

Tablas de Contingencia y el Control por Tercera Variable

Unidad 5: Estadística Descriptiva Bivariada

Gabriel Sotomayor

2025-11-17



Objetivos de la Sesión de Hoy

- **Construir, interpretar y dar formato a tablas de contingencia** (tablas cruzadas).
- **Dominar** el cálculo y la interpretación de **porcentajes condicionales** (de fila, de columna y totales).
- **Introducir** el concepto de control estadístico mediante una tercera variable (Z).
- **Comprender y diagnosticar una asociación espuria.**



1. La Tabla de Contingencia

De Comparar Grupos a Cruzar Categorías

En la clase anterior, vimos cómo analizar la relación entre una variable categórica y una cuantitativa.

Hoy, nos enfocamos en el caso **Categórica (X) → Categórica (Y)**, uno de los análisis más clásicos e importantes en sociología.

- **Pregunta clave:** ¿La probabilidad de pertenecer a una categoría de la variable Y depende de la categoría de la variable X a la que se pertenece?
- **Herramienta Central:** La **tabla de contingencia** (también llamada tabla cruzada o de doble entrada).



Anatomía de una Tabla de Contingencia

Una tabla de contingencia cruza las categorías de una variable explicativa (X) con las de una variable respuesta (Y), y muestra la **frecuencia** de casos en cada celda.

Ejemplo con ENUT 2023: - Pregunta: ¿Se asocia el **sexo** (X) con la **participación en trabajo doméstico no remunerado** (p_tdnr_dt, Y)?

Frecuencias Absolutas: Participación en Trabajo Doméstico por Sexo

p_tdnr_dt_factor	Hombre	Mujer	Total
No	739	350	1089
Sí	11118	16270	27388
Total	11857	16620	28477

Problema: Hay más mujeres que hombres en la muestra que contestó el cuestionario de uso del tiempo. Comparar los conteos directos (9,742 vs. 17,646) puede ser engañoso.



2. Uso de Porcentajes

Tres Tipos de Porcentajes

En una tabla de contingencia, podemos calcular tres tipos de porcentajes:

1. **Porcentajes Totales:** $(\text{frecuencia de la celda} / N \text{ Total}) * 100$.

- *Responde a:* “¿Qué porcentaje del total de la muestra son, por ejemplo, hombres que no participan en trabajo doméstico?”.
- *Utilidad:* Bueno para describir la composición de la muestra, pero **malo para analizar relaciones**.

2. **Porcentajes de Fila:** $(\text{frecuencia de la celda} / \text{Total de la Fila}) * 100$.

- *Responde a:* “De todas las personas que participan en trabajo doméstico, ¿qué porcentaje son hombres y qué porcentaje son mujeres?”.

3. **Porcentajes de Columna:** $(\text{frecuencia de la celda} / \text{Total de la Columna}) * 100$.

- *Responde a:* “Del total de hombres, ¿qué porcentaje participa en trabajo doméstico?”.



Comparando grupos

Para analizar una relación explicativa ($X \rightarrow Y$), debemos seguir un procedimiento simple pero fundamental:

1. Calcula los porcentajes en la dirección de la variable explicativa (X).
2. Compara los porcentajes en la dirección opuesta.

- **En la práctica:** Si tu variable explicativa (X) está en las **columnas**, calcula **porcentajes de columna** (cada columna suma 100%).
- Luego, para ver si hay una relación, **compara los porcentajes a través de las filas**.

Si los porcentajes son diferentes entre las columnas, significa que la distribución de Y **depende** de la categoría de X. ¡Hay una asociación!



Ejemplo de comparación

Pregunta: ¿Se asocia el **sexo** (X) con la **participación en trabajo doméstico** (Y)?

Participación en Trabajo Doméstico por Sexo (%)

p_tdnr_dt_factor	Hombre	Mujer	Valor Perdido
No	739 (6.2%)	350 (2.1%)	0 (-)
Sí	11,118 (93.8%)	16,270 (97.9%)	0 (-)
Valor Perdido	0 (0.0%)	0 (0.0%)	0 (-)

Interpretación:

1. **Calculamos** porcentajes de columna (X = **sexo_factor** está en las columnas).
2. **Comparamos** los porcentajes a través de la fila “Sí participa”.
3. **Conclusión:** Un **97.9%** de las mujeres participa en trabajo doméstico, en comparación con un **93.8%** de los hombres. Como estos porcentajes son diferentes, concluimos que **existe una asociación** entre el sexo y la participación en trabajo doméstico en la muestra.



3. El Control por una Tercera Variable

¿Es “Real” la Relación?

El análisis bivariado es solo el primer paso. Una pregunta fundamental en sociología es:

¿La relación que observamos entre X e Y se mantiene, o es un “espejismo” creado por una tercera variable (Z) que no hemos considerado?

Para responder a esto, usamos una técnica central en el análisis de datos: el **control estadístico**.

La idea es **estratificar**: rehacemos nuestro análisis de la relación entre X e Y, pero esta vez por separado para cada categoría de la variable de control Z. Esto nos permite ver si la relación original se comporta de la misma manera en diferentes contextos.



Un Ejemplo Clásico: La Asociación Espuria

Antes de usar nuestros datos, veamos un ejemplo conceptual clásico para entender la lógica.

- **Relación Bivariada Original ($X \rightarrow Y$):** En una ciudad, se observa que a mayor **número de bomberos** (X) en un incendio, mayores son los **daños económicos** (Y).
 - *Conclusión apresurada:* “Los bomberos causan los daños”.
- **Introducimos una Variable de Control (Z):** La **magnitud del incendio** (pequeño / grande).
- **Análisis Estratificado:**
 - **Tabla 1 (Solo para incendios pequeños):** Vemos que no hay relación entre el número de bomberos y los daños.
 - **Tabla 2 (Solo para incendios grandes):** Vemos que tampoco hay relación.
- **Conclusión Final:** La relación original era **espuria** (engañosa). La magnitud del incendio (Z) era la **causa común** que explicaba tanto el aumento de bomberos (X) como el aumento de los daños (Y).



Aplicando el Control en la Práctica (ENUT 2023)

Paso 1: La Relación Original (X → Y)

Analicemos la relación bivariada entre **sexo** (X) y la **participación en trabajo de cuidados no remunerado** (p_tcnr_dt, Y).

Participación en Trabajo de Cuidados por Sexo (%)

p_tcnr_dt_factor	Hombre	Mujer	Valor Perdido
No	8,409 (70.9%)	9,811 (59.0%)	0 (-)
Sí	3,448 (29.1%)	6,809 (41.0%)	0 (-)
Valor Perdido	0 (0.0%)	0 (0.0%)	0 (-)

Interpretación Inicial: Observamos una clara brecha de género. Un **41.0%** de las mujeres participa en trabajo de cuidados, en comparación con solo el **29.1%** de los hombres. Esto representa una diferencia de casi 12 puntos porcentuales en la participación.



Aplicando el Control en la Práctica (ENUT 2023)

Paso 2: Introduciendo la Variable de Control (Z)

Pregunta: Pero, ¿es esta brecha de género la misma para todas las personas? ¿O podría depender de su relación con el mercado laboral? Por ejemplo, entre las personas que ya tienen un trabajo remunerado, la distribución de cuidados podría ser diferente.

- **Variable de Control (Z):** *cae* (Condición de Actividad Económica).
- **Estrategia:** Repetiremos la tabla anterior, pero ahora por separado para “Personas ocupadas” y “Personas inactivas”.



Aplicando el Control en la Práctica (ENUT 2023)

Paso 3: Las Tablas Parciales o Estratificadas

Tabla Parcial 1: Solo Personas Ocupadas (%)

p_tcnr_dt_factor	Hombre	Mujer	Valor Perdido
No	5,011 (66.4%)	4,290 (53.2%)	0 (-)
Sí	2,540 (33.6%)	3,778 (46.8%)	0 (-)
Valor Perdido	0 (0.0%)	0 (0.0%)	0 (-)

Tabla Parcial 2: Solo Personas Inactivas (%)

p_tcnr_dt_factor	Hombre	Mujer	Valor Perdido
No	2,628 (80.7%)	4,832 (64.8%)	0 (-)
Sí	630 (19.3%)	2,629 (35.2%)	0 (-)
Valor Perdido	0 (0.0%)	0 (0.0%)	0 (-)



Conclusión del Análisis de Control

Al comparar las tablas parciales, observamos un patrón de **Especificación (o Interacción)**:

1. **Entre las personas ocupadas:** La brecha de género en la participación de cuidados **se mantiene**, y de hecho, es similar a la brecha general. Un **46.8%** de las mujeres ocupadas participa en cuidados, frente a un **33.6%** de los hombres ocupados (una diferencia de 13.2 puntos).
2. **Entre las personas inactivas:** La brecha de género también **se mantiene**, siendo incluso ligeramente mayor. La participación de las mujeres inactivas es del **35.2%**, mientras que la de los hombres inactivos es solo del **19.3%** (una diferencia de 15.9 puntos).

Conclusión Sociológica: La relación entre sexo y trabajo de cuidados no es espuria, sino robusta y **especificada** por la condición de actividad. La brecha de género existe tanto dentro como fuera del mercado laboral, pero se manifiesta de forma diferente: aunque en términos absolutos la participación en cuidados es mayor entre las personas ocupadas, la **desigualdad relativa** (la diferencia porcentual entre hombres y mujeres) es aún más pronunciada entre quienes están fuera de la fuerza de trabajo.



Patrones del Análisis de Control

Al introducir una variable de control (Z), la relación original entre X e Y puede:

1. **Replicarse:** La relación se mantiene igual en todas las tablas parciales. Esto refuerza nuestra confianza en la relación original.
2. **Desaparecer (Explicación / Asociación Espuria):** La relación desaparece en las tablas parciales. Esto sugiere que Z es una causa común que explica la relación original (como en el ejemplo de los bomberos).
3. **Especificarse (Interacción):** La relación es **diferente** en las distintas categorías de Z, siendo más fuerte para un grupo que para otro. **Esto es lo que acabamos de encontrar en nuestro ejemplo de la ENUT:** la relación entre sexo y trabajo de cuidados fue especificada por la condición de actividad.



Cierre y Próximos Pasos

Resumen de la sesión de hoy:

- Las **tablas de contingencia** son la herramienta central para analizar la relación entre dos variables categóricas.
- El procedimiento estandar es **calcular porcentajes en la dirección de la variable explicativa** y comparar en la dirección opuesta.
- El **control por una tercera variable** es un paso crucial del análisis sociológico para probar la robustez de nuestras asociaciones y evitar conclusiones espurias.

En el práctico de hoy:

- Utilizarán el paquete **janitor** para crear tablas de contingencia ponderadas y estratificadas, aplicando la regla de los porcentajes.

Adelanto de la próxima clase:

- Sintetizaremos el análisis bivariado, reforzando las estrategias visuales y cuantificando la relación entre variables cuantitativas con el **coeficiente de correlación**.

