

Tarea 07

Nombres y código:

Ruiz Ortiz Cesar Adrián 208020817

Padilla Martin Juan Pablo 217294261

Álvarez Gutiérrez David Alejandro 217294016

Sección: D01

Materia: MINERIA DE DATOS

Profesor: ISRAEL ROMAN GODINEZ

CARRERA: INGENIERIA EN INFORMATICA

CICLO ESCOLAR: 2020 A



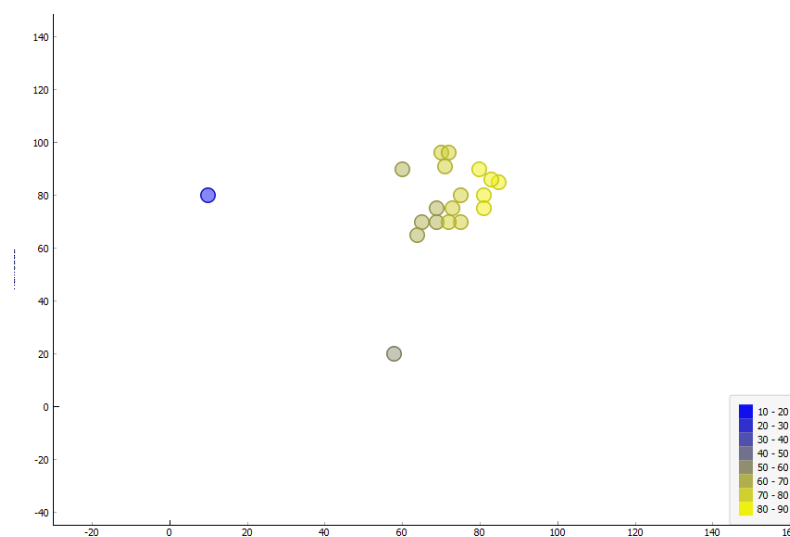
Descripción del problema

Con base en la tabla 1 determine, de manera visual (usando Orange ML) y analítica, la correlación que existe entre los atributos. Reporte en un documento las imágenes de los análisis visuales y describa sus observaciones. En el caso del análisis analítico de los atributos, realice el desarrollo de los cálculos en papel y con calculadora.

Temperatura vs Humedad: no existe correlación el índice es muy cercano al cero

	temperatura	humedad		
	15.895	85	85	8
	10.895	80	90	13
	13.895	83	86	9
	-0.105	69	75	-2
	0.895	70	96	19
	11.895	81	80	3
	-4.105	65	70	-7
	-5.105	64	65	-12
	2.895	72	95	18
	-9.105	60	90	13
	-0.105	69	70	-7
	5.895	75	80	3
	-59.105	10	80	3
	5.895	75	70	-7
	2.895	72	70	-7
	11.895	81	75	-2
	-11.105	58	20	-57
	3.895	73	75	-2
	1.895	71	91	14
suma	1313	1463	sumatoria	878
media	69.1052632	77	n * desv a * desv b	5110.6094
desv std	16.1516932	16.653328	coeficiente de correlacion de pearson	0.1717995

De otra forma podemos validar si existe con la grafica de dispercion



Pronóstico vs Viento: no existe correlación ya que el índice de correlación es cercano al 0

Handwritten calculations on graph paper showing the Chi-square test for independence between Wind (Viento) and Forecast (Pronóstico).

Contingency Table:

Pronóstico \ Viento	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	5	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Chi-square calculation:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where O is observed frequency and E is expected frequency.

Expected frequencies (E) for each cell are calculated as:

$$E_{ij} = \frac{\text{Row Total}_i \times \text{Column Total}_j}{\text{Grand Total}}$$

Grand Total = 20

Row Totals: 4, 5, 3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

Column Totals: 4, 3, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

Expected frequencies (E) for each cell:

Pronóstico \ Viento	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.8	0.6	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.4	0.6	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.2	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.1	0.15	0.1	0.05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

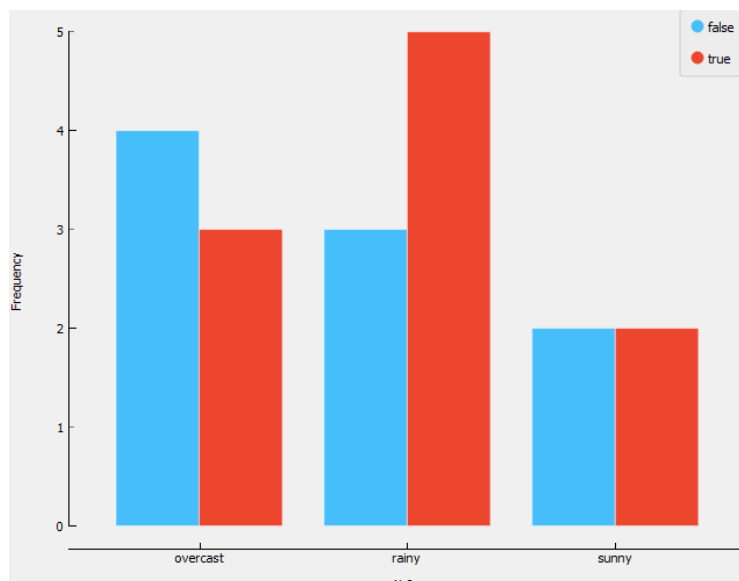
Chi-square value: $\chi^2 = 1.3$

Degrees of freedom: $df = (20 - 1) = 19$

Significance level: $\alpha = 0.05$

Conclusion: Since $\chi^2 < \chi^2_{critical}$, we fail to reject the null hypothesis. There is no significant correlation between Wind and Forecast.

La forma de la distribución de las columnas no sigue algún patrón definido por lo cual podemos validar que no hay alguna correlación



Pronóstico vs Jugar: no existe correlacion en estas tuplas debido a que el coeficiente se acerca mas a 0 que a 1

pronostico	jugar						
sunny	no	1	2.52631579	-1.52631579	2.32963989	0.92214912	
rainy	no	5	5.05263158	-0.05263158	0.00277008	0.00054825	
overcast	yes	6	4.42105263	1.57894737	2.49307479	0.56390977	
sunny	no	3	1.47368421	1.52631579	2.32963989	1.58082707	
rainy	yes	3	2.94736842	0.05263158	0.00277008	0.00093985	
rainy	yes	1	2.57894737	-1.57894737	2.49307479	0.96670247	
rainy	no						
overcast	yes						
sunny	no						
rainy	yes						
overcast	yes						
rainy	yes						
overcast	yes						
rainy	yes						
overcast	yes						
overcast	no						
rainy	yes						
rainy	no						

s	r	o	
1	5	6	12
3	3	1	7
4	8	7	19

No podemos validar alguna forma visual de corroborar la correlación por medio de la siguiente grafica también no podemos validar una correlación entre ellos

