Universidad Autónoma del Estado de México

Facultad de Turismo y Gastronomía

Licenciatura en Gastronomía

Unidad de aprendizaje: Estadística Aplicada a la gastronomía .
Pruebas de correlación
Periodo: Séptimo.

Autor: Claudia Villegas Marín

# Objetivo del material didáctico

identificar los requisitos para la elección ya aplicación de pruebas paramétricas y no paramétricas de correlación aplicadas al análisis de datos a través del SPSS en casos y ejercicios prácticos de investigación

# Pruebas correlación paraméticas y no paramétricas

- Coeficiente de correlación producto momento de Pearson
- Coeficiente de rangos ordenados de Spearman
- Coeficiente de contingencia
- Coeficiente de correlación Phi

# CORRELACIÓN

*Es frecuente que en la investigación o en el ejercicio de la profesión interese estudiar en una misma población los valores de dos variables estadísticamente distintas, con el fin de ver si existe alguna relación entre ellas, es decir, si los cambios en una se asocian a cambios en los valores de la otra.*

# Ejemplos de asociación entre variables:

- Inteligencia con aprendizaje
- Altura con peso (Kilogramos)
- Índice de pobreza con índice de desnutrición
- Nivel de Agresión con grado de frustración
- Índice de desempleo con índice de robos.

Dos variables se correlacionan cuando ambas covarian

Variación conjunta en diagrama de dispersión

La magnitud de la covariación se dispone en índices o coeficientes de correlación

La covariación puede ser positiva o negativa

Si es positiva la correlación es directa

Si es negativa la correlación es inversa

# *VALORES DEL ÍNDICE DE CORRELACIÓN*

- El índice de correlación fluctúa entre:

**1 y – 1**

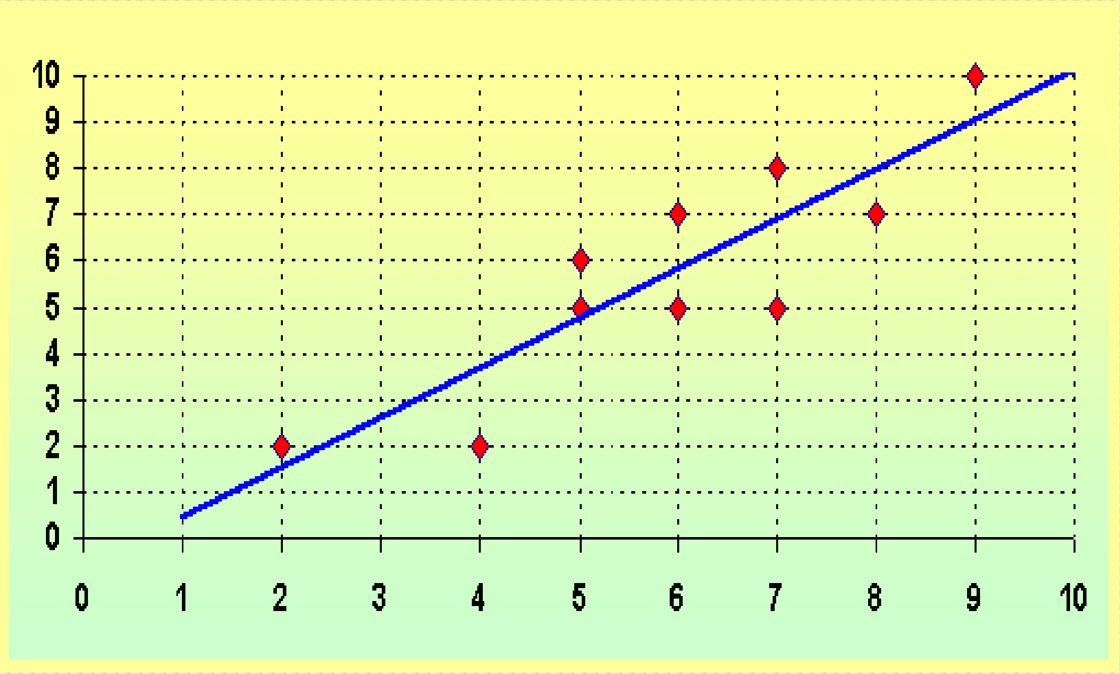
**1 = correlación directa perfecta**

**-1= correlación inversa perfecta**

**0 = no hay correlación**

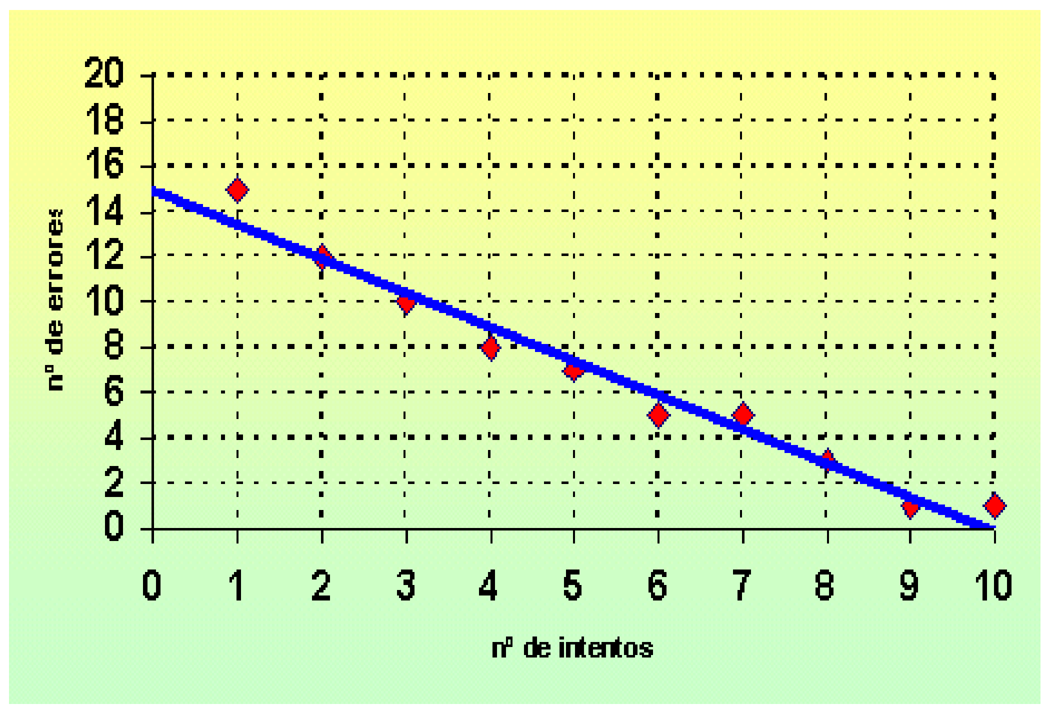
# CORRELACIÓN DIRECTA PERFECTA

- Diagrama de dispersión



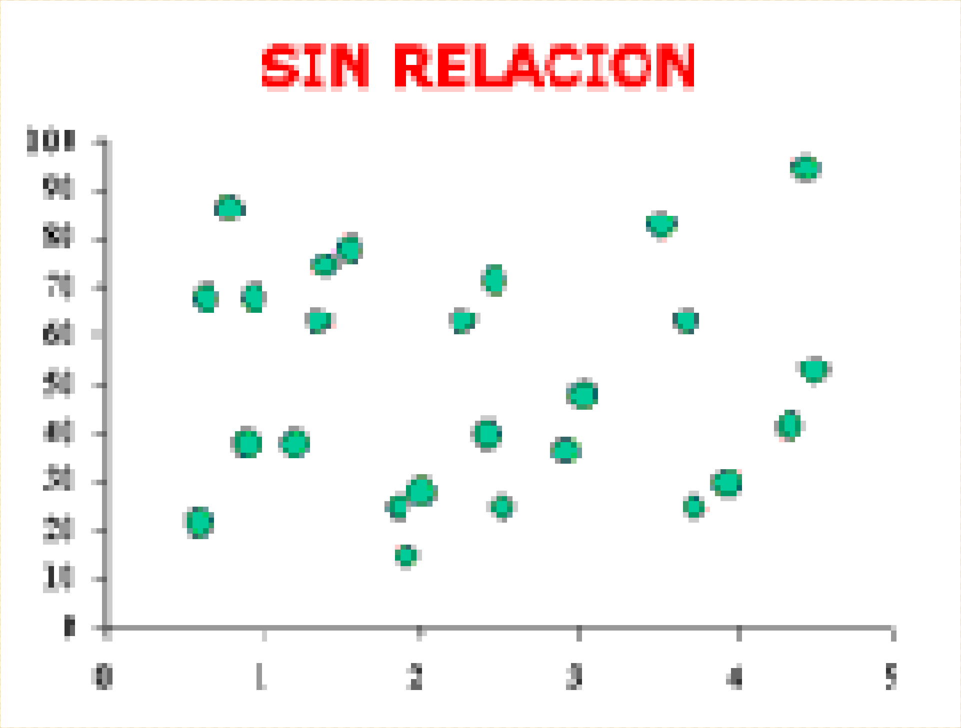
# CORRELACIÓN INVERSA PERFeCTA

- Diagrama de dispersión



# NO HAY CORRELACIÓN

- *Diagrama de dispersión*



# Coeficiente de correlación producto momento de Pearson

# Requisitos

- *Nivel de medición al menos intervalar.*
- *Las variables distribuidas normalmente*
- *Relación lineal entre las dos variables en el diagrama de dispersión*
- *Muestreo probabilístico*

# Hipótesis

- **Hipótesis no direccionadas**
- H0: $r=0$
- Hi: $r \neq 0$
- **Hipótesis direccionadas**
- Unilateral derecha
- H0: $r=0$
- Hi: $r>0$
- Unilateral Izquierda
- H0: $r=0$
- Hi: $r<0$

# CORRELACIÓN "PEARSON"

- Fórmula para calcular el índice de correlación "Pearson"

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

# CORRELACIÓN "PEARSON"

- *Posibles resultados:*
- *Valor de la correlación tendiente o cercana a **"1"** será una correlación alta, con relación directa entre las dos vairables.*
- *Valor de la correlación tendiente o cercana a **"-1"** será una correlación alta, con relación inversa entre las dos vairables.*

# CORRELACIÓN “PEARSON”

- *Si por el contrario el índice de correlación resultante se aleja de 1 acercándose a* ***“0”*** *será una correlación baja o nula.*

- *Sin embargo para decidir si hay o no correlación hay una regla de decisión*

# CORRELACIÓN "PEARSON"

- *Para tomar la decisión se de debe hacer lo siguiente:*
- *Comparar el valor de "r" (calculado) con el valor de "r" (de tablas).*

$gl = n \cdot$

- *Para conocer el valor de "r" de tablas se fija el nivel de significancia y los grados de libertad.*
  - gl = n - 2

# Decisión de hipótesis

- Si se trabaja con el valor calculado y se compara con el valor teórico
  - Si $r_o \geq r_T$ entonces se rechaza H0
  - Si $r_o < r_T$ entonces se acepta H0
- Si se trabaja en SPSS con probabilidad exacta
  - Si p ≤ α entonces se rechaza H0
  - Si p > α entonces se acepta H0

# CORRELACIÓN DE "PEARSON"

- Interpretación convencional:

| CORRELACIÓN | VALOR O RANGO |
|---|---|
| **Perfecta** | $\lvert r \rvert = 1$ |
| **Excelente** | $0.9 <= \lvert r \rvert < 1$ |
| **Buena** | $0.8 <= \lvert r \rvert < 0.9$ |
| **Regular** | $0.5 <= \lvert r \rvert <0.8$ |
| **Baja** | $\lvert r \rvert< 0.5$ |

# Correlación y causalidad

- No se puede interpretar una correlación como causa efecto, o que una variable pueda determinar la aparición de la otra, solo puede interpretarse la covariación, cuando una cambia la otra también cambia

# Coeficiente de correlación de rangos ordenados de Spearman

# CORRELACIÓN DE SPEARMAN

- Propósito de la Prueba: Determina el grado de asociación de dos variables.
- Características para ser empleada:
  - Dos variables en nivel de medición ordinal
- Alternativa no paramétrica más viable cuando no se cumplen criterios para aplicar correlación de Pearson

- Requisitos:
- 1. Una correlación lineal, relaciones lineales entre X y Y
- 2. Las variables X y Y deben ordenarse o colocarse por rangos
- 3. Los miembros de la muestra deben haber sido preferentemente extraídos aletoriamente de una población mayor.

# Hipótesis

- **Hipótesis no direccionadas**
- H0: $r_s$=0
- Hi: $r_s \neq 0$
- **Hipótesis direccionadas**
- Unilateral derecha
- H0: $r_s$=0
- Hi: $r_s$>0
- Unilateral Izquierda
- H0: $r_s$=0
- Hi: $r_s$<0

# Fórmula:

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N}$$

- donde:

rs = el coeficiente de correlación por rangos ordenados

d = la diferencia de rangos entre las variables X y Y

N = número total de casos

# Método:

Colocar por rangos a los entrevistados sobre las variables X y Y. Se asigna el rango 1 al de mayor puntuación, el 2 al que le sigue, el 3 al siguiente y así sucesivamente. En caso de haber puntuaciones iguales, se obtiene un promedio del rango

Encontrar la diferencia entre los rangos X y Y (d), elevar al cuadrado la diferencia ($d^2$) y sumar estos cuadrados ($\Sigma d^2$)

Sustituir el resultado del paso 2 en la fórmula para el coeficiente de correlación por rangos ordenados

# Criterios para la toma de decisión de hipótesis

- Si se trabaja con el valor calculado y se compara con el valor teórico
  - Si $r_{so} \geq r_{sT}$ entonces se rechaza H0
  - Si $r_{so} < r_{sT}$ entonces se acepta H0
- Si se trabaja en SPSS con probabilidad exacta
  - Si $p \leq \alpha$ entonces se rechaza H0
  - Si $p > \alpha$ entonces se acepta H0

# Coeficiente de Contingencia

# Requisitos Coeficiente de Contingencia

- Se aplica cuando se quieren relacionar dos variables categóricas polítomicas
- Se trabaja con tablas de contingencia de más de 2x2
- El nivel de medición puede ser nominal y ordinal
- Basado en la distribución $X^2$
- Frecuencias esperadas mayores a 5
- Nota: las hipótesis para esta prueba son no direccionadas

# Esquema

**Tabla de contingencia TRANS2 * vive en Barcelona**

| | | | vive en Barcelona | | Total |
|---|---|---|---|---|---|
| | | | si | no | |
| TRANS2 | Metro | Recuento | 41 | 12 | 53 |
| | | Frecuencia esperada | 32,5 | 20,5 | 53,0 |
| | Bus | Recuento | 15 | 14 | 29 |
| | | Frecuencia esperada | 17,8 | 11,2 | 29,0 |
| | Tren | Recuento | 2 | 11 | 13 |
| | | Frecuencia esperada | 8,0 | 5,0 | 13,0 |
| | Coche | Recuento | 8 | 3 | 11 |
| | | Frecuencia esperada | 6,8 | 4,2 | 11,0 |
| | Otros | Recuento | 4 | 4 | 8 |
| | | Frecuencia esperada | 4,9 | 3,1 | 8,0 |
| Total | | Recuento | 70 | 44 | 114 |
| | | Frecuencia esperada | 70,0 | 44,0 | 114,0 |

# Formulas

- Formula para probar la significancia

$$\chi^2_{(F-1)(C-1)} = \sum_{i=1}^{F}\sum_{j=1}^{C}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

- Formula para calcular el coeficiente de contingencia

$$C=\sqrt{\frac{\chi^2}{(\chi^2+n)}}$$

# Criterios de decisión de hipótesis

- Si se trabaja con el valor calculado y se compara con el valor teórico
  - Si ${X_0}^2 \geq {X_t}^2$ entonces se rechaza H0
  - Si${X_0}^2 < {X_t}^2$ entonces se acepta H0
- Si se trabaja en SPSS con probabilidad exacta
  - Si p ≤ α entonces se rechaza H0
  - Si p > α entonces se acepta H0

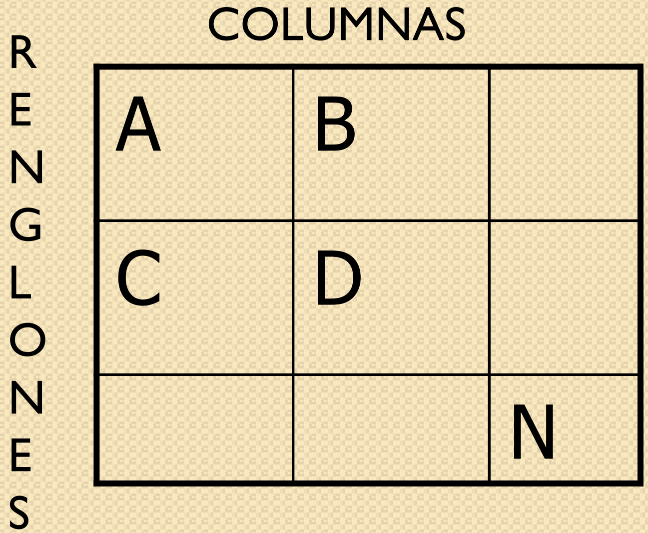
# Coeficiente de Correlación Phi

# Requisitos Coeficiente de correlación Phi

- Cuando se quiere asociar dos variables categóricas dicotómicas
- Nivel de medición nominal
- Basado en la distribución $X^2$
- Frecuencias observadas mayores a 5

- Nota: las hipótesis para esta prueba son no direccionadas

- Estructura



# Formula

Para calcular el coeficiente:

$$r_\phi = \frac{|AD - BC|}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

Para probar significancia:

$$x^2 = \frac{N(|AD - BC| - N/2)^2}{(A+B)(C+D)(A+C)(B+D)}$$

# Criterios de decisión de hipótesis

- Si se trabaja con el valor calculado y se compara con el valor teórico
  - Si ${X_0}^2 \geq {X_t}^2$ entonces se rechaza H0
  - Si${X_0}^2 < {X_t}^2$ entonces se acepta H0
- Si se trabaja en SPSS con probabilidad exacta
  - Si p ≤ α entonces se rechaza H0
  - Si p > α entonces se acepta H0

# Regresión lineal

# Regresión lineal

- Permite determinar si dos variables están relacionadas linealmente
- Permite razonar que si se conoce el comportamiento de una variable se puede predecir el comportamiento de la otra, con la cual esta relacionada la primera
- Variable independiente=variable predictora= x
- Variable dependiente=variable criterio= y

# Diferencias de regresión lineal y correlación

- Una regresión corresponde a un modelo predictivo
- Una correlación corresponde a una prueba de hipótesis

# Regresión lineal simple

- El diagrama de dispersión permite detectar si las variables pueden ajustarse más o menos bien a una línea recta
- $Y_i = B_0 + B_1 * X_1$
- Pendiente
- $B_1 = \frac{((n*\sum XY)-(\sum X*\sum Y))}{((n*\sum X^2)-(\sum X)^2)}$
- Intercepto
- $B_0 = (\sum Y - (B_1 * \sum X))/n$

- El análisis de regresión lineal es el cuadrado del coeficiente de correlación de Pearson

- $R^2 = \dfrac{Suma\ de\ cuadrados\ de\ los\ pronosticos}{Suma\ de\ cuadrados\ total}$

- Se interpreta como la proporción de datos que es posible predecir de la variable predictora a la variable criterio

# Supuestos de la regresión lineal

- Considerar que sus valores son fijos, determinados por el investigador
- Considerar que se trata de variables aleatorias, al igual que la variable dependiente o criterio
- Existe linealidad parcial entre Y (criterio) y cada predictora $X_i$
- Para cada observación de Y los errores de pronostico tienen distribución normal, con media 0

- Los errores son independientes entre sí, de las predictoras y de los pronósticos
- La varianza de los errores es la misma para cualquier valor de las X (es uniforme a lo largo de los valores pronosticados
- Las variables predictoras no están correlacionadas entre sí (ausencia de colinealidad)

# Referencias consultadas


- Campbell, D., y J. Stanley (1993 ed.). Diseños experimentales y cuasiexperimentales
- en la investigación social, Amorrortu, Buenos Aires.
- Dixon, W. y F. Massey (1969). Introduction to statistical analysis, McGraw-Hill,
- Nueva York.
- Estrada, A., C. Batanero y J. Fortuny (2004). Un estudio sobre conocimientos de
- estadística elemental de profesores en formación, Educación Matemática, vol.
- 16, núm. 1, pp. 89-111.
- Nadelsticher, A. (1983). Técnicas para la construcción de cuestionarios de actitudes
- y opción múltiple, Instituto Nacional de Ciencias Penales, México.
- Naiman, A., R. Rosenfeld y G. Zirkel (1996). Understanding statistics, McGrawHill,
- Nueva York.
- Ruiz, H. C., E. Castillo, A. Lupercio, I. Galicia y S. Juárez (2006). "Alfabetización
- ambiental en primaria y secundaria", Ciencia y Desarrollo, vol. 32, núm. 200,
- pp. 60-66.
- Silva, A. (1998). Investigación asistida por computadoras en las ciencias sociales y
- de la salud, Facultad de Estudios Superiores Iztacala-Universidad Nacional
- Autónoma de México, México.