Si consideri la distribuzione doppia relativa ai caratteri Fatturato (FATT) e Fatturato Estero (FATEST) le cui intensità sono state raggruppate in classi:

Fatturato *(X)*	Fatturato Estero *(Y)*					
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%	Totale
≤ 200	6	9	7	2	1	25
200 -| 300	2	1	1	1	0	5
300 -| 400	1	2	3	1	0	7
400 -| 500	3	0	2	0	0	5
>500	2	0	4	2	0	8
Totale	14	12	17	6	1	50

Si misuri l'indipendenza assoluta attraverso l'indice di contingenza quadratica Φ^2 e l'indipendenza relativa attraverso l'indice *V* di Cramer.

$$\Phi^2 = \sum_{ij} \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} - 1 \qquad V = \sqrt{\frac{\Phi^2}{\min\{r-1, c-1\}}}$$

Per il calcolo dell'indice Φ^2 è consigliabile procedere nel modo seguente:

1. Si costruisce la tabella delle n_{ij}^2

Fatturato (X)	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	36	81	49	4	1
200 -| 300	4	1	1	1	0
300 -| 400	1	4	9	1	0
400 -| 500	9	0	4	0	0
>500	4	0	16	4	0

2. Si costruisce la tabella delle $n_{i\bullet}n_{\bullet j}$

Fatturato (X)	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	350	300	425	150	25
200 -| 300	70	60	85	30	5
300 -| 400	98	84	119	42	7
400 -| 500	70	60	85	30	5
>500	112	96	136	48	8

3. Si costruisce la tabella delle $\dfrac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}}$

Fatturato *(X)*	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	0,10	0,27	0,12	0,03	0,04
200 -| 300	0,06	0,02	0,01	0,03	0,00
300 -| 400	0,01	0,05	0,08	0,02	0,00
400 -| 500	0,13	0,00	0,05	0,00	0,00
>500	0,04	0,00	0,12	0,08	0,00

4. Si calcola l'indice Φ^2 sottraendo "1" alla somma degli elementi della tabella costruita al punto 3.

$$\Phi^2 = \sum_{ij} \frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1 = 1,24 - 1 = 0,24$$

L'indice V di Cramer sarà dato da:

$$V = \sqrt{\frac{\Phi^2}{\min\{r-1, c-1\}}} =$$

$$\sqrt{\frac{0,24}{\min\{5-1, 5-1\}}} = \sqrt{\frac{0,24}{4}} = 0,25$$

Calcolare l'indice Φ^2 e l'indice *V* ipotizzando che la distribuzione doppia delle aziende in base al Fatturato ed al Fatturato estero sia la seguente:

Fatturato *(X)*	Fatturato Estero *(Y)*					
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%	Totale
≤ 200	6	0	0	0	0	6
200 -| 300	0	1	0	0	0	1
300 -| 400	0	0	3	0	0	3
400 -| 500	0	0	0	5	0	5
>500	0	0	0	0	7	7
Totale	6	1	3	5	7	22

1. tabella delle n_{ij}^2

Fatturato *(X)*	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	36	0	0	0	0
200 -| 300	0	1	0	0	0
300 -| 400	0	0	9	0	0
400 -| 500	0	0	0	25	0
>500	0	0	0	0	49

2. tabella delle $n_{i\bullet}n_{\bullet j}$

Fatturato (X)	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	36	6	18	30	42
200 -| 300	6	1	3	5	7
300 -| 400	18	3	9	15	21
400 -| 500	30	5	15	25	35
>500	42	7	21	35	49

3. Si costruisce la tabella delle $\dfrac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}}$

Fatturato (X)	Fatturato Estero *(Y)*				
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%
≤ 200	1,00	0,00	0,00	0,00	0,00
200 -| 300	0,00	1,00	0,00	0,00	0,00
300 -| 400	0,00	0,00	1,00	0,00	0,00
400 -| 500	0,00	0,00	0,00	1,00	0,00
>500	0,00	0,00	0,00	0,00	1,00

4. Si calcola l'indice Φ^2 sottraendo 1 alla soma degli elementi della tabella costruita al punto 3.

$$\Phi^2 = \sum_{ij} \frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1 = 5 - 1 = 4$$

L'indice V di Cramer sarà dato da:

$$V = \sqrt{\frac{\Phi^2}{\min\{r-1, c-1\}}} =$$

$$\sqrt{\frac{4}{\min\{5-1, 5-1\}}} = \sqrt{\frac{4}{4}} = 1$$

Questa situazione corrisponde a quella di MASSIMA DIPENDENZA

....

Calcolare l'indice Φ^2 e l'indice *V* ipotizzando che la distribuzione doppia delle aziende in base al Fatturato ed al Fatturato estero sia la seguente:

Fatturato *(X)*	Fatturato Estero *(Y)*					
	≤ 5%	**5%-|10%**	**10%-|25%**	**25%-|50%**	>50%	Totale
≤ 200	5	5	5	5	5	25
200 -| 300	5	5	5	5	5	25
300 -| 400	5	5	5	5	5	25
400 -| 500	5	5	5	5	5	25
>500	5	5	5	5	5	25
Totale	25	25	25	25	25	125

1. tabella delle n_{ij}^2

Fatturato (X)	Fatturato Estero (Y)				
	≤ 5%	5%-|10%	10%-|25%	25%-|50%	>50%
≤ 200	25	25	25	25	25
200 -| 300	25	25	25	25	25
300 -| 400	25	25	25	25	25
400 -| 500	25	25	25	25	25
>500	25	25	25	25	25

2. tabella delle $n_{i\bullet}n_{\bullet j}$

Fatturato (X)	Fatturato Estero (Y)				
	≤ 5%	5%-|10%	10%-|25%	25%-|50%	>50%
≤ 200	625	625	625	625	625
200 -| 300	625	625	625	625	625
300 -| 400	625	625	625	625	625
400 -| 500	625	625	625	625	625
>500	625	625	625	625	625

3. Si costruisce la tabella delle $\frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}}$

Fatturato (X)	Fatturato Estero (Y)				
	≤ 5%	5%-|10%	10%-|25%	25%-|50%	>50%
≤ 200	0,04	0,04	0,04	0,04	0,04
200 -| 300	0,04	0,04	0,04	0,04	0,04
300 -| 400	0,04	0,04	0,04	0,04	0,04
400 -| 500	0,04	0,04	0,04	0,04	0,04
>500	0,04	0,04	0,04	0,04	0,04

4. Si calcola l'indice Φ^2 sottraendo 1 alla soma degli elementi della tabella costruita al punto 3.

$$\Phi^2 = \sum_{ij} \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} - 1 = 1 - 1 = 0$$

L'indice V di Cramer sarà dato da:

$$V = \sqrt{\frac{\Phi^2}{\min\{r-1, c-1\}}} =$$

$$\sqrt{\frac{0}{\min\{5-1, 5-1\}}} = \sqrt{\frac{0}{4}} = 0$$

Questa situazione corrisponde a quella di INPENDENZA.

....

L'indice di contingenza quadratica e quello di Cramer possono essere calcolati per qualsiasi distribuzione doppia.

Si calcolino i due indici per le seguenti tabelle:

Dimesione dell'azienda (X)	Settore Merceologico *(Y)*				
	Alimentare	Bevande	Health Care	Ice Packaging	Totale
Piccola	1	1	1	4	7
Medio-Piccola	6	0	2	5	13
Media	4	0	4	1	9
Medio-Grande	6	1	3	1	11
Grande	4	1	2	3	10
Totale	21	3	12	14	50

Risultati: $\Phi^2 = 0,23; \quad V = 0.28$

Fatturato *(X)*	Settore Merceologico *(Y)*				
	Alimentare	Bevande	Health Care	Ice Packaging	Totale
≤ 200	11	1	6	7	25
200 -| 300	1	1	1	2	5
300 -| 400	5	0	1	1	7
400 -| 500	1	1	2	1	5
>500	3	0	2	3	8
Totale	21	3	12	14	50

Risultati: $\Phi^2 = 0,18; \quad V = 0.24$

Calcolare la covarianza e la correlazione tra i caratteri “Fatturato” e “Numero di Addetti” per le prime 20 aziende incluse nel campione:

$$s_{XY} = \overline{xy} - \overline{x} \bullet \overline{y}$$

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}} = \frac{\overline{xy} - \overline{x} \bullet \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}$$

I dati da considerare sono i seguenti:

Azienda	FATT (X)	ADD94 (Y)	xy	x^2	y^2
1	1.021	2.600	2.654.600	1.042.441	6.760.000
2	109	292	31.828	11.881	85.264
3	233	323	75.259	54.289	104.329
4	199	1.320	262.680	39.601	1.742.400
5	354	640	226.560	125.316	409.600
6	145	135	19.575	21.025	18.225
7	467	1.176	549.192	218.089	1.382.976
8	177	225	39.825	31.329	50.625
9	161	326	52.486	25.921	106.276
10	158	378	59.724	24.964	142.884
11	115	192	22.080	13.225	36.864
12	108	385	41.580	11.664	148.225
13	1.444	1.477	21.32.788	2.085.136	2.181.529
14	493	933	459.969	243.049	870.489
15	185	208	38.480	34.225	43.264
16	285	678	193.230	81.225	459.684
17	242	576	139.392	58.564	331.776
18	386	205	79.130	148.996	42.025
19	981	2.001	19.62.981	962.361	4.004.001
20	105	2.664	279.720	11.025	7.096.896
Totale	7.368	16.734	9.321.079	5.244.326	26.017.332

$$\overline{x} = \frac{1}{n}\sum_j x_j = \frac{1}{20}7.368 = 368,4$$

$$\overline{y} = \frac{1}{n}\sum_j y_j = \frac{1}{20}16.734 = 836,7$$

$$\overline{xy} = \frac{1}{n}\sum_j x_j y_j = \frac{1}{20}9.321.079 = 466.053,95$$

$$s_{XY} = \overline{xy} - \overline{x} \bullet \overline{y} =$$
$$= 466.053,95 - 368,4 \bullet 836,7 = 157.813,57$$

$$\overline{x^2} = \frac{1}{n}\sum_j x^2{}_j = \frac{1}{20}5.244.326 = 262.213,3$$

$$\overline{y^2} = \frac{1}{n}\sum_j y^2{}_j = \frac{1}{20}26.017.332 = 1.300.867$$

$$r_{XY} = \frac{\overline{xy} - \overline{x} \bullet \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}} =$$

$$= \frac{157.813,57}{\sqrt{\left(262.213,3 - 368,4^2\right)\left(1.300.867 - 836,7^2\right)}} = 0,572$$

Calcolare la covarianza e la correlazione tra i caratteri Fatturato e Numero di Addetti per tutte le 50 aziende incluse nel campione.

(Risultati: $s_{XY} = 279.403,72$, $r_{XY} = 0,803$).

....

Data la seguente distribuzione doppia:

Fatturato *(X)*	Addetti *(Y)*					
	≤ 200	**200-|300**	**300-|550**	**550-|950**	>950	Totale
≤ 200	5	9	7	2	2	25
200 -| 300	1	1	1	2	0	5
300 -| 400	1	2	1	3	0	7
400 -| 500	0	0	0	2	3	5
>500	0	1	0	2	5	8
Totale	7	13	9	11	10	50

Calcolare la covarianza e la correlazione.

Trattandosi di due distribuzioni in classi bisogna considerare i punti medi. La covarianza sarà data da:

$$s_{XY} = \overline{xy} - \overline{x} \bullet \overline{y} = \sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij} - \overline{x} \bullet \overline{y}$$

Per calcolare il termine $\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij}$ è consigliabile costruire la tabella delle $\hat{x}_i \hat{y}_j n_{ij}$, ossia:

	100	250	425	750	3.608
100	50.000	225.000	297.500	150.000	721.600
250	25.000	62.500	106.250	375.000	0
350	35.000	175.000	148.750	787.500	0
450	0	0	0	675.000	4.870.800
1.256	0	314.000	0	1.884.000	22.658.240

La somma degli elementi all'interno di tale tabella è pari a $\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij}$, per cui

$$\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij} = 33.561.140$$

da cui $\overline{xy} = \dfrac{\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij}}{n} = \dfrac{33.561140}{50} = 671.223$

Per il calcolo della covarianza e della correlazione è utile considerare la seguente tabella:

$\hat{x}_i$	n_i	$\hat{y}_j$	n_j	$\hat{x}_i n_i$	$\hat{y}_j n_j$	$\hat{x}_i^2$	$\hat{x}_i^2 n_i$	$\hat{y}_j^2$	$\hat{y}_j^2 n_j$
100	**25**	**100**	**7**	**2.500**	**700**	**10.000**	**250.000**	**10.000**	**70.000**
250	**5**	**250**	**13**	**1.250**	**3.250**	**62.500**	**312.500**	**62.500**	**812.500**
350	**7**	**425**	**9**	**2.450**	**3.825**	**122.500**	**857.500**	**180.625**	**1.625.625**
450	**5**	**750**	**11**	**2.250**	**8.250**	**202.500**	**1.012.500**	**562.500**	**6.187.500**
1.256	**8**	**3.608**	**10**	**10.048**	**36.080**	**1.577.536**	**12.620.288**	**13.017.664**	**130.176.640**
				18.498	**52.105**		**15.052.788**		**138.872.265**

$$\overline{x} = \frac{1}{n}\sum_i \hat{x}_i n_i = \frac{1}{50}18.498 = 369,96$$

$$\overline{y} = \frac{1}{n}\sum_j \hat{y}_j n_j = \frac{1}{50}52.105 = 1.042,1$$

$$s_{XY}^2 = \overline{xy} - \overline{x} \bullet \overline{y} =$$

$$= 671.223 - 369,96 \bullet 1.042,1 = 285.687$$

$$\overline{x^2} = \frac{1}{n}\sum_i \hat{x}_i^2 n_i = \frac{1}{50}15.052.788 = 301.056$$

$$\overline{y^2} = \frac{1}{n}\sum_j \hat{y}_j^2 n_j = \frac{1}{50}138.872.275 = 2.777.445$$

$$r_{XY} = \frac{\overline{xy} - \overline{x} \bullet \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}} =$$

$$= \frac{285.687}{\sqrt{\left(301.056 - 369,96^2\right)\left(2.777.445 - 1.042,1^2\right)}} = 0,542$$