

Laboratorio 5: Tablas de Contingencia

1. El tipo de bebida ordenada con los alimentos en un restaurante es independiente de la edad del cliente? Se toma una encuesta al azar de 309 clientes del restaurante de donde resulta la siguiente información:

Edad	Bebida Preferida			
	Café/Té	Refresco	Otros	
21-34	26	95	18	139
35-55	41	40	20	101
>55	24	13	32	69
	91	148	70	

Utilice un nivel de significancia del 0.01 para determinar si las 2 variables son independientes.

H_0 : La bebida ordenada es independiente de la edad del cliente

H_a : La bebida ordenada no es independiente de la edad del cliente

fesp.	Café-Té	Refresco	Otros	$\frac{(f_o - f_e)^2}{f_e}$	Café/Té	Refresco	Otros
21-34	40.9353	66.5761	31.4887		5.449	12.135	5.778
35-55	29.7413	48.3754	22.8803	\Rightarrow	4.259	1.450	0.363
>55	20.3204	33.0485	15.6311		0.666	12.162	17.142

$$\chi^2_E = \sum_{i,j} \frac{(f_o - f_e)^2}{f_e} = 59.405$$

Rechazamos H_0 si $\chi^2_E > \chi^2_{\alpha, (r-1)(c-1)}$

$$\chi^2_E > \chi^2_{0.01, (3-1)(3-1)}$$

$$\chi^2_E > \chi^2_{0.01}$$

$$59.405 > 13.2767 \checkmark$$

\therefore Rechazamos H_0

Hay evidencia estadística que afirma que las variables no son independientes, es decir, la bebida ordenada depende de la edad.

2. El periódico El Reforma busca encontrar la proporción de la gente que está a favor de que López Obrador ha tenido una buena gestión como presidente de los Estados Unidos Mexicanos. Para ello, realizó una encuesta telefónica a distintos estados al interior de la república preguntando lo siguiente:

¿Está usted de acuerdo con que el actual presidente y su equipo de trabajo, autoproclamados la "4ta transformación" ha hecho un buen trabajo en materia económica y social? Responda Sí o No.

Se obtuvieron los siguientes resultados.

	Sí	No	
Nuevo León	80	170	250
Tabasco	120	130	250
Jalisco	102	148	250
Guangjuato	95	155	250
	397	630	100

Pruebe si hay alguna diferencia significativa en la proporción de individuos que están a favor de que el actual presidente ha hecho un buen trabajo en materia económica y social para estos 4 estados del interior de la república mexicana. Utilice un $\alpha = 0.01$

H_0 : La proporción de individuos a favor es igual en todos los estados
 H_a : La proporción de individuos a favor difiere en algún estado.

fesp.	Sí	No
Nuevo León	99.25	150.75
Tabasco	99.25	150.75
Jalisco	99.25	150.75
Guangjuato	99.25	150.75

$\frac{f_o - f_e}{f_e}$	Sí	No
Nuevo León	3.73	2.4581
Tabasco	1.338	2.8561
Jalisco	0.076	0.0501
Guangjuato	0.181	6.1198

$$\chi^2_e = \sum \frac{(f_o - f_e)^2}{f_e} = 13.8142$$

Rechazamos H_0 si $\chi^2_e > \chi^2_{\alpha, k-1}$

$$13.8142 > \chi^2_{0.01, 3}$$

$$13.8142 > 11.3448$$

\therefore Rechazamos H_0

\Rightarrow Al menos una proporción difiere

	$p_1 = 0.32$	$p_2 = 0.48$	$p_3 = 0.408$	$p_4 = 0.38$
$p_1 = 0.32$	/	/	/	/
$p_2 = 0.48$	① R. Ho	/	/	/
$p_3 = 0.408$	② No R. Ho	④ No R. Ho	/	/
$p_4 = 0.38$	③ No R. Ho	⑤ No R. Ho	⑥ No R. Ho	/

$$① |0.48 - 0.32| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{0.48(0.52)}{250} + \frac{0.32(0.68)}{250}}$$

$$0.16 \geq 0.14506 \checkmark$$

$$② |0.408 - 0.32| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{(0.408)(0.592)}{250} + \frac{(0.32)(0.68)}{250}}$$

$$0.088 \geq 0.1443 \times$$

$$③ |0.38 - 0.32| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{(0.38)(0.62)}{250} + \frac{(0.32)(0.68)}{250}}$$

$$0.06 \geq 0.1434 \times$$

$$④ |0.408 - 0.48| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{(0.408)(0.592)}{250} + \frac{0.48(0.52)}{250}}$$

$$0.072 \geq 0.1492 \times$$

$$⑤ |0.48 - 0.38| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{(0.48)(0.52)}{250} + \frac{0.38(0.62)}{250}}$$

$$0.10 \geq 0.1423 \times$$

$$⑥ |0.408 - 0.38| \geq \sqrt{\chi_{0.01,3}^2} \sqrt{\frac{0.408(0.592)}{250} + \frac{0.38(0.62)}{250}}$$

$$0.028 \geq 0.1471 \times$$

∴ Los datos dan evidencia de que la única proporción de individuos a favor que es diferente es la de Nuevo León con la de Tabasco

3. Un estudiante de la UANL de manera diaria toma el metro como medio de transporte para llegar a sus clases. De manera diaria toma el metro en la estación Tapia y se baja en la estación Universidad. Un día sintió que los tiempos entre estaciones eran diferentes, por lo que decidió recolectar información para poder encontrar evidencia de que su creencia era cierta. Durante 5 días, recolectó la siguiente información de los tiempos entre cada estación (minutos):

Tapia - San Nicolás	San Nicolás - Anáhuac	Anáhuac - Universidad
1.15	1.14	1.2
1.1	1.17	1.32
1.06	1.09	1.25
1.1	1.1	1.18
1.15	1.13	1.21

La creencia de este estudiante acerca de que los tiempos entre estaciones no son los mismos es correcta? En caso de que sea correcta, ¿cuáles son las estaciones que tienen diferencias significativas en sus tiempos de recorrido? Utilice una prueba de la mediana. Use $\alpha = 0.05$

H_0 : Los tiempos entre estaciones tienen la misma mediana

H_a : Al menos el tiempo entre dos estaciones difiere de las otras en su mediana

Como son 15 datos la mediana es el dato en posición 8

1.06 1.09 1.1 1.1 1.1 1.13 1.14 1.15 1.15 1.17 1.18 1.2 1.21 1.25 1.32

$Me = 1.15$

	T-SN	SN-A	A-U	Totales
> 1.15	0	1	5	6
≤ 1.15	5	4	0	9
	5	5	5	15

$$T = \frac{(15)^2}{(6)(9)} \left[\frac{1}{5} + \frac{25}{9} \right] - \frac{(15)(6)}{9} = 11.666$$

Rechazamos H_0 si $\chi^2_c > \chi^2_{d.f., \alpha}$
 $\chi^2_c > \chi^2_{0.05, 2}$

$$11.666 > 5.9914$$

\therefore Rechazamos H_0

Como se rechazó H_0 , se hacen comparaciones individuales

T-SN vs. SN-A

$Me = 1.115$

	T-SN	SN-A	Totales
> 1.115	2	3	5
≤ 1.115	3	2	5
	5	5	10

$$T = \frac{100}{(cs)(cs)} \left[\frac{4}{5} + \frac{9}{5} \right] - \frac{(10)(cs)}{5} = 0.4$$

R. H_0 si $T > \chi^2_{\alpha-0.05, 1}$

$0.4 < 3.8414$, \Rightarrow No Rechazamos H_0

T-SN vs A-U

$Me = 1.165$

	T-SN	A-U	Totales
> 1.165	0	5	5
≤ 1.165	5	0	5
	5	5	10

$$T = \frac{100}{(cs)(cs)} \left[\frac{25}{5} \right] - \frac{(10)(cs)}{5} = 10$$

R. H_0 si $T > \chi^2_{\alpha-0.05, 1}$

$10 > 3.8414 \Rightarrow$ Rechazamos H_0

SN-A vs A-U

$Me = 1.175$

	SN-A	A-U	Totales
> 1.175	0	5	5
≤ 1.175	5	0	5
	5	5	10

$$T = \frac{100}{(cs)(cs)} \left[\frac{25}{5} \right] - \frac{10(cs)}{5} = 10$$

R. H_0 si $T > \chi^2_{\alpha, 1}$

$10 > 3.8414 \Rightarrow$ Rechazamos H_0

\therefore Hay evidencia que afirma que

\rightarrow La mediana del tiempo entre Tapia y San Nicolás y entre San Nicolás y Anáhuac no difiere

\rightarrow La mediana difiere entre Tapia-San Nicolás y Anáhuac-Universidad así como entre San Nicolás-Anáhuac y Anáhuac-Universidad