****

***Frontespizio***

**Università degli studi di Roma**

**Unitelma Sapienza**

**Dipartimento di Scienze giuridiche ed economiche**

**Corso di Laurea TRIENNALE in SCIENZE DELL’ECONOMIA E DELLA GESTIONE AZIENDALE**

**Tesi di Laurea in**

**[Statistica]**

**Classificazione binaria tramite la regressione logistica: caso di studio su processi bancari**

Relatore:

Ch.mo Prof. **Pasquale Sarnacchiaro**

Candidato:

Americo Costantini

Anno Accademico 2016/2017

CAPITOLO PRIMO

TITOLO DEL CAPITOLO

wcnipwbifhvbfiebvihefbvihfbvfiehvbefihbvqeifhvbefihbveifhbvqiefbveihfbviefhbviefhqbviqfehvbihefbvihv

# Paragrafo

àldkwncdljwnvjrnvfjnvfjlnvjlfnvlajfnvfdjnvdljvnlfjvnfljvnfljvnfldjnjlvndfaljvndfljvndfljvndfljvndfljvnldfjnv

# Secondo paragrafo

### Terzo sotto

### Sottoparagrafo

### Sottoparagrafo

### Sottoparagrafo

# Paragrafi

### sotto

CAPITOLO secondo

LA REGRESSIONE LOGISTICA

# La funzione di ipotesi nella regressione logistica

### Perché non la regressione lineare

L’intento della classificazione binaria è predire l’uscita di una variabile dicotomica *Y* che può assumere soltanto due valori:

*Y*=

Un tipico approccio per la predizione è quello di stimare la probabilità che *Y* assuma il valore *1* in base ai valori del vettore delle variabili indipendenti, che denoteremo con ***X***. Come noto, la probabilità che un evento accada può assumere soltanto valori tra *0* e *1[[1]](#footnote-1)*, quindi a questo scopo la regressione lineare è inadatta, in quanto modello volto a predire il valore di una variabile quantitativa che non può essere sottoposta a una restrizione di intervallo di questo tipo[[2]](#footnote-2).

Inoltre una variabile casuale dicotomica ha una distribuzione di Bernoulli (o binomiale), che prevede una varianza pari a , dove è la probabilità che l’evento accada. Tale varianza dipende da un valore non costante (perché dipendente dal valore dei regressori), e quindi non è costante. Se ne conclude ulteriormente l’inadeguatezza dell’utilizzo della regressione lineare, che regge sull’assunto della omoschedasticità e quindi della omogeneità della varianza per tutte le combinazioni di valori delle variabili indipendenti[[3]](#footnote-3).

Insomma, l’obiettivo è quello di individuare una funzione che a partire dal vettore di valori delle variabili indipendenti ***X*** fornisca una probabilità che *Y = 1* come valore atteso di *Y*:

[[4]](#footnote-4)

### La funzione logistica

Una funzione adatta allo scopo di legare ***X*** a *Y* è la cosiddetta funzione logistica[[5]](#footnote-5):

continua da pagina 5 del libro e magari mettici pure una immagine

Capitolo terzo

Titolo capitolo terzo

# Paragrafo 1

### Sottop

lkcjwndournornforuvnreouvneouvneouvnouvntouvntouvntouvntountouvntrouvnrtouvnrtouvntorun[[6]](#footnote-6)

BIBLIOGRAFIA

1. Fai citazione dal manuale di statistica. [↑](#footnote-ref-1)
2. Fai citazione dal pdf di rimini almeno [↑](#footnote-ref-2)
3. citare libro che parla di questo requisito [↑](#footnote-ref-3)
4. rifai citazione da rimini su non varianza anche dell’errore [↑](#footnote-ref-4)
5. citazione da p.5 di kleineibaum [↑](#footnote-ref-5)
6. [↑](#footnote-ref-6)