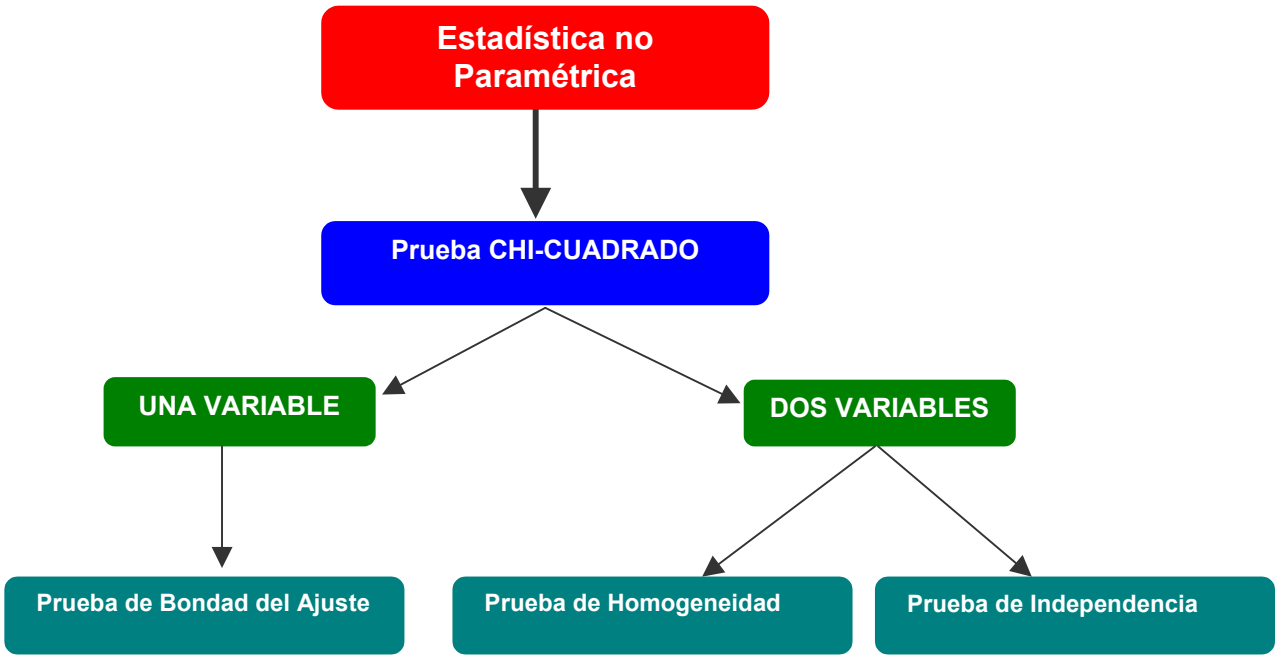
ESTADÍSTICA NO PARAMÉTRICA: PRUEBA CHI-CUADRADO χ^2

Autores: Juan Francisco Monge Ivars (jmonje@uoc.edu), Ángel A. Juan Pérez (ajuanp@uoc.edu)

ESQUEMA DE CONTENIDOS



OBJETIVOS

El objetivo de este e-block es el estudio de varias cuestiones en relación con v.a. cualitativas ó cuantitativas cuyos datos están recogidos en forma de tabla de frecuencias. El denominador común a todas ellas es que su tratamiento estadístico está basado en la misma distribución teórica: la distribución χ^2 (chi-cuadrado ó ji-cuadrado). En esencia se van a abordar tres tipos de problemas:

a) Prueba de Bondad de Ajuste, consiste en determinar si los datos de cierta muestra corresponden a cierta distribución poblacional. En este caso es necesario que los valores de la variable en la muestra y sobre la cual queremos realizar la inferencia esté dividida en clases de ocurrencia, o equivalentemente, sea cual sea la variable de estudio, deberemos categorizar los datos asignado sus valores a diferentes clases o grupos.

b) Prueba de Homogeneidad de varias muestras cualitativas, consiste en comprobar si varias muestras de una carácter cualitativo proceden de la misma población (por ejemplo: ¿estas tres muestras de alumnos provienen de poblaciones con igual distribución de aprobados?. Es necesario que las dos variables medibles estén representadas mediante categorías con las cuales construiremos una tabla de contingencia.

c) Prueba de Independencia, consistente en comprobar si dos características cualitativas están relacionadas entre sí (por ejemplo: ¿el color de ojos está relacionado con el color de los cabellos?). Aunque conceptualmente difiere del anterior, operativamente proporciona los mismos resultados. Este tipo de contrastes se aplica cuando deseamos comparar una variable en dos situaciones o poblaciones diferentes, i.e., deseamos estudiar si existen diferencias en las dos poblaciones respecto a la variable de estudio.

CONOCIMIENTOS PREVIOS

Este math-block supone ciertos conocimientos básicos de estadística (inferencia y probabilidad), así como conocimientos básicos del software estadístico MINITAB.

CONCEPTOS FUNDAMENTALES

- **Muestra**: Parte de una población que se toma cuando es imposible acceder a toda ella. La elección de la muestra se hace con la intención de, a partir de la información que ella proporciona, extender sus resultados a toda la población a la que representa.
- **Muestra aleatoria**: (Muestra elegida al azar) Aquella muestra tomada de la población en la que todo individuo tiene la misma probabilidad de resultar elegido para ella, y esto con independencia entre individuos.
- **Función de Distribución**: Función que hace corresponder a cada uno de los valores de una variable aleatoria la probabilidad de que tal variable aleatoria tome un valor igual o inferior al dado.
- **Función de Probabilidad**: Función que hace corresponder a cada uno de los valores de la variable aleatoria discreta su probabilidad.
- **Contraste de hipótesis**: Conjunto de reglas tendentes a decidir cuál de dos hipótesis –la nula ó la alternativa- debe aceptarse en base al resultado obtenido en una muestra. Es de dos colas cuando la alternativa es la negación de la nula. De una cola en caso contrario.

- **Variable aleatoria:** Toda función que toma diversos valores numéricos, dependiente de los resultados de un fenómeno aleatorio, con distintas probabilidades.

- **Variable aleatoria discreta**. Las variables aleatorias discretas son aquellas que presentan un número finito de valores, constituyen una sucesión numerable.

- **Variable aleatoria continua**. Las variables aleatorias continuas pueden tomar un número infinito de valores en un intervalo determinado.

- **Variable categórica.** Una variable categórica es una variable que clasifica cada individuo de una población en una de las varias clases mutuamente excluyentes en que ésta se divide.

- **Variable numérica.** Corresponde a los datos expresados en una escala continua numérica.

PRUEBA DE BONDAD DE AJUSTE

INTRODUCCIÓN

Estamos interesados en determinar si los datos disponibles de una muestra aleatoria simple de tamaño *n* corresponden a cierta distribución teórica. El primer paso a realizar consiste en descomponer el recorrido de la distribución teórica en un número finito de subconjuntos: A_1, A_2, ..., A_k. Después, clasificar las observaciones muestrales, según el subconjunto a que pertenezcan. Y, por último, comparar las frecuencias observadas de cada A_i con las probabilidades que les corresponderían con la distribución teórica a contrastar.

OBJETIVOS

- Comprender la importancia de este método para medir si los datos resultantes de una muestra provienen de una distribución teórica.
- Metodología útil para validar las hipótesis sobre la distribución teórica en la población que se realiza en la estadística paramétrica, i.e., contrastes de hipótesis, intervalos de confianza, regresión lineal, etc.

CONOCIMIENTOS PREVIOS

- Se debe poseer unos conocimientos mínimos de Inferencia Estadística, i.e., Estadística Descriptiva, Intervalos de Confianza y Contrastes de Hipótesis.

- Es preciso conocer el manejo de algún paquete estadístico y recomendable algunas nociones del paquete MINITAB.

BONDAD DEL AJUSTE (I)

Supongamos que tenemos un número k de clases en las cuales se han ido registrado un total de n observaciones (n será pues el tamaño muestral). Denotaremos las **frecuencias observadas** en cada clase por O_1, O_2, ..., O_k (O_i es el número de valores en la clase A_i). Se cumplirá:

$$O_1 + O_2 + ... + O_k = n$$

Lo que queremos es comparar las frecuencias observadas con las **frecuencias esperadas** (teóricas), a las que denotaremos por E_1, E_2, ..., E_k. Se cumplirá:

$$E_1 + E_2 + ... + E_k = n$$

	FRECUENCIA OBSERVADA	FRECUENCIA ESPERADA
CLASE 1	O_1	E_1
CLASE 2	O_2	E_2
...	...	...
CLASE K	O_K	E_K
Total	n	N

Se tratará ahora de decidir si las frecuencias observadas están o no en concordancia con las frecuencias esperadas (es decir, si el número de resultados observados en cada clase corresponde

aproximadamente al número esperado). Para comprobarlo, haremos uso de un contraste de hipótesis usando la distribución Chi-cuadrado:

El estadístico de contraste será $\chi^{2^*} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

Observar que este valor será la suma de k números no negativos. El numerador de cada término es la diferencia entre la frecuencia observada y la frecuencia esperada. Por tanto, cuanto más cerca estén entre sí ambos valores más pequeño será el numerador, y viceversa. El denominador permite relativizar el tamaño del numerador.

Las ideas anteriores sugieren que, cuanto menor sean el valor del estadístico χ^{2^*}, más coherentes serán las observaciones obtenidas con los valores esperados. Por el contrario, valores grandes de este estadístico indicarán falta de concordancia entre las observaciones y lo esperado. En este tipo de contraste se suele rechazar la hipótesis nula (los valores observados son coherentes con los esperados) cuando el estadístico es mayor que un determinado valor crítico.

Notas:

(1) El valor del estadístico χ^{2^*} se podrá aproximar por una distribución Chi-cuadrado cuando el tamaño muestral n sea grande (n > 30), y todas las frecuencias esperadas sean iguales o mayores a 5 (en ocasiones deberemos agrupar varias categorías a fin de que se cumpla este requisito).

(2) Las observaciones son obtenidas mediante muestreo aleatorio a partir de una población particionada en categorías.

BONDAD DEL AJUSTE (II)

Un **experimento multinomial** es la generalización de un experimento binomial:

1. Consiste en n pruebas idénticas e independientes.
2. Para cada prueba, hay un número k de resultados posibles.
3. Cada uno de los k posibles resultados tiene una probabilidad de ocurrencia p_i asociada ($p_1 + p_2 + ... + p_k = 1$), la cual permanece constante durante el desarrollo del experimento.
4. El experimento dará lugar a un conjunto de frecuencias observadas (O_1, O_2, ..., O_k) para cada resultado. Obviamente, $O_1 + O_2 + ... + O_k = n$.

En ocasiones estaremos interesados en comparar los resultados obtenidos al realizar un experimento multinomial con los resultados esperados (teóricos). Ello nos permitirá saber si nuestro modelo teórico se ajusta bien o no a las observaciones. Para ello, recurriremos a la distribución Chi-cuadrado, la cual nos permitirá realizar un **contraste sobre la bondad del ajuste**.

Concretamente, usaremos el estadístico $\chi^{2^*} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ con **k – 1** grados de libertad.

Podemos calcular cada frecuencia esperada (teórica) multiplicando el número total de pruebas n por la probabilidad de ocurrencia asociada, es decir:

$$E_i = n * p_i \qquad i = 1, ..., k$$

CASOS PRÁCTICOS

EJEMPLO:

En cierta máquina Expendedora de Refrescos existen 4 canales que expiden el mismo tipo de bebida. Estamos interesados en averiguar si la elección de cualquiera de estos canales se hace de forma aleatoria o por el contrario existe algún tipo de preferencia en la selección de alguno de ellos por los consumidores. La siguiente tabla muestra el número de bebidas vendidas en cada uno de los 4 canales durante una semana. Contrastar la hipótesis de que los canales son seleccionados al azar a un nivel de significación del 5%.

Canal	Número de bebidas consumidas mediante este expendedor
1	13
2	22
3	18
4	17

SOLUCIÓN:

Para realizar el contraste de Bondad de Ajuste debemos calcular las frecuencias esperadas de cada suceso bajo la hipótesis de uniformidad entre los valores. Si la selección del canal fuera aleatoria, todos los canales tendrían la misma probabilidad de selección y por lo tanto la frecuencia esperada de bebidas vendidas en cada uno de ellos debería ser aproximadamente la misma. Como se han vendido en total 70 refrescos, la frecuencia esperada en cada canal es

$E_i = n * p_i = 70 * ¼ = 17.5 \qquad i = 1, ..., k$

El estadístico del contraste sería:

$$\chi^2 = \frac{(13-17.5)^2}{17.5} + \frac{(22-17.5)^2}{17.5} + \frac{(18-17.5)^2}{17.5} + \frac{(17-17.5)^2}{17.5} = 2.3428$$

Este valor debemos compararlo con el valor crítico de la distribución χ^2 con (4-1)=3 grados de libertad. Este valor es: $\chi^2_{0.95}(3) = 7.81$

Puesto que el valor del estadístico (2.34) es menor que el valor crítico, no podemos rechazar la hipótesis de que los datos se ajustan a una distribución uniforme. Es decir, que los canales son seleccionados aleatoriamente entre los consumidores.

EJEMPLO:

Estamos interesados en comprobar la perfección de un dado cúbico (un dado normal de 6 caras). Para esto realizamos 100 lanzamientos del dado anotando los puntos obtenidos en cada lanzamiento. A la vista de los resultados obtenidos, ¿podemos concluir que el dado no es perfecto?. Nivel de significación (5%)

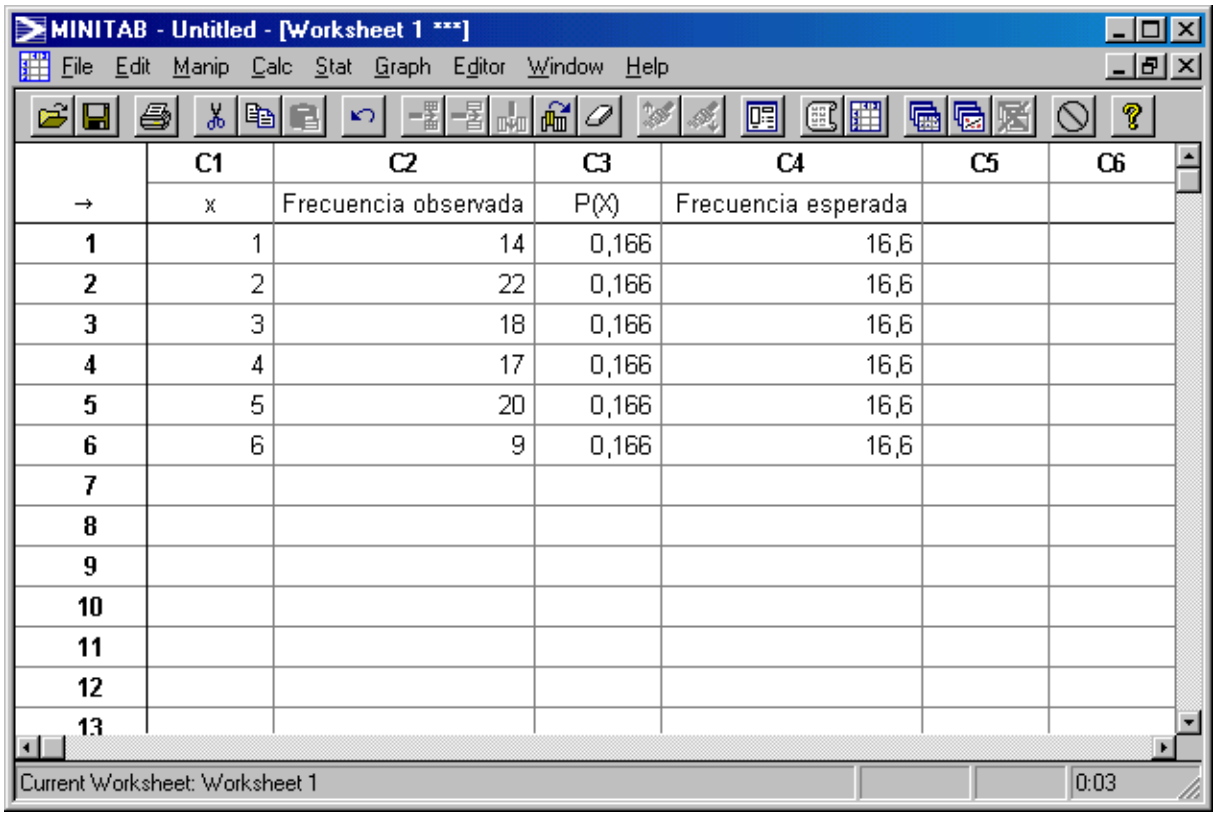
Puntuación en el dado	Número de veces que se obtiene la puntuación.
1	14
2	22

3	18
4	17
5	20
6	9

SOLUCIÓN:

Si el dado estuviera equilibrado, en el resultado de lanzarlo sucesivamente se deberían obtener aproximadamente el mismo número de veces cada una de las caras del dado. En este ejercicio debemos contrastar si la distribución del dado es una distribución uniforme, con probabilidad de obtener cada una de las caras igual a 1/6.

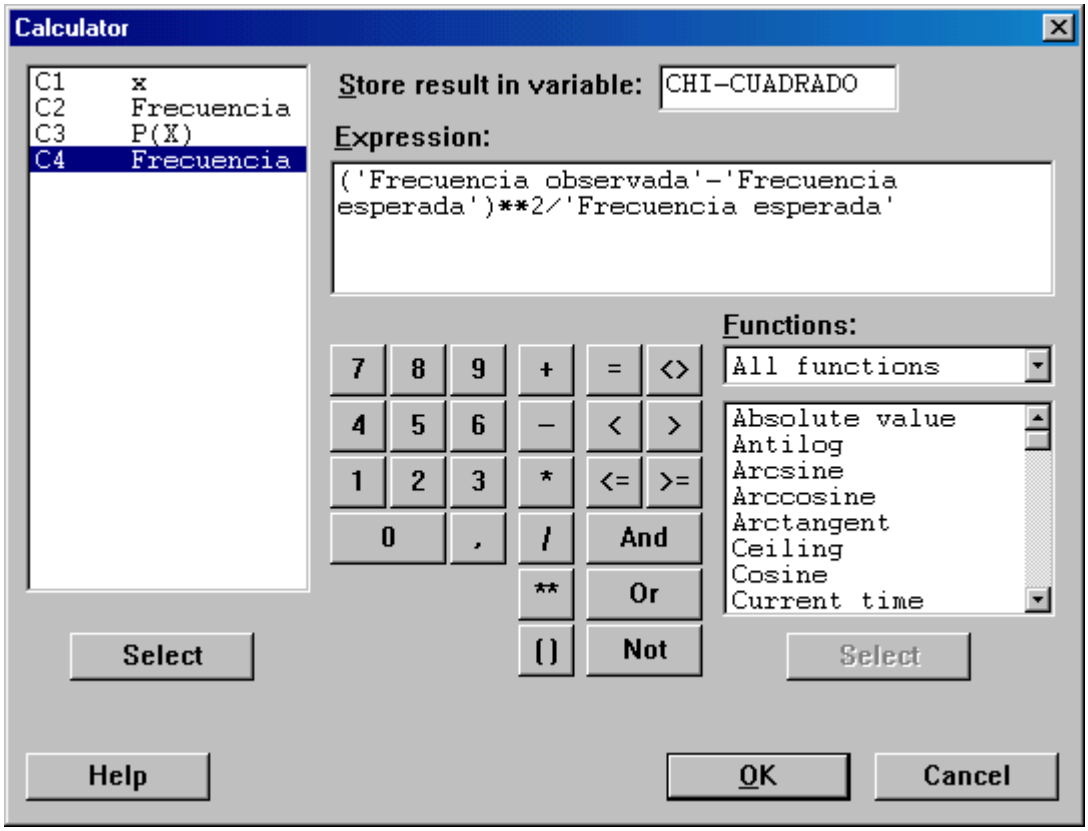
Podemos calcular de una forma muy sencilla el número esperado de resultados obtenidos en cada clase multiplicando la probabilidad de obtener cada una de las caras (p = 1/6) por el número de lanzamientos (n = 100).

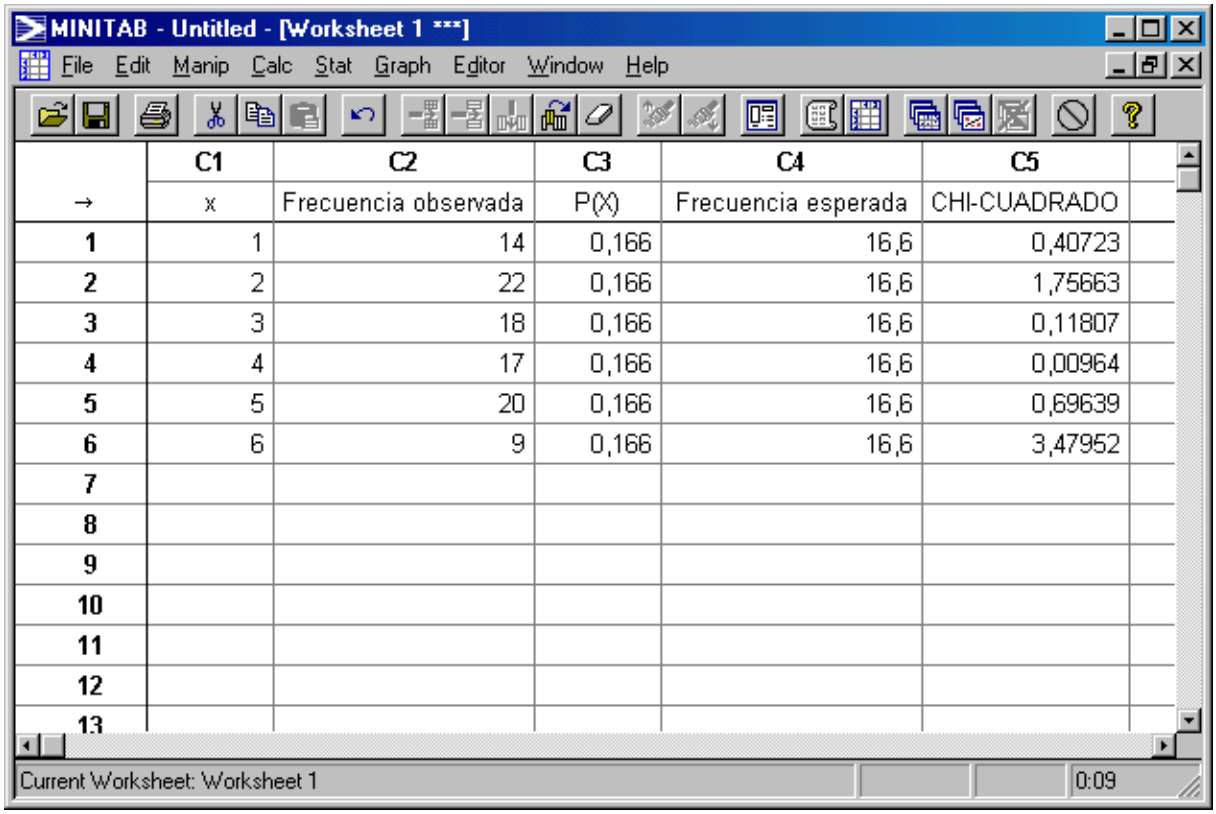


	C1	C2	C3	C4	C5	C6
→	x	Frecuencia observada	P(X)	Frecuencia esperada		
1	1	14	0,166	16,6		
2	2	22	0,166	16,6		
3	3	18	0,166	16,6		
4	4	17	0,166	16,6		
5	5	20	0,166	16,6		
6	6	9	0,166	16,6		
7						
8						
9						
10						
11						
12						
13						

Podemos observa que los valores observados y esperados no parecen coincidir, por lo tanto, a priori parece haber evidencias de irregularidades en el dado. Calculemos el estadístico χ^2 con ayuda del *Calculator* de MINITAB.

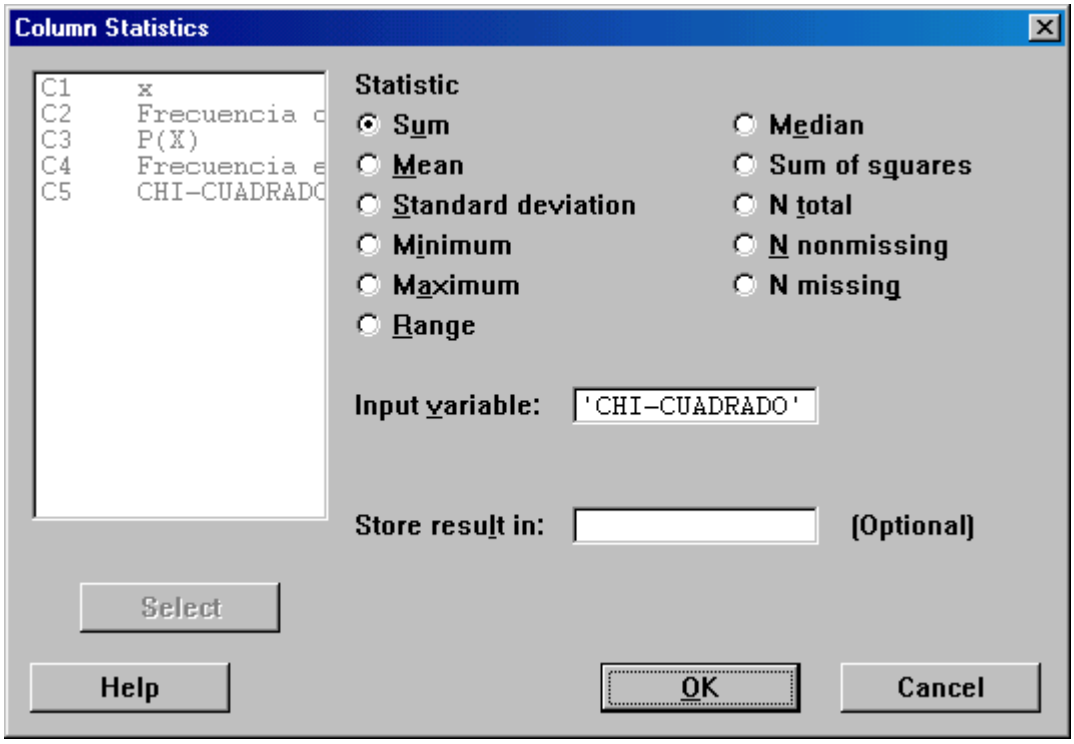
Calc > Calculator





	C1	C2	C3	C4	C5
→	x	Frecuencia observada	P(X)	Frecuencia esperada	CHI-CUADRADO
1	1	14	0,166	16,6	0,40723
2	2	22	0,166	16,6	1,75663
3	3	18	0,166	16,6	0,11807
4	4	17	0,166	16,6	0,00964
5	5	20	0,166	16,6	0,69639
6	6	9	0,166	16,6	3,47952
7					
8					
9					
10					
11					
12					
13					

Calc > Column Statistics



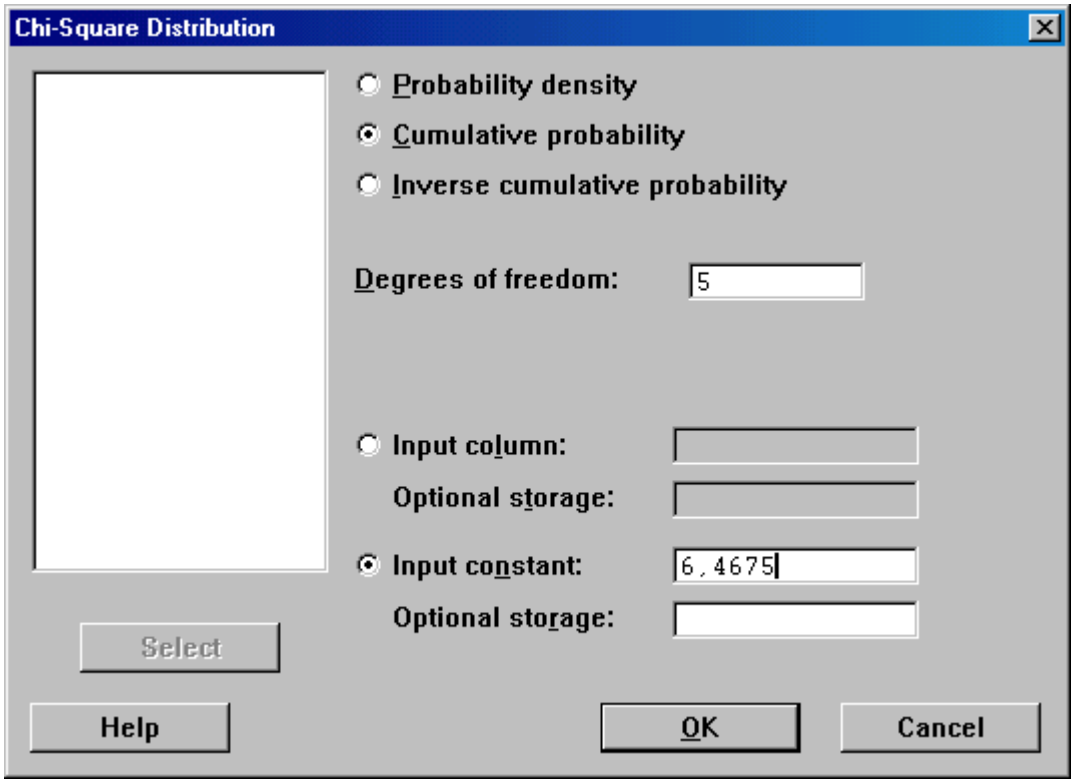
Con el resultado:

Column Sum

Sum of CHI-CUADRADO = 6,4675

Calculemos finalmente el p-valor asociado a este estadístico. En este caso, como trabajamos con un contraste unilateral, p-valor= $P(\chi^2 > 6{,}4675) = 1 - P(\chi^2 < 6{,}4675)$ donde χ^2 sigue una distribución Chi-cuadrado con k-1=5 grados de libertad. Por tanto:

Calc > Probability Distributions > Chi-square



Cumulative Distribution Function

```
Chi-Square with 5 DF

          x      P( X <= x)
    6,4675         0,7367
```

Así pues, p-valor = 1 – 0,7367 = 0, 2633 . Por tanto, podemos considerar que el p-valor no es significativo. Concluiremos, a pesar de las evidencias que habían en un principio, que no hay evidencias para rechazar que el dato fuera correcto, i.e., no podemos rechazar la distribución uniforme para los posibles resultados del dado.

PRUEBA DE HOMOGENEIDAD

INTRODUCCIÓN

Estamos interesados en determinar si los datos correspondientes a dos o más muestras aleatorias provienen de la misma población. Nuevamente el conjunto de posibles valores de las observaciones se divide en k conjuntos disjuntos: A_1, A_2, ..., A_k.; clasificando en ellos las observaciones de cada muestra. Si n_{ij} representa el número de observaciones de la muestra i que pertenecen al conjunto A_j , los datos pueden tabularse en lo que se denomina una tabla de contingencia.

Muestra	A_1	A_2	...	A_k.	Total
1	n_{11}	n_{12}		n_{1k}	$n_{1.}$
2	n_{21}	n_{22}		n_{2k}	$n_{2.}$
...					
m	n_{m1}	n_{m2}		n_{mk}	$n_{m.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.k}$	n

La hipótesis de que las m poblaciones son homogéneas, se traduce en que cada conjunto A_j debe tener una probabilidad teórica p_j, desconocida, pero que no varia de la población i a la población i'. Esto debe verificarse para todas las categorías, i.e., las categorías deben ser homogéneas en las diversas muestras.

OBJETIVOS

- Comprender la importancia de este método para medir si dos muestras aleatorias provienen de la misma población. Notar que en la estadística no paramétrica, como es este contraste, no se realizan contrastes sobre parámetros de la población (contraste de igualdad de medias),i.e., se realizan contrastes sobre la población origen.
- Metodología muy útil para comparar diversas muestras y extraer conclusiones sobre la igualdad en las distribuciones poblacionales de cada una de ellas.

CONOCIMIENTOS PREVIOS

- Se debe poseer unos conocimientos mínimos de Inferencia Estadística, i.e., Estadística Descriptiva, Intervalos de Confianza y Contrastes de Hipótesis.
- Es preciso conocer el manejo de algún paquete estadístico y recomendable algunas nociones del paquete MINITAB.

CONCEPTOS FUNDAMENTALES

Del mismo modo que la Prueba de Bondad de Ajuste, en este caso debemos comparar las frecuencias observadas en cada una de las muestras y para cada categoría con las frecuencias bajo el supuesto de homogeneidad en las poblaciones. En este caso las frecuencias observadas corresponde al número de individuos de la muestra i en la clase j, i.e., n_{ij}.

El estadístico de contraste será $$\chi^{2^*} = \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}$$

Donde e_{ij} es la frecuencia esperada bajo el supuesto de homogeneidad, que puede representarse como $n_i\, p_j$, es decir, el número de individuos en la muestra i por la probabilidad de que ocurra la característica j en la población. Para el cálculo de las probabilidades de pertenecer un individuo a cada una de las categorías podemos utilizar:

$$p_i = n_{.j} / n$$

Por lo tanto : $e_{ij} = n_{i.} \cdot n_{.j} / n$

Observar que este valor será la suma de n*k números no negativos. El numerador de cada término es la diferencia entre la frecuencia observada y la frecuencia esperada. Por tanto, cuanto más cerca estén entre sí ambos valores más pequeño será el numerador, y viceversa. El denominador permite relativizar el tamaño del numerador.

Las ideas anteriores sugieren que, cuanto menor sean el valor del estadístico χ^{2*}, más coherentes serán las observaciones obtenidas con los valores esperados. Por el contrario, valores grandes de este estadístico indicarán falta de concordancia entre las observaciones y lo esperado. En este tipo de contraste se suele rechazar la hipótesis nula (los valores observados son coherentes con los esperados) cuando el estadístico es mayor que un determinado valor crítico.

Notas:

(3) El valor del estadístico χ^{2*} se podrá aproximar por una distribución Chi-cuadrado cuando el tamaño muestral n sea grande (n > 30), y todas las frecuencias esperadas sean iguales o mayores a 5 (en ocasiones deberemos agrupar varias categorías a fin de que se cumpla este requisito).

(4) Las observaciones son obtenidas mediante muestreo aleatorio en cada muestra a partir de una población particionada en categorías.

Concretamente, usaremos el estadístico $\chi^{2^*} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ con (n-1)(k – 1) grados de libertad.

CASOS PRÁCTICOS

EJEMPLO :

Estamos interesados en estudiar la fiabilidad de cierto componente informático con relación al distribuidor que nos lo suministra. Para realizar esto, tomamos una muestra de 100 componentes de cada uno de los 3 distribuidores que nos sirven el producto comprobando el número de defectuosos en cada lote. La siguiente tabla muestra el número de defectuosos en para cada uno de los distribuidores.

	Componentes Defectuosos	Componentes correctos	
Distribuidor 1	16	94	100
Distribuidor 2	24	76	100
Distribuidor 3	9	81	100
	49	251	300

SOLUCIÓN:

Debemos realizar un contraste de homogeneidad para concluir si entre los distribuidores existen diferencias de fiabilidad referente al mismo componente.

	Componentes Defectuosos	Componentes correctos	
Distribuidor 1	16 (16.33)	94 (83.66)	100
Distribuidor 2	24 (16.33)	76 (83.66)	100
Distribuidor 3	9 (16.33)	81 (83.66)	100
	49	251	300

Las frecuencias esperadas bajo homogeneidad son las representadas entre paréntesis.

El estadístico del contraste será:

$$\chi^2 = \frac{(16-16.33)^2}{16.33} + \frac{(24-16.33)^2}{16.33} + \frac{(9-16.33)^2}{16.33} +$$
$$+ \frac{(94-83.66)^2}{83.66} + \frac{(76-83.66)^2}{83.66} + \frac{(81-83.66)^2}{83.66} = 8.9632$$

Este valor del estadístico Ji-cuadrado es mayor que el valor para el nivel de significación del 5%, por lo tanto debemos concluir que no existe homogeneidad y por lo tanto que hay diferencias entre los tres distribuidores. $\chi^2_{0.05}(2) = 5.99$

EJEMPLO:

Estamos interesados en estudiar la relación entre cierta enfermedad y la adicción al tabaco. Para realizar esto seleccionamos una muestra de 150 individuos, 100 individuos no fumadores y 50 fumadores. La siguiente tabla muestra las frecuencias de enfermedad en cada grupo (Completar la tabla).

	Padecen la Enfermedad	No Padecen la enfermedad	
Fumadores	12	88	100
No Fumadores	25	25	50
	37	113	150

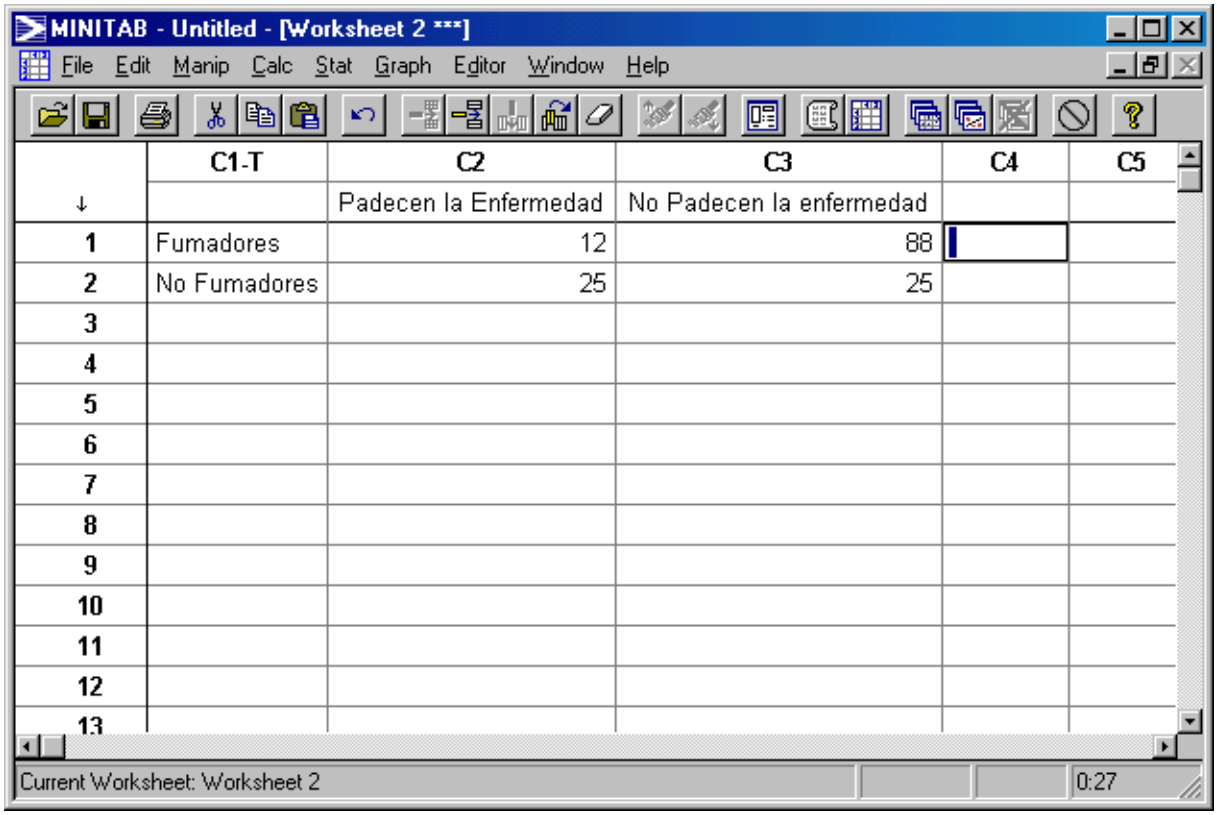
Realizar un contraste de homogeneidad y obtener las conclusiones sobre la relación entre las variables.

SOLUCIÓN:

Para considerar este contraste como un contraste de Homogeneidad suponemos que las personas fumadoras y las personas no fumadoras constituyen dos poblaciones diferenciadas. Un estudio similar consistiría en considerar a los fumadores y no fumadores como una característica de una población y por lo tanto este ejemplo podría plantearse como un contraste de independencia, ver PRUEBA DE INDEPENDENCIA.

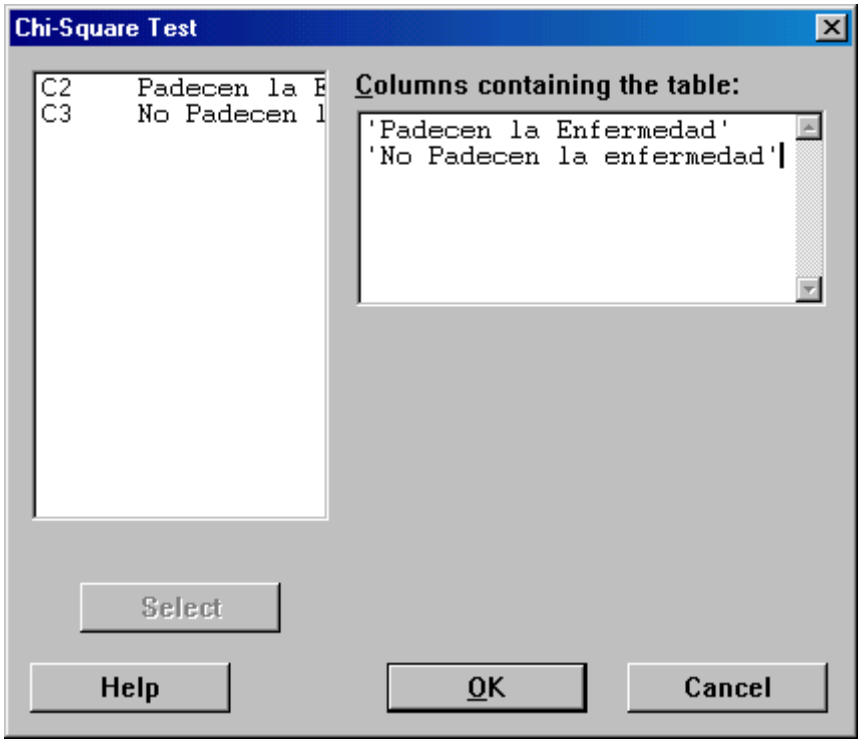
En este ejemplo queremos contrastar la hipótesis de que las proporciones de enfermos en ambas poblaciones (Fumadores y No Fumadores) es la misma.

La representación de la tabla de contingencia en Minitab debe ser la misma que la anterior:



Minitab realiza los cálculos por nosotros:

Stat > Tables > Chi-square Test:




```
Expected counts are printed below observed counts

        Padecen No Padece    Total
    1         12        88      100
           24,67     75,33

    2         25        25       50
           12,33     37,67

Total         37       113      150

Chi-Sq =  6,505 +  2,130 +
         13,009 +  4,260 = 25,903
DF = 1, P-Value = 0,000
```

En los resultados aparecen las frecuencias esperadas bajo el supuesto de homogeneidad. Con un p-valor de 0,000 hay suficiente evidencia en contra de que la hipótesis nula sea cierta. Por tanto, la rechazaríamos, i.e.; parece evidente que los fumadores tienen una mayor propensión a padecer la enfermedad.

PRUEBA DE INDEPENDENCIA

INTRODUCCIÓN

Estamos interesados en determinar si dos cualidades o variables referidas a individuos de una población están relacionadas. Se diferencia de los contrastes anteriores en que en este caso estamos interesados en ver la relación existente entre dos variables de una misma población, no queremos contrastar la distribución teórica de una variable (prueba de bondad de ajuste) ni en comparar la distribución de una única variable en dos poblaciones (prueba de homogeneidad).

OBJETIVOS

- Comprender la importancia de este método para medir relaciones entre variables si realizar supuesto adicionales sobre las distribuciones de estas.
- Alternativa muy potente para medir relaciones entre variables categóricas, donde no es posible aplicar los métodos clásicos de Inferencia Estadística como la Regresión Lineal. También es aplicable a variables cuantitativas si no se verifican los supuestos necesarios a satisfacer por otras técnicas estadísticas.
- Identificar las diferencias conceptuales entre el test de homogeneidad y el Test de Independencia.

CONOCIMIENTOS PREVIOS

- Se debe poseer unos conocimientos mínimos de Inferencia Estadística, i.e., Estadística Descriptiva, Intervalos de Confianza y Contrastes de Hipótesis.
- Es preciso conocer el manejo de algún paquete estadístico y recomendable algunas nociones del paquete MINITAB.

PRUEBA DE INDEPENDENCIA

Supongamos que de n elementos de una población se han observado dos características X e Y, obteniéndose una muestra aleatoria simple bidimensional $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$. Sobre la base de dichas observaciones se desea contrastar si las características poblacionales X e Y son independientes o no. Para ello se dividirá el conjunto de posibles valores de X en k conjuntos disjuntos $A_1,A_2,...,A_k$; mientras que el conjunto de posibles valores Y será descompuesto en r conjuntos disjuntos: $B_1,B_2,...,B_r$. Al clasificar os elementos de la muestra, aparecerá un cierto número de ellos, n_{ij}, en cada una de las $k \times r$ clases así constituidas, dando lugar a una tabla de contingencia de la forma:

	A_1	A_2	...	$A_{k.}$	Total
B_1	n_{11}	n_{12}		n_{1k}	$n_{1.}$
B_2	n_{21}	n_{22}		n_{2k}	$n_{2.}$
...					
B_r	n_{r1}	n_{r2}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.k}$	n

Al igual que para el Test de homogeneidad, el estadístico del contraste será $\chi^{2^*} = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}$ con (k-1)(r-1) grados de libertad.

Donde: $e_{ij} = n_{i.} \cdot n_{.j} / n$

EJEMPLO:

Para estudiar la dependencia entre la práctica de algún deporte y la depresión, se seleccionó una muestra aleatoria simple de 100 jóvenes, con los siguientes resultados:

	Sin depresión	Con depresión
Deportista	38	9
No deportista	31	22

Determinar si existe independencia entre la actividad del sujeto y su estado de ánimo. Nivel de significación (5%)

SOLUCIÓN:

Debemos primero calcular las frecuencias esperadas bajo el supuesto de independencia. La tabla de frecuencias esperadas sería:

	Sin depresión	Con depresión	
Deportista	32.43	14.57	47
No deportista	36.57	16.43	53
	69	31	100

Calculamos ahora el estadístico del contraste:

$$\chi^2 = \frac{(38-32.43)^2}{32.43} + \frac{(9-14.57)^2}{14.57} + \frac{(31-36.57)^2}{36.57} + \frac{(22-16.43)^2}{16.43} = 5.82$$

Este valor debemos compararlo con el percentil de la distribución χ^2 con (2-1)(2-1)=1 grado de libertad. $\chi^2_{0.95}(1) = 3.84$.

Por lo tanto como el valor del estadístico es superior al valor crítico, concluimos que debemos rechazar la hipótesis de independencia y por lo tanto asumir que existe relación entre la depresión e los hábitos deportistas del individuo.

EJEMPLO:

Un estudio que se realizó con 81 personas referente a la relación entre la cantidad de violencia vista en la televisión y la edad del televidente produjo los siguientes resultados.

	16-34	34-55	55 ó más
Poca violencia	8	12	21
Mucha Violencia	18	15	7

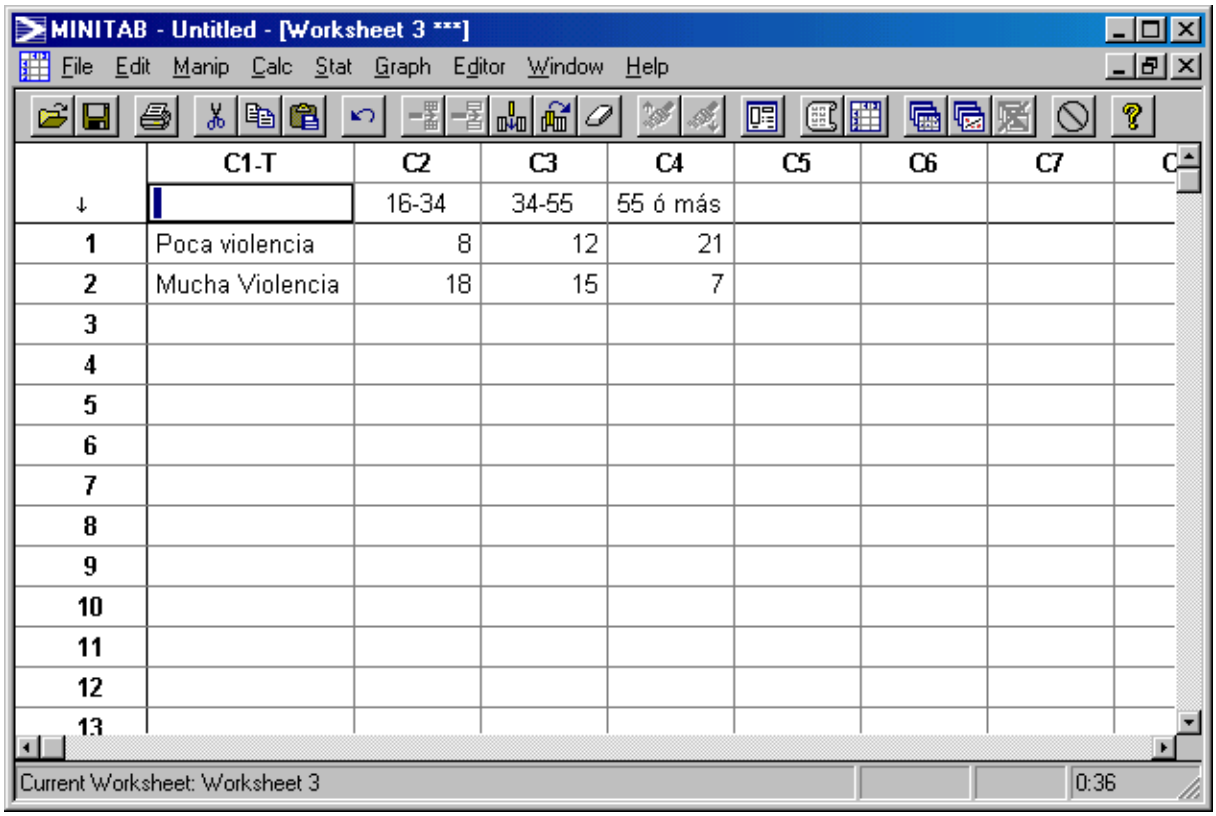
¿Indican los datos que ver violencia en la televisión depende de la edad del televidente, a un nivel de significación del 5%?

SOLUCIÓN:

Debemos realizar un test de independencia para ver si existe relación entre la violencia vista en televisión con el grupo de edad al que pertenece el individuo.

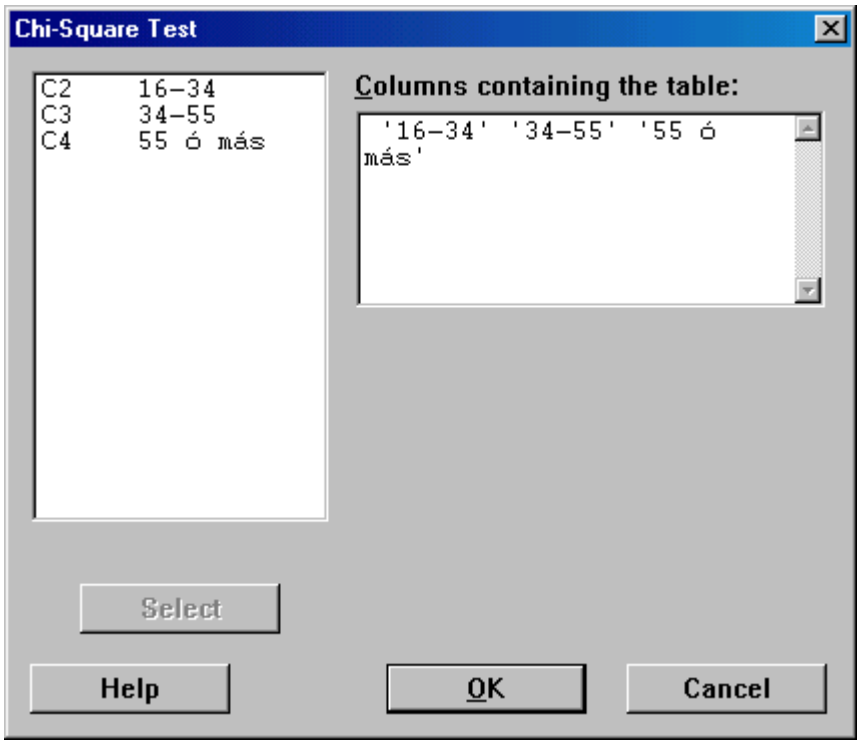
Dado que el test de Independencia, no difiere del test de Homogeneidad a nivel operacional, el desarrollo es análogo al ejercicio de Minitab de la sección anterior.

Introducimos los valores de la tabla de contingencia del siguiente modo:



	C1-T	C2	C3	C4	C5	C6	C7	C
↓		16-34	34-55	55 ó más				
1	Poca violencia	8	12	21				
2	Mucha Violencia	18	15	7				
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								

Stat > Tables > Chi-Square Test:



Chi-Square Test

Expected counts are printed below observed counts

	16-34	34-55	55 ó más	Total
1	8	12	21	41
	13,16	13,67	14,17	
2	18	15	7	40
	12,84	13,33	13,83	
Total	26	27	28	81

Chi-Sq = 2,024 + 0,203 + 3,289 +
2,074 + 0,208 + 3,371 = 11,169
DF = 2, P-Value = 0,004

El valor del estadístico del contraste es 11,169. El p-valor asociado a este valor es 0,004. Por lo tanto a un nivel de significación del 0.005 deberemos rechazar la hipótesis nula de independencia, y por lo tanto concluir que existe diferencias entre el tipo de televisión consumida y la edad del televidente.

BIBLIOGRAFÍA

[1] Baró, J. y Alemany, R. (2000): "Estadística II". Ed. Fundació per a la Universitat Oberta de Catalunya. Barcelona.

[2] Peña Sánchez de Rivera, D. (1987): "Estadística. Modelos y Métodos. Volumen 2". Alianza Editorial. Madrid. ISBN: 84-206-8110-5

[3] Johnson, R. R. (1996): "Elementary statistics". Belmont, etc. : Duxbury, cop

[4] R. Vélez y A. García: "Cálculo de Probabilidades y Estadística Matemática". Ciencias Matemáticas. UNED.

[5] A. Martín Andrés y J. de D. Luna del Castillo: " 50 ± 10 horas de Bioestadística" Ediciones Norma.