

Michael Sullivan III

FONDAMENTI DI STATISTICA

Edizione italiana a cura di Emma Zavarrone
Università IULM, Milano

© 2011 Pearson Italia, Milano-Torino

*Authorized translation from the English language edition, entitled: **Fundamentals of Statistics, 3rd Edition**, by Michael Sullivan, published by Pearson Education, Inc., publishing as Pearson, Copyright © 2011*

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Italian language edition published by Pearson Italia S.p.A., Copyright © 2011.

Le informazioni contenute in questo libro sono state verificate e documentate con la massima cura possibile. Nessuna responsabilità derivante dal loro utilizzo potrà venire imputata agli Autori, a Pearson Italia S.p.A. o a ogni persona e società coinvolta nella creazione, produzione e distribuzione di questo libro.

Per i passi antologici, per le citazioni, per le riproduzioni grafiche, cartografiche e fotografiche appartenenti alla proprietà di terzi, inseriti in quest'opera, l'editore è a disposizione degli aventi diritto non potuti reperire nonché per eventuali non volute omissioni e/o errori di attribuzione nei riferimenti.

I diritti di riproduzione e di memorizzazione elettronica totale e parziale con qualsiasi mezzo, compresi i microfilm e le copie fotostatiche, sono riservati per tutti i paesi.

LA FOTOCOPIATURA DEI LIBRI È UN REATO Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da AIDRO, corso di Porta Romana n. 108, 20122 Milano, e-mail segreteria@aidro.org e sito web www.aidro.org.

Edizione italiana a cura di: Emma Zavarrone

Traduzione: Carmelo Giarratana, Michela Gnaldi, Francesco Paolo Natale, Anna Simonetto, Marika Vezzoli

Realizzazione editoriale: Carmelo Giarratana

Progetto grafico di copertina: Achilli Ghizzardi Associati – Milano

Stampa: EcoBook – Rho (MI)

Tutti i marchi citati nel testo sono di proprietà dei loro detentori.

978-88-7192-635-3

978-88-7192-760-2

Printed in Italy

1^a edizione: settembre 2011

Ristampa

00 01 02 03 04

Anno

11 12 13 14 15

Sommario

Prefazione all'edizione italiana	XI
Prefazione	XII
Introduzione alle applet	XX

PARTE I OTTENERE LE INFORMAZIONI NECESSARIE **1**

Capitolo 1 Raccolta delle informazioni **2**

1.1 Primi passi verso la statistica	2
1.1.1 Definizione di statistica	2
1.1.2 Spiegare il processo della statistica	4
1.1.3 Distinguere tra variabili qualitative e quantitative	6
1.1.4 Distinguere tra variabili discrete e continue	8
1.1.5 Determinare la scala di misurazione di una variabile	9
1.2 Campionamento casuale semplice	11
1.2.1 Campionamento	11
1.2.2 Ottenere un campione casuale semplice	12
1.2.3 Ottenere un campione casuale semplice	12
1.3 Altri metodi di campionamento	16
1.3.1 Ottenere un campione stratificato	16
1.3.2 Ottenere un campione sistematico	17
1.3.3 Ottenere un campione a grappolo	18
1.4 Errore di campionamento	22
1.4.1 Spiegare le fonti degli errori di campionamento	22

PARTE II STATISTICA DESCRITTIVA 31

Capitolo 2 Organizzare e sintetizzare i dati 32

2.1 Organizzare i dati qualitativi	32
2.1.1 Organizzare i dati qualitativi in tabelle	32
2.1.2 Costruire un grafico a barre	34
2.1.3 Costruire grafici a torta	39
2.2 Organizzare i dati quantitativi: le rappresentazioni più utilizzate	41
2.2.1 Organizzare i dati discreti in tabelle	41
2.2.2 Costruire un istogramma per i dati discreti	42
2.2.3 Organizzare i dati quantitativi continui in tabelle	42
2.2.4 Costruire un istogramma per i dati continui	46
2.2.5 Disegnare un grafico a punti	48
2.2.6 Identificare la forma di una distribuzione	48
2.2.7 Disegnare un grafico per serie storiche	49
2.3 Rappresentazioni grafiche errate	50
2.3.1 Descrivere che cosa può rendere un grafico fuorviante o ingannevole	50

Capitolo 3 Sintetizzare numericamente i dati 64

3.1 Misure di tendenza centrale	64
3.1.1 Calcolare la media aritmetica	65
3.1.2 Calcolare la mediana	67
3.1.3 Spiegare che cosa s'intende per statistica robusta	68
3.1.4 Calcolare la moda	71
3.2 Misure di dispersione	72
3.2.1 Calcolare l'intervallo di variazione	74
3.2.2 Calcolare la varianza	74
3.2.3 Calcolare lo scarto quadratico medio (deviazione standard)	78
3.2.4 Uso degli intervalli tipici per descrivere una distribuzione a forma campanulare	80
3.2.5 Utilizzare la disuguaglianza di Chebyshev per descrivere una distribuzione	82
3.3 Misure di tendenza centrale e di dispersione per dati raggruppati in classi	82
3.3.1 Approssimare la media per la distribuzione in classi	82
3.3.2 Calcolare la media pesata	84
3.3.3 Approssimare la varianza e la deviazione standard per una distribuzione in classi	85
3.4 Misure di posizione e outlier	87
3.4.1 Determinare e interpretare gli z-score	87
3.4.2 Interpretare i percentili	88
3.4.3 Determinare e interpretare i quartili	89
3.4.4 Determinare e interpretare l'intervallo interquartile	91
3.4.5 Controllare la presenza di outlier in una distribuzione	92

3.5	I cinque numeri della statistica e i boxplot	94
3.5.1	Determinare la sintesi dei 5 numeri	94
3.5.2	Rappresentare e interpretare i boxplot	95
Capitolo 4 Descrivere la relazione tra due variabili		104
4.1	Grafico a dispersione e correlazione	104
4.1.1	Disegnare e interpretare i diagrammi a dispersione	105
4.1.2	Proprietà del coefficiente di correlazione lineare	107
4.1.3	Calcolare e interpretare il coefficiente di correlazione lineare	110
4.1.4	Determinare se esiste una relazione lineare tra due variabili	113
4.1.5	Spiegare la differenza tra correlazione e causalità	113
4.2	Regressione ai minimi quadrati	114
4.2.1	Trovare l'equazione della retta ai minimi quadrati e utilizzarla per effettuare delle previsioni	116
4.2.2	Interpretare il coefficiente angolare e l'intercetta della retta ai minimi quadrati	120
4.2.3	Calcolare la somma dei quadrati dei residui	121
4.3	Coefficiente di determinazione	123
4.3.1	Calcolare e interpretare il coefficiente di determinazione	123
PARTE III PROBABILITÀ E DISTRIBUZIONI DI PROBABILITÀ		133
Capitolo 5 Introduzione alla probabilità		134
5.1	Regole delle probabilità	134
5.1.1	Applicazione delle regole della probabilità	136
5.1.2	Calcolare e interpretare le probabilità utilizzando il metodo empirico	138
5.1.3	Calcolare e interpretare le probabilità utilizzando il metodo classico	140
5.1.4	Uso delle simulazioni per ottenere dati basati sulle probabilità	144
5.2	La regola additiva e il concetto di evento complementare	144
5.2.1	Utilizzare la regola additiva per gli eventi disgiunti	144
5.2.2	Regola generale dell'addizione	147
5.2.3	Calcolo della probabilità di un evento complementare	149
5.3	Regola della moltiplicazione e indipendenza	151
5.3.1	Identificare gli eventi indipendenti	151
5.3.2	Regola della moltiplicazione per eventi indipendenti	152
5.3.3	Calcolo delle probabilità mediante la regola dell'evento complementare	154
5.4	Probabilità condizionata e regola generale della moltiplicazione	155
5.4.1	Calcolare le probabilità condizionate	155
5.4.2	Calcolo delle probabilità utilizzando la regola generale della moltiplicazione	158

5.5	Collegiamo tutto: quale metodo utilizzare?	160
5.5.1	Determinare la regola della probabilità più appropriata da utilizzare	160
5.5.2	Scelta della tecnica più appropriata di calcolo combinatorio	164
Capitolo 6 Come utilizzare le distribuzioni di probabilità discrete		172
6.1	Variabili casuali discrete	172
6.1.1	Distinguere una variabile casuale discreta da una variabile casuale continua	172
6.1.2	Identificare distribuzioni di probabilità discrete	173
6.1.3	Costruire istogrammi di probabilità	175
6.1.4	Calcolare e interpretare la media di una variabile casuale discreta	176
6.1.5	Interpretare la media di una variabile casuale discreta come valore atteso	178
6.1.6	Calcolare la varianza e la deviazione standard di una variabile casuale discreta	179
6.2	Distribuzione di probabilità binomiale	180
6.2.1	Determinare se un esperimento probabilistico è un esperimento binomiale	180
6.2.2	Calcolare media e deviazione standard di una variabile casuale binomiale	187
6.2.3	Costruire istogrammi di probabilità binomiale	188
Capitolo 7 La distribuzione di probabilità normale		194
7.1	Proprietà della distribuzione normale	194
7.1.1	Utilizzo della distribuzione di probabilità uniforme	194
7.1.2	Disegnare la curva normale	196
7.1.3	Proprietà della curva normale	198
7.1.4	Il ruolo dell'area nella funzione di densità normale	198
7.1.5	Descrizione della relazione tra variabile casuale normale e variabile casuale normale standardizzata	202
7.2	Distribuzione normale standardizzata	204
7.2.1	Calcolo dell'area sottesa dalla curva normale standardizzata	205
7.2.2	Calcolare gli z-score per un'area specifica	209
7.2.3	Interpretare l'area sottesa dalla curva normale standardizzata come probabilità	214
7.3	Applicazioni della distribuzione normale	215
7.3.1	Calcolare e interpretare l'area sottesa dalla curva normale	215
7.3.2	Calcolare il valore di una variabile casuale normale	219
7.4	Verificare l'ipotesi di normalità	222
7.4.1	Utilizzo dei grafici di probabilità normale per verificare la normalità	222
7.5	Approssimazione normale della distribuzione di probabilità binomiale	225
7.5.1	Approssimazione delle probabilità binomiali utilizzando la distribuzione normale	225

PARTE IV INFERENZA: DAI CAMPIONI ALLA POPOLAZIONE 233

Capitolo 8 Distribuzioni campionarie 234

8.1 Distribuzione della media campionaria	234
8.1.1 Descrivere la distribuzione della media campionaria: campioni da popolazioni normali	235
8.1.2 Descrivere la distribuzione della media campionaria: campioni da una popolazione che non è normale	240
8.2 Distribuzione della frequenza relativa	246
8.2.1 Descrivere la distribuzione della frequenza relativa campionaria	246
8.2.2 Calcolare le probabilità della frequenza relativa campionaria	250

Capitolo 9 Come utilizzare gli intervalli di confidenza per media e percentuale 256

9.1 Logica nella costruzione degli intervalli di confidenza per la media della popolazione quando la deviazione standard è nota	256
9.1.1 Calcolare una stima puntuale della media della popolazione	257
9.1.2 Costruire e interpretare un intervallo di confidenza per la media della popolazione	257
9.1.3 Spiegare il ruolo del margine di errore nella costruzione di un intervallo di confidenza	267
9.1.4 Determinare l'ampiezza campionaria necessaria a stimare la media della popolazione all'interno di uno specifico margine di errore	268
9.2 Intervalli di confidenza per la media della popolazione quando la deviazione standard non è nota	270
9.2.1 Proprietà della distribuzione t di Student	270
9.2.2 Determinare i valori di t	272
9.2.3 Costruire e interpretare un intervallo di confidenza per la media della popolazione	273
9.3 Intervalli di confidenza per la frequenza relativa della popolazione	275
9.3.1 Ottenere una stima puntuale per la frequenza relativa della popolazione	276
9.3.2 Costruire e interpretare un intervallo di confidenza per la frequenza relativa della popolazione	276
9.3.3 Determinare l'ampiezza campionaria necessaria per stimare la frequenza relativa della popolazione avendo fissato uno specifico margine di errore	279
9.4 Quale procedura usare?	281
9.4.1 Determinare l'intervallo di confidenza più adatto	281

Capitolo 10	Come utilizzare la verifica di ipotesi	288
10.1	Linguaggio della verifica di ipotesi	288
10.1.1	Determinare l'ipotesi nulla e l'ipotesi alternativa	289
10.1.2	Spiegare l'errore di I e II tipo	292
10.1.3	Formulare le conclusioni della verifica di ipotesi	294
10.2	Verifica di ipotesi per la media della popolazione con deviazione standard della popolazione nota	295
10.2.1	Spiegare la logica della verifica di ipotesi	295
10.2.2	Verifica di ipotesi sulla media di una popolazione con σ noto utilizzando il metodo classico	299
10.2.3	Verifica di ipotesi sulla media di una popolazione con σ noto utilizzando il metodo <i>p-value</i>	303
10.2.4	Distinguere tra significatività statistica e significatività pratica	308
10.3	Verifica di ipotesi per la media della popolazione con deviazione standard della popolazione non nota	309
10.3.1	Verifica di ipotesi per la media della popolazione con σ non noto	309
10.4	Verifica di ipotesi per una proporzione della popolazione	316
10.4.1	Verifica di ipotesi sulla proporzione della popolazione	316
10.5	Quale metodo usare?	321
10.5.1	Determinare la verifica di ipotesi appropriata	321
Capitolo 11	Come utilizzare il test chi-quadrato	328
11.1	Test per la bontà di adattamento	328
11.1.1	Effettuare un test per la bontà di adattamento	328
11.2	Test di indipendenza	337
11.2.1	Effettuare un test di indipendenza	337
Appendice A	Tabelle	350
Indice analitico		364

Prefazione all'edizione italiana

Lo studio della statistica in ambito umanistico non sempre è affrontato con serenità dagli studenti, i quali in gran misura conservano ricordi ostili nei confronti delle discipline di calcolo e trasferiscono questi preconcetti allo studio della statistica.

Questo testo è dedicato a tutti gli studenti che affrontano con spirito di sacrificio lo studio della disciplina della statistica, ma che al contempo sono desiderosi di maneggiare i “numeri” e analizzare in modo indipendente la realtà circostante.

L'esperienza didattica mi ha insegnato che si ottengono risultati eccellenti se la sensibilità di questi studenti è adeguatamente coniugata alla metodologia statistica. Tuttavia, un classico manuale non sempre riesce a supportare questi successi poiché può risultare molto arduo bilanciare aspetti teorici, aspetti applicativi, casi di studio, approfondimenti ulteriori con le conoscenze pregresse degli studenti. *Fondamenti di Statistica* si configura non solo come un equilibrato compromesso tra gli aspetti precedenti, ma fornisce un metodo di studio molto simile a quello utilizzato dagli studenti.

Le esigenze didattiche degli studenti rappresentano il nucleo attorno al quale è stato sviluppato il libro e sono state sintetizzate nell'adozione di: una ridotta formalizzazione matematica, che contribuisce ad abbattere tutte le barriere psicologiche preesistenti; un metodo di studio *a step*, molto utile nella fase di comprensione; un ricorso al simbolismo per indicare i concetti più importanti e l'utilizzo di Excel, dettagli molto importanti in fase di apprendimento; un'ampia scelta di diversi esempi legati alla realtà italiana, sviluppati su dati provenienti da Istat, Banca d'Italia e Reuters; una verifica dell'apprendimento, sviluppata seguendo un approccio gerarchico e declinata in quattro livelli: riconoscere i concetti, applicare i concetti, consolidare l'apprendimento e collegare i concetti.

Fondamenti di Statistica, più che un classico manuale di statistica dedicato agli argomenti cardini della statistica descrittiva e inferenziale, può essere definito uno strumento di apprendimento e come tale dedica ampio spazio agli esercizi (per la maggior parte provvisti di soluzione, aspetto molto gradito agli studenti) disponibili su formato cartaceo e on-line (sito web <http://hpe.pearson.it/sullivan>) dove sono riportati esercizi aggiuntivi, casi di studio e dataset da utilizzare con i principali software statistici. Sempre in questa direzione segnalo l'opportunità di autoverifica on-line dell'apprendimento attraverso MyStatLab.

Desidero infine ringraziare tutti coloro, traduttori e staff della Pearson, che con impagabile pazienza mi hanno supportato in questo lungo percorso di adattamento e revisione, resta inteso che la responsabilità dello scritto è unicamente mia, per cui sarò molto grata a chiunque mi segnalerà refusi e incongruenze.

Emma Zavarrone
Milano, 1 luglio 2011

Prefazione

Per i docenti

Un libro di testo per spiegare una materia importante e interessante

La statistica è una materia importante ed è una delle mie passioni. La passione per questa materia e il desiderio di creare un manuale utile per me, i miei studenti e la mia scuola mi hanno indotto a scrivere la prima edizione di questo testo. Le motivazioni iniziali continuano a sostenere il mio lavoro di revisione del testo per riflettere i cambiamenti avvenuti negli studenti, nella comunità statistica e nel mondo che ci circonda.

Quando ho iniziato a scrivere, utilizzavo i manoscritti di questo testo in classe. I miei studenti fornivano preziosi e appropriati consigli; io apportavo le opportune modifiche sulla base dei loro commenti. Per molti aspetti, questo testo è stato scritto dagli studenti e per gli studenti. Ho ricevuto anche suggerimenti costruttivi da numerosi docenti di statistica, che mi hanno permesso di affinare l'esposizione dei concetti nel libro e nelle mie lezioni. Continuo a ricevere preziosi suggerimenti da docenti e studenti, consentendo a questo testo di perfezionarsi con spiegazioni più chiare e concise per aiutare gli studenti a pensare in modo statistico.

Nelle ultime due edizioni, ho compiuto uno sforzo speciale per supportare le direttive GAISE (*Guidelines for Assessment and Instruction in Statistics Education*) per il corso introduttivo ai college sostenuto dall'ASA (*American Statistical Association*). Il rapporto GAISE include sei raccomandazioni per il corso:

1. dare enfasi alla letteratura statistica e sviluppare il pensiero statistico;
2. utilizzare dati reali nell'insegnamento della statistica;
3. ribadire i concetti;
4. promuovere l'apprendimento attivo;
5. utilizzare le tecnologie informatiche per sviluppare la comprensione dei concetti;
6. utilizzare la verifica per migliorare e valutare l'apprendimento degli studenti.

La nuova edizione del libro rispecchia queste importanti direttive.

Strumenti didattici

- **Che cosa ci serve sapere.** Questa sezione, che si trova all'inizio dei paragrafi principali, indica agli studenti gli argomenti necessari per capire i temi che si accingono a studiare.
- **Esempi.** Esempi interessanti forniscono spiegazioni chiare e concise agli studenti secondo lo schema collaudato *Problema, Metodo e Soluzione*. Il *Problema* presenta lo scenario dell'esempio; il *Metodo* descrive il processo alla base degli strumenti utilizzati per risolvere il problema; la *Soluzione* è quella che si ottiene utilizzando gli strumenti descritti nel metodo.
- **Esempi step-by-step.** Questi esempi guidano gli studenti a trovare la soluzione di un problema attraverso 3 semplici passi da seguire.
- **Risolvi il problema.** Queste indicazioni, che si trovano alla fine di molti esempi, suggeriscono agli studenti un problema appropriato ai concetti esposti.
- **Excel step-by-step.** Per alcuni esempi viene fornita la procedura dettagliata per trovare la soluzione con Excel.
- **Riepilogo e verifiche.** Questa sezione alla fine di ogni capitolo include il riassunto del capitolo, un elenco delle formule utilizzate nel capitolo, una serie di problemi le cui soluzioni sono disponibili sul sito web.
- **Problemi alla fine del capitolo.** Questi problemi, di difficoltà crescente, sono suddivisi in varie categorie: riconoscere i concetti, consolidare l'apprendimento, applicare i concetti e collegare i concetti.
- **Collegare i concetti.** Questa serie di esercizi aiuta gli studenti a collegare i temi appena studiati con quelli precedenti.
- **Casi di studio.** Per ogni capitolo è disponibile on-line un caso di studio che aiuta gli studenti ad applicare le loro conoscenze a casi reali.

Flessibilità ai vari corsi di studio

Per soddisfare le esigenze dei vari corsi di studio, questo libro è stato organizzato per essere flessibile. Gli argomenti elencati sotto il titolo "Che cosa ci serve sapere" all'inizio di ciascun paragrafo indicano le interdipendenze tra i vari argomenti. Le due più comuni varianti all'interno di un corso introduttivo di statistica sono l'analisi della regressione e la teoria della probabilità.

- **Correlazione e regressione.** Il testo è stato scritto con la parte che descrive i dati bivariati (Capitolo 4) dopo quella che descrive i dati univariati (Capitolo 3). I docenti che preferiscono posporre la discussione sui dati bivariati, possono ignorare temporaneamente il Capitolo 4 per riprenderlo dopo aver trattato gli argomenti del Capitolo 11.
- **Probabilità.** Il testo consente di trattare la probabilità in modo più o meno approfondito. I docenti che desiderano minimizzare la teoria della probabilità possono limitarsi a trattare il Paragrafo 5.1, ignorando i successivi paragrafi. Per un trattamento un po' più approfondito, utilizzare i Paragrafi 5.1-5.3. I docenti che volessero trattare il test chi-quadrato per l'indipendenza, utilizzeranno i Paragrafi 5.1-5.3. Infine, i docenti che volessero trattare le probabilità binomiali, tratteranno l'indipendenza nel Paragrafo 5.3 e le combinazioni nel Paragrafo 5.5.

Collegare i concetti

Quando gli studenti iniziano a studiare la statistica, spesso fanno fatica ad avere una visione d'insieme di tutti i concetti. Uno degli obiettivi di questo libro consiste nell'aiutare gli studenti non solo a imparare importanti concetti e metodi della statistica, ma anche a collegarli insieme. Questa serie di esercizi indicati nella tabella aiuta gli studenti a collegare i temi appena studiati.

Capitolo	Problema	Concetti utilizzati
1	1.37	Campionamento causale semplice e altre tecniche di campionamento
1	1.38	Campionamento casuale
1	1.39	Popolazione, variabili, scala di misurazione, campionamento, errore di campionamento
2	2.28	Rappresentazione grafica dei dati
2	2.29	Rappresentazione grafica dei dati
3	3.37	Dati continui e discreti, istogrammi, forma di una distribuzione, media, mediana, moda
4	4.30, 4.31	Grafici a dispersione, correlazione, regressione lineare
5	5.37	Variabili, sintesi grafica dei dati, esperimenti, probabilità
5	5.38	Variabili, distribuzioni di frequenze relative, grafici a barre, media, deviazione standard, probabilità
7	7.37	Distribuzione di frequenze relative, istogrammi, media e deviazione standard di dati raggruppati, probabilità normali
8	8.18	Distribuzioni di probabilità, media e deviazione standard di una variabile casuale, distribuzioni campionarie
9	9.26	Errore di campionamento, intervalli di confidenza
10	10.30	Grafici a barre, intervalli di confidenza, verifica di ipotesi
10	10.31	Istogramma di frequenze relative, moda, media, deviazione standard, teorema del limite centrale, verifica di ipotesi
11	11.10	Test per la bontà di adattamento, test di indipendenza

Supplementi on-line



All'indirizzo <http://hpe.pearson.it/sullivan>, nella sezione Companion WebSite, sono disponibili numerosi materiali didattici a corredo del testo:

- molti **esercizi aggiuntivi**;
- le **soluzioni** di alcuni degli esercizi proposti sia sul libro sia on-line;
- **dataset** in formato .txt, che possono essere aperti sia con Excel sia con Spss;
- **casi di studio** in inglese;
- **argomenti aggiuntivi** in italiano e in inglese;
- 20 **applet**, a cui sono collegati alcuni esercizi del libro, per aiutare gli studenti a capire gli argomenti fondamentali;
- il **plug-in DDXL** per Excel che aiuta a risolvere alcuni esempi.

Guida alla lettura

Decidere informati è un problema pratico posto a inizio capitolo per mettere in risalto quelli che sono gli argomenti introdotti di seguito e utili alla sua risoluzione.

CAPITOLO 1

Raccolta delle informazioni

DECIDERE INFORMATI Introduzione

Quale film dovremmo andare a vedere?
È lunedì mattina e stiamo già pensando al film di venerdì sera. Le recensioni dei critici ci sembrano tutte molto severe pertanto decidiamo di basarci sul parere delle persone che hanno già visto il film. A tal fine predisponiamo una lista di domande (questionario) da somministrare a chi ha già visto il film. Questo capitolo ci fornirà gli strumenti essenziali per scoprire se le risposte fornite dai nostri conoscenti hanno confermato le recensioni espresse dai critici. Si rimanda al paragrafo conclusivo alla fine del capitolo.

Sommario del capitolo

- 1.1 Primi passi verso la statistica
- 1.2 Campionamento casuale semplice
- 1.3 Altri tipi di campionamento
- 1.4 Errore di campionamento

1.1 Primi

L'intera trattazione si basa sull'uso di **esempi** in cui vengono schematizzati i dati del **Problema**, il **Metodo** e la **Soluzione**.

ESEMPIO 2.12

Disegnare un grafico di una serie storica

Problema: i dati nella Tabella 2.16 rappresentano il prezzo di chiusura delle azioni della *Tod's* per ogni mese, a partire da gennaio 2006 fino a dicembre 2007. Dobbiamo rappresentare la serie storica con un grafico adeguato.

Metodo

Step 1: disegniamo i punti per ciascun mese, indicando la data sull'asse orizzontale e il corrispondente prezzo di chiusura sull'asse verticale.

Step 2: colleghiamo i punti con segmenti rettilinei.

Soluzione: la Figura 2.10 mostra il grafico della serie storica. Il trend nei due anni analizzati non promette bene per gli investitori in azioni *Tod's*; in dettaglio si evidenzia che dopo il mese di luglio 2007 è possibile notare una preoccupante inversione dell'andamento del prezzo.

Un commento a latere **sintetizza** i concetti esposti.

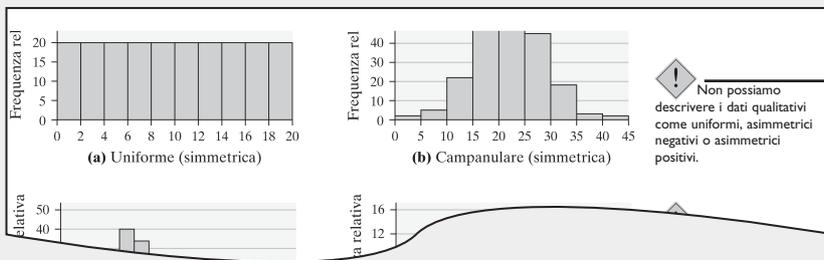
In sintesi
Se contiamo per ottenere il valore di una variabile quantitativa, questa è discreta. Se misuriamo per ottenere il valore di una variabile quantitativa, questa è continua. Quindi, per decidere se una variabile è discreta o continua, chiediamoci se dobbiamo contare o misurare.

(a) Il numero di teste ottenuto lanciando cinque volte una moneta non tr
(b) Il numero di automobili che arrivano al McDrive tra le 12:00 e le 13:0
(c) La distanza che una Toyota Prius 2007 può percorrere con un pieno d

Metodo: una variabile è discreta se i suoi valori risultano da un conteggio, continua se il suo valore è misurato.

Soluzione

(a) Il numero di teste ottenute lanciando cinque volte una moneta non tr
riabile discreta, perché contiamo le teste ottenute. I valori possibili che qu
assumere (ovvero il do
(b) Il numero di



Non possiamo descrivere i dati qualitativi come uniformi, asimmetrici negativi o asimmetrici positivi.

Una icona a latere focalizza l'attenzione su alcuni aspetti fondamentali.

Metodo

Step 1: disegniamo i punti per ciascun mese, indicando la data sull'asse orizzontale e il corrispondente prezzo di chiusura sull'asse verticale.

Step 2: colleghiamo i punti con segmenti rettilinei.

Soluzione: la Figura 2.10 mostra il grafico della serie storica. Il trend nei due anni analizzati non promette bene per gli investitori in azioni Tod's; in dettaglio si evidenzia che dopo il mese di luglio 2007 è possibile notare una preoccupante inversione dell'andamento del prezzo.

Risolvi il Problema 2.24

Per stimolare lo studente a mettere in pratica subito i concetti appresi, spesso gli esempi sono seguiti da un riferimento al **problema** attinente all'argomento studiato.

vengono indicati come **serie storica**. Il prezzo di chiusura delle azioni della Tod's rilevato mensilmente da gennaio 2006 a dicembre 2007 rappresenta un esempio di serie storica.

Il grafico di una serie storica si ottiene indicando sull'asse orizzontale il tempo in cui la variabile è misurata e sull'asse verticale il corrispondente valore della variabile, rilevato a quella data. Poi si uniscono i punti tramite segmenti rettilinei.

Definizione

Rappresentare graficamente una serie storica è molto utile per identificare l'andamento o tendenza di fondo (*trend*) dei dati nel tempo.

Le **definizioni** sono messe ben in evidenza nel testo.

Excel step-by-step

Ottenere un campione casuale semplice

Step 1: Attivate il pacchetto degli **Strumenti di analisi**. Per farlo, aprite il menu **Strumenti** e scegliete **Componenti aggiuntivi**. Selezionate la casella accanto a **Strumenti di analisi** e fate clic su OK.

Step 2: Aprite il menu **Strumenti** e selezionate **Analisi dati**. Selezionate **Generazione di un numero casuale** e fate clic su OK.

Per alcuni esempi viene fornita la procedura dettagliata per trovare la **soluzione con Excel**. I dataset sono disponibili sul sito web.

A chiusura del capitolo le **conclusioni** al problema posto inizialmente.

DECIDERE INFORMATI ● Conclusione

Quale film dovremmo andare a vedere?

Redigere una lista di domande non è semplice; è uno dei compiti più difficili nella costruzione di un questionario. Le domande devono essere formulate senza che gli intervistati possano fraintenderle. Inoltre, le domande dovrebbero essere espresse in maniera tale da agevolare il ricercatore a condurre analisi scientificamente valide. Desideriamo creare un questionario che possa essere usato per prendere una decisione informata per vedere un particolare film. Selezioniamo un film che desideriamo vedere. Se è ancora presente nelle sale, assicuriamoci che da almeno un paio di settimane sia in cartellone, in modo che sia stato visto da un certo numero di persone. Potremmo includere domande riguardanti la demografia degli intervistati (come l'età, il genere, il livello di istruzione e così via). Poniamo tante domande quante crediamo siano necessarie per ottenere un'opinione sul film. Le domande possono essere aperte o chiuse. Distribuiamo il questionario almeno a 20 persone che hanno visto il film, selezionandole a caso. Mentre distribuiamo il questionario, teniamo traccia di quegli individui che non hanno visto il film. In particolare teniamo traccia delle loro informazioni demografiche. Successivamente, procediamo a riassumere i risultati della ricerca. Sulla base di tali risultati,

andremo a vedere il film? Perché? Ora guardiamo il film. Ci è piaciuto? L'indagine ha previsto in maniera accurata il nostro livello di gradimento nei riguardi del film? Ora rispondiamo alle seguenti domande.

- (a) Che metodo di campionamento abbiamo utilizzato? Perché? Avevamo una lista della popolazione?
- (b) Abbiamo avuto qualche problema di fraintendimento delle nostre domande da parte degli intervistati? Come si può risolvere questa questione?
- (c) Che ruolo ha avuto la demografia degli intervistati nel condizionare la nostra opinione? Perché?
- (d) Ha giocato un ruolo la demografia degli individui che non hanno visto il film mentre stavamo decidendo di andare al cinema?
- (e) Consultiamo la recensione di un film scritta da un critico cinematografico. La sua opinione è in accordo con la nostra? A quali fattori potrebbero essere associati gli eventuali elementi di somiglianza o differenza nelle nostre opinioni rispetto a quelle del critico?
- (f) Descriviamo i problemi riscontrati nel distribuire il questionario. Se dovessimo riproporre questa indagine, cambieremo qualcosa? Perché?

Questa sezione alla fine di ogni capitolo include il **riassunto** del capitolo, un elenco delle **formule** utilizzate nel capitolo, una serie di **problemi** le cui soluzioni sono disponibili sul sito web <http://hpe.pearson.it/sullivan>.

Riepilogo e verifiche

Riassunto

Abbiamo definito la statistica come la scienza attraverso cui i dati vengono raccolti, organizzati, riassunti e analizzati in modo da dedurre le peculiarità riguardanti la popolazione. La statistica fornisce anche una misura di confidenza sulle conclusioni che traiamo. La statistica descrittiva si occupa di organizzare e riassumere le informazioni, mentre la statistica inferenziale trae conclusioni su una popolazione basandosi sui risultati di un campione. La popolazione è un insieme di unità statistiche delle quali si desidera ottenere informazioni specifiche e il campione è un sottoinsieme della popolazione.

I dati o modalità sono le osservazioni di una variabile e possono essere di natura qualitativa o quantitativa. I dati quantitativi sono discreti o continui.

Possiamo ottenere i dati da quattro diverse fonti: censimento, fonti esistenti, studi osservazionali o disegno degli esperimenti (che non abbiamo approfondito in questo testo). Un censimento elencherà tutti gli individui di una popolazione, congiuntamente con certe caratteristiche. Poiché i costi di un censimento sono elevati, molti ricercatori opteranno per il campione.

Abbiamo introdotto cinque metodi di campionamento: casuale semplice, stratificato, sistematico, a grappolo e non probabilistico. Tutti i metodi di campionamento, eccetto quello non probabilistico, permettono di applicare la statistica inferenziale. Il campionamento non probabilistico porta tipicamente a un campione non rappresentativo e a risultati distorti.

Questi **problemi**, di difficoltà crescente, sono suddivisi in varie categorie: riconoscere i concetti, consolidare l'apprendimento, applicare i concetti e collegare i concetti.

Problemi

Riconoscere i concetti

- 1.1 Definire la statistica.
- 1.2 Spiegare la differenza tra popolazione e campione.
- 1.3 Un _____ è una persona o un oggetto che è membro della popolazione che si sta studiando.
- 1.4 La statistica _____ consiste nell'organizzare e riassumere i dati, mentre l' _____ statistica generalizza i risul-

tati ottenuti per un campione e li estende all'intera popolazione misurandone l'affidabilità.

- 1.5 Una _____ è una sintesi numerica del campione.
- 1.6 Un _____ è una sintesi numerica della popolazione.
- 1.7 Spiegare la differenza tra variabili qualitative e quantitative.
- 1.8 Spiegare la differenza tra variabili continue e discrete.

tra l'altezza di un bambino e la sua circonferenza della testa. A tale proposito seleziona, tra i suoi pazienti, 11 bambini di 3 anni e per ognuno di essi registra l'altezza e la circonferenza della testa.

Altezza (cm)	Circonferenza della testa (cm)	Altezza (cm)	Circonferenza della testa (cm)
70.485	44.45	67.31	43.942
62.23	43.434	65.58	44.45
64.77	43.434	67.945	43.942
66.04	43.942	67.495	44.45
63.5	42.926	69.85	44.45
70.485	44.704		

Fonte: Denise Slucki, studente al Joliet Junior college.

- Se il pediatra volesse usare l'altezza per prevedere la circonferenza della testa, quale sarebbe la variabile esplicativa e quale sarebbe la variabile risposta?
- Disegna il grafico a dispersione.
- Calcola il coefficiente di correlazione lineare tra l'altezza e la circonferenza della testa dei bambini.
- Esiste una relazione lineare tra le variabili considerate?

Collegare i concetti

- 4.26 Avvia l'applet per la correlazione.
- Aggiungi 10 punti nell'angolo in basso a sinistra dell'applet in modo che siano allineati con un coefficiente angolare positivo e la correlazione lineare tra i punti sia circa pari a 0.8. Fai clic sul pulsante "Mostra r" per vedere il coefficiente di correlazione.
 - Aggiungi un altro punto nell'angolo in alto a destra dell'applet in modo che sia approssimativamente in linea con i 10 punti precedentemente posizionati nell'angolo in basso a sinistra. Descrivi come cambia il coefficiente di correlazione lineare.
 - Trascina verso il basso il punto posizionato in alto a destra. Verifica come si è modificato il coefficiente di correlazione lineare e commenta l'impatto che un singolo punto può avere sul coefficiente di correlazione.
- 4.27 Avvia l'applet per la correlazione. Aggiungi 10 punti in modo da formare una U rovesciata. Certamente esiste una correlazione tra x e y , ma qual è il valore del coefficiente di correlazione lineare? Questo esempio ci fa concludere che un coefficiente di correlazione basso non significa che non esiste alcuna relazione tra due variabili, ma solamente che non esiste una relazione *lineare* tra le due variabili.
- 4.28 Avvia l'applet per la correlazione.
- Aggiungi 10 punti in modo che seguano un andamento

La serie di esercizi **Collegare i concetti** aiuta gli studenti a collegare i temi appena studiati con quelli precedenti.

Gli esercizi contrassegnati con l'icona  fanno riferimento a una delle 20 **applet** disponibili on-line, progettate per aiutare gli studenti a capire numerosi argomenti fondamentali della statistica.



Le soluzioni di alcuni degli esercizi proposti nel testo e numerosi altri problemi riepilogativi e dedicati a specifici argomenti sono disponibili on-line all'indirizzo <http://hpe.pearson.it/sullivan>. Tra questi anche i problemi che fanno riferimento ai dataset in formato .txt per l'uso in Excel e Spss.

Con questo **richiamo on-line** si ricorda allo studente di consultare il sito web <http://hpe.pearson.it/sullivan> dove compaiono, oltre alla soluzione di alcuni degli esercizi, molti altri utili materiali.

CONSUMER REPORTS® Strato impermeabile per scantinati

Uno strato impermeabilizzante può essere un sistema economico e facile per far fronte all'umidità negli scantinati. Ma quanto è efficace? In uno studio, la rivista *Consumer Reports* ha analizzato nove isolanti per stabilire la loro efficacia nel controllare le infiltrazioni di acqua attraverso le fondamenta in cemento.

Per confrontare la capacità dei prodotti di controllare le infiltrazioni di acqua, sono state applicate due mani di ciascun prodotto a lastre ricavate da un blocco di cemento. Per verificarne la validità statistica, il procedimento è stato ripetuto almeno sei volte. In ogni esperimento, sono stati posti simultaneamente quattro blocchi (ognuno ricoperto con un prodotto diverso) in una camera di alluminio rettangolare. Vedi la figura successiva.

La camera è stata sigillata e riempita d'acqua, sottoponendo i blocchi a pressioni idrostatiche crescenti. L'acqua che usciva a ogni aumento di pressione è stata incanalata alla base dell'apertura della camera, raccolta e pesata.

La tabella successiva contiene un campione di dati raccolti per due dei prodotti analizzati. Usando questi dati:

- calcolare la media, la mediana e la moda pesate dell'acqua raccolta per il prodotto A;
- calcolare la deviazione standard pesata dell'acqua raccolta per il prodotto A;
- calcolare la media, la mediana e la moda pesate dell'acqua raccolta per il prodotto B;
- calcolare la deviazione standard pesata dell'acqua raccolta per il prodotto B.

Ti sembra che ci sia una differenza nella capacità di questi due prodotti di controllare le infiltrazioni di acqua? Perché?

Nota ai lettori: in molti casi il protocollo sperimentale e i metodi di analisi sono molto più complessi di quelli descritti in questi esempi. I dati e le descrizioni sono semplificati per rendere il materiale più appropriato ai lettori.



Prodotto	Ripetizione	Peso dell'acqua raccolta (grammi)
A	1	91.2
A	2	91.2
A	3	90.9
A	4	91.3
A		90.8

Un caso reale analizzato dalla rivista americana **Consumer Reports** mostra come utilizzare ciò che si è appreso nella pratica quotidiana.

Introduzione alle applet

Le 20 applet disponibili on-line sono state progettate per aiutare gli studenti a capire numerosi argomenti fondamentali della statistica. Nella terza edizione, le applet sono state notevolmente migliorate modificando l'interfaccia utente e aggiungendo nuove funzionalità. La dimensione dei caratteri è stata aumentata per consentire di proiettare con maggior chiarezza le applet nelle aule.

- L'applet **Campione da una popolazione** consente all'utente di selezionare campioni di varie dimensioni da un'ampia gamma di forme di popolazioni, incluse le popolazioni uniformi, campanulari, asimmetriche e binarie (che includono un intervallo di valori per la probabilità di un 1). Gli studenti possono modificare qualsiasi popolazione di default per creare una distribuzione personalizzata trascinando il mouse sulla popolazione. I piccoli campioni vengono creati con un'animazione per aiutare gli studenti a capire il concetto fondamentale del campionamento. I campioni più grandi vengono creati senza animazione in modo che le caratteristiche di questi campioni possano essere rapidamente confrontate con quelle della popolazione.
- L'applet **Distribuzioni campionarie** è una variante della precedente applet che permette di aggiungere i valori delle statistiche scelte dall'utente per ciascun campione. Gli studenti possono esaminare le distribuzioni risultanti e vedere come le caratteristiche della distribuzione campionaria, quali il centro e la dispersione, siano influenzate dalla dimensione campionaria e dalla forma della popolazione. Gli studenti possono confrontare inoltre le distribuzioni campionarie con statistiche differenti, quali la media e la mediana campionarie.
- L'applet **Numeri casuali** consente agli studenti di scegliere un campione casuale da un intervallo di valori interi definito dall'utente. Gli studenti possono utilizzare l'applet per studiare la probabilità di base considerando la frequenza relativa di particolari risultati tra i campioni. Possono anche scegliere i campioni da una lista di valori per un campionamento interattivo.
- Le sei applet **Esperimenti di probabilità ripetuti molte volte** simulano il lancio di un dado, di una moneta e le fluttuazioni dei mercati azionari. Gli studenti possono scegliere il numero di simulazioni, con o senza animazione. La frequenza relativa dell'evento che interessa viene rappresentata in un diagramma in funzione del numero di simulazioni. All'aumentare del numero di simulazioni, la convergenza della frequenza relativa della probabilità reale dell'evento diventerà evidente.
- L'applet **Media e mediana** consente agli studenti di costruire un dataset in modo interattivo facendo clic su un grafico che visualizza la media e la mediana dei dati, consentendo agli studenti di analizzare gli effetti della forma e degli outlier sulla media e sulla mediana. L'applet **deviazione standard** offre un analogo tipo di analisi. La particolare forma impilata dell'applet consente di confrontare facilmente dataset con deviazioni standard differenti.
- Tre applet aiutano gli studenti a capire gli intervalli di confidenza. L'applet **Intervalli di confidenza per una proporzione** consente agli studenti di simulare intervalli di confidenza al 95% e al 99% per la proporzione di una popolazione. Gli intervalli di confidenza sono rappresentati in un grafico che illustra la loro relazione in funzione dell'ampiezza e della loro natura casuale. La dimensione campionaria e la proporzione reale sottostante sono specificate dall'utente, che può così analizzare l'effetto di questi valori sulle probabilità fornite dall'applet. Due applet **Intervalli di confidenza per una media** consentono di svolgere analisi analoghe. La prima può essere utiliz-

zata per capire l'impatto che la dimensione campionaria e la forma della distribuzione hanno sulle performance degli intervalli t della media. La seconda consente agli studenti di confrontare le performance degli intervalli t e z per varie forme di distribuzioni e dimensioni campionarie.

- Le applet **Verifica di ipotesi per una proporzione** e **Verifica di ipotesi per una media** consentono agli studenti di capire come le ipotesi di base influenzano le performance delle verifiche di ipotesi. Queste applet rappresentano in un grafico le statistiche delle verifiche e i corrispondenti p -value per i dati generati sotto le varie condizioni specificate dall'utente. Le proporzioni che rifiutano l'ipotesi nulla consentono agli studenti di determinare come le condizioni specificate influiscono sul livello reale di significatività delle verifiche.
- L'applet **Correlazione a occhio** consente agli studenti di stimare il valore del coefficiente di correlazione in base a un grafico a dispersione dei dati simulati. Analogamente, l'applet **Regressione a occhio** consente di valutare interattivamente la retta di regressione per i dati simulati.
- L'applet **Distribuzione binomiale** genera dei campioni dalla distribuzione binomiale che ha i parametri specificati dall'utente. Variando i parametri, gli studenti possono capire come questi parametri influiscono sulla distribuzione binomiale.
- L'applet **Let's Make a Deal** permette agli studenti di simulare i risultati del gioco di un popolare show televisivo degli anni '70 negli USA.

Applet 	Capitolo	Attività/Problema sul libro
Campione da una popolazione	1	Utilizzare l'applet per aiutare gli studenti a capire come i campioni vengono ottenuti dalle popolazioni. L'applet può essere utilizzata anche per capire la variabilità di campionamento.
Media e mediana	3	Problemi 3.35 e 3.36
Deviazione standard	3	Utilizzare l'applet per aiutare gli studenti a capire questa misura di dispersione.
Correlazione a occhio	4	Problemi 4.26, 4.27 e 4.28
Regressione a occhio	4	Problema 4.29
Simulare la probabilità di ottenere testa lanciando una moneta non truccata	5	Problema 5.36 punti (a)-(d)
Simulare la probabilità di ottenere testa lanciando una moneta truccata	5	Problema 5.36 punto (e)
Distribuzione binomiale	6	Problema 6.19
Distribuzioni campionarie	8	Problemi 8.15, 8.16, 8.17 e 8.19
Intervalli di confidenza per una media (impatto del livello di confidenza)	9	Problemi 9.15, 9.16 e 9.17
Intervalli di confidenza per una media (impatto della deviazione standard non nota)	9	Problema 9.18
Intervalli di confidenza per una proporzione	9	Problemi 9.20 e 9.21
Verifica di ipotesi per una media	10	Utilizzare l'applet per illustrare i concetti di verifica di ipotesi e di p -value e le conseguenze di non soddisfare i requisiti del modello.
Verifica di ipotesi per una proporzione	10	Utilizzare l'applet per illustrare i concetti di verifica di ipotesi e di p -value e le conseguenze di non soddisfare i requisiti del modello.

Avviso ai lettori

Si segnala che:

- le cifre decimali sono separate dal punto anziché dalla virgola secondo la notazione anglosassone;
- nei numeri decimali le cifre dopo il punto di norma sono tre, ma in alcuni capitoli sono state ridotte a due in conformità con il testo originale; dove non è espressamente indicato procedere con tre cifre decimali.