

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Facultad de Ciencias

ALGUNOS MODELOS MATEMATICOS DE PATRONES DE  
ASENTAMIENTO Y SU USO EN ARQUEOLOGIA

Tesis presentada por  
MA. TERESA GOMEZ LOPEZ  
Para la obtención de la licenciatura  
en matemáticas.  
México D. F.

Diciembre de 1978



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis padres

A mis hermanas

Agradezco la revisión y sugerencias a este trabajo de Yoko Sugiura, sin su ayuda no hubiera sido posible conocer el nivel adecuado de la exposición.

Deseo también agradecer la cooperación en la comprensión de algunos - problemas a José Luis Sánchez y a Ruben Hernández

Gracias a Guillermo Espinosa por el estímulo y dirección en la investigación de este trabajo.

# I N D I C E

INTRODUCCION	i
CAPITULO 1.- ALGUNAS DISTRIBUCIONES DE PROBABILIDAD	1
1. Distribución Binomial	2
2. Distribución Poisson	14
3. Distribución XI Cuadrada	20
4. Distribución Normal	25
CAPITULO 2.- ANALISIS DE PATRONES DE PUNTOS	34
1. Métodos de cuadrantes	36
2. Prueba de bondad de ajuste xi cuadrada	37
2.1. Aplicación de la prueba xi cuadrada	37
2.2. Limitaciones de la prueba xi cuadrada	41
3. Prueba de la razón varianza/media	43
3.1. Aplicación de la prueba varianza/media	
3.2. Limitaciones de la prueba varianza/media	
4. Problemas de los métodos de cuadrantes	48
CAPITULO 3.- ANALISIS AL VECINO MAS CERCANO	51
1. Medida de la distancia al vecino más cercano de Clark y Evans	51
2. Distribución de la distancia al punto más cercano en una población dispersada aleatoriamente	52
3. Prueba de significancia	56
4. Aplicación de la prueba a un patrón hipotético	57
5. Problemas del proceso	59
5.1. Nota de Diggle a la prueba de Clark y Evans	61
6. Estadístico de Pielou	62
6.1. Aplicación del estadístico	63
CAPITULO 4.- ALGUNAS APLICACIONES DE LAS PRUEBAS	65
1. Aplicaciones de las pruebas a patrones hipotéticos	66
2. Aplicación de las pruebas en Arqueología	73
CONCLUSION	79
TABLAS	80
BIBLIOGRAFIA	83

## I N T R O D U C C I O N .

El principal objetivo de este trabajo es hacer accesible a los arqueólogos algunos métodos que existen para un estudio más detallado y sistemático de patrones espaciales basados en modelos matemáticos. Este trabajo surgió de la necesidad que la arqueóloga Yoko Sugiura -- tenía de asesoría matemática.

Los métodos para el análisis de patrones espaciales que han sido usados recientemente en arqueología, han sido introducidos y adaptados de otros campos de estudio como geografía y ecología en donde se han desarrollado.

El patrón de distribución es uno de los principales instrumentos de la investigación arqueológica. Por ejemplo la distribución de artefactos dentro de un sitio puede ser usada para definir las localizaciones de actividades.

De las pruebas que se mostrarán para el estudio de patrones de asentamiento, se señalan los requisitos para su aplicación y algunas de sus ventajas y desventajas, como el campo es muy extenso no es posible cubrir todos los aspectos ni considerar todos los problemas -- asociados .

Una de las formas para realizar un estudio del patrón de distribución anteriormente, consistía en la interpretación visual. Este tipo de estudio resulta subjetivo y no existe una forma sistemática y cuantitativa de probarlo. Surge la necesidad de desarrollar el análisis espacial de datos arqueológicos debido a que se cuenta con grandes

cantidades de información distribucional que no es fácil de examinar sin algún avance en los métodos disponibles.

Se verá que la forma de algunos patrones es comparable a la distribución Poisson, que es el modelo matemático que sirve de base para las pruebas de aleatoriedad.

Es necesario diferenciar los conceptos de distribución de probabilidad y distribución de puntos, el primero es un modelo matemático y el segundo es utilizado en un sentido coloquial como sinónimo de patrón, arreglo, etc.

Algo importante de señalar es que estas pruebas únicamente forman parte de un estudio arqueológico completo del proceso espacial (cambio que sufre la distribución espacial tanto en escala regional como dentro de un sitio dado), pues la identificación de un patrón estructurado no significa que todo se ha explicado, sino únicamente es una ayuda en la interpretación del proceso espacial, que produjo el patrón.

El arqueólogo interesado en ampliar sus conocimientos de estadística puede consultar el libro de Hoel(1977), el cual es de fácil acceso para la persona que no tiene conocimientos profundos en la materia.

En el primer capítulo veremos algunas distribuciones de probabilidad que se utilizan en las pruebas de aleatoriedad y en las pruebas de significancia.

Es una breve presentación que pretende ser objetiva para que el arqueólogo pueda comprender mejor las técnicas que se emplean en el análisis de patrones.

En el segundo y tercer capítulos se presentan algunas pruebas sencillas para el estudio de patrones de puntos, sus condiciones necesarias, sus desarrollos y los problemas de su aplicación.

En el último capítulo se muestra la aplicación de las pruebas a ejemplos sintéticos y algunas aplicaciones que se han hecho en arqueología y sus conclusiones.

Un interés básico de este trabajo es que el arqueólogo se interese en estas técnicas de tal manera que seleccione datos adecuados para su uso.

## CAPITULO 1

### ALGUNAS DISTRIBUCIONES DE PROBABILIDAD.

Una distribución de probabilidad es un modelo matemático para la distribución de frecuencia real. En este capítulo veremos ciertas distribuciones que se utilizarán en los capítulos siguientes.

Generalmente el conjunto de datos que se puede obtener es sólo una muestra de la población, esta muestra proporciona información sobre la población, y a partir de ella se puede obtener una distribución empírica, la cual sugiere un tipo de distribución de la población; algunas veces se encuentran varias clases de distribuciones empíricas que infieren un modelo general de distribuciones, conocidas como distribuciones de frecuencia teóricas. Estas distribuciones intentan interpretar una distribución real.

Veremos la distribución binomial, la distribución Poisson, la distribución  $\chi^2$  cuadrada y la distribución normal; de una forma superficial intentando que se conozca como se comportan.

## 1. DISTRIBUCION BINOMIAL.

La distribución binomial es debida a Santiago Bernoulli alrededor del año 1700.

Consideremos un experimento en el cual, sus resultados únicamente pueden clasificarse como ocurrencia o no ocurrencia de un determinado evento A; (como por ejemplo la ocurrencia de un águila en el lanzamiento de una moneda, o la presencia de un dos al arrojar un dado). Si ocurre el evento A le llamaremos éxito y si no ocurre fallo; el experimento se repetirá  $n$  veces, y el número de éxitos que resulten de las  $n$  repeticiones se denotará con la letra  $x$ .

Como ejemplo de un experimento de este tipo, considérese el lanzamiento de una moneda, supongamos que el éxito sea obtener un águila; si el experimento se repite dos veces ( $n=2$ ),  $x$  será el número de águilas que se obtiene de los dos experimentos. El problema es calcular las probabilidades de los diferentes valores que toma  $x$  y conocer como se distribuyen.

Primero consideremos los resultados posibles del experimento, que en este caso son cuatro y registramos el número de águilas que es el valor que toma  $x$ ; estos datos se muestran en la tabla 1. Obsérvese que los valores que  $x$  puede tomar son: 0, 1, 2.

Tabla 1

Resultado	AA	AS	SA	SS
Valor de $x$	2	1	1	0

(Águila= A, Sol= S).

La probabilidad de obtener uno de los cuatro resultados por ejemplo, dos águilas sería  $1/4$ , ya que se espera que cada uno de los cuatro resultados ocurra con la misma frecuencia, por lo cual, a cada uno se le asigna la misma probabilidad. Denotamos con  $P$  la probabilidad, así que la probabili-

dad de dos águilas quedaría  $P(2) = 1/4$  ; la probabilidad de cero águilas se denota  $P(0) = 1/4$  .

Nótese que la probabilidad de obtener un águila está compuesta de dos posibles resultados AS y SA, por lo cual, su probabilidad se obtiene sumando las probabilidades de los resultados de que esta compuesta, y como cada uno de ellos ocurre con probabilidad de  $1/4$  , la probabilidad de un águila sería,

$$P(1) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}.$$

Las probabilidades de los valores que toma  $x$ , se presentan en la tabla 2.

Tabla 2

x	0	1	2
P(x)	1/4	1/2	1/4

La gráfica de la distribución de  $x$  correspondiente a la tabla 2 se representa en la figura 1.

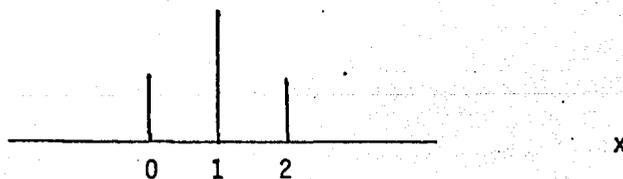


Fig. 1 Distribución para el número de águilas que se obtiene al lanzar una moneda dos veces.

Otro ejemplo de un experimento repetitivo que consideraremos es el lanzamiento de un dado, suponiendo como éxito la obtención de un seis; el experimento se repetirá tres veces ( $n=3$ ), y  $x$  será el número de veces que aparezca un seis, en los tres experimentos; nos interesa conocer la distribución de  $x$  .

Procediendo como en el ejemplo anterior registramos si ocurre un éxito o un fallo, que denotaremos con E y F respectivamente; en los ocho resultados posibles obtenemos los valores de  $x$  que se muestran en la tabla 3.

Tabla 3

Resultado	EEE	EEF	EFE	EFF	FEE	FEF	FFE	FFF
Valor de $x$	3	2	2	1	2	1	1	0

Los valores que  $x$  toma son: 0,1,2,3. Nos interesa conocer la probabilidad de cada uno de estos valores. Como cada una de las tiradas del dado es independiente, la probabilidad de un éxito es  $1/6$ , ya que se desea obtener una determinada cara del dado, el cual tiene seis; se espera que cada una de las seis caras ocurra con la misma probabilidad. La probabilidad de fallo es  $5/6$ , que es la suma de las probabilidades de las cinco caras restantes que no presentan un éxito; por lo cual, la probabilidad de que ocurran dos éxitos y un fallo, la obtenemos multiplicando las probabilidades correspondientes a cada uno de ellos, es decir,

$$P(\text{EEF}) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$$

Así también se obtiene la probabilidad de un éxito y dos fallos.

$$P(\text{EFF}) = \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$$

La probabilidad de éxito, fallo, éxito quedaría,

$$P(\text{EFE}) = \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right).$$

De esta manera se pueden obtener cada uno de los ocho resultados que se muestran en la tabla 4.

Tabla 4

Resultado	EEE	EEF	EFE	FEE	EFF	FEF	FFE	FFF
probabilidad	$\left(\frac{1}{6}\right)^3$	$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$\left(\frac{5}{6}\right)^3$

Ahora procederemos a calcular las probabilidades de los valores que toma  $x$ : 0,1,2,3. La probabilidad de que  $x$  sea igual a 1, consta de tres posibilidades EFF, FEF y FFE; y se obtiene sumando la probabilidad de cada una de estas componentes; es decir,

$$P(1) = \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 + \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 + \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 = 3 \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$$

La probabilidad de dos éxitos

$$P(2) = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) = 3 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$$

De esta forma se obtienen las probabilidades para los valores de  $x$ , los cuales se representan en la tabla 5, que proporciona la distribución de  $x$ , y su gráfica correspondiente se presenta en la figura 2.

Tabla 5

$x$	0	1	2	3
$P(x)$	$\left(\frac{5}{6}\right)^3$	$3 \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$3 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right)^3$

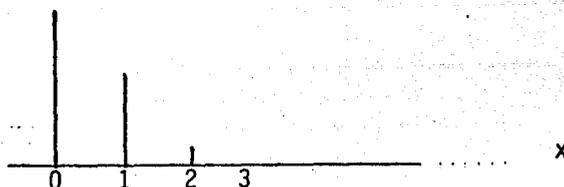


Fig. 2 Distribución para el número de veces que se obtiene un seis al lanzar un dado tres veces.

En los ejemplos vistos, el proceso que se utilizó consistió en observar los posibles resultados del experimento y se calculó la probabilidad para cada uno de ellos, asociando a cada resultado el valor de la  $x$  correspondiente, y para obtener las probabilidades de valores compuestos de  $x$  se suman las probabilidades de los resultados de que están compuestos. Estas probabilidades dan la distribución de probabilidad de  $x$ .

Como no siempre es fácil realizar los cálculos para problemas de este tipo es necesaria una formulación general; para ello, considerando el problema en general, se tendría un experimento en el cual, los resultados se clasifican como éxito o fallo, supóngase que se conoce la probabilidad de obtener un éxito y la correspondiente probabilidad de obtener un fallo, a la primera la designaremos  $p$  y a la segunda  $q$ ; obtendríamos que  $p+q = 1$ , de donde  $q = 1-p$ .

El experimento se repite  $n$  veces y supongamos que se obtuvieron los resultados de la siguiente manera,  $x$  éxitos y  $n-x$  fallos; se tendría una sucesión así,

$$\underbrace{EEE\dots E}_x \underbrace{FFF\dots F}_{n-x}$$

Dado que cada experimento es independiente, la probabilidad de obtener esta sucesión es,

$$p: \underbrace{p.p\dots p}_x \underbrace{q.q.q\dots q}_{n-x} = p^x q^{n-x}$$

Si obtuviéramos otra sucesión como la siguiente en donde, primero ocurrieran dos fallos,  $x$  éxitos y los restantes fueran fallos, se tendría:

$$\underbrace{FFEE\dots E}_x \underbrace{FF\dots F}_{n-x-2}$$

La probabilidad de obtener esta sucesión es la misma que la probabilidad de obtener la primera; ya que,

$$q \cdot \underbrace{p \cdot p \dots p}_{x} \cdot \underbrace{q \cdot q \dots q}_{n-x-2} = p^x q^{n-x}$$

Si obtuviéramos otra sucesión, la probabilidad correspondiente sería la misma; por lo tanto, cualquiera que sea la forma en que este combinada la serie que tenga  $x$  éxitos y  $n-x$  fallos, la probabilidad de obtenerla es la misma.

El número de maneras en que el experimento puede ocurrir, es igual a el número de sucesiones diferentes de  $n$  elementos, en que pueden combinarse  $x$  letras E y  $n-x$  letras F; que es igual a el número de formas diferentes en que se pueden escoger  $x$  posiciones de las  $n$  disponibles para colocar la letra E y las  $n-x$  posiciones que restan para colocar la letra F para ello existe la fórmula siguiente:

$$C_n^x = \frac{n!}{x! (n-x)!} \quad (1)$$

donde  $n!$  se lee  $n$  factorial y es el producto de los números 1 hasta  $n$ ; como ejemplo,  $6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6$ ;  $0!$  es definido como 1.

Como cada una de las sucesiones representa una de las formas particulares en que puede suceder el experimento, y cada una de ellas tiene la misma probabilidad de ocurrir ( $p^x q^{n-x}$ ), entonces la probabilidad de obtener  $x$  éxitos se obtiene sumando esta probabilidad tantas veces como sucesiones existan, que son  $\frac{n!}{x! (n-x)!}$ ; por lo tanto se obtiene la fórmula siguiente, llamada distribución binomial.

$$P(x) = \frac{n!}{x! (n-x)!} p^x q^{n-x} \quad (2)$$

Veremos que utilizando la fórmula (2) es más fácil resolver los problemas anteriores. En el ejemplo del lanzamiento de la moneda, la probabilidad de éxito es  $1/2$  y la probabilidad de fallo es  $1/2$ , o sea  $p = 1/2$ ,  $q = 1/2$  y  $n=2$ ;  $x$  toma los valores  $0,1,2$ ; sustituyendo estos valores en la fórmula (2) obtenemos:

Para  $x=0$

$$P(0) = \frac{2!}{0! (2-0)!} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{2-0} = \frac{2 \cdot 1}{1 \cdot (2 \cdot 1)} \cdot \frac{1}{4} = \frac{1}{4}$$

Para  $x=1$

$$P(1) = \frac{2!}{1! (2-1)!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{2-1} = 2 \cdot \frac{1}{4} = \frac{1}{2}$$

Para  $x=2$

$$P(2) = \frac{2!}{2! (2-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{2-2} = \frac{1}{4}$$

Obsérvese que estas probabilidades son las mismas que obtuvimos antes.

En el segundo ejemplo, del lanzamiento del dado; la probabilidad de éxito es igual a  $1/6$  y la probabilidad de fallo es  $5/6$ , o sea,  $p = 1/6$ ,  $q = 5/6$   $n=3$  y  $x$  toma los valores  $0,1,2,3$ ; sustituyendo estos datos en la fórmula (2) obtenemos las siguientes probabilidades.

Para  $x=0$

$$P(0) = \frac{3!}{0! (3-0)!} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{3-0} = \left(\frac{5}{6}\right)^3$$

Para  $x=1$

$$P(1) = \frac{3!}{1! (3-1)!} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{3-1} = 3 \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$$

Para  $x=2$

$$P(2) = \frac{3!}{2! (3-2)!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{3-2} = 3 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$$

Para  $x=3$

$$P(3) = \frac{3!}{3! (3-3)!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{3-3} = \left(\frac{1}{6}\right)^3$$

Estas probabilidades son las mismas que obtuvimos en el segundo ejemplo.

Propiedades de la Distribución Binomial.

Dado nuestro conjunto de datos, es útil conocer cuál es el promedio.

Supóngase que se tienen  $n$  números y los denotamos por  $x_1, x_2, \dots, x_n$  siendo  $x_1$  el primer número y  $x_n$  el último. El promedio de este conjunto de números que se denota  $\bar{X}$  está dado por la fórmula:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3)$$

A  $\bar{X}$  se le conoce como la media aritmética o simplemente media; una expresión más cómoda para la fórmula anterior, es la siguiente:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (4)$$

donde  $x_i$  denota el  $i$ -ésimo número,  $\Sigma$  (sigma) es el símbolo de suma y junto con los índices  $i=1$  e  $i=n$ , indican que se sumen todos los números  $x$  de 1 hasta  $n$ .

Si los datos estuvieran en forma de frecuencias, es decir, que los números  $x_1, x_2, \dots, x_k$  se presentan  $f_1, f_2, \dots, f_k$  veces respectivamente, o sea con frecuencias  $f_1, f_2, \dots, f_k$  como en la tabla 6.

Tabla 6

$x$	0	1	2	3
$f$	45	23	12	1

La obtención de la media con la fórmula 4 sería muy laboriosa, ya que, por ejemplo, el número 1 tendría que ser sumado 23 veces, debido a ello si  $x_1$  ocurre  $f_1$  veces, entonces se puede multiplicar  $x_1$  por el número de frecuencias  $f_1$ ; si así lo hacemos para cada uno de los números, la media puede ser obtenida sumando los productos  $x_i f_i$ , y dividirlos por el total de números;

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{n} \quad (5)$$

en forma simplificada la fórmula es;

$$\bar{X} = \frac{\sum_{i=1}^k x_i f_i}{n} \quad (6)$$

Como ejemplo aplicaremos la fórmula 6 para la obtención de la media en el conjunto de 81 datos de la tabla 6.

$$\bar{X} = \frac{0 \cdot 45 + 1 \cdot 23 + 2 \cdot 12 + 3 \cdot 1}{81} = \frac{50}{81}$$

Otra medida aritmética que se puede obtener es la variación del conjunto de datos respecto a su media; se desea obtener una medida de variación que indique qué tanto se alejan de su media los datos del conjunto; para esto, se pueden considerar la diferencia de cada uno de los datos con respecto a la media, es decir,

$$x_1 - \bar{X}, x_2 - \bar{X}, \dots, x_n - \bar{X}.$$

Cuando las  $x_i$  son mayores que  $\bar{X}$  se obtienen números positivos, mientras que si las  $x_i$  son menores que  $\bar{X}$  se obtendrían números negativos; pero como sólo se desea obtener distancias positivas desde la media, podemos considerar

Los cuadrados de estas distancias y promediándolos, obtenemos así una medida de la variación de un conjunto de números con respecto a su media, que se denota  $s^2$  y se llama variancia.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} \quad (7)$$

Si los datos están en forma de frecuencias, la fórmula correspondiente es la siguiente:

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{X})^2 f_i}{n} \quad (8)$$

Algunas veces es más útil considerar las fórmulas (7) y (8) teniendo como divisor  $n-1$  en lugar de  $n$ ; cuando  $n$  es grande el valor que se obtiene de  $s$  con las dos fórmulas es muy cercano.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} \quad (9)$$

Para datos clasificados.

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{X})^2 f_i}{n-1} \quad (10)$$

Una forma más fácil para calcular las fórmulas anteriores se obtiene desarrollando el binomio y agrupando los términos, quedando las fórmulas (9) y (10) en la forma siguiente.

$$s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right] \quad (11)$$

Para datos clasificados.

$$s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^k x_i^2 f_i - \frac{\left( \sum_{i=1}^k x_i f_i \right)^2}{n} \right] \quad (12)$$

Obtendremos la variancia del conjunto de datos de la tabla 6 aplicando la fórmula (12).

$$\sum_{i=1}^k x_i^2 f_i = 0^2 \cdot 45 + 1^2 \cdot 23 + 2^2 \cdot 12 + 3^2 \cdot 1 = 80$$

$$\frac{\left( \sum_{i=1}^k x_i f_i \right)^2}{n} = \frac{(0 \cdot 45 + 1 \cdot 23 + 2 \cdot 12 + 3 \cdot 1)^2}{81} = \frac{(50)^2}{81} = \frac{2500}{81}$$

$$s^2 = \frac{1}{81-1} (80 - 30.8) = 0.6141$$

Si sacamos la raíz cuadrada positiva de la variancia se obtiene un valor llamado desviación estándar que se denota por  $s$ .

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (13)$$

y para datos clasificados sería:

$$s = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{n-1}} \quad (14)$$

Ahora consideraremos como se puede calcular la media y la variancia para una distribución teórica.

Las fórmulas (6) y (7) pueden escribirse de la forma siguiente:

$$\bar{x} = \sum_{i=1}^k x_i \frac{f_i}{n} \quad s^2 = \sum_{i=1}^k (x_i - \bar{x})^2 \frac{f_i}{n}$$

La media y la variancia correspondientes a una distribución de probabilidad teórica, las denotaremos con las letras griegas  $\mu$  y  $\sigma^2$  y las obtenemos substituyendo en las fórmulas anteriores, el cociente  $f_i/n$  (llamado frecuencia relativa) correspondiente al valor  $x$  por la probabilidad  $P(x_i)$ , por tanto,

$$\mu = \sum_{i=1}^k x_i P(x_i) \quad \text{y} \quad \sigma^2 = \sum_{i=1}^k (x_i - \mu)^2 P(x_i)$$

para la segunda fórmula desarrollando el binomio se obtiene,

$$\sigma^2 = \sum_{i=1}^k x_i^2 P(x_i) - \mu^2$$

Como ejemplo de una aplicación de las fórmulas anteriores, obtendremos la media y la variancia de la distribución de la figura 1, en donde los valores posibles de  $x$  son  $x_1=0$ ,  $x_2=1$  y  $x_3=2$ , con sus correspondientes probabilidades  $P(x_1)=1/4$ ,  $P(x_2)=1/2$  y  $P(x_3)=1/4$ .

$$\mu = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

$$\sigma^2 = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} - 1 = \frac{1}{2}$$

Es posible realizar cálculos similares con la distribución binomial para obtener la media y la variancia, que pueden emplearse en cualquier problema de tipo binomial. Debido a que los cálculos son complicados, no se harán, únicamente se darán las fórmulas que se obtienen de ellos, que son las siguientes:

$$\mu = np$$

$$\sigma^2 = npq$$

Como ejemplo de su aplicación, las utilizaremos con los datos de la tabla 2, en donde  $n=2$ ,  $p=1/2$  y  $q=1/2$ .

$$\mu = 2 \left( \frac{1}{2} \right) = 1$$

$$\sigma^2 = 2 \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) = \frac{1}{2}$$

## 2. DISTRIBUCION POISSON.

Si en la distribución binomial la probabilidad de éxito  $p$ , en un experimento, es muy pequeña, mientras el número de experimentos  $n$  es muy grande, de tal forma que  $np = \lambda$  permanece fijo, entonces la distribución binomial se aproxima a la distribución Poisson con media  $\lambda$ .

La distribución Poisson tiene la fórmula siguiente:

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

en donde  $x$  toma los valores: 0, 1, 2, . . . etc.

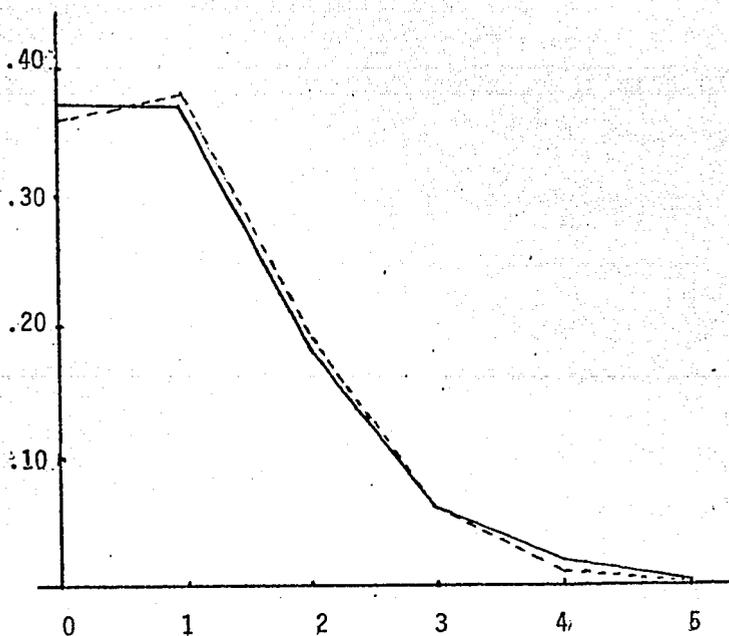
Esta aproximación la podemos observar a través de un ejemplo. En la tabla 5 se considera la distribución binomial que toma los valores de  $n$  igual a 3, 4, 6, 10 y 20; mientras  $p$  toma los valores: 1/3, 1/4, 1/6, 1/10 y 1/20; a su vez se obtiene la distribución poisson con  $np = \lambda = 1$ .

Obsérvese que cuando  $n$  va creciendo y  $p$  va disminuyendo, los valores obtenidos con la distribución binomial se van aproximando más a los valores de la distribución Poisson.

Con los datos de las dos últimas columnas de la tabla 5 se presentan en la figura 3 las gráficas de la distribución Poisson y de la distribución binomial; obsérvese que ésta última se aproxima bastante a la gráfica de la distribución Poisson.

Tabla 5

x	Binomial				Poisson	
	n=3, p=1/3	n=4, p=1/4	n=6, p=1/6	n=10, p=1/10	n=20, p=1/20	$\lambda=np=1$
0	0.296	0.316	0.335	0.34868	0.358485	0.367879
1	0.444	0.422	0.402	0.38742	0.377353	0.367879
2	0.222	0.211	0.201	0.19371	0.188676	0.183940
3	0.037	0.047	0.053	0.05740	0.059582	0.061313
4		0.004	0.008	0.01116	0.013327	0.015328
5			0.001	0.00149	0.002245	0.003066
6			0.000	0.00014	0.000295	0.000511
7				0.00001	0.000031	0.000073
8					0.000003	0.000009
9					0.000000	0.000001
10					0.000000	0.000000

Fig. 3 Gráfica de la distribución Poisson con  $\lambda=1$ , con línea continua.

Gráfica de la distribución binomial con  $n=20$ ,  $p=1/20$ , con línea punteada.

También para la distribución Poisson se puede calcular la variancia y la media de manera similar como se hizo para el ejemplo de tipo binomial; debido a que los cálculos son bastante complicados; únicamente consideraremos las fórmulas que se obtienen a partir de ellos, que son las siguientes:

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

La distribución Poisson es útil cuando se tiene un gran número de objetos que se encuentran distribuidos al azar sobre un área grande; con la Poisson es posible obtener la probabilidad de encontrar  $x$  objetos en una subregión del área.

Definición formal de la distribución Poisson.

Un evento  $S$  que ocurre en el espacio obedece los siguientes postulados:

- a) Independencia: el número de veces que ocurre  $S$  en una región del espacio es independiente del número de ocurrencias de  $S$  en cualquier otra región ajena del espacio.
- b) Proporción: el número promedio de ocurrencias por unidad de espacio es una constante, denotada por  $\lambda$ , y no cambia con el espacio.
- c) Falta de Agrupamiento: la probabilidad de que dos o más ocurrencias se den simultáneamente, puede suponerse que es cero. Bhattacharyya y Johnson (1977).

Veremos como podemos obtener la serie Poisson a partir de la distribución binomial.

Supóngase que se tiene un área  $A$ , la cual contiene  $n$  puntos y es muestreada por un cuadrante de área  $a$ .

Consideremos la proporción  $\frac{A}{a} = r$ , y la densidad de los puntos que se define como la razón del número de puntos entre el área, o sea,  $x = \frac{n}{A}$ .

La probabilidad de encontrar un punto en el cuadrante es:

$$\frac{a}{A} = \frac{1}{r} = p$$

por lo tanto la probabilidad de no encontrarlo es:

$$1 - \frac{1}{r} = q,$$

utilizando la distribución binomial,

$$P(x) = \frac{n!}{x! (n-x)!} p^x q^{n-x}$$

La probabilidad de no encontrar ningún punto es por lo tanto:

$$P(0) = \frac{n!}{0! (n-0)!} \left(\frac{1}{r}\right)^0 \left(1 - \frac{1}{r}\right)^{n-0} = \left(1 - \frac{1}{r}\right)^n$$

Pero como  $n = Ax = \frac{aA}{a}x = axr$ , entonces sustituyendo se obtiene:

$$\left(1 - \frac{1}{r}\right)^n = \left(1 - \frac{1}{r}\right)^{axr}$$

Si  $r$  es grande, entonces se puede demostrar que:

$$\left(1 - \frac{1}{r}\right)^r = \frac{1}{e} = e^{-1}$$

donde  $e = 2.71828$ , es la base de los logaritmos naturales.

Por lo tanto,

$$\left(1 - \frac{1}{r}\right)^{axr} = e^{-ax}$$

es la probabilidad de que el cuadrante no contenga puntos.

La probabilidad de que un cuadrante contenga un punto es:

$$\begin{aligned} P(1) &= \frac{n \cdot (n-1)!}{1! \cdot (n-1)!} \left(\frac{1}{r}\right)^1 \left(1 - \frac{1}{r}\right)^{n-1} \\ &= n \cdot \frac{1}{r} \left(1 - \frac{1}{r}\right)^{n-1} \\ &= n \cdot \frac{1}{r} \cdot \frac{1}{\left(1 - \frac{1}{r}\right)} \left(1 - \frac{1}{r}\right)^n \end{aligned}$$

sustituyendo  $n=axr$ ,

$$= axr \cdot \frac{1}{r} \cdot \frac{1}{\left(1 - \frac{1}{r}\right)} \cdot e^{-ax}$$

Consideremos  $t = \frac{1}{1 - \frac{1}{r}}$  de la expresión anterior. Si  $r$  es muy grande

entonces  $1/r$  se aproxima a 0 y  $1-1/r$  se aproxima a 1; por lo tanto,  $t$  se aproxima a 1 y la expresión quedaría:

$$ax \cdot e^{-ax}$$

que es la probabilidad de que un cuadrante contenga un punto.

Análogamente la probabilidad de encontrar dos puntos en un cuadrante

La obtenemos por la binomial.

$$\begin{aligned}
 P(2) &= \frac{n!}{2! (n-2)!} \left(\frac{1}{r}\right)^2 \left(1 - \frac{1}{r}\right)^{n-2} \\
 &= \frac{n(n-1)(n-2)!}{2! (n-2)!} \left(\frac{1}{r}\right)^2 \left(1 - \frac{1}{r}\right)^{n-2} \\
 &= \frac{n(n-1)}{2!} \left(\frac{1}{r}\right)^2 \frac{1}{\left(1 - \frac{1}{r}\right)^2} \left(1 - \frac{1}{r}\right)^n
 \end{aligned}$$

sustituyendo  $n=axr$

$$\begin{aligned}
 &= \frac{axr(axr-1)}{2!} \left(\frac{1}{r}\right)^2 \frac{1}{\left(1 - \frac{1}{r}\right)^2} \cdot e^{-ax} \\
 &= \frac{ax^2r^2 - axr}{2! \cdot r^2} \cdot \frac{1}{\left(1 - \frac{1}{r}\right)^2} \cdot e^{-ax} \\
 &= \frac{ax(ax-1/r)}{2!} \cdot \frac{1}{\left(1 - \frac{1}{r}\right)^2} \cdot e^{-ax}
 \end{aligned}$$

Si  $r$  crece entonces  $1/r$  tiende a 0 y la expresión queda:

$$\frac{(ax)^2}{2!} \cdot e^{-ax}$$

que es la probabilidad de que un cuadrante contenga dos puntos.

Similarmente la probabilidad de que un cuadrante contenga tres puntos, puede demostrarse que es igual a:

$$\frac{(ax)^3}{3!} \cdot e^{-ax}$$

y en general las probabilidades de que un cuadrante contenga 0, 1, 2, ...,  $n$ , ... etc. puntos están dadas por la serie:

$$e^{-ax}, axe^{-ax}, \frac{(ax)^2}{2!} e^{-ax}, \frac{(ax)^3}{3!} e^{-ax}, \dots, \frac{(ax)^n}{n!} e^{-ax} \dots$$

Esta serie es conocida como la serie Poisson y usualmente es expresada en la forma siguiente:

$$e^{-\lambda}, \lambda e^{-\lambda}, \frac{\lambda^2}{2!} e^{-\lambda}, \dots, \frac{\lambda^n}{n!} e^{-\lambda}$$

donde  $\lambda = ax$  es el número promedio de puntos por cuadrante.

### 3. DISTRIBUCION XI CUADRADA

Algunas veces se quiere comparar si una distribución empírica corresponde a una distribución teórica, veamos un ejemplo.

Considérese el experimento de arrojar un dado; sus seis resultados posibles los representamos por seis celdas o casillas; el experimento se repite 120 veces y las frecuencias observadas las anotamos en las celdas correspondientes. si el dado no está alterado, entonces la probabilidad de que salga cada cara es  $1/6$  y dado que se realiza el experimento 120 veces, se espera que aparezca 20 veces cada cara; el problema es determinar si las frecuencias observadas son compatibles con las frecuencias esperadas, si suponemos que el dado está bien hecho, es decir, es homogéneo.

Los datos se presentan en la tabla 6;  $o_i$ ,  $e_i$ , denotan frecuencias observadas y esperadas respectivamente.

Tabla 6

	1	2	3	4	5	6
$o_i$	16	28	15	10	24	25
$e_i$	20	20	20	20	20	20

Necesitamos una forma de comparar las frecuencias observadas con las frecuencias esperadas para conocer que tanto difieren, para esto se pueden considerar las diferencias de los valores observados con respecto a los valores esperados, véase la tabla 7.

Tabla 7

	$o_i$	$e_i$	$o_i - e_i$
1	16	20	-4
2	28	20	8
3	15	20	-5
4	10	20	-10
5	24	20	4
6	25	20	5

De la tabla se observa que en algunos casos se tienen diferencias negativas, pero sólo nos interesan los valores positivos, ya que si se dejaran así podrían cancelarse; por lo tanto, se considera el cuadrado de la diferencia. Para evitar problemas con las unidades en que se realizan las medidas de las frecuencias, se divide el cuadrado de las diferencias por la frecuencia esperada, quedando cocientes de la forma:

$$\frac{(o_i - e_i)^2}{e_i}$$

Estos cocientes los sumamos para obtener la desviación total y conocer de esta manera la medida de la discrepancia entre las frecuencias esperadas y las observadas. A esta medida se le llama  $\chi^2$  cuadrada ( $\chi^2$ ) y queda definida por la fórmula:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k} = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

En donde  $o_i$  y  $e_i$  denotan las frecuencias observadas y esperadas respectivamente,  $k$  denota el número de celdas,  $N$  la frecuencia total y se tiene que:  $\sum e_i = \sum o_i = N$

Ahora calcularemos el valor de  $\chi^2$  para los datos de la tabla 6.

$$\chi^2 = \frac{(16-20)^2}{20} + \frac{(28-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(25-20)^2}{20} = 12.3$$

Podemos observar que si el valor de  $\chi^2$  es cero, entonces existe una concordancia exacta entre los valores observados y los esperados.

Si  $\chi^2$  es mayor que cero, entonces las frecuencias observadas y esperadas no coinciden, y entre mayor sea el valor de  $\chi^2$  más grandes son las discrepancias entre las frecuencias.

Así  $\chi^2$  da una medida de la correspondencia entre las frecuencias observadas y las esperadas.

Ya que se tiene la medida  $\chi^2$  es necesario conocer que tan significativa es; para ello supongamos que el experimento anterior de lanzar 120 veces un dado, se repitiera muchas veces con un dado homogéneo; y en cada una de ellas se calculara el valor de  $\chi^2$ ; obtendríamos un conjunto de valores de  $\chi^2$  que podría clasificarse en una tabla de frecuencia relativa y una gráfica de  $\chi^2$ .

Entonces podríamos comparar si el valor que se obtuvo de  $\chi^2 = 12.3$  es muy grande con respecto al valor usual de los valores  $\chi^2$  que se obtienen en los experimentos con un dado homogéneo.

Si el porcentaje de experimentos para los cuales  $\chi^2$  mayor que 12.3 es muy pequeño por ejemplo, 5% se podría concluir que las frecuencias observadas no concuerdan con las frecuencias esperadas; lo que implica que

nuestra suposición no es correcta; y concluiríamos que el dado no es homogéneo.

Nos interesa saber como es la distribución teórica de  $\chi^2$ , a la cual se le conoce con el nombre de distribución  $\chi^2$ , que es el mismo nombre que se le dá a  $\chi^2$ , esto crea algo de confusión, por lo cual nótese que se diferencian en su escritura.

La gráfica de la distribución xi cuadrada tiene la forma siguiente:

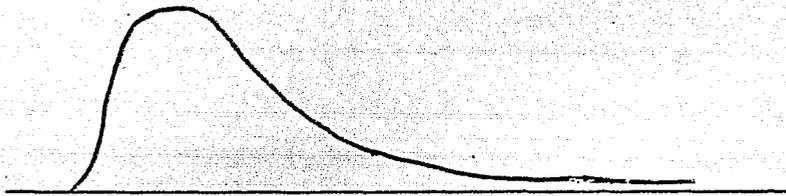


Fig. 4 Gráfica de la distribución xi cuadrada.

Una característica de esta distribución es que su forma depende solamente del número de celdas; en la figura 5 se muestran las gráficas de seis curvas correspondientes al número de celdas variable de 2 a 7.

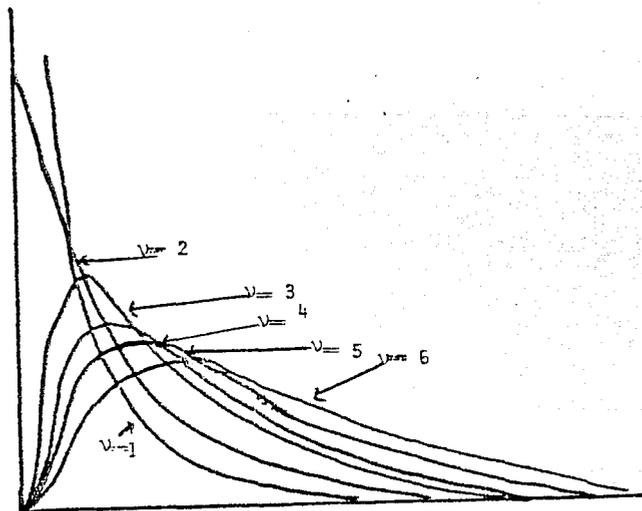


Fig. 5 Distribución xi cuadrada para varios grados de libertad.

Obsérvese que en el ejemplo la suma de las seis frecuencias observadas en la tabla 6 es igual a 120, la sexta frecuencia, se puede obtener a partir de las cinco primeras frecuencias. A el número de frecuencias independientes se le llama número de grados de libertad, se denota por  $\nu = k-1$ , donde  $k$  es el número de celdas; por medio de este parámetro se identifica la distribución  $\chi^2$  cuadrado.

En nuestro ejemplo existen cinco celdas independientes, así que se tienen  $\nu = 6-1 = 5$  grados de libertad.

La distribución  $\chi^2$  cuadrada cuenta con una tabla de probabilidades con grados de libertad; por ejemplo, el valor de  $\chi^2$  cuadrada que se obtiene de la tabla con el 5% y 5 grados de libertad; se busca en la columna .05 y en la línea que marca  $\nu = 5$ , el valor que se obtiene es  $\chi^2 = 11.1$ .

El porcentaje que se escoge como en este caso el 5%, se denomina nivel de significancia; es posible considerar cualquier valor pero, en general se seleccionan los niveles de significancia del 1% y 5%. Si se selecciona el 5% indica que hay aproximadamente 5 ocasiones en 100 que se rechazaría la hipótesis cuando debería ser aceptada, o sea se está con un 95% de confianza de que se toma la decisión adecuada. En tal caso se dice que la hipótesis ha sido rechazada a nivel de significancia de 0.05, lo que significa que se puede cometer error con una probabilidad de 0.05. El nivel de significancia se denota con  $\alpha$ .

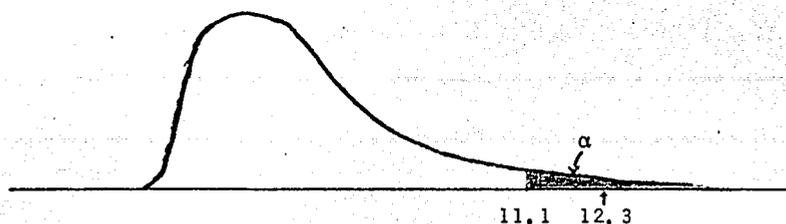


Fig. 6 Gráfica de la distribución  $\chi^2$  cuadrada para el problema del dado.

En la figura 6 se indica el valor que obtuvimos de  $\chi^2 = 12.3$  y el valor  $\chi^2 = 11.1$ , que corta la cola de la derecha, dejando de ese lado el 5% del área de la distribución.

Como los valores grandes de  $\chi^2$  corresponden a una concordancia mala, el conjunto de los valores de  $\chi^2$  mayores que 11.1 se elige como región crítica de la prueba. El valor que obtuvimos  $\chi^2 = 12.3$  se encuentra en la región crítica, por lo tanto la hipótesis de que el dado es homogéneo se rechaza.

Debido a que  $\chi^2$  es únicamente una aproximación a una distribución  $\chi^2$  cuadrada; la prueba  $\chi^2$  debe utilizarse solamente cuando esta aproximación es buena. La experiencia y la teoría indican que la aproximación es generalmente satisfactoria siempre que las frecuencias esperadas en todas las celdas tengan por lo menos un valor igual a 5.

Si la frecuencia esperada de una celda es menor que 5, esta frecuencia se puede sumar con una o más frecuencias de las otras celdas hasta obtener la condición necesaria.

#### 4. DISTRIBUCION NORMAL.

Una distribución muy usual para los casos en donde los datos consisten de medidas, presenta una gráfica parecida a la de la figura 7, en donde se muestra la distribución de frecuencia de la altura de 8585 hombres, las medidas están en pulgadas.

Este tipo de gráfica es común de muchas distribuciones empíricas; obsérvese que se encuentra un número mayor de datos en el centro de la gráfica que en los extremos, presenta una forma muy similar a la de una campana.

Si se tomara una muestra mayor y los intervalos se consideraran más pequeños, se llegaría a un tipo de gráfica simétrica; que sería la gráfica de

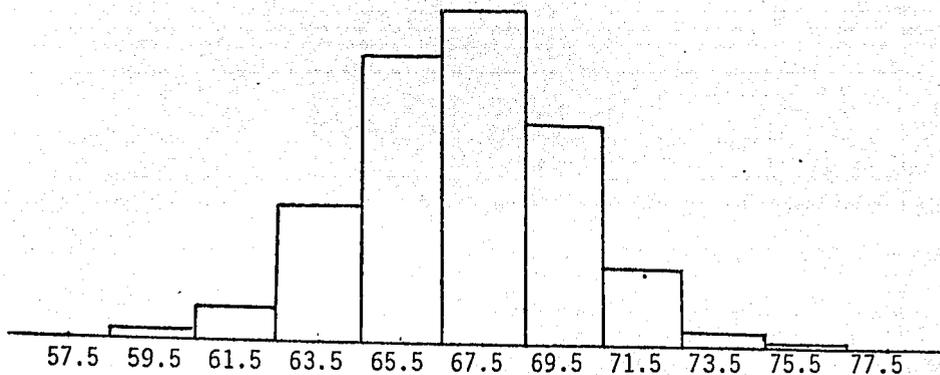


Fig. 7 Gráfica de la distribución de frecuencia de la altura de 8585 hombres.

una distribución teórica que serviría como modelo a las distribuciones empíricas. Esta distribución es la distribución Normal y es la distribución teórica más importante en estadística, aparece con mucha frecuencia, además muchas de las técnicas utilizadas en estadística aplicada se basan en la - distribución Normal. Su gráfica es una curva como se muestra en la figura 8. El área total bajo la curva es igual a uno.

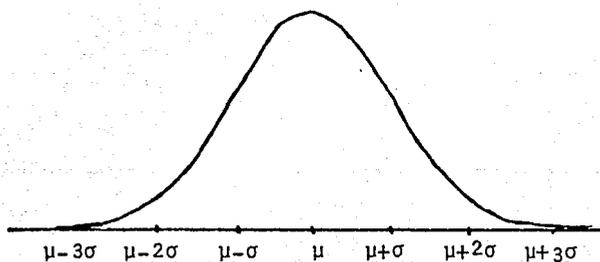


Fig. 8 Curva Normal.

La media  $\mu$  se encuentra en el punto de simetría sobre el eje  $x$ ; este eje se ha marcado en unidades de  $\sigma$  a partir de  $\mu$ . El área es muy pequeña - bajo la curva más allá de los valores  $3\sigma$  y  $-3\sigma$ .

Se caracteriza esta distribución porque con los valores  $\mu$  y  $\sigma$ , es posible determinar la localización y la forma de la curva normal, el valor  $\mu$  señala el centro de la curva, en tanto que el valor de  $\sigma$  determina la dispersión de sus valores, de tal forma que si se tienen dos gráficas con la misma media pero, en la segunda gráfica se aumenta el valor de  $\sigma$ , la curva se reduce en altura, como se muestra en las figuras 9 y 10.

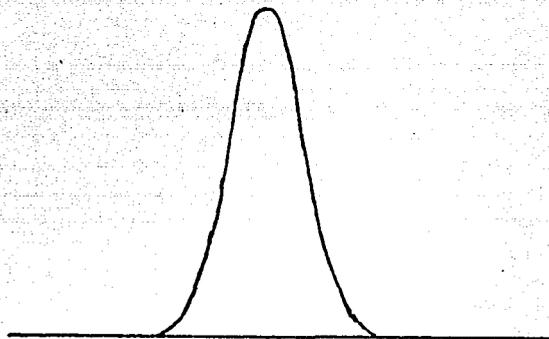


Fig. 9 Distribución normal con  $\mu = 0$  y  $\sigma = 1$ .



Fig. 10 Distribución normal con  $\mu = 0$  y  $\sigma = 3$ .

Debido a que la forma de una curva normal queda determinada por su desviación estandar, es posible reducir cualquier curva normal a una curva patrón o estandar, por medio de un cambio de variable; y dado que la curva normal más sencilla es la que tiene media 0 y desviación estandar igual a 1, se

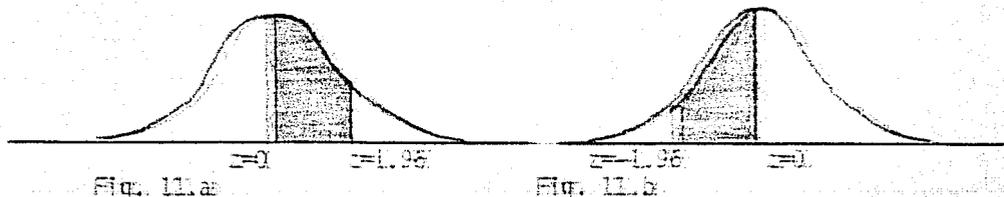
utiliza la tabla de distribución de esta curva normal, para encontrar el área bajo una parte cualquiera de la curva.

Como el área limitada por la curva normal es igual a uno entonces, el área bajo la curva entre dos puntos  $x=a$  y  $x=b$ , donde  $a$  es menor que  $b$ , representa la probabilidad de que  $x$  se encuentre entre  $a$  y  $b$  y es denotada por  $P(a < x < b)$ .

Veamos unos ejemplos, supóngase que se quiere encontrar el área bajo la curva normal en los siguientes intervalos.

a) entre  $z=0$  y  $z=1.96$ .

Utilizando la tabla de la distribución normal, se busca en la columna encabezada por  $z$  hacia abajo hasta el valor 1.9, entonces por esa renglón hacia la derecha hasta la columna encabezada por 6. El valor que se encuentra es 0.4750, éste es el área pedida y representa la probabilidad de que  $z$  esté comprendida entre 0 y 1.96, denotada  $P(0 < z < 1.96)$ . Obsérvese la figura 11.a



b) entre  $z=-1.96$  y  $z=0$ .

Por simetría el área pedida es igual a el área entre  $z=0$  y  $z=1.96$ , el valor 0.4750 representa la probabilidad de que  $z$  esté entre  $-1.96$  y  $0$ , denotado por  $P(-1.96 < z < 0)$ . Véase figura 11.b

c) a la izquierda de  $z=-1.96$ .

El área pedida la podemos obtener restando el área a la izquierda de  $z=0$  el área entre  $z=-1.96$  y  $z=0$ . El área a la izquierda de  $z=0$  es igual a 0.50 porque es exactamente la mitad del área total que vale 1, por lo tanto tenemos:

que  $0.5 - 0.4750 = 0.025$ . Esta es el área a la izquierda de  $-1.96$ . Observe figura 11.c.

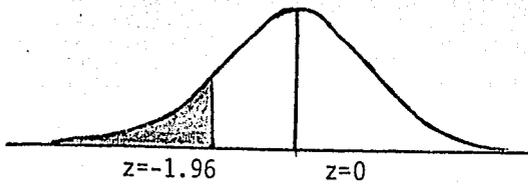


Fig. 11.c

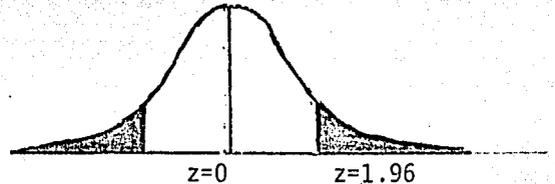


Fig. 11.d

d) área a la derecha de  $z=1.96$  y a la izquierda de  $z=-1.96$ . Véase figura 11.d.

El área pedida se obtiene restando a el área total dos veces el área entre  $z=0$  y  $z=1.96$ . Queda  $1-2(0.4750) = 0.05$ . Los valores extremos del intervalo se conocen con el nombre de puntos críticos.

Ahora consideraremos el caso en el cual se conoce el área y se quiere determinar los puntos críticos.

a) Determinar el valor de  $z$  cuando el área entre 0 y  $z$  es 0.4750.

En la tabla de la normal el valor 0.4750 se localiza a la derecha del renglón que empieza por 1.9 y bajo la columna encabezada por 6. Entonces la  $z$  pedida es igual a 1.96.

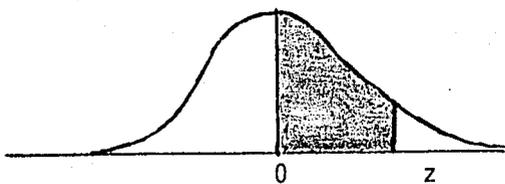


Fig. 12.a

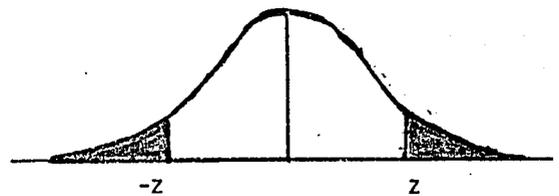


Fig. 12.b

b) Determinar el valor  $z$  cuando el área en los extremos de la figura 12.b es igual a 0,025

Para obtener el valor deseado observamos que cada extremo en la figura tiene un área igual a 0.025. Entonces si a la mitad del área le restamos este

valor obtenemos el área entre  $z=0$  y el valor desconocido de  $z$ , en la tabla de la normal buscamos este valor.  $0.5-0.025=0.475$ , y se encuentra que  $z=1.96$  por lo tanto  $-z=-1.96$ . Estos son los puntos críticos pedidos.

Usualmente la tabla de la distribución  $\chi^2$  cuadrada incluye valores únicamente hasta 30 grados de libertad, para valores mayores se puede utilizar la aproximación de Fisher  $\sqrt{2X^2} - \sqrt{2V-1}$  que se distribuye como una Normal con media cero y varianza uno.

Si consideramos,

$$T = \sqrt{2X^2} - \sqrt{2V-1}$$

despejando  $X^2$  se tiene:

$$X^2 = \frac{1}{2} (T + \sqrt{2V-1})^2$$

a) Por ejemplo si queremos hallar los valores críticos de  $X^2$  tales que el área de cada extremo de la figura 13.b, sea 0.025 y con 80 grados de libertad.

Tenemos que:

$T = 1.96$  para una prueba de dos colas con un nivel de significancia del 0.05 (véase ejemplo anterior).

Por lo tanto

$$X_1^2 = \frac{1}{2} (-1.96 + \sqrt{2(80)-1})^2 = 56.706135$$

$$x_2^2 = \frac{1}{2} (1.96 + \sqrt{2(80)-1})^2 = 106.13545$$

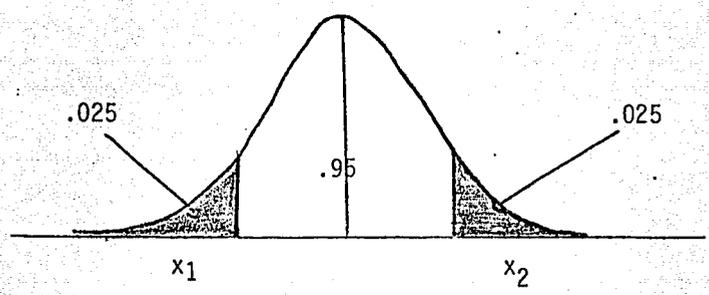
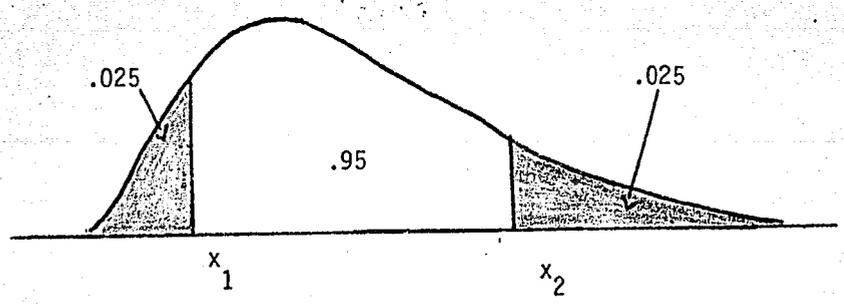


Fig. 13.a

b) Hallar los valores críticos tales que el área de cada extremo sea 0.025 y con 100 grados de libertad. véase figura 13.b.



En este caso tendríamos

$$x_1^2 = \frac{1}{2} (-1.96 + \sqrt{2(100)-1})^2 = 73.77152$$

$$x_2^2 = \frac{1}{2} (1.96 + \sqrt{2(100)-1})^2 = 129.0699$$

La distribución Normal es muy usual en antropología especialmente en antropología física, en donde grupos de datos como estaturas, pesos y medidas de algunos músculos se aproximan bastante a un tipo de distribución normal. Consideremos un ejemplo tomado de datos de la antropóloga física Ma. Eugenia Villanueva. Se tiene un conjunto de los pesos en kilogramos de un grupo de 50 niños cuya edad va de 6 años 8 meses a 7 años 8 meses.

Los datos se muestran en la tabla siguiente:

Tábla

Pesos (Kgs.)	frecuencia
16 - 19	9
20 - 23	26
24 - 27	10
28 - 31	5

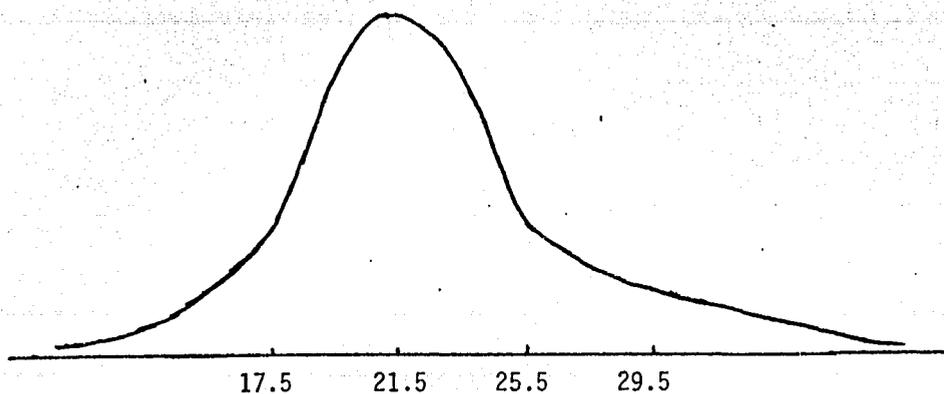


Fig. 14 Distribución de pesos.

Es importante señalar que cuando se utilicen las distribuciones de probabilidad es necesario observar si se cumplen las condiciones necesarias en cada caso, para tener confianza sobre los resultados que se obtengan.

En los capítulos siguientes haremos uso de estas distribuciones para el análisis de los patrones espaciales.

Al final del trabajo se dan las tablas correspondientes a cada una de las distribuciones que se utilizarán para las pruebas de significancia.

## CAPITULO 2

### ANALISIS DE PATRONES DE PUNTOS.

En este capítulo trataremos con el análisis de patrones de puntos distribuidos sobre superficies planas, aunque es también aplicable a patrones de puntos distribuidos a lo largo de una línea o distribuidos a través de un volumen; por propósitos prácticos se explicará en un espacio bidimensional.

En un contexto arqueológico estos puntos pueden ser: artefactos de un sitio, artefactos dispersos sobre un área o sitios arqueológicos.

Clark y Evans señalan que "el patrón de distribución de una población de puntos es una característica fundamental de ella, pero es un rasgo distintivo que es extremadamente difícil de describir en términos precisos y significativos". (Clark y Evans 1952, 445).

Las distribuciones exhibidas por poblaciones de objetos presentan una variedad infinita de patrones; por ejemplo, tres tipos de distribuciones se presentan en la figura 1, entre ellas se encuentran una infinidad de distribuciones posibles.

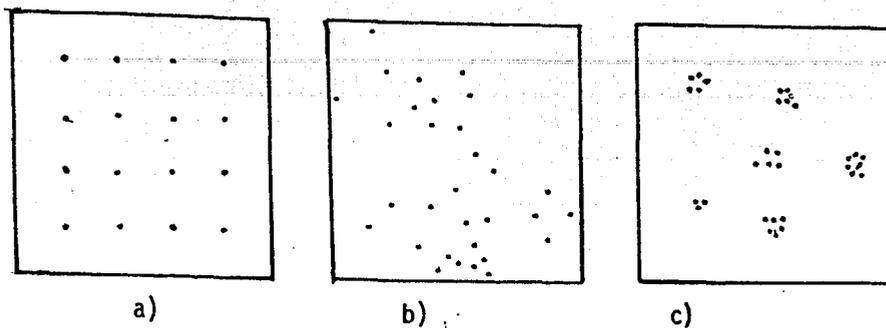


Fig. 1 a) Puntos regularmente espaciados.  
 b) Puntos distribuidos al azar.  
 c) Puntos distribuidos en grupos.

Existen problemas para encontrar métodos descriptivos adecuados para el estudio del patrón de distribución de una población, en general se utilizan muestras de la población como fuente de información; ello introduce cierta inexactitud y desviación en los datos.

Los métodos para el análisis de patrones de puntos han sido desarrollados principalmente en ecología y geografía, en estos campos el estudio de patrones espaciales parte de la suposición de aleatoriedad. Cuando no se tiene forma de predecir un resultado se le denomina aleatorio o al azar.

Clark y Evans definen "en una distribución aleatoria de un conjunto de puntos sobre un área dada se supone que todo punto tiene la misma probabilidad de ocurrencia sobre cualquier subárea, además que cualquier subárea de tamaño específico tiene la misma probabilidad de recibir un punto que cualquier otra subárea del mismo tamaño y que el lugar de cada punto no ha sido influenciado por ningún otro". (Clark y Evans 1952, 446).

La distribución al azar puede utilizarse como una norma y ver la manera

y grado en el cual una distribución empírica se desvía de ella.

Consideremos la aplicación práctica de la distribución Poisson a distribuciones espaciales de puntos, la probabilidad de que un área de tamaño especificado contenga exactamente  $x$  puntos es dada por esta distribución.

Así para una distribución aleatoria de puntos la probabilidad de que un área de tamaño específico seleccionada al azar contenga  $x$  puntos esta dada por la distribución Poisson, donde  $\lambda$  es el número promedio de puntos por el área especificada.

Este proceso teórico podemos utilizarlo como la norma contra la cual un patrón particular puede ser analizado.

La experiencia de campo muestra que la mayoría de las poblaciones estudiadas no tienen un patrón aleatorio, entonces la desviación de un patrón aleatorio puede servir como base de comparación; aunque es difícil medir la desviación.

Los métodos que han sido desarrollados para demostrar la ocurrencia de distribuciones no aleatorias, pueden dividirse en métodos de cuadrantes y métodos de distancias.

#### 1. METODOS DE CUADRANTES.

Para los métodos de cuadrantes el área de estudio es dividida por una malla uniforme de cuadrados o rectángulos, aunque es posible utilizar otras figuras como triángulos o exágonos. Se registra el número de observaciones en cada cuadrante de la malla haciendo una tabla de frecuencias que muestre el número observado de cuadrantes que contienen 0, 1, 2, . . . , etc. puntos.

Si el patrón es aleatorio el número de cuadrantes conteniendo 0, 1, 2, .., etc. puntos es determinado por la distribución Poisson. Basandose en esto se han ideado pruebas de aleatoriedad que veremos en la sección siguiente.

## 2. PRUEBA DE BONDAD DE AJUSTE XI CUADRADA.

Quando se tiene alguna indicación por medios probabilísticos o por algún otro medio sobre la distribución de una población es posible ajustarla a una distribución teórica. El método generalmente consiste en emplear la media de la población o de la muestra y comparar las frecuencias observadas con las frecuencias esperadas por medio de una prueba de bondad de ajuste xi cuadrada.

Las frecuencias observadas las obtenemos de los datos de la población y las frecuencias esperadas se obtienen a partir de suponer una cierta hipótesis  $H$ . Recordemos el ejemplo de la sección 3 del capítulo 1, que consistía en el lanzamiento de un dado; el problema era saber si las frecuencias observadas concordaban con las frecuencias esperadas si suponíamos la hipótesis de que el dado era homogéneo. Se aplicó la prueba xi cuadrada para probar la bondad de ajuste. Ahora aplicaremos esta prueba en el estudio de un patrón espacial sintético.

### 2.1. Aplicación de la prueba xi cuadrada.

Supóngase que se tiene una población de 50 puntos con un patrón de distribución como se muestra en la figura 2; lo dividimos con una malla de cuadrados y contamos el número de cuadrantes que contienen 0, 1, 2, . . ., etc. puntos, obteniendo así las frecuencias observadas; por ejemplo, se encuentran 40 cuadrantes vacíos, 36 cuadrantes con un punto, etc. las otras frecuencias se muestran en la tabla 1; con estos datos calcularemos el promedio de puntos por cuadrante.

Como hipótesis suponemos que el patrón de distribución es aleatorio (aunque no existe una razón lógica para hacer esta suposición sobre cualquier otra).

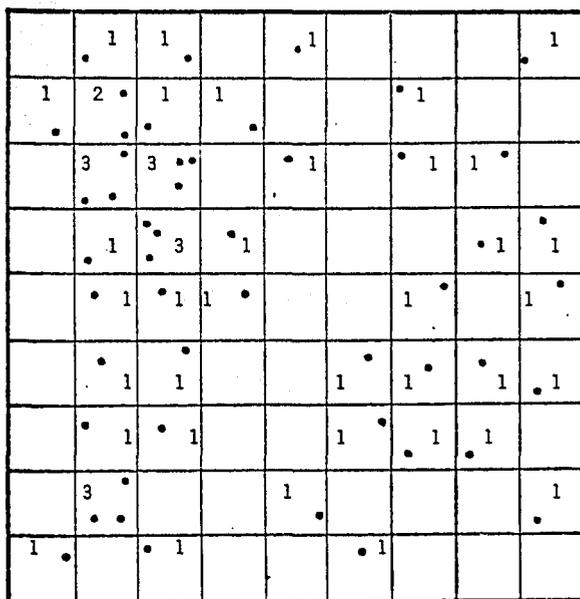


Fig. 2 Patrón de distribución sintético.

Tabla 1

No. de puntos por cuadrante	Frecuencia	
	$f_j$	$x_j f_j$
0	40	0
1	36	36
2	1	2
3	4	12
	$n = 81$	$\Sigma x_j f_j = 50$

Por medio de la distribución Poisson, donde  $\lambda$  es el número promedio de puntos por cuadrante, podemos obtener las frecuencias esperadas es decir, - el número de cuadrantes esperados que contienen 0, 1, 2, . . . , etc. puntos si el patrón es aleatorio.

El valor de la media que se obtiene es:

$$\bar{X} = \frac{\sum x_i f_i}{n} = \frac{50}{81} = 0.6173$$

Utilizando la distribución Poisson 
$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

con la media obtenida  $\lambda = 0.6173$ , calculamos el primer término de la serie - Poisson  $e^{-\lambda}$ , se obtiene de la tabla de valores de  $e^{-\lambda}$  que  $e^{-0.6173} = 0.5394$  que es la probabilidad de que no halla puntos en un cuadrante, por lo tanto, el número esperado de cuadrantes vacíos lo obtenemos multiplicando esta probabilidad por el número total de cuadrantes, es decir,  $81(0.5394) = 43.692$ , - las frecuencias esperadas restantes son calculadas como se muestra en la tabla 2.

Para la prueba xi cuadrada es necesario que las frecuencias esperadas - tengan al menos un valor igual a 5, por lo que, para este ejemplo agruparemos las frecuencias esperadas con dos o más puntos, quedando como se muestra en la tabla 3.

Para calcular el valor  $\chi^2$  se aplicó la fórmula 
$$\chi^2 = \frac{\sum_{j=1}^k (o_j - e_j)^2}{e_j}$$

$$\chi^2 = \frac{(40-43.692)^2}{43.692} + \frac{(36-26.9706)^2}{26.9706} + \frac{(5-10.3374)^2}{10.3374} =$$

$$0.3120 + 3.0229 + 2.7558 = 6.0907.$$

Tabla 2

No. de puntos por cuadrante	Número de cuadrantes	
	esperados	observados
0	$81e^{-\lambda} =$ $81(0.5394) = 43.692$	40
1	$81\lambda e^{-\lambda} =$ $(43.692)(0.617) = 26.9706$	36
2	$\frac{81\lambda^2 e^{-\lambda}}{2!} =$ $(26.971)\frac{(0.617)}{2} = 8.3245$	1
3	$\frac{81\lambda^3 e^{-\lambda}}{3!} =$ $(8.325)\frac{(0.617)}{6} = 0.856$	4
4	$\frac{81\lambda^4 e^{-\lambda}}{4!} =$ $(0.856)\frac{(0.617)}{24} = 0.022$	0

Tabla 3

No. de puntos por cuadrante	Número de cuadrantes			
	$e_i$	$o_i$	$o_i - e_i$	$\chi^2$
0	43.6920	40	-3.6920	0.3120
1	26.9706	36	9.0294	3.0229
2	8.3245	5	-5.3374	2.7558
> 2	2.0129			
	81	81		6.0907

Para saber si el valor de  $\chi^2$  es significativo recurriremos a la tabla de la distribución  $\chi^2$  cuadrada, pero antes calcularemos los grados de libertad. En la tabla 3 existen dos celdas independientes pero, cuando se utiliza un parámetro de la población, como en nuestro caso la media  $\lambda$  para calcular las frecuencias esperadas se le resta al número de celdas independientes y así obtenemos los grados de libertad  $\nu = n - 1 - k$ , donde  $k$  es el número de parámetros de la población utilizados. Para el ejemplo  $\nu = 3 - 1 - 1 = 1$  y con un nivel de significancia de 0.05 buscamos en la tabla de la distribución  $\chi^2$  cuadrada, obteniendo el valor  $\chi^2 = 3.84$ .

Como el valor que obtuvimos  $\chi^2 = 6.09$  es mayor que  $\chi^2 = 3.84$ , se encuentra en la región crítica, por lo tanto, rechazamos la hipótesis de que el patrón es aleatorio y concluimos que el patrón de distribución de la figura 2 no es aleatorio.

## 2.2. Limitaciones de la prueba $\chi^2$ cuadrada.

Cuando se han obtenido las frecuencias esperadas, antes de aplicar la prueba  $\chi^2$  cuadrada, debemos verificar si cada uno de los valores esperados tiene como valor mínimo 5, si esta condición no se cumple es necesario combinar los valores esperados menores que 5, de tal forma que cada celda tenga un valor esperado mayor o igual que 5.

Otra limitación se presenta cuando el valor de la media es bajo y se tiene que agrupar a los cuadrantes en ocupados o vacíos. Consideremos un ejemplo donde la media  $\lambda = 0.1543$  y con las frecuencias que se muestran en la tabla 4.

Tabla 4

No. de puntos por cuadrante	frecuencia	
	$e_1$	$o_1$
0	277.67	278
1	42.84	42
2	3.30	4
> 2	0.19	0

Obsérvese que en la tercera fila se tiene un valor inferior a 5, por lo que, es necesario agrupar las frecuencias como se muestra en la tabla 5.

Quedando únicamente el número de cuadrantes vacíos y ocupados.

Tabla 5

No. de puntos por cuadrante	frecuencia	
	$e_1$	$o_1$
0	277.67	278
1	46.33	46

Si calculamos  $v = n - 1 - k = 2 - 1 - 1 = 0$ , vemos que no existen grados de libertad para aplicar la prueba  $\chi^2$  cuadrada.

Como en este caso no es posible utilizar esta prueba entonces es necesario acudir a otras pruebas de aleatoriedad.

### 3. PRUEBA DE LA RAZON VARIANZA/MEDIA.

Esta razón es también conocida como coeficiente de dispersión o varianza relativa.

Recuérdese que en la distribución Poisson la media y la varianza son iguales, esto sugiere utilizar la razón varianza/media ( $s^2/\bar{X}$ ) como una medida de desviación de una distribución aleatoria. Si el valor de la razón es mayor que uno se indica una distribución agrupada y si el valor de la razón es menor que uno se indica una distribución regular.

Cuando el patrón de distribución es aleatorio se espera que el valor de la razón ( $s^2/\bar{X}$ ) este muy cercano a uno.

El valor de la razón  $s^2/\bar{X}$  es:

$$\frac{1}{(n-1)\bar{X}} \sum_{i=1}^n (x_i - \bar{X})^2$$

Si obtenemos un valor de la razón  $s^2/\bar{X}$  que varía de uno muy poco, nos interesa conocer si esta variación es significativa o el valor es producido por un patrón aleatorio.

La significancia de la diferencia entre el valor de la razón  $s^2/\bar{X}$  y la unidad se puede probar por medio del índice de dispersión, debido al hecho que el índice

$$I = \frac{s^2}{\bar{X}} (n-1) = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\bar{X}}$$

es la suma de  $n$  términos de la forma  $(o-e)^2/e$ , donde  $o$  y  $e$  son las frecuencias observadas y esperadas respectivamente.

Esta suma se distribuye aproximadamente como una  $\chi^2$  con  $n-1$  grados de libertad, donde  $n$  es el número de cuadrantes.

Tomando como hipótesis que el patrón es aleatorio, la probabilidad de obtener cualquier valor del índice de dispersión puede entonces ser obtenida consultando la tabla de la distribución  $\chi^2$  cuadrada.

Pielou señala "sin embargo es frecuentemente irracional postular que un patrón es aleatorio, pueden por lo menos no existir campos particulares para favorecer la hipótesis de aleatoriedad sobre cualquier otro patrón imaginable. La razón  $s^2/\bar{X}$  no debe entonces ser vista como una prueba de criterio sino meramente como un estadístico descriptivo de un patrón de población". (Pielou 1969, 91).

Además considera que si el valor de  $s^2/\bar{X}$  es muy cercano a uno, no se debe inmediatamente concluir que el patrón es aleatorio en el sentido que los puntos son independientes y el número esperado por cuadrante es el mismo para todos los cuadrantes. Esta conclusión está justificada únicamente si existen razones a priori para apoyarla como verdadera y si la aplicación de la prueba no dá razones para rechazarla.

### 3.1. Aplicación de la prueba varianza/media.

Ahora aplicaremos la prueba de la razón  $s^2/\bar{X}$  al patrón de distribución de la figura 2, partiendo de la hipótesis de que es aleatorio.

La media que se obtuvo fué  $\bar{X} = 0.6173$ .

Calculamos la varianza con la fórmula

$$s^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^k f_i x_i^2 - n\bar{X}^2}{n-1}$$

porque los datos se encuentran en forma de frecuencias, realizamos los cálculos como se muestra en la tabla 6.

Tabla 6

No de puntos por cuadrante	No. de cuadrantes observados			
$x_i$	$f_i$	$x_i^2$	$f_i x_i^2$	$f_i x_i$
0	40	0	0	0
1	36	1	36	36
2	1	4	4	2
3	4	9	36	12
	$n = 81$		$\sum f_i x_i^2 = 76$	$\sum f_i x_i = 50$

sustituyendo en la fórmula de la varianza obtenemos:

$$\sum_{i=1}^k f_i x_i^2 - n\bar{x}^2 = 76 - 81(0.6173)^2 = 45.1359$$

$$s^2 = \frac{45.1359}{80} = 0.5642$$

Con la media y la varianza obtenemos el siguiente valor de la razón:

$$\frac{s^2}{\bar{X}} = \frac{0.5642}{0.6173} = 0.9140$$

Para probar si el valor de esta razón varía significativamente de la unidad, calculamos el índice de dispersión.

$$I = \frac{s^2}{\bar{x}} (n-1) = 0.9140(80) = 73.12$$

El índice se distribuye como una  $\chi^2$  con 80 grados de libertad, con un nivel de significancia de 0.05 para dos colas se obtienen los valores críticos  $x_1 = 56.7061$  y  $x_2 = 106.1354$ . Obsérvese que el valor del índice no se encuentra en la región crítica, por lo tanto, se concluye que el patrón es aleatorio.

La aproximación de Fisher obtenida es:

$$\sqrt{(2x^2)} - \sqrt{(2v-1)} = \sqrt{2(73.12)} - \sqrt{2(80)-1} = \sqrt{146.24} - \sqrt{159} = -0.5166.$$

de la tabla de la normal la probabilidad correspondiente es 0.6844. Por lo tanto se concluye el mismo resultado.

Greig-Smith indica que "si la desviación de la distribución aleatoria tiende a una distribución regular, el valor del índice de dispersión se juzga significativo si la probabilidad de la tabla de la distribución  $\chi^2$  cuadrada es inusitadamente grande; es decir, la probabilidad del 95% corresponde al 5% para una distribución agrupada y el 99% corresponde al 1%. Si se sigue el proceso para un número grande de grados de libertad, se utiliza el valor de  $\sqrt{2x^2} - \sqrt{2v-1}$  y se requiere una probabilidad baja para la significancia". (Greig-Smith 1964, 63).

### 3.2. Limitaciones de la prueba varianza/media.

Puede ocurrir que la varianza y la media sean iguales y el patrón no sea aleatorio. La razón  $s^2/\bar{x}$  deriva de un aspecto particular de la desviación de una distribución Poisson, la ocurrencia anormal de una varianza alta o baja. - Considérese el siguiente caso hipotético de Evans citado por Greig-Smith(1964, 69):

Tabla 7

No. de puntos por cuadrante	frecuencia	
	o	e
0	20	37.16
1	76	37.16
2	0	18.58
3	0	6.19
4	0	1.55
5	5	0.31
5	0	0.05

A pesar de la evidencia de no aleatoriedad, la varianza y la media son iguales ( $s^2=1$  y  $\bar{x}=1$ ). De la aplicación de la prueba  $\chi^2$  cuadrada se obtiene un valor de 63.24 con dos grados de libertad y una probabilidad mucho menor que 0.1 %.

Greig-Smith señala que "una combinación de este tipo de distribución -- regular en general con grupos ocasionales de puntos no es muy común en una vegetación natural y podría en cualquier caso ser reconocida sin la prueba. -- El tipo más común de no-aleatoriedad es la ocurrencia de un número excesivo de cuadrantes vacíos y pocos con gran número de puntos". (Greig-Smith 1964, 70).

Los resultados de la prueba  $s^2/\bar{x}$  deben ser interpretados con precaución cuando la media es muy baja, que es el mismo cuidado que se tiene que tener con la prueba  $\chi^2$  cuadrada. El ejemplo enfatiza que en algunos casos -- una de las dos pruebas la  $\chi^2$  cuadrada o la  $s^2/\bar{x}$  da evidencia de no aleatoriedad, cuando la otra falla.

## 4. PROBLEMAS DE LOS METODOS DE CUADRANTES.

El principal problema con el uso de cuadrantes es que los resultados dependen del tamaño del cuadrante usado; esto lo podemos observar a través de los siguientes ejemplos.

Se aplicará la prueba  $\chi^2$  al patrón de distribución de la figura 2, pero ahora con una malla de cuadrantes de mayor tamaño como se muestra en la figura 3. La hipótesis que suponemos es que el patrón de distribución es aleatorio.

Las frecuencias observadas y la media se muestran en la tabla 8.

Tabla 8

No. de puntos por cuadrante	0	1	2	3	4	5	6	7
No. de cuadrantes	8	13	13	0	1	0	0	1
No. promedio por cuadrante	1.3889							

Las frecuencias esperadas y el valor de  $\chi^2$  se muestran en la tabla 9.

Tabla 9

No. de puntos por cuadrante	frecuencia		$\chi^2$
	e	o	
0	8.9767	8	0.1062
1	12.4676	13	0.0227
2	8.6564	14.55	0.0135
>2	5.8993		
	<u>36</u>	<u>36</u>	<u>0.1424</u>

Se tiene un grado de libertad y con un nivel de significancia de 0.05 se obtiene el valor  $\chi^2 = 3.84$ , como el valor  $\chi^2 = 0.1424$  es menor que el obtenido de la tabla, no se encuentra en la región crítica y por lo tanto se acepta la hipótesis que el patrón es aleatorio.

El resultado obtenido con el cuadrante más pequeño fué que el patrón no es aleatorio, sin embargo con el cuadrante de mayor tamaño se obtiene el resultado contrario; observamos de esta manera que el resultado de la prueba de bondad de ajuste  $\chi^2$  cuadrada varía dependiendo del tamaño del cuadrante utilizado.

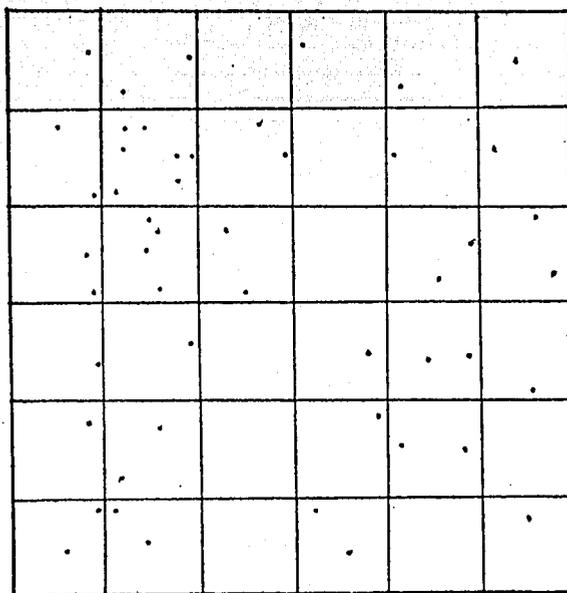


Fig. 3 Patrón de distribución sintético.

a) 

b) 

- a) tamaño del cuadrante utilizado antes.  
 b) tamaño de cuadrante utilizado ahora.

En general al cambiar el tamaño del cuadrante se altera la magnitud de  $\lambda$ , la cual es proporcional al tamaño del cuadrante.

Con el uso de cuadrantes se ha visto que los resultados se ven seriamente afectados por el tamaño del cuadrante; por lo que el investigador debe decidir cuidadosamente el tamaño adecuado de acuerdo a su estudio. Para lo cual no puede derivarse ningún criterio a partir del modelo matemático.

Otro problema con el uso de cuadrantes es la forma del cuadrante, algunos autores señalan que su efecto es menor que el del tamaño del cuadrante. Por ejemplo, Stiteler y Patil han mostrado que para un patrón de puntos localizados en los vértices de una malla de cuadrados, el muestreo con cuadrantes en forma de rectángulos delgados puede producir una distribución con una razón varianza/media mayor que uno, que podría ser considerada como evidencia de agrupación, más que de regularidad. Aunque ésta puede ser una situación extrema. (Hodder y Orton 1976, 37).

Otra desventaja señalada por Pielou es que cuando se realiza el registro de los cuadrantes en una tabla de frecuencia, no se guarda el registro de la posición relativa de los cuadrantes.

## CAPITULO 3

### ANALISIS AL VECINO MAS CERCANO.

Para el análisis del patrón de una población, además de los métodos de cuadrantes existen los métodos de distancias en los cuales se mide la distancia de un punto a otro; esta distancia proporciona una variable para la medición de espacio, que evita el efecto del tamaño del cuadrante, ya que no se utilizan cuadrantes; por esta razón son más apropiadas las pruebas basadas sobre medidas de distancias. Una prueba importante es proporcionada por Clark y Evans (1954).

#### 1. MEDIDA DE LA DISTANCIA AL VECINO MAS CERCANO DE CLARK Y EVANS.

Uno de los artículos más tempranos en el uso de medidas al vecino más cercano para detectar aleatoriedad o no aleatoriedad de un patrón de puntos es el de Clark y Evans (1954). En este artículo ellos desarrollan una medida sencilla para el estudio de patrones espaciales.

Se explica la medida para un espacio bidimensional es decir, con relación a poblaciones sobre superficies planas. Clark y Evans proponen "una medida de la manera y el grado en el cual la distribución de puntos en una población sobre un área dada, difiere de una distribución aleatoria". (Clark y Evans 1954, 446).

La medida básica que se considera es la distancia  $r$  de un punto a su vecino más cercano sin importar la dirección; se registran estas distancias utilizando todos los puntos de la población o una muestra de ella y con este conjunto de datos se obtiene la distancia media al vecino más cercano. También se calcula la distancia media al vecino más cercano que se esperaría si el patrón es aleatorio. La razón de estas dos medidas la observada y la esperada es la medida propuesta para medir la desviación de aleatoriedad y requiere de una prueba de significancia.

Para el cálculo de la distancia media esperada, si el patrón de la población es aleatorio; se considera la distancia de un punto a su vecino más cercano y se quiere encontrar la distribución de probabilidad de esta distancia. Su desarrollo se basa en la distribución Poisson la cual da para una distribución aleatoria de puntos, la probabilidad de que un área de tamaño específico seleccionada al azar, contenga  $x$  puntos. En los métodos de cuadrantes se utilizó como área especificada el cuadrante, pero ahora se considerará el sector de un círculo.

## 2. DISTRIBUCION DE LA DISTANCIA AL PUNTO MAS CERCANO EN UNA POBLACION DISPERSADA ALEATORIAMENTE.

Sea un sector de un círculo de radio  $r$  formado al dividir el círculo en  $k$  sectores iguales. Si  $d$  es la densidad promedio de la distribución entonces,

$$\lambda = \frac{\pi r^2 d}{k}$$

es el número promedio de puntos por área. Sustituyendo  $\lambda$  en la distribución Poisson, se obtiene la probabilidad de encontrar exactamente  $x$  puntos en un área arbitraria de  $\frac{\pi r^2}{k}$  unidades.

$$\frac{(\pi r^2 dk^{-1})^x e^{-\pi r^2 dk^{-1}}}{x!}$$

Por lo tanto, la probabilidad de que un área de  $\pi r^2 k^{-1}$  unidades no contenga puntos se obtiene cuando  $x$  toma el valor cero y es:

$$e^{-\pi r^2 dk^{-1}}$$

Si el área especificada es el sector de un círculo alrededor de un punto seleccionado al azar, la probabilidad que el sector no contenga puntos dentro de la distancia  $r$ , es la misma

$$e^{-\pi r^2 dk^{-1}}$$

La función de distribución requerida es la función de  $r$  que es igual a la probabilidad de la distancia al vecino más cercano menor o igual que  $r$ . Esta es la probabilidad que un sector de radio  $r$  centrado sobre un punto contenga al menos un punto es decir, que el sector no este vacío. Por lo tanto la función es:

$$1 - e^{-\pi r^2 dk^{-1}} \quad (1)$$

Los siguientes resultados no se demostrarán debido a que es necesario utilizar el cálculo integral.

Diferenciando la función (1) con respecto a  $r$ , se obtiene la función de probabilidad

$$2\pi r dk^{-1} e^{-\pi r^2 dk^{-1}} dr. \quad (2)$$

Utilizando la expresión (2) multiplicada por  $r$ , e integrando se puede demostrar que la media esperada denotada  $\bar{r}_e$  es igual a  $\frac{\sqrt{k}}{2\sqrt{d}}$ .

La varianza se puede demostrar que es igual a  $\frac{(4-\pi)k}{4\pi d}$ .

Y el error estandar de  $\bar{r}_e$ , denotado  $\sigma_{\bar{r}_e}$  es posible demostrar que es igual a

$$\sqrt{\frac{(4-\pi)k}{4\pi dN}} = \frac{0.26136 \sqrt{k}}{\sqrt{N d}}$$

donde  $N$  es el número de medidas hechas.

Para el sector considerado  $\bar{r}_e = \frac{1}{2\sqrt{d}}$

En una población de  $N$  puntos con densidad  $d$ , en la cual se ha medido la distancia de cada punto a su vecino más cercano, la distancia media observada, denotada  $\bar{r}_a$  se puede representar:

$$\bar{r}_a = \frac{\sum r_i}{N}$$

y la distancia media esperada si el patrón es aleatorio es:

$$\bar{r}_e = \frac{1}{2\sqrt{d}}$$

La razón

$$R = \frac{\bar{r}_a}{\bar{r}_e}$$

es utilizada como la medida del grado en el cual la distribución observada

se aproxima o desvía de una distribución aleatoria.

Para una distribución aleatoria el valor de  $R$  está muy próximo a uno; para un patrón regular el valor de  $R$  es mayor que uno y para un patrón agregado el valor de  $R$  es menor que uno.

Bajo condiciones de máxima agregación,  $R = 0$  dado que, todos los puntos ocupan el mismo lugar y la distancia al vecino más cercano es por lo tanto cero.

Bajo condiciones de máximo espacio, los puntos se encuentran distribuidos en un patrón hexagonal donde, cada punto (excepto los que están en la periferia de la población) estará equidistante de otros seis puntos. En esta distribución la distancia media al vecino más cercano estará maximizada.

Sea  $r_u$  la distancia constante entre vecinos más cercanos en esta distribución hexagonal. En tal distribución uniforme cada punto se puede demostrar

que ocupa un área de  $\frac{r_u^2 \sqrt{3}}{2}$ .

La densidad de la población es  $d = \frac{2}{r_u^2 \sqrt{3}}$

Despejando  $r_u$  se obtiene  $r_u = \frac{\sqrt{2}}{\sqrt[4]{3} \sqrt{d}}$

Dado que  $r_u$  es el valor posible más grande de  $\bar{r}_a$ , entonces el valor máximo de la razón es:

$$R = \frac{r_u}{\bar{r}_e} = \frac{\sqrt{2^3}}{\sqrt[4]{3} \sqrt{k}}$$

este valor es aproximadamente igual a  $\frac{2.1491}{\sqrt{k}}$

Por lo tanto, el rango del valor de R va desde 0 para una distribución con máxima agregación, a 2.1491 para una distribución de máxima uniformidad.

Clark y Evans señalan que "en cualquier distribución dada, la distancia media observada al vecino más cercano es R veces tan grande como podría esperarse en una distribución aleatoria de la misma densidad". (Clark y Evans 1954, 447).

Obsérvese que

$$R = \frac{\bar{r}_a}{\bar{r}_e}$$

implica que  $\bar{r}_a = R \bar{r}_e$  que es lo señalado por los autores.

### 3. PRUEBA DE SIGNIFICANCIA.

Para verificar si esta medida es confiable es necesario probar si el valor observado  $\bar{r}_a$  difiere significativamente del valor esperado  $\bar{r}_e$ . Se puede probar la significancia por medio de la distribución normal, utilizando para la prueba la fórmula siguiente:

$$c = \frac{\bar{r}_a - \bar{r}_e}{\sigma_{\bar{r}_e}}$$

donde c es la variable estandar de la normal,  $\sigma_{\bar{r}_e}$  es el error estandar de la distancia media al vecino más cercano en una población distribuida aleatoriamente con la misma densidad que la población observada.

Para una población de densidad d el valor de  $\sigma_{\bar{r}_e}$  es:

$$\frac{0.26136}{\sqrt{Nd}}$$

donde N es el número de medidas de distancias realizadas.

El valor de  $c$  se busca en la tabla de la distribución normal, cuando el valor de  $N$  no es pequeño ( $N \geq 100$ ); en caso de que  $N$  sea pequeña es necesario utilizar una distribución Pearson tipo III. (la cual no se ve en este trabajo)

Clark y Evans comentan que puede ocurrir que en un patrón no aleatorio, se obtengan iguales los valores de  $\bar{r}_a$  y  $\bar{r}_e$ , en este caso la prueba  $R$  falla, entonces la prueba de desviación se puede realizar por medio de una prueba  $\chi^2$  comparando las distribuciones de frecuencias de las distancias observadas y esperadas al vecino más cercano (aunque en la mayoría de las investigaciones de poblaciones naturales no se requiere de una prueba tan sensitiva).

#### 4. APLICACION DE LA MEDIDA A UN PATRON HIPOTETICO.

Considérese un patrón con 50 puntos distribuidos en un cuadrado de  $90 \times 90$  mm<sup>2</sup> como se muestra en la figura 1.

La distancia al vecino más cercano de cada punto ha sido medida y la distancia media observada al vecino más cercano obtenida es:

$$\bar{r}_a = \frac{\sum r_i}{N} = \frac{227}{50} = 4.54$$

La densidad  $d = \frac{N}{A} = \frac{50}{(90)(90)} = 0.006173$

La media esperada  $\bar{r}_e = \frac{1}{2\sqrt{d}} = \frac{1}{2\sqrt{0.006173}} = 6.3639$

La razón  $R = \frac{\bar{r}_a}{\bar{r}_e} = \frac{4.54}{6.36} = 0.7138$

Como  $R$  resulta menor que 1 el patrón tiende a ser agrupado.

Para la prueba de significancia, primero se calcula el error estandar

$$\sigma_{\bar{r}_e} = \frac{0.26136}{\sqrt{Nd}} = \frac{0.26136}{0.5555627} = .4704419$$

de donde

$$c = \frac{\bar{r}_a - \bar{r}_e}{\sigma_{\bar{r}_e}} = -3.876952$$

Buscando este valor en la tabla de la distribución normal, se obtiene una probabilidad muy baja, lo que nos significa que el patrón no es aleatorio.

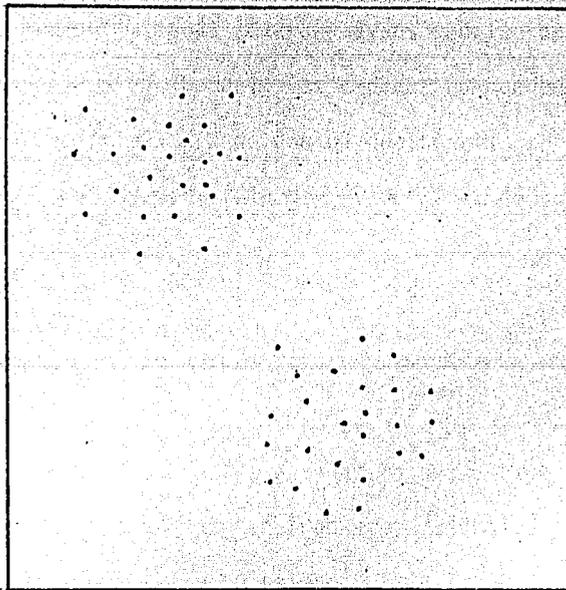


Fig. 1 Patrón de distribución sintético.

## 5. PROBLEMAS DEL PROCESO.

Puede ocurrir que en una distribución el vecino más cercano de un punto, se encuentre fuera del área especificada, esta distancia debe ser medida y -tomada en cuenta para los cálculos, pero los puntos que están fuera del área especificada no deben utilizarse como centros de medidas.

Si dos puntos seleccionados como centros de medición están más cercanos que cualquier otro punto, en este caso la misma distancia se toma dos veces.

La desviación de la distancia media esperada al vecino más cercano en una distribución aleatoria, se basa en la hipótesis que el área ocupada por la población es infinita. Pero en la práctica se aplica la medida a poblaciones finitas que ocupan áreas limitadas. Clark y Evans señalan que " la presencia de una frontera más lejana en la cual, las medidas no pueden ser hechas tenderán a hacer el valor de  $\bar{r}_a$  más grande que el que podría ser obtenido si un área infinita fuera considerada". (Clark y Evans 1954, 450).

Por ello se debe seleccionar el área de estudio con cuidado de tal manera, que el área este situada dentro del área total cubierta por la población total.

Delimitar el tamaño del área de estudio es una de las principales dificultades; en algunas ocasiones existen fronteras naturales (costas, barrancas, etc.), sin embargo frecuentemente se selecciona la frontera.

En la prueba el resultado del valor de R está sujeto a que tanto del área alrededor esta incluida en el análisis. Obsérvese el siguiente ejemplo - de Hodder y Orton(1976) del efecto de una frontera impuesta, como se muestra en la figura 2. El área de estudio es el cuadrado interior, en donde se han colocado puntos al azar.

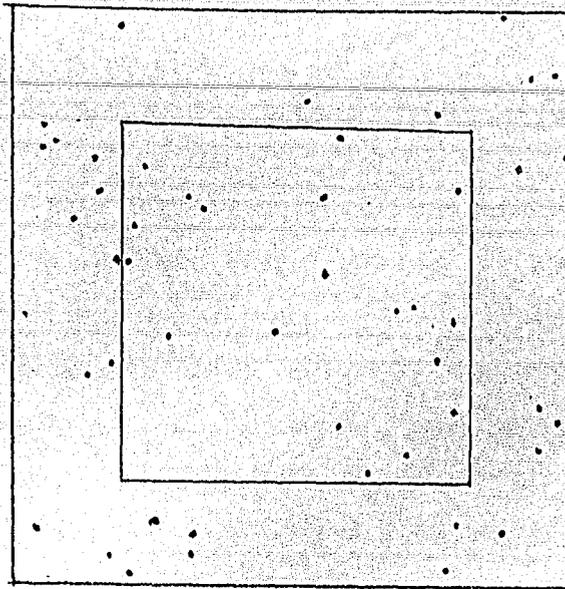


Fig. 2. El efecto de una frontera impuesta sobre un área de estudio. Hodder y orton (1976).

El valor de  $R$  obtenido fué 1.25 que indica un patrón que tiende a la uniformidad. Si se añade una banda alrededor con puntos colocados al azar, con la misma densidad y se calcula nuevamente el valor de  $R$  del área central, pero en este caso los vecinos cercanos de algunos de los puntos se encuentran en la banda. El valor de  $R$  obtenido es 1.09 el cual indica un patrón aleatorio. Se realizaron tres ensayos más en los cuales, el patrón de puntos en la banda fué de nuevo simulado, proporcionaron valores de  $R$  de 1.05, 1.04, y 1.04.

De este ejemplo se puede observar que la presencia de una frontera afecta seriamente los resultados del estudio.

### 5.1. Nota de Diggle a la prueba de Clark y Evans.

Peter J. Diggle realiza una nota sobre la prueba de Clark y Evans (Hodder y Orton 1976, 246-248). En primer lugar señala el problema del efecto de la frontera indicando que Clark y Evans recomiendan que las medidas deben ser tomadas únicamente de los puntos dentro de una subsección de la población localizada centralmente. "Notar, sin embargo, que el proceso alternativo de ignorar simplemente cualquier distancia mayor que la distancia -- del punto correspondiente al borde la población conduce a una desviación en favor de la retención de distancias pequeñas al vecino más cercano. Así la subpoblación localizada centralmente debe seleccionarse conservativamente y el tamaño de la población disponible puede ser reducido considerablemente". (Hodder y Orton 1976, 246-247).

El segundo problema que señala es la no independencia de las diferentes distancias al vecino más cercano. "Por ejemplo, Pielou (1969, 121) ha demostrado que, en una población aleatoria, dos puntos estarán más cercanos entre sí que a cualquier tercer punto con probabilidad aproximada de 0.62, y en tanto Clark y Evans establecen que 'tales distancias dobles no introducen desviación y ambas medidas deben ser usadas en los cálculos', la estructura de correlación algo compleja entre las diferentes  $x_i$  pueden afectar adversamente la aproximación Normal a la distribución de  $\bar{X}$ , y casi seguramente incrementará su varianza". (Hodder y Orton 1976, 247).

Para estudiar este problema se llevó a cabo un experimento Monte Carlo, de sus resultados Diggle indica que "la varianza estimada fué consistentemente mayor que el valor  $0.0683/np$  supuesto por Clark y Evans, sugiriendo que la prueba está expuesta a dar resultados falsamente significativos; el elemento de inexactitud así introducido es, sin embargo, de una naturaleza comparativamente menor". (Hodder y Orton 1976, 248).

## 6. ESTADISTICO DE PIELOU.

En la aplicación de la prueba de Clark y Evans cuando el número de puntos es menor que 100, no se puede utilizar para la prueba de significancia la distribución Normal; en este caso se utiliza la distribución Pearson tipo III (distribución Gamma), pero su aplicación no es fácil. Sin embargo, algunos autores han desarrollado una serie de estadísticos, los cuales relacionan directamente las distancias al vecino más cercano con la distribución  $\chi^2$  cuadrada. Entre ellos se encuentra el estadístico de Pielou (1969).

Su derivación es algo complicada y requiere del uso del cálculo integral. Por esta razón únicamente se presentan los resultados a que llega. El desarrollo está dado por Pielou (1969, 111-114).

Pielou considera la distancia desde un punto aleatorio a su vecino más cercano y desea encontrar la distribución de probabilidad de esta distancia suponiendo que el patrón de la población es aleatorio.

La función de distribución considerada es:

$$F(r) = P(\text{distancia al vecino más cercano menor o igual que } r)$$

Esta es la probabilidad que un círculo de radio  $r$  centrado sobre el punto contenga al menos otro punto, o sea que el círculo no esté vacío. Esta distribución se obtiene de forma similar a como la obtuvimos para un sector del círculo, sólo que ahora se considera como área específica un círculo de radio  $r$  y se utiliza  $e^{-\lambda r^2}$  en vez de  $e^{-\pi r^2 dk^{-1}}$  para la probabilidad que el círculo no contenga otro punto. Entonces la función de distribución es:

$$F(r) = 1 - e^{-\lambda r^2} \quad (1)$$

donde  $\lambda = \pi d$  representa el número promedio de puntos encontrados en un círculo de radio 1, en el área de estudio.

Utilizando como variable el cuadrado de la distancia por ser más sencillo y conveniente que el uso de la distancia misma, se tiene sustituyendo  $w = r^2$  en la expresión (1) la siguiente función de distribución:

$$F(w) = 1 - e^{-\lambda w} \quad (2)$$

Pielou(1969,113) realizando un desarrollo de la expresión (2) obtiene el estadístico siguiente:

$$\chi^2 = 2N\lambda\bar{w} \quad (3)$$

donde  $w = r^2$  y  $N$  es el número de distancias medidas. Esta expresión se distribuye como una distribución  $\chi^2$  cuadrada con  $2N$  grados de libertad.

### 6.1. Aplicación del estadístico.

Se aplicará el estadístico al patrón de distribución de la figura 1, el cual tiene 50 puntos.

La media del cuadrado de las distancias es:

$$\bar{w} = \frac{\sum_{i=1}^N r_i^2}{N} = 21.92$$

La densidad  $d = 0.006173$  por lo tanto,

$$\lambda = \pi d = (3.1416)(0.006173) = 0.019393$$

Sustituyendo  $\lambda$  y  $\bar{w}$  en la expresión (3) se obtiene:

$$\chi^2 = 2N\lambda\bar{w} = 2(50)(0.019393)(21.93) = 42.50$$

Se tienen  $v=2N= 2(50)= 100$  grados de libertad. Considerando un nivel de significancia de 0.05 se obtienen los puntos críticos  $x_1 = 73.77152$  y  $x_2 = 129.069$  como el valor de  $\chi^2$  se encuentra en la región crítica entonces se rechaza la hipótesis de aleatoriedad y se concluye que el patrón no es aleatorio.

## CAPITULO 4

### ALGUNAS APLICACIONES DE LAS PRUEBAS.

Las pruebas de aleatoriedad  $\chi^2$  cuadrada, varianza/media y la de Clark y Evans, se han desarrollado básicamente en ecología y han sido aplicadas - en problemas de patrones de distribución de vegetación natural. Ejemplos de este campo pueden verse en los artículos de Clark y Evans (1954) y Greig--Smith (1952,1964). En este capítulo en la primera parte consideraremos unos ejemplos sintéticos a los cuales se aplicarán las pruebas, observaremos si los resultados obtenidos coinciden.

En la segunda parte se mostrarán algunas aplicaciones prácticas de las pruebas que se han realizado en arqueología.

Los arqueólogos se han interesado por la aplicación de estas pruebas - en el análisis de patrones de asentamiento.

## 1. APLICACIONES DE LAS PRUEBAS A PATRONES HIPOTETICOS.

Considerense los cinco patrones hipotéticos que se muestran en las figuras 1, 2, 3, 4 y 5; a los cuales se aplican las pruebas  $\chi^2$  cuadrada, varianza/media, la de Clark y Evans y el estadístico de Pielou.

Cada uno de los patrones tiene 50 puntos dentro de un área cuadrada de  $90 \times 90 \text{ mm}^2$ . Para las pruebas  $\chi^2$  cuadrada y razón varianza/media ( $s^2/\bar{X}$ ), el área es dividida en 81 cuadrantes de  $10 \times 10 \text{ mm}^2$  cada uno.

Las frecuencias observadas en cada patrón se muestran en la tabla 1.

Tabla 1

No. de puntos por cuadrante	No. de cuadrantes observados				
	patrón I	II	III	IV	V
0	64	45	31	36	67
1	2	23	50	40	0
2	4	12	0	5	3
3	6	1	0	0	3
4	3	0	0	0	5
5	2	0	0	0	3

En la aplicación de la prueba  $\chi^2$  cuadrada se obtuvieron los resultados que se muestran en la tabla 2.

Tabla 2

Patrón	I	II	III	IV	V
$\chi^2$	34.4864	1.3096	34.9880	10.4045	40.7022

Con un nivel de significancia de 0.05 y  $\nu=1$  grado de libertad, se obtiene como punto crítico 3.84; por lo tanto, el único de los valores de  $\chi^2$  que es menor que el valor crítico es el correspondiente al patrón II y concluimos que es un patrón aleatorio, en tanto que los valores de  $\chi^2$  de los restantes patrones se encuentran en la región crítica y resultando entonces que los patrones I, III, IV y V no son aleatorios.

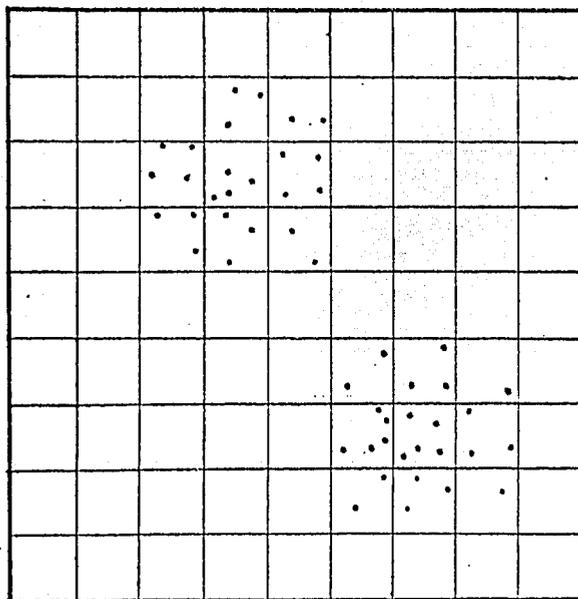


Fig. 1 Patrón de distribución sintético I.

De la aplicación de la razón varianza/media se obtienen los resultados de la tabla 3.

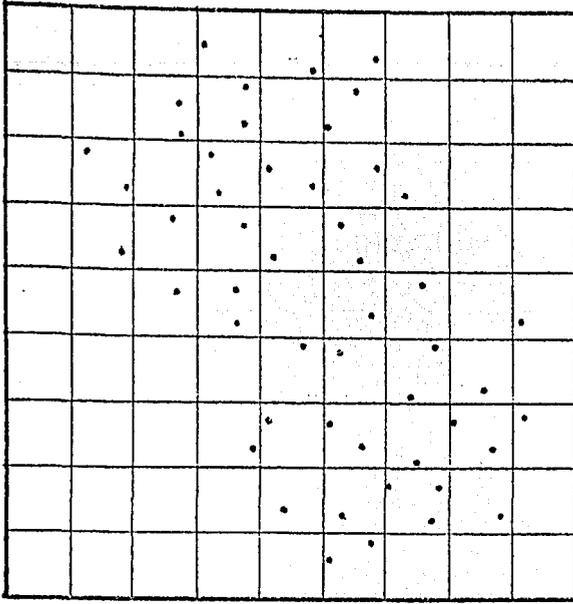


Fig. 2 Patrón de distribución sintético II.

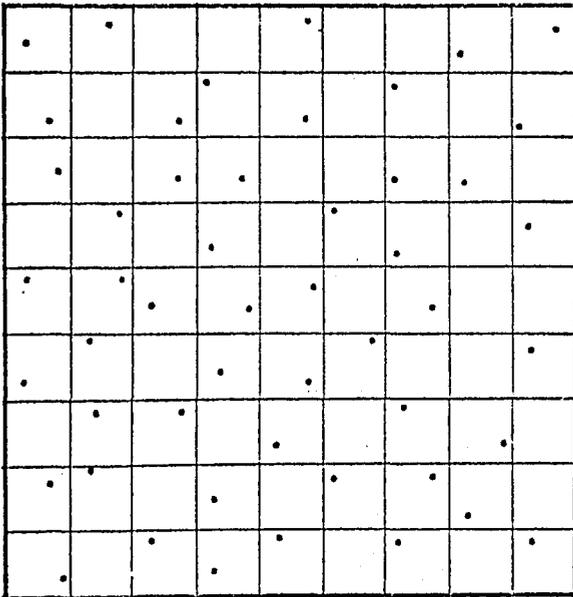


Fig. 3 Patrón de distribución sintético III

Tabla 3

Patrón	I	II	III	IV	V
$\bar{X}$	0.6173	0.6173	0.6173	0.6173	0.6173
$s^2$	1.7392	0.6142	0.2392	0.3642	2.0392
$s^2/\bar{X}$	2.8174	0.9950	0.3875	0.5899	3.3034
n	81	81	81	81	81
I	225.39	79.598	30.999	47.199	264.27
T	8.62	0.01	-4.74	-2.89	10.38
P	baja	-0.504	baja	0.002	baja
v	80	80	80	80	80
puntos críticos	$x_1=56.7061$ y $x_2=106.1355$ con $\alpha=0.05$				

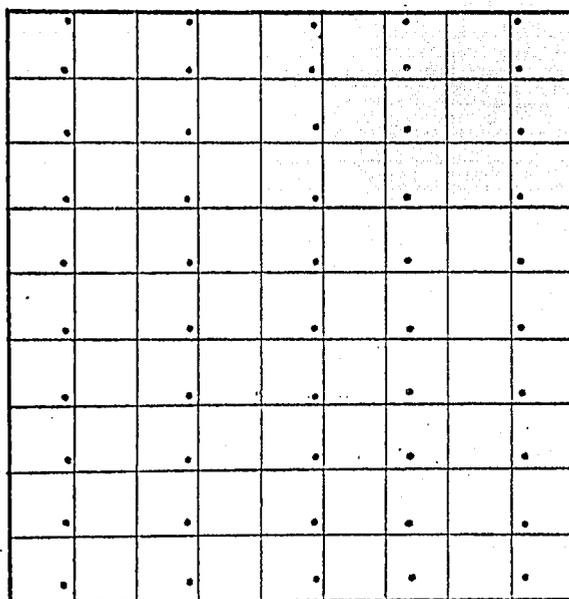


Fig. 4 Patrón de distribución sintético IV.

De los datos de la tabla 3 se puede observar que el valor de la razón  $(s^2/\bar{X})=0.9950$  del patrón II está próximo a la unidad y su valor del índice no se encuentra dentro de la región crítica, por lo tanto es aleatorio.

Para el patrón I  $(s^2/\bar{X})= 2.8174$  y para el patrón V,  $(s^2/\bar{X})= 3.3034$ , indican que los patrones I y V tienden a un tipo de patrón agregado. Los valores para los patrones III y IV  $(s^2/\bar{X})= 0.3875$  y  $(s^2/\bar{X})= 0.5899$  indican que los patrones III y IV tienden a ser regulares. Ninguno de los valores de la razón se encuentra cercano a la unidad y en cada caso el valor del índice se encuentra en la región crítica.

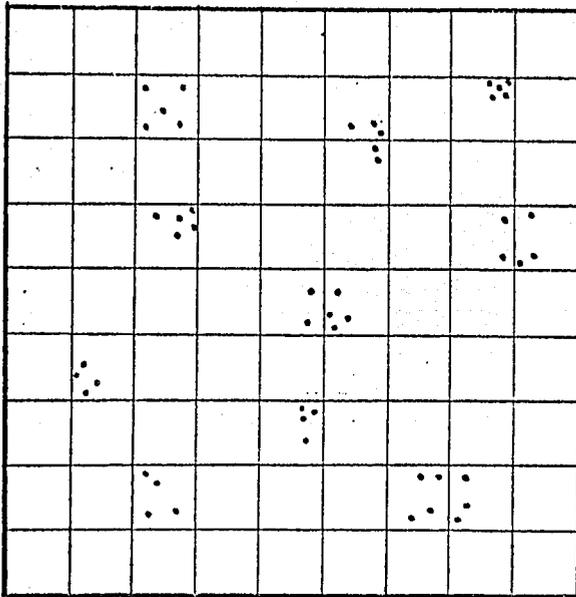


Fig. 5 Patrón de distribución sintético V.

Para la aplicación de la prueba de Clark y Evans se cuenta con 50 distancias medibles para cada patrón; se obtienen los resultados que se muestran en seguida:

$$A = 8100 \text{ mm}^2$$

$$d = 0.006173$$

$$\sqrt{Vd} = 0.078568$$

$$\sigma_{\bar{r}_e} = 0.471442$$

Tabla 4

Patrón	I	II	III	IV	V
$\Sigma r$	227	343	529	450	142
$\bar{r}_a$	4.54	6.86	10.58	9	2.84
$\bar{r}_e$	6.36388	6.36388	6.36388	6.36388	6.36388
R	0.71340	1.07796	1.66251	1.41423	0.44627
c	3.87695	1.05458	8.96204	5.60349	3.52388
Probabilidad de una mayor diferencia entre $\bar{r}_e$ y $\bar{r}_a$ .	.000002	0.28914	.000002	.000002	.000002

El valor de R para el patrón II no se encuentra lejos de la unidad, que es el valor esperado para una distribución perfectamente aleatoria y la prueba de significancia de la diferencia entre la distancia media esperada y la observada da un valor de  $c=1.05458$  indicando que la mayor desviación de lo esperado ~~de~~ ocurrir 28.914% de las veces únicamente por oportunidad. Así que el patrón II resulta ser aleatorio.

Para el patrón I,  $R=0.7134$  y para el patrón V,  $R=0.4463$  estos valores indican agregación, en tanto que para el patrón III,  $R=1.6625$  y para el patrón IV,  $R=1.4192$ , indican desviación del patrón aleatorio en la direc-

ción de regularidad. Ninguno de los valores de R esta próximo a 1, y en cada caso la probabilidad de obtener en una distribución aleatoria de la misma densidad un valor de  $\hat{c}$  como el observado es menor que .000002.

En la aplicación del estadístico  $\chi^2$  se obtuvieron los resultados - presentados en la tabla 5

Tabla 5

Patrón	I	II	III	IV	V
$\Sigma r^2$	1096	2489	5849	4050	444
$\lambda$	.019393	.019393	.019393	.019393	.019393
$\bar{w}$	21.92	49.78	116.98	81	8.88
$\chi^2$	42.508	96.538	226.859	157.083	17.221
$v$	100	100	100	100	100

Puntos  
críticos

$x_1 = 73.77152$  y  $x_2 = 129.0699$  con  $\alpha = 0.05$

El valor de  $\chi^2 = 96.538$  de el patrón II, con un nivel de significancia de 0.05 para una prueba de dos colas, no se encuentra en la región crítica, - por lo tanto, aceptamos la hipótesis de aleatoriedad para el patrón II.

Sin embargo, los valores de  $\chi^2$  para los patrones I, III, IV y V (véase tabla 5) se encuentran en la región crítica y por lo tanto no son aleatorios, tendiendo los patrones I y V en dirección de agregación y los patrones III y IV en dirección de regularidad.

Con la aplicación de las diferentes pruebas a cada uno de los patrones llegamos a conclusiones similares, obsérvese la tabla 6.

Tabla 6

Prueba	patrón				
	I	II	III	IV	V
$\chi^2$	34.4864	1.3096	34.9880	10.4045	40.7022
$s^2/\bar{X}$	2.8174	0.9950	0.3875	0.5899	3.3034
R	0.7134	1.0780	1.6625	1.4142	0.4463
$\chi^2$	42.508	96.538	226.859	157.083	17.221

## 2. APLICACION DE LAS PRUEBAS EN ARQUEOLOGIA.

Las pruebas de aleatoriedad han sido utilizadas por los arqueólogos para el estudio de patrones de distribución de artefactos dentro de un sitio, de patrones de artefactos dispersos sobre un área y patrones de distribución de sitios.

La principal ventaja de las pruebas basadas en métodos de cuadrantes es que la información arqueológica es frecuentemente coleccionada en esta forma.

Por ejemplo Dacey(1973) ha utilizado la prueba de la razón varianza/media para el análisis de los patrones de distribución de tres tipos de artefactos líticos: raspadores, raspadores carenados y buriles que se encontraron en el sitio de 63 m<sup>2</sup>. Sde Divshon del área de Avdat. El sitio fué dividido en 63 cuadrantes de 1 m<sup>2</sup> cada uno, los resultados obtenidos se muestran en la tabla 7.

Tabla 7

	raspadores	raspadores carenados	buriles
$\bar{X}$	1.83	0.43	1.29
$s^2$	2.59	0.50	1.73
$s^2/\bar{X}$	1.38	1.16	1.34
n	63	63	63
I	87.00	73.33	84.67
$P(\chi^2_{n-1} \leq I)$	.98	.84	.97

Utilizando un nivel de significancia del 0.05, los datos de la tabla no apoyan la hipótesis que los patrones de los raspadores y de los buriles sean aleatorios, en tanto que el patrón de los raspadores carenados resulta ser aleatorio.

Para la distribución de los raspadores carenados Dacey vuelve a aplicar la prueba pero, con un cambio de tamaño del cuadrante, en este caso se consideraron cuadrantes rectangulares de 1x3 m.; los datos obtenidos se presentan en la tabla 8.

Tabla 8

$\bar{X}$	1.29
$s^2$	2.20
$s^2/\bar{X}$	1.71
n	21
I	35.99
$P(\chi^2_{n-1} \leq I)$	.98

En esta segunda aplicación, como se puede observar de los datos de la tabla la prueba no apoya la hipótesis de que estén distribuidos aleatoriamente.

Dacey concluye. "Estas pruebas de  $s^2/\bar{X}$  no proporcionan evidencia convincente de que los mapas de los tres tipos de artefactos líticos tengan patrones aleatorios". (Dacey 1973, 324).

Se han aplicado dos tamaños diferentes de cuadrantes a uno solo de los patrones y se han obtenido resultados diferentes, es necesario preguntarse. ¿Cuál de los dos resultados obtenidos para el patrón de los raspadores carenados es el más apropiado para el análisis práctico de los datos arqueológicos?

Thomas(1972, 698-699) ha utilizado la razón  $s^2/\bar{X}$  para analizar la distribución de artefactos sobre un área.

La prueba de Clark y Evans ha sido utilizada para estudiar las distribuciones de sitios arqueológicos. Por ejemplo, Hodder(1972) la utilizó para la identificación de un patrón aleatorio en la primera etapa de su estudio del asentamiento de los pueblos amurallados del Romano-Británico.

Del análisis del patrón de distribución de los pueblos amurallados obtuvo los siguientes resultados:

Tabla 9

A	367 cm <sup>2</sup>
n	25
d	0.0653
$\bar{r}_a$	2.644
$r_e$	1.957
R	1.351
$\sigma_{r_e}$	0.2046

## c 3.352

Como la muestra es pequeña utilizó la Pearson tipo III y obtuvo que la probabilidad de una mayor diferencia entre  $\bar{r}_a$  y  $\bar{r}_e$  es 0.069% que indica que el patrón no es aleatorio y el valor de R señala que tiende hacia un patrón regular.

Hodder señala que "este resultado indica que estadísticamente es altamente probable que el espacio relativo de los pueblos amurallados en el área seleccionada para el análisis no es aleatorio, sino sugiere más bien - que existe alguna relación inter-sitio. Además el valor de R muestra que esta relación tiende hacia un rechazamiento mutuo y espaciación uniforme". (Hodder 1972, 892).

Whallon (1974) aplica el análisis del vecino más cercano a un conjunto de datos arqueológicos, selecciona cuatro tipos de artefactos líticos dispersados sobre un piso de ocupación Protomagdeliano en el Abri Pataud en el sureste de Francia.

La ocupación protomagdeliana fué escavada sobre un área de 12 m. de largo por 2.5 o 3 m. de ancho. En esta área se registraron de los artefactos del tipo (1) 32, del tipo (2) 15, del tipo (3) 20 y del tipo (4) 20.

Los resultados de la aplicación de la prueba de Clark y Evans y del estadístico de Pielou para cada uno de los patrones se muestran en la tabla. El número de observaciones originales ha sido reducido por la eliminación de objetos que están más cercanos a la frontera del área de estudio - que a otro objeto, también se eliminan algunos objetos que no tienen vecino. Esto se refiere a la eliminación de objetos aislados en las esquinas de un

área formada irregularmente de tal forma que no existe distancia en línea recta de ellos a otro objeto que no cruce una frontera. El criterio de "no vecino" también detecta y elimina objetos que están localizados afuera del área de análisis a través de medidas o registrando error.

Tabla 10

Artefacto	(1)	(2)	(3)	(4)
N	24	10	11	11
A	393249.9	39.3249.9	39.3249.9	39.3249.9
d	.00061	.000025	.000028	.000028
$\bar{r}_a$	28.68	53.34	41.78	46.50
$\bar{r}_e$	64.00	99.15	94.54	94.54
R	0.4481	0.5380	0.4419	0.4919
$\sigma_{\bar{r}_e}$	5.9142	13.3822	11.05	11.05
c	-5.973	-3.386	-4.774	-4.348
Puntos críticos de la normal	$x_1 = -1.96$ y $x_2 = 1.96$ con $\alpha = 0.05$			
v	48	20	22	22
$\chi^2$	9.917	5.627	4.744	6.45

Estos puntos pueden eliminarse por el criterio de la cercanía a una frontera, pero el criterio de "no vecino" puede ser útil para la detección de errores en las coordenadas.

En cada caso la media observada de las distancias al vecino más cercano es menor que la media de las distancias esperadas. Los cuatro patrones resultan ser significativamente agregados usando un nivel de significancia de 0.05. Whallon observa que " los estadísticos  $\chi^2$  y las aproxi

maciones normales calculadas de ellos son más conservativas en su detección de agregación espacial que las de Clark y Evans. En algunas situaciones, los primeros podrían indudablemente fallar para rechazar la hipótesis de aleatoriedad donde las pruebas de Clark y Evans podrían indicar agregación". (Whallon 1974, 31-33).

## C O N C L U S I O N

En los capítulos precedentes se han visto que las pruebas para el análisis de patrones de puntos son útiles en arqueología debido a que - las distribuciones de sitios y artefactos son grupos importantes de datos arqueológicos.

Se han señalado algunos problemas de la aplicación de estas pruebas, destacando que requieren datos adecuados para su uso y un problema general que se presenta en su aplicación es que las pruebas estadísticas que han sido usadas requieren que sean hechas ciertas suposiciones sobre los datos, en muchos casos en arqueología no es posible esperar que las suposiciones necesarias se satisfagan razonablemente. Por lo cual resulta que las pruebas no pueden ser utilizadas con el rigor posible en otras disciplina y deben ser observadas con el cuidado y la precaución - que se han indicado en cada caso.

Quizás una forma satisfactoria de probar la validéz de los patrones observados, es coleccionar datos adicionales, para ver si los nuevos datos - concuerdan con las tendencias observadas.

La aproximación subjetiva en el análisis de distribuciones de puntos es inadecuada para la investigación de patrones; los métodos vistos permiten medir el grado de desviación de la distribución observada de un patrón aleatorio; las pruebas describen los patrones en una forma que no sería posible subjetivamente.

Es deseable que los arqueólogos se interesen por las posibilidades que ofrecen estas técnicas, para coleccionar en lo futuro datos adecuados para - su aplicación y para contribuir al estudio de las distribuciones que se presentan en arqueología.

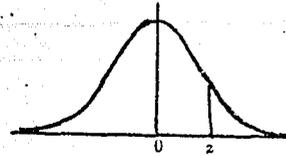
TABLA DE LA DISTRIBUCION XI CUADRADA \*

n	'99	'98	'95	'90	'80	'70	'50	'30	'20	'10	'05	'02	'01	'001
1	0.157	0.628	0.0393	0.158	0.642	1.148	1.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.201	0.404	1.103	1.211	1.446	1.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	1.115	1.185	3.352	5.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345	16.266
4	2.297	4.429	7.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.467
5	5.554	7.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.515
6	8.72	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.990	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.827
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703
32	16.362	17.783	20.072	22.271	25.148	27.373	31.336	35.665	38.466	42.585	46.194	50.487	53.486	62.487
34	17.789	19.275	21.664	23.952	26.938	29.242	33.336	37.795	40.676	44.903	48.602	52.995	56.061	65.247
36	19.233	20.783	23.269	25.643	28.735	31.115	35.336	39.922	42.879	47.212	50.999	55.489	58.619	67.985
38	20.691	22.304	24.884	27.343	30.537	32.992	37.335	42.045	45.076	49.513	53.384	57.969	61.162	70.703
40	22.164	23.838	26.509	29.051	32.345	34.872	39.335	44.165	47.269	51.805	55.759	60.436	63.691	73.402
42	23.650	25.383	28.144	30.765	34.157	36.755	41.335	46.282	49.456	54.090	58.124	62.892	66.206	76.084
44	25.148	26.939	29.787	32.487	35.974	38.641	43.335	48.396	51.639	56.369	60.481	65.337	68.710	78.750
46	26.657	28.504	31.439	34.215	37.795	40.529	45.335	50.507	53.818	58.641	62.830	67.771	71.201	81.400
48	28.177	30.080	33.098	35.949	39.621	42.420	47.335	52.616	55.993	60.907	65.171	70.197	73.683	83.037
50	29.707	31.664	34.764	37.689	41.449	44.313	49.335	54.723	58.164	63.167	67.505	72.613	76.154	86.661
52	31.246	33.256	36.437	39.433	43.281	46.209	51.335	56.827	60.332	65.422	69.832	75.021	78.616	89.272
54	32.793	34.856	38.116	41.183	45.117	48.106	53.335	58.930	62.496	67.673	72.153	77.422	81.069	91.872
56	34.350	36.464	39.801	42.937	46.955	50.005	55.335	61.031	64.658	69.919	74.468	79.815	83.513	94.461
58	35.913	38.078	41.492	44.696	48.797	51.906	57.335	63.129	66.816	72.160	76.778	82.201	85.950	97.039
60	37.485	39.699	43.188	46.459	50.641	53.809	59.335	65.227	68.972	74.397	79.082	84.580	88.379	99.607
62	39.063	41.327	44.889	48.226	52.487	55.714	61.335	67.322	71.125	76.630	81.381	86.953	90.802	102.166
64	40.649	42.960	46.595	49.996	54.336	57.620	63.335	69.416	73.276	78.860	83.675	89.320	93.217	104.716
66	42.240	44.599	48.305	51.770	56.188	59.527	65.335	71.508	75.424	81.085	85.965	91.681	95.626	107.258
68	43.838	46.244	50.020	53.548	58.042	61.436	67.335	73.600	77.571	83.308	88.250	94.037	98.028	109.791
70	45.442	47.893	51.739	55.329	59.898	63.346	69.334	75.689	79.715	85.527	90.531	96.388	100.425	112.317

\* Tabla tomada de Statistical Tables de R.A. Fisher y Frank Yates, publicado por Oliver & Boyd, Ltd., Edimburgo y Londres. p 47

AREAS  
BAJO LA  
CURVA NORMAL \*

DE 0 a z



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

\* Tabla tomada del apendice de Estadísticas de la serie Schaum de Murray

VALORES DE  $e^{-\lambda}$  \*

$$(0 < \lambda < 1)$$

$\lambda$	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.9048	0.8958	0.8869	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.8187	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7408	0.7334	0.7261	0.7189	0.7118	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6703	0.6636	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.6065	0.6005	0.5945	0.5886	0.5827	0.5770	0.5712	0.5655	0.5599	0.5543
0.6	0.5488	0.5434	0.5379	0.5326	0.5273	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4966	0.4916	0.4868	0.4819	0.4771	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4493	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4066	0.4025	0.3985	0.3946	0.3906	0.3867	0.3829	0.3791	0.3753	0.3716

$$(\lambda = 1, 2, 3, \dots, 10)$$

$\lambda$	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$	0.36788	0.13534	0.04979	0.01832	0.006738	0.002479	0.000912	0.000335	0.000123	0.000045

Nota: Para obtener los valores de  $e^{-\lambda}$  para otros valores de  $\lambda$ , emplear las leyes de los exponentes.

Ejemplo:  $e^{-3.48} = (e^{-3.00})(e^{-0.48}) = (0.04979)(0.6188) = 0.03081$ .

\* Tabla tomada del apéndice de Estadística de la serie Shaun de Murray y Spiegel, publicada por libros Mc. Graw - Hill, P. 348

## BIBLIOGRAFIA.

- Bhattacharyya, G. and Johnson, R.A., *Statistical concepts and methods*, USA, John Wiley & Sons, 1977.
- Clark, P.J. and Evans, F.C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, 35, 445-453.
- Dacey, M.F. (1973). Statistical tests of spatial association in the locations of tool types. *American Antiquity*, 38, 320-328.
- Greig-Smith, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany, London*, 16, 293-316.
- Greig-Smith, P., *Quantitative plant ecology*, London, Methuen, 1964.
- Hodder, I.R. (1972). Locational models and the study of Romano-British settlement. In Clarke, D.L. (ed.) *Models in archaeology*, pp. 887-909. Methuen, London.
- Hodder, I. and Orton, C. *Spatial analysis in archaeology*. Cambridge University Press, Cambridge.
- Hoel, P.G., *Estadística elemental*, México, Compañía continental, S.A., 1977.
- Pielou, E.C., *An introduction to mathematical ecology*, London, (1969).
- Thomas, D.H. (1972). A computer simulation model of Great Basin Shoshonean subsistence and settlement patterns. In Clarke, D.L. (ed.), *Models in Archaeology*, pp. 671-704. Methuen, London.

Whallon, R. (1974). Spatial analysis of occupation floors, II,  
The application of nearest neighbour analysis. *American  
Antiquity*, 39, 16-34.