

ESERCIZIO 1

La tabella seguente contiene i dati relativi alla composizione degli occupati in Italia relativamente ai tre macrosettori di attività (agricoltura, industria e altre attività) negli anni 1971 e 1981.

SETTORE DI ATTIVITA' ECONOMICA	OCCUPATI	
	1971	1981
Agricoltura	3'598	2'655
Industria	7'561	7'532
Altre attività	8'339	10'481
	19'498	20'668

In particolare i dati rappresentano la media annua (in migliaia) di occupati per ciascun settore.

- 1) Commentare brevemente la tabella (unità statistiche, tipo di tabella, tipo di variabile).
- 2) Calcolare le frequenze relative (usando un'approssimazione a due cifre decimali) e percentuali.
- 3) Calcolare le frequenze relative (usando un'approssimazione a tre cifre decimali).
- 4) Rappresentare graficamente i dati.

SVOLGIMENTO

1)

La tabella in questione può essere interpretata sia come distribuzione di frequenza che come **distribuzione di quantità o di intensità**. Infatti se il collettivo considerato è quello degli addetti allora la tabella può essere letta come una distribuzione di frequenza relativamente al carattere qualitativo Settore di attività economica. Il carattere assume, nella tabella riportata, le tre modalità Agricoltura, Industria e Altre attività.

Se invece si considera come collettivo di riferimento allora la tabella può essere vista come distribuzione di quantità del carattere quantitativo Numero di addetti.

E' comunque possibile lavorare sulla tabella per costruire le distribuzioni delle frequenze relative e percentuali.

2)

La seguente tabella riporta, accanto al numero di occupati, le colonne con le frequenze relative (con un'approssimazione alla seconda cifra decimale) e delle frequenze percentuali (per le quali non si riportano cifre decimali).

SETTORE DI ATTIVITA' ECONOMICA	OCCUPATI		FREQ. RELATIVE		FREQ. PERCENT.	
	1971	1981	1971	1981	1971	1981
Agricoltura	3'598	2'655	0.18	0.13	18%	13%
Industria	7'561	7'532	0.39	0.36	39%	36%
Altre attività	8'339	10'481	0.43	0.51	43%	51%
	19'498	20'668	1.00	1.00	100%	100%

L'utilizzo delle frequenze assolute permette di capire le variazioni numeriche nel fenomeno di interesse ma, al fine di confrontare le variazioni nella

composizione del carattere, è necessario utilizzare le frequenze relative o le frequenze percentuali.

Le prime (seconde) non sono niente altro che le frequenze assolute normalizzate in modo che il totale valga 1 (100). In particolare esse vengono calcolate, rispettivamente, come:

$$\text{freq. relativa} = \text{parte} / \text{totale}$$

$$\text{freq. percentuale} = \text{parte} / \text{totale} \times 100$$

e non sono nient'altro che la soluzione delle seguente proporzioni:

$$\text{valore osservato} : \text{totale} = \text{valore normalizzato} : 1 \text{ (frequenze relative)}$$

$$\text{valore osservato} : \text{totale} = \text{valore normalizzato} : 100 \text{ (frequenze assolute)}$$

Da evidenziare la necessità di ricorrere ad un'approssimazione in relazione al numero di decimali considerati. Si consideri a titolo di esempio, la frequenza relativa degli occupati in Agricoltura per l'anno 1971:

$$3958 / 19498 = 0.18453\dots$$

L'approssimazione (nel caso dell'esempio usando due cifre decimali) avviene secondo le seguenti regole:

- a) se la cifra decimale successiva alla precisione scelta (la terza nel caso in questione) è minore di 4 il numero viene troncato (approssimazione per difetto)
- b) se la cifra decimale successiva alla precisione scelta è maggiore di 5 la cifra precedente viene incrementata di un'unità (approssimazione per eccesso)
- c) se la cifra decimale successiva alla precisione scelta è uguale a 5 è indifferente scegliere una delle due regole precedenti. Nella pratica spesso si utilizza l'approssimazione per eccesso. Una regola leggermente più complessa richiede, nel caso di cifra pari a 5, l'uso dell'approssimazione per eccesso se la cifra decimale successiva è seguita da cifre almeno una delle quali diverse da zero e indifferentemente l'uno o l'altro tipo di arrotondamento nel caso di cifre decimali successive tutte uguali a zero.

3)

L'utilizzo di tre cifre decimali per la distribuzione delle frequenze relative mette in risalto come in alcuni casi sia necessario effettuare la cosiddetta quadratura dei risultati:

SETTORE DI ATTIVITA' ECONOMICA	OCCUPATI		FREQ. RELATIVE	
	1971	1981	1971	1981
Agricoltura	3'598	2'655	0.185	0.128
Industria	7'561	7'532	0.388	0.364
Altre attività	8'339	10'481	0.428	0.507
	19'498	20'668	1.001	0.999

QUADRATURA

vale a dire fare in modo che venga rispettato il vincolo di somma ad uno per le frequenze relative (e di somma a 100 per le frequenze percentuali).

A tale fine conviene ritoccare (aumentando o diminuendo opportunamente) il valore più grande risultante dalle divisioni. Nel caso della tabella la quadratura

si ottiene utilizzando il numero 0.427 in luogo di 0.428 per gli occupati del 1971 e il numero 0.508 in luogo di 0.507 per gli occupati del 1981. La scelta di ritoccare il numero più grande si basa sulla minimizzazione dell'errore relativo.

In particolare si definisce errore assoluto la differenza tra il valore ottenuto in seguito all'operazione di quadratura e il valore originario. Con riferimento ai dati del 1981 l'errore assoluto che si commetterebbe scegliendo uno qualunque dei tre numeri è uguale ed è pari a:

$$\text{Agricoltura} \rightarrow 0.129 - 0.128 = 0.001$$

$$\text{Industria} \rightarrow 0.365 - 0.364 = 0.001$$

$$\text{Altre Attiv.} \rightarrow 0.508 - 0.507 = 0.001$$

L'errore relativo è invece definito come l'errore relativo rapportato al valore originario, ovvero:

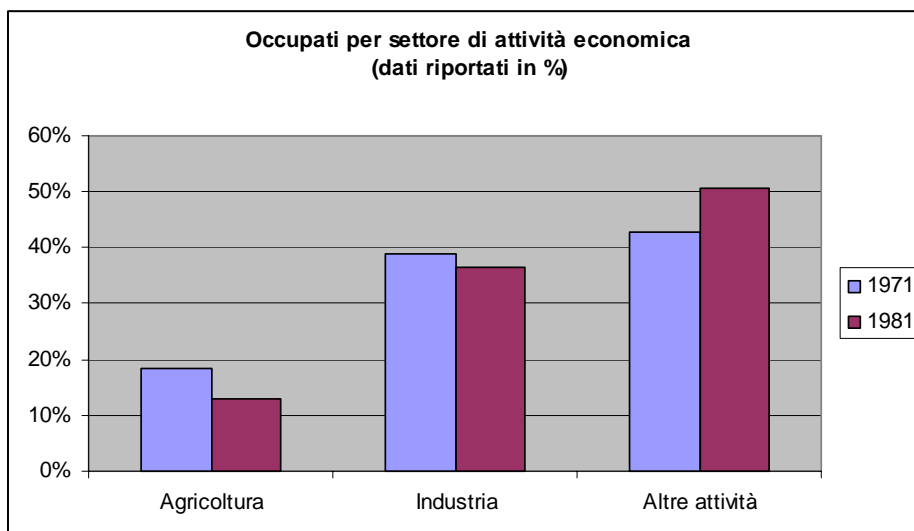
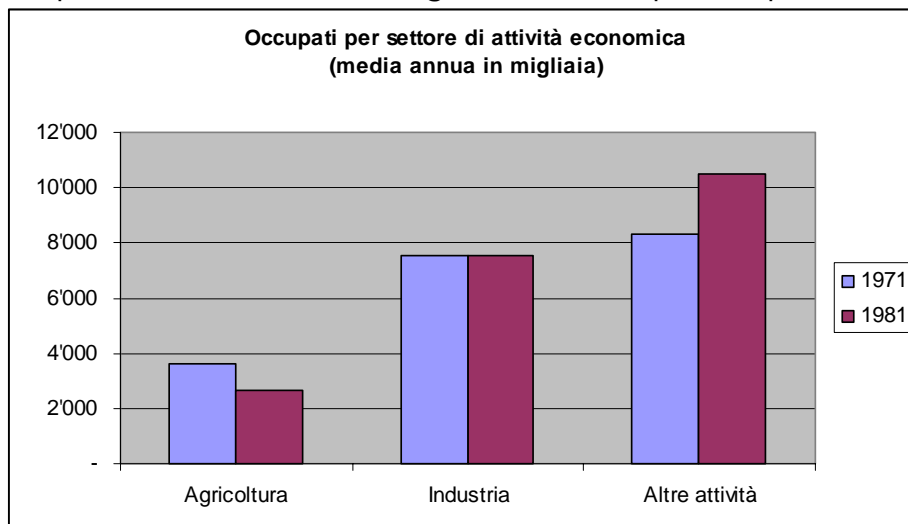
$$\text{Agricoltura} \rightarrow (0.129 - 0.128) / 0.128 = 0.0078$$

$$\text{Industria} \rightarrow (0.365 - 0.364) / 0.364 = 0.0027$$

$$\text{Altre Attiv.} \rightarrow (0.508 - 0.507) / 0.507 = 0.0020$$

4)

Di seguito si riportano i diagrammi a barre (a nastri) accostati per gli occupati utilizzando, rispettivamente, i dati originari e le frequenze percentuali.



ESERCIZIO 2

Il direttore del personale ha somministrato un breve questionario a 20 operaie di una data fabbrica nel quale si chiedeva

- il grado di pesantezza del lavoro svolto.
- il numero di figli a carico.
- lo stipendio medio orario.
- la disponibilità di un'auto propria.

I risultati sono riportati nella tabella che segue:

	Grado di pesantezza ⁽¹⁾	Numero figli	Stipendio Medio Orario	Disponibilità auto propria ⁽²⁾
1	2	0	22.5	0
2	2	1	23.0	1
3	3	3	18.5	1
4	4	2	18.3	1
5	1	0	15.0	0
6	3	1	25.7	0
7	3	0	24.2	1
8	1	2	16.7	1
9	2	2	17.9	1
10	3	1	15.0	0
11	4	2	24.6	0
12	2	0	26.8	1
13	3	2	21.5	1
14	5	2	20.3	1
15	3	2	23.6	1
16	5	2	18.4	1
17	2	1	18.9	0
18	4	0	19.4	0
19	4	3	19.3	1
20	1	1	26.0	0

⁽¹⁾ 1-->Poco faticoso, 2-->Moderatamente faticoso, 3-->Faticoso, 4-->Molto Faticoso, 5-->Estremamente faticoso

⁽²⁾ 0 --> No, 2 --> Sì

- 1) Commentare la tabella (unità statistiche, tipologia variabili, tipo di codifica).
- 2) Costruire, per ognuno dei quattro caratteri le distribuzioni di frequenza assolute, relative, percentuali e cumulate (assolute, relative e percentuali). In particolare per la variabile Stipendio medio orario utilizzare:
 - a. Divisione in classi equi-ampie (ampiezza classi 4)
 - b. Divisione in classi equi-ampie (numero di classi 4)
 - c. Divisione in classi equi-frequenti (frequenza 5).
- 3) Quali differenze è possibile riscontrare tra i due procedimenti di categorizzazione?
- 4) Ripetere la divisione in classi equi-frequenti (punto 2b) nel caso in cui lo stipendio medio orario dell'impiegato 16 è pari a 18.3 anziché all'effettivo 18.4.
- 5) Mostrare come è possibile ritornare alla distribuzione di frequenza assoluta partendo rispettivamente:
 - a. dalla distribuzione di frequenza relativa.
 - b. dalla distribuzione di frequenza percentuale.
 - c. dalla distribuzione di frequenza cumulata.

SVOLGIMENTO

1)

La tabella è una classica matrice dei dati unità x variabili. In particolare la matrice ha dimensioni 20 righe x 4 colonne (la prima colonna contiene un identificativo numerico e non può essere utilizzato ai fini di sintesi successive dei dati).

Le unità statistiche su cui è effettuata la rilevazione sono le 20 operaie della fabbrica in questione mentre i caratteri possono essere classificati come di seguito riportato:

Grado di pesantezza → variabile qualitativa o mutabile (ordinale)

Numero di figli → variabile quantitativa o variabile (discreta)

Stipendio medio orario → variabile quantitativa o variabile (continua)

Disponib. auto propria → variabile qualitativa o mutabile (sconnessa, in particolare dicotomica)

Per quanto riguarda la codifica i valori numerici attribuiti alla variabile grado di pesantezza (uno per ogni modalità che il carattere può assumere) non possono essere utilizzati ai fini dell'elaborazione. Essendo la variabile di tipo ordinale ha però senso confrontarli in termini di grandezza del valore.

La codifica utilizzata per la variabile Dispon. auto propria è la codifica che viene solitamente utilizzata nel caso di variabili dicotomiche, dove lo 0 indica l'assenza del carattere sulla particolare unità di rilevazione ed 1 la presenza dello stesso.

2)

Di seguito sono riportate le frequenze assolute, relative e percentuali per la variabile Grado di pesantezza.

Grado di pesantezza	Frequenza Assoluta	Frequenza Relativa	Frequenza Percentuale
1	3	0.15	15%
2	5	0.25	25%
3	6	0.3	30%
4	4	0.2	20%
5	2	0.1	10%
	20	1	100%

Essendo la variabile di tipo ordinale è possibile calcolare anche le frequenze cumulate. La seguente tabella riporta le frequenze assolute cumulate:

Grado di pesantezza	Frequenza Assoluta	Frequenza Assoluta Cumulata
1	3	3
2	5	8
3	6	14
4	4	18
5	2	20
	20	

La successiva tabella riporta le frequenze relative cumulate (funzione di ripartizione empirica) e le frequenze percentuali cumulate:

Grado di pesantezza	Frequenza Relativa	Frequenza Relativa Cumulata	Frequenza Percentuale Cumulata
1	0.15	0.15	15%
2	0.25	0.4	40%
3	0.3	0.7	70%
4	0.2	0.9	90%
5	0.1	1	100%
	1		

Per il carattere Disponibilità di auto propria, essendo questo di tipo sconnesso, è possibile calcolare solo le frequenze assolute, relative e percentuali.

Disponibilità auto propria	Frequenza Assoluta	Frequenza Relativa	Frequenza Percentuale
0	8	0.4	40%
1	12	0.6	60%
	20	1	100%

Il carattere quantitativo discreto Numero figli permette il calcolo delle frequenze assolute, relative e percentuali (sia semplici che cumulate), come riportate nella seguente tabella:

Numero figli	FREQUENZA			FREQUENZA		
	Assoluta	Relativa	Percentuale	Assoluta	Relativa	Percentuale
0	5	0.25	25%	5	0.25	25%
1	5	0.25	25%	10	0.5	50%
2	8	0.4	40%	18	0.9	90%
3	2	0.1	10%	20	1	100%
	20	1	100%			

Poiché le modalità assumibili dal carattere sono in numero ridotto le distribuzioni di frequenza risultanti sono abbastanza informative. Non è lo stesso per la variabile quantitativa continua Stipendio medio orario, la cui tabella di frequenze semplici risulterebbe:

Stipendio Medio Orario	Frequenza Assoluta
15.0	2
16.7	1
17.9	1
18.3	1
18.4	1
18.5	1
18.9	1
19.3	1
19.4	1
20.3	1
21.5	1
22.5	1
23.0	1
23.6	1
24.2	1
24.6	1
25.7	1
26.0	1
26.8	1

ovvero consisterebbe nell'elenco dei singoli valori assunti dalla variabile (ad eccezione del valore 15.0 che si ripete per le unità 5 e 10).

E' necessario quindi ricorrere alla distribuzione in classi del carattere (nel caso di distribuzioni in classi bisogna avere l'accortezza di specificare se i valori estremi della classe sono o meno compresi nella classe stessa).

2a)

La distribuzione in classi equiampie richiede semplicemente che venga specificata l'ampiezza desiderata per le classi o (caso più comune) il numero di classi in cui si intende dividere il carattere.

Nel primo caso, supponendo di volere utilizzare classi di ampiezza 4, la suddivisione è di agevole determinazione una volta noti il valore minimo e massimo del carattere, come di seguito indicato:

valore minimo = 15.0 valore max = 26.8 ampiezza classi = 4

classe 1 → da 15.0 a 19.0 (15.0 + 4)

classe 2 → da 19.0 a 23.0 (19.0 + 4)

classe 3 → da 23.0 a 27.0 (23.0 + 4)

e si procede poi alla costruzione della tabella di frequenza operando come visto in precedenza nel caso di una variabile qualitativa (ha naturalmente senso, essendo la categorizzazione intrinsecamente ordinale, calcolare anche i vari tipi di frequenze cumulate).

	Classi Stipendio	Frequenza Assoluta
1	[15, 19 [8
2	[19, 23 [5
3	[23, 27 [7
		20

2b)

Nel secondo caso è necessario determinare l'ampiezza delle classi prima di procedere, ovvero:

Ampiezza = (max – min) / num. Classi

Nell'esempio si ha quindi:

Ampiezza = (26.8 – 15.0) / 4 = 2.95

Essendo il carattere di tipo continuo è possibile utilizzare direttamente tale valore per la suddivisione in classi (nel caso discreto è invece necessario procedere ad un'approssimazione utilizzando un numero intero).

Le classi risultanti sono di seguito riportate:

classe 1 → da 15.00 a 17.95 (15.00 + 2.95)

classe 2 → da 17.95 a 20.90 (17.95 + 2.95)

classe 3 → da 20.90 a 23.85 (20.90 + 2.95)

classe 4 → da 23.85 a 26.80 (23.85 + 2.95)

	Classi Stipendio	Frequenza Assoluta
1	[15.00, 17.95 [4
2	[17.95, 20.90 [7
3	[20.90, 23.85 [4
4	[23.85, 26.80]	5
		20

2c)

Nel caso di classi equifrequenti è necessario procedere all'ordinamento dei dati per poi passare agevolmente alla suddivisione in classi:

Stipendio Medio Orario		
15.0	Classe 1	
15.0		
16.7		
17.9		
18.3		
18.4	Classe 2	
18.5		
18.9		
19.3		
19.4		
20.3	Classe 3	
21.5		
22.5		
23.0		
23.6		
24.2	Classe 4	
24.6		
25.7		
26.0		
26.8		

3)

Una differenza fondamentale nei due procedimenti di categorizzazione (a parte naturalmente le ovvie differenze nei risultati) consiste nel fatto che la distribuzione in classi equiampie richiede solo la conoscenza del minimo e del massimo e la specifica dell'ampiezza delle classi o la determinazione di tale valore a partire dal numero di classi desiderate. Il caso di classi equifrequenti richiede invece che i dati vengano inizialmente ordinati per poi procedere successivamente alla suddivisione in classi.

Inoltre, siccome nel caso di classi equifrequenti l'ampiezza delle classi sarà solitamente differente, è necessario calcolare le densità di frequenza prima di passare alla costruzione dell'istogramma:

densità di frequenza = frequenza classe / ampiezza classe

4)

Se si ripete il procedimento con la sostituzione del valore indicato si ottiene la seguente suddivisione in classi:

Stipendio Medio Orario		
15.0	Classe 1	
15.0		
16.7		
17.9		
18.3		
18.3		
18.5	Classe 2	
18.9		
19.3		
19.4		
20.3		
21.5	Classe 3	
22.5		
23.0		
23.6		
24.2		
24.6	Classe 4	
25.7		
26.0		
26.8		

Le classi risultanti non soddisfano strettamente il vincolo di uguale frequenza in quanto la prima classe contiene 6 unità (poiché, in seguito alla sostituzione effettuata, il quinto ed il sesto valore coincidono) e, di conseguenza, l'ultima classe contiene 4 unità.

5)

5a)

Il passaggio dalle frequenze relative a quelle assolute richiede la conoscenza della numerosità del collettivo di riferimento. In tale caso basta infatti moltiplicare la frequenza relativa per la numerosità del collettivo per ottenere la corrispondente frequenza assoluta.

5b)

Analogamente nel caso del passaggio da frequenze percentuali a frequenze assolute basta moltiplicare la frequenza percentuale per la numerosità del collettivo e dividere per 100.

5c)

Il passaggio dalle frequenze assolute cumulate a quelle relative si ottiene semplicemente effettuando la differenza tra la frequenza cumulata fino alla modalità di interesse e quelle della modalità precedente nell'ordinamento (naturalmente le due frequenze coincidono nel caso della prima modalità). Analogamente al caso precedente, per passare dalle frequenze relative/percentuali cumulate a quelle assolute, è necessario conoscere la numerosità del collettivo di riferimento.

ESERCIZIO 3

Il responsabile del personale di un'azienda pubblica conosce la distribuzione degli impiegati secondo la qualifica funzionale che è riportata nella seguente tabella:

Qualifica funzionale	Frequenza
Prima	58
Seconda	308
Terza	287
Quarta	71
Quinta	52
Sesta	28
Settima	12
	816

- 1) Sapendo che l'azienda ha bandito un concorso per quattro posti riservati agli interni con una qualifica non inferiore alla V, calcolare il numero di possibili candidati.
- 2) Calcolare le frequenze assolute cumulate e la funzione di ripartizione empirica ed utilizzarle per rispondere al quesito 1.

SVOLGIMENTO

1)

Il numero di possibili candidati è pari semplicemente alla somma delle frequenze relative associate alle categorie V, VI e VII, ovvero:

$$52 + 28 + 12 = 92$$

2)

La seguente tabella contiene le frequenze cumulate (assolute e relative).

Qualifica funzionale	Frequenza	Frequenza Relativa	Frequenza Assoluta Cumulata	Frequenza Relativa Cumulata
Prima	58	0.07	58	0.07
Seconda	308	0.38	366	0.45
Terza	287	0.35	653	0.80
Quarta	71	0.09	724	0.89
Quinta	52	0.06	776	0.95
Sesta	28	0.03	804	0.99
Settima	12	0.01	816	1.00
	816	1		

Per calcolare il numero di possibili candidati è possibile semplicemente sottrarre al valore complessivo (816) il numero di dipendenti con categoria non superiore alla quarta (724):

$$816 - 724 = 92$$

Lavorando sulle frequenze relative cumulate è possibile invece ottenere solo il valore percentuale dei dipendenti interessati al concorso, ovvero:

$$1 - 0.89 = 0.11$$

ovvero l'11% dei dipendenti è interessato al concorso. Per ottenere il numero di dipendenti è necessario avere l'informazione sulla numerosità del collettivo di interesse:

$$0.11 * 816 = 92$$