David Montaño Castro
Tarea 8. Prueba de independencia con tablas de contingencia. Prueba Chi Cuadrada
Tema: Pruebas No Paramétricas, bondad de ajuste.

Se presentan los siguientes datos en una tabla. En las columnas, se tienen los sexos 'hombre' y 'mujer'; en los reglones, se tienen diferentes niveles educativos.

Nivel Educativo	Hombre	Mujer	Total
No estudió	**16**	**57**	**73**
Primaria Incompleta	**105**	**105**	**210**
Primaria completa	**114**	**158**	**272**
Secundaria	**215**	**204**	**419**
Bachillerato	**244**	**317**	**561**
Superior	**259**	**262**	**521**
Total	**953**	**1103**	**2056**

Se plantea la siguiente cuestión:

¿Podría decirse que el sexo de una persona determina o tiene relación con el grado de estudios con los que esta/este cuenta?

Ho: El nivel educativo es independiente del sexo de una persona.

Vs

Ha: El nivel educativo no es independiente del sexo de una persona.

A lo que equivale:

Sea Pi la probabilidad de pertenecer al i-ésimo renglón y Pj la probabilidad de pertenecer a la j-ésima columna, para toda i,j.

$$Ho: P_{ij} = P_i P_j$$

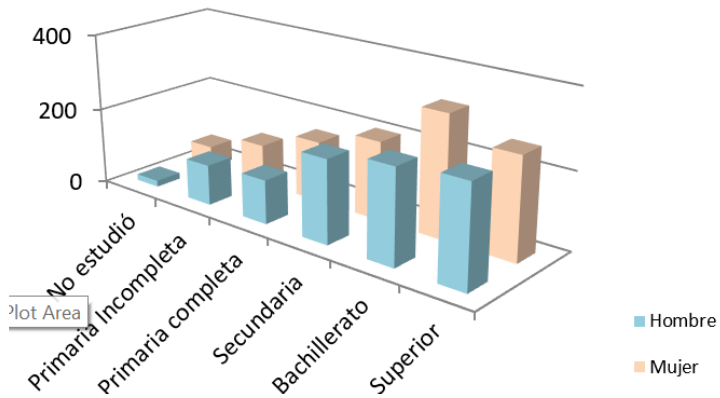
Vs

$$Ha: P_{ij} \neq P_i P_j$$

Se establece un nivel de confianza de Alpha = .05

Por simple inspección, se puede notar que, a excepción del nivel secundaria, las mujeres tienden a asistir más a la escuela y terminarla que los hombres. Esta observación a primera vista puede indicar que existe un grado dependencia; sin embargo, hay que notar que existen 150 registros más de mujeres que de hombres. Esto puede dar una falsa percepción con respecto a la interpretación de los datos. Es por eso que necesita probarse formalmente la hipótesis.

David Montaño Castro
Tarea 8. Prueba de independencia con tablas de contingencia. Prueba Chi Cuadrada
Tema: Pruebas No Paramétricas, bondad de ajuste.



Si se realiza un análisis vertical (por columna), se obtienen los siguientes porcentajes:

Nivel Educativo	Hombre	Mujer	Total
No estudió	**1.7%**	**5.2%**	**3.6%**
Primaria Incompleta	**11.0%**	**9.5%**	**10.2%**
Primaria completa	**12.0%**	**14.3%**	**13.2%**
Secundaria	**22.6%**	**18.5%**	**20.4%**
Bachillerato	**25.6%**	**28.7%**	**27.3%**
Superior	**27.2%**	**23.8%**	**25.3%**
Total	**100.0%**	**100.0%**	**100.0%**

Concerniente a los hombres, se puede decir que:

- La mayoría de ellos estudia el nivel superior; aún más, los hombres parecen presentar valores crecientes en comparación al nivel educativo. Solo el 1.7% no estudio.

Con respecto a las mujeres, se puede decir que:

- La mayoría de las mujeres estudia el nivel bachillerato, el segundo nivel más estudiado es el superior. Se puede ver que el 5.2% de las mujeres no tienen estudios.

Comparando ambos sexos:

- Pese a que hay un mayor número de registros de mujeres, es claro que los hombres tienden a estudiar un nivel más alto que las mujeres. Siguiendo esta idea, las mujeres son más propensas a no tener estudios. Las diferencias entre cada nivel educativo son mínimas, casi totalmente explicadas por la no consistencia del numero de datos entre sexos.

- Lo anterior contrasta la idea que se planteó al inicio sobre que las mujeres estudiaban más que los hombres.

Analizando el total:

- La mayoría de las personas cuenta con un nivel educativo de bachillerato y, afortunadamente, solo el 3.6% de la población entrevistada no estudio.

Posterior a ese análisis y ahora con más idea de lo que se puede esperar, se procede a calcular los valores esperados, esto para no perder objetividad por causa de la no consistencia de registros entre hombres. Los valores son calculados de la siguiente manera:

$$E_{ij} = \frac{m_i n_j}{N}$$

Distribución de Electores por Nivel Educativo y Sexo

Observadas

Nivel Educativo	Hombre	Mujer	Total
No estudió	16	57	73
Primaria Incompleta	105	105	210
Primaria completa	114	158	272
Secundaria	215	204	419
Bachillerato	244	317	561
Superior	259	262	521
Total	953	1103	2056

En donde mi corresponde al reglón, nj representa la columna j y N representa el Total.

Distribución de Electores por Nivel Educativo y Sexo

Esperadas

Nivel Educativo	Hombre	Mujer	Total
No estudió	33.84	39.16	73.00
Primaria Incompleta			210.00
Primaria completa	=D$12*$F8/F12		272.00
Secundaria	194.22	224.78	419.00
Bachillerato	260.04	300.96	561.00
Superior	241.49	279.51	521.00
Total	953.00	1103.00	2056.00

De donde se obtiene:

Nivel Educativo	Hombre	Mujer	Total
No estudió	33.84	39.16	**73.00**
Primaria Incompleta	97.34	112.66	**210.00**
Primaria completa	126.08	145.92	**272.00**
Secundaria	194.22	224.78	**419.00**
Bachillerato	260.04	300.96	**561.00**
Superior	241.49	279.51	**521.00**
Total	**953.00**	**1103.00**	**2056.00**

David Montaño Castro
Tarea 8. Prueba de independencia con tablas de contingencia. Prueba Chi Cuadrada
Tema: Pruebas No Paramétricas, bondad de ajuste.

NOTA: En valor de los totales de cada columna y fila deben de ser los mismos que los de la primera tabla. Si no, hay algún calculo mal formulado.

Estadística de prueba: Chi cuadrada con grados de libertad (# Columnas - 1) * (# Reglones - 1)

$$T^2 = \sum_{i=1}^{c}\sum_{j=1}^{k}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

En donde Oij es la observación real y Eij es la observación esperada.

Se calcula de la siguiente manera:

Observadas

Nivel Educativo	Hombre	Mujer	Total
No estudió	16	57	73
Primaria Incompleta	105	105	210
Primaria completa	114	158	272
Secundaria	215	204	419
Bachillerato	244	317	561
Superior	259	262	521
Total	953	1103	2056

Distribución de Electores por Nivel Educativo y Sexo

Esperadas

Nivel Educativo	Hombre	Mujer	Total
No estudió	33.84	39.16	73.00
Primaria Incompleta	97.34	112.66	210.00
Primaria completa	126.08	145.92	272.00
Secundaria	194.22	224.78	419.00
Bachillerato	260.04	300.96	561.00
Superior	241.49	279.51	521.00
Total	953.00	1103.00	2056.00

Distribución de Electores por Nivel Educativo y Sexo

Ji cuadrada

Nivel Educativo	Hombre	Mujer	Total
No estudió	9.403	8.124	17.527
Primaria Incompleta	0.603	0.521	1.124
Primaria completa	=POWER(D8-D20,2)/D20		2.157
Secundaria	2.224	1.922	4.146
Bachillerato	0.989	0.854	1.843
Superior	1.269	1.096	2.365
Total	15.645	13.517	29.162

David Montaño Castro
Tarea 8. Prueba de independencia con tablas de contingencia. Prueba Chi Cuadrada
Tema: Pruebas No Paramétricas, bondad de ajuste.

Teniendo la tabla completa:

Nivel Educativo	Hombre	Mujer	Total
No estudió	**9.403**	**8.124**	17.527
Primaria Incompleta	**0.603**	**0.521**	1.124
Primaria completa	**1.157**	**1.000**	2.157
Secundaria	**2.224**	**1.922**	4.146
Bachillerato	**0.989**	**0.854**	1.843
Superior	**1.269**	**1.096**	2.365
Total	**15.645**	**13.517**	29.162

Donde **29.162** será el valor de la estadística de prueba.

Auxiliándose de la función **CHISQ.DIST.RT()** se obtiene el valor P-Value:

Grados de L	**5**	(6-1)*(2-1)
Probabilidad	**0.0000216**	

El P-Value es de **0.0000216**, mucho menor comparado con el valor de Alpha (.05). Por lo tanto, se cuenta con suficiente prueba estadística para **rechazar la hipótesis nula. Esto es, existe un grado de dependencia entre el sexo y el nivel educativo.** Una explicación plausible puede estar relacionada con el cómo anteriormente se veían las mujeres privadas de la educación por creencias machistas.

También se acostumbra a calcular algunos coeficientes para la mejor interpretación de la prueba:

Coeficiente de Contingencia

$$CC = \sqrt{\frac{T^2}{T^2 + N}}$$ **0.118259943**

Este coeficiente mide el grado de asociación que depende del número de columnas y reglones, por lo tanto, su interpretación puede llegar a ser ambigua. Este coeficiente nunca llega a ser 1.

Cota Superior

$$Cota = \sqrt{\frac{R-1}{R}}$$ **0.707106781**

Este coeficiente está dado por el mínimo número entre reglones y columnas, en el cual sí se puede alcanzar el valor 1. Por lo tanto, se puede concluir que existe un grado de asociación bastante alto entre ambas variables.

La V de Cramér

$$V = \sqrt{\frac{T^2}{N(m-1)}}$$

0.119095676

Este valor tiene la ventaja de que sí alcanza el valor máximo de 1 sin tomar en cuenta la estructura de la tabla. m es el mínimo entre el número de columnas y reglones.