlaciones y para construirla se usa el coeficiente Alpha de Cronbach reportado anteriormente, con el fin de determinar la correlación promedio entre las variables (ítems).

Tabla. 3. Correlación promedio entre variables

Dimensión	Alpha de Cronbach	Número de ítems (k)	$\bar{r} = \frac{1}{\left(\frac{k}{\alpha}\right) - (k - 1)}$
PAR	0,75	6	0,33

Fuente: Hurtado (2020).

Selección del software apropiado para ejecutar la simulación: utilizó para la simulación el Software R.

Ejecución de la simulación: Para realizar la simulación se necesita crear una matriz de correlaciones, la cual se obtiene de la información obtenida de la tabla 3.

Tabla. 4. Matriz de correlaciones para la simulación

	Item1	Item2	Item3	Item4	Item5	Item6
Item1	1					
Item2	0,33	1				
Item3	0,33	0,33	1			
Item4	0,33	0,33	0,33	1		
Item5	0,33	0,33	0,33	0,33	1	
Item6	0,33	0,33	0,33	0,33	0,33	1

Fuente: Hurtado (2020).

Mediante la aplicación de funciones y librerías del programa R se elaboraran las líneas de código necesarias para ejecutar la simulación. Las variables que componen la dimensión PAR están correlacionadas y una forma experimental de establecer esa relación con éste tipo de variables es usando la función *mvrnorm* () de la librería MASS de R. Para ejecutar la simulación se necesita ejecutar las siguientes instrucciones y líneas de código:

- 1) La librería MASS y función *mvrnorm* (). La función *mvrnorm* () genera datos aleatorios para una distribución normal multivariada.

Se instalar la librería con el comando `install.packages("MASS")`; luego se carga la función MASS mediante el comando `library(MASS)`

- 2) Creación de la matriz poblacional de correlaciones.

Se crea la matriz de correlaciones (`matriz_correla`) usando la información contenida en la tabla 4

```
matriz_correla=matrix(c(1,0.33,0.33,0.33,0.33,0.33,0.33,1,0.33,0.33,0.33,0.33,0.33,0.33,1,0.33,0.33,0.33,0.33,0.33,0.33,1,0.33,0.33,0.33,0.33,0.33,1,0.33,0.33,0.33,0.33,0.33,1), nrow=6)
```

- 3) Se crea una matriz que contiene la media de todas las variables que miden la dimensión PAR. Para ello se usa la tabla 2:

```
Medias=c(0.35,0.73,0.52,0.60,0.49,0.45)
```

- 4) Creación de una semilla aleatoria.

Se usa `unaseed` (semilla) aleatoria. Conel comando `set.seed(3240692)`.

5) Se simulan los datos correlacionados, usando la función:

```
simula_correla_data=mvrnorm(n=700, mu= medias, Sigma=
matriz_correla)
```

6) Se obtiene una matriz de datos para la dimensión PAR, las primeras filas de la matriz puede ser visualizada usando el comando head()

Al usar la función head (simula_correla_data) se muestran las primeras filas de la matriz de datos:

A continuación se muestran las dos primeras filas generadas (tabla 5):

Tabla. 5. Primeras filas de la matriz de datos generada por simulación

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.75126	1.27504	1.71804	0.65156	0.41884	0.24371
[2,]	-0.17110	-0.77265	0.65680	0.77688	0.20526	0.11662

Fuente: Hurtado (2020).

La matriz de datos tiene como encabezado la identificación de columnas [,i] y el identificador de filas [j,] que genera R. También los datos de la matriz se muestran con decimales, cuando deberían ser los códigos de respuestas para cada ítem de la dimensión PAR. El problema se resuelve estableciendo criterios o umbrales correspondientes a las categorías de respuesta.

7) Se genera un data frameo base de datos con encabezados que correspondan a los ítems.

```
simula_correla_data=as.data.frame (simula_correla_data)
```

```
names (simula_correla_data)=c("item1", "item2", "item3", "item4", "it
em5", "item6")
```

Se muestra las dos primeras filas de la base de datos (tabla 6):

Tabla. 6. Primeras filas de la base de datos generada por simulación

	Item1	Item2	Item3	Item4	Item5	Item6
1	0.75126	1.27504	1.71804	0.65156	0.41884	0.24371
2	-0.17110	-0.77265	0.65680	0.77688	0.20526	0.11662

Fuente: Hurtado (2020).

8) Para resolver el problema de datos con valores decimales se transforman todas las variables a sus categorías reales y también se asegura que se cumplan los valores mínimo y máximo de éstas.

Se aplica la función `mutate()` que permite transformar las variables, la cual se encuentra en el paquete “`dplyr`”. Luego se carga la librería del paquete con el comando: `library(dplyr)`

A continuación se presentan las líneas de comando para lograr las transformaciones en las variables de la base de datos generada:

```

simula_correla_data=mutate(simula_correla_data,
item1=round(item1),
item2=round(item2),
item3=round(item3),
item4=round(item4),
item5=round(item5),
item6=round(item6),
item1=ifelse(item1<0,0,ifelse(item1>4,4,item1)),
item2=ifelse(item2<0,0,ifelse(item2>4,4,item2)),
item3=ifelse(item3<0,0,ifelse(item3>4,4,item3)),
item4=ifelse(item4<0,0,ifelse(item4>4,4,item4)),
item5=ifelse(item5<0,0,ifelse(item5>4,4,item5)),
item6=ifelse(item6<0,0,ifelse(item6>4,4,item6)),

```

Se muestra las dos primeras filas de la matriz de datos transformada (tabla 7):

Tabla. 7. Primeras filas de la base de datos transformada

	Item1	Item2	Item3	Item4	Item5	Item6
1	1	1	2	1	0	0
2	0	0	1	1	0	0

Fuente: Hurtado (2020).

Finalmente, se obtiene una matriz de datos que contiene las posibles respuestas simuladas de 700 sujetos. También se utilizó la hoja de cálculo Excel para recuperar los datos simulados en R a una base de datos en formato Excel y el programa estadístico SPSS (versión 21) para procesar la distribución de frecuencias de los ítems de la dimensión PAR.

Resultados

Se generó el modelo de la dimensión PAR del instrumento SCL-90-R, la cual consta de 6 ítems; el modelo factorial se muestra en la figura 1 también se obtuvo la ecuación estadístico matemático, mediante la ecuación matricial (1). También se describieron en el modelo las características psicométricas de la dimensión PAR, usando para ello información reportada en otras investigaciones.

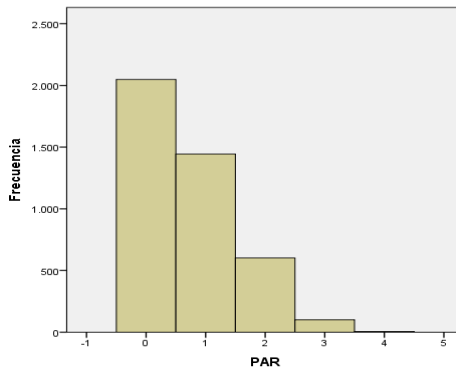
Otro resultado obtenido se relaciona con la caracterización y descripción teórico-práctica de cada etapa de una simulación computacional; se generaron las líneas de código en R para simular los datos relacionados con el modelo seleccionado.

Mediante la simulación computacional se obtuvo una muestra de 700 datos, es decir, la respuesta simulada de 700 sujetos sobre los ítems que reflejan la incidencia de la dimensión PAR sobre 6 ítems; el tamaño de esa muestra puede modificarse, con la finalidad de tener una muestra mayor o menos según sean las necesidades en investigaciones futuras.

En la figura 2 se muestra el gráfico de la distribución de frecuencias de las categorías de la dimensión PAR con los datos simulados. Ahora bien, si se compara ésta distribución de frecuencias con aquellas obtenidas con datos reales en los trabajos de Sánchez et al. (2020), Sánchez y Ledesma (2009), se observa la similitud de las tres distribuciones, figuras 3 y 4 respectivamente. El rasgo importante es la asimetría que muestra la distribución de los datos simulados; los datos simulados conservan las propiedades psicométricas de las distribuciones de donde se obtuvieron las características para el

modelo establecido.

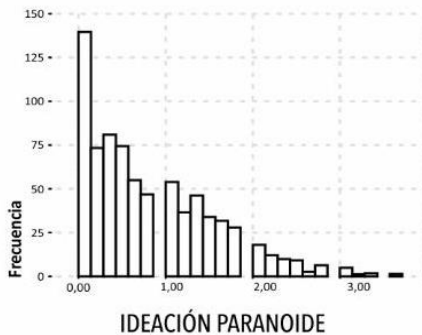
Figura 2. Distribución de frecuencias de la dimensión PAR por simulación



Fuente: Hurtado (2020).

A continuación se muestra la distribución de frecuencias de la dimensión **PAR** (Ideación Paranoide) obtenidas por Sánchez et al. (2020).

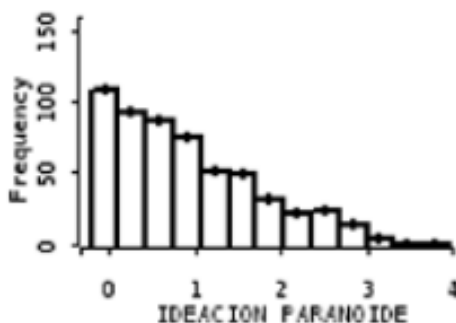
Figura 3. Distribución de frecuencias de la dimensión Ideación Paranoide



Fuente: Gráfico tomado de “Comportamiento psicométrico del inventario de síntomas revisado SCL-90 R. Administración remota durante confinamiento por Covid-19 en el Gran Mendoza”. Universidad de Congreso, Facultad de Ciencias de la Salud. [Documento en línea]. Disponible: <https://preprints.scielo.org/index.php/scielo/preprint/view/802>. [Consulta: 2020, junio 15].

A continuación se muestra la distribución de frecuencias de la dimensión PAR (Ideación Paranoide) obtenidas por Sánchez y Ledesma (2009).

Figura 4. Distribución de frecuencias de la dimensión Ideación Paranoide Sánchez y Ledesma



Fuente: Gráfico tomado de “Análisis psicométrico del inventario de síntomas revisado (SCL-90-R) en población clínica”. Revista argentina de clínica psicológica, Vol. XVIII, Número 3, pp. 265-274. Documento en línea]. Disponible: <https://www.redalyc.org/pdf/2819/281921775007.pdf>. [Consulta: 2020, junio 15].

Con éste informe se ha explicado el alcance que tiene el modelado y la simulación computacional de datos para disponer de datos sin tener que recurrir a su obtención empírica, con la finalidad de hacer experimentación sobre el modelo.

Conclusiones

En general producir datos simulados ayuda al investigador a manipular el modelo propuesto con la finalidad de conocer por aproximación el comportamiento que pudiera tener la realidad al variar una de las variables o relaciones que tienen las variables que configuran el modelo que sustenta una teoría estudiada, sin tener que intervenir directamente en la realidad empírica.

Con ésta investigación se llevó a la práctica la generación de datos por simulación computacional, y se verificó que los datos simulados conser-

van las propiedades que tienen los datos obtenidos de la realidad empírica. La aproximación de los datos simulados con los datos reales se logra a medida que el modelo establecido para la simulación refleja las características propuestas por la teoría.

En la investigación se han descrito las características principales del modelo seleccionado para la simulación, en éste caso la dimensión **PAR** del instrumento SCL-90-R; esas características tienen que ver con las distribuciones de las variables que conforman ese modelo y las correlaciones entre esas variables.

Se han caracterizado de manera clara y pedagógica las etapas que se deben llevar a cabo para ejecutar una simulación con el Método Monte Carlo.

Los datos producidos por la simulación conservan con mucha aproximación las características que tienen los datos obtenidos por recolección mediante la aplicación del instrumento SCL-90-R.

En éste documento se han explicado los alcances que tiene la simulación de datos para la investigación.

Queda abierta para otros investigadores continuar con ésta línea de investigación basada en la simulación de datos.

Referencias bibliográficas

- Armatte, Michel (2006). **“La noción de modelo en las ciencias sociales”**.EMPIRIA. Revista de metodología de ciencias sociales, 11, 33-70.
- Carballo, Nahir, Jaimes, Edgar, Pineda, Neida, Mendoza, José, Garcés, Yolimar y Rodríguez, Hilda (2016). **Estimación del tamaño muestral requerido en la modelización con indicadores de calidad de vida**. Universidad Nacional Experimental de los Llanos Ezequiel Zamora Y Universidad de los Andes. [Documento en línea]. Disponible: <http://www.postgradovipi.50webs.com/archivos/agrollania/2016/agro11.pdf>. [Consulta: 2.021, Agosto 20].
- Durán, Juan (2015). **“Nociones de simulación computacional: simulaciones y modelos científicos”**.Argumentos de la razón técnica, 18, 87-110.

- García, Pio (2015). **“Caracterización de una noción de simulación a partir de prácticas experimentales”**.Principia, 19 (2), 217-234.
- Gentle, James (1985). **Monte Carlo Methods**.The encyclopedia of statistical sciences.Ins S. Kotz y N. L. Johnson (Ed.). Wiley. New York. Vol. 5, pp. 612-617.
- Hernández, Roberto, Fernández, Carlos y Baptista, Pilar (2014). **Metodología de la Investigación**. 6ta Edición. McGraw-Hill/Interamericana Editores, S.A, DE C.V. México.
- Jackson, Dennis (2009). **“Sample size and number of parameters estimates in maximum likelihood in confirmatory factor analysis: A montecarlo investigation”**.Structural Equation Modelling: A Multidisciplinary Journal, 8(2), 205-223.
- Matsumoto, Makoto, Nishimura, Takuji (1998). **“Mersenne Twister: A 623-dimensionality equidistributed uniform pseudo-random number generator”**.ACM Transactions on modeling and computer simulation, 8, 3-30.
- Paxton, Pamela, Curran, Patrick, Bollen, Kenneth, Kirby, Jim y Chen, Feinian(2.001). **“Monte Carlo experiments: design and implementation”**.StructuralEquationModeling, 8(2), 287-312.
- Robles, José, Andreu, José, y Peña, María (2002). **CRL-90-R: Aplicación y análisis de sus propiedades psicométricas en una muestra de sujetos clínicos españoles**. Psicopatología clínica, legal y forense, Vol. 2, N° 1, pp.5-19. [Documento en línea]. Disponible: <https://dialnet.unirioja.es/descarga/articulo/2518060.pdf>. [Consulta: 2.020, Agosto 21].
- Rodriguez, Leonardo y Roggero, Pascal (2014). **La modelización y simulación computacional como metodología de investigación social**. [Documento en línea]. Disponible: <http://journals.openedition.org/polis/10568>. [Consulta: 2.019, noviembre 23].
- Sánchez, Jesús, Vela, María, Neila, Marcos, Ragone, Fabiola, Quiero, Alicia, Mesa, Zarina, Lucero, Yesica, Rossi, Denis, Maschi, Celia y Belamate, Carla (2020). **Comportamiento psicométrico del inventario de síntomas revisado SCL-90 R. Administración remota durante confinamiento por Covid-19 en el Gran Mendoza**. Universidad de Congreso, Facultad de Ciencias de la Salud. [Documento en línea].

Disponible:<https://preprints.scielo.org/index.php/scielo/preprint/view/-802>. [Consulta: 2020, junio 15].

- Sánchez, Roberto y Ledesma, Rubén (2009). **Análisis psicométrico del inventario de síntomas revisado (SCL-90-R) en población clínica.** Revista argentina de clínica psicológica, Vol. XVIII, Número 3, pp. 265-274. Documento en línea. Disponible: <https://www.redalyc.org/pdf/2819/281921775007.pdf>. [Consulta: 2020, junio 15].
- Seoane, María, Arriasecq, Irene y Greca, Ileana (2015). **Simulaciones computacionales: un análisis fenomenográfico.** Revista de la enseñanza de la física, Vol. 27, N° Extra, 289-296. [Documento en línea]. Disponible: <https://revistas.unc.edu.ar/index.php/revista-EF/article/view/12618>. [Consulta: 2.020, Julio 07].