UNIVERSIDAD DE ARTEMISA

ESTADÍSTICA II

MSc. Rodolfo Pérez Léon
email: rperezleon@uart.edu.cu
rplrpl55@gmail.com

UNIVERSIDAD DE ARTEMISA

Tema 2. Prueba no paramétrica chi-cuadrado.

Sumario.

2.1. Introducción. Antecedentes históricos de la prueba Chi-cuadrado
2.2. Distribución Chi-Cuadrado. Características.
2.3. Prueba Chi-Cuadrado. Tablas de contingencia.
2.4. Prueba no paramétrica de independencia de variables.
2.5. Prueba de independencia hi-Cuadrado con App Estadística.
2.6. Prueba de homogeneidad o de proporcionalidad de varias muestras

Objetivos: Que el alumno se capaz de:

Aplicar las pruebas Chi-Cuadrado en la solución de ejercicios de la profesión.

Método: Expositivo.

Medios: Tiza y pizarra, Libro de texto, MDD.

Forma de enseñanza: Conferencia

Bibliografía:

- Probabilidades y Estadística para ingenieros. J.E Freund, I.R. Miller, R. Jonhnson.
- Materiales Didácticos Digitales entregados por el profesor

2.1. Introducción, antecedentes históricos

2.1. Introducción, antecedentes históricos de la distribución Chi-Cuadrado

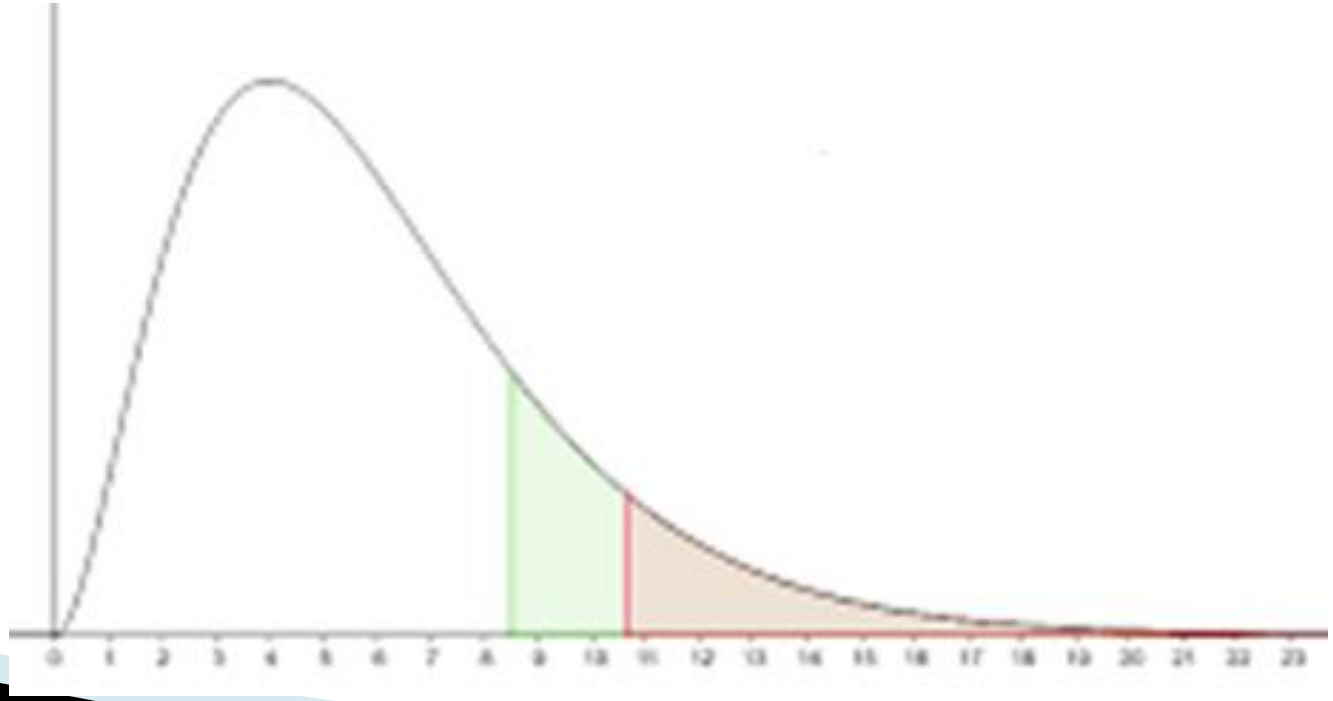
- El matemático Karl Pearson (1857−1936)
- Karl Pearson en 1900 propuso uno de los primeros Test Estadísticos que desde la óptica de las distribuciones de la probabilidad sirve para calcular si los resultados estadísticos de un experimento se alejan significativamente o no de los resultados esperados del modelo teórico, test que actualmente es conocido como el «Test Chi Cuadrado».
- se denota mediante la letra griega minúscula ji elevada al cuadrado (χ^2)

2.2. Distribución CHI CUADRADO. Características

- La distribución Chi-Cuadrado es una distribución asimétrica

- Sólo toma valores positivos y es asintótica con respecto al eje de las x positivas ($0 < \chi_2 < +\infty$)

- El área comprendida entre la curva y el eje de las x es 1 ó 100%



- Las pruebas Chi-Cuadrado se utilizan para determinar si las frecuencias observadas (O) en las categorías difieren significativamente de las frecuencias esperadas (E).

- No se utiliza en números solo se categoriza. (la prueba Chi-cuadrada no considera relaciones causales).

2.3. Prueba CHI CUADRADO
Tablas de Contingencia (TC)

¿QUÉ ES LA PRUEBA CHI CUADRADA?

- Es una prueba estadística para evaluar hipótesis acerca de la relación entre dos variables categóricas, también para evaluar si hay homogeneidad o igual proporción entre diferentes muestras de esas variables

- Se simboliza por $\chi 2$

- Hipótesis a probar : independencia o igual proporciones
  (H_0 : son independientes y H_1 : no son independientes)
  (H_0 : $p_1 = p_2 = \ldots p_n$ H_1 : $p_1, p_2, \ldots p_n$ no son iguales)

- Variables involucradas: 2, esta prueba no considera relaciones causales

- Nivel de medición de las variables: Nominal u ordinal organizadas en una Tabla de Contingencia

EJEMPLO DE UNA TABLA DE CONTINGENCIA 2x2

Dos variables : voto y sexo

Cada variable con dos categorías

		VOTO	
		candidato a	candidato b
sexo	masculino		
	femenino		

PROCEDIMIENTO

- Se calcula a través de una tabla de contingencia o tabulación cruzada.
- Es una tabla de dos dimensiones y cada dimensión contienen una variable
- Cada variable se subdivide en dos o más categorías.

Se utiliza

- **PRUEBA DE BONDAD DE AJUSTE:** compara la distribución de una muestra con la distribución teórica que se supone que describe a la población de la cual se extrajo.
- **PRUEBA DE INDEPENDENCIA DE DOS VARIABLES:** comprueba si dos o más variables que se miden en una misma muestras están o no relacionadas entre sí. Siempre hablamos de variables cualitativas.
- **PRUEBA DE HOMOGENEIDAD:** comprueba si dos o más muestras pertenecen a la misma población.

2.4. Pruebas no paramétrica de independencia de dos variables

Ejemplo 1. Independencia de dos variables (1 muestra)

Un taller de una industria azucarera fabrica piezas metálicas para la zafra y desea evaluar si existe una relación entre el turno de producción (mañana, tarde y noche) y la calidad de las piezas producidas. Para ello, **se seleccionó una muestra** de 1040 piezas y se clasificaron según el turno de fabricación y la calidad de las piezas, obteniendo la siguiente tabla de contingencia de 3x2:

Turnos categorías: mañana, tarde, noche.

Calidad de las piezas categorías: no defectuosas, defectuosas.

TABLA DE CONTINGENCIA

		calidad de las piezas	
		no defectuosas	**defectuosa**
turnos	**mañana**	180	100
	tarde	190	280
	noche	170	120

Prueba de hipótesis X^2: (Probar si hay relación entre las variables)

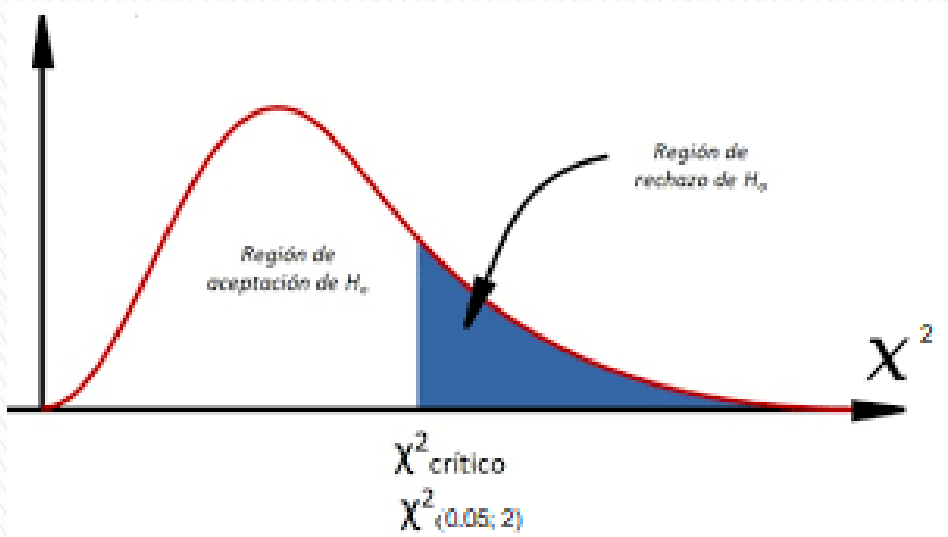
1. Planteamiento de la Hipótesis

H_0: No existe relación entre el turnos y la calidad de las piezas (son independiente).

H_1: Existe relación entre el turnos y la calidad de las pieza (no son independiente).

2. Nivel de significación

α = 0.05



3. Regla de decisión

Si $X^2_{calculada} > X^2_{tabla}$ se rechaza (H0)

4. Calcular $X^2_{calculada}$ y buscar $X^2_{critica}$

4. Calculo $X^2_{calculada}$

Tabla de frecuencias observadas (O):

	calidad de las piezas			TOTAL
		no defectuosas	**defectuosas**	
turnos	**mañana**	180	100	**280**
	tarde	190	280	**470**
	noche	170	120	**290**
	TOTAL	**540**	**500**	**1040**

La Chi-cuadrada es una comparación entre las tablas de **frecuencias observadas** y la denominada tabla de **frecuencias esperadas (**la tabla que esperaríamos encontrar si las variables fueran estadísticamente independientes o no estuvieran relacionadas**)**.

4. Calculo $X^2_{calculada}$

Tabla de frecuencias esperadas (E): Fe= Tf x Tc/ Tg

La **frecuencia esperada** de cada celda, casilla o recuadro, se calcula mediante la siguiente fórmula aplicada a la tabla de frecuencias observadas:

N = es el número total de frecuencias observadas.

Fe = (marginal de la fila)(marginal de columna) / N.

		calidad de las piezas		
		no defectuosas	**defectuosas**	Marginal de filas
turnos	**mañana**	**(280) (540) / 1040**	**(280) (500) / 1040**	**280**
	tarde	**(470) (540) / 1040**	**(470) (500) / 1040**	**470**
	noche	**(290) (540) / 1040**	**(290) (500) / 1040**	**290**
marginal de columnas		**540**	**500**	**1040**

4. Calculo $X^2_{calculada}$

Frecuencia observada:

		calidad de las piezas		TOTAL
		no defectuosas	defectuosas	
turnos	mañana	180	100	**280**
	tarde	190	280	**470**
	noche	170	120	**290**
	TOTAL	**540**	**500**	**1040**

Frecuencia esperada:

		calidad de las piezas		TOTAL
		no defectuosas	defectuosas	
turnos	mañana	145,4	134,6	**280**
	tarde	244,0	226,0	**470**
	noche	150,6	139,4	**290**
	TOTAL	**540**	**500**	**1040**

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Donde:
O i: frecuencia observada en cada celda
E i: frecuencia esperada en cada celda

Calculo $X^2_{calculada}$

$$X^2 = \frac{\sum (O - E)^2}{E}$$

Celda	O	E	O-E	$(O-E)^2$	$(O-E)^2 / E$
Mañana / no defectuosa	180	145,4	34,6	1197,16	8,23
tarde/ no defectuosas	190	244,4	-54,4	2959,36	12,11
noche / no defectuosas	170	150,6	19,4	376,36	2,50
mañana / defectuosa	100	134,6	-34,6	1197,16	8,69
tarde / defectuosa	280	226,0	54,0	2916,00	12,80
noche / defectuosa	120	139,4	-19,4	376,36	2,70
				X^2 =	**47,33**

4. Buscar $X^2_{crítica}$

Se comienza por calcular los grados de libertad.

G.L. = (Nº de filas - 1)(Nº de columnas - 1).

Para el ejemplo: **Nº de filas** = 3 y **Nº de columnas** = 2; entonces G.L. = (3-1)(2-1) = 2.

Luego, acudimos a la “tabla de distribución de Chi-cuadrado”, eligiendo nuestro nivel de significación (α = 0,05).

TABLA 3-Distribución Chi Cuadrado χ^2

P = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, ν = Grados de Libertad

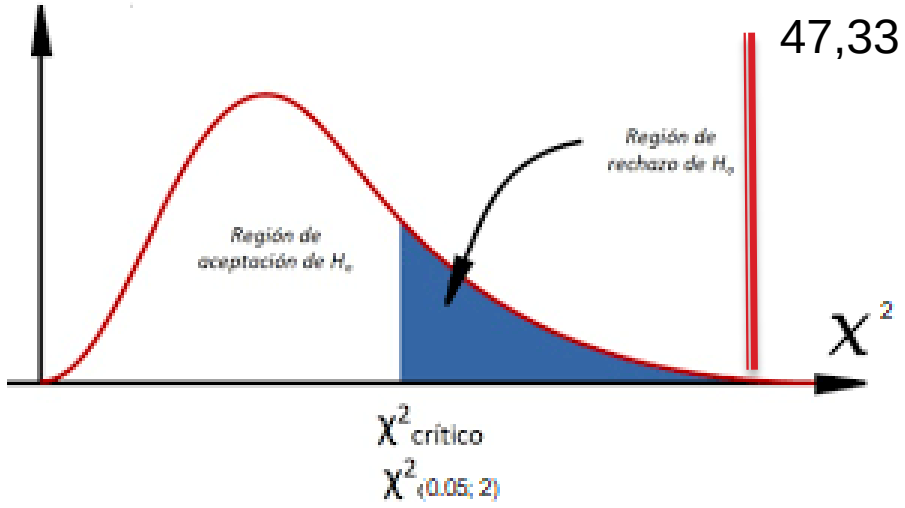
ν/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7007	12,8987	12,1836
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993

Aplicación:

Para el nivel de significación de α = **0,05** y **g.l. = 2**, el X^2 de tabla es **5,9915** (ver tabla).

$$X^2_{caculado} = 47{,}33$$

$$X^2_{tabla} = 5{,}9915$$



5,9915

Prueba de hipótesis X^2: (Probar si hay relación entre las variables)

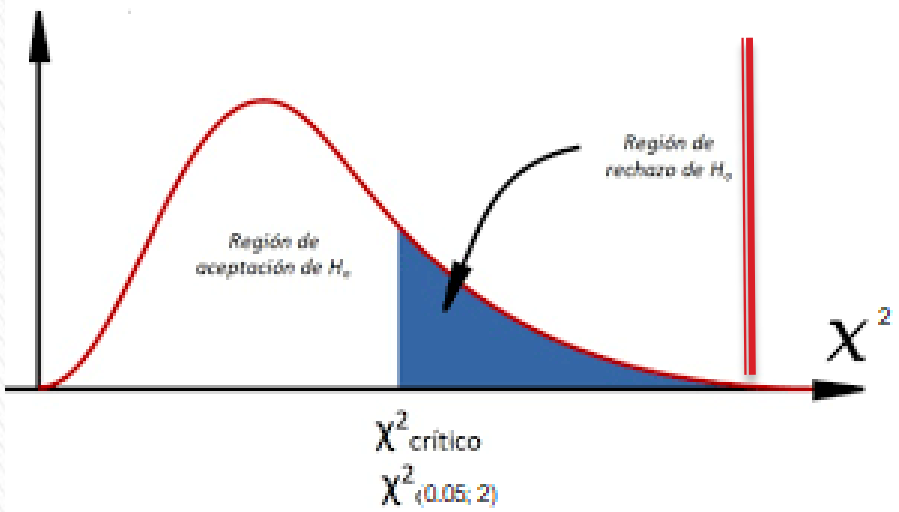
Planteamiento de la Hipótesis

H_0: No existe relación entre el turnos y la calidad de las piezas.

H_1: Existe relación entre el turnos y la calidad de las piezas.

Nivel de significación

$\alpha = 0.05$



Regla de decisión

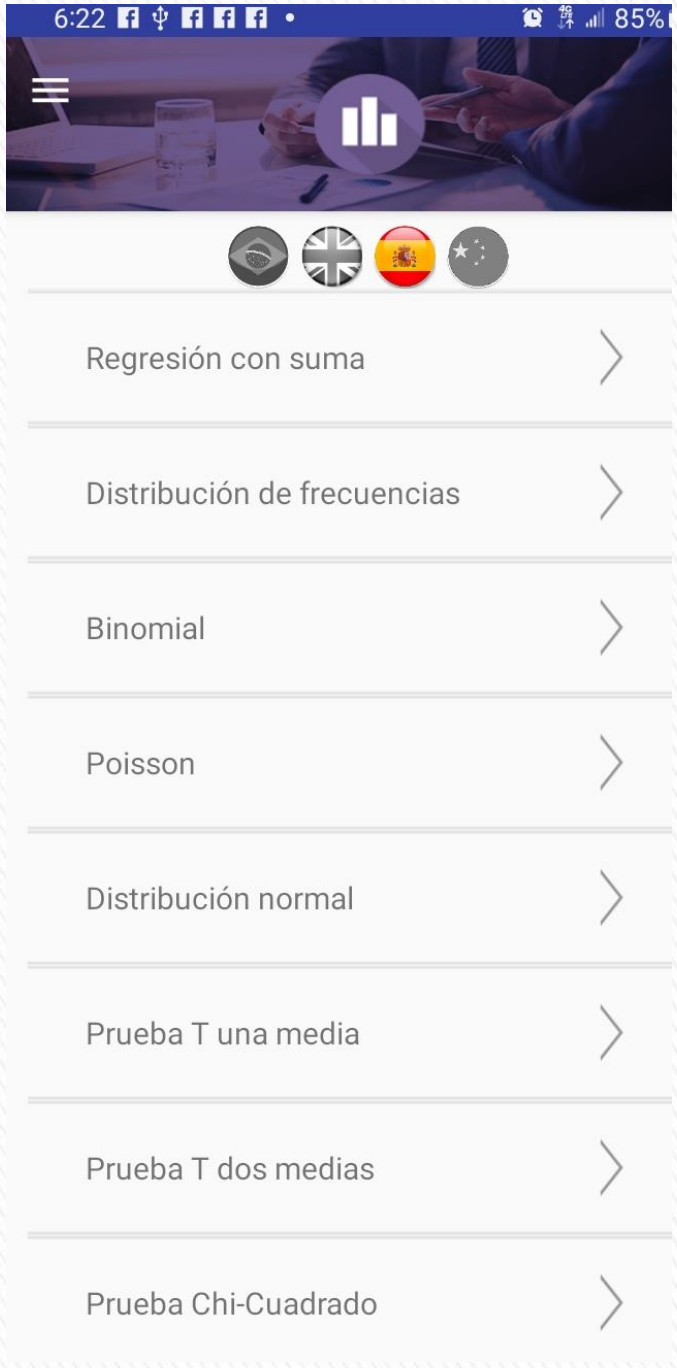
Si $X^2_{calculada} > X^2_{tabla}$ se rechaza (H_0)

Calculo $X^2_{calculada}$

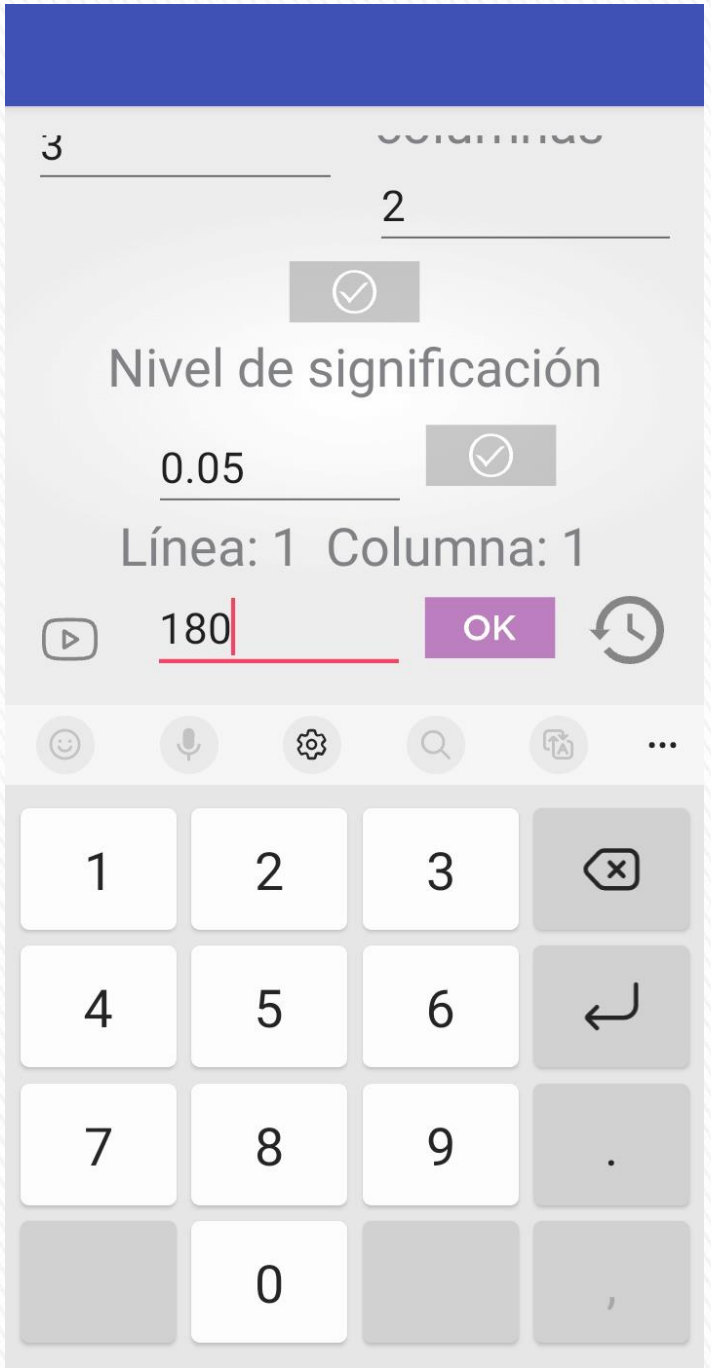
5,9915

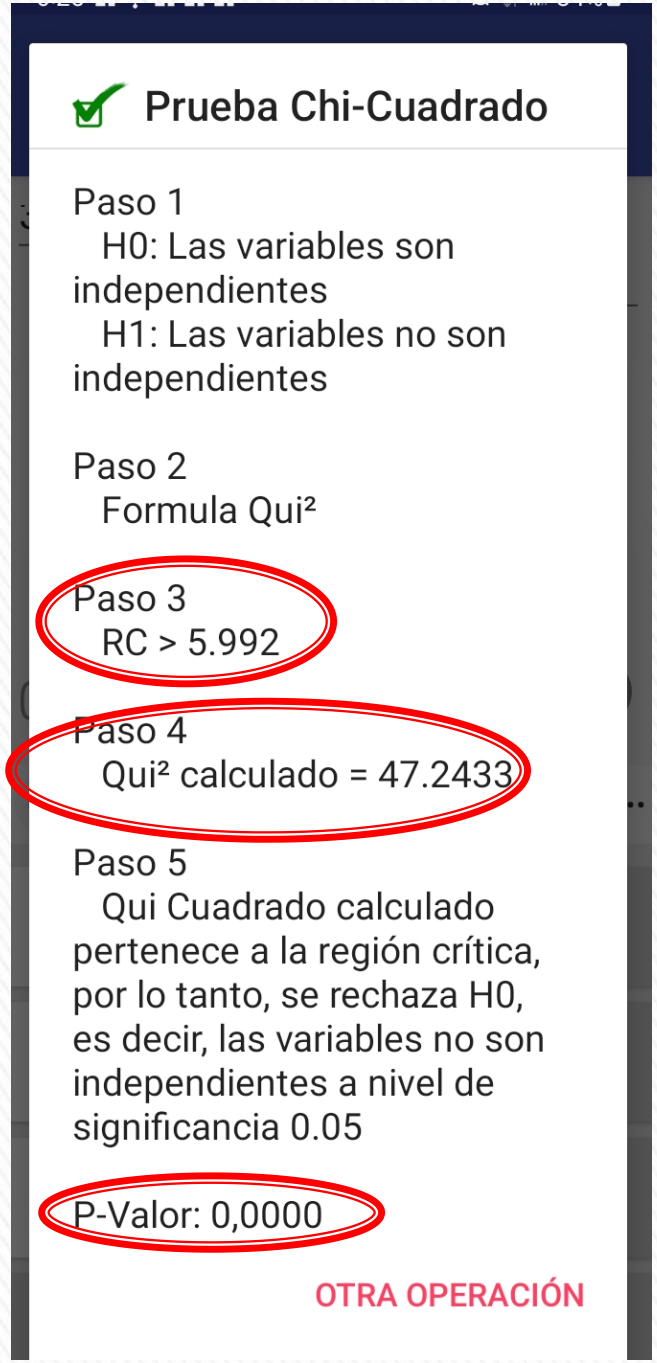
47.33 > 5,9915 Se rechaza Ho R/ Existe relación entre las variables

2.5. Prueba de independencia CHI CUADRADO con la App ESTADÍSTICA


6:22
85%
Regresión con suma
Distribución de frecuencias
Binomial
Poisson
Distribución normal
Prueba T una media
Prueba T dos medias
Prueba Chi-Cuadrado


Nº de líneas
3
Nº de columnas
2
Nivel de significación
0.05
OK
Valores obtenidos
1 2 3
4 5 6
7 8 9
0


3
2
Nivel de significación
0.05
Línea: 1 Columna: 1
180
OK
1 2 3
4 5 6
7 8 9
0


Prueba Chi-Cuadrado
Paso 1
H0: Las variables son independientes
H1: Las variables no son independientes
Paso 2
Formula Qui²
Paso 3
RC > 5.992
Paso 4
Qui² calculado = 47.2433
Paso 5
Qui Cuadrado calculado pertenece a la región crítica, por lo tanto, se rechaza H0, es decir, las variables no son independientes a nivel de significancia 0.05
P-Valor: 0,0000
OTRA OPERACIÓN

Prueba de hipótesis X^2: (Probar si hay relación entre las variables)

Planteamiento de la Hipótesis

H_0: No existe relación entre el turnos y los niveles de educación.

H_1: Existe relación entre el turnos y niveles de educación.

Nivel de significación

$\alpha = 0.05$

Regla de decisión

Si Pvalor < α rechazo (H_0)

Toma de decisión

0.000 **<** 0.05 **entonces** se rechaza (H_0),

R/ Se acepta que hay relación entre el **turnos y el calidad de las piezas**.

2.6. Prueba Chi- Cuadrado de homogeneidad o de proporción de varias muestras.

Ejemplo 2. Prueba de proporción de varias muestras

Un taller de una industria azucarera fabrica piezas metálicas para la zafra y desea evaluar si existe una relación entre los turno de producción (mañana, tarde y noche) y la calidad de las piezas producidas. Para ello, **se seleccionó 3 muestras** (una de cada turno) y se clasificaron las piezas según la calidad de ellas (no defectuosas y no defectuosas), obteniendo la siguiente tabla de contingencia de 3x2:

Turnos categorías: mañana, tarde, noche.

Calidad de las piezas categorías: no defectuosas, o defectuosas.

TABLA DE CONTINGENCIA

		calidad de las piezas	
		no defectuosas	**defectuosa**
turnos	**mañana**	180	100
	tarde	190	280
	noche	170	120

Prueba de hipótesis X^2: (Probar si hay relación entre las variables)

Planteamiento de la Hipótesis

H_0: $p_1 = p_2 = p_3$ (la proporciones de piezas defectuosas son iguales por turnos)

H_1: p_1, p_2, p_3 diferentes (las proporciones de piezas defectuosas no son iguales por turnos)

Nivel de significación

$\alpha = 0.05$

Regla de decisión

Si Pvalor < α rechazo (H_0)

Toma de decisión

0.000 **<** 0.05 **entonces** se rechaza (H_0),

R/ La proporción de piezas defectuosos no son iguales en cada turno.

Conclusiones

CONCLUSIONES

- **La prueba Chi-Cuadrado** es una prueba no paramétrica para **variables categóricas.**
- **PRUEBA DE BONDAD DE AJUSTE:** compara la distribución de una muestra con la distribución teórica que se supone que describe a la población de la cual se extrajo.
- **PRUEBA DE INDEPENDENCIA DE DOS VARIABLES:** comprueba si dos o más variables que se miden en una misma muestras están o no relacionadas entre sí. Siempre hablamos de variables cualitativas.
- **PRUEBA DE HOMOGENEIDAD:** comprueba si dos o más muestras pertenecen a la misma población o si sus proporciones son iguales o no

CONCLUSIONES

- En la prueba de independencia para construir la TC, al emplear una sola muestra, solo se conoce el total general pero no se conoce de antemano los marginales ni de las filas ni de las columnas.
- Mientras que en la prueba de Homogeneidad o de proporciones como son varias muestras se conocen los marginales que coinciden con el tamaño de las muestras
- Mientras mayor sea la diferencia entre los valores observados y los esperados, mayor será el valor de $\chi 2$ y aumentará por tanto la probabilidad de rechazar la hipótesis de nula
- No se puede usar la prueba de $\chi 2$ cuando el valor esperado en alguna celda es menor que 5.

Estudio independiente

Un taller de la industria deportiva desea evaluar si existe una relación entre la materia prima (MP1, MP2 y MP3) empleada en la fabricación de pelotas de béisbol y su calidad (cumplen la norma y no cumplen la norma). Para ello, evalúan 500 pelotas de cada materia prima y se conforma la siguiente tabla de contingencia. Habrá igual proporción de pelotas que cumplen la norma entre cada materia prima.:

TABLA DE CONTINGENCIA

		calidad de las pelotas de béisbol	
		cumplen la norma	**no cumplen la norma**
Materia Prima	**MP 1**	485	15
	MP 2	480	20
	MP 3	475	25

Muchas gracias por su atención