

Curso Estadística IV
Sociología
Universidad
Alberto Hurtado

Profesora
Carolina Aguilera
caguilera@uahurtado.cl



Ayudantes
Vicente Díaz – vidiazam@alumnos.uahurtado.cl
Miguel Tognarelli – mtognare@alumnos.uahurtado.cl



Clase 2

13 de agosto

- Módulo 2
- CHI CUADRADO (χ^2)
- Ejercicio práctico



Módulo 2

Incentivo para la realización de ejercicios prácticos

El módulo 2 de las clases está destinado a la práctica de los conocimientos teóricos en el programa R Studio, según una pauta y materiales entregados por la profesora.

Para incentivar la realización de estos ejercicios, se propone que:

- Sean realizados en grupos de máximo 3 personas (puede realizarse de manera individual)
- La realización correcta del ejercicio se premiará con 2 décimas para el control siguiente.
- Para hacer la entrega se debe entregar el archivo de script de R, junto al archivo de la base de datos. Alternativamente se puede entregar un documento de R Markdown o en Quarto con todo el flujo del trabajo.

¿Qué es la prueba χ^2 ?



Permite saber si dos variables nominales u ordinales están asociadas (es decir tienen relación entre sí).

Por ejemplo

- género y nivel de ingreso
- nivel educacional y nivel de ingreso

¿Qué es la prueba χ^2 ?



Permite saber si dos variables nominales u ordinales están asociadas (es decir tienen relación entre sí).

Es una medida de probabilidad... “con un 95% (otro nivel) de confianza una variable está asociada a otra”

Ejemplo: base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
TOTAL	3	8	4	2	

Precaución: este ejemplo tiene muy pocos casos y para un ejercicio real se requiere que hay al menos 5 casos para cada casilla.

En general para que la prueba resulte se requieren al menos $12 \times n^{\circ}$ categorías variable 1 x n° cat variable 2. En nuestro caso tenemos: $12 \times 2 \times 4 = 96$ casos (y tenemos solo 17).

Ejemplo: base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
TOTAL	3	8	4	2	

¿Cómo comparo las menciones de “hombres” con las de “mujeres” si el grupo de hombres es más pequeño que el de mujeres?

Tengo que usar PROPORCIONES. Las proporciones me permiten comparar. Se suelen usar porcentajes o simplemente la razón entre el n de menciones y el total de ese grupo (marginales)

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombr e	16,70%	50%	16,70%	16,70%	100%
Mujer	18,20%	45,50%	27,30%	9,09%	100%

Valores “marginales”, son las frecuencias expresadas como una proporción del total (en este caso en relación a las filas, es decir al género)

$$1/6\% = 16,7 \%$$

16,7% = lo que representa “1” en un total de 6 menciones

Aquí los valores para hombres y mujeres son comparables)

Ejemplo: base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	16,7	50	16,7	16,7	100
Mujer	18,2	45,5	27,3	9,09	100

Si miramos las menciones, parecen haber diferencias entre hombres y mujeres, y el resultado es diferente cuando los miramos en términos PROPORCIONALES

Ejemplo: base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	<i>TOTAL</i>
Hombre	16,7	50	16,7	16,7	100
Mujer	18,2	45,5	27,3	9,09	100

¿Cómo sabemos si estas diferencias son o no significativas (no se deben al muestreo sino que a diferencias fuertes entre hombres y mujeres?)

Las comparamos con una distribución “ideal” donde lo que pasa a hombres y mujeres es lo mismo (es decir NO hay asociación). Esta se llamará H_0 ó Hipótesis nula.

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

Base observada

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	<i>TOTAL</i>
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
Total	3	8	4	2	17

A diagram illustrating the calculation of a summary statistic. It shows four arrows originating from the 'Total' row and column labels. One arrow points from the 'Total' row to the value '17'. Another arrow points from the 'Total' column to the value '17'. A third arrow points from the 'Total' row to the value '2'. A fourth arrow points from the 'Total' column to the value '2'. These values represent the sum of the observed frequencies across all categories.

La distribución de tiempos que demoran las personas SIN IMPORTAR GENERO (3; 8; 4 ; 2)

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

Base observada

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	<i>TOTAL</i>
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
Total	3	8	4	2	17

La PROPORCIÓN (para cada caso) de la distribución de tiempos que demoran las persona SIN IMPORTAR GENERO
Total "tiempo" / "Total casos" (3/17; 8/17; 4/17; 2/17)

"tiempo" / "Total casos"
 $3/17 = 17,6\%$
 $8/17 = 47,1\%$
 $4/17 = 23,5\%$
 $2/17 = 11,8\%$

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

Frecuencias Observadas vs Frecuencias esperadas (si no hubiera asosiación)

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	3	8	4	2	6
Mujer					11
Total	3	8	4	2	17

Considerando que tenemos 6 hombres y 11 mujeres, **¿cuáles serían los valores esperados en cada categoría**, si no hubiese diferencia entre es hombre o mujer? (se tendrían que mantener las PROPORCIONES)

Se debe “ponderar” el “peso” que tiene la categoría en la muestra; para que 6 sea equivalente (comparable) a 11

$$3 \times 6/17 = 1,1$$

$$8 \times 6/17 = 2,8$$

$$4 \times 6/17 = 1,4$$

$$2 \times 6/17 = 0,7$$

$$3 \times 11/17 = 1,9$$

$$8 \times 11/17 = 5,2$$

$$4 \times 11/17 = 2,6$$

$$2 \times 11/17 = 1,3$$

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

Base fercuencias observadas vs Frecuencia esperada (si no hubiera asosiación)

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1,1	2,8	1,4	0,7	6
Mujer	1,9	5,2	2,6	1,3	11
Total	3	8	4	2	17

Considerando que tenemos 6 hombres y 11 mujeres, **¿cuáles son los valores esperados en cada categoría**, sin que importe si es hombre o mujer?

Se mantiene la proporción sin importar el género

$$6 \times 3/17 = 1,1$$

$$6 \times 8/17 = 2,8$$

$$6 \times 4/17 = 1,4$$

$$6 \times 2/17 = 0,7$$

$$11 \times 3/17 = 1,9$$

$$11 \times 8/17 = 5,2$$

$$11 \times 4/17 = 2,6$$

$$11 \times 2/17 = 1,3$$

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL		16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6	Hombre	1,1	2,8	1,4	0,7	6
Mujer	2	5	3	1	11	Mujer	1,9	5,2	2,6	1,3	11
Total	3	8	4	2	17	Total	3	8	4	2	17

Lo observado

Lo esperado en caso
que el género no cambie las frecuencias
observadas (H_0 = Hipótesis nula)

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
Total	3	8	4	2	17

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1,1	2,8	1,4	0,7	6
Mujer	1,9	5,2	2,6	1,3	11
Total	3	8	4	2	17

Lo observado (F_o)

Lo esperado (F_e)

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minut	31-45 minut	46-60 minut	Más de 60	TOTAL
Homb re	1	3	1	1	6
Mujer	2	5	3	1	11
Total	3	8	4	2	17

	16-30 minuto s	31-45 minuto s	46-60 minuto s	Más de 60 minuto	TOTAL
Hombr e	1,1	2,8	1,4	0,7	6
Mujer	1,9	5,2	2,6	1,3	11
Total	3	8	4	2	17

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	1 - 1,1	3 - 2,8	1 - 1,4	1 - 0,7
Mujer	2 - 1,9	5 - 5,2	3 - 2,6	1 - 1,3

$F_o - F_e$. (para cada celda)

¿valores negativos? > elevamos al cuadrado para conservar la “idea de distancia” y además se “premia” las distancias grandes sobre las pequeñas.

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	1 - 1,1	3 - 2,8	1 - 1,4	1 - 0,7
Mujer	2 - 1,9	5 - 5,2	3 - 2,6	1 - 1,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Sumamos las diferencias para tener una medida de cuanto se distancia la tabla de observados de la tabla de frecuencias esperadas.

Para sumar tomamos distancias (es decir pasamos valores negativos en positivos, elevando al cuadrado). Elevar al cuadrado además “premia” las distancias grandes sobre las pequeñas.

$\sum (F_o - F_e)^2$ (para todas las celdas)

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Para tener un valor de “diferencia” que NO DEPENDA DE LAS UNIDADES, dividimos cada una de las distancias por el valor de la frecuencia esperada. Así se normaliza la diferencia

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

F_o = Frecuencia observada en cada celda

F_e = Frecuencia esperada en cada celda

Sumamos las diferencias, elevamos ese número al cuadrado y lo dividimos por la F_e . Luego sumamos los valores para cada celda de la tabla

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Para tener un valor de “diferencia” que NO DEPENDA DE LAS UNIDADES, dividimos cada una de las distancias por el valor de la frecuencia esperada. Así se normaliza la diferencia

χ^2 = Chi cuadrado

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

Es una medida de probabilidad de las diferencias entre las frecuencias observadas y las frecuencias esperadas.

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Para tener un valor de “diferencia” que NO DEPENDA DE LAS UNIDADES, dividimos cada una de las distancias por el valor de la frecuencia esperada. Así se normaliza la diferencia

$$\sum \frac{(F_o - F_e)^2}{F_e} = (-0,1^2/1,1) + (0,2^2/2,8) + (-0,4^2/1,4) + (0,3^2/0,7) + (0,1^2/1,9) + (-0,2^2/5,2) + (0,4^2/2,6) + (-0,3^2/1,3) \\ = 0,41$$

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

$\chi^2 = 0$ no hay asociación. "Hipótesis nula"

$\chi^2 > 0$ puede haber asociación

$\chi^2 = 0,41$ ¿implica o no asociación?

Debemos recurrir a la tabla de distribución de χ^2 , ya que el valor depende de "los grados de libertad" de la tabla.

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos	TOTAL
Hombre	1	3	1	1	6
Mujer	2	5	3	1	11
Total	3	8	4	2	17

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

¿Qué son “los grados de libertad” de la tabla.

Dados los valores “marginales” fijos (que es lo que me va a dar la tabla de frecuencias esperadas), es posible tener “libertad” para los valores cierta cantidad la celdas (no todas). En este caso en la fila de “hombres”, fijando el marginal de 6, tengo libertad para 3 de las 4 celdas (la 4a está determinada por las otras 3 y el marginal). Lo mismo vale para las columnas.

Por tanto, los grados de libertad se calculan como el $(nº \text{ de filas} - 1) \times (nº \text{ columnas} - 1)$
 $GL = (r-1) \times (c-1) = (4-1) \times (2-1) = 3 \times 1 = 3$

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

$$\chi^2 = 0,41$$

$$GL = 3$$

Debemos recurrir a la tabla de distribución de χ^2

Consideramos un valor de confianza de 0,05

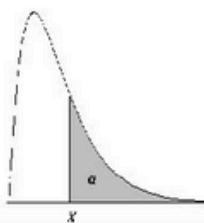
Para que haya asociación, $\chi^2 \geq 7,8147$

Es decir, con una probabilidad de 95% NO HAY ASOCIACIÓN ENTRE GÉNERO Y TIEMPO



Tabla de la distribución chi-cuadrado.

La tabla contiene los valores x tales que $P[\chi^2_n \geq x] = \alpha$
en función de los grados de libertad (n).



n	0,99	0,98	0,975	0,95	0,90	0,80	0,50	0,20	0,10	0,05	0,025	0,02	0,01	0,001
1	0,0002	0,0006	0,0010	0,0039	0,0158	0,0642	0,4549	1,6424	2,7055	3,8415	5,0239	5,4119	6,6349	10,8274
2	0,0201	0,0404	0,0506	0,1026	0,2107	0,4463	1,3863	3,2189	4,6052	5,9915	7,3778	7,8241	9,2104	13,8150
3	0,1148	0,1848	0,2158	0,3518	0,5844	1,0052	2,3660	4,6416	6,2514	7,8147	9,3484	9,8374	11,3449	16,2660
4	0,2971	0,4294	0,4844	0,7107	1,0636	1,6488	3,3567	5,9886	7,7794	9,4877	11,1433	11,6678	13,2767	18,4662
5	0,5543	0,7519	0,8312	1,1455	1,6103	2,3425	4,3515	7,2893	9,2363	11,0705	12,6325	13,3882	15,0863	20,5147
6	0,8721	1,1344	1,2373	1,6354	2,2041	3,0701	5,3481	8,5581	10,6446	12,5916	14,4494	15,0332	16,8119	22,4575
7	1,2390	1,5643	1,6899	2,1673	2,8331	3,8223	6,3458	9,8032	12,0170	14,0671	16,0128	16,6224	18,4753	24,3213
8	1,6465	2,0325	2,1797	2,7326	3,4895	4,5936	7,3441	11,0301	13,3616	15,5073	17,5345	18,1682	20,0902	26,1239
9	2,0879	2,5324	2,7004	3,3251	4,1682	5,3801	8,3428	12,2421	14,6837	16,9190	19,0228	19,6790	21,6660	27,8767
10	2,5582	3,0591	3,2470	3,9403	4,8652	6,1791	9,3418	13,4420	15,9872	18,3070	20,4832	21,1608	23,2093	29,5879
11	3,0535	3,6087	3,8157	4,5748	5,5778	6,9887	10,3410	14,6314	17,2750	19,6752	21,9200	22,6179	24,7250	31,2635
12	3,5706	4,1783	4,4038	5,2260	6,3038	7,8073	11,3403	15,8120	18,5493	21,0261	23,3367	24,0539	26,2170	32,9092
13	4,1069	4,7654	5,0087	5,8919	7,0415	8,6339	12,3398	16,9848	19,8119	22,3620	24,7356	25,4715	27,6882	34,5274
14	4,6604	5,3682	5,6287	6,5706	7,7895	9,4673	13,3393	18,1508	21,0641	23,6848	26,1189	26,8727	29,1412	36,1239
15	5,2294	5,9849	6,2621	7,2609	8,5468	10,3070	14,3389	19,3107	22,3071	24,9958	27,4884	28,2595	30,5780	37,6978
16	5,8122	6,6142	6,9077	7,9818	9,3122	11,1521	15,3385	20,4651	23,5418	26,2962	28,8453	29,6332	31,9999	39,2518
17	6,4077	7,2550	7,5642	8,6718	10,0852	12,0023	16,3382	21,6148	24,7690	27,5871	30,1910	30,9950	33,4087	40,7911
18	7,0149	7,9062	8,2307	9,3904	10,8649	12,8570	17,3379	22,7595	25,9894	28,8693	31,5264	32,3462	34,8052	42,3119
19	7,6327	8,5670	8,9065	10,1170	11,6509	13,7158	18,3376	23,9004	27,2036	30,1435	32,8523	33,6874	36,1908	43,8194
20	8,2804	9,2367	9,5908	10,8508	12,4426	14,5784	19,3374	25,0375	28,4120	31,4104	34,1636	35,0196	37,5863	45,3142
21	8,8972	9,9145	10,2829	11,5913	13,2396	15,4446	20,3372	26,1711	29,6151	32,6706	35,4789	36,3434	38,9322	46,7963
22	9,5425	10,6000	10,9823	12,3380	14,0415	16,3140	21,3370	27,3015	30,8133	33,9245	36,7807	37,6595	40,2894	48,2676
23	10,1957	11,2926	11,6885	13,0905	14,8480	17,1885	22,3369	28,4288	32,0089	35,1725	38,0756	38,9683	41,6383	49,7276
24	10,8563	11,9918	12,4011	13,8484	15,6587	18,0618	23,3367	29,5533	33,1962	36,4150	39,3641	40,2703	42,9798	51,1790
25	11,5240	12,6973	13,1197	14,6114	16,4734	18,9397	24,3366	30,6752	34,3816	37,6525	40,6465	41,5660	44,3140	52,6187
26	12,1982	13,4086	13,8439	15,3792	17,2919	19,8202	25,3365	31,7946	35,5632	38,8851	41,9231	42,8558	45,6416	54,0511
27	12,8785	14,1254	14,5734	16,1514	18,1139	20,7030	26,3363	32,9117	36,7412	40,1133	43,1945	44,1399	46,9628	55,4751
28	13,5647	14,8475	15,3079	16,9279	18,9392	21,5880	27,3362	34,0266	37,9159	41,3372	44,4608	45,4188	48,2782	56,8918
29	14,2564	15,5745	16,0471	17,7084	19,7677	22,4751	28,3361	35,1394	39,0875	42,5569	45,7223	46,6926	49,5878	58,3006
30	14,9535	16,3062	16,7908	18,4927	20,5992	23,3641	29,3360	36,2502	40,2560	43,7730	46,9792	47,9618	50,8922	59,7022

Ejemplo base de datos formulario

¿Existe asociación entre género y tiempo destinado a ir a la universidad?

La prueba chi cuadrado compara lo observado con lo esperado

	16-30 minutos	31-45 minutos	46-60 minutos	Más de 60 minutos
Hombre	-0,1	0,2	-0,4	0,3
Mujer	0,1	-0,2	0,4	-0,3

Comparación = medir las diferencias en cada celda = restar los valores para cada celda (y luego agregarlos)

$F_o - F_e$. (para cada celda)

Hay un 95% de probabilidad de obtener ese valor de Chi-cuadrado para una tabla de contingencia con 3 GL en donde, a nivel poblacional, no existe asociación entre las variables, o sea, en donde la H_0 es cierta.

Es decir, con un 95% de confianza NO HAY ASOCIACIÓN ENTRE GÉNERO Y TIEMPO

Precaución: este es solo un ejemplo que tiene muy pocos casos, y es posible que ello conduzca a errores.

¿Existe asociación?

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

Cuando las frecuencias observadas (F_o) son parecidas a las frecuencias esperadas (F_e) entonces χ^2 va a ser pequeños (suma de distancias pequeñas)

χ^2 en casos de “no asociación” tiene un valor “crítico” que se expresa en una tabla de distribución de χ^2 . Se trata de una distribución de probabilidad, donde los valores dependen del intervalo de confianza y de los grados de libertad.

- Cuando χ^2 observado > χ^2 (esperado, lo que aparece en la tabla) se debe rechazar la H_o y aceptar la H_A
- Cuando χ^2 observado < χ^2 (esperado, lo que aparece) se debe aceptar la H_o y rechazar la H_A

VI. Distribución Chi cuadrado y valor crítico

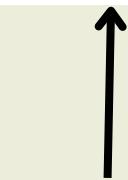
Interpretar a partir del valor crítico de “p”

R Studio y otros programas entregan la información así:

statistic	p	parameter	Method
21.61	< .001***	4	Pearson's Chi-squared test



Valor de X^2
observado



GL (Grados de
Libertad de la
Tabla)

Valor de p (probabilidad de encontrar una
muestra sin asociación con ese valor de X^2)

VI. Distribución Chi cuadrado y valor crítico

Cuando valor-p del χ^2 observado < valor-p “crítico” (de la tabla) = se acepta la H alternativa (hay asociación)

Cuando valor-p del χ^2 observado > valor-p “crítico” (de la tabla) = se rechaza H

Módulo 2 Creación de tablas de contingencia en R y cálculo e interpretación de χ^2

- Realizar la primera parte del ejercicio propuesto por el profesor Kevin Carrasco **con otra base de datos de su elección**. <https://estadisticaiv.netlify.app/practicos/02-content> (hasta antes de “Análisis de correspondencias simple”).
 - Puede consultar **bases de datos en**
<https://www.dropbox.com/scl/fo/qhnirhox1xujcyrbhs8u/Alpo5Ee5uCe0WSutCz5P5-A?rlkey=psit3plh3t8gnbeeixdm6my1y&dl=0>
 - Puede también usar el siguiente documento de ejemplo:
https://www.dropbox.com/scl/fi/s9890rhe9gs0i5chdkta0/20_08_Tabla-de-Contingencia-y-cálculo-de-chi-cuadrado_-2.pdf?rlkey=eoic6fvkrjkwohz01gc8viab8&dl=0
- Para la tabla de contingencia **calcule de manera manual el valor de χ^2** (puede usar calculadora, Excel, R u otro). **Saque una foto** de ese proceso para enviar. Compare el resultado con lo calculado por R Studio.
- Use la tabla de distribución de χ^2 para interpretar el resultado
- **Suba el resultado al buzón de Tareas** de U Campus para obtener 2 décimas en el próximo control (el ejercicio debe estar correctamente realizado).