

# APLICACIONES DE LA DISTRIBUCIÓN $\chi^2$

Por José Antonio ROMERO LÓPEZ  
Profesor de Matemáticas  
del Instituto de Educación Secundaria  
"Isidro de Arcenegui y Carmona",  
de Marchena, Sevilla-Spain.  
E-MAIL: [rljatf@teleline.es](mailto:rljatf@teleline.es)

1. Biografías de Von Neumman y de la familia Bernuilli
2. Aplicaciones de la distribución chi-cuadrado.
  - 2.1. Test de bondad del ajuste.
  - 2.2. Test de independencia.
  - 2.3. Test de homogeneidad.
3. Problemas resueltos.
4. Resolución de problemas con el ordenador.
5. Problemas propuestos.
6. Resumen.
7. Bibliografía.

## 1. Biografías de Von Neumann y de la Familia Bernuilli:

---

### *Neumann, John von*



#### Vida

- 1903

matemático estadounidense nacido en Hungría, que desarrolló la rama de las matemáticas conocida como **teoría de juegos**. Nació en Budapest y estudió en Zurich y en las universidades de Berlín y Budapest.

- 1930

Viajó a Estados Unidos en 1930 para unirse al claustro de la Universidad de Princeton.

- 1933

A partir de 1933 se incorporó al Instituto de Estudios Avanzados de Princeton (Nueva Jersey).

- 1937

Adquirió la nacionalidad estadounidense en 1937 y durante la II Guerra Mundial ejerció como asesor en el proyecto de la bomba atómica de Los Álamos.

- 1955

En marzo de 1955 fue nombrado miembro de la Comisión de Energía Atómica de los Estados Unidos.

## Obra

Von Neumann fue un gran matemático. Destacó por sus aportaciones fundamentales a la teoría cuántica, especialmente el concepto de anillos de operadores (actualmente conocido como álgebra de Neumann) y también por su trabajo de iniciación de las matemáticas aplicadas, principalmente la estadística y el análisis numérico. También es conocido por el diseño de computadoras electrónicas<sup>1</sup> de gran velocidad. En 1952 diseñó la primera computadora que utilizaba un programa archivado flexible, el *MANIAC I*.

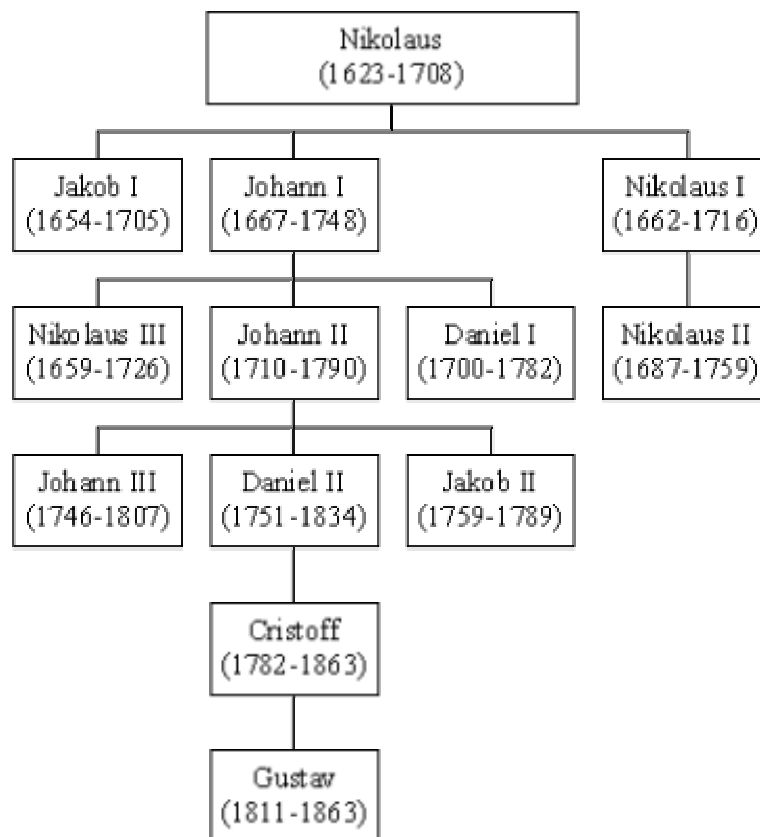
## Notas

- 1En 1956, la Comisión de Energía Atómica le concedió el premio Enrico Fermi por sus notables aportaciones a la teoría y al diseño de las computadoras electrónicas.

## Familia Bernouilli

El nombre Bernouilli está asociado una familia de sabios suizos que sobresalieron por sus aportaciones a las matemáticas y la física. Jamás ha habido una familia de científicos que haya aportado tanto como ellos. La familia de Jacobo Bernouilli procedía de la ciudad de Francfort del Main donde se habían visto obligados a refugiarse huyendo de las persecuciones religiosas de la ciudad de Amberes durante la segunda mitad del siglo XVI. Un descendiente suyo, Nicolas Bernouilli, decidió en 1583 trasladarse a Basilea donde nació Nikolaus que llegó a gozar de cierto prestigio en la ciudad.

El árbol genealógico de la familia Bernouilli es el siguiente:



Destacamos como matemáticos a:

**Jakob I:** estudió teología, aunque su afición natural lo inclinaron a los estudios de las matemáticas y la astronomía. Se dedicó a dar clases de física y matemáticas en Basilea, siendo uno de sus alumnos su hermano Johann I.

Escribió *Ars coniectandi* (Arte de la conjetura) a finales del siglo XVII, publicado por su sobrino Nikolaus ocho años después de su muerte. En esta obra enuncia la ley de los grandes números: "La frecuencia relativa de un suceso tiende a estabilizarse en torno a un número, a medida que el número de pruebas del experimento crece indefinidamente". Dicha obra consta de cuatro partes. La primera contiene los estudios de Huygens; la segunda se ocupa de las variaciones, permutaciones y combinaciones; en la tercera se aplican los teoremas de la teoría de permutaciones al cálculo de probabilidades y en la cuarta, a las aplicaciones de éste a cuestiones de la vida política y social.

**Johann I:** desoyendo los consejos de su padre para que estudiase comercio, estudió medicina doctorándose en 1694. No obstante ejerció diez años el profesorado en matemáticas y física en Groninga hasta que en 1705 regresa a Basilea para encargarse de una cátedra de griego, pero habiendo muerto su hermano Jacobo, le fue confiada la cátedra de matemáticas de éste, la cual desempeñó durante cuarenta y dos años, al cabo de los cuales murió. Las universidades de Leiden, Padua, Groninga y Berlín hicieronle diferentes ofertas llamándole a su seno, más nunca fueron aceptadas.

**Nikolaus II:** fue discípulo de Johann I y Jacob I, se dedicó al estudio de las matemáticas y de la jurisprudencia. A los diez y seis años se dedica a auxiliar a su padre, Johann. En 1716 es nombrado profesor en Padua; en 1722 se hizo cargo de la cátedra de lógica de Basilea y en 1731 de otra de enseñanza del Código. Así mismo, estudia pintura en Augsburgo, perteneciendo al gremio de pintores de su ciudad natal.

**Nikolaus III:** estudió matemáticas y derecho, teniendo como discípulo a su hermano Daniel I. Se encargó en 1723 de una cátedra de derecho de Berna.

**Daniel I:** aunque consiguió el título médico en 1721, su mayor afición fueron las matemáticas. Fue profesor de matemáticas en la Academia Rusa de San Petersburgo en 1725 junto con su hermano Nicolas. Nicolas muere al año y medio de vivir en Rusia, y Daniel, transcurridos los cinco años de permanencia en San Petersburgo fuése a vivir otra vez a Basilea donde enseña anatomía, botánica, física y filosofía especulativa.

**Johann II:** se dedicó a la enseñanza de las matemáticas en la cátedra que había hecho famosa su tío y padre en Basilea. La Academia de Ciencias de París le premió tres memorias escritas en colaboración con su hermano Daniel. Fue miembro de la Academia de Berlín.

**Jacob II :** Ocupa la cátedra de física de su tío Daniel había poseído. Dedícose a viajar hasta que en San Petersburgo se casa con una nieta de Euler, estableciéndose en dicha ciudad, a cuya Academia

perteneció, así como a la Sociedad física de Basilea y a la Sociedad real de Turín. Sus primeros trabajos hicieron concebir de él grandes esperanzas, pero murió prematuramente de apoplejía a los treinta años, bañándose en el Neva.

**Johann III:** Se dedicaron a la enseñanza de las matemáticas, astronomía y filosofía. A los diez y nueve años fue nombrado astrónomo de la Academia de Berlín. En 1779 es nombrado director de la clase de matemáticas de la Academia de Berlín. Fue miembro de las academias de Estocolmo y San Petersburgo.

Otros miembros de la familia:

**Cristobal:** Naturalista suizo, fue profesor de la Universidad de Halle y de Basilea, donde fundó un establecimiento particular de enseñanza.

**Jerónimo:** Naturalista suizo que recoge numerosos materiales para un gabinete de historia para un gabinete de historia natural que a su muerte legó a Basilea, y escribió importantes trabajos sobre mineralogía.

**Juan Jacobo:** Arqueólogo suizo, que ejerce como profesor en la Universidad de Basilea.

**Nicolas I:** Pintor suizo; estudia en Estrasburgo y en París, residiendo algunos años en Italia. En el museo de Basilea se conserva un retrato de su hermano Jacobo.

## 2. Aplicaciones de la distribución $\chi^2$ :

Además de la importancia de la distribución  $\chi^2$  como la distribución a la que se ajusta la distribución muestral de la varianza de una población normal, y su posterior aplicación en el contraste de la varianza, otras aplicaciones importantes de dicha distribución son:

- 2.1. Test de Bondad de Ajuste.
- 2.2. Test de Independencia.
- 2.3. Test de Homogeneidad.

A continuación se especifica la notación que se utilizará a lo largo del desarrollo de la unidad y a la que se hará referencia.

**n** = tamaño de la muestra correspondiente.

**1- $\alpha$**  = nivel de confianza

**$\alpha$**  = nivel de riesgo o de significación.

$\chi^2_{\alpha}(n-1)$  es el valor de una variable chi-cuadrado con **(n-1)** grados de libertad que deja a su derecha un área (probabilidad) de valor  **$\alpha$** .

OBSERVACIÓN: En el caso de dos poblaciones se utilizarán subíndices.

### 2.1 Test de bondad de ajuste.

Consideramos una población y el carácter X que presenta las siguientes modalidades  $x_1, x_2, \dots, x_k$  excluyentes con sus correspondientes probabilidades  $p_1, p_2, \dots, p_k$ . Tenemos una muestra de tamaño n en la que observamos el carácter X y nos planteamos hasta qué punto esta muestra se puede considerar como perteneciente a una población con una distribución teórica ya conocida.

Independientemente de la distribución teórica que consideremos siempre existirán diferencias entre los valores teóricos esperados y los valores observados. El problema está en saber en qué medida dichos valores son debidos al azar o a que los datos no se ajustan a la distribución teórica considerada.

Si denotamos por:

$O_i$  = nº de elementos de la muestra con el carácter  $x_i$ .

$p_i$  = probabilidad teórica de que la variable aleatoria X tome el valor  $x_i$  verificándose que  $\sum_{i=1}^k p_i = 1$ .  
Si tenemos una muestra de tamaño n, el número de elementos que cabe esperar que tomen el valor

$x_i$  es:  $e_i = n \cdot p_i$  verificándose que  $\sum_{i=1}^k e_i = n$

Podremos formar la siguiente tabla:

Variable X	$x_1$	$x_2$	.....	$x_k$
Frecuencias observadas	$O_1$	$O_2$	.....	$O_k$
Frecuencias esperadas	$e_1$	$e_2$	.....	$e_k$

Consideramos como hipótesis nula e hipótesis alternativa a:

$H_0$ : la distribución empírica se ajusta a la distribución teórica considerada.

$H_1$ : Se rechaza el ajuste.

Evidentemente, si aceptamos la hipótesis nula(aceptamos el ajuste), las diferencias entre los valores observados y los valores esperados son debidas al azar y podemos decir que no existen evidencias para rechazar dicha hipótesis; en otro caso diremos existen diferencias significativas para el nivel de significación marcado entre ambas distribuciones, no pudiendo atribuirse las diferencias entre las distribuciones empíricas y observadas al azar.

El estadístico que se utilizará para dicho contraste será:

$$T = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^k \frac{O_i^2}{e_i} - n$$

Pearson demostró que la distribución de dicho estadístico es una  $\chi^2$  con  $k-1$  grados de libertad en el caso de no existir discrepancias entre los valores observados y los esperados.

Se acepta  $H_0$  si:  $T < \chi^2_{\alpha}(k-1)$  (REGIÓN DE ACEPTACIÓN)

Se rechaza  $H_0$  si:  $T \geq \chi^2_{\alpha}(k-1)$  (REGIÓN CRÍTICA).

A la hora de aplicarlo correctamente tenemos que realizar las siguientes consideraciones:

A.- Las frecuencias esperadas de las distintas modalidades deben ser superiores a cinco; en caso de ocurrir se deben agrupar clases contiguas en una sola clase hasta lograr que la nueva frecuencia sea mayor que cinco. Esto supone cambiar la distribución teórica con la consiguiente pérdida de información.

B.- Si para obtener las frecuencias esperadas se necesitan estimar  $p$  parámetros entonces los grados de libertad de la  $\chi^2$  son  $k-p$  si son independientes y  $k-p-1$  si son excluyentes las modalidades.

C.- Se puede aplicar a las distribuciones continuas como discretas.

Ejemplo : Se desea probar si el número de rayos gamma emitidos por segundo por cierta sustancia radiactiva es una variable aleatoria que tiene la distribución de Poisson de parámetro  $\lambda = 2,4$ . Utilice los siguientes datos obtenidos en 300 intervalos de segundo para probar dicha hipótesis a un nivel de significación 0,05.

Nº de rayos gamma	0	1	2	3	4	5	6	7 o más
Valores observados	20	48	65	75	45	34	9	4

La población en estudio son los rayos gamma emitidos por una sustancia radiactiva.

Se trata de contrastar las hipótesis:

$H_0$ : La distribución observada sigue una distribución  $P(2,4)$ .

$H_1$ : La distribución observada no sigue una distribución  $P(2,4)$

El nivel de confianza es  $1-\alpha = 0,95$  puesto que el nivel de significación es  $\alpha = 0,05$  y el tamaño de la muestra  $n=300$

El estadístico del contraste será el definido anteriormente, para ello deberemos calcular los valores esperados bajo la hipótesis  $H_0$ .

Calculamos los valores esperados para una Poisson de parámetro  $\lambda = 2,4$ . Se tiene que las probabilidades de una Poisson de parámetro  $\lambda = 2,4$  son (mirando en la tabla de la distribución de Poisson):

Valores de X	0	1	2	3	4	5	6	7 o más
Probabilidades	0,0907	0,2177	0,2613	0,2090	0,1254	0,0602	0,0241	0,0116

Para hallar los valores esperados bastará con multiplicar las probabilidades por el número de observaciones, obteniéndose:

Valores de X	0	1	2	3	4	5	6	7 o más
Valores esperados	27,21	65,31	78,39	62,7	37,62	18,06	7,23	3,48

Como se observa, existe un valor esperado menor que 5, por lo que debemos de agrupar dos clases contiguas, en este caso, las clases 6 y 7 o más, con lo cual las distribuciones quedarían:

Valores de X	0	1	2	3	4	5	6 o más
Valores esperados	27,21	65,31	78,39	62,7	37,62	18,06	10,71
Valores observados	20	48	65	75	45	34	13

Aplicando la fórmula resulta que el valor del estadístico para dicho contraste vale

$$T = \frac{(20 - 27,21)^2}{27,21} + \dots + \frac{(13 - 10,71)^2}{10,71} = 27,2047$$

Calculamos el valor del punto crítico de una chi-cuadrado con un valor de  $\alpha = 0,05$  (nivel de significación) y  $(7-1)$  grados de libertad, es decir,  $\chi^2_{0,05}(6) = 12,592$  mirando en las tablas

La región crítica es:  $T$  mayor o igual que  $12,592$ , es decir, el intervalo  $(12,592, +\infty)$

Como el valor del estadístico  $T$  es mayor que el punto crítico se rechaza la hipótesis  $H_0$ .  $T=27,2047 > 12,592 = \chi^2_{0,05}(6)$

Luego tenemos evidencias de que la distribución empírica no sigue una distribución  $P(2,4)$ .

OBSERVACIÓN: En el caso de que el ajuste sea por una distribución normal es preferible aplicar otro test más potente: test de Kolmogorov.

## 2.2. Contraste de dependencia o independencia de caracteres.

Deseamos saber si dos caracteres  $X$  e  $Y$  de una población son dependientes o independientes.

Suponemos que las modalidades que presentan cada una de las variables  $X$  e  $Y$  son:

$X: x_1, x_2, \dots, x_k$

$Y: y_1, y_2, \dots, y_m$

y se ha tomado una muestra de tamaño  $n$  midiéndose dichas características  $X$  e  $Y$  en cada uno de los elementos de la muestra.

Si denotamos por:

$O_{ij}$  = nº de elementos que presentan la característica  $x_i$  e  $y_j$

$e_{ij}$  = nº de elementos esperados que presenten los valores  $x_i$  ,  $y_j$  si las variables son independientes.

Podríamos formar la siguiente tabla de contingencia en la que aparecen las frecuencias empíricas y las teóricas:

Y	$y_1$	.....	$Y_j$	.....	$y_m$	Frecuencias absolutas X
X						
$x_1$	$O_{11}$ $e_{11}$	.....	$O_{1j}$ $e_{1j}$	.....	$O_{1m}$ $e_{1m}$	$O_{x1}$
.....	.....	.....	.....	.....	.....	.....
$x_i$	$O_{i1}$ $e_{i1}$	.....	$O_{ij}$ $e_{ij}$	.....	$O_{im}$ $e_{im}$	$O_{xi}$
.....	.....	.....	.....	.....	.....	.....
$x_k$	$O_{k1}$ $e_{k1}$	.....	$O_{kj}$ $e_{kj}$	.....	$O_{km}$ $e_{km}$	$O_{xk}$
Frecuencias absolutas Y	$O_{y1}$	.....	$O_{yj}$	.....	$O_{ym}$	$n$

Para el cálculo de las frecuencias teóricas podemos utilizar la siguiente fórmula si las dos variables son independientes

$$e_{ij} = p_{ij} \cdot n = \frac{O_{xi}}{n} \cdot \frac{O_{yj}}{n} \cdot n = \frac{(total\ de\ la\ fila\ i) \cdot (total\ de\ la\ columna\ j)}{n}$$

$$i=1,2,\dots,k \quad j=1,2,\dots,m$$

Consideramos como hipótesis nula e hipótesis alternativa a:

$H_0$ :  $X$  e  $Y$  son independientes.

$H_1$ :  $X$  e  $Y$  no son independientes.

Si aceptamos la hipótesis nula, podemos considerar que no tenemos evidencias que nos hagan suponer una dependencia entre las dos variables a un nivel de confianza de  $1-\alpha$ .

Consideramos como estadístico del contraste:

$$T = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^k \sum_{j=1}^m \frac{O_{ij}^2}{e_{ij}} - n$$

La distribución de dicho estadístico es una  $\chi^2$  con  $(k-1)(m-1)$  grados de libertad en caso de que las variables sean independientes a un nivel de confianza  $1-\alpha$ .

Se acepta  $H_0$  si:  $T < \chi^2_{\alpha}(k-1)(m-1)$  (REGIÓN DE ACEPTACIÓN)

Se rechaza  $H_0$  si:  $T \geq \chi^2_{\alpha}(k-1)(m-1)$  (REGIÓN CRÍTICA).

Ejemplo El consejo de administración de Telefónica desea conocer si la opinión, Y, de sus accionistas respecto a una posible fusión es independiente del número de acciones, X, que poseen. Una muestra de 500 accionistas proporciona la siguiente tabla:

Y X	A favor	En contra	Indecisos	Total
Menos de 200	25	18	21	64
200- 1000	93	62	67	222
Más de 1000	82	70	62	214
Total	200	150	150	500

Contrastar a un nivel de confianza del 99,5% la independencia de las variables X e Y.

La población en estudio son los accionistas de Telefónica y deseamos ver si existe dependencia entre el número de acciones y la opinión acerca de una posible fusión.

Se trata de un test no paramétrico donde las hipótesis nula y alternativa son:

$H_0$ : X e Y son independientes

$H_1$ : X e Y no son independientes.

El nivel de confianza es  $1-\alpha = 0,995$ , luego  $\alpha = 0,005$  y el tamaño muestral  $n=500$

Calculamos los valores esperados  $e_{ij}$  bajo la hipótesis nula (independencia de X e Y) aplicando la fórmula  $\frac{(total\ de\ la\ fila\ i) \cdot (total\ de\ la\ columna\ j)}{n}$

donde n es el tamaño de la muestra, 500.

Por ejemplo  $e_{11}=64.200/500=25,6$        $e_{12}=64.150/500=19,2$  .....

La tabla de los valores esperados sería:

Y	A favor	En contra	Indecisos	Total
X				
Menos de 200	25,6	19,2	19,2	64
200-1000	88,8	66,6	66,6	222
Más de 1000	85,6	64,2	64,2	214
Total	200	150	150	500

El valor del estadístico experimental T vale:

$$T = \frac{(25 - 25,6)^2}{25,6} + \dots + \frac{(62 - 64,2)^2}{64,2} = 1,53$$

El valor del punto crítico es el valor de una chi-cuadrado con  $(3-1) \cdot (3-1) = 4$  grados de libertad y  $1-\alpha = 0,955$  que mirando en las tablas 6 nos da:  $\chi^2_{0,005}(4) = 14,860$

La región crítica es, es decir, rechazamos  $H_0$  si:  $T \geq \chi^2_{0,005}(4)$

Como  $T=1,53$  es menor que 14,86 se acepta  $H_0$  y podemos decir que no tenemos evidencias de que X e Y sean dependientes y se acepta la hipótesis de que la opinión de los accionistas es independiente del número de acciones que poseen con un riesgo del 0,5%.

### 2.3. Test de homogeneidad de varias muestras.

Se trata de determinar si varias muestras que estudian el mismo carácter A han sido tomadas o no de la misma población, respecto de dicha característica A.

Supongamos que tenemos k muestras de tamaños  $n_1, n_2, \dots, n_k$  siendo  $y_1, y_2, \dots, y_k$  los elementos de cada muestra que presentan una determinada característica A y el resto no la presentan.

Si suponemos que todas las muestras provienen de la misma población, la proporción de elementos que presentan la característica A sería:

$$p = \frac{y_1 + y_2 + \dots + y_k}{n_1 + n_2 + \dots + n_k}$$

Si suponemos que las muestras provienen de la misma población, los valores esperados para la característica A en cada muestra serían:  $n_1 \cdot p, n_2 \cdot p, n_3 \cdot p, \dots, n_k \cdot p$ .

Podríamos formar la siguiente tabla de contingencia en la que aparecen los valores observados y los valores esperados:

Muestras	Presentan el carácter A Se esperan con el carácter A	No presentan el carácter A Se esperan sin el carácter A	Tamaño de las muestras
Primera muestra	$y_1$ $n_1 \cdot p$	$n_1 - y_1$ $n_1(1-p)$	$n_1$
.....	.....	.....	.....
i-ésima muestra	$y_i$ $n_i \cdot p$	$n_i - y_i$ $n_i(1-p)$	$n_i$
.....	.....	.....	.....
k-ésima muestra	$y_k$ $n_k \cdot p$	$n_k - y_k$ $n_k(1-p)$	$n_k$

Consideramos como hipótesis nula e hipótesis alternativa a:  
 $H_0$ : todas las muestras provienen de la misma de la población.  
 $H_1$ : se rechaza que provengan de la misma población.

Si aceptamos la hipótesis nula, podemos considerar que las muestras provienen de la misma población y las diferencias entre los valores observados y los valores esperados son debidas al azar.

$$\text{El estadístico que se utilizará será: } T = \frac{1}{p(1-p)} \sum_{i=1}^k \frac{(y_i - n_i \cdot p)^2}{n_i}$$

La distribución de dicho estadístico es una  $\chi^2$  con  $k-1$  grados de libertad en el caso de no existir discrepancias entre los valores observados y los esperados a un nivel de confianza  $1-\alpha$

Se acepta  $H_0$  si:  $T < \chi^2_{\alpha}(k-1)$  (REGIÓN DE ACEPTACIÓN)

Se rechaza  $H_0$  si:  $T \geq \chi^2_{\alpha}(k-1)$  (REGIÓN CRÍTICA)

En el caso de que los elementos de las muestras se clasifiquen en más de dos categorías, el análisis se realiza como en el caso de un test de independencia o dependencia entre variables, donde la tabla que se obtendría sería similar a la anterior, por filas aparecen las muestras y por columnas las distintas categorías. El estadístico sería el mismo que en el caso de independencia de variables y los valores esperados se calcularían de igual forma y la hipótesis nula será  $H_0$ : todas las distribuciones se distribuyen homogéneamente.

Ejemplo: En un hospital se somete a examen la eficacia de cinco medicamentos a un determinado número de pacientes que aparece reflejado en la siguiente tabla, determinándose si al final del tratamiento mejoran o no.

Tratamiento	A	B	C	D	E	Total
Nº de pacientes	50	52	46	54	48	250
Pacientes mejorados	11	9	8	17	7	50

¿Existe diferencia entre los diferentes medicamentos a un nivel de significación  $0,05$  ?

La población son los medicamentos analizados y deseamos comprobar si alguno de los medicamentos es más eficaz que los otros, es decir, si alguno de ellos puede considerarse que muestra diferencias significativas respecto de los demás.

Se trata de contrastar las hipótesis

$H_0$ : no existen diferencias significativas entre los cuatro medicamentos, es decir, provienen de la misma población.

$H_1$ : existen diferencias significativas en alguno de ellos.

El nivel de confianza es  $1-\alpha = 0,95$ , luego  $\alpha = 0,05$  y el tamaño muestral  $n=250$

Si denominamos

$y_i$  = número de pacientes mejorados con el medicamento  $i$ .

$n_i$  = número de pacientes que utilizan el medicamento  $i$ .

Y suponemos que no hay diferencias significativas entre los cuatro medicamentos, es decir, provienen de la misma población ( $H_0$  cierta) tendríamos que:

$$p = \frac{y_1 + y_2 + \dots + y_k}{n_1 + n_2 + \dots + n_k} = 50/250 = 0,2$$
 será la proporción de pacientes que mejoran si todas las muestras provienen de la misma población, es decir, son homogéneas.

Calculando los valores esperados,  $n_i \cdot p$ , tendríamos la siguiente tabla:

Tratamiento	Pacientes mejorados Pacientes esperados que mejoren	Pacientes no mejorados Pacientes esperados sin mejoría	Tamaño de la muestra
A	11 $50 \cdot 0,2 = 10$	39 $50 \cdot (1 - 0,2) = 40$	50
B	9 $52 \cdot 0,2 = 10,4$	43 $52 \cdot (1 - 0,2) = 41,6$	52
C	8 $46 \cdot 0,2 = 9,2$	38 $46 \cdot (1 - 0,2) = 36,8$	46
D	17 $54 \cdot 0,2 = 10,8$	37 $54 \cdot (1 - 0,2) = 43,2$	54
E	7 $48 \cdot 0,2 = 9,6$	41 $48 \cdot (1 - 0,2) = 37,4$	48

$$T = \frac{1}{0,2(1-0,2)} \left( \frac{(11-10)^2}{50} + \dots + \frac{(7-9,6)^2}{48} \right) = 5,8855$$

El valor del estadístico es:  $T =$

El valor del punto crítico es el valor de una chi-cuadrado con  $(5-1) = 4$  grados de libertad y 0,05 que mirando en las tablas 6 nos da:  $\chi^2_{0,05}(4) = 9,488$

La región crítica es, es decir, rechazamos  $H_0$  si:  $T \geq \chi^2_{0,05}(4) = 9,488$

Luego como  $T=5,8855$  es menor que 9,488 aceptamos  $H_0$ , es decir, no existe diferencia entre los diferentes medicamentos con un nivel de confianza del 95%, para la mejora de los pacientes al finalizar el tratamiento.

### 3. Problemas Resueltos.

1. La calificación final de los estudiantes de un curso de estadística se clasifica por carreras. ¿Se podría concluir que existe una asociación entre la carrera y la calificación final con un nivel de significación del 0,05?

	Psicología	Medicina	Farmacia
Sobresaliente	11	28	22
Notable	20	34	30
Aprobado	22	8	13
Suspenso	6	4	9

2. En un laboratorio se observó el número de partículas  $\square$  que llegan a una determinada zona procedentes de una sustancia radiactiva en un corto espacio de tiempo siempre igual, anotándose los resultados en la siguiente tabla:

Nº de partículas	0	1	2	3	4	5
Nº de periodos de tiempo	120	200	140	20	10	2

Se pide:

Ajustar a una distribución de Poisson.

Calcular la probabilidad con que llegan.

Verificar si el ajuste es correcto mediante una  $\chi^2$ , con un nivel de confianza  $\alpha = 0,05$

3. Una compañía de venta de libros quiere saber si el volumen de ventas, V, de sus distribuidores es independiente del carácter C, de los mismos. Para ello, recoge los siguientes datos de una muestra de 250 vendedores:

V C	Bajo	Medio	Alto
Antipáticos	38	29	9
Normales	30	42	7
Simpáticos	32	59	4

Realiza un test de independencia a nivel de significación de  $\alpha = 0,1$

Puedes utilizar la opción *Resolución de problemas con el ordenador* [\*] para comprobar que las soluciones son:

1.- Test de independencia:

Valor del estadístico del contraste T= ;

Valor del punto crítico=

2.- Test de Bondad de Ajuste:

Valor del estadístico del contraste T= ;

Valor del punto crítico=

3.- Test de independencia:

Valor del estadístico del contraste T= ;

Valor del punto crítico=

## **[\*] Resolución de problemas con el ordenador**

Puedes resolver los problemas propuestos (o cualquier otro del mismo tipo), seleccionando la opción adecuada:

Para su resolución se utiliza un programa en [JAVA](#) (test1.html) en el que se pide que se seleccione el tipo de problema a resolver, el nivel de significación deseado y los datos del problema. Como resultado nos muestra el valor del estadístico experimental y teórico del contraste y la decisión que se debe tomar. Para ello basta con pulsar en JAVA.

## 5. Problemas propuestos:

1. Disponemos de un lote de 400 dispositivos electrónicos, cada uno de los cuales consta de 3 circuitos, observándose el siguiente número de defectos en cada dispositivo electrónico. Aproximar dicha distribución por una binomial, viendo si es un buen ajuste con un nivel de significación de 0,05.

Nº de circuitos defectuosos	0	1	2	3
Nº de dispositivos	49	153	148	50

2. Un partido político desea conocer si existe relación entre el grado de aceptación de su programa electoral(X) y el sueldo medio(Y) expresado en miles de ptas. Para ello selecciona al azar a 50 personas y obtiene los siguientes datos:

X Y	A favor	En contra	Indecisos	Total
Menos de 100	3	2	2	7
Entre 100 y 300	9	6	7	22
Más de 300	7	7	7	21
Total	19	15	16	50

Aceptaría el partido político que el sueldo medio de los electores explicase el grado de aceptación del programa electoral a un nivel de confianza del 95%.

3. Se han aplicado tres métodos de empaquetado de dispositivos electrónicos durante un periodo de 4 meses; al final de cada mes se realiza un recuento del número de paquetes estropeados, obteniéndose los siguientes datos:

Meses	1	2	3	4	Total
Método					
A	6	8	8	8	30
B	10	12	9	14	45
C	9	11	13	15	48
Total	25	31	30	37	123

Comprobar con un nivel de confianza del 95% si los tres métodos de envasado son igualmente buenos (homogéneos) o existen diferencias significativas entre ellos.

## 6. Resumen:

- Aplicaciones de la distribución  $\chi^2$

- **Test de bondad de ajuste a una distribución teórica:**

$$\text{Estadístico: } T = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} \quad \text{Región crítica: } T \geq \chi_{\alpha}^2(k-1)$$

- **Test de independencia o dependencia de dos variables:**

$$\text{Estadístico: } T = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \quad \text{Región crítica } T \geq \chi_{\alpha}^2(k-1)(m-1)$$

- **Test de  $\square$ omogeneidad de muestras:**

$$\text{Estadístico: } T = \frac{1}{p(1-p)} \sum_{i=1}^k \frac{(y_i - n_i \cdot p)^2}{n_i} \quad \text{Región crítica } T \geq \chi_{\alpha}^2(k-1)$$

## 7. Bibliografía:

- **Encuestas y precios.** Nortes Checa, A. Ed. Síntesis Madrid 1995
- **Prensa y educación matemática.** Fernández Cano, A. Rico Romero, L. Ed. Síntesis Madrid 1992.
- **Azar y probabilidad.** Díaz Godino, J. Batanero Bernabeú, M. Cañizares Castellano, M. Ed. Síntesis Madrid 1996.
- **Introducción a la teoría de probabilidades y sus aplicaciones.** Feller, W. Ed. Limusa-Waley. México 1973
- **Cálculo de probabilidades.** Baró Llinás, J. Ed. Paramón Barcelona 1987.
- **Hacer estadística.** Sanchís, C. Salillas, J. Riera, T. Fontanet, G. Ed. Alhambra. Madrid 1987.
- **Introducción a la Estadística** Wonnacott, T.H Wonnacott, R.J. Ed. Limusa México 1989.
- **Introducción a la Estadística.** Nieto de Alba, U. Ed. Aguilar Madrid 1972
- **Introducción a la Estadística teórica.** Arnáiz, G. ED. Lex-Nova Valladolid 1978.
- **Introducción a la Estadística para las Ciencias Sociales.** Peña, D. Romo, J. Ed. Mc-Graw-Hill Madrid 1997
- **Estadística elemental moderna.** García, A. Ed. Ariel. Barcelona 1973
- **Métodos Estadísticos.** Ríos, S. Ed. del Castillo Madrid 1974
- **Métodos Estadísticos Aplicados I y II** Visauta, B. Batalle, P. Ed. PPU Barcelona 1986.
- **Fundamentos y métodos de Estadística.** López Cachero, M. Ed. Pirámide Madrid 1993.
- **Análisis estadístico.** Ríos, S. Ed. Paraninfo. Madrid 1984.
- **Ejercicios resueltos de Estadística.** Sierra, M.A. Ed. Ceura Madrid 1987.
- **Curso de Estadística descriptiva** Calot, G. Ed. Paraninfo Madrid 1988.
- **Métodos matemáticos de Estadística.** Cramer, H. Ed. Aguilar. Madrid 1968.
- **Métodos y aplicaciones del muestreo.** Azorín, F. Sánchez-Crespo, J.L. Ed. Alianza Madrid 1986
- **Curso de muestreo y aplicaciones.** Azorín, F. Ed. Aguilar Madrid 1972.
- **Técnicas de muestreo.** Cochran, W.G. Ed. CECSA México 1971.
- **Inferencia Estadística.** Baró Llinás, J. Ed. Paramón Barcelona 1988
- **Fundamentos de Estadística.** Arnal, J.I. Omedes, A. Daimón Barcelona 1981
- **Estadística aplicada a las Ciencias Humanas.** Echevarría, B. Daimón Barcelona 1982
- **Métodos estadísticos aplicados a las Ciencias Sociales.** Glass, Stanley. Ed. Prentice-Hall Internacional Madrid 1978
- **Estadística elemental.** Johnson, R. Ed. Trillas México 1976
- **Estadística aplicada a las Ciencias Sociales.** Alcaide, A. Ed. Pirámide. Madrid 1979
- **Iniciación Estadística.** Ríos, S. Ed. Paraninfo Madrid 1991.
- **Estadística descriptiva.** Baró Llinás. Ed. Paramón Barcelona 1985
- **Diccionario práctico Estadística.** Sierra Bravo, R. Ed. Paraninfo Madrid 1991.

