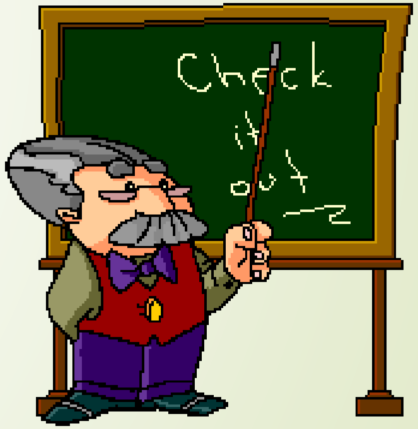
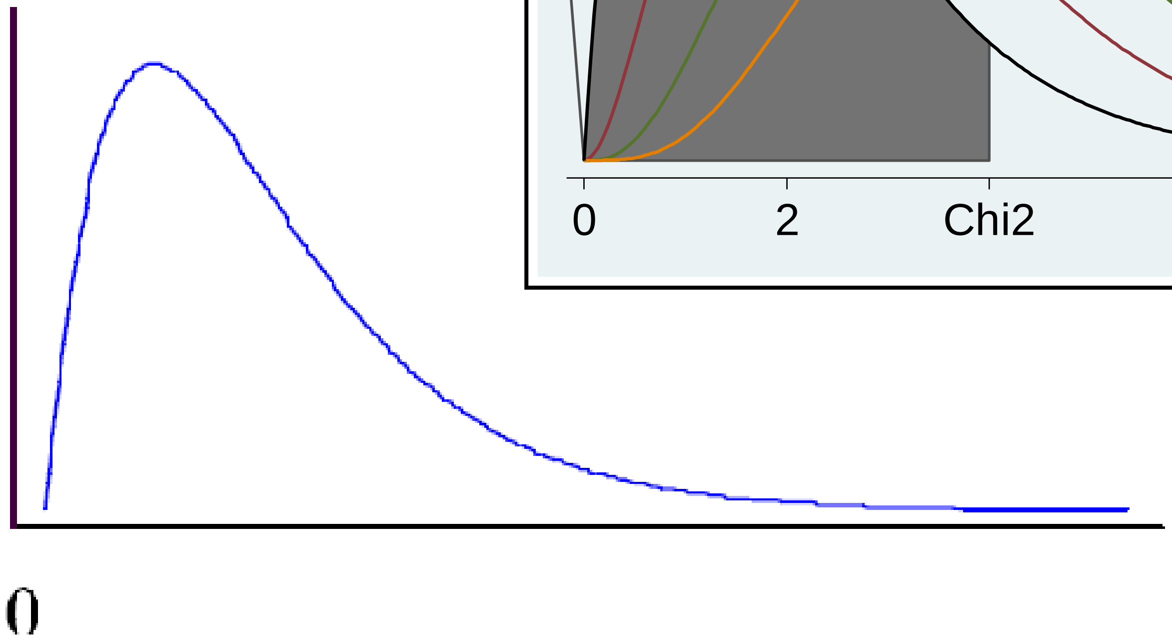
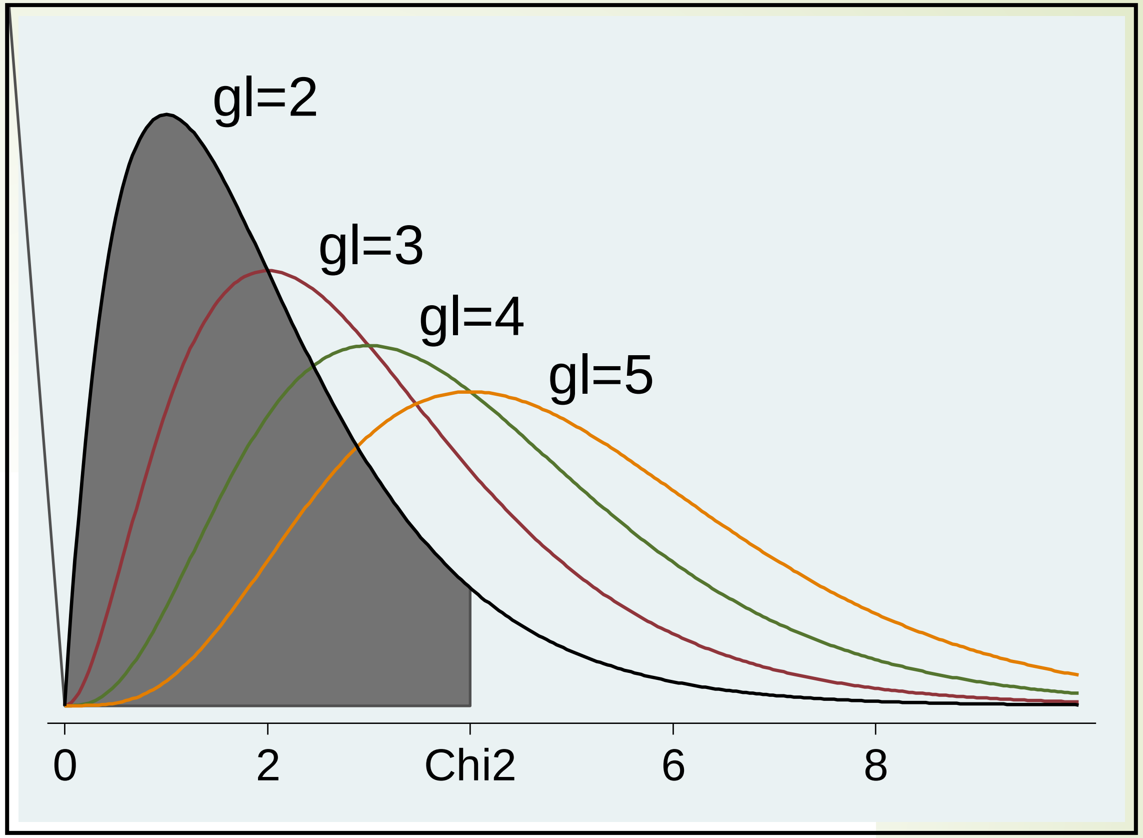
Pruebas de Hipótesis
No Paramétricas



CARACTERÍSTICAS DE LA DISTRIBUCIÓN DE $\chi 2$

- χ^2 es <u>**siempre positivo**</u> porque es una suma de cuadrados.
- Varia desde 0 a ∞ (no tiene valores negativos)
- Familia de distribuciones que dependen de 1 parámetro (**gl=grados de libertad**)





χ^2_n : Chi-cuadrado n grados de libertad

Pruebas de Hipótesis

Pruebas de Hipótesis No Paramétricas

Las pruebas de hipótesis anteriores se llaman paramétricas porque suponen conocida la distribución de la población y la hipótesis es acerca de los parámetros de dicha distribución.

Las Pruebas de hipótesis aplicables a los niveles de medición nominal u ordinal, se denominan Pruebas No Paramétricas ó libres de distribución.

La Estadística No Paramétrica es: una rama de la Estadística que estudia las pruebas y modelos estadísticos, cuya distribución no puede ser definida a priori, pues son los datos observados los que la determinan.

La Prueba de Chi Cuadrado es denotada por la letra griega X^2, es frecuentemente una prueba para probar hipótesis concernientes a la diferencia entre un conjunto de frecuencias observadas de una muestra y un conjunto correspondientes de las secuencias teóricas o esperadas.

Al probar una hipótesis mediante el uso de la distribución X^2, podemos determinar si las diferencias entre los conjuntos de frecuencias son significativas, o si las diferencias son demasiados grandes para ser atribuibles a fluctuaciones de la muestras.

Pruebas de Hipótesis

Pruebas de Hipótesis No Paramétricas

Las pruebas de hipótesis no Paramétricas más utilizadas son:

Bondad de Ajuste

Homogeneidad

Independencia

Pruebas de Hipótesis

Prueba Ji^2 de la Bondad del Ajuste

Se parte de una muestra aleatoria de tamaño n, proveniente de una población cuya distribución de probabilidad es desconocida.

La hipótesis nula indica que los **datos se ajustan** a una distribución (binomial, Poisson ,normal ,hipergeométrica ,cualquiera de los estudiados, u otro diseñado).

Las n observaciones se acomodan en una tabla de frecuencias con k intervalos de clase.

Sea O_i la i-ésima frecuencia de clase.

De la distribución de probabilidad propuesta se calcula la frecuencia esperada E_i en el i-ésimo intervalo de clase

El estadístico de prueba es

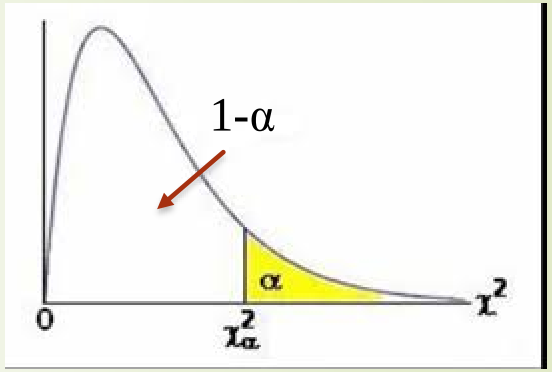
$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

El cual tiene una distribución Ji^2 con k-1-r grados de libertad si la población sigue la distribución propuesta. (donde r es el número de parámetros de la población a estimar)

Pruebas de Hipótesis

Prueba Ji^2 de la Bondad del Ajuste

- Conforme aumenta el número de grados de libertad, la distribución se aproxima a la de tipo normal.
- La aproximación mejora a medida que n es más grande.



- La hipótesis debe rechazarse si el valor del estadístico de prueba es

$$\chi^2 > \chi^2_{1-\alpha,k-1-r}$$

- En la hipótesis se plantea que los datos se ajustan a una distribución conocida.
- Precaución: Si las frecuencias esperadas son muy pequeñas el estadístico χ^2 no reflejará el alejamiento entre lo observado y lo esperado. (Se considera que valores menores de 5 son pequeños)

- Si en una prueba resultan frecuencias esperadas pequeñas, se pueden combinar intervalos de clase adyacentes para aumentar estos valores, ya que no es necesario que los anchos de clase sean del mismo tamaño.

Pruebas de Hipótesis

Prueba de Bondad del Ajuste

Ejemplo 1.- Se desea saber si un dado cúbico está perfectamente balanceado. Con este fin se arroja 300 veces el dado. Si las frecuencias observadas en éstas seis categorías son:35,40, 32, 60, 68, y 65 respectivamente. ¿Debería llegarse a la conclusión de que el dado está perfectamente balanceado? Utilice α=0,01

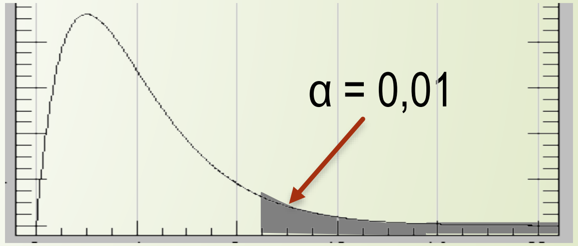
1.-
H_0: El dado está perfectamente balanceado
H_1: El dado no está perfectamente balanceado

2.- $\alpha = 0,01$

3.- $$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Grados de libertad: k-1-r= 6- 1- 0 =5

4.-



$\chi^2_{5,\,0,99} = 15,95$

Regla de decisión

Rechazo H_0 si χ^2cal ≥15,95
No rechazo H_0 si χ^2cal <15,95

5.- Cálculos

Resultados posibles o categorías	O_i Frecuencias observadas	Probabilidad	$E_i = n.p$ Frecuencias esperadas	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	35					
2	40					
3	32					
4	60					
5	68					
6	65					
Total	300					

5.- Cálculos

Resultados posibles o categorías	O_i Frecuencias observadas	Probabilidad	$E_i=n.p$ Frecuencias esperadas	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	35	1/6	300.1/6=50	-15	225	4,5
2	40	1/6	50	-10	100	2
3	32	1/6	50	-18	324	6,48
4	60	1/6	50	10	100	2
5	68	1/6	50	18	324	6,48
6	65	1/6	50	15	225	4,5
Total	**300**	**1**	**300**			**25,96**

6.- Como χ^2cal >15,95 **se rechaza H_0**

Conclusión: Con un nivel de significación de 0,01 tengo evidencias para rechazar la hipótesis nula y suponer que el dado no está perfectamente balanceado.

Pruebas de Hipótesis

Prueba de Bondad del Ajuste

Ejemplo 2.- Se sabe que el número de defectos en determinados tensiómetros sigue una distribución de Poisson. Se reúne una muestra aleatoria de 60 tensiómetros y se observa el número de defectos. Los resultados obtenidos son los siguientes:

Número de defectos	Frecuencia observada
0	32
1	15
2	9
3 ó más	4

¿Muestran estos datos suficiente evidencia para decir que provienen de una distribución Poisson?. Realice la prueba correspondiente con α= 0,05.

Pruebas de Hipótesis

Prueba de Bondad del Ajuste

Ejemplo 2.-

1.- H_0: La distribución de los defectos en los tensiómetros se ajusta a la distribución de Poisson.
H_1: La distribución de los defectos en los tensiómetros no se ajusta a la distribución de Poisson

2.- α = 0,05

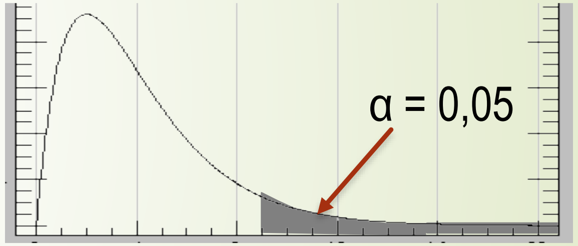
3.-

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Grados de libertad: k-1-r= ….- 1- 1 =…….

Dejamos pendiente el calculo de grados de libertad ya que se deben corregir las categorías si se observan valores esperados menores o iguales que 5.

4.-



$$\chi^2_{\dots;\,0,95} = \dots\dots$$

Regla de decisión

Rechazo H_0 si χ^2cal ≥…..
No rechazo H_0 si χ^2cal <……

Ejemplo 2.-continuación

La media de la distribución Poisson propuesta en este ejemplo es desconocida y debe estimarse a partir de los datos contenidos en la muestra.

$$\lambda = \frac{(32)(0) + (15)(1) + (9)(2) + (4)(3)}{60} = 0{,}75$$

A partir de la distribución Poisson con parámetro 0,75, pueden calcularse las probabilidades asociadas con cada valor de X, utilizando:

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-0{,}75}0{,}75^{x}}{x!}$$

Se calculan las probabilidades y se multiplican por 60 para obtener los valores esperados. Por ejemplo, para encontrar:

$$\mathrm{E(0) = n.p(0)} = 60 * 0{,}472 = 28{,}32$$

Ejemplo 2.-continuación

Número de defectos	Probabilidad	Frecuencia observada		Frecuencia esperada	
0	0,472	32		28,32	
1	0,354	15		21,24	
2	0,133	9	13	7,98	10,44
3 (o más)	0,041	4		2,46	
Total		60		60	

Observando la tabla de distribución de probabilidad de la variable, se tiene que la suma de las frecuencias esperadas coincide con la suma de las frecuencias observadas.
Si la frecuencia esperada en alguna celda es menor que 5, se combinan celdas hasta lograr "5" o más.

Número de defectos	Frecuencia observada	Frecuencia esperada
0	32	28,32
1	15	21,24
2 ó más	13	10,44

En este caso, los grados de libertad serán:
K -1 -r = **3** - 1 - 1= 1, debido a que la media de la distribución Poisson λ fue estimada a partir de los datos.

$$\chi^2 = \sum_{j=1}^{n} \frac{(O_j - E_j)^2}{E_j}$$

$$\chi^2 calculado = \frac{(32-28{,}32)^2}{28{,}32} + \frac{(15-21{,}24)^2}{21{,}24} + \frac{(13-10{,}44)^2}{10{,}44} = 2{,}94$$

Ejemplo 2.- Continuación

1.- H_0: La distribución de los defectos en los tensiómetros se ajusta a la distribución de Poisson.
H_1: La distribución de los defectos en los tensiómetros no se ajusta a la distribución de Poisson

2.- $\alpha = 0,05$

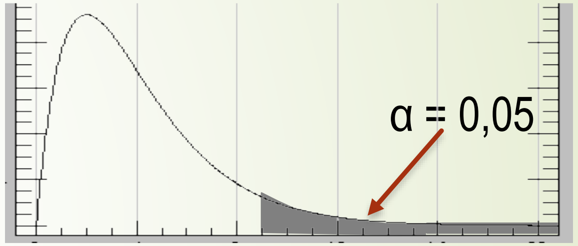
3.-

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Grados de libertad: k -1 - r = 3- 1- 1 = 1.

Dejamos pendiente el calculo de grados de libertad ya que se deben corregir las categorías si se observan valores esperados menores o iguales que 5.

4.-



$$\chi^2_{1;\,0,95} = 3,84$$

<u>Regla de decisión</u>

Rechazo H_0 si χ^2cal $\geq 3,84$
No rechazo H_0 si χ^2cal $< 3,84$

6.- Como χ^2cal <3,84 **no se rechaza H_0**

<u>Conclusión</u>: con un nivel de significación α= 0,05 no tengo evidencias suficientes para rechazar la hipótesis nula, por lo que se puede decir que la distribución de los defectos en los tensiómetros se ajusta a la distribución de Poisson con λ= 0,75

Pruebas de Hipótesis

Prueba de Independencia

La ***Prueba de Independencia*** es: “una prueba estadística de proporciones de frecuencias que se utiliza para determinar si la pertenencia de una variable a categorías es diferente como función de la pertenencia a la categoría de una segunda variable”.

Al probar una hipótesis se involucra una tabla de contingencia, primero se calcula las correspondientes frecuencias esperadas o teóricas de acuerdo con la hipótesis. La suma de todas las frecuencias esperadas debe ser igual a las sumas de todas las frecuencias observadas.

Las tablas de contingencia son usadas frecuentemente en pruebas de independencia. Este tipo de pruebas nos dirá si son o no independientes (o no relacionadas), las dos bases de clasificación usadas respectivamente hileras y columnas.

Tablas de Contingencia

Una tabla de contingencia es una herramienta que nos permite poner a prueba si *dos criterios de clasificación de una misma muestra* son independientes o no. Ejemplo de tabla:

Hepatitis	**VACUNACIÓN**		**Total**
	Si	**No**	
Si	11	70	81
No	538	464	1002
Total	549	534	1083

Tablas de Contingencia: Independencia

		VARIABLE DE CLASIFICACIÓN COLUMNAS					
		1	**2**	**3**	**..........**	**c**	**Total**
VARIABLE CLASIFICACIÓN FILAS	**1**	O_{11}	O_{12}	O_{13}	....	O_{1c}	r_1
	2	O_{21}	O_{22}	O_{23}	....	O_{2c}	r_2
	3	O_{31}	O_{32}	O_{33}	....	O_{3c}	r_3
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
	r	O_{r1}	O_{r2}	O_{r3}	....	O_{rc}	r_r
	Total	c_1	c_2	c_3	....	c_c	n

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

$$E_{ij} = n \frac{c_j}{n} \frac{f_j}{n}$$

$$gl = (f - 1) \times (c - 1)$$

Para recorrer una tabla de contingencia la distribución chi cuadrado debe recorrerla por lo que será la siguiente:

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

r = número de categorías de la variable en las filas
c = número de categorías de la variable en las columnas
Oij = número observado en entrada ij
Eij = número esperado en la entrada ij

Este estadístico Chi-cuadrado tiene (f-1)*(c-1) grados de libertad.
Para calcular el valor esperado se utiliza en concepto de independencia que es el que se plantea en la hipótesis nula.

Tablas de Contingencia

Una tabla de contingencia es una herramienta que nos permite poner a prueba si *dos criterios de clasificación de una misma muestra* son independientes o no. Ejemplo de tabla:

Hepatitis	VACUNACIÓN		Total
	Si	No	
Si	11	**70**	81
No	538	464	1002
Total	549	534	1083

grados de libertad=(f-1)*(c-1) =(2-1) (2-1)=1

Para entender los grados de libertad: consideremos los números que están dentro de la tabla, son 4, si elegimos en valor que está marcado en negrita (**70**), conociendo los totales de filas y columnas, por diferencia se calculan los demás valores. De allí es que sólo se puede elegir libremente un valor.

Pruebas de Hipótesis

Ejemplo: Prueba de Independencia

Un investigador quiere probar si la escolaridad de una persona está relacionada con su clasificación de desempeño en el trabajo. Para ello obtiene una muestra aleatoria de 100 a partir de la población de todos los empleados. Los datos se muestran en la siguiente tabla:

Desempeño	Escolaridad		Total
	Universitarios	No Universitarios	
Bueno	15	15	30
Malo	25	45	70
Total	40	60	100

Contraste la hipótesis utilizando $\alpha = 0{,}05$.

Pruebas de Hipótesis

Prueba de Independencia

Ejemplo 1.-

1.- H_0: El desempeño en el trabajo **es independiente** de la escolaridad
H_1: El desempeño en el trabajo no es independiente de la escolaridad.

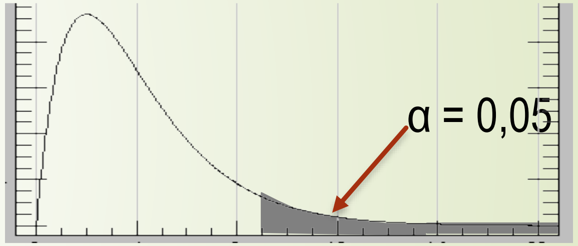
2.- $\alpha = 0{,}05$

3.-

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

Grados de Libertad= (f-1)*(c-1)=(2-1)(2-1)=1

4.-



$\chi^2_{1;\,0{,}95} = 3{,}84$

<u>Regla de decisión</u>

Rechazo H_0 si $\chi^2 cal \geq 3{,}84$

No rechazo H_0 si $\chi^2 cal < 3{,}84$

5.- Cálculos

Desempeño	Escolaridad		Total
	Universitarios	No Universitarios	
Bueno	15	15	30
Malo	25	45	70
Total	40	60	100

	O_i Frecuencias observadas	$E_i = n\ .\ p$ Frecuencias Esperadas
Bueno(B) y Universitario (U)	15	
Malo (M) y Universitario (U)	25	
Bueno (B) y no Universitario (NU)	15	
Malo (M) y no Universitario (NU)	45	
Total	**100**	

E_{11}= n. P(bueno ∩ Universitario)

5.- Cálculos

	O_i Frecuencias observadas	$E_i = n \, . \, p$ Frecuencias Esperadas
Bueno(B) y Universitario (U)	15	$100.\frac{30}{100}\frac{40}{100} = 12$
Malo (M) y Universitario (U)	25	$100.\frac{70}{100}\frac{40}{100} = 28$
Bueno (B) y no Universitario (NU)	15	$100.\frac{30}{100}\frac{60}{100} = 18$
Malo (M) y no Universitario (NU)	45	$100.\frac{70}{100}\frac{60}{100} = 42$
Total	**100**	**100**

E_{11}= n. P(bueno ∩ Universitario)
Por ser sucesos independientes
P(B∩U)=P(B).P(U)=30/100.40/100
P(B∩U)=0,12
E_{11}= 100.0,12=12

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

$$\chi^2 calculado = \frac{(15-12)^2}{12} + \frac{(25-28)^2}{28} + \frac{(15-18)^2}{18} + \frac{(45-42)^2}{42} = 1{,}78$$

6.- Como χ^2cal <3,84 **no se rechaza H_0**

<u>Conclusión</u>: Con un nivel de significación del 5% no tengo evidencias suficientes para rechazar la hipótesis nula del que el desempeño en el trabajo (Bueno-Malo) es independiente de la escolaridad (Universitario-No Universitario).

Pruebas de Hipótesis

Prueba de Homogeneidad

Otro caso en que usamos una tabla de contingencia es para realizar la prueba de Homogeneidad. En esta prueba se comparan dos o más poblaciones independientes. En cada una de estas poblaciones se toma una muestra, y los individuos de la misma se clasifican según una variable *Y* que puede tomar *m* valores posibles y_1, y_2.....y_m.,es decir que una de las dos variables involucradas es controlada por el investigador, quedando de este modo los totales de filas o de columnas predeterminados.

Se cuenta con r poblaciones x_1, x_2,...,x_r. Sea p_{ij} la proporción de individuos que, en la población x_i tiene como valor de $Y=y_j$.

Pruebas de Hipótesis

Prueba de Homogeneidad

Un *contraste de homogeneidad* se utiliza cuando se desean contrastar las dos hipótesis siguientes:

H_0: $p_{1j} = p_{2j} = \ldots\ldots = p_{mj}$ para todo *j;* dicho de otro modo, todas las poblaciones tienen idéntica distribución para la variable Y.

H_1: algunas de estas proporciones son diferentes. Dicho de otro modo, la distribución de la variable Y en alguna de estas poblaciones es diferente.

El principal objetivo de realizar este contraste es comprobar que las distribuciones de todas las poblaciones son iguales o si hay alguna que difiere. Es decir comprobar que los datos de las distintas muestras son homogéneos.

Para recorrer una tabla de contingencia la distribución chi cuadrado debe recorrerla por lo que utilizaremos la misma variable pivotal o estadístico de la prueba que en el caso de independencia y será la siguiente:

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

r = número de categorías de la variable en las filas
c = número de categorías de la variable en las columnas
Oij = número observado en entrada ij
Eij = número esperado en la entrada ij

Este estadístico Chi-cuadrado tiene (f-1)*(c-1) grados de libertad (Igual que en independencia).
El valor esperado se calcula de diferente manera ya que está fijado de antemano o el total de filas o el de columnas.

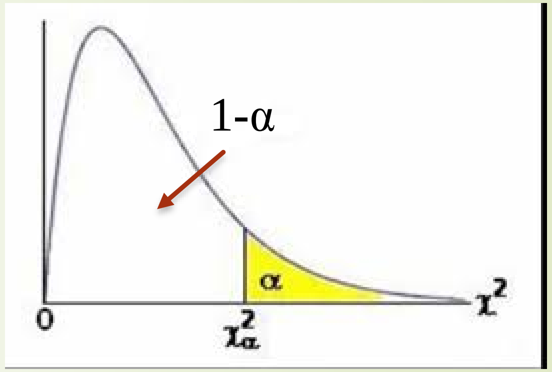
Pruebas de Hipótesis

Homogeneidad

- La región crítica se ubica a la derecha, ya que la variable pivotal o estadístico de la prueba es el mismo que en independencia.

La hipótesis debe rechazarse si el valor del estadístico de prueba es

$$\chi^2 > \chi^2_{1-\alpha,(f-1)(c-1)}$$



- Precaución: Si las frecuencias esperadas son muy pequeñas el estadístico χ^2 no reflejará el alejamiento entre lo observado y lo esperado. (Se considera que valores menores de 5 son pequeños)

Pruebas de Hipótesis

Ejemplo: Prueba de Homogeneidad

Se desea saber si la distribución de los grupos sanguíneos es similar en los individuos de dos poblaciones. Para ello se elige una muestra aleatoria de cada una de ellas, obteniéndose los siguientes datos ¿Qué decisión se debe tomar? Use $\alpha=0,05$

	Grupos sanguíneos				
	A	B	AB	0	**Total**
Muestra Población 1	90	80	110	20	**300**
Muestra Población 2	200	180	240	30	**650**
Total	**290**	**260**	**350**	**50**	**950**

Pruebas de Hipótesis

Prueba de Homogeneidad

Ejemplo.-

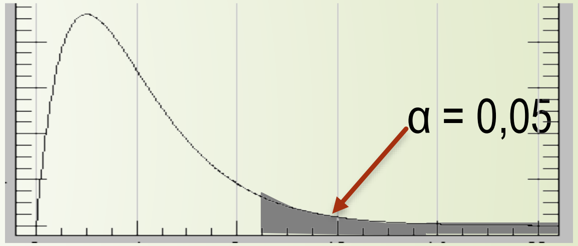
1.- H_0: La proporción de los diferentes grupos sanguíneos (A,B, AB y 0) es la misma en las dos poblaciones
H_1: La proporción de los diferentes grupos sanguíneos (A,B, AB y 0) no es la misma en las dos poblaciones

2.- $\alpha = 0{,}05$

3.-
$$\chi^2 = \sum_{i=1}^{f}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

Grados de Libertad= (f-1)*(c-1)=(2-1)(4-1)=3

4.-



$\chi^2_{3;\,0{,}95} = 7{,}81$

Regla de decisión

Rechazo H_0 si $\chi^2 cal \geq 7{,}81$

No rechazo H_0 si $\chi^2 cal < 7{,}81$

5.- Cálculos

	Grupos sanguíneos				Total
	A	B	AB	0	
Muestra Población 1	90	80	110	20	**300**
Muestra Población 2	200	180	240	30	**650**
Total	**290**	**260**	**350**	**50**	**950**

	Valores esperados			
	A	B	AB	0
Muestra población 1	91,5789	82,105	110,53	15,789
Muestra población 2	198,421	177,89	239,47	34,21

$$E_{11}=300.\frac{290}{950}=91{,}5789$$

$$E_{12}=300.\frac{260}{950}=82{,}105$$

$$E_{13}=300.\frac{350}{950}=110{,}53$$

$$E_{14}=300.\frac{50}{950}=15{,}78$$

$$E_{21}=650.\frac{290}{950}=198{,}421$$

$$E_{22}=650.\frac{260}{950}=177{,}89$$

$$E_{23}=650.\frac{350}{950}=239{,}47$$

$$E_{24}=650.\frac{50}{950}=34{,}21$$

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

$$\chi^2 calculado=\frac{(90-91{,}6)^2}{91{,}6}+\frac{(80-82{,}1)^2}{82{,}1}+\frac{(110-110{,}5)^2}{110{,}5}+\frac{(20-15{,}8)^2}{15{,}8}+\frac{(200-198{,}4)^2}{198{,}4}+\frac{(180-177{,}9)^2}{177{,}9}+\frac{(240-239{,}4)^2}{239{,}4}+\frac{(30-34{,}2)^2}{34{,}2}=1{,}76$$

6.- Como χ^2cal <7,81 **no se rechaza H_0**

Conclusión: Con un nivel de significación del 5% no hay evidencias suficientes para rechazar la hipótesis nula de que la distribución de los grupos sanguíneos (A,B,AB y 0) es similar en los individuos de dos poblaciones.

Gracias