## Revisione di elementi di calcolo delle probabilità

1) Supponete si traggano due dadi e lasciate che *M* indichi la somma del numero di punti sui due dadi (quindi *M* è un numero compreso tra 1 e 12) non truccati. La tabella seguente elenca tutti i possibili risultati per la variabile casuale *M* insieme ai relativi valori della distribuzione delle probabilità e della distribuzione delle probabilità cumulative.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Risultato (somma di punti) | 2 = 1 + 1 | 3= 1 + 2= 2 + 1 | 4= 1 + 3= 3 + 1= 2 + 2 | 5= 1 + 4= 4 + 1= 2 + 3= 3 + 2 | 6= 1 + 5= 5 + 1= 2 + 4= 4 + 2= 3 + 3 | 7= 1 + 6= 6 + 1= 2 + 5= 5 + 2= 3 + 4= 4 + 3 | 8= 2 + 6= 6 + 2= 3 + 5= 5 + 3= 4 + 4 | 9= 3 + 6= 6 + 3= 2 + 7= 7 + 2 | 10= 4 + 6= 6 + 4= 5 + 5 | 11 = 5 + 6= 6 + 5 | 12= 6 + 6 |
| Distribuzione di probabilità | 0,028= $\frac{1}{6}×\frac{1}{6}$ | 0,056 = $\frac{1}{6}×\frac{1}{6}×2$  | 0,083 = $\frac{1}{6}×\frac{1}{6}×3$ | 0,111= $\frac{1}{6}×\frac{1}{6}×4$ | 0,139 = $\frac{1}{6}×\frac{1}{6}×5$ | 0,167 = $\frac{1}{6}×\frac{1}{6}×6$ | 0,139 = $\frac{1}{6}×\frac{1}{6}×5$ | 0,111= $\frac{1}{6}×\frac{1}{6}×4$ | 0,083= $\frac{1}{6}×\frac{1}{6}×3$ | 0,056 = $\frac{1}{6}×\frac{1}{6}×2$  | 0,028= $\frac{1}{6}×\frac{1}{6}$ |
| Distribuzione di probabilità cumulativa | 0,028 | 0,083 | 0,167 | 0,278 | 0,417 | 0,583 | 0,722 | 0,833 | 0,912 | 0,972 | 1,000 |

1. Disegnare entrambe le distribuzioni mostrate nella tabella di cui sopra.

Risposta:

1. Calcolare il valore atteso e la deviazione standard per *M*.

Risposta: 7,0; 2,42.

2) Qual è la probabilità dei risultati seguenti?

1. Pr(*M* = 7)
2. Pr(*M* = 2 o *M* = 10)
3. Pr(*M* = 4 o *M*4)
4. Pr(*M* = 6 e *M* = 9)
5. Pr(*M* < 8)
6. Pr(*M* = 6 o *M* > 10)

Risposta:

(a) 0,167 ovvero ;

(b) 0,111 ovvero ;

(c) 1;

(d) 0;

(e) 0,583;

(f) 0,222 ovvero .

 3) Le probabilità e le frequenze relative sono concetti legati l'uno all'altro in quanto la probabilità di un risultato è la frequenza con cui il risultato si verifica nel lungo periodo. Quindi i concetti delle distribuzioni congiunte, marginali e di probabilità condizionale derivano dai concetti relativi delle distribuzioni di frequenza.

Supponiamo dunque che siate interessati a indagare il rapporto tra l'età dei capi-famiglia e i guadagni settimanali delle famiglie. Si considerino a tale scopo dei dati sul numero di occorrenze di diversi valori di queste variabili raggruppate per età e reddito. I dati sono relativi a 1.744 individui. Si pensi a questi individui come a una popolazione che si desidera descrivere, piuttosto che un campione tramite il quale si desideri dedurre i tratti di una popolazione più grande. Dopo aver ordinato i dati in classi rispetto ai valori assunti dalle due variabili, si genera la tabella seguente:

## Tabella 1

## Frequenze assolute congiunte di età e reddito per una popolazione di 1.744 famiglie

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  | **Età** **del capofamiglia**  |  |  |
|  |  |  |  |  |  |  |
| **Reddito familiare**  | 16-Under 20 | 20-Under 25 | 25-sotto 45 | 45-sotto 65 | 65 e > |
|  | $0-sotto $200 | 80 | 76 | 130 | 86 | 24 |
|  | $200-sotto $400 | 13 | 90 | 346 | 140 | 8 |
|  | $400-sotto $600 | 0 | 19 | 251 | 101 | 6 |
|  | $600-sotto $800 | 1 | 11 | 110 | 55 | 1 |
|  | $800 e > | 1 | 1 | 108 | 84 | 2 |

 La mediana del gruppo di reddito che guadagna non meno di $800 è $1.050.

* 1. Calcolare le frequenze relative congiunte e marginali. Tracciare la distribuzione del reddito cumulativa.

Risposta: Le frequenze relative congiunte e marginali sono indicate nella tabella in basso. Quelle congiunte sono nelle celle all'interno del bordo. Quelle marginali del reddito familiare sono nella colonna esterna a destra che riporta i totali per riga. Quelle marginali dell'età sono nella riga esterna in basso che riporta i totali per colonna. Ad esampio, il 5,2 per cento (= 90/1.744) degli individui sono di età compresa tra 20 e 24 anni e guadagnano tra $200 e $400, mentre il 21,6 per cento (= (0+19+251+101+6)/ 1.744 = 377/1.744) degli individui guadagna fra $400 e $600.

## Tabella 2

## Frequenze relative congiunte e marginali di età e reddito, 1.744 famiglie

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  | **Età** **del capofamiglia**  |  |  |  |
|  |  |  |  |  |  |  |  |
| **Reddito familiare**  | 16-Under 20 | 20-Under 25 | 25-sotto 45 | 45-sotto 65 | 65 e > | Totale |
|  | $0-sotto $200 | 0,046 | 0,044 | 0,075 | 0,049 | 0,014 |  0,227 |
|  | $200-sotto $400 | 0,007 | 0,052 | 0,198 | 0,080 | 0,005 |  0,342 |
|  | $400-sotto $600 | 0,000 | 0,011 | 0,144 | 0,058 | 0,003 |  0,216 |
|  | $600-sotto $800 | 0,001 | 0,006 | 0,063 | 0,032 | 0,001 |  0,102 |
|  | $800 e > | 0,001 | 0,001 | 0,062 | 0,048 | 0,001 |  0,112 |
|  | **Totale** | 0,055 | 0,114 | 0,542 | 0,267 | 0,024 |  |

* 1. Calcolare le frequenze relative di reddito condizionali per le due categorie di età 16-Under 20, e 45-under 65. Calcolare il reddito medio delle famiglie per entrambe le categorie di età.

Risposta: Le frequenze calcolate sono riportate nella tabella qui sotto.

 Sulla base di tali frequenze, il reddito medio delle famiglie per la categoria di età 16-Under 20 è di circa $144, mentre è di circa $489 per la categoria di età 45-under 65.

## Tabella 3

## Frequenze relative di reddito per le categorie di età "16-Under 20" e "45-sotto 65"

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  | **Età** **del capofamiglia**  |
|  |  |  |  |  |
| **Reddito familiare**  | 16-Under 20 |  | 45-sotto 65 |
|  | $0-sotto $200 | 83.64% |  | 18.35% |
|  | $200-sotto $400 | 12.73% |  | 29.96% |
|  | $400-sotto $600 | 0.00% |  | 21.72% |
|  | $600-sotto $800 | 1.82% |  | 11.99% |
|  | $800 e > | 1.82% |  | 17.98% |

Qui di seguito è disponibile un link a un foglio elettronico di MS Excel con i calcoli:



* 1. Se il reddito delle famiglie e l'età del capo della famiglia fossero distribuiti in modo indipendente, che cosa vi aspettereste riguardo al rapporto tra le due distribuzioni condizionali di cui sopra e tra queeste e la distribuzione marginale del reddito? È questo il caso?

Risposta: Dovrebbero essere identiche tra di loro e alla distribuzione marginale. Chiaramente non è questo il caso. Si confrontino ad esempio la distribuzione condizionale del reddito per le due classi di età di cui sopra con la distribuzione nella ultima colonna di Tabella 2. Quindi le due variabili (reddito delle famiglie e l'età del capo della famiglia non sono indipendentemente distribuiti).

* 1. Il tuo libro di testo ti ha dato una definizione alternativa di indipendenza che non comporta distribuzioni di frequenza relativa condizionale. Qual è la definizione? Pensi che l'età e il reddito sono indipendenti qui, utilizzando questa definizione?

Risposta: La definizione alternativa è basata sulla condizione seguente: per ciascun valore di *x* e *y*. Possiamo verificare se vale questa condizione moltiplicando le due probabilità marginali per ciascuna possibile coppia di valori di *X* e *Y* e vedere se il risultato è uguale alla loro probabilità congiunta. Per esempio, poiché  e , il loro prodotto è 0,117, che non è uguale alla loro probabilità congiunta (ovvero 0,144). Per chi abia già studiato statistica inferenziale, si ricorda che, poiché stiamo trattando i dati come riferiti a una popolazione piuttosto che a un campione, non dobbiamo preoccuparci di verificare se la differenza tra 0,117 e 0,144 sia dovuta a errore di campionamento (ovevro se siano differenti in maniera statisticamente significativa).

4) Avete probabilmente sentito parlare della cosiddetta teoria "catch-up" sostenuta da storici dell'economia, per cui le nazioni che sono più indietro nel reddito pro-capite tendono a crescere più velocemente in seguito. Per mettere la teoria alla prova, si raccolgono i dati sul reddito pro capite per due anni, 1960 e 1990, per 24 paesi OCSE, espresso come frazione del reddito pro-capite degli USA. Si pensi a questi paesi come una popolazione che si desidera descrivere, piuttosto che un campione da cui si desidera dedurre il comportamento di una popolazione più grande. I dati rilevanti per questa domanda sono i seguenti:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 0,023 | 0,770 | 1,030 | 0,018 | 0,00053 | 0,593 | 1,0609 |
| 0,014 | 1,000 | 1,000 | 0,014 | 0,00020 | 1,000 | 1,0000 |
| .... | .... | .... | .... | .... | .... | .... |
| 0,041 | 0,200 | 0,450 | 0,008 | 0,00168 | 0,040 | 0,2025 |
| 0,033 | 0,130 | 0,230 | 0,004 | 0,00109 | 0,017 | 0,0529 |
| 0,625 | 13,220 | 17,800 | 0,294 | 0,01877 | 8,529 | 13,9164 |

Dove  E  sono reddito pro-capite relativamente agli Stati Uniti negli anni 1960 e 1990, rispettivamente, e *Y* è il tasso di crescita annuo medio *X* nel periodo 1960-1990. I numeri nell'ultima riga rappresentano le somme delle colonne di cui sopra.

A. Calcolare la varianza e la deviazione standard di  E . Perchè si possa concludere che un effetto catch-up sia presente, quale rapporto ci deve essere tra le due deviazioni standard? È questo il caso?

Risposta: Le varianze di  E  sono rispettivamente 0,0520 e 0,0298, con deviazioni standard di 0,2279 e 0,1726. Perchè si possa concludere che un effetto catch-up sia presente, la deviazione standard dovrebbe ridursi nel tempo. Questo è effettivamante il caso.

B. Calcolare la correlazione tra *Y* e . Quale segno deve avere il coefficiente di correlazione perché sia la prova di un effetto catch-up?

Risposta: Il coefficiente di correlazione è – 0,88. Deve essere negativo perché si possa parlare di un effetto catch-up. Se i paesi che erano relativamente avanti nel periodo iniziale in termini di reddito pro-capite crescono di relativamente meno nel tempo, poi alla fine i ritardatari recuperano (catch-up).

5) Per volontà di Alfred Nobel, ci sono cinque premi Nobel assegnati ogni anno per meriti eccezionali nella chimica, nella fisica, nella fisiologia o nella medicina, nella letteratura e nella pace. Nel 1968, la banca di Svezia aggiunse un premio in scienze economiche in memoria di Alfred Nobel. La tabella di seguito descrive la distribuzione di probabilità congiunta relativa ai caratteri "essere destinatari di un premio Nobel in economia o in una delle altre cinque discipline" e "cittadinanza" dei destinatari di premi Nobel nel periodo 1969-2001. Si pensi ai dati come volti a descrivere una popolazione (quella dei destinatari), piuttosto che un campione da cui si desidera dedurre il comportamento di una popolazione più grande.

Distribuzione congiunta di premi Nobel in discipline economiche e non-economia, e cittadinanza, 1969-2001

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cittadino degli Stati Uniti () | Cittadino non statunitense () | Totale |
| Premio Nobel di economia () | 0,118 | 0,049 | 0,167 |
| Fisica, chimica, medicina, letteratura e premio Nobel di pace () | 0,345 | 0,488 | 0,833 |
| Totale | 0,463 | 0,537 | 1,00 |

A. Calcolare  e interpretare il numero risultante.

Risposta: , ovvero il 53,7 per cento dei vincitori del premio Nobel erano cittadini non statunitensi.

B. Calcola e interpreta  E .

Risposta:  vuol dire che il 58,6 per cento dei vincitori del premio Nobel nelle discipline non di economia erano cittadini non-USA.  vuol dire che il 29,3 per cento dei vincitori del premio Nobel di economia erano cittadini non-USA.

C. Un vincitore del premio Nobel casualmente selezionato riporta che è un cittadino non statunitense. Qual è la probabilità che questa persona abbia vinto il premio Nobel per l'economia? Qual è la probabilità che abbia vinto un premio Nobel nelle altre cinque discipline?

Risposta: C' è una probabilità di 9,1 per cento che abbia vinto il premio Nobel per l'economia e una probabilità di 90,9 per cento che abbia vinto un premio Nobel in una delle altre cinque discipline.

D. Mostrare che la distribuzione congiunta sarebbe simile se le due categorie fossero indipendenti.

Risposta:

Distribuzione congiunta di premi Nobel in discipline economiche e non-economia, e cittadinanza, 1969-2001, sotto assunzione di indipendenza

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cittadino degli Stati Uniti () | Non= Cittadino degli Stati Uniti () | Totale |
| Premio Nobel di economia () | 0,077 | 0,090 | 0,167 |
| Fisica, chimica, medicina, letteratura e premio Nobel di pace () | 0,386 | 0,447 | 0,833 |
| Totale | 0,463 | 0,537 | 1,00 |

Per i calcoli dettagliati, si veda il foglio di calcolo al link di seguito:



# 6) Si acceda al sito del libro di testo di Stock & Watson al link seguente: [**http://www.pearsonglobaleditions.com/Stock**](http://www.pearsonglobaleditions.com/Stock)

#  Si ottenga il dataset CPS (ch8\_cps. xls) relativo ad un esempio trattato nel capitolo 8, aprendolo in un programma di foglio di calcolo come Excel. Per l'esercizio, utilizzare le prime osservazioni di 500 relative ai guadagni orari medi (Ahe). Si utilizzino statistiche riepilogative, quali media, meadiana, varianza e asimmetria, per descrivere la distribuzione dei guadagni. Si produce inoltre una distribuzione di frequenza ("istogramma") utilizzando classi di misura dei guadagni di ragionevoli dimensioni.

 Risposta:

|  |
| --- |
| *Ahe* |
|  |  |
| Dire | 19,79 |
| Errore standard | 0,51 |
| Mediana | 16,83 |
| Modalità | 19,23 |
| Deviazione standard | 11,49 |
| Varianza del campione | 131,98 |
| Curtosi | 0,23 |
| Asimmetria | 0,96 |
| Gamma | 58,44 |
| Minimo | 2,14 |
| Massimo | 60,58 |
| Somma | 9897,45 |
| Conteggio | 500,00 |

# La media è $19,79. La mediana ($16,83) è inferiore alla media, suggerendo che la media è stata tirata su da individui con guadagni orari media relativamente elevati. Ciò è confermato dalla misura di asimmetria, che è positiva e quindi suggerisce una distribuzione con una coda lunga alla destra. La varianza è $2131,96, mentre la deviazione standard è $11,49.

# Per generare la distribuzione di frequenza in Excel, è necessario innanzitutto definire il numero di classi di misura e dunque l'ampiezza dei relativi intervalli. Una volta che avete deciso sul numero, la differenza tra il minimo e il massimo nei dati suggerisce la larghezza della classe. Nel seguito, ho optato per 8 intervalli con una ampiezza di ciascuna classe pari a $8 (poiché i salari minimi in California sono attualmente $8 e circa lo stesso in altri Stati degli Stati Uniti).

# La tabella riporta le frequenze assolute, e le relative frequenze possono essere calcolate in modo semplice e ovvio.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| *Estremo superiore delle classi di misura ("bins", che non vuol dire "bidoni"!)* | *Frequenza* | *Frequenza relativa* |
| 8 | 50 | 0,1 |
| 16 | 187 | 0,374 |
| 24 | 115 | 0,23 |
| 32 | 68 | 0,136 |
| 40 | 38 | 0,076 |
| 48 | 33 | 0,066 |
| 56 | 8 | 0,016 |
| 66 | 1 | 0,002 |
| > 66 | 0 |  |

# La sostituzione delle frequenze relative nella tabella dell'istogramma produce quindi il grafico seguente (dopo aver eliminato gli spazi tra le barre).

7) Nel considerare l'acquisto di un certo stock, si attribuiscono le seguenti probabilità di possibili variazioni del prezzo delle azioni nel corso dell'anno successivo.

|  |  |
| --- | --- |
| Cambio di prezzo delle azioni duranteProssimi dodici mesi (%) | Probabilità |
| + 15 | 0,2 |
| + 5 | 0,3 |
| 0 | 0,4 |
| -5 | 0,05 |
| -15 | 0,05 |

Qual è il valore atteso, la varianza e la deviazione standard? Qual è il risultato più probabile? Si produca uno schizzo della funzione di distribuzione cumulativa.

.

Risposta: E(Y) = 3,5; 8,49; = 2,91; molto probabilmente: 0.