

***Distribuzioni doppie di frequenze
Analisi dell'associazione
tra due caratteri qualitativi***

1

Distribuzione doppia di frequenze

- Si considerano n unità e 2 caratteri, X e Y .
- La distribuzione doppia di frequenze (o tabella a doppia entrata) contiene l'insieme delle frequenze congiunte n_{ij}
- n_{ij} : il numero di unità che presentano la modalità i del carattere X e la modalità j del carattere Y .
- Presenta H righe e K colonne.

2

**Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»**

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
	TOT	2	6	9	3	20

Totale delle unità

3

**Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»**

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Ci sono 2 aziende appartenenti al settore agricolo e con un numero di dipendenti pari a 1				1	6
					1	5
	TOT	2	6	9	3	20

Totale delle unità

4

**Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»**

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
TOT	2	6	9	3	20	

Ci sono 4 aziende appartenenti al settore industriale e con un numero di dipendenti pari a 2

Totale delle unità

5

**Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»**

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
	TOT	2	6	9	3	20

6

Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
TOT	2 + 6 + 9 + 3 = 20					

7

Distribuzione doppia (tabella a doppia entrata)
«Settore di appartenenza» e «Numero dipendenti»

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
TOT	2	6	9	3	20	

$$H = 4$$

$$K = 4$$

8

**Distribuzione marginale del carattere
«Settore di appartenenza»**

		Numero dipendenti				<i>TOT</i>
		1	2	3	4	
Settore di appartenenza	<i>Agricolo</i>	2	1	0	0	3
	<i>Industriale</i>	0	4	1	1	6
	<i>Servizi</i>	0	1	4	1	6
	<i>Tecnologico</i>	0	0	4	1	5
	<i>TOT</i>	2	6	9	3	20

9

**Distribuzione marginale del carattere
«Settore di appartenenza»**

<i>Settore (X)</i>	<i>Freq.</i>
<i>Agricolo</i>	3
<i>Industriale</i>	6
<i>Servizi</i>	6
<i>Tecnologico</i>	5
<i>TOT</i>	20

10

**Distribuzione marginale del carattere
«Numero dipendenti»**

		Numero dipendenti				
		1	2	3	4	TOT
Settore di appartenenza	Agricolo	2	1	0	0	3
	Industriale	0	4	1	1	6
	Servizi	0	1	4	1	6
	Tecnologico	0	0	4	1	5
	TOT	2	6	9	3	20

11

**Distribuzione marginale del carattere
«Numero dipendenti»**

<i>Num dip (Y)</i>	<i>Freq.</i>
1	2
2	6
3	9
4	3
TOT	20

12

**Distribuzione del carattere
«Settore di appartenenza» condizionata a $Y = 2$**

		Numero dipendenti				<i>TOT</i>
		1	2	3	4	
Settore di appartenenza	<i>Agricolo</i>	2	1	0	0	3
	<i>Industriale</i>	0	4	1	1	6
	<i>Servizi</i>	0	1	4	1	6
	<i>Tecnologico</i>	0	0	4	1	5
	<i>TOT</i>	2	6	9	3	20

13

**Distribuzione del carattere
«Settore di appartenenza» condizionata a $Y = 2$**

<i>Settore (X)</i>	<i>Freq.</i>
<i>Agricolo</i>	1
<i>Industriale</i>	4
<i>Servizi</i>	1
<i>Tecnologico</i>	0
<i>TOT</i>	6

14

**Distribuzione del carattere
«Numero dipendenti» condizionata a X = Servizi**

		Numero dipendenti				
		1	2	3	4	TOT
Settore di appartenenza	<i>Agricolo</i>	2	1	0	0	3
	<i>Industriale</i>	0	4	1	1	6
	<i>Servizi</i>	0	1	4	1	6
	<i>Tecnologico</i>	0	0	4	1	5
	TOT	2	6	9	3	20

15

**Distribuzione del carattere
«Numero dipendenti» condizionata a X = Servizi**

<i>Num dip (Y)</i>	<i>Freq.</i>
1	0
2	1
3	4
4	1
TOT	6

16

Frequenze percentuali

- Frequenze percentuali

$$p_{ij} = 100 \frac{n_{ij}}{n}$$

- Frequenze percentuali di riga: si divide ogni frequenza per il totale di riga e si moltiplica per 100
- Frequenze percentuali di colonna: si divide ogni frequenza per il totale di colonna e si moltiplica per 100

17

Distribuzione doppia (tabella a doppia entrata) «Settore di appartenenza» e «Numero dipendenti»

		Numero dipendenti				
		1	2	3	4	TOT
Settore di appartenenza	<i>Agricolo</i>	2	1	0	0	3
	<i>Industriale</i>	0	4	1	1	6
	<i>Servizi</i>	0	1	4	1	6
	<i>Tecnologico</i>	0	0	4	1	5
	TOT	2	6	9	3	20

18

Frequenze percentuali

Numero dipendenti							
Settore di appartenenza		1	2	3	4	TOT	
	<i>Agricolo</i>	100(2/20)					
	<i>Industriale</i>						
	<i>Servizi</i>						
	<i>Tecnologico</i>						
TOT							

19

Frequenze percentuali

Numero dipendenti						
Settore di appartenenza		1	2	3	4	TOT
	<i>Agricolo</i>	10	5	0	0	15
	<i>Industriale</i>	0	20	5	5	30
	<i>Servizi</i>	0	5	20	5	30
	<i>Tecnologico</i>	0	0	20	5	25
	TOT	10	30	45	15	100

20

Frequenze percentuali

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	<i>Agricolo</i>	10	5	0	0	15
	<i>Industriale</i>	0	20	5	5	30
	<i>Servizi</i>	0	5	20	5	30
	Il 20% delle aziende appartiene al settore dei servizi e ha 3 dipendenti				5	25
				15	100	

21

Frequenze percentuali

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	<i>Agricolo</i>	10	5	0	0	15
	<i>Industriale</i>	0	20	5	5	30
	<i>Servizi</i>	0	5	20	5	30
	<i>Tecnologico</i>	0	0	20	5	25
Il 5% delle aziende appartiene al settore tecnologico e ha 4 dipendenti				15	100	

22

Frequenze percentuali di riga

Numero dipendenti							
Settore di appartenenza		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>TOT</i>	
	<i>Agricolo</i>	100(2/3)					
	<i>Industriale</i>						
	<i>Servizi</i>						
	<i>Tecnologico</i>						
<i>TOT</i>							

23

Frequenze percentuali di riga

Numero dipendenti						
Settore di appartenenza		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>TOT</i>
	<i>Agricolo</i>	66,67	33,33	0	0	100
	<i>Industriale</i>	0	66,67	16,67	16,67	100
	<i>Servizi</i>	0	16,67	66,67	16,67	100
	<i>Tecnologico</i>	0	0	80	20	100
	<i>TOT</i>	10	30	45	15	100

24

Frequenze percentuali di riga

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	66,67	33,33	0	0	100
	Industriale	0	66,67	16,67	16,67	100
	Servizi	0	16,67	66,67	16,67	100
	TOT	10	30	45	15	100

Il 66,67% delle aziende del settore industriale ha 2 dipendenti

25

Frequenze percentuali di colonna

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	Agricolo	100(2/2)				
	Industriale					
	Servizi					
	Tecnologico					
TOT						

26

Frequenze percentuali di colonna

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	<i>Agricolo</i>	100	16,67	0	0	15
	<i>Industriale</i>	0	66,67	11,11	33,33	30
	<i>Servizi</i>	0	16,67	44,44	33,33	30
	<i>Tecnologico</i>	0	0	44,44	33,33	25
	TOT	100	100	100	100	100

27

Frequenze percentuali di colonna

		Numero dipendenti				
Settore di appartenenza		1	2	3	4	TOT
	<i>Agricolo</i>	100	16,67	0	0	15
	<i>Industriale</i>	0	66,67	11,11	33,33	30
	<i>Servizi</i>	0	16,67	44,44	33,33	30
	<i>Tecnologico</i>	0	0	44,44	33,33	25
	TOT	100	100	100	100	100

L'11,11% delle aziende con 3 dipendenti appartiene al settore industriale

28

Tabella a doppia entrata (in simboli)

		<i>Y</i>					
		<i>y</i> ₁	...	<i>y</i> _{<i>j</i>}	...	<i>y</i> _{<i>K</i>}	<i>TOT</i>
<i>X</i>	<i>x</i> ₁	<i>n</i> ₁₁	...	<i>n</i> _{1<i>j</i>}	...	<i>n</i> _{1<i>K</i>}	<i>n</i> _{1·}

	<i>x</i> _{<i>i</i>}	<i>n</i> _{<i>i</i>1}	...	<i>n</i> _{<i>i</i><i>j</i>}	...	<i>n</i> _{<i>i</i><i>K</i>}	<i>n</i> _{<i>i</i>·}

	<i>x</i> _{<i>H</i>}	<i>n</i> _{<i>H</i>1}	...	<i>n</i> _{<i>H</i><i>j</i>}	...	<i>n</i> _{<i>H</i><i>K</i>}	<i>n</i> _{<i>H</i>·}
	<i>TOT</i>	<i>n</i> _{·1}	...	<i>n</i> _{·<i>j</i>}	...	<i>n</i> _{·<i>K</i>}	<i>n</i>

29

Tabella a doppia entrata

		<i>Y</i>					
		<i>y</i> ₁	...	<i>y</i> _{<i>j</i>}	...	<i>y</i> _{<i>K</i>}	<i>TOT</i>
<i>X</i>	<i>x</i> ₁	<i>n</i> ₁₁	...	<i>n</i> _{1<i>j</i>}	...	<i>n</i> _{1<i>K</i>}	<i>n</i> _{1·}

	<i>x</i> _{<i>i</i>}	<i>n</i> _{<i>i</i>1}	...	<i>n</i> _{<i>i</i><i>j</i>}	...	<i>n</i> _{<i>i</i><i>K</i>}	<i>n</i> _{<i>i</i>·}

	<i>x</i> _{<i>H</i>}	<i>n</i> _{<i>H</i>1}	...	<i>n</i> _{<i>H</i><i>j</i>}	...	<i>n</i> _{<i>H</i><i>K</i>}	<i>n</i> _{<i>H</i>·}
	<i>TOT</i>	<i>n</i> _{·1}	...	<i>n</i> _{·<i>j</i>}	...	<i>n</i> _{·<i>K</i>}	<i>n</i>

30

Tabella a doppia entrata

		<i>Y</i>					<i>TOT</i>
		<i>y</i> ₁	...	<i>y</i> _{<i>j</i>}	...	<i>y</i> _{<i>K</i>}	
<i>X</i>	<i>x</i> ₁	<i>n</i> ₁₁	...	<i>n</i> _{1<i>j</i>}	...	<i>n</i> _{1<i>K</i>}	<i>n</i> _{1·}

	<i>x</i> _{<i>i</i>}	<i>n</i> _{<i>i</i>1}	...	<i>n</i> _{<i>i</i><i>j</i>}	...	<i>n</i> _{<i>i</i><i>K</i>}	<i>n</i> _{<i>i</i>·}

	<i>x</i> _{<i>H</i>}	<i>n</i> _{<i>H</i>1}	...	<i>n</i> _{<i>H</i><i>j</i>}	...	<i>n</i> _{<i>H</i><i>K</i>}	<i>n</i> _{<i>H</i>·}
	<i>TOT</i>	<i>n</i> _{·1}	...	<i>n</i> _{·<i>j</i>}	...	<i>n</i> _{·<i>K</i>}	<i>n</i>

31

Ci sono n_{11} unità che presentano la modalità x_1 del carattere X e la modalità y_1 del carattere Y .

Tabella a doppia entrata

		<i>Y</i>					<i>TOT</i>
		<i>y</i> ₁	...	<i>y</i> _{<i>j</i>}	...	<i>y</i> _{<i>K</i>}	
<i>X</i>	<i>x</i> ₁	<i>n</i> ₁₁	...	<i>n</i> _{1<i>j</i>}	...	<i>n</i> _{1<i>K</i>}	<i>n</i> _{1·}

	<i>x</i> _{<i>i</i>}	<i>n</i> _{<i>i</i>1}	...	<i>n</i> _{<i>i</i><i>j</i>}	...	<i>n</i> _{<i>i</i><i>K</i>}	<i>n</i> _{<i>i</i>·}

	<i>x</i> _{<i>H</i>}	<i>n</i> _{<i>H</i>1}	...	<i>n</i> _{<i>H</i><i>j</i>}	...	<i>n</i> _{<i>H</i><i>K</i>}	<i>n</i> _{<i>H</i>·}
	<i>TOT</i>	<i>n</i> _{·1}	...	<i>n</i> _{·<i>j</i>}	...	<i>n</i> _{·<i>K</i>}	<i>n</i>

32

Tabella a doppia entrata

Ci sono $n_{1\cdot}$ unità che presentano la modalità x_1 del carattere X .

		Y					TOT
		y_1	...	y_j	...	y_K	
X	x_1	n_{11}	...	n_{1j}	...	n_{1K}	$n_{1\cdot}$

	x_i	n_{i1}	...	n_{ij}	...	n_{iK}	$n_{i\cdot}$

	x_H	n_{H1}	...	n_{Hj}	...	n_{HK}	$n_{H\cdot}$
	TOT	$n_{\cdot 1}$...	$n_{\cdot j}$...	$n_{\cdot K}$	n

33

Tabella a doppia entrata

		Y					TOT
		y_1	...	y_j	...	y_K	
	x_1	n_{11}	...	n_{1j}	...	n_{1K}	$n_{1\cdot}$

		n_{i1}	...	n_{ij}	...	n_{iK}	$n_{i\cdot}$
	
	x_H	n_{H1}	...	n_{Hj}	...	n_{HK}	$n_{H\cdot}$
	TOT	$n_{\cdot 1}$...	$n_{\cdot j}$...	$n_{\cdot K}$	n

Ci sono $n_{\cdot 1}$ unità che presentano la modalità y_1 del carattere Y .

34

Tabella a doppia entrata

		<i>Y</i>					<i>TOT</i>
		<i>y</i> ₁	...	<i>y</i> _{<i>j</i>}	...	<i>y</i> _{<i>K</i>}	
<i>X</i>	<i>x</i> ₁	<i>n</i> ₁₁	...	<i>n</i> _{1<i>j</i>}	...	<i>n</i> _{1<i>K</i>}	<i>n</i> _{1.}

	<i>x</i> _{<i>i</i>}	<i>n</i> _{<i>i</i>1}	...	<i>n</i> _{<i>i</i><i>j</i>}	...	<i>n</i> _{<i>i</i><i>K</i>}	<i>n</i> _{<i>i</i>.}

	<i>x</i> _{<i>H</i>}	<i>n</i> _{<i>H</i>1}	...	<i>n</i> _{<i>H</i><i>j</i>}
<i>TOT</i>	<i>n</i> _{.1}	...	<i>n</i> _{.<i>j</i>}	...	<i>n</i> _{.<i>K</i>}	<i>n</i>	

Totale delle unità

35

Analisi dell'associazione tra due caratteri

- Tra due caratteri si ha:
- indipendenza logica se *a priori* non può esistere una relazione di causa-effetto (es.: altezza e reddito);
- indipendenza statistica quando la conoscenza della modalità di uno dei due caratteri non migliora la previsione della modalità dell'altro.

36

Indipendenza tra X e Y

- Si consideri una tabella a doppia entrata con caratteri X e Y.
- Il carattere X è indipendente da Y se la distribuzione percentuale condizionata di X non cambia.
- Le frequenze percentuali delle distribuzioni condizionate sono uguali tra loro e uguali alle frequenze percentuali della distribuzione marginale.
- Se X è indipendente da Y, allora Y è indipendente da X.

37

Tabella a doppia entrata nel caso di indipendenza

		Genere (Y)		
		<i>Masch</i>	<i>Femm</i>	<i>TOT</i>
Colore Occhi (X)	<i>Cast</i>	3	10	13
	<i>Azzur</i>	9	30	39
	<i>TOT</i>	12	40	52

38

**Distribuzione del carattere Colore Occhi
condizionata a Genere = M**

X	n_{i1}	p_{i1}
<i>Cast</i>	3	25
<i>Azzur</i>	9	75
<i>TOT</i>	12	

39

**Distribuzione del carattere Colore Occhi
condizionata a Genere = F**

X	n_{i1}	p_{i1}
<i>Cast</i>	10	25
<i>Azzur</i>	30	75
<i>TOT</i>	40	

40

Distribuzione marginale del carattere Colore Occhi

<i>X</i>	n_{i1}	p_{i1}
<i>Cast</i>	13	25
<i>Azzur</i>	39	75
<i>TOT</i>	52	

41

- Se due caratteri sono indipendenti, allora

$$n_{ij} = \frac{n_i \cdot n_j}{n}$$

(totale riga x totale colonna) / totale

42

		Genere (Y)		
		<i>Masch</i>	<i>Femm</i>	<i>TOT</i>
Colore Occhi (X)	<i>Cast</i>	3	10	13
	<i>Azzur</i>	9	30	39
	<i>TOT</i>	12	40	52

$$9 = \frac{39 \cdot 12}{52}$$

43

- Infatti,

$$100 \frac{9}{12} = 100 \frac{39}{52}$$

da cui

$$9 = \frac{39 \cdot 12}{52}$$

44

- In generale,

$$100 \frac{n_{ij}}{n_{.j}} = 100 \frac{n_{i.}}{n}$$

da cui

$$n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

45

Associazione perfetta

1. Dipendenza perfetta
2. Interdipendenza perfetta

46

- Un carattere Y dipende perfettamente da X se ad ogni modalità di X è associata una sola modalità di Y .

47

Tabella 6: tabella a doppia entrata nel caso di dipendenza perfetta di Y da X

		Y			<i>TOT</i>
		y_1	y_2	y_3	
X	x_1	10	0	0	10
	x_2	0	20	0	20
	x_3	0	0	30	30
	x_4	0	0	45	45
	<i>TOT</i>	10	20	75	105

Non è vero il contrario: al valore y_3 non è associato un unico valore di X .

48

- Tra due caratteri X e Y sussiste interdipendenza perfetta se ogni modalità di X è associata ad una sola modalità di Y e viceversa.

49

Tabella 7: tabella a doppia entrata nel caso di interdipendenza perfetta

		Y			TOT
		y_1	y_2	y_3	
X	x_1	10	0	0	10
	x_2	0	20	0	20
	x_3	0	0	30	30
	TOT	10	20	30	60

50

Tabella 7: tabella a doppia entrata nel caso di interdipendenza perfetta

		Y			TOT
		y ₁	y ₂	y ₃	
X	x ₁	10	0	0	10
	x ₂	0	20	0	20
	x ₃	0	0	30	30
	TOT	10	20	30	60

51

Misura dell'associazione tra due caratteri qualitativi

- Si considerano n unità e 2 caratteri qualitativi, X e Y.
- I due casi estremi sono:
 1. Indipendenza

$$n_{ij} = \frac{n_i \cdot n_j}{n}$$

2. Associazione perfetta
 - Dipendenza perfetta
 - Interdipendenza perfetta

52

**«Condizione occupazionale» e «Residenza»
1418 giovani**

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	240	166	102	508
	<i>In cerca di lavoro</i>	80	94	214	388
	<i>Non cercano</i>	120	154	248	522
	<i>TOT</i>	440	414	564	1418

53

- La condizione occupazionale è indipendente dal luogo di residenza?
- Confronto tra le frequenze congiunte osservate n_{ij} e le frequenze congiunte teoriche (quelle che osserverei che vi fosse indipendenza tra i due caratteri)
- Le frequenze congiunte teoriche sono date da

$$n'_{ij} = \frac{n_i \cdot n_j}{n}$$

54

Frequenze teoriche

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	$\frac{508 \cdot 440}{1418}$			508
	<i>In cerca di lavoro</i>				388
	<i>Non cercano</i>				522
	<i>TOT</i>	440	414	564	1418

55

Frequenze teoriche

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	157,63	$\frac{508 \cdot 414}{1418}$		508
	<i>In cerca di lavoro</i>				388
	<i>Non cercano</i>				522
	<i>TOT</i>	440	414	564	1418

56

Frequenze teoriche

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	157,63	148,32	$\frac{508 \cdot 564}{1418}$	508
	<i>In cerca di lavoro</i>				388
	<i>Non cercano</i>				522
	<i>TOT</i>	440	414	564	1418

57

Frequenze teoriche

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	157,63	148,32	202,05	508
	<i>In cerca di lavoro</i>	120,39	113,28	154,33	388
	<i>Non cercano</i>	161,98	152,40	207,62	522
	<i>TOT</i>	440	414	564	1418

58

- Si definiscono le contingenze

$$c_{ij} = n_{ij} - n'_{ij}$$

59

Contingenze

		Residenza			
		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
Cond. Occupazionale	<i>Occupati</i>				
	<i>In cerca di lavoro</i>				
	<i>Non cercano</i>				
	<i>TOT</i>				

60

Contingenze

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	240-157,63			
	<i>In cerca di lavoro</i>				
	<i>Non cercano</i>				
	<i>TOT</i>				

61

Contingenze

		Residenza			
Cond. Occupazionale		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
	<i>Occupati</i>	82,37	17,68	-100,05	0
	<i>In cerca di lavoro</i>	-40,39	-19,28	59,67	0
	<i>Non cercano</i>	-41,98	1,60	40,38	0
		<i>TOT</i>	0	0	0

62

- L'indice di associazione Chi-quadrato di Pearson misura l'associazione tra due caratteri qualitativi

$$\chi^2 = \sum_{i=1}^H \sum_{j=1}^K \left[\frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} \right]$$

- Se i due caratteri sono indipendenti, $\chi^2 = 0$
- Se i due caratteri non sono indipendenti, $\chi^2 > 0$

63

Calcolo di χ^2

		Residenza			
		<i>Nord</i>	<i>Centro</i>	<i>Sud-Isole</i>	<i>TOT</i>
Cond. Occupazionale	<i>Occupati</i>	$\frac{82,37^2}{157,63}$			
	<i>In cerca di lavoro</i>				
	<i>Non cercano</i>				
	<i>TOT</i>				

64

Calcolo di χ^2

		Residenza			TOT
		Nord	Centro	Sud-Isole	
Cond. Occupazionale	Occupati	43,04	+ 2,11	+ 49,54	
	In cerca di lavoro	13,55	+ 3,28	+ 23,08	
	Non cercano	10,88	+ 0,02	+ 7,85	
	TOT				

65

$$\chi^2 = \sum_{i=1}^3 \sum_{j=1}^3 \left[\frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} \right]$$

$$\chi^2 = \frac{82,37^2}{157,63} + \frac{17,68^2}{148,32} + \dots + \frac{84038^2}{207,62}$$

$$\chi^2 = 153,35$$

66

L'indice V di Cramer

- Per avere un indice compreso tra 0 e 1, Cramer ha proposto

$$V = \sqrt{\frac{\chi^2/n}{\min[(H-1), (K-1)]}}$$

- H è il numero delle righe (modalità di X)
- K è il numero delle colonne (modalità di Y)
- $0 \leq V \leq 1$

67

Interpretazione dell'indice V di Cramer

- $V = 0$ → indipendenza tra i caratteri
- $0 < V < 1$ → associazione
- $V = 1$ → associazione perfetta (dipendenza o interdipendenza)

68

Calcolo dell'indice V di Cramer

$$\chi^2 = 153,35$$

$$V = \sqrt{\frac{153,35/1418}{\min[(3 - 1), (3 - 1)]}}$$

$$V = \sqrt{\frac{0,108}{2}} = 0,233$$

Debole dipendenza tra i due caratteri