**Dipartimento di Giurisprudenza ed Economia**

*Corso di Laurea in Scienze Economiche*

# Statistica Economica

a.a. 2016/17

Prof. Massimo Finocchiaro Castro

massimo.finocchiaro@unirc.it

# Programma

- Introduzione e richiami (Capitolo 1)
  - Domande economiche e dati economici

- Richiami di probabilità (Capitolo 2)

- Richiami di statistica (Capitolo 3)

- Elementi fondamentali dell'analisi di regressione (Capitoli 4, 5, 6, 7,8,9)
  - Regressione lineare con un singolo regressore
  - Regressione lineare con un singolo regressore: verifica di ipotesi e intervalli di confidenza
  - Regressione lineare con regressori multipli
  - Verifica di ipotesi ed intervalli di confidenza nella regressione multipla
  - Funzioni di regressioni non lineari
  - Valutazioni di studi basati sulla regressione multipla

- Regressione con dati panel (Capitolo 10)

<u>Testo di riferimento</u>:

Stock, J.H. e Watson, M.W. (2016). Introduzione all'Econometria, 4/ed., Pearson Education, Milano.

# Posologia

- Lezioni: lunedì 9-11; Mercoledì 9-11; Giovedì 9-11.
- Ricevimento: Mercoledì 13-14.30.
- Alla fine di ogni settimana saranno assegnati degli esercizi da correggere in classe la settimana successiva.
- Modalità di esame: Prova Scritta (domande ed esercizi).
- Sarà effettuata una prova intermedia (14/11) ed uno scritto finale (12/12).
- Saranno previsti dei questionari da svolgere in classe durante le lezioni a fini puramente auto-valutativi.

# Controindicazioni

- Corso non adatto a “furbi”, svogliati e amanti della chiacchiera.
- È obbligatorio il silenzio durante le spiegazioni.
- Si richiede la partecipazione alle discussioni ed almeno la tentata risoluzione degli esercizi.
- Si consiglia la frequenza e la partecipazione alle prove intermedie.

# Econometria

- Scienza che sottopone a verifica le teorie economiche
- Insieme di strumenti usati ai fini della previsione
- Adattamento di modelli economico-matematici ai dati del mondo reale
- Scienza che usa i dati storici per fare raccomandazioni di politica economica per il governo o gli affari

La teoria economica suggerisce interessanti relazioni, spesso con implicazioni di politica economica, ma quasi mai è in grado di segnalare la dimensione quantitativa degli effetti causali

- Quale è l’elasticità di prezzo delle sigarette?
- Quale è l’effetto della riduzione della dimensione delle classi sul rendimento di uno studente?
- Quale è l’effetto dell’istruzione sul reddito
- Quale è l’effetto dell’aumento di un punto percentuale del tasso d’interesse fissato dalla BCE sul tasso di crescita dell’Italia?

- La nostra attenzione si focalizzerà sull’uso dei metodi statistici ed econometrici al fine di quantificare gli effetti causali

- Idealmente vorremmo poter usare dati sperimentali, cioè ottenuti da esperimenti disegnati per valutare un trattamento, una specifica azione di politica economica, un effetto causale.

- Solitamente si possono usare soltanto dati non sperimentali osservando il comportamento reale al di fuori di un contesto sperimentale

- Tuttavia nel caso dei dati sperimentali può risultare difficile scindere l’effetto “trattamento” da altri fattori rilevanti

## In questo corso potrete

- Imparare i metodi per stimare gli effetti causali usando dati non sperimentali

- Familiarizzare con gli strumenti da utilizzare per altri scopi, ad esempio fornire previsioni economiche

- Focalizzare l’attenzione sulle applicazioni mantenendo al minimo la teoria

- Capire ed interpretare i risultati empirici riportati da quotidiani e telegiornali

# Esempi di indagini econometriche

<u>Problema</u> da analizzare: Riforma del sistema di istruzione pubblica

– Scuola Elementare: migliorare apprendimento

<u>Proposta</u>: Ridurre la dimensione delle classi migliora la qualità dell'istruzione nella scuola elementare

- Si?
- Di quanto? Possiamo misurare il beneficio per compararlo con i costi?
- Come si misura?
- Quale è la dimensione ottimale?
- Ci possono essere cause alternative per spiegare un maggiore apprendimento nelle classi la cui dimensione è stata ridotta?

# Esempi di indagini econometriche

Problema: Discriminazione (di vario tipo) nella concessione dei mutui per le abitazioni.

- La teoria dice che non dovrebbe esserci alcuna discriminazione nella concessione dei mutui
- La FED di Boston mostra che nei primi anni Novanta al 28% dei richiedenti afro-americani veniva rifiutato il mutuo contro il 9% dei richiedenti bianchi.

C'è discriminazione?

  – Bisogna valutare l'effetto della razza o altro sulla probabilità di ottenere un mutuo controllando (tenendo costante) tutti gli altri fattori che sono importanti per la concessione di un mutuo

# Esempi di indagini econometriche

<u>Problema</u>: Ridurre il consumo di sigarette

<u>Proposta</u>: aumentare le imposte sul tabacco

- La teoria economica suggerisce che un aumento delle imposte sul tabacco causa una diminuzione nei consumi dello stesso.
- Di quanto?
  - Bisogna guardare l'elasticità della domanda di tabacco
  - Problema di causalità simultanea
    - Se la domanda è alta (molti fumatori) potrebbe condizionare i politici a tenere prezzi bassi.

# Esempi di indagini econometriche

Quale sarà il tasso d’inflazione il prossimo anno?

- Problema di previsione e quindi non riguarda la stima di un effetto causale

Che cosa determina la crescita delle nazioni?

- La teoria economica suggerisce la possibilità di convergenza, mentre l’evidenza empirica sembra suggerire il contrario. Abbiamo teorie ragionevoli?

# Dati Economici

- Dati Sperimentali
  - Provengono da esperimenti disegnati per valutare un trattamento oppure per valutare un effetto causale (ad es. una particolare politica economica).
  - Gruppo di controllo e gruppo di trattamento

- Dati <u>non</u> Sperimentali
  - Raccolti tramite indagini campionarie
  - I livelli di trattamento non sono assegnati in maniera causale
  - È difficile distinguere gli effetti del trattamento da altri effetti

# Dati Economici

- Dati Sezionali (Cross Section)
  - Dati su entità diverse osservati per un solo periodo
  - Il numero delle entità definisce la dimensione del database

- Serie Temporali (Time Series)
  - Dati per una singola entità osservati in momenti diversi
  - Il numeri di momenti (anni, mesi, giorni) in cui la singola entità viene osservata definisce la dimensione del database

- Dati Panel (Longitudinali)
  - Dati su entità diverse ognuna delle quali è osservata in due o più periodi.
  - Il numero totale di osservazioni a disposizione è:

    n°entità*n°periodi

# Variabili Casuali e Distribuzioni di Probabilità

## Definizioni

- Risultati: esiti potenziali (mutuamente esclusivi) di un processo casuale
  - Es.: il mio pc si blocca 0 volte, 1 volta, 2 volte, mai
- Probabilità: proporzione di volte in cui un determinato risultato si verifica nel lungo periodo
  - Se la probabilità che il computer non si blocchi è pari all'80%, significa che scrivendo un numero elevato di documenti l'80% di questi sarà completato

# Variabili Causali e Distribuzioni di Probabilità

## Definizioni

- Spazio Campionario: insieme di tutti i risultati possibili

- Evento: sottoinsieme di uno spazio campionario, cioè può essere composto da uno o più risultati.
  - L’evento “il pc non si bloccherà più di una volta” contiene due risultati: “un blocco” e “nessun blocco”
- Variabili casuali: indicatore numerico di un risultato casuale
  - Numero di volte in cui il pc si blocca

- Variabile casuale discreta: 0,1,2,..

- Variabile casuale continua: $[-\infty,+\infty]$

# Distribuzione di Probabilità
## Variabile Casuale Discreta

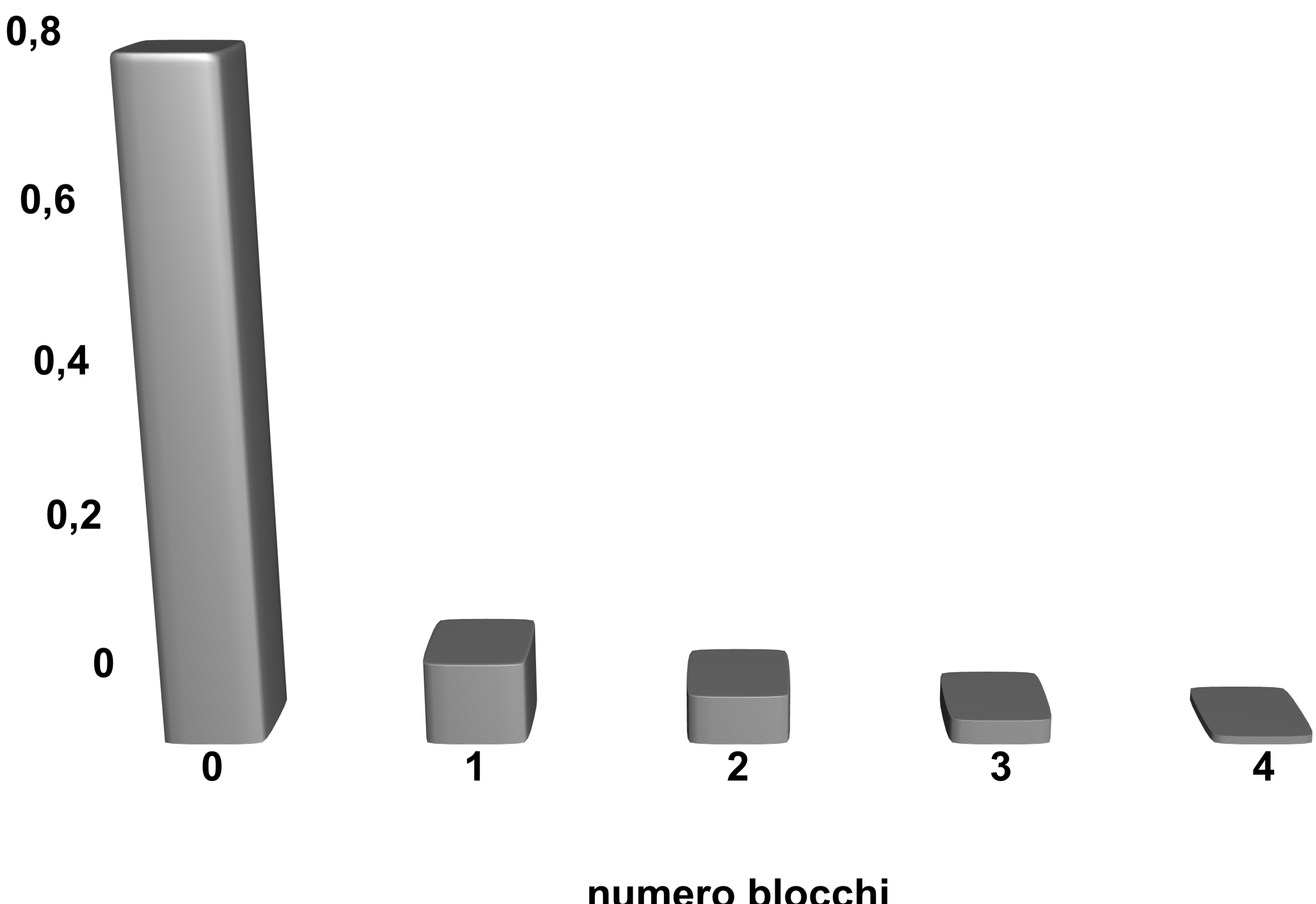
- Definizione: elenco di tutti i possibili valori della variabile e delle probabilità con cui si verificano.

| Numero blocchi | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Distribuzione di Probabilità | 0,8 | 0,1 | 0,06 | 0,03 | 0,01 |
| Distribuzione di probabilità cumulata | 0,8 | 0,9 | 0,96 | 0,99 | 1 |

- Probabilità di eventi: utilizzando la distribuzione di probabilità.
  - Es. Pr(M=1 o M=2) = Pr(M=1)+Pr(M=2) = 0,1+0,06 = 0,16 (16%)

- Distribuzione di probabilità cumulata (o Funzione di ripartizione)
  - Probabilità che una variabile casuale sia minore o uguale ad un particolare valore.
  - Es. Probabilità che ci sia almeno un blocco = Pr(M≤1) = Pr(M=0)+Pr(M=1) = 0,8+0,1 = 0,9 (90%)

# Distribuzione di Probabilità del numero di blocchi del pc



# Distribuzione di Bernoulli

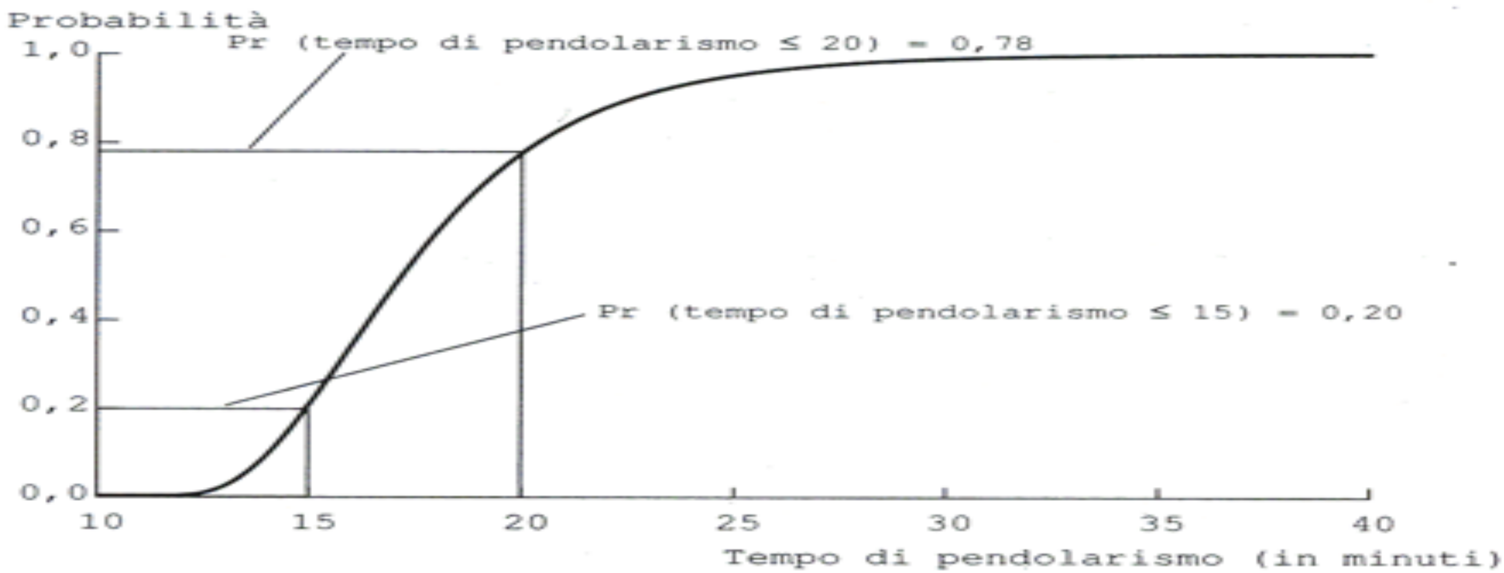
- Un particolare caso di variabile binaria è la variabile casuale di Bernoulli che può assumere soltanto i valori 0 oppure 1.

- La sua distribuzione è chiamata Distribuzione di Bernoulli
  – Es. Definiamo F come la faccia che esce quando lanciamo una moneta, dove F=1 se esce testa e F=0 se esce croce. I risultati possibili e le probabilità sono:

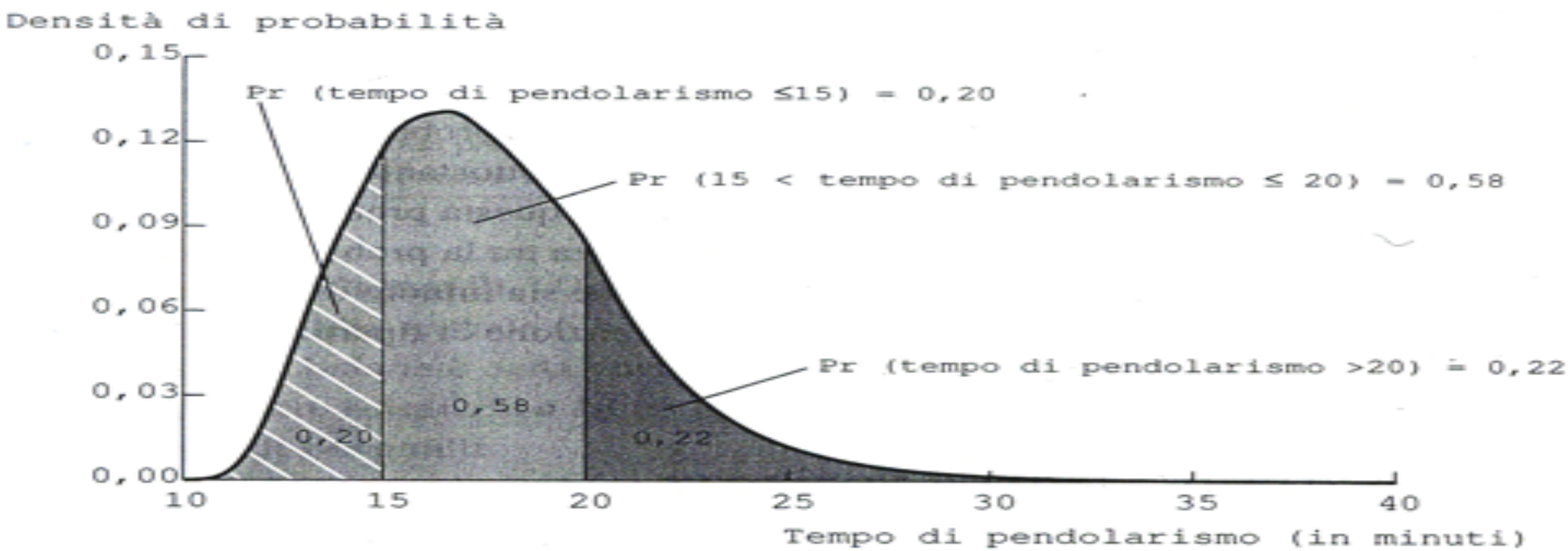
$$F = \begin{cases} 1 \text{ con probabilità p} \\ 0 \text{ con probabilità } 1 - \text{p} \end{cases} \qquad (2.1)$$

# Distribuzione di Probabilità di una variabile casuale continua

- Possiamo applicare gli stessi concetti illustrati per le variabili casuali discrete.
- La funzione di ripartizione di una variabile casuale continua è la probabilità che la variabile sia minore o uguale ad un certo valore.
  - Il tempo impiegato da uno studente/lavoratore per guidare da casa a scuola/ufficio.
- Una differenza è data dalla distribuzione di probabilità
  - In questo caso parleremo di Funzione di Densità di Probabilità (p.d.f.)
- L’area sottostante la p.d.f. tra due punti qualsiasi rappresenta la probabilità che la variabile casuale cada tra quei due punti.



(a) Funzione di ripartizione del tempo di pendolarismo



# Valore Atteso, Media e Varianza

- Valore Atteso (media di Y; $\mu_Y$; E(Y)) è il valore medio della variabile casuale.
  - Nel caso di variabili casuali discrete, il valore atteso è calcolato come media ponderata dei possibili risultati con i pesi pari alla probabilità di tali risultati.
  - Esempio: Consideriamo nuovamente il caso del blocco del pc. Il valore atteso di M è il numero medio di blocchi ponderato con la frequenza con la quale avviene un certo numero di blocchi. Avremo E(M)=0*0.8+1*0.1+2*0,06+3*0.04+4*0.01=0.35 (2.1)
- Variabile casuale di Bernoulli
  - E(F)=1*p+0*(1-p)=p (2.3)
- Variabile casuale continua
  - Integrali

# Valore Atteso, Media e Varianza

- Data una variabile casuale che assume k valori, $y_1, y_2, .. y_k$ e che la probabilità che Y assuma il valore $y_1$ sia $p_1$ e così via. Il valore atteso di Y è

$$E(Y)=y_1p_1+y_2p_2+\ldots+y_kp_k=\sum_{i=1..k}(y_ip_i) \qquad (2.4)$$

- La varianza e la deviazione standard sono misure di dispersione di una distribuzione di probabilità.
  - Varianza: $\sigma^2_Y=\text{var}(Y)=E[(Y-\mu_Y)^2]=\sum_{i=1\ldots k}(y_i-\mu_Y)^2p_i$ (2.6)
  - Deviazione standard: $\sqrt{\text{var}(Y)}=\sigma_Y$.
- Esempio del blocco pc:
  - $Var(M)=(0-0,35)^2*0,8+(1-0,35)^2*0,1+(2-0,35)^2*0,06+(3-0,35)^2*0,03+(4-0,35)^2*0,01=0,6475$
  - $\sigma_Y=\sqrt{(0.6475}\approx 0,8$
- Variabile di Bernoulli: $\mu_F=p$; $Var(F)=(0-p)^2*(1-p)+(1-p)^2*p=p(1-p)$; $\sigma_F=\sqrt{p(1-p)}$

# Esempio prestito

- Supponiamo di prestare 100 euro ad un amico ad un tasso d'interesse del 10%.
- Se il prestito viene restituito otteniamo 110 euro.
- Il rischio che il nostro amico sia insolvente è pari all'1%.
- L'ammontare rimborsato è una variabile casuale che è uguale a 110 euro con la probabilità 0,99 ed a 0 euro con probabilità 0,01.
- Considerando molti prestiti avremo un valore atteso di rimborso pari a
  - 110 euro*0,99+0 euro*0.01=108,90 euro
  - Conviene?

# Esempio con una funzione lineare di due Variabili

- Supponiamo un'imposizione fiscale pari al 20% del proprio reddito che prevede anche un trasferimento esentasse di 2000\$. In termini formali abbiamo,

$$Y=2000+0,8X \qquad (2.8)$$

- Assumiamo che Y (reddito netto) sia una variabile casuale con media $\mu_Y$ e varianza $\sigma^2_Y$. Anche X (reddito lordo) sarà una variabile casuale. Possiamo calcolare la media e la variazione standard di Y.
  - $E(Y)= \mu_Y=2000+0,8\mu_X$ (2.9)
  - $Var(Y)=(Y-\mu_Y)^2$ Allora $Y-\mu_Y=2000+0,8X-2000-0,8\mu_X=0,8(X-\mu_X)$, quindi $E[(Y-\mu_Y)^2]=E\{[0,8(X-\mu_X)^2]\}=0,64\ E[(X-\mu_X)^2]$. Quindi $var(Y)=0,64var(X)$
  - La deviazione standard è $\sigma_Y=0,8\ \sigma_X$

# Distribuzioni Congiunte

- La distribuzione di probabilità congiunta di due variabili casuali discrete X e Y misura la probabilità che le variabili assumano certi valori (x e y) simultaneamente.
  - La somma di tali probabilità sarà sempre uguale ad uno
  - Possiamo quindi scrivere $\Pr(X=x, Y=y)$

## Distribuzione congiunta di condizioni meteorologiche e tempo di percorrenza

- Le condizioni meteorologiche influenzano il tempo di percorrenza di un pendolare.
  - Y è variabile casuale binaria uguale ad 1 se il tempo di percorrenza è minore di 20 minuti ed a zero altrimenti
  - X è variabile casuale binaria uguale a zero se piove ed a 1 altrimenti
  - Considerando le due variabili si ottengono 4 possibili risultati
- La distribuzione di probabilità congiunta è la frequenza con la quale ciascuno di questi 4 risultati si verifica considerando molti viaggi.

# Esempio

| | Pioggia (X=0) | No Pioggia (X=1) | Totale |
|---|---|---|---|
| Tragitto lungo (Y=0) | **0.15** | **0.07** | **0.22** |
| Tragitto breve (Y=1) | **0.15** | **0.63** | **0.78** |
| Totale | **0.3** | **0.7** | **1** |

# Distribuzione marginale e condizionata

- La distribuzione di probabilità marginale di una variabile casuale Y coincide con la distribuzione di probabilità di Y.
- La distribuzione di una variabile casuale Y condizionatamente ai valori assunti da un'altra variabile casuale X è definita distribuzione condizionata di Y data X.
  - $\Pr(Y=y|X=x)= \Pr(X=x,Y=y)/\Pr(X=x)$ (2.15)

- Quale è la probabilità di un tempo di percorrenza lungo (Y=0) se si sa che sta piovendo (X=0)?

  - $\Pr(Y=0|X=0)=\Pr(X=0,Y=0)/\Pr(X=0)=0{,}15/0{,}30=0{,}50$

- Esempio. Supponiamo di andare in biblioteca per scrivere il nostro articolo. Ci viene assegnato un pc in maniera casuale.

- Metà dei pc sono nuovi e metà sono vecchi. La probabilità che il pc si blocchi ha una distribuzione condizionata all'età del pc utilizzato.

- A=1 pc nuovo, 0 altrimenti

- Probabilità congiunta di M=A=0?

- Probabilità condizionata che utilizzando un pc vecchio non ci siano blocchi
  - Pr(M=0, A=0)/Pr(A=0)=0,35/0,50=0,70

# Esempio di Distribuzione di Probabilità Condizionata

| Distribuzione Congiunta | | | | | | |
|---|---|---|---|---|---|---|
| | M=0 | M=1 | M=2 | M=3 | M=4 | Totale |
| Computer Vecchio | 0.35 | 0.065 | 0.05 | 0.025 | 0.01 | 0.5 |
| Computer Nuovo | 0.45 | 0.035 | 0.01 | 0.005 | 0 | 0.5 |
| Totale | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 | 1 |

Distribuzione Condizionata di M data A(=1 se è nuovo, =0 se è vecchio)

| | M=0 | M=1 | M=2 | M=3 | M=4 | Totale |
|---|---|---|---|---|---|---|
| Pr(M\|A=0) | 0.7 | 0.13 | 0.1 | 0.05 | 0.02 | 1 |
| Pr(M\|A=1) | 0.9 | 0.07 | 0.02 | 0.01 | 0 | 1 |

# Media e Varianza Condizionata

- La media condizionata di Y data X è la media della distribuzione condizionata di Y data X. È il valore medio di Y quando X=x.
  - $E(Y|X=x)=\sum_{i=1...k}Pr(Y=y_i|X=x)$ (2.16)
- Riferendoci alla tabella precedente calcoliamo la media condizionata dei blocchi dato un computer nuovo:
  - E(M|A=1)=0*0.9+1*0.07+2*0.02+3*0.01+4*0= 0.14
- E se volessimo calcolare il numero medio dei blocchi?
  - Dobbiamo utilizzare la legge delle interazioni iterate
  - E(Y)=E[E(Y|X)] (2.18)

  Esempio: E(M)=E(M|A=0)*Pr(A=0)+E(M|A=1)*Pr(A=1)=0.56*0.5+0.14*0.5=0.35

# Media e Varianza Condizionata

- La varianza di Y condizionata a X è la varianza della distribuzione condizionata di Y data X.
  - $Var(Y|X=x)=\sum_{i=1...k}[y_i-E(Y|X=x)^2]\ Pr(Y=y_i|X=x)$ (2.19)
  - Esempio:var(M| A=0)=$(0-0.56)^2*0.7+(1-0.56)^2*0.13+(2-0.56)^2*0.1+(3-0.56)^2*0.05+(4-0.56)^2*0.02 \approx 0.99$

- Due variabili casuali X e Y sono indipendentemente distribuite (indipendenti), se conoscere il valore di una di esse non fornisce alcuna informazione circa l'altra.
  - X e Y sono indipendentemente distribuite se per tutti i valori x e y

- Considerando la distribuzione congiunta delle due variabili abbiamo
  - $Pr(Y=y|X=x)=Pr(Y=y)$ oppure $Pr(X=x,Y=y)=Pr(X=x)*Pr(Y=y)$ (2.20 e 2.21)

# Covarianza

- La covarianza è una misura dell'intensità con la quale due variabili casuali si muovono insieme.
  - $Cov(X,Y)=\sigma_{XY}=E[(X-\mu_X)(Y-\mu_Y)]=\sum^{k}_{i=1.}\sum^{l}_{j=1}(x_j-\mu_X)(y_i-\mu_Y)Pr(X=x_j,Y= y_i)$ (2.22)
- Ad esempio, supponiamo che quando X è maggiore della sua media, Y tenda ad essere maggiore della propria media e viceversa. In entrambi i casi la covarianza è positiva.
- Se le due variabili casuali tendono a muoversi nella stessa (opposta) direzione, la covarianza è positiva (negativa).
- Se le due variabili casuali sono indipendenti che valore assumerà la covarianza?
- L'unita di misura della covarianza può rendere problematica l'interpretazione dei suoi risultati. Unità di misura di X per quella di Y.

# Correlazione

- La Correlazione è una misura alternativa della dipendenza tra Y e X che risolve il problema dell'unità di misura della covarianza.
  – $\mathrm{Corr}(X,Y)=\mathrm{Cov}(X,Y)/\sqrt{[\mathrm{var}(X)\mathrm{var}(Y)]}=\sigma_{XY}/(\sigma_X\sigma_Y)$ (2.23)
  – $-1 \leq \mathrm{Corr}(X,Y) \geq 1$
- Correlazione e media condizionata
  – Se $E(Y|X)= \mu_Y$, allora $\mathrm{cov}(Y,X)=0$ e $\mathrm{corr}(Y,X)=0$

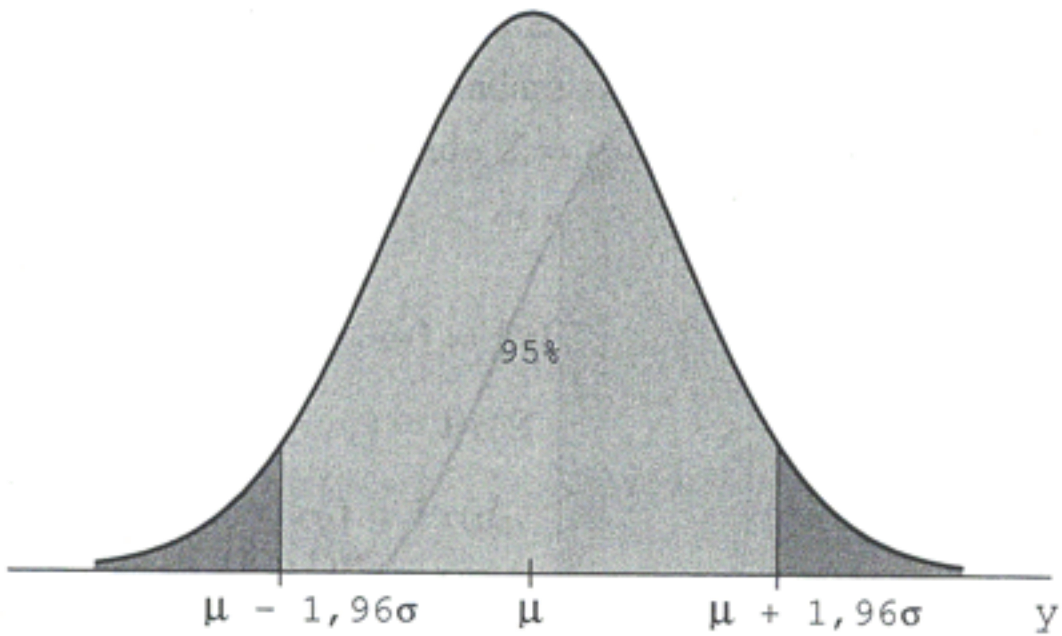
  (il percorso inverso non è necessariamente vero)

# Media e Varianza di somme di variabili casuali

- $E(X+Y)=E(X)+E(Y)=\mu_X+\mu_Y$ (2.26)
- $Var(X+Y)=var(X)+var(Y)+2cov(X,Y)=\sigma^2_X+\sigma^2_Y+2\sigma_{XY}$ (2.27)
- Se X e Y sono indipendenti la cov(X,Y)=0, quindi
  - $Var(X+Y)\ var(X)+var(Y)+2cov(X,Y)=\sigma^2_X+\sigma^2_Y$ (2.28)
- $E(a+bX+cY)=a+b\mu_X+c\mu_Y$ (2.29)
- $Var(a+bY)=b^2\sigma^2_Y$ (2.30)
- $Var(aX+bY)=a^2\sigma^2_X+b^2\sigma^2_Y+2ab\sigma_{XY}$ (2.31)
- $E(Y^2)=\sigma^2_Y+\mu^2_Y$ (2.32)
- $Cov(a+bX+cVY)=b\sigma_{XY}+c\sigma_{YV}$ (2.33)
- $E(XY)=\sigma_{XY}+\mu_X\mu_Y$ (2.34)

# La Distribuzione Normale

**Figura 2.3:** *la densità di probabilità normale*




*La funzione di densità di probabilità normale con media $\mu$ e varianza $\sigma^2$ è una curva di forma campanulare, centrata in $\mu$. L'area sotto la funzione di densità normale compresa tra $\mu - 1,96\,\sigma$ e $\mu + 1,96\,\sigma$ è $0,95$. La distribuzione normale è indicata con $N(\mu, \sigma^2)$*

# Distribuzione Normale Standard

- È la distribuzione normale con N(0,1).

- Le variabili con tale distribuzione sono usualmente indicate con Z e la funzione di ripartizione normale standard è indicata con Φ.
  - $\Pr(Z \leq c) = \Phi(c)$, dove c è una costante.

- La probabilità di una variabile normale deve essere standardizzata prima di poterla calcolare: cioè bisogna prima sottrarre la media e dividere poi il risultato per la deviazione standard.

# Matematicamente...

- Supponiamo $Y \sim N(\mu,\sigma^2)$. Standardizzando Y otteniamo $Z = (Y - \mu)/\sigma$.
- Siano $c_1$ e $c_2$ due numeri tali che $c_1 < c_2$ e sia $d_1 = (c_1 - \mu)/\sigma$ e $d_2 = (c_2 - \mu)/\sigma$. Allora possiamo scrivere,

- $\Pr(Y \leq c_2) = \Pr(Z \leq d_2) = \Phi(d_2)$ (2.38)

- $\Pr(Y \geq c_1) = \Pr(Z \geq d_1) = 1 - \Phi(d_1)$ (2.39)

- $\Pr(c_1 \leq Y \leq c_2) = \Pr(d_1 \leq Z \leq d_2) = \Phi(d_2) - \Phi(d_1)$ (2.40)

# Esempio

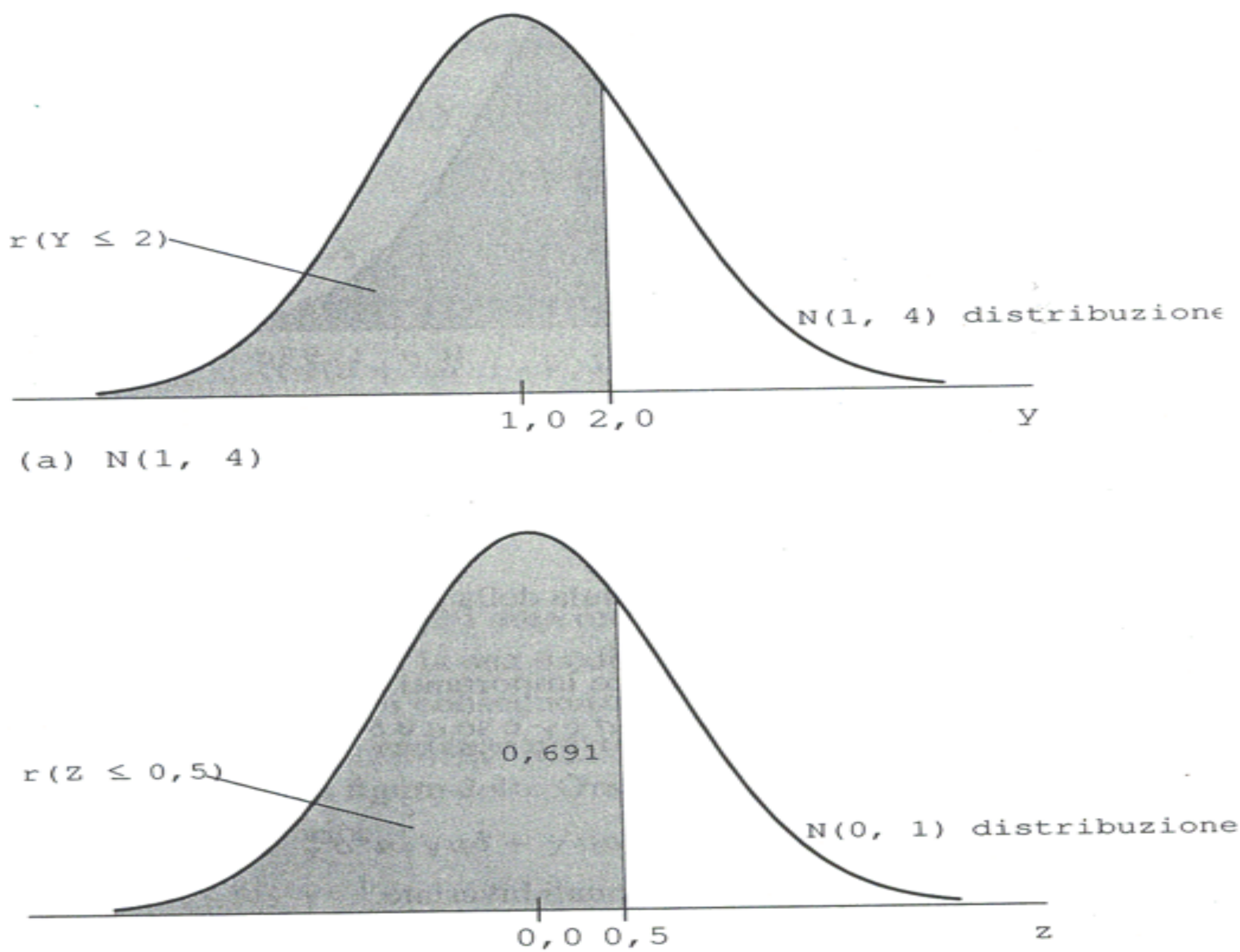
- Supponiamo che Y sia distribuita secondo una N(1,4). Calcoliamo la probabilità che $Y \leq 2$.

- Y standardizzata è $(Y-1)/\sqrt{4}=1/2(Y-1)$. Quindi la variabile casuale $\frac{1}{2}(Y-1)$ e distribuita secondo N(0,1).

- La probabilità $Y \leq 2$ è equivalente a $\frac{1}{2}(Y-1) \leq 1/2(2-1)=1/2$.

- Possiamo quindi scrivere,
  - $\Pr(Y \leq 2)=\Pr[1/2(Y-1) \leq 1/2]=\Pr(Z \leq 1/2)=\Phi(0.5)=0.691$ (2.36)
    - Il valore 0.691 è preso dalla tavola delle Z

**Figura 2.4:** *calcolo della probabilità che* $Y \leq 2$ *quando* $Y$ *si distribuisce come una* $N(1,4)$



# La distribuzione chi-quadrato ($\chi^2$)

- È la distribuzione della somma dei quadrati di m variabili casuali, indipendenti e normalmente distribuite.

- m è chiamato numero di gradi di libertà della distribuzione chi-quadrato.

- $Z^2_1 + Z^2_2 + Z^2_3$ ha una distribuzione chi-quadrato con 3 gradi di libertà (m=3).

- I valori della distribuzione sono riportati in tavola 3 SW.

# La distribuzione $F_{m,\infty}$

- È la distribuzione di una variabile casuale chi-quadrato con m gradi di libertà, divisa per m.

- $(Z^2_1 + Z^2_2 + Z^2_3)/3$ ha una distribuzione $F_{3,\infty}$.

- I valori della distribuzione sono riportati in tavola 4 SW.

# La distribuzione t di Student

- La distribuzione t di Student con m gradi di libertà è la distribuzione del rapporto di due variabili casuali indipendenti.
  - $Z/\sqrt{(W/m)}$, dove Z è una variabile casuale normale standard e W è una variabile casuale con distribuzione chi-quadrato con m gradi di libertà.

- La forma della t è simile a quella della distribuzione normale. Per $m \geq 30$ la t è ben approssimata dalla distribuzione normale standard.

# Campionamento Casuale

- Campionamento casuale:selezionare in modo casuale un campione da una popolazione più ampia.
  - Effetto:rendere la media campionaria una variabile casuale con, quindi, una distribuzione campionaria.
- Campionamento casuale semplice: n soggetti scelti a caso da una popolazione i cui membri hanno la stessa probabilità di essere scelti.
  - Esempio studente pendolare
- I valori delle osservazioni sono trattati come variabili casuali.
- La distribuzione marginale di ogni valore $Y_i$ è la stessa per ogni $i=1,...,n$. Quindi tali variabili casuali sono identicamente distribuite.

# Campionamento Casuale

- Conoscere il valore di $Y_1$ non fornisce alcuna informazione su $Y_2$ e la distribuzione condizionata di $Y_2$ data $Y_1$ coincide con la distribuzione marginale di $Y_2$. Cioé $Y_1$ è distribuita indipendentemente dalle altre variabili casuali.
- Quando $Y_1,..,Y_n$ vengono dalla stessa distribuzione e sono indipendentemente distribuite sono indipendentemente e identicamente distribuite (i.i.d.).

# Distribuzione campionaria della media campionaria

- La media campionaria di n osservazioni $Y_1,..,Y_n$ è

$$\bar{Y} = \frac{1}{n}(Y_1 + ... + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- Importante: l'estrazione di un campione casuale ha l'effetto di rendere la media campionaria una variabile casuale. La media campionaria varia da una campione ad un altro

- La distribuzione di $\bar{Y}$ è detta distribuzione campionaria di $\bar{Y}$

- Possiamo ottenere le seguenti formule per il valore medio, la varianza e la deviazione standard della media campionaria

$$E(\bar{Y}) = \mu_Y$$

$$\text{var}(\bar{Y}) = \sigma^2_{\bar{Y}} = \frac{\sigma^2_n}{n}$$

$$std.dev(\bar{Y}) = \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$$

## Approssimazioni alla distribuzione campionaria per grandi campioni

- Per caratterizzare le distribuzioni campionarie si possono seguire due approcci,
  - Esatto: formula che valga per ogni valore di n (distribuzione esatta o in campioni finiti).
    - Problemi se la distribuzioni di Y non è normale.
  - Approssimato: elevata numerosità campionaria (distribuzione asintotica).
- Mentre le distribuzioni esatte sono molto complesse e dipendono dalla distribuzione di Y, le distribuzioni asintotiche sono semplici.
- L’approssimazione tramite la distribuzione normale è alla base di quasi tutti i testi di econometria.

# Legge dei grandi numeri e consistenza

- La legge dei grandi numeri stabilisce che, sotto condizioni generali, $\bar{Y}$ è prossima a $\mu_Y$ con probabilità molto alta quando n è grande.
- Le condizioni perché valga la legge dei grandi numeri sono che le $Y_i$ siano i.i.d. e che la varianza di $Y_i$ sia finita.
- La convergenza in probabilità (consistenza) è la proprietà per cui $\bar{Y}$ è prossima a $\mu_Y$ con probabilità crescente al crescere di n.

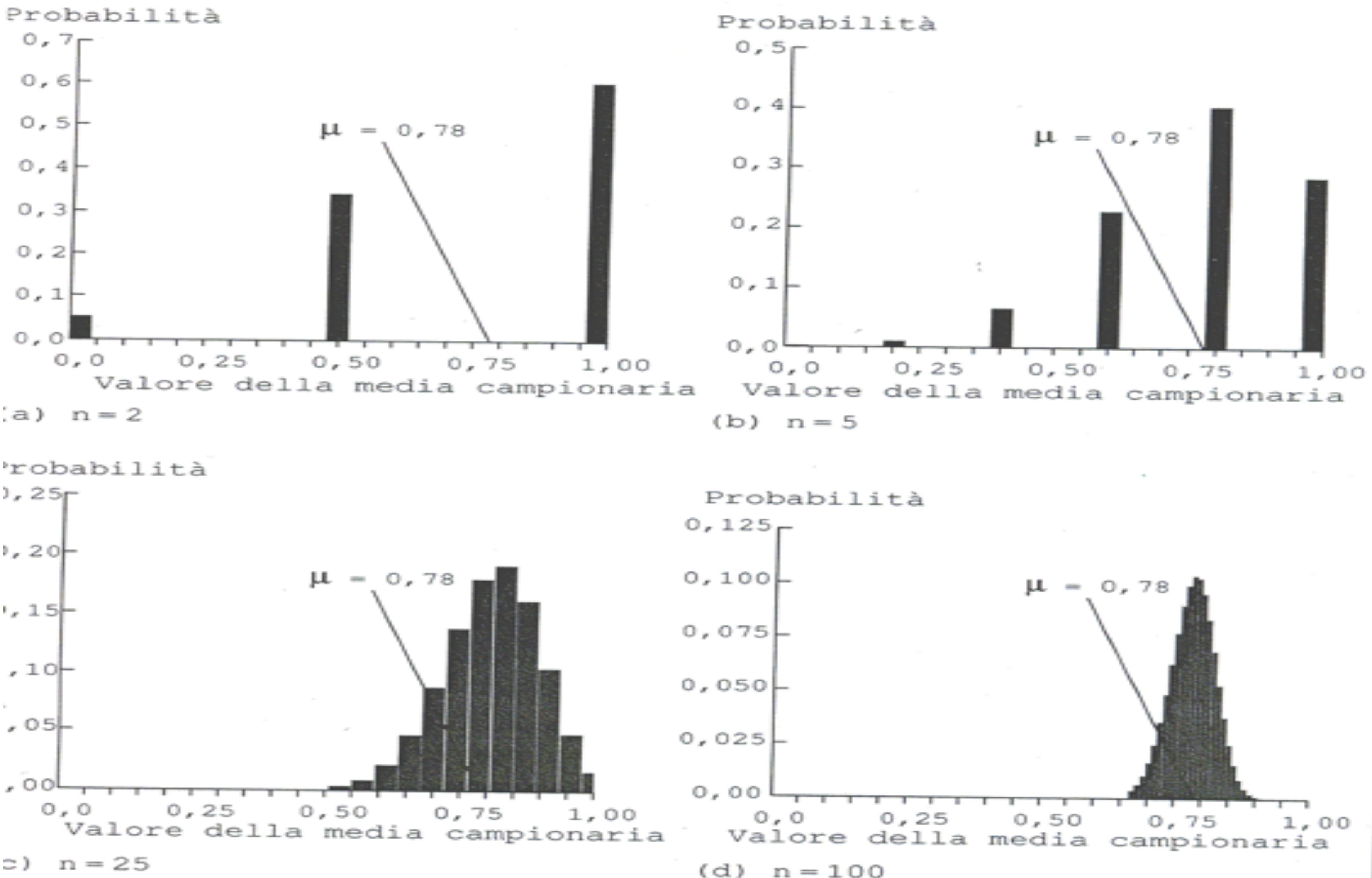
# Esempio con una distribuzione di Bernoulli

- La figura mostra la distribuzione campionaria di un viaggio lungo per varie dimensioni del campione.

- Dalla tabella precedente sappiamo che la probabilità di un viaggio breve $Y_i$=1 è $\mu_Y$ =0.78.

- Al crescere di n la media campionaria si avvicina sempre di più a $\mu_Y$.

**Figura 2.6:** *distribuzione campionaria della media campionaria di $n$ variabili casuali di Bernoulli*



# Il teorema limite centrale

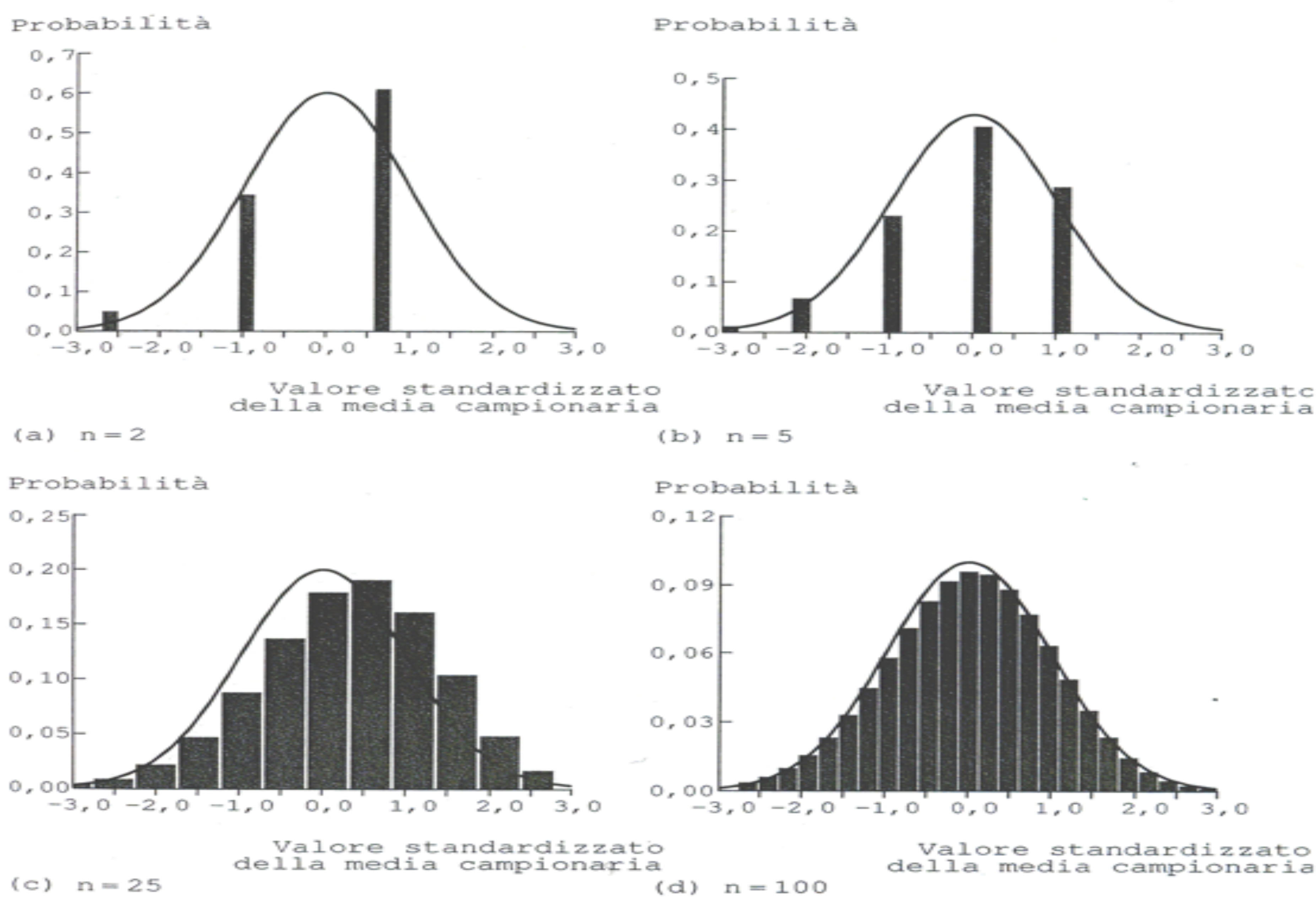
- Il teorema limite centrale afferma che, sotto condizioni generali, la distribuzione di $\bar{Y}$ è ben approssimata da una distribuzione normale quando n è grande. Si distribuisce asintoticamente secondo una normale.

- Bisogna notare che il teorema non fa alcuna menzione al tipo di distribuzione delle osservazioni.

- Un esempio è fornito dalla figura 2.7 dove viene presentata la forma standardizzata della distribuzione di Bernoulli descritta in figura 2.6.

- La distribuzione della versione standardizzata di $\bar{Y}$ è
  - $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ ed è approssimata da N(0,1) per n grande

**Figura 2.7:** *distribuzione della media campionaria di $n$ variabili casuali di Bernoulli con $p = 0,78$*



Distribuzione della media campionaria standardizzata di n estratti da una distribuzione asimmetrica

**Figura 2.8:** *distribuzione della media campionaria standardizzata di $n$ estratti da una distribuzione asimmetrica*