

Tratamiento de datos cuantitativos en investigación científica

DOI: <https://doi.org/10.64121/rece.2026v4.bmn017>

Roberto Carlos Barriga Granados¹, Carlos Heriberto Mendoza Perez², Víctor Fernando Nieto del Valle³

Introducción

En la investigación científica, el análisis cuantitativo ocupa un lugar central al facilitar la obtención de conclusiones objetivas a partir de datos numéricos. Este enfoque, asociado al paradigma positivista, contrasta con el paradigma cualitativo al basarse en mediciones numéricas y procedimientos estadísticos para validar hipótesis. La claridad metodológica en el tratamiento de datos es esencial para garantizar la calidad y la reproducibilidad del estudio. Un proceso sistemático de codificación, organización y análisis numérico de los datos permite al investigador sintetizar la información de la muestra y, posteriormente, inferir resultados generales. Así, antes de aplicar cualquier prueba estadística, es necesario planificar con rigor cómo se recogerán los datos y cómo se prepararán para su análisis. Por ejemplo, la estadística descriptiva resume el comportamiento de los datos mediante tablas y gráficas, sin intentar extraer conclusiones más allá de la muestra estudiada, mientras que la estadística inferencial utiliza los datos muestrales para obtener generalizaciones sobre la población. En consecuencia, contar con una metodología clara desde la recolección hasta el análisis garantiza que las conclusiones sean válidas y sólidas, evitando sesgos y errores en la interpretación de los resultados.

Procesos de codificación y organización de datos

Una vez recolectados los datos (a través de cuestionarios, mediciones experimentales, encuestas, etc.), el primer paso es la codificación. Codificar consiste en

asignar códigos numéricos a las distintas categorías o respuestas observadas, de manera que puedan procesarse estadísticamente. Por ejemplo, si en una encuesta se registra la variable “Sexo” de los participantes, se puede codificar “hombre”=1 y “mujer”=2. Este procedimiento facilita el manejo de datos y permite también definir códigos especiales (por ejemplo, el valor 9) para observaciones faltantes. Una vez codificados, se organizan los datos en una hoja de cálculo (Excel u otro formato similar) o directamente en un software estadístico. La disposición habitual es colocar cada caso o individuo en una fila y cada variable en una columna. De este modo, la estructura de la base de datos queda clara: una fila corresponde a un sujeto o elemento de la muestra y cada columna, a una variable medida (por ejemplo, edad, sexo codificado, resultados de un test, etc.). A menudo se emplea Excel para ingresar los datos iniciales: cada celda contiene un valor numérico o codificado. Es importante entonces verificar la consistencia de los datos (revisar rangos válidos, ausencia de duplicados, codificación homogénea) antes de proceder al análisis. En paralelo, se debe llevar un documento (manual de códigos) en el que se explique qué representa cada código numérico para cada variable.

Análisis descriptivo con herramientas básicas

Con los datos organizados, el siguiente paso es realizar un análisis exploratorio o descriptivo. Aquí se calculan medidas como la media, la mediana, la varianza o la desviación estándar para variables numéricas, y se construyen tablas

de frecuencia para variables categóricas. Por ejemplo, se pueden obtener la media de edad, el porcentaje de hombres y mujeres, o la frecuencia de respuestas para cada categoría de una pregunta. También, se elaboran gráficos (histogramas, diagramas de barras o de sectores, diagramas de caja, etc.) que permitan visualizar la distribución y las características principales de los datos. Estos gráficos y tablas ayudan a resumir y describir el comportamiento de las variables sin efectuar inferencias más allá de la muestra. Un ejemplo, un histograma mostraría cómo se distribuyen las edades en la muestra; un diagrama de barras reflejaría las frecuencias de cada categoría de una encuesta, facilitando la interpretación rápida de los datos.

Uso de software estadístico

Para procesar grandes volúmenes de datos y efectuar cálculos más complejos, se utilizan paquetes especializados. SPSS, R y Jamovi son ejemplos de software comúnmente empleados. SPSS ofrece menús gráficos para análisis estadísticos, R es un entorno de programación estadística de código abierto muy poderoso, y Jamovi es un software libre con interfaz intuitiva que combina lo mejor de R sin requerir conocimientos de programación. Es decir, Jamovi es “un software gratuito y de libre distribución” diseñado con una interfaz gráfica de usuario (GUI) que permite aprovechar las capacidades de R sin necesidad de escribir código. Estas herramientas permiten importar la hoja de datos (por ejemplo, un archivo Excel o CSV), definir el tipo de cada variable (cuantitativa o categórica), y luego aplicar análisis descriptivos automáticos (resúmenes estadísticos, tablas y gráficos). Su uso agiliza el proceso, pero requiere un entrenamiento mínimo para interpretar correctamente los resultados que se obtienen. En general, se recomienda a los investigadores que aprendan al menos un paquete de este tipo para asegurar precisión y reproducibilidad en el análisis estadístico.

Estadística inferencial: pruebas básicas

Una vez conocidos los resultados descriptivos, se procede a las pruebas de hipótesis o estadística inferencial, que permiten extraer conclusiones más allá de la muestra. Entre las pruebas más comunes en investigación cuantitativa se encuentran:

- **Prueba t de Student:** Se usa para comparar la media de una variable numérica entre dos grupos independientes (por ejemplo, comparar el rendimiento medio en un examen entre hombres y mujeres). En este caso, la prueba t evalúa si la diferencia observada entre las medias podría atribuirse al azar. El método consiste en comparar la media de cada grupo considerando su variabilidad, asumiendo que los grupos son independientes.
- **ANOVA (Análisis de Varianza):** Generaliza la prueba t cuando hay tres o más grupos. Un ANOVA de un factor contrasta la igualdad de medias en “k” grupos distintos sobre una variable cuantitativa. Así, comprueba si al menos dos de esas medias de poblaciones independientes difieren significativamente. En la práctica se verifica la hipótesis nula de que todas las medias son iguales contra la alternativa de que alguna difiere.
- **Chi-cuadrado (ji-cuadrado):** Prueba no paramétrica útil cuando ambas variables son categóricas. La prueba de chi-cuadrado evalúa la diferencia entre las frecuencias observadas y esperadas bajo la hipótesis de independencia. En esencia, cuantifica si las diferencias en una tabla de contingencia son mayores de lo esperado por azar, respondiendo

preguntas como “¿varían juntas las dos variables categóricas, o son independientes?”.

- **Otras pruebas:** Según el diseño pueden usarse correlación de Pearson o Spearman (para medir asociación entre variables cuantitativas), regresión lineal, pruebas no paramétricas (por ejemplo, U de Mann-Whitney o Kruskal-Wallis) si no se cumplen supuestos, entre otras.

Estas pruebas se ejecutan típicamente dentro del software elegido (p. ej. SPSS, R o Jamovi) seleccionando las variables de interés y el tipo de contraste. El resultado incluye estadísticos de prueba (t , F , χ^2 , etc.) y valores de significación (p -valores) que indican si rechazar o no la hipótesis nula. El uso de estas herramientas estadísticas debe ajustarse a los requisitos de cada prueba (como normalidad de los datos, homogeneidad de varianzas, tamaño de muestra, etc.) para obtener conclusiones válidas.

Ejemplo ilustrativo

A modo de ejemplo, consideremos un estudio hipotético sobre el rendimiento académico de estudiantes. Supongamos que se recolectaron datos de 10 alumnos en una prueba y se registraron las variables “ID” (identificador de estudiante), “Sexo” (1=Varón, 2=Mujer), “Edad” (años) y “Nota” (puntaje numérico del examen). Estos datos se organizan en una planilla de Excel con la estructura recomendada (casos en filas, variables en columnas). Por ejemplo:

ID	Sexo (1=Varón, 2=Mujer)	Edad (años)	Nota (examen)
1	1	19	85.0
2	2	18	90.5

ID	Sexo (1=Varón, 2=Mujer)	Edad (años)	Nota (examen)
3	1	20	78.0
4	2	21	88.0
5	1	19	92.5
6	2	20	75.0
7	1	18	81.0
8	2	19	89.0
9	1	20	83.5
10	2	21	95.0

En este ejemplo, la codificación se aplicó a “Sexo” (1=Varón, 2=Mujer). Con los datos cargados, se calculan estadísticos descriptivos: por ejemplo, la media de las notas, la desviación estándar, la mediana de edad, etc. También se pueden elaborar un histograma de las notas o un diagrama de cajas por sexo para visualizar la distribución.

Para el análisis inferencial, imaginemos que se quiere saber si la media de las notas difiere entre varones y mujeres. Se realizaría entonces una *prueba t de muestras independientes* con la variable “Nota” agrupada por “Sexo”. Si el p -valor resultante es menor que el nivel de significación (usualmente 0,05), se concluye que existe una diferencia estadísticamente significativa entre los grupos. Asimismo, si existieran más de dos grupos (por ejemplo, comparando tres secciones de aula), se emplearía un ANOVA. En cambio, si se quisiera investigar si la proporción de estudiantes que aprobaron el examen (definido como nota ≥ 60) difiere por sexo, se podría transformar la información en una tabla de contingencia (aprobados vs reprobados por sexo) y aplicar la prueba de chi-cuadrado de independencia.

Este proceso completo (codificación, verificación, análisis descriptivo e inferencial) ilustra cómo los datos originales se transforman en conclusiones sólidas. El uso de software estadístico simplifica estos pasos: basta importar la tabla de datos simulados al programa elegido (p.ej. jamovi o R) para ejecutar los análisis mencionados y obtener tablas de salida con resultados numéricos, gráficos y pruebas de hipótesis.

Conclusiones

El tratamiento adecuado de datos cuantitativos es clave para el éxito de cualquier investigación científica cuantitativa. Un correcto proceso metodológico —que incluye planificación de la recolección, codificación cuidadosa, estructuración ordenada en bases de datos y análisis estadístico pertinente— asegura que los resultados sean robustos y reproducibles. En primer lugar, la codificación uniforme de respuestas (p.ej. variables categóricas con valores numéricos) y la organización de los datos según el formato recomendado facilitan la automatización del análisis. Luego, las herramientas informáticas especializadas (como SPSS, R o Jamovi) permiten calcular con rapidez medidas descriptivas y ejecutar pruebas inferenciales con un esfuerzo reducido, siempre y cuando el investigador entienda las hipótesis y supuestos detrás de cada prueba.

La meticulosidad en el manejo de los datos contribuye directamente al rigor metodológico: evita errores de procesamiento, minimiza los sesgos de entrada y permite auditar fácilmente cada paso (por ejemplo, revisando tablas de frecuencia o líneas de código). Se recomienda además documentar todos los pasos (codificación, transformaciones, supuestos estadísticos) y realizar comprobaciones cruzadas (como la validación de datos ingresados). De esta forma, los investigadores noveles pueden asegurarse de que sus análisis sean consistentes.

En síntesis, los datos cuantitativos cobran valor cuando se tratan con rigor: la claridad en la codificación y organización posibilita resúmenes estadísticos efectivos (descriptivos) y la realización de pruebas estadísticas adecuadas (inferenciales). El empleo de software estadístico amplía las capacidades analíticas del investigador, pero no sustituye la necesidad de una base metodológica sólida. Para quienes comienzan en la investigación, es aconsejable seguir guías de análisis de datos, aprovechar tutoriales sobre los programas estadísticos y, sobre todo, siempre verificar la calidad de los datos antes de extraer conclusiones. Con estos cuidados, el tratamiento de datos cuantitativos fortalece la objetividad y credibilidad de los hallazgos científicos.

Palabras claves: Análisis cuantitativo, Codificación de datos, Estadística inferencial.

Autores:

¹Roberto Carlos Barriga Granados: Docente de la Universidad Michoacana de San Nicolás de Hidalgo.

Orcid: 0009-0001-5000-3782

Contacto: roberto.carlos.barriga@umich.mx

²Carlos Heriberto Mendoza Perez: Docente de la Universidad Michoacana de San Nicolás de Hidalgo.

Orcid: 0009-0005-9118-1189

Contacto: carlos.mendoza@umich.mx

³Victor Fernando Nieto del Valle: Doctor en Educación por UNIVIM, Docente de la Universidad Michoacana de San Nicolás de Hidalgo, adscrito al Colegio Primitivo y Nacional de San Nicolás de Hidalgo de la UMSNH.

Orcid: 0000-0002-7999-3548

Contacto: fnieto@umich.mx

Referencias bibliográficas

Estadística descriptiva: 2.1 Codificación, Introducción de datos. Universidad de Valencia (cursos de Estadística).

García Bellido, R.; González Such, J.; Jornet, J.M. (s.f.). SPSS: Prueba t para muestras independientes. Universidad de Valencia.

Bakieva, M.; González Such, J.; Jornet, J.M. (s.f.). SPSS: ANOVA de un factor. Universidad de Valencia.

Escuela de Salud Pública, Universidad de Chile (2022). *Jamovi, un software estadístico útil para la práctica docente.*

Prueba Chi-cuadrado: Comprender y aplicar esta herramienta estadística. Mind The Graph (blog de difusión científica).