

---

# Metodología II: Análisis de Datos

---

Prof. Reinaldo Mayol  
Derecho

---

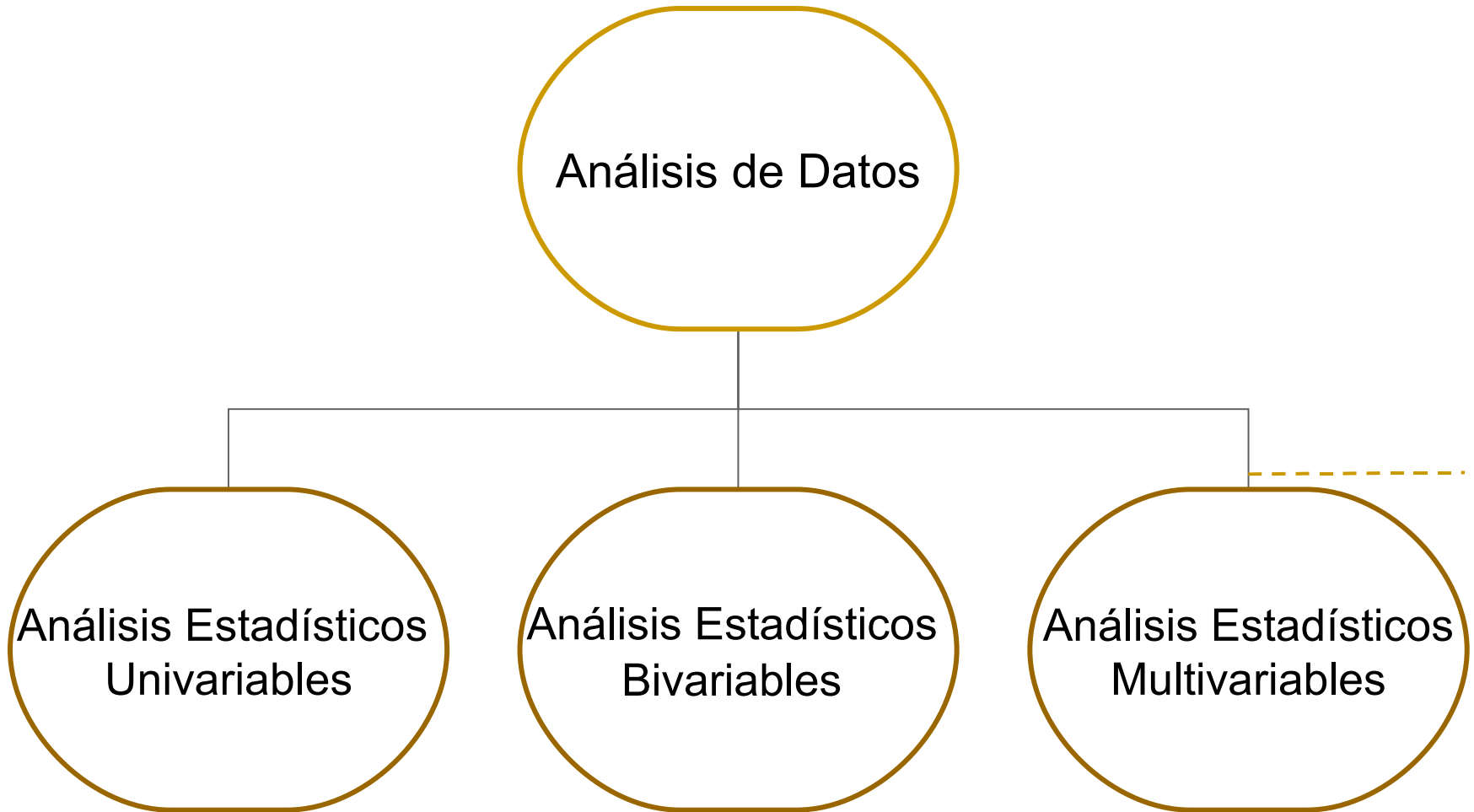
# Por donde vamos?

- Luego de obtenidos los datos, el siguiente paso es realizar el análisis de los mismos.
  - Aunque ha sido presentado en este curso de manera secuencial es importante que antes de obtener los datos se piense que tipo de análisis se realizará y que información es importante recolectar, incluyendo el formato de la misma
-

# Depuración de la Información

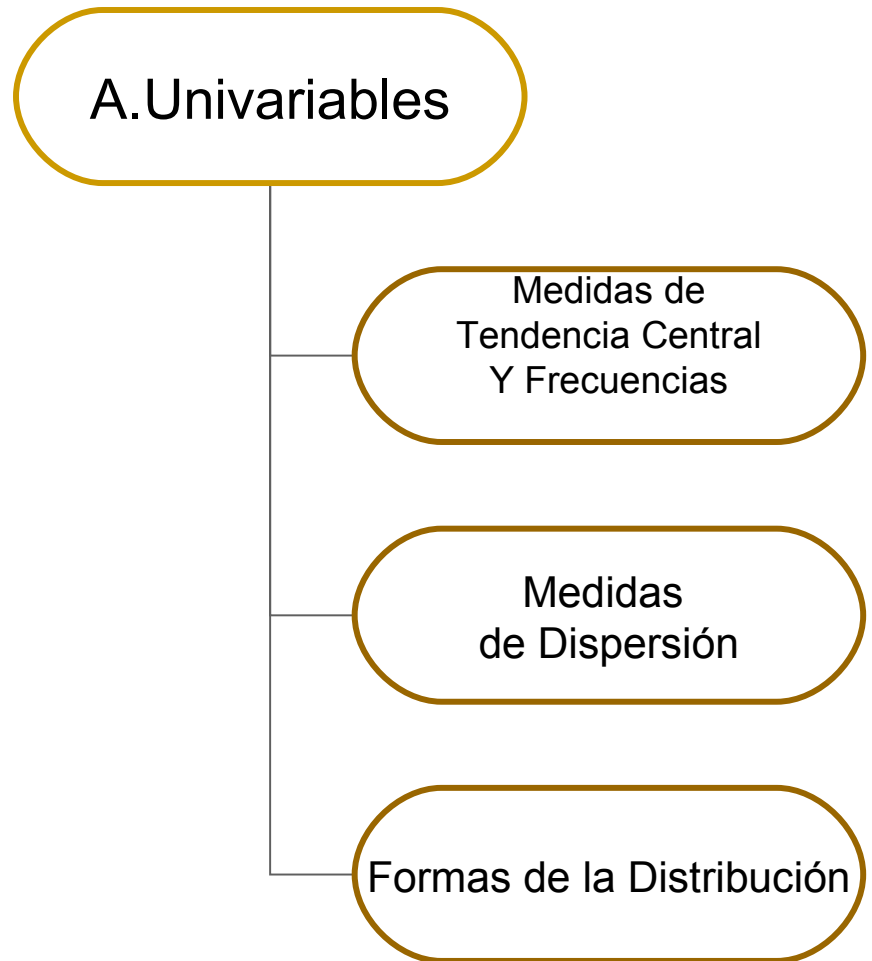
- El objetivo es encontrar anomalías en los datos recolectados.
  - Ej. Si una variable contiene el sexo de los participantes en un experimento codificado como 1: Hombres 0: Mujeres y aparecen otros valores ( 3,4,etc) significa que ha ocurrido un error.
  - Si no es posible corregir esos valores debe asumirse la pérdida de los mismos.( missing values)
  - Igual condición puede suceder, por ejemplo si de una variable no se tienen todos los valores para cada caso.

# Tipos de Análisis



# Análisis Univariados ( exploratorios)

- Suelen ser la antesala para la realización de otros análisis.
- Se concentran en una sola variable y buscan encontrar y representar características propias de las mismas.



---

# Medidas de Tendencia Central

- Media
- Mediana
- Moda



# Medidas de Frecuencia

- Frecuencia Absoluta
- Frecuencia Relativa
- Frecuencia Acumulada

Calificación $x_i$	Frecuencia $f_i$	Frecuencia Acumulada	Frecuencia Relativa	Frec. Rel. Acumulada
2.4	1	1	0.05	0.05
2.7	1	2	0.05	0.10
3.6	1	3	0.05	0.15
5.5	2	5	0.10	0.25
5.7	1	6	0.05	0.3
6.0	1	7	0.05	0.35
6.3	1	8	0.05	0.4
6.4	1	9	0.05	0.45
6.5	3	12	0.15	0.6
7.0	2	14	0.10	0.7
7.2	1	15	0.05	0.75
7.4	1	16	0.05	0.8
7.6	1	17	0.05	0.85
7.8	1	18	0.05	0.9
8.0	1	19	0.05	0.95
8.6	1	20	0.05	1.0

---

# Medidas de Dispersión

- Rango o Recorrido
  - Desviación Típica
  - Varianza
-



---

# Medidas de la forma de la distribución

- Asimetría
- Curtosis



# Ejemplo, Variable Sueldo

Individuo	Sueldo
1	100
2	200
3	150
4	400
5	1000
6	1600
7	400
8	1000
9	---

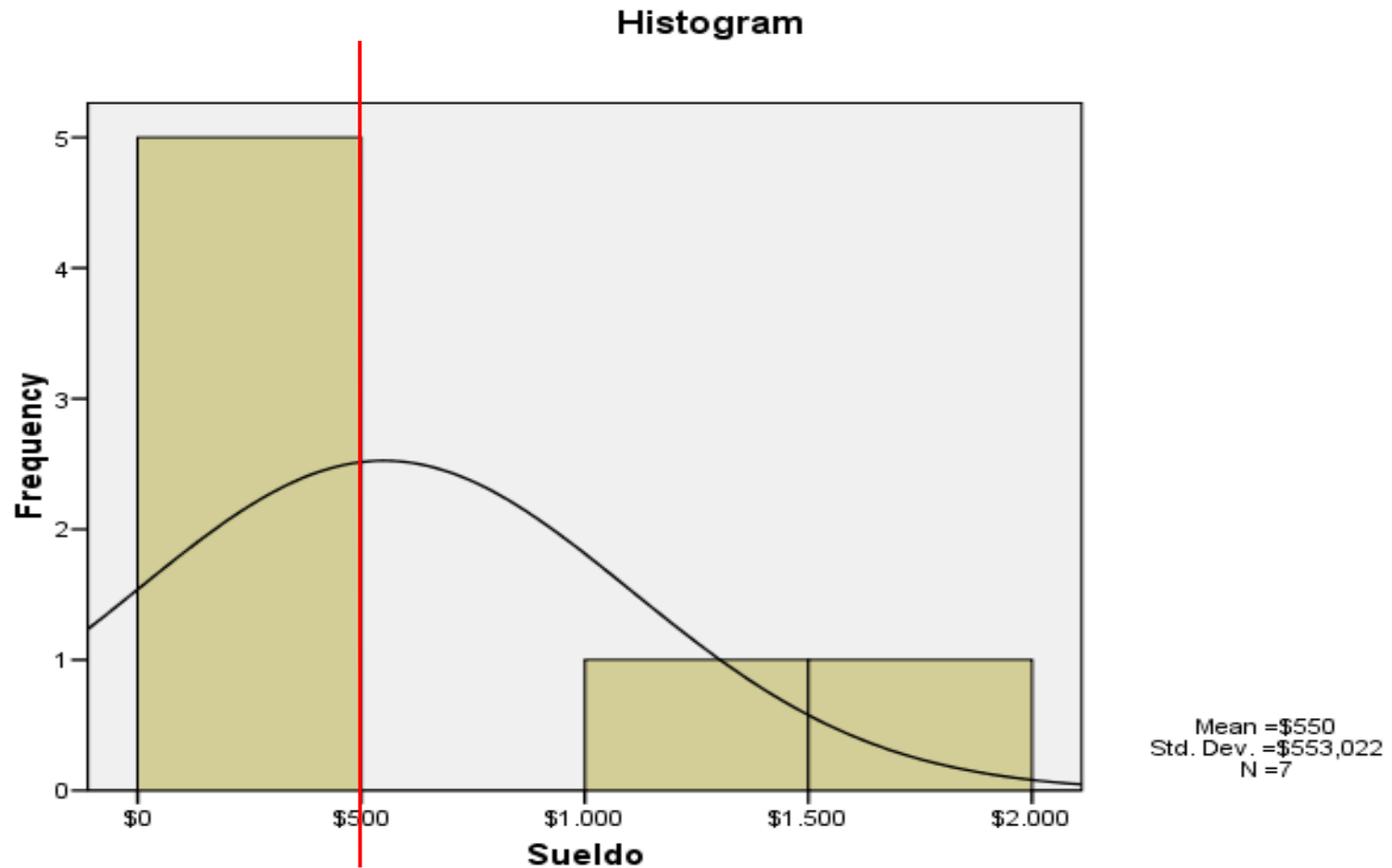
## Statistics

Sueldo		
N	Valid	7
	Missing	1
Mean		\$550.00
Median		\$400.00
Mode		\$400
Std. Deviation		\$553.022
Variance		305833,3
Skewness		1,440
Std. Error of Skewness		,794
Kurtosis		1,281
Std. Error of Kurtosis		1,587
Range		\$1,500
Minimum		\$100
Maximum		\$1,600
Sum		\$3,850

## Sueldo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	\$100	1	12,5	14,3	14,3
	\$150	1	12,5	14,3	28,6
	\$200	1	12,5	14,3	42,9
	\$400	2	25,0	28,6	71,4
	\$1,000	1	12,5	14,3	85,7
	\$1,600	1	12,5	14,3	100,0
	Total	7	87,5	100,0	
Missing	\$0	1	12,5		
Total		8	100,0		

# Histograma Variable Sueldo



# Ejemplo 2. Variable Temperatura

Día	Temperatura
1	30
2	31
3	30
4	32
5	29
6	30
7	31
8	31
9	32

## Statistics

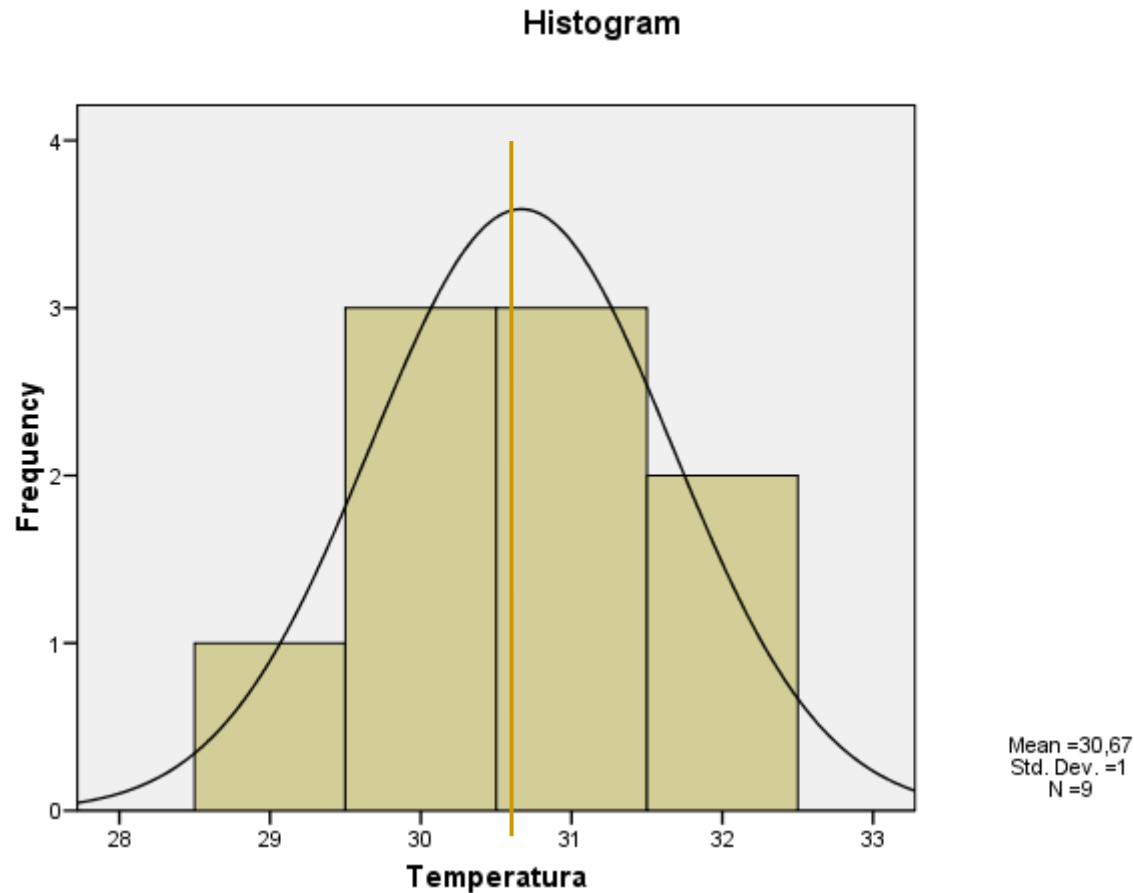
Temperatura		
N	Valid	9
	Missing	0
Mean		30,67
Median		31,00
Mode		30 <sup>a</sup>
Std. Deviation		1,000
Variance		1,000
Skewness		-,107
Std. Error of Skewness		,717
Kurtosis		-,643
Std. Error of Kurtosis		1,400
Range		3
Minimum		29
Maximum		32
Sum		276

a. Multiple modes exist. The smallest value is shown

## Temperatura

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	29	1	11,1	11,1	11,1
	30	3	33,3	33,3	44,4
	31	3	33,3	33,3	77,8
	32	2	22,2	22,2	100,0
	Total	9	100,0	100,0	

# Histograma Variable Temperatura



# Ejemplo 3 Variable humedad relativa

Día	Humedad
1	90
2	91
3	91
4	92
5	96
6	96
7	68
8	50
9	90

## Statistics

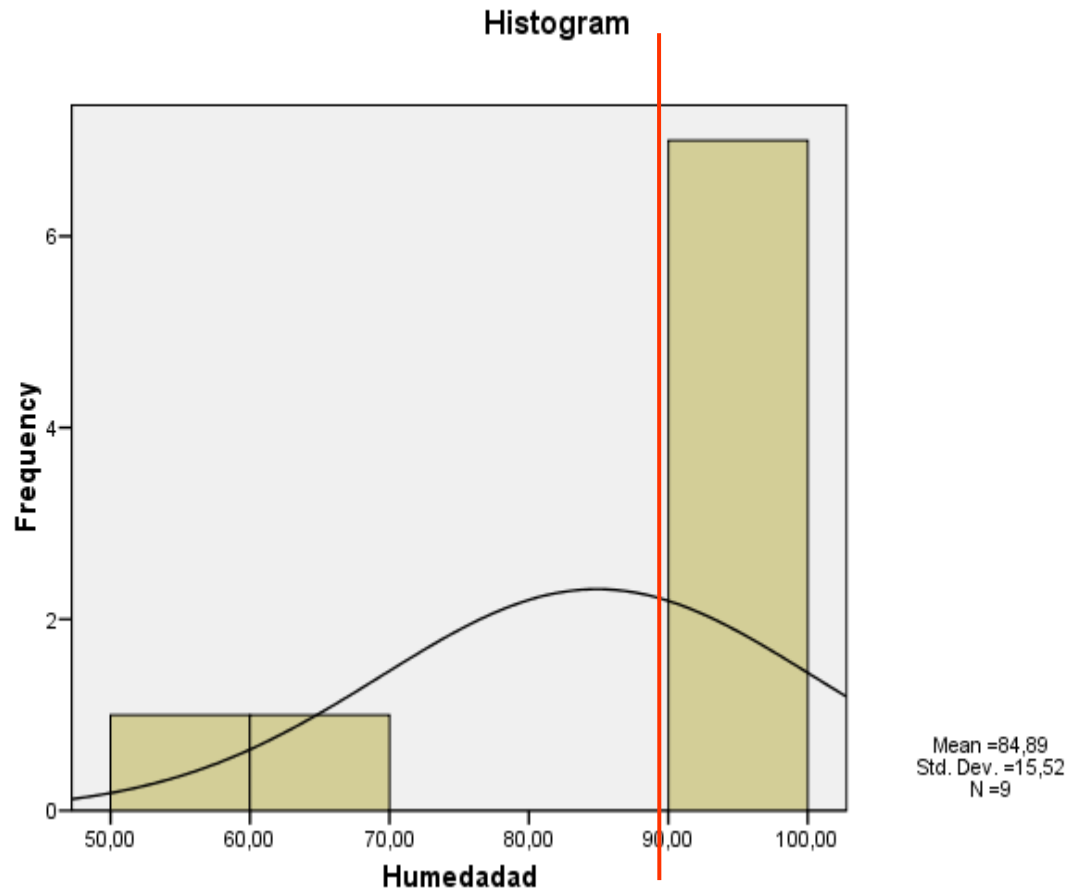
Humedadad		
N	Valid	9
	Missing	0
Mean		84,8889
Median		91,0000
Mode		90,00 <sup>a</sup>
Std. Deviation		15,51970
Variance		240,861
Skewness		-1,868
Std. Error of Skewness		,717
Kurtosis		2,817
Std. Error of Kurtosis		1,400
Range		46,00
Minimum		50,00
Maximum		96,00
Sum		764,00

a. Multiple modes exist. The smallest value is shown

## Humedadad

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	50,00	1	11,1	11,1	11,1
	68,00	1	11,1	11,1	22,2
	90,00	2	22,2	22,2	44,4
	91,00	2	22,2	22,2	66,7
	92,00	1	11,1	11,1	77,8
	96,00	2	22,2	22,2	100,0
	Total	9	100,0	100,0	

# Histograma, Variable Humedad



---

# Análisis Bivariante

- Luego de realizar los análisis exploratorios corresponde realizar los análisis bivariantes con dos fines:
    - Fines Descriptivos: Describir al conjunto de la población observada
    - Fines Explicativos: Analizar la posible existencia de relaciones causales entre dos variables ( dependiente e independiente)
-



# Tablas de Contingencia

- Por **tablas de contingencia** se entiende aquellas de doble entrada, donde se realiza una clasificación de la muestra de acuerdo a un doble criterio de clasificación
- Cada valor de la tabla es el cruce de una fila y una columna.( variable dependiente vs. Independiente)
- Una variación porcentual es significativa si es mayor al 5 % ( dependiendo del error)

Humedad \* Lluvia Crosstabulation

Count

		Lluvia		Total
		0	1	
Humedad	70	1	0	1
	80	3	0	3
	84	0	2	2
	85	0	1	1
	95	0	2	2
Total		4	5	9

- 1: Lluvia
- 0: No llueve

- **Si existen mayores niveles de humedad (>84) llueve**

---

# Covariación

- Dos variables se encuentran relacionadas si sus parámetros varían conjuntamente.
  - Para medir esta relación se utilizan métodos *estadísticos de contingencia*:
    - *Variables Nominales: Chi Cuadrado, C Pearson, etc*
    - *Variables Ordinales: Rho de Spearman, Tau-A, etc.*
    - *Variables de Intervalo: Coeficiente de Correlación producto-momento de Pearson*
-

---

# Análisis Multivariables

- Buscan relaciones entre más de dos variables de manera simultánea.
  - En las ciencias sociales el análisis bivariante es por regla general insuficiente o simplista
  - Las técnicas de AMV se pueden agrupar en dos grupos:
    - Técnicas de Dependencia
    - Técnicas de Interdependencia
-

---

Fin

---