LA DESCRIPCION DE LA INFORMACION GEOGRAFICA



Ventajas de las medidas numéricas.-

Posiblemente el paso más básico en cualquiera ciencia consiste en describir los hechos tanto numéricamente como en palabras. Expresar los hechos en forma numérica es a menudo laboriosa, más dificil y menos sutil que describirlos verbalmente, pero las medidas numéricas tienen varias ventajas; primero, son más precisas. Es posible, por Ej. describir una unidad como grande o muy grande, o aún enorme, pero si se midiera el número de habitantes o el área de la unidad, sería posible indicar todas las gamas de tamaño. Los números son tambien más exactos. Dos observadores pueden diferir en sus ideas de lo que constituye una ciudad "grande", pero si se les pide que midan el número de habitantes, deben llegar a la misma conclusión. Las comparaciones se hacen más fáciles con el uso de números: es posible decir no sólo que una unidad es más grande que otra, pero exactamente cuanto más grande. Los números se pueden manejar mucho más fácilmente, que planteamientos verbales, se pueden sumar o dividir para dar medidad secundarias útiles, como por Ej. la densidad media de la población.

Finalmente, si el problema que se puede expresar numéricamente, hay una amplia gama de herramientas matemáticas y estadísticas para ayudar a solucionarlo. Cuando se emplean bien estas "herramientas" nos pueden decir más sobre el problema que lo que podríamos averiguar a través de un estudio intensivo, por muy sensitivo que fuera.

POBLACIONES Y MUESTRAS

Un objeto de estudio común en la Geografía es un grupo de entidades distribuidas sobre un área, por ejemplo, un grupo de cuencas hidrográficas, en un estudio de geomorfología plural, o un grupo de fundos en una región en un estudio de geogra fía agrícola.

A veces es posible estudiar todos los sistemas hidrográficos o todos los fundos que abarca la investigación, pero generalmente no lo es, debido a la gran cantidad de información geográfica disponible, a menudo se acostumbra seleccionar una pequeña parte para estudiarla, y entonces aplicar las conclusiones obtenidad a todo el sistema. Así, se podría describir uno o dos fundos en detalle, y estos un estudio de casos, sería una guía para la organización de probablemente miles de otros fundos en la misma área. Hay que entonces distinguir entre todos los objetos que interesan a la investigación (llamada la "Poblacion") y el número menor de objetos que efectivamente se estudian (llamado la "Muestra").

Es necesario hacer además, otra ~~distribución~~ distinción entre estudios que intentan describir y aplicar las características de una muestra, y otro tipo de estudio que saca conclusiones generales sobre las características de la población a partir de las características de la muestra. Por ejemplo, una selección de, digamos, 50 pedimentos en un área desértica puede constituir un tema digno de estudio por si mismo, o la información obtenida puede usarse para formular afirmaciones sobre los procesos de la formación de cadenas bajo condiciones áridas en general. Puede que esta diferenciación aparezca algo pedante, pero es importante porque tambien divide a los 2 tipos de herramienta estadística disponible. La estadística "descriptiva" se usa, como su nombre implica, para describir cierto grupo de datos; mientras la estadística "inferencial" (o pruebas estadísticas) nos ayudan a formular conclusiones generales a partir de ciertos datos.

METODOS GEOGRAFICOS DE MUESTREO

Si tomamos el ejemplo de un estudio de fundos en una región, es posible mostrar una manera en que difieren la geografía cuantitativa y la geografía en sus maneras de abordar el problema. El método tradicional podría consistir en seleccionar un pequeño número de fundos que según la opinión experta del investigador son típicos de los fundos de la región. En seguida estos se analizan con gran detalle. Una crítica posible a este método es que las opiniones de los expertos no siempre coincide y, por lo tanto podrá ser preferible usar un método más objetivo para identificar lo que es típico.

Un método alternativo, entonces, para seleccionar la muestra es el azar, con la esperanza de que si se pueden incluír suficientes fundos, el resultado será típico de la región. Tal muestreo al azar o "aleatorio" puede lograrse, por ejemplo, sacando los nombres de los fundos de un sombrero que contenga todos los nombres. Un método alternativo podría ser asignándole a cada fundo un número y usando tablas de números aleatorios para seleccionar, digamos, "fuentes" de ellos para estudiarlos. Pero otros métodos de muestreo han sido desarrollados en la geografía para tomar en cuenta el hecho de que individuos geográficos se encuentran distribuidos sobre un área.

<u>MUESTRAS ALEATORIAS Y SISTEMATICAS</u>

Supongamos que fuera necesario obtener una selección imparcial de 30 localidades dentro de un área cuadrada, para el estudio de la variación de tipos de suelo.

Se puede tomar una muestra aleatoria marcando el área en un mapa y trazando un reticulado de coordenadas norte-sur y este-oeste sobre el área. Dos números aleatorios de una tabla de números aleatorios ubicarán entonces un punto para efectuar un muestreo, si el primer número representa la coordenada norte-sur y el segundo la este-oeste. Así treinta pares de números aleatorios definirán treinta localidades totalmente imparciales. Sin embargo, la desventaja de la muestra aleatoria es la posibilidad de que resulte una distribución muy dispareja, con algunas áreas bien cubiertas y otras muy mal representadas (Fig. 1). Una solución para este problema es la <u>Muestra sistemática</u> en que el primer punto se elije en forma aleatoria, pero todos los puntos subsiguientes están ubicados sobre un trazado regular (Fig.2). Así se asegura una distribución pareja de puntos, pero existe el pequeño peligro de que la regularidad de los puntos del muestreo podría coincidir con la distribución del fenómeno que se estudia. Por ejemplo, la distribución del uso de la tierra en una ciudad podría seguir el trazado rectangular de las calles y manzanas, y una muestra sistemática podría dar una imagen muy falsa del área entera.

MUESTRA ALEATORIA ESTRATIFICADA

Un término medio acertado que combine el caracter imparcial de la muestra aleatoria y la buena distribución de la muestra sistemática es el método <u>aleatorio estratificado</u>, en que el área de estudio primero se divide en áreas más pequeñas (~~Stra-ba~~ Estratos) y se selecciona cierto número de puntos de cada Stratum. El Area podría dividirse en forma regular (Fig.3) o según la distribución de una variable que se supone influirá sobre el fenómeno que s estudia. Por Ej., una investigación de tipo de suelo en un área que contiene mesetas, laderas abruptas y fondos de valles, podría tener el mismo número de muestras aleatorias de cada uno de los tres tipos de sitio (Fig.4).

Hay por supuesto, otros métodos para el muestreo de datos geográficos:

- métodos que emplean ubicaciones de punto,
- métodos que sobreponen areas menores ("cuadrados") al área de estudio.
- métodos que estudian los fenómenos a lo largo de una línea (transversal). Un resumen se encuentra en P. Haggett: Locational Analysis in Human Geography. Chap ter 7 y en S. Gregory: Statistical Methods and the geographer. Chapter 7

## MEDIDAS DE PROMEDIO

### La medida aritmética

Habiendo obtenido una serie de datos, ya sea de la población entera (por ejemplo en un estudio de Chile, usando datos al nivel provincial) o de una muestra (por ejemplo de 100 comunas), el próximo paso consiste en describir la información de una manera eficiente. La manera más directa de comprimir cierta cantidad de datos para representarlos por una sola medida significante es, desde luego, calculando el valor promedio. Para la mayoría de los propósitos geográficos, la <u>media aritmética</u>, es la mejor medida de tendencia central o promedio. La media aritmética se calcula de la fórmula:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

donde $\bar{X}$ es la medida aritmética, Xi es el valor de la variable X, y N es el número de observaciones. El ~~símbolo~~ subíndice i puede significar cualquier número entre 1 y N. En palabras, la fórmula dice: Sume todos los valores de X desde el primero hasta el último y divida por el número de observaciones. La población media de una provincia chilena se obtiene sumando la población de las 25 provincias y dividiendo por 25.

### La Mediana y la Moda

Otras medidas de promedio no son tan conocidad pero pueden ser útiles en ciertas circunstancias, la <u>mediana</u> es el valor del medio cuando todas las observaciones están ordenadas según su tamaño. Obsérvense las cifras siguientes:

10  20  40  <u>40</u>  40  60  490

Su media aritmética es 100, pero la mediana (el cuarto valor, contando desde cada extremo) es 40, una medida mucho más representativa cuando una de las cifras es muy extrema.

En el caso de las cifras de población de las provincias, por ejemplo, el valor muy alto de Santiago podría falsear la media aritmética, y la mediana que no toma en cuenta los valores extremos podría resultar, en este caso, un mejor valor representativo del promedio. La tercera medida, la moda, se usa más comunmente para datos para los cuales no conviene usar i la media aritmética ni la mediana. Si por ejemplo, de 50 fundos 15 practican predominantemente los cultivos propios de tierras arables, 10 se dedican a la lechería y 25 a las hortalizas, entonces no es posible calcular un tipo de agricultura promedio o mediano, pero es posible decir que el tipo que ocurre más frecuentemente (la moda) es el cultivo de hortalizas.

## MEDIDAS DE VARIABILIDAD

La media aritmética, la mediana y la moda miden el promedio de cierta información dada, pero el promedio tiene poco valor si no sabemos si las observaciones están cerca del promedio, o si hay mucha variación alrededor del promedio. Para obtener una impresión de la distribución de las observaciones alrededor del promedio, se acostumbra dibujar un histograma. Esto se hace contando el número de observaciones que caen dentro de una serie de categorías de tamaño separadas por intervalos iguales y dibujar los resultados como un diagrama.

Veamos el caso de 11 medidas de precipitación anual, tomadas en 2 lugares distintos.

| | A | B |
|---|---|---|
| 1963 | 61 | 62 (cm) |
| 1964 | 56 | 55 |
| 1965 | 54 | 54 |
| 1966 | 51 | 46 |
| 1967 | 54 | 53 |
| 1968 | 59 | 64 |
| 1969 | 56 | 57 |
| 1970 | 49 | 51 |
| 1971 | 52 | 48 |
| 1972 | 58 | 59 |
| 1973 | 55 | 56 |

La media aritmética de la precipitación anual en ambos lugares es 55 cm. No es fácil, si sólo se miran los números, obtener una buena idea de la varialidad de la precipitación en ambos lugares, pero si la información se presenta en dos histogramas (Fig.5) se puede obtener una impresión muy clara. Las observaciones en el lugar A están mucho más agrupadas alrededor del promedio que aquellas del lugar B, o en otras palabras, la precipitación en B es más variable.

Existen diversas medidas de variabilidad, de las cuales la más ~~servible~~ común es el rango. El rango se obtiene simplemente calculando la diferencia entre el valor más alto en el más bajo. Existen otras versiones más complicadas del rango, tales como el rango percentil, pero todas estas medidas presentan la desventaja que se derivan no de todo el grupo de observaciones, sino de los dos valores extremos.

<u>Desviación Standard</u>

La medida más satisfactoria de varialidad es la <u>desviación standard</u>, que es un índice del valor medio en que cada observación se desvía del promedio. Primero se resta el promedio de cada observación, para dar la cantidad en que cada observación se desvía del promedio. Alrededor de la mitad de estas desviaciones serán negativas, así es que se cuadran todas las desviaciones para eliminar los signos menos; luego estas desviaciones ~~cuadradas~~ elevadas al cuadrado se suman (si no se hubieran ~~cuadrado~~ elevado al cuadrado la suma sería 0) y se divide por el número de observaciones. Esta desviación media cuadrática se llama <u>VARIANZA</u> ($S^2$).

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

La varianza de la precipitación en el lugar A es 11.4 $cm^2$ y en el lugar B 27,5 $cm^2$. Por supuesto que la varianza se podría usar como una medida de varialidad (y a veces se usa así) pero como es un valor cuadrado (aquí $cm^2$), es usual extraer la raíz cuadrada para obtener una medida de variabilidad en las mismas unidades de la conformación original. La raiz cuadrada de la varianza es la desviación standard (S).

$$S = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}}$$

En el ejemplo de precipitación, la desviación standard para el lugar A es 3.38 cm y para el lugar B 5.25 cm.

Notación para muestras y poblaciones

Una breve explicación sobre notación se necesita aquí debido a que en algunos casos la desviación standard se representa con la letra griega sigma ($\sigma$). Tambien, la media aritmética a veces se representa por mu ($\mu$).

La convención es que las letras griegas se refieren a los promedios y desviaciones standard de la población y las letras latinas se refieren a estas medidas de una muestra.

| | Población | Muestra |
|---|---|---|
| Media | $\mu$ | $\bar{X}$ |
| Varianza | $\sigma^2$ | $S^2$ |
| desviación standard. | $\sigma$ | $S$ |

Puntaje Standard

Antes de dejar la desviación standard, describiremos una medida aliada, el puntaje stardard · generalmente es imposible compararar números que estan medidos en unidades diferentes como el puntaje standard lo hace posible en ciertos casos. En 1963, la precipitación anual del lugar B era de 62 cm, una cifra 7 cm más alta que el valor medio de 55 cm. Supongamos tambien que cada año el número de días en que la temperatura cayó bajo 0ºC en el lugar B fue registrado, y el promedio fue de 35 días y en 1963 fueron de 41 días. En 1963, entonces, hubo 7 cm. más lluvia que el promedio y 6 días con helada sobre el promedio.

¿Cual fue el acontecimiento mas raro?. Es muy difícil contestar tales preguntas sin recurrir a la desviación standard que supondremos es de 2.5 para los días con helada, marando número de precipitación, 7 cm de lluvia sobre el promedio es lo mismo que decir 7/3.38 o 2.06 desviaciones standard sobre la media.

De la misma manera, 6 días más que comunmente es lo mismo que 6/2.5 o 2.4 desviación standard sobre la media. Los números 2.06 y 2.4 son porcentaje standard (reprentados por el símbolo Z ), que son números sin unidades que se pueden comparar directamente, la fórmula es:

$$Z_i = \frac{X_i - \bar{X}}{S}$$

donde $Z_i$ es el puntaje standard para Xi , Debido a que el número de días con helada esta a más desviaciones standard sobre su media (tiene un puntaje standard más alto) que la precipitación para 1963, podemos concluír que 1963 se destaca más por su alto número de días fríos que por su precipitación.

Analisis del vecino más cercano

Además de las medidas estadísticas básicas que se mencionan arriba, existen muchas otras medidas, para describir en un sólo número los conceptos que requerirían muchas palabras o una expresión verbal . Una de estas medidas, que primero fue desarrollada en la ecología vegetal, pero que fue empleada posteriormente en muchos estudios geográficos, es el índice del vecino más cercano.

El problema consiste en cómo describir la distribución de un grupo de puntos. Los puntos pueden indicar la ubicación de cierta planta en una ladera a la distribución de estaciones meteorológicas en una provincia, o la distribución de ciudades en un país. ¿Como es posible medir la distribución de los puntos?. A un extremo está el caso de los puntos muy concentrados [illegible] y al otro los puntos, los puntos podrían tener el máximo distanciamiento formando un trazado reticular. Es claro que la mayoría de las distribuciones caerán entre estos dos extremos (Fig.6)

Sería útil poder disponer de un índice que pudiera evaluar dos distribuciones distintas de puntos, y juzgar cual era la más dispersa, cuál la más agrupada, y por cuanto. Esto se logra por la medida del vecino más cercano.

El indice se calcula midiendo primero la distancia desde cada punto al punto más cercano (el vecino más cercano). Se suman estas distancias, y se dividen por el número de puntos para dar la distancia media $\bar{D}$ . Esta cifra entonces se pone en la fórmula.

$$R_n = 2\,\bar{D}\sqrt{N/A} \qquad R_n = 2\,\bar{D}\sqrt{N/A}$$

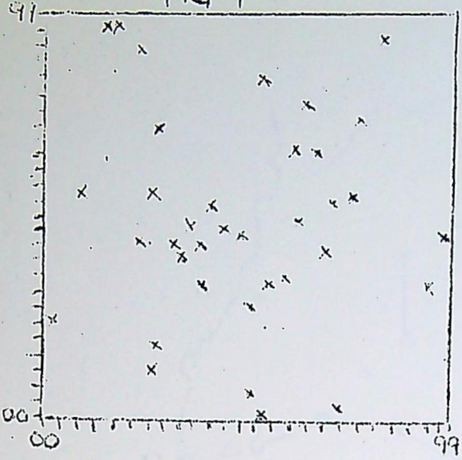
en que $R_n$ es la medida del vecino más cercano, $\bar{D}$ es la distancia media, N es el número de puntos y A es el area de estudio. La medida A debe estar en las mismas unidades que D pero al cuadrado.

La medida del vecino más cercano varía entre 0 y 2,15. Si ocurre la cantidad máxima de agrupamiento (en que todos los puntos están ubicados en la misma ubicación) entonces $R_n$ es 0. Para los puntos que estan distribuidos con un distanciamiento máximo según un trazado regular, $R_n$ es 2,15. Entre estos dos extremos, una distribución al azar que no es ni concentrada ni regular da un $R_n$ de 1.

Ej: Caso de distribución de tiendas al cambiar el tiempo.

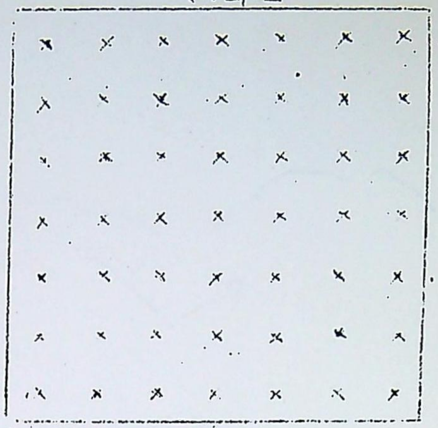
Tb.

FIG 1




MUESTRA ALEATORIA

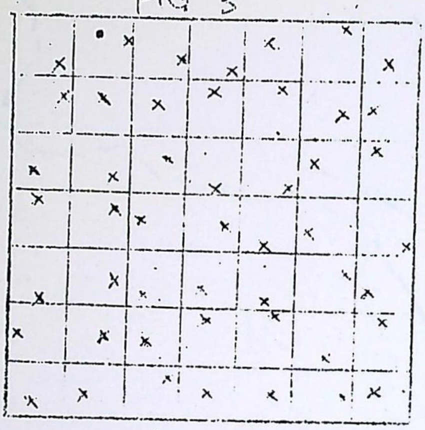
FIG 2



MUESTRA SISTEMÁTICA

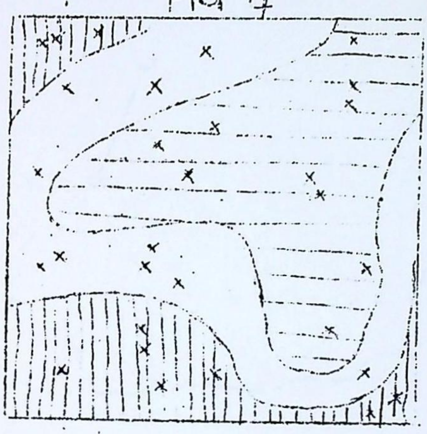
No muy recomendable.

FIG 3



Una de las mejores muestras.

FIG 4



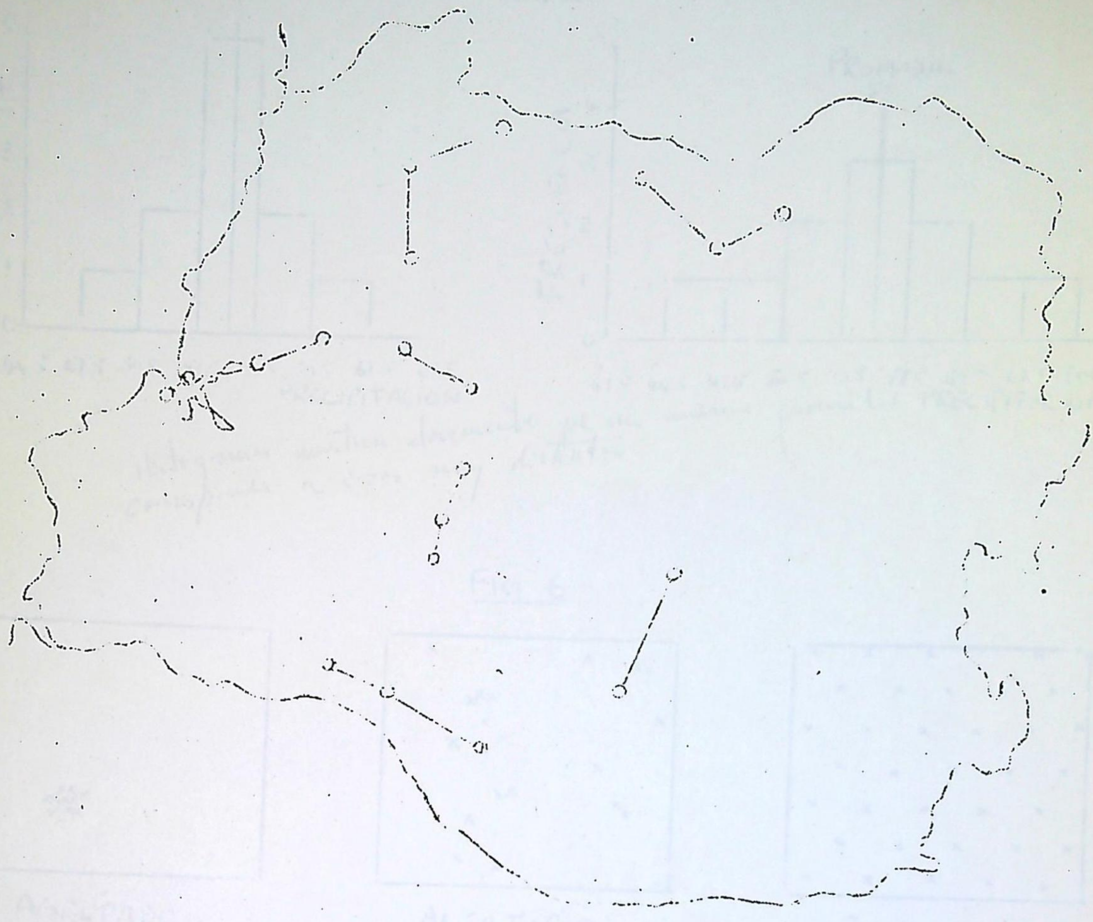
TIPOS DE MUESTRA ALEATORIA ESTRATIFICADA

MESETAS

LADERAS ABRUPTAS

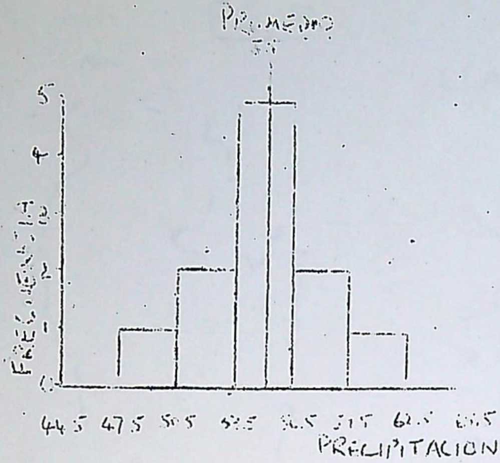
FONDO DE VALLE

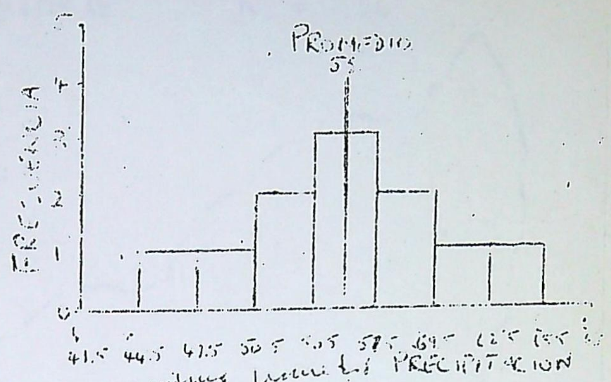
FIG 7 VALDIVIA $R_n = 0{,}86$



$$R_n = 2 \times 1{,}29 \sqrt{20/181} = 0{,}86$$

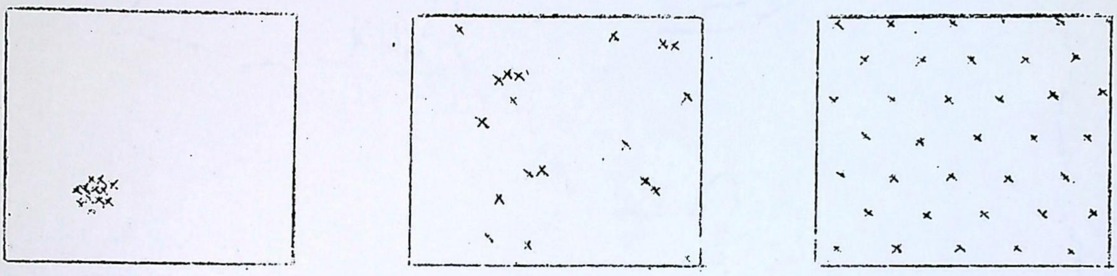
FIG. 5




Histogramas muestran claramente que un mismo promedio corresponde a casos muy distintos.

FIG 6



AGRUPADO $R_n = 0$

ALEATORIO $R_n = 1$

REGULAR $R_n = 2,15$

FIG. 9

LLANQUIHUE $R_n = 0,66$