

Chi Cuadrado y tablas de contingencia aplicado en SPSS

Chi-Square and contingency tables applied in SPSS

Tabelas de qui-quadrado e contingência aplicadas no SPSS

Chango Pilamunga, Byron Orlando

Unidad Educativa Francisco Flor

byron.changop@educacion.gob.ec

<https://orcid.org/0009-0009-6170-4128>



Chancusig López, Marcelo Bernave

Universidad de las Fuerzas Armadas ESPE

mbchancusig@espe.edu.ec

<https://orcid.org/0009-0006-8794-4232>



García Monar, Katty Rocío

Unidad Educativa Corina Parral de Velasco Ibarra

rocio.garcia@educacion.gob.ec

<https://orcid.org/0009-0007-2111-513X>



Haro Sarango, Alexander Fernando

Instituto Superior Tecnológico España

alexander.haro@iste.edu.ec

<https://orcid.org/0000-0001-7398-2760>



DOI / URL: <https://doi.org/10.55813/gaea/ccri/v5/nE3/329>

Como citar:

Chango Pilamunga, B. O., Chancusig López, M. B., García Monar, K. R., & Haro Sarango, A. F. (2024). Chi Cuadrado y tablas de contingencia aplicado en SPSS. Código Científico Revista De Investigación, 5(E3), 499–514.

Recibido: 20/03/2023

Aceptado: 10/04/2024

Publicado: 30/04/2024

Resumen

Este estudio investiga la independencia entre variables categóricas utilizando la prueba Chi Cuadrado y tablas de contingencia en SPSS. Se explora la robustez de SPSS en el manejo de grandes volúmenes de datos, facilitando análisis estadísticos complejos accesibles para usuarios de distintos niveles de experiencia. La metodología empleada es no experimental, transversal y descriptiva, centrada en evaluar las diferencias estadísticamente significativas en las tablas de contingencia. El análisis se realizó mediante la creación de tablas para cada par de variables de interés, aplicando la prueba Chi Cuadrado y analizando residuos estandarizados y ajustados para identificar contribuciones significativas al chi cuadrado total. Los resultados confirmaron asociaciones estadísticamente significativas entre las variables estudiadas, destacando la importancia de una interpretación cautelosa de los valores p. Las conclusiones subrayan la utilidad del test de Chi Cuadrado en aplicaciones donde no se cumplen supuestos paramétricos, además de resaltar el papel crítico de SPSS en la mejora de la eficacia del análisis de datos en ciencias sociales y empresariales, promoviendo un enfoque más crítico y amplio en la interpretación de resultados estadísticos.

Palabras clave: Prueba Chi Cuadrado, Tablas de Contingencia, SPSS, Análisis de Datos Categóricos, Estadística.

Abstract

This study investigates the independence between categorical variables using the Chi-Square test and contingency tables in SPSS. It explores the robustness of SPSS in handling large volumes of data, facilitating complex statistical analyses accessible to users of different levels of experience. The methodology employed is non-experimental, cross-sectional, and descriptive, focusing on evaluating statistically significant differences in contingency tables. The analysis was performed by creating tables for each pair of variables of interest, applying the Chi-square test and analyzing standardized and adjusted residuals to identify significant contributions to the total chi-square. The results confirmed statistically significant associations between the variables studied, highlighting the importance of a cautious interpretation of p-values. The conclusions underline the usefulness of the Chi-Square test in applications where parametric assumptions are not met, as well as highlighting the critical role of SPSS in improving the efficiency of data analysis in social sciences and business, promoting a more critical and comprehensive approach to the interpretation of statistical results.

Keywords: Chi-Square Test, Contingency Tables, SPSS, Categorical Data Analysis, Statistics.

Resumo

Este estudo investiga a independência entre variáveis categóricas usando o teste do qui-quadrado e tabelas de contingência no SPSS. Ele explora a robustez do SPSS no tratamento de grandes volumes de dados, facilitando análises estatísticas complexas acessíveis a usuários de diferentes níveis de experiência. A metodologia empregada é não-experimental, transversal e descritiva, com foco na avaliação de diferenças estatisticamente significativas em tabelas de contingência. A análise foi realizada criando-se tabelas para cada par de variáveis de interesse, aplicando-se o teste do qui-quadrado e analisando-se os resíduos padronizados e ajustados para identificar contribuições significativas para o qui-quadrado total. Os resultados confirmaram associações estatisticamente significativas entre as variáveis estudadas, destacando a importância de uma interpretação cautelosa dos valores de p. As conclusões enfatizam a utilidade do teste do qui-quadrado em aplicações em que as suposições paramétricas não são atendidas, além de destacar o papel fundamental do SPSS na melhoria da eficiência da análise

de dados nas ciências sociais e nos negócios, promovendo uma abordagem mais crítica e abrangente da interpretação dos resultados estatísticos.

Palavras-chave: Teste de qui-quadrado, Tabelas de contingência, SPSS, Análise de dados categóricos, Estatística.

Introducción

SPSS es conocido por su capacidad para manejar grandes volúmenes de datos y realizar análisis estadístico complejo de manera accesible. Su interfaz gráfica, junto con la posibilidad de usar sintaxis de comandos, hace que sea una opción preferida tanto para novatos como para usuarios avanzados en el ámbito de la investigación social y de negocios. Las funciones de gestión de datos permiten a los usuarios seleccionar casos, reestructurar archivos, crear datos derivados y documentar el conjunto de datos, facilitando así el análisis integral de la información (Bala, 2016; Castañeda, 2010; Pedroza et al., 2007).

SPSS facilita la descripción y exploración de datos a través de estadísticas descriptivas, tablas de frecuencia, y gráficos, incluyendo histogramas y diagramas de dispersión. Esta capacidad es esencial para entender la distribución y características clave de las variables antes de proceder a análisis más complejos (Alili & Krstev, 2019; Nel, 2014).

La prueba Chi Cuadrado, desarrollada inicialmente por Karl Pearson en 1900, es fundamental en estadísticas por su capacidad de aplicarse en diversas disciplinas, incluyendo psicología, ciencias del comportamiento, y, de manera crucial, en ciencias sociales y empresariales. Esta prueba es especialmente útil para testear la independencia de dos variables y evaluar qué tan bien un modelo teórico se ajusta a un conjunto de datos (Hernández de la Rosa et al., 2017; Rossi, 2010).

En el ámbito de las ciencias sociales y la gestión empresarial, la prueba es valorada por su robustez y la detallada información que proporciona sobre el comportamiento de los grupos estudiados, como resalta McHugh (2013) en su análisis sobre las ventajas de la prueba Chi

Cuadrado. Esta herramienta permite la evaluación de variables dicotómicas y la realización de estudios de múltiples grupos sin requerir igualdad de varianzas entre ellos, lo que la hace adecuada para aplicaciones donde no se cumplen los supuestos paramétricos.

En ciencias sociales, la prueba Chi Cuadrado se utiliza frecuentemente para explorar relaciones entre variables sociodemográficas y diversos fenómenos comportamentales, permitiendo a los investigadores comprender y modelar complejidades en la dinámica social. Por ejemplo, un estudio sobre la influencia de la arquitectura en la asistencia a iglesias utilizó la prueba Chi Cuadrado para confirmar la relación entre el diseño arquitectónico y el patrocinio de las iglesias, demostrando cómo factores estéticos y espirituales pueden influir en el crecimiento de la asistencia (Cedron, 2017; Mendivelso & Rodríguez, 2018; Olanusi & Samuel, 2023).

Una tabla de contingencia es una herramienta estadística que permite observar la relación entre dos o más variables categóricas. Por ejemplo, para estudiar cómo el género afecta las preferencias de compra online, se puede crear una tabla que cruce género (masculino, femenino) con categorías de productos preferidos (digitales, no digitales) como se hizo en el estudio de Pascual-Miguel et al. (2015). En este estudio, se encontraron diferencias significativas en las expectativas de esfuerzo y la intención de compra entre géneros, influenciadas además por el tipo de producto.

El análisis de las tablas de contingencia generalmente involucra la prueba Chi Cuadrado para determinar si las diferencias observadas entre grupos son estadísticamente significativas. Por ejemplo, se utilizó el análisis de correspondencias para visualizar las conexiones entre variables categóricas como comportamientos de compra durante la pandemia de COVID-19. Este tipo de análisis ayuda a identificar patrones significativos y diferencias en comportamientos de compra entre grupos de edad y preferencias de servicio (Matejková & Matušek, 2023).

En el ámbito del marketing y la gestión empresarial, el análisis de tablas de contingencia se utiliza para segmentar el mercado y adaptar las estrategias de producto y comunicación. Por ejemplo, el estudio de Kanwal et al. (2021) exploró las diferencias de género en el comportamiento de compra en línea, lo que proporciona información crucial para diseñar plataformas de e-business que apelen específicamente a segmentos de consumidores según su género.

La crítica principal a la dependencia excesiva de los valores p es que pueden llevar a conclusiones erróneas si no se manejan adecuadamente. Los investigadores han señalado que un enfoque excesivo en la significancia estadística puede resultar en la publicación de estudios con efectos pequeños pero estadísticamente significativos, los cuales podrían no ser relevantes en la práctica (Leek & Peng, 2015). La American Statistical Association ha resaltado la necesidad de ir más allá de los umbrales tradicionales de significancia estadística y considerar una variedad de evidencia estadística y no estadística en la toma de decisiones científicas y políticas (Wasserstein & Lazar, 2016).

Metodología

Diseño de la Investigación

En este estudio se emplea un diseño no experimental, transversal y descriptivo para investigar la relación entre variables categóricas mediante el uso del test de Chi Cuadrado de Independencia. Este enfoque permite evaluar si las diferencias observadas en las tablas de contingencia son estadísticamente significativas, proporcionando insights sobre la asociación entre las variables.

Procedimiento de Análisis

El análisis se llevó a cabo utilizando el software IBM SPSS Statistics. Los pasos específicos incluyeron:

1. **Creación de Tablas de Contingencia:** Se configuraron tablas para cada par de variables categóricas de interés, detallando las frecuencias observadas.
2. **Cálculo del Chi Cuadrado de Independencia:** Se aplicó el test Chi Cuadrado para evaluar la independencia entre las variables. El cálculo se realizó utilizando la función específica del SPSS para este test, ajustando los grados de libertad según el número de categorías de cada variable.
3. **Análisis de Residuos:** Para profundizar en las diferencias específicas entre categorías, se analizaron los residuos estandarizados y ajustados, identificando aquellas células de la tabla que contribuyen significativamente a la chi cuadrada total.

Criterios de Significancia

Se estableció un nivel de significancia de 0.05 para todas las pruebas estadísticas. Valores de p-valor inferiores a este umbral indican una asociación estadísticamente significativa entre las variables.

Resultados

En primera instancia cargamos los datos en el sistema SPSS, es importante considerar que, para el cálculo de la prueba de chi cuadrado (χ^2) se utiliza especialmente para analizar la relación entre variables de tipo categóricas, no para variables cuantitativas. Esta prueba se basa en la distribución chi cuadrado y se emplea esencialmente para evaluar si existe una asociatividad estadística significativa entre las categorías (Ver Figura 1):

Figura 1

Carga de los datos en el sistema SPSS

Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	
Identificador	Numérico	8	0	Nº Encuestado	Ninguna	Ninguna	16	Centro	Escala	Er
E1	Numérico	8	0	Obesidad	{1, Presencia}...	Ninguna	16	Centro	Escala	Er
E2	Numérico	8	0	Hipertensión	{1, Presencia}...	Ninguna	16	Centro	Escala	Er

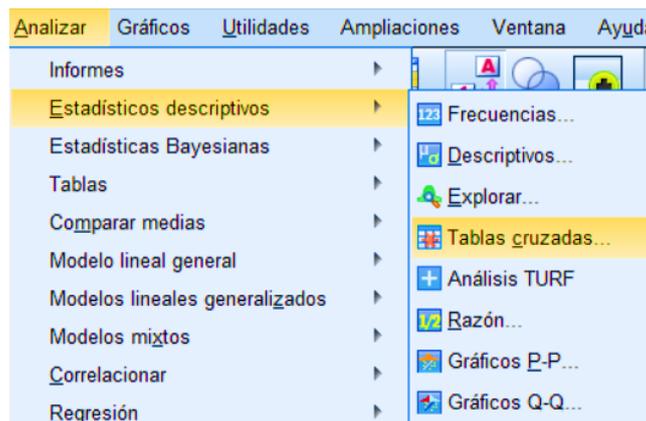
Identificador	E1	E2
1	Ausencia	Ausencia
2	Presencia	Ausencia
3	Ausencia	Ausencia
4	Ausencia	Presencia

Nota: Autores (2024)

A continuación, en la barra de opciones nos orientamos en **>Analizar<** luego **>Estadísticos descriptivos<** finalmente en la opción **>Tablas cruzadas<** (Ver Figura 2).

Figura 2

Barra de opciones

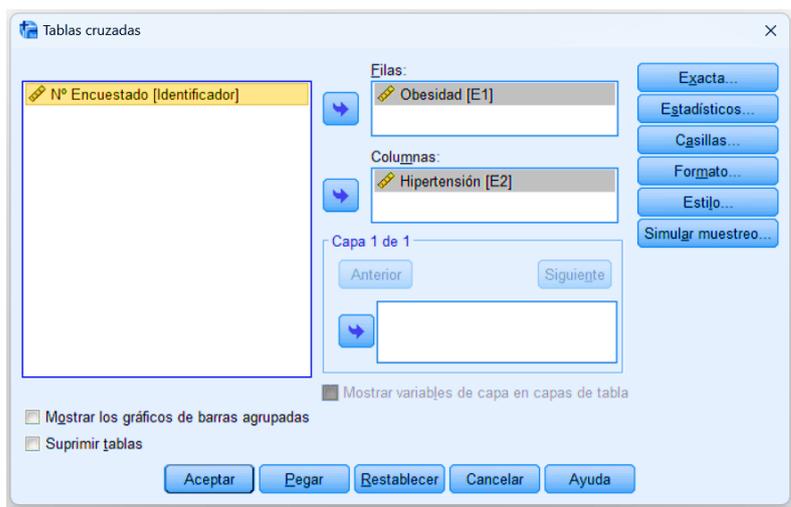


Nota: Autores (2024)

Inmediatamente, nos aparecerá la siguiente ventana en la opción **>Tablas cruzadas<**, donde, en la **>Fila<** procedemos a insertar la variable cualitativa cualesquiera que considere el investigador, la misma acción se repite en recuadro **>Columna<**, importante ratificar que debe ser cualitativa (Ver Figura 3).

Figura 3

Ventana explorar

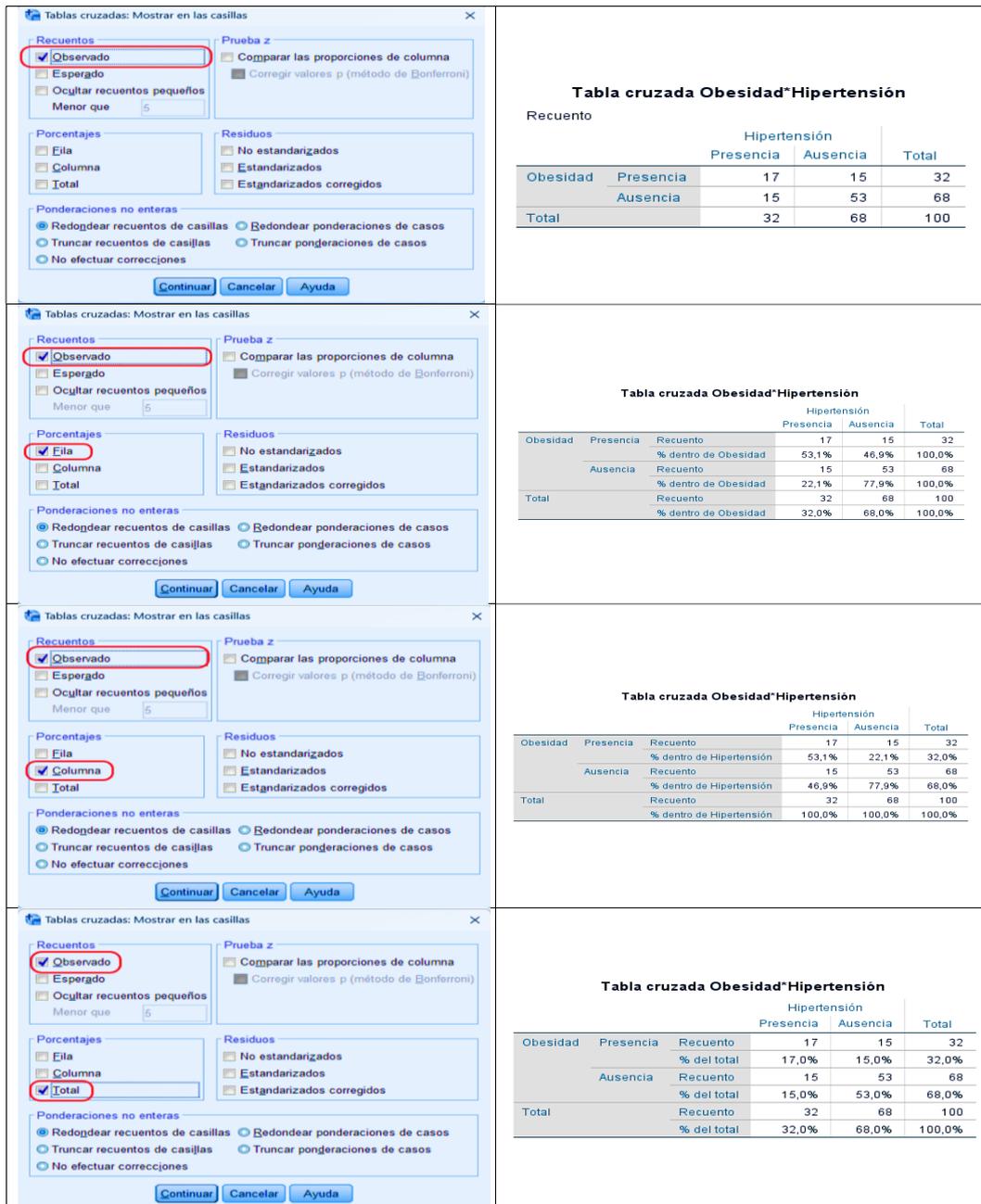


Nota: Autores (2024)

Seguidamente, indagaremos los distintos botones que nos ofrece la opción > **Tablas cruzadas**<, en el tercero denominado >**Casillas**< se procederá a generar una interacción entre cada uno para divisar los resultados que se pueden generar (Ver Figura 4).

Figura 4

Tablas cruzadas: Mostrar en las casillas



Nota: Autores (2024)

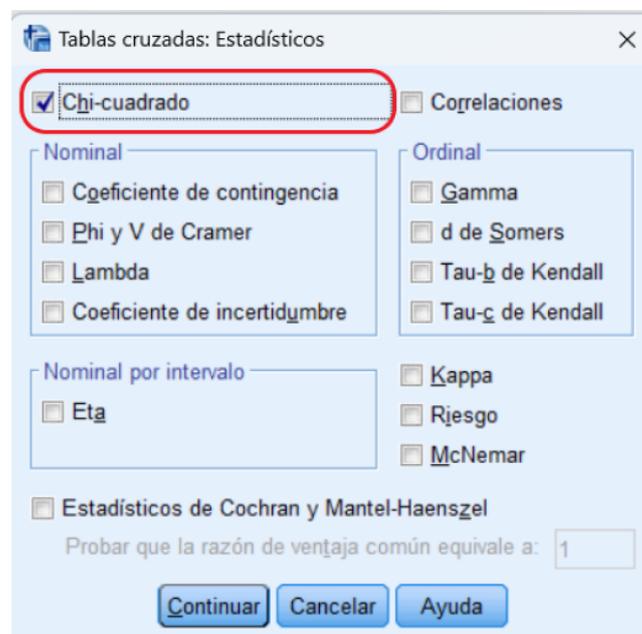
En el primer caso donde solo se selecciona >Observado< se divide el recuento de las categorías una frente a la otra en forma de conteo, sin embargo, cuando empezamos a seleccionar las diversas opciones disponibles en el recuadro porcentaje la formulación de la tabla cambia debido a que se adiciona el valor proporcional de una variable categórica frente a

la otra. Cuando seleccionamos >Fila< el proporcional es dentro de obesidad, mientras que, cuando se selecciona >Columna< este circunda en hipertensión, dando diferentes visiones de las variables; finalmente, cuando preferimos >Total< se orienta en el cruce de los totales.

Seguidamente, exploraremos en el siguiente botón que nos ofrece la opción >Tablas cruzadas<, en el denominado >Estadísticos<, es en este punto donde procederemos a generar el test. Dentro de la ventana que se despliega seleccionamos la denominada >Chi-cuadrado< (Ver Figura 5).

Figura 5

Tablas cruzadas: Estadísticos



Nota: Autores (2024)

Luego de haber dado click en la opción >Continuar< y luego >Aceptar< se despliega todo el modelo tal y con las especificaciones mencionadas. En primera instancia saldrá el >Resumen de procesamiento de casos< en donde podemos visualizar la cantidad de datos procesados en el análisis y la cantidad de datos omitidos por algún error encontrado, un ejemplo podría ser cuadros vacíos o datos fuera de las especificaciones adecuadas (Ver Figura 6).

Figura 6

Resumen de procesamiento de casos

Resumen de procesamiento de casos

	Válido		Casos Perdido		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
	Obesidad * Hipertensión	100	100,0%	0	0,0%	100

Nota: Autores (2024)

Continuadamente, se forma una tabla cruzada entre las variables cualitativas eje de estudio. La tabla cruzada proporciona información sobre la relación entre las variables Obesidad e Hipertensión. Se observa que, de un total de 32 individuos con obesidad, 17 también tienen hipertensión, mientras que 15 no la tienen. En el grupo de 68 individuos sin obesidad, 15 presentan hipertensión y 53 no. Los porcentajes del total indican que el 17% del total de individuos tienen ambas condiciones, un 15% tiene obesidad sin hipertensión, un 15% tiene hipertensión sin obesidad y el 53% no presenta ninguna de las dos condiciones. Estos datos sugieren una posible relación entre obesidad e hipertensión en la muestra analizada (Ver Figura 7)

Figura 7

*Tabla cruzada Obesidad*Hipertensión*

Tabla cruzada Obesidad*Hipertensión

		Hipertensión		Total	
		Presencia	Ausencia		
Obesidad	Presencia	Recuento	17	15	32
		% del total	17,0%	15,0%	32,0%
Ausencia	Recuento	15	53	68	
	% del total	15,0%	53,0%	68,0%	
Total	Recuento	32	68	100	
	% del total	32,0%	68,0%	100,0%	

Nota: Autores (2024)

Por consiguiente, damos interpretación a la prueba Chi-cuadrado. Se presentan los resultados de diversas pruebas de chi-cuadrado. En primer lugar, el valor del chi-cuadrado de Pearson es 9,651, con 1 grado de libertad (df), y muestra una significación asintótica bilateral

de 0,002. Se observa que este valor se mantiene coherente en otras pruebas como la corrección de continuidad (8.276) y la razón de verosimilitud (9.377), ambas con 1 grado de libertad. Asimismo, se muestra la significación exacta bilateral de 0.003 y un valor de significación exacta unilateral de 0.002 en la prueba exacta de Fisher. La prueba de asociación lineal por lineal arroja un valor de 9.555 con 1 grado de libertad y una significación bilateral de 0.002. Se reporta un total de 100 casos válidos en el análisis. Se destaca que en ningún caso se esperan recuentos menores que 5 en las celdas (0 casillas), siendo el recuento mínimo esperado de 10.24. Adicionalmente, se especifica que la prueba exacta de Fisher solo se ha calculado para una tabla 2x2. En conjunto, estos resultados indican que existen relaciones significativas entre las variables analizadas en el estudio, respaldando la relevancia de las asociaciones encontradas. (Ver Figura 8):

Figura 8

Pruebas de significancia

Pruebas de chi-cuadrado					
	Valor	df	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	9,651 ^a	1	,002		
Corrección de continuidad ^b	8,276	1	,004		
Razón de verosimilitud	9,377	1	,002		
Prueba exacta de Fisher				,003	,002
Asociación lineal por lineal	9,555	1	,002		
N de casos válidos	100				

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 10,24.

b. Sólo se ha calculado para una tabla 2x2

Nota: Autores (2024)

Discusión

La crítica de Leek y Peng (2015) sobre la dependencia excesiva de los valores p es particularmente relevante en el contexto de nuestro estudio. Mientras que la prueba Chi

Cuadrado proporciona un indicativo claro de las asociaciones estadísticamente significativas, la interpretación de estos resultados requiere cautela. Wasserstein y Lazar (2016) destacan la necesidad de ir más allá de los umbrales tradicionales de significancia estadística, sugiriendo que los investigadores deben considerar un espectro más amplio de evidencia estadística y no estadística.

Comparativamente, herramientas como el análisis de regresión logística y modelos lineales generalizados podrían ofrecer insights adicionales sobre la relación entre variables, especialmente cuando los supuestos de la prueba Chi Cuadrado no se cumplen completamente. Rossi (2010) argumenta que, aunque el Chi Cuadrado es invaluable para datos categóricos, su aplicación puede ser limitada en escenarios donde la asociación entre variables es más compleja o cuando los tamaños de muestra son pequeños.

La versatilidad de SPSS y la prueba Chi Cuadrado se demostró ampliamente en este estudio, corroborando las afirmaciones de McHugh (2013) sobre su robustez. Sin embargo, Cedrón (2017) y Mendivelso y Rodríguez (2018) señalan que los resultados pueden variar significativamente dependiendo del campo de aplicación. En ciencias empresariales, por ejemplo, la dinámica de las variables puede ser más volátil y menos predecible que en contextos más controlados como los estudios psicológicos o de comportamiento.

Una limitación notable de este estudio es la generalización de los resultados. Como menciona Alili y Krstev (2019), la extrapolación de los hallazgos a poblaciones más amplias requiere cautela debido a la variabilidad en los diseños de investigación y en las características de las muestras. Para futuras investigaciones, se recomienda la implementación de estudios longitudinales que puedan proporcionar una comprensión más profunda de las dinámicas a largo plazo entre las variables estudiadas.

Conclusión

El presente estudio demostró la eficacia de la prueba Chi Cuadrado y las tablas de contingencia aplicadas en SPSS para analizar la independencia entre variables categóricas en diversos contextos sociales y empresariales. La aplicación de estas herramientas estadísticas facilitó la comprensión de las relaciones entre variables sociodemográficas y fenómenos comportamentales, permitiendo a los investigadores modelar complejidades en la dinámica social y tomar decisiones estratégicas informadas.

Los resultados obtenidos validan la robustez de la prueba de Chi Cuadrado, subrayando su utilidad en la evaluación de asociaciones entre grupos sin la necesidad de cumplir con los supuestos paramétricos típicos de otros test estadísticos. Este enfoque no solo proporciona un método riguroso para el análisis estadístico de datos categóricos, sino que también enfatiza la importancia de interpretar adecuadamente los valores de p , evitando las trampas de las inferencias erróneas basadas exclusivamente en la significancia estadística.

Además, la utilización de SPSS como herramienta de análisis ha demostrado ser indispensable por su capacidad para manejar grandes volúmenes de datos y ofrecer una interfaz accesible tanto para usuarios novatos como avanzados. Este estudio resalta cómo el uso adecuado de software estadístico puede influir positivamente en la calidad y eficacia del análisis de datos, facilitando la exploración, visualización y comprensión de las relaciones entre variables.

Referencias bibliográficas

- Alili, A., & Krstev, D. (2019). USING SPSS FOR RESEARCH AND DATA ANALYSIS. Knowledge International Journal. <https://doi.org/10.35120/kij3203363a>
- Bala, J. (2016). Contribution of SPSS in Social Sciences Research. International Journal of Advanced Research in Computer Science, 7. <https://doi.org/10.26483/IJARCS.V7I6.2773>
- Castañeda, M. (2010). Procesamiento de datos y análisis estadísticos utilizando SPSS: Un libro práctico para investigadores y administradores educativos. EDIPUCRS.
- Cedrón, L. A. A. (2017). PRUEBA CHI-CUADRADO EN LA ESTADÍSTICA NO PARAMÉTRICA. Ciencias, 1(1), Article 1. <https://doi.org/10.33326/27066320.2017.1.829>
- Hernández de la Rosa, Y., Hernández Moreno, V. J., Batista Hernández, N. E., & Tejada Castañeda, E. (2017). ¿Chi cuadrado o Ji cuadrado? Medicentro Electrónica, 21(4), 294–295.
- Kanwal, M., Burki, U., Ali, R., & Dahlstrom, R. (2021). Systematic review of gender differences and similarities in online consumers' shopping behavior. Journal of Consumer Marketing. <https://doi.org/10.1108/jcm-01-2021-4356>
- Leek, J., & Peng, R. (2015). Statistics: P values are just the tip of the iceberg. Nature, 520, 612–612. <https://doi.org/10.1038/520612a>
- Matejková, E., & Matušek, V. (2023). The use of correspondence analysis in exploring consumer purchasing behavior.
- McHugh, M. L. (2013). The Chi-square test of independence. Biochemia Medica, 23, 143–149. <https://doi.org/10.11613/BM.2013.018>
- Mendivelso, F., & Rodríguez, M. (2018). Prueba Chi-Cuadrado de independencia aplicada a tablas 2xN. Revista Médica Sanitas, 21, 92–95. <https://doi.org/10.26852/01234250.6>
- Nel, Q. (2014). Estadística con SPSS 22. Editorial Macro.
- Olanusi, D. J. A., & Samuel, S. A. (2023). Application of Chi-Square Test to Determine Architectural Impact on Church Patronage. International Journal of Research and Innovation in Social Science. <https://doi.org/10.47772/ijriss.2023.7646>
- Pascual-Miguel, F., Agudo-Peregrina, Á. F., & Chaparro-Peláez, J. (2015). Influences of gender and product type on online purchasing. Journal of Business Research, 68, 1550–1556. <https://doi.org/10.1016/J.JBUSRES.2015.01.050>

- Pedroza, H., Dicoovskyi, L., Agricultura (IICA), I. I. de C. para la, & Instituto Nicaragüense de Tecnología Agropecuaria, M. (Nicaragua) (INTA). (2007). Sistema de análisis estadísticos con SPSS. Instituto Nicaragüense de Tecnología Agropecuaria (INTA). <https://repositorio.iica.int/handle/11324/4106>
- Rossi, J. (2010). Chi-Square Test. 1–1. <https://doi.org/10.1002/9780470479216.CORPSY0174>
- Wasserstein, R., & Lazar, N. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70, 129–133. <https://doi.org/10.1080/00031305.2016.1154108>