

*UNIVERSITA' DEGLI STUDI DI PADOVA*

FACOLTA' DI SCIENZE STATISTICHE

**CORSO DI LAUREA**

**IN STATISTICA, ECONOMIA E FINANZA**

RELAZIONE FINALE

**ANALISI STATISTICA DELLE VOTAZIONI  
PRESIDENZIALI AMERICANE  
IN FLORIDA NEL 2000**

RELATORE: PROF. ALESSANDRA SALVAN

LAUREANDO: MARIO VALERIN

ANNO ACCADEMICO 2004-05



# Indice

<b>1. Introduzione</b>	1
<b>2. Le elezioni presidenziali in Florida nel 2000</b>	5
2.1 Il voto a Buchanan nella contea di Palm Beach . . . . .	5
2.2 <i>Misvotes, undervotes e overvotes</i> . . . . .	6
<b>3. Primi ragionamenti</b>	15
3.1 Una valutazione statistica del voto a Palm Beach . . . . .	15
<b>4. Possibili modelli</b>	25
4.1 Trasformazioni della variabile risposta . . . . .	25
4.2 Scelta dei regressori. . . . .	27
4.3 Test di eteroschedasticità . . . . .	29
4.4 Altre diagnostiche di adattamento del modello . . . . .	31
4.5 Test della normalità dei dati . . . . .	33
<b>5. Approfondimenti</b>	35
5.1 Prevedere il voto di Palm Beach . . . . .	35
<b>6. Conclusioni</b>	37
<b>Ringraziamenti</b>	39
<b>Bibliografia</b>	40



# Capitolo 1

## INTRODUZIONE

In questa relazione si espongono e sintetizzano alcuni contributi recenti della letteratura statistica volti ad analizzare i risultati delle elezioni presidenziali del 2000 negli Stati Uniti. In particolare, ci si è concentrati sui risultati della Florida e della contea di Palm Beach. Le principali fonti utilizzate sono i seguenti articoli apparsi sul volume 17 del 2002 della rivista *Statistical Science* : Casella e Gill (2002), Agresti e Presnell (2002), Smith (2002).

Le discipline toccate sono molto diverse tra loro: le scienze politiche e la statistica. Si è scelta la votazione della Florida in particolare perché, oltre al clamore suscitato, ha generato un considerevole interesse essendo relativa ad uno stato abbastanza equamente diviso tra le due maggiori fazioni politiche, repubblicani e democratici, ed anche sufficientemente grande da influenzare in modo significativo le elezioni nazionali.

Le elezioni presidenziali americane prevedono diverso peso ai diversi stati nell'ambito dell'elezione finale. Infatti, il particolare meccanismo elettorale americano prevede che ciascuno stato disponga di un numero stabilito di voti elettorali che corrispondono ad un numero di rappresentanti, chiamati elettori presidenziali, da inviare a Washington per l'elezione formale del presidente.

I voti elettorali di uno stato vengono attribuiti per intero al candidato che ottiene, in quello stato, la maggioranza relativa. Questo comporta la vanificazione dei suffragi ottenuti dagli altri candidati che non hanno la possibilità di sommarli ai voti ottenuti in altri stati. Si pensi che nel 2000 in Florida Al Gore perse i voti degli elettori presidenziali di quello stato per sole 537 preferenze, mentre a livello nazionale ottenne globalmente oltre mezzo milione di voti in più del suo rivale Bush.

Particolare attenzione ha attirato la situazione insolita della contea di Palm Beach nei confronti del candidato conservatore di estrema destra Pat Buchanan. In tale regione vivono infatti persone di classe medio alta con inclinazioni liberali e quindi l'alto consenso ricevuto dal candidato suddetto ha sollevato molti dubbi e perplessità.

D'altro canto, ricerche più datate (Aldrich, Niemi, Rabinowitz e Rohde, 1982) danno risalto alle caratteristiche di complessità, confusione ed incertezza che hanno sempre contraddistinto le politiche elettorali americane. Prova ne sono le schede relative alle votazioni del 1996 per l'eliminazione degli aiuti statali, dove lo scopo della tornata elettorale non era scritto chiaramente da nessuna parte, provocando, naturalmente, estrema confusione negli elettori.

Lo studio statistico è uno strumento particolarmente appropriato ed attraente rispetto alla disciplina delle scienze politiche che qui non cureremo. Infatti, il comportamento dell'elettore è uno dei pochi aspetti dell'uomo che si può realmente osservare e misurare concretamente. Ci sono infatti sia la componente sistematica, rigorosa, dettata da determinati schemi, sia quella stocastica, squisitamente probabilistica, e sono così ben miscelate da essere distinte con difficoltà.

Le scienze statistiche ci vengono in aiuto offrendo strumenti teorici per gestire i problemi tipici derivanti da uno studio su elezioni ed

elettori: errori di rilevazione, dati mancanti, eteroschedasticità, forme non parametriche ecc. Da non dimenticare, infine, che le previsioni fornite dai modelli statistici possono sempre essere verificate in quanto ci sarà sempre, si spera, una votazione futura.

Nel capitolo 2 sono stati descritti i primi dati riguardanti l'argomento in discussione, mentre nel capitolo successivo sono considerate le prime valutazioni dal punto di vista statistico degli stessi. Nel capitolo 4 sono stati presi in considerazione i modelli statistici applicabili a queste rilevazioni elettorali, quindi nel successivo sono stati effettuati dei ragionamenti predittivi riguardo il voto di Palm Beach. Infine, a conclusione di questa relazione sono state tratte delle considerazioni finali in merito a questo fatto che ha fatto discutere il mondo contemporaneo.



## Capitolo 2

### LE ELEZIONI PRESIDENZIALI IN FLORIDA NEL 2000

#### 2.1 Il voto a Buchanan nella contea di Palm Beach

Come già accennato precedentemente, a causa del meccanismo di assegnazione dei voti ad ogni candidato, stato per stato e non a livello nazionale, George Walker Bush, grazie a soli 537 voti in più nello stato della Florida è riuscito ad ottenere tutti i voti degli elettori presidenziali di quello stato che sono ben 25, su un totale di 538, ottenendo quindi la vittoria su Al Gore per soli 4 voti (271 a 267).

L'attenzione di tutta l'opinione pubblica, non solo americana, si rivolse al controverso conteggio effettuato nella contea di Palm Beach in particolare, ma anche in altre contee dello stato della Florida. Il dato sorprendente, e sicuramente fuori da ogni modello di regressione statistica, è l'alto favore elettorale ottenuto dal candidato del *Reform Party*, Pat Buchanan, rispetto a tutte le altre contee dello stesso stato, ma anche rispetto al candidato dello stesso partito alle precedenti elezioni del 1996 (Ross Perot, ex-IBM amico di Thomas Watson Jr. figlio del fondatore della stessa IBM). Ciò è evidenziato dal grafico di Figura 2.1:

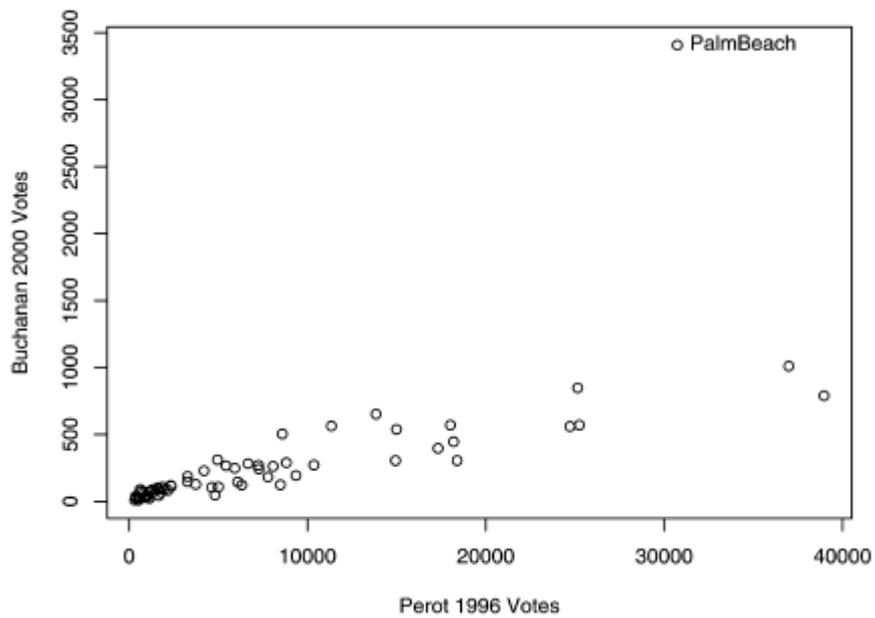


Fig. 2.1: Voti totali, contea per contea, dei candidati del *Reform party*, Buchanan nel 2000 e Perot nel 1996

Grafico a p.437 di *Statistical Science* vol.17 (2002)

## 2.2 *Misvotes, undervotes e overvotes*

Oltre alla osservazione anomala, come viene classificata in statistica descrittiva, i media americani si concentrarono anche sul fenomeno dell' *undervote*, schede bianche senza nessuna preferenza, nelle contee che avevano usato un determinato tipo di scheda elettorale (*punch-card ballot*, cioè una scheda elettorale da perforare). Tale scheda é riportata fedelmente in Figura 2.2:

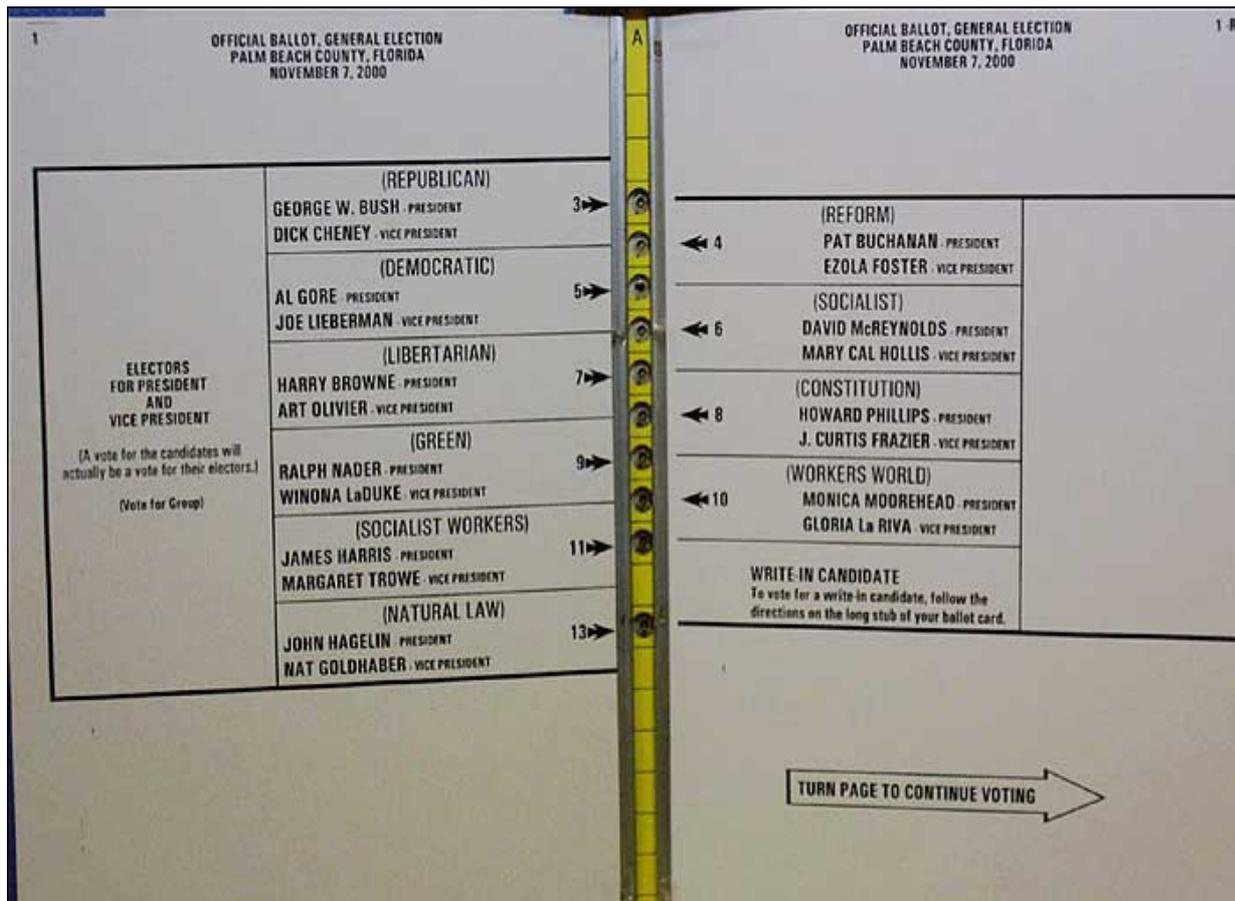


Figura 2.2: Scheda elettorale ufficiale usata nella contea di PalmBeach

Tratta da [www.skirsch.com/politics/election2000/reallocation.htm](http://www.skirsch.com/politics/election2000/reallocation.htm)

Meno risalto, invece, é stato dato al fenomeno opposto, *overvote*, ovvero schede con più di una preferenza, considerate comunque nulle, che in realtà potrebbe essere molto più significativo.

Andando nel dettaglio dei numeri, si nota che la contea di Palm Beach (che adottava la *butterfly-ballot* della Figura 2.2) ha assegnato 3407 preferenze a Buchanan, cioè circa 3 volte e mezza la più alta preferenza ottenuta in un'altra contea senza quel tipo di scheda (circa 1000 voti a Pinellas, che ha all'incirca lo stesso numero di elettori, si veda la tabella 3.2).



alternativa, si possono analizzare i dati relativi ai voti, a livello contea, come una risposta binomiale usando dei modelli di regressione logistica, inoltre il modello binomiale può essere adattato usando il metodo di quasi-verosimiglianza o modelli misti come il beta-binomiale (Agresti e Presnell, 2002). Usando qualsiasi modello ragionevole il messaggio sembra rimanere il medesimo: il voto di Buchanan nella contea di Palm Beach è altamente anomalo.

Si cerca ora di capire quali possano essere state le cause di questa anomalia.

Da un punto di vista squisitamente politico possiamo dire che, un probabile elettore di Gore potrebbe aver forato il buco n°4 perché è alla fine della linea sotto la quale appare la dicitura *Democratic* e quindi dare il voto erroneamente a Buchanan: è proprio quello che l'analisi statistica dei dati ha indicato (cfr. Agresti e Presnell, 2002 e [www.skirsch.com/politics/election2000/reallocation.htm](http://www.skirsch.com/politics/election2000/reallocation.htm)).

Inoltre, essendo il secondo buco forabile, e poiché il nome di Gore appare come secondo nella scheda (colonna di sinistra), molti hanno sicuramente pensato di votare il candidato dei democratici bucando il foro n°4.

Se leggiamo le istruzioni della scheda elettorale, si evince che si vota per i candidati, al plurale; “Presidente e vice”: allora più di una scelta, quindi più di un foro. Sotto c'è scritto “vota per il gruppo”: allora anche qui più di una preferenza. Quindi, i probabili elettori di Gore possono aver forato i buchi n°5 e 6 (Gore ed il vice Lieberman), altri il n°4 (Partito Democratico, il gruppo) e il n°5 (il presidente associato a quel gruppo). Anche gli elettori di Bush nella medesima contea possono aver assegnato un doppio voto per Bush ed il suo vice Cheney (corrispondono infatti ai fori n°3 e 4).

Ecco quindi perché il n°4 (Buchanan) è risultato così votato. Si spiega inoltre un fenomeno così rilevante di *overvote*, cioè di preferenze doppie, infatti i fori più perforati sono stati nell'ordine: 4/5, 5/6 ed infine 3/4.

Un altro fenomeno osservato è il seguente. Se sono stati dati voti a Buchanan erroneamente (perché inserito tra Bush e Gore), anche chi seguiva Gore (David McReynolds, socialista) ha avuto molti voti (302 solo a Palm Beach), la metà dei 622 ottenuti in tutto lo stato (ben 67 contee). Nel caso del candidato socialista, il modello di regressione adottato indicava un valore atteso di circa il 10% di quello in realtà ottenuto (Adams, 2001).

Analizzando ulteriormente il fenomeno dell'*overvote*, si nota che le schede con più di una preferenza sono state ben 19235 (a Palm Beach), il 4,2% dei votanti contro le sole 3073 relativamente alle schede standard (non *butterfly*, non multicolonne e non multipagine) del 1996. Del totale di schede *overvote*, ben 15371 (80%) riportavano la preferenza per Gore, 8689 per Buchanan, 4567 per McReynolds e solo 3571 per Bush.

Inoltre, a sostegno di quanto detto prima, 5330 schede riportavano Buchanan e Gore (buchi n°4 e 5) e 2908 Gore e McReynolds (buchi n°5 e 6). È evidente che questo alto fenomeno di *overvote* è costato molti voti validi soprattutto a Gore, molti meno a Bush.

Il fenomeno dell'*undervote* (schede bianche) è stato impugnato dai legali di Gore per chiedere il riconteggio delle schede, in quanto nelle contee con lettore ottico tale fenomeno in media era lo 0,3% (trascurabile), mentre nelle contee con le *punch-card* l'1,5% (in particolare a Palm Beach 2,2%), ma la Corte Suprema ha negato tale possibilità.

Si spiega quindi la possibile causa dell'alto riscontro elettorale ricevuto da Pat Buchanan nella contea di Palm Beach, ma come mai

c'è stato un così alto *overvote*? E questo non solo in quella contea, ma anche in quella di Duval (7,5% : 21942 schede) e Gadsden (11,6%).

La risposta sembra stare nella complessità delle schede:

- nella contea di Duval i candidati erano in una scheda elettorale composta da 2 pagine; nei distretti a predominanza afro-americana di questa contea più del 20% delle schede sono state scartate, quindi non conteggiate.
- la contea di Gadsden è quella con la più alta concentrazione di afro-americani di tutta la Florida (New York Times, 18/11/2000).

In ultima analisi, il Miami Herald (11 maggio, 2001) ha rivisto 111000 *overvotes* e 61000 *undervotes* della Florida. Gore era segnato su 84197 schede e Bush su 37731. Solo sul 3% di queste schede la volontà dell'elettore era chiara, perché aveva scritto il nome del candidato prescelto anche in basso a destra della scheda dove si leggeva di scrivere il candidato *write-in candidate*, ma con un vantaggio del primo rispetto al secondo di 682 voti, maggiore dei famosi 537.

Come accennato precedentemente, i media americani, ed anche i legali di Gore, dettero più importanza all'*undervote*, ma l'*overvote* era circa due volte il primo. In aggiunta, nelle 18 contee della Florida che usarono le due colonne, fu addirittura 5 volte maggiore rispetto alle contee dove c'era la scheda elettorale con una sola colonna.

Il notevole incremento dell'*overvote* per le schede di tipo multicolonna è avallato anche dal grafico di Figura 2.4, sempre relativo alla complessità della scheda elettorale.

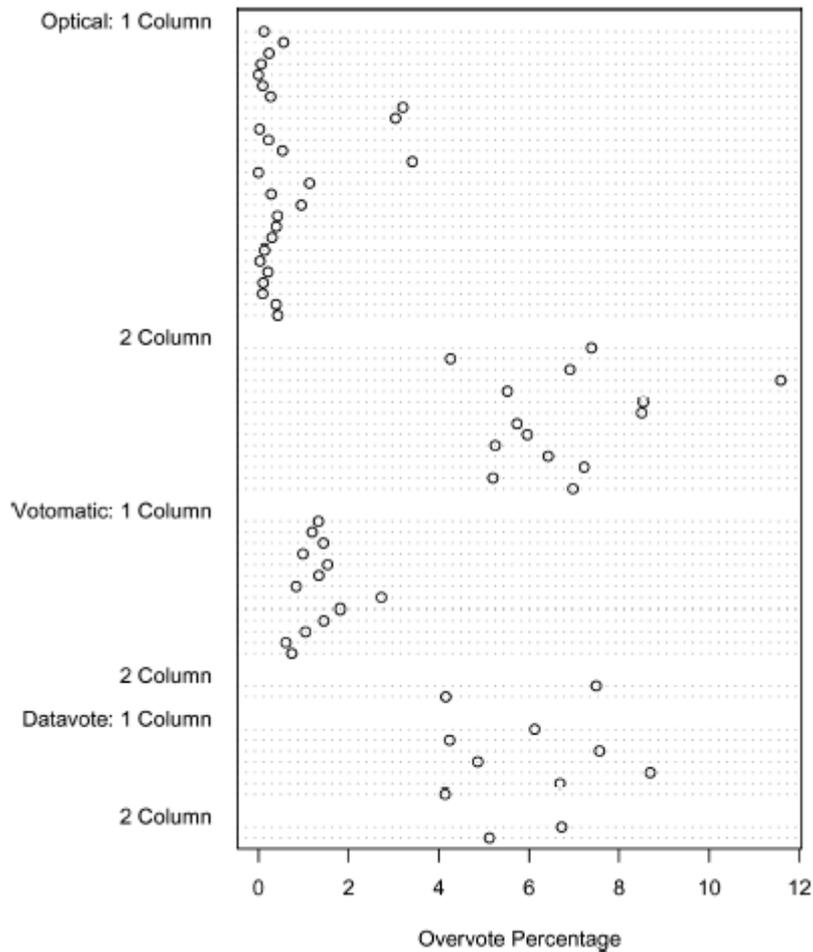


Figura 2.4: Percentuale di *overvote* su 65 contee, non sono state considerate Union e Martin

Grafico a p.439 di *Statistical Science* vol.17 (2002)

E' inoltre considerato il meccanismo di conteggio: sia la tecnologia più avanzata, con il lettore ottico, che quella con la perforazione delle schede. *Optical* e *Votomatic*, perforazione manuale, hanno rilevato la stessa percentuale relativa di *overvote*, facendo quindi propendere l'opinione pubblica per un *design* sbagliato della scheda elettorale come causa di tale fenomeno. Ad ogni modo, da qualunque punto di vista si analizzi l'argomento, risulta evidente che, a causa di voti assegnati erroneamente ad un altro candidato per colpa di:

- schede *butterfly ballot*
- schede multicolonna
- schede multipagine
- confusione degli elettori
- mancanza di attenzione degli elettori stessi,

Al Gore sia rimasto danneggiato di parecchi suffragi e l'ordine di grandezza va dalle migliaia alle decine di migliaia di voti.



## Capitolo 3

### PRIMI RAGIONAMENTI

#### 3.1 Una valutazione statistica del voto a Palm Beach

Finora si é parlato della differenza veramente minima (poco più di 500 voti) che ha permesso al candidato dei repubblicani di battere l'avversario dei democratici. Un ulteriore fatto notevole é che il numero di voti annullati, in tutto lo stato della Florida, si é avvicinato ai 6 milioni.

Di questa cifra, oltre 400 mila (relativi a tutti e 10 i candidati) non sono stati validati nella sola contea di Palm Beach, che é la seconda più popolosa di tutto lo stato, dopo quella di Miami-Dade.

Come é stato discusso nel paragrafo 2.2, il *design* della scheda elettorale, la *butterfly-ballot*, ha confuso molti elettori, i quali, volendo votare Gore, hanno erroneamente bucato il foro relativo a Pat Buchanan, portando il punteggio di questo candidato ad oltre 3400, voti quando una stima del valore atteso su tutto lo stato era al massimo di 1300.

Si nota immediatamente che i 2000 voti di scarto avrebbero dato a Gore la vittoria nello stato della Florida e quindi anche la Casa Bianca.

Molti articoli di quei giorni attestarono che la previsione del risultato di Buchanan in quella contea fosse molto minore dei 1300 voti. Ciò

teneva anche conto del fatto che il candidato democratico riportò il 62% dei voti totali della contea. Di conseguenza, il candidato dell'estrema destra non poteva aver guadagnato così tante preferenze. Considerazioni come questa ribadirono la convinzione dell'opinione pubblica che il vincitore dello stato della Florida fosse Gore e non Bush, ma sollevarono anche dubbi sulla corretta applicazione dei metodi di regressione statistica ai risultati di un'elezione politica.

L'obiettivo di questo e dei prossimi capitoli è presentare l'analisi dettagliata di regressione condotta in Smith (2002). La variabile risposta è rappresentata dai voti al candidato del *Reform Party* su tutte le 67 contee della Florida. Le variabili esplicative sono sia i voti per gli altri candidati sia le variabili demografiche.

L'analisi effettuata da Smith (2002) pone particolare attenzione ai problemi di eteroschedasticità (varianza della variabile risposta non costante), che risulta inevitabile confrontando contee con grande disparità di popolazione (da Lafayette e Liberty con meno di 7 mila votanti a Miami-Dade con più di 2 milioni). Per rimediare a ciò l'autore ha applicato varie trasformazioni dei dati.

Altre problematiche considerate sono la selezione delle variabili, la sovradisersione in un'analisi binomiale o di Poisson e svariate forme diagnostiche e test di adattamento. Il modello stimato sulla base dei dati relativi alle 66 contee della Florida, esclusa Palm Beach, è stato poi utilizzato per prevedere il risultato del candidato conservatore di quella contea. Tutti i risultati indicano una stima o intervallo di confidenza ben al di sotto del dato osservato, avallando quindi l'ipotesi che il discusso 3407 sia un enorme valore anomalo.

Ci si concentrerà ora sull'aspetto statistico della questione. L'analisi di seguito presentata è un esempio di come possono essere applicate le moderne tecniche di regressione statistica ad un fatto reale di grande interesse pubblico.

I dati usati da Smith (2002) sono i risultati elettorali diramati dalla *Florida Division of Elections* e i dati demografici provenienti dallo *U.S. Census Bureau*. Di seguito, sono riportate le tabelle 3.1 e 3.2 con i dati, contea per contea, dei 10 candidati a presidente degli Stati Uniti nello stato della Florida. Tali dati sono stati i primi ad essere resi pubblici, quindi possono differire di poco da quelli certificati definitivamente, ma dopo il completamento dello studio di Smith (2002). Di tutti i candidati ne sono stati considerati solo la metà: Harris, Hagelin, McReynolds, Phillips e Moorehead tutti insieme costituiscono meno dello 0.1% dei voti, quindi risultano assolutamente trascurabili.

County	Bush	Gore	Brow	Nade	Har	Hag	Buc	Mc	Ph	Mo
Alachua	34,124	47,365	658	3226	6	42	263	4	20	21
Baker	5,610	2,392	17	53	0	3	73	0	3	3
Bay	38,637	18,850	171	828	5	18	248	3	18	27
Bradford	5,414	3,075	28	84	0	2	65	0	2	3
Brevard	115,185	97,318	643	4470	11	39	570	11	72	76
Broward	177,323	386,561	1212	7101	50	129	788	34	74	124
Calhoun	2,873	2,155	10	39	0	1	90	1	2	3
Charlotte	35,426	29,645	127	1462	6	15	182	3	18	12
Citrus	29,765	25,525	194	1379	5	16	270	0	18	28
Clay	41,736	14,632	204	562	1	14	186	3	6	9
Collier	60,433	29,918	185	1399	7	34	122	4	10	29
Columbia	10,964	7,047	127	258	1	7	89	2	8	5
Desoto	4,256	3,320	23	157	0	0	36	3	8	2
Dixie	2,697	1,826	32	75	0	2	29	0	3	2
Duval	152,098	107,864	952	2757	37	162	652	15	58	41
Escambia	73,017	40,943	296	1727	6	24	502	3	110	20
Flagler	12,613	13,897	60	435	1	4	83	3	3	12
Franklin	2,454	2,046	17	85	1	3	33	0	3	2
Gadsden	4,767	9,735	24	139	3	4	38	4	7	6
Gilchrist	3,300	1,910	52	97	0	1	29	0	2	4
Glades	1,841	1,442	12	56	0	3	9	1	0	1
Gulf	3,550	2,397	21	86	2	4	71	2	2	9
Hamilton	2,146	1,722	12	37	4	1	23	8	7	4
Hardee	3,765	2,339	17	75	0	2	30	0	2	3
Hendry	4,747	3,240	11	103	3	1	22	2	7	2
Hernando	30,646	32,644	116	1501	8	26	242	4	10	22
Highlands	20,206	14,167	64	545	6	16	127	3	7	8
Hillsborough	180,760	169,557	1138	7490	35	217	847	29	68	154
Holmes	5,011	2,177	18	94	1	7	76	3	6	2
Indian River	28,635	19,768	122	950	4	13	105	2	13	10
Jackson	9,138	6,868	40	138	0	2	102	1	4	7
Jefferson	2,478	3,041	14	76	2	1	29	1	0	0
Lafayette	1,670	789	6	26	2	0	10	1	1	0
Lake	50,010	36,571	204	1460	4	36	289	1	21	15
Lee	106,141	73,560	538	3587	30	81	305	5	34	96
Leon	39,053	61,425	330	1932	9	28	282	7	16	31
Levy	6,858	5,398	92	284	1	1	67	1	10	12
Liberty	1,317	1,017	12	19	0	3	39	0	1	2

Tabella 3.1: Dati votazioni per contea, parte I

Tabella a p.444 di *Statistical Science* vol.17 (2002)

I dati demografici delle tabelle 3.3 e 3.4 includono le seguenti variabili:

- Pop → popolazione della contea nel 1997
- Whi → percentuale di bianchi nel 1996
- Bla → percentuale di neri nel 1996

- Hisp→ percentuale di ispanici nel 1996  
(bianchi+neri+ispanici>100 perché gli ispanici includono altre razze)
- $\geq 65$ → percentuale di popolazione di età maggiore o uguale a 65 anni (calcolata dalla popolazione di tale età nel 1996 divisa la totale del 1997)
- HS → percentuale di popolazione con livello di istruzione High School (1990)
- Coll→ percentuale di popolazione con livello di istruzione College (1990)
- Inc → reddito pro-capite (1994)

County	Bush	Gore	Brow	Nade	Har	Hag	Buc	Mc	Ph	Mo
Madison	3,038	3,014	18	54	0	2	29	1	1	5
Manatee	57,952	49,177	242	2491	5	35	271	3	19	26
Marion	55,141	44,665	662	1809	13	26	563	6	22	49
Martin	33,970	26,620	109	1118	14	29	112	7	20	14
Miami-Dade	289,492	328,764	760	5352	87	119	560	35	69	124
Monroe	16,059	16,483	162	1090	1	26	47	0	3	7
Nassau	16,280	6,879	62	253	0	7	90	4	3	3
Okaloosa	52,093	16,948	313	985	4	15	267	2	33	20
Okeechobee	5,057	4,588	21	131	1	4	43	1	3	4
Orange	134,517	140,220	891	3879	13	65	446	7	41	46
Osceola	26,212	28,181	309	732	10	20	145	5	10	33
Palm Beach	152,846	268,945	743	5564	45	143	3407	302	188	103
Pasco	68,582	69,564	413	3393	19	83	570	14	16	77
Pinellas	184,823	200,629	1230	10022	41	442	1013	27	72	170
Polk	90,180	75,193	365	2062	8	59	532	5	46	36
Putnam	13,447	12,102	114	377	2	7	148	3	10	12
Santa Rosa	36,274	12,802	131	724	1	13	311	1	43	19
Sarasota	83,100	72,853	431	4069	11	94	305	5	15	59
Seminole	75,677	59,174	550	1946	6	38	194	5	18	26
St. Johns	39,546	19,502	210	1217	4	11	229	2	12	13
St. Lucie	34,705	41,559	165	1368	4	12	124	10	13	29
Sumter	12,127	9,637	53	306	2	2	114	0	3	17
Suwannee	8,006	4,075	52	180	2	4	108	0	9	5
Taylor	4,056	2,649	4	59	0	3	27	1	8	1
Union	2,332	1,407	15	33	1	0	37	0	1	0
Volusia	82,214	97,063	442	2903	8	36	496	5	20	69
Wakulla	4,512	3,838	30	149	2	3	46	1	0	6
Walton	12,182	5,642	68	265	3	11	120	2	7	18
Washington	4,994	2,798	32	93	0	2	88	0	9	5

Tabella 3.2: Dati votazioni per contea, parte II

Tabella a p.445 di *Statistical Science* vol.17 (2002)

Analizzando la Figura 3.1, che mette a confronto la percentuale dei voti a Buchanan con 12 variabili esplicative, notiamo le prime correlazioni:

1. aumentando la popolazione, Pop, il voto a Buchanan decresce (correlazione fortemente negativa), cfr. Fig. 3.1(a)
2. al crescere della popolazione ispanica, Hisp, il voto decresce (correlazione fortemente negativa), cfr. Fig. 3.1(d)
3. correlazione negativa con le variabili  $\geq 65$  e HS, cfr. Fig. 3.1(e), (f)
4. correlazione negativa più accentuata con la variabile Coll e con l'aumentare del reddito procapite, cfr. Fig. 3.1(g), (h)
5. correlazione negativa con i voti a Gore ed al candidato dei verdi Nader, cfr. Fig. 3.1(j), (l)
6. correlazione positiva invece con i voti a Bush, cfr. Fig. 3.1(i)

Nulla invece si può dire riguardo ad ipotetiche correlazioni tra il voto a Buchanan e la percentuale della popolazione bianca o nera (Whi e Bla). Non è chiara nemmeno la relazione con i voti al candidato dei liberali. Chiara invece sembra la posizione, indicata in ogni grafico da una  $x$ , della contea di Palm Beach all'interno dello stato: un'osservazione decisamente anomala almeno in 2/3 dei grafici riportati.

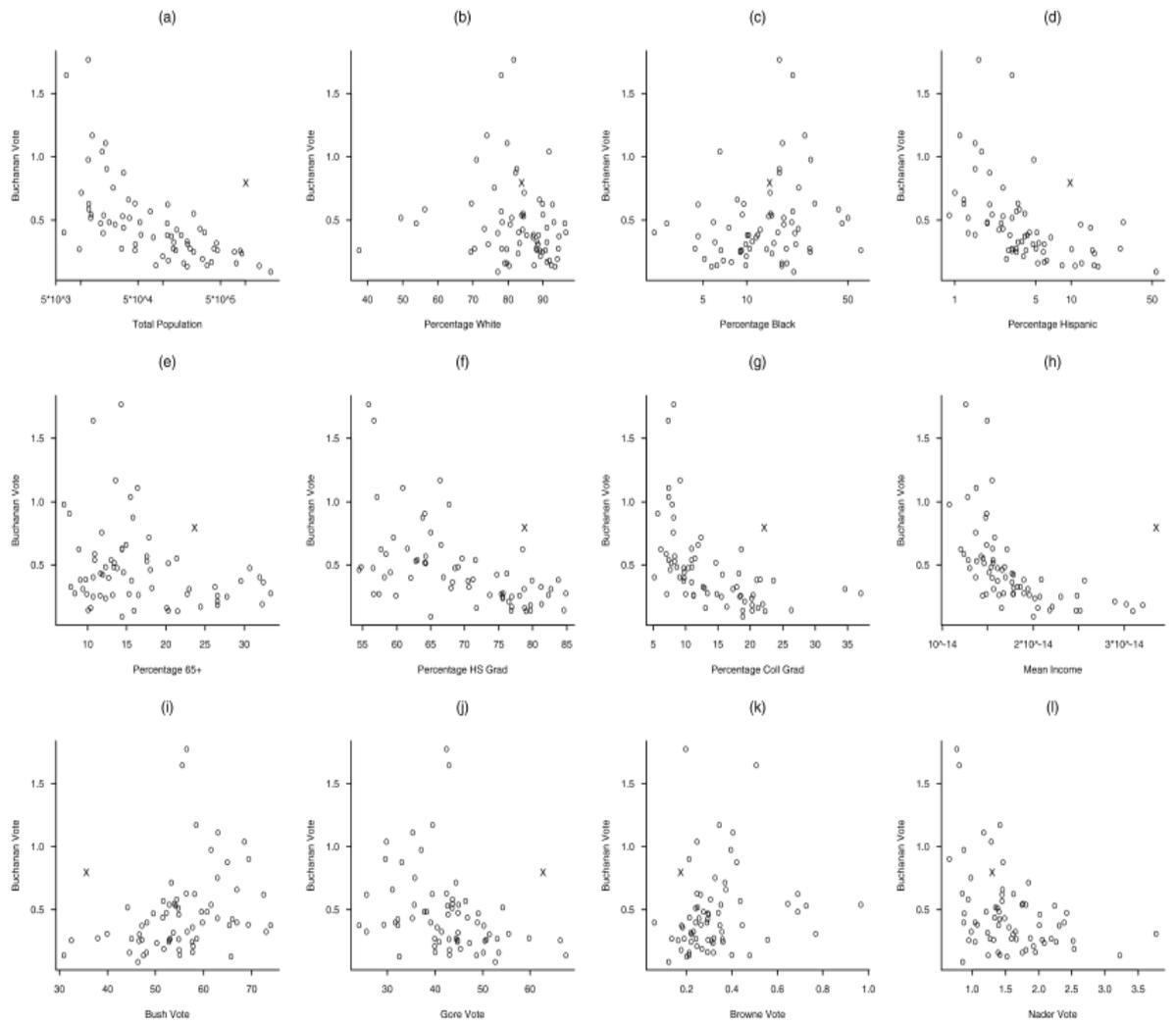


Figura 3.1: Percentuale dei voti a Buchanan rispetto a 12 variabili esplicative, Palm Beach é indicata con una x.

Grafici a p.448 di *Statistical Science* vol.17 (2002)

County	Pop	Whi	Bla	Hisp	≥ 65	HS	Coll	Inc
Alachua	198,326	74.4	21.8	4.7	9.4	82.7	34.6	19,412
Baker	20,761	82.4	16.8	1.5	7.7	64.1	5.7	14,859
Bay	146,223	84.2	12.4	2.4	11.9	74.7	15.7	17,838
Bradford	24,646	76.1	22.9	2.6	11.8	65.0	8.1	13,681
Brevard	460,977	88.3	9.2	4.1	16.5	82.3	20.4	19,567
Broward	1,470,758	80.3	17.5	10.9	20.3	76.8	18.8	24,706
Calhoun	12,337	81.6	16.9	1.6	14.3	55.9	8.2	12,570
Charlotte	133,681	94.3	4.4	3.4	33.4	75.7	13.4	18,977
Citrus	112,454	96.2	2.8	2.5	30.7	68.6	10.4	16,060
Clay	135,179	91.0	6.0	3.5	7.9	81.2	17.9	18,598
Collier	195,731	93.3	5.7	17.1	21.5	79.0	22.3	30,906
Columbia	52,856	78.3	20.5	1.9	12.3	69.0	11.0	15,349
Desoto	26,259	80.6	18.1	12.1	18.0	54.5	7.6	16,544
Dixie	12,563	89.8	9.5	1.2	14.4	57.7	6.2	12,035
Duval	732,622	69.4	27.5	3.4	10.7	76.9	18.4	20,686
Escambia	282,604	73.3	22.7	2.6	11.7	76.2	18.2	17,661
Flagler	46,128	88.5	9.8	5.9	23.0	78.7	17.3	15,613
Franklin	10,133	84.5	14.5	1.0	17.8	59.5	12.4	15,735
Gadsden	45,441	37.6	61.8	2.9	11.6	59.9	11.2	14,416
Gilchrist	13,367	90.0	9.3	2.1	13.0	63.0	7.4	12,865
Glades	9,698	79.6	13.7	10.1	15.3	57.4	7.1	14,789
Gulf	13,926	73.9	25.2	1.1	13.6	66.4	9.2	15,482
Hamilton	12,521	56.3	43.0	3.6	10.9	58.4	7.0	12,357
Hardee	22,113	93.1	5.9	28.4	13.3	54.8	8.6	16,812
Hendry	31,634	78.2	18.8	26.6	9.9	56.6	10.0	17,823
Hernando	125,537	94.4	4.6	4.0	29.6	70.5	9.7	16,062
Highlands	76,854	87.1	11.6	6.7	32.4	68.2	10.9	17,655
Hillsborough	909,444	82.8	14.9	16.0	12.3	75.6	20.2	20,167
Holmes	18,382	91.7	6.5	1.7	15.5	57.1	7.4	12,790
Indian River	99,215	89.2	9.9	3.9	26.6	76.5	19.1	28,977
Jackson	45,706	69.5	29.6	3.5	14.4	61.6	10.9	15,519
Jefferson	13,232	49.4	50.1	1.3	13.4	64.1	14.7	15,574
Lafayette	6,289	83.0	16.4	5.1	10.7	58.2	5.2	13,663
Lake	196,214	88.2	10.9	3.8	26.3	70.6	12.7	18,269
Lee	387,091	91.1	7.8	5.9	24.4	76.9	16.4	22,053
Leon	215,170	70.4	27.3	3.1	8.4	84.9	37.1	16,705
Levy	32,254	84.4	14.2	2.6	17.6	62.8	8.3	13,745
Liberty	6,703	78.1	20.9	3.1	10.7	56.7	7.3	14,896

Tabella 3.3: Dati demografici per contea, parte I

Tabella a p.446 di *Statistical Science* vol.17 (2002)

County	Pop	Whi	Bla	Hisp	≥ 65	HS	Coll	Inc
Madison	17,558	53.9	45.6	1.9	13.8	56.5	9.7	13,002
Manatee	237,159	89.8	9.0	5.8	27.8	75.6	15.5	23,031
Marion	237,308	84.3	14.6	4.0	21.4	69.6	11.5	14,502
Martin	116,087	91.8	6.9	6.2	26.6	79.7	20.3	31,996
Miami-Dade	2,044,600	77.0	21.2	54.4	14.4	65.0	18.8	20,014
Monroe	81,919	92.3	6.2	15.8	15.9	79.7	20.3	25,160
Nassau	54,096	87.3	11.9	1.5	9.8	71.2	21.5	20,874
Okaloosa	167,580	85.3	10.3	4.2	9.1	83.8	21.0	18,959
Okeechobee	33,102	91.1	7.5	14.8	14.6	59.1	9.8	15,162
Orange	783,974	79.1	17.5	12.3	10.4	78.8	21.2	20,469
Osceola	142,128	90.7	6.6	15.3	13.2	73.7	11.2	16,256
Palm Beach	1,018,524	83.9	14.4	9.8	23.7	78.8	22.1	33,518
Pasco	320,253	96.5	2.3	4.4	32.0	66.9	9.1	16,924
Pinellas	871,766	89.1	9.0	3.1	26.6	78.1	18.5	24,796
Polk	448,646	83.3	15.4	5.3	18.2	68.0	12.9	17,824
Putnam	70,430	78.1	20.9	3.4	17.6	64.3	8.3	14,250
Santa Rosa	114,481	92.6	4.6	2.0	8.9	78.5	18.6	17,127
Sarasota	301,644	94.0	5.1	2.8	32.3	81.3	21.9	30,205
Seminole	344,729	87.4	9.8	8.4	10.1	84.6	26.3	21,815
St. Johns	112,707	88.7	10.1	3.0	15.6	79.9	23.6	25,637
St. Lucie	179,559	79.6	19.0	5.2	20.1	71.7	13.1	16,483
Sumter	39,428	81.0	18.1	3.1	20.3	64.3	7.8	14,606
Suwannee	33,077	82.2	16.9	2.0	15.8	63.8	8.2	14,773
Taylor	18,718	77.1	21.5	1.3	12.7	62.1	9.8	15,459
Union	12,359	71.0	27.8	4.8	7.0	67.7	7.9	10,783
Volusia	419,797	88.0	10.5	5.0	22.7	75.4	14.8	17,778
Wakulla	19,172	83.9	14.9	0.9	10.9	71.6	10.9	15,570
Walton	37,914	88.9	8.6	1.2	14.9	66.5	11.9	14,866
Washington	20,221	79.7	17.6	1.5	16.4	60.9	7.4	13,732

Tabella 3.4: Dati demografici per contea, parte II

Tabella a p.447 di *Statistical Science* vol.17 (2002)



## Capitolo 4

### POSSIBILI MODELLI

#### 4.1 Trasformazioni della variabile risposta

La principale difficoltà tecnica di adattamento di un modello di regressione ai dati disponibili è la grande variabilità della numerosità della popolazione delle varie contee.

Sia  $N_i$  il numero totale dei voti validi e  $y_i$  il numero delle preferenze a Buchanan per la contea  $i$ -esima,  $i = 1, \dots, 67$ . Se si assume per  $y_i$  un modello binomiale con indice  $N_i$  e probabilità di successo  $p_i$ , si ha:

$$(4.1) \quad \text{Var}(y_i) = N_i p_i (1 - p_i)$$

dove  $p_i$  è la proporzione attesa dei suffragi elettorali a Buchanan nella contea  $i$ -esima. Dal momento che le stime di  $p_i$  si aggirano attorno a 0.003 (cfr. p.445 di *Statistical Science* vol.17, 2002) la (4.1) diventa:

$$(4.2) \quad \text{Var}(y_i) \approx N_i p_i$$

Inoltre, dato che  $p_i$  è molto piccolo ed  $N_i$  molto grande, la distribuzione binomiale può essere approssimata da una Poisson per la quale la (4.2) è esatta.

Di conseguenza qualsiasi tentativo di fare un'analisi della regressione con il metodo dei minimi quadrati si scontra con l'eteroschedasticità dei dati, non soddisfacendo, quindi, una delle ipotesi base per poter applicare lo stesso metodo dei minimi quadrati. Un modo per tentare di risolvere il problema è trasformare la variabile risposta in modo da considerare un modello della forma:

$$(4.3) \quad h(y_i) = \sum_j x_{ij} \beta_j + \varepsilon_i$$

dove  $h$  è la funzione di trasformazione,  $\{ x_{ij}, 1 \leq i \leq n, 1 \leq j \leq p \}$  sono i regressori, o variabili esplicative e gli  $\{\varepsilon_i\}$  sono gli errori casuali. Una possibile trasformazione è  $h(y_i) = \sqrt{y_i}$  che approssimativamente stabilizza la varianza di una variabile casuale di Poisson. Un'altra possibilità potrebbe essere la trasformazione logaritmica, come applicata in varie altre analisi effettuate da altri ricercatori, ma questa trasformazione non stabilizza la varianza, quindi si rimane con il problema dell'eteroschedasticità.

Si possono però fondere entrambe le trasformazioni suddette nella famiglia di trasformazioni di Box e Cox (1964):

$$(4.4) \quad h(y_i) = C_\lambda \frac{y_i^\lambda - 1}{\lambda}$$

dove  $C_\lambda$  è una costante scalare e  $\lambda$  può essere un qualsiasi numero reale. Per  $\lambda \rightarrow 0$  si ha la trasformazione logaritmica.

Nei primi passi dell'analisi sono state usate tre possibili variabili risposta:  $y_i$ ,  $\sqrt{y_i}$ , e  $\log(y_i / N_i)$ .

## 4.2 Scelta dei regressori (o variabili esplicative)

Dopo aver applicato le trasformazioni del paragrafo precedente, l'analisi della regressione continua ricercando quali variabili esplicative sono significative da un punto di vista statistico (e reale):

- dai grafici della Fig. 3.1 del capitolo precedente si può evincere che alcuni regressori sono più significativi se espressi in scala logaritmica, quindi non sono state considerate le loro versioni lineari,
- inoltre è stata trascurata anche la proporzione dei voti a Gore, in quanto tale variabile è fortemente correlata alla proporzione dei voti a Bush (la somma di queste due quantità si avvicina spesso al 100% dei voti validi).

Ecco la tabella con la lista completa di tutte le variabili esplicative considerate:

<b>Variabili esplicative</b>	<b>Definizioni</b>
lpop	Log della popolazione totale
whit	Proporzione di bianchi
lblac	Log della proporzione di neri
lhispanic	Log della proporzione di ispanici
o65	Proporzione di pop. sopra i 65 anni
hsed	Proporzione di pop. High School
coll	Proporzione di pop. College
inco	Reddito pro-capite
pbush	Proporzione di voti per Bush
pbrow	Proporzione di voti per Browne
pnader	Proporzione di voti per Nader

Tabella 4.1: Variabili esplicative

Si ricorda che l'osservazione di Palm Beach, evidenziata come anomala nel capitolo precedente, é stata omessa da tutte le analisi per la scelta dei regressori. La scelta delle variabili esplicative può essere fatta usando qualsiasi tecnica utilizzata per la regressione multipla. I criteri usati da Smith in questa analisi sono il  $C_p$  di Mallows (1973) e la selezione *backward*.

Le variabili scelte tramite questi due metodi sono elencate nella seguente tabella 4.2:

<b>Variabile risposta</b>	<b>Metodo di selezione</b>	<b>Variabili selezionate</b>
$y_i$	$C_p$	lpop, whit, lhis, o65, hsed, coll, pbush, pnade
$y_i$	<i>Backward</i>	lpop, whit, lhis, o65, hsed, pbush, pnade
$\sqrt{y_i}$	$C_p$	whit, lhis, o65, hsed, inco, pbrow
$\sqrt{y_i}$	<i>Backward</i>	whit, lhis, o65, hsed, inco, pbrow
$\log(y_i / N_i)$	$C_p$	lpop, lhis, hsed, inco, pbush
$\log(y_i / N_i)$	<i>Backward</i>	lhis, hsed, inco, pbush

Tabella 4.2: Variabili esplicative selezionate tramite i metodi  $C_p$  di Mallows e *backward*

Si nota subito che la trasformazione della variabile risposta  $y_i$  tramite la radice quadrata produce gli stessi regressori con i due diversi metodi. Ciò non si verifica con le altre due trasformazioni.

In aggiunta, la somma dei quadrati dei residui per i tre modelli selezionati dal metodo  $C_p$  di Mallows sono:

- 127298 con 57 gradi di libertà per la variabile risposta  $y_i$   
(si ricorda che i gradi di libertà sono dati dal n° di contee considerate, 66, meno gli 8 regressori scelti dal metodo applicato e la costante,  $66-9=57$ )
- 72712 con 59 gradi di libertà per la trasformazione della variabile risposta,  $\sqrt{y_i}$  (dalla tabella 6 vediamo che i regressori corrispondenti sono 6 più la costante,  $66-7=59$ )
- 91933 con 60 gradi di libertà per la trasformazione della variabile risposta,  $\log(y_i / N_i)$  (sempre dalla tabella precedente estraiamo i 5 regressori più la costante,  $66-6=60$ )

queste osservazioni fanno chiaramente propendere per il modello che ha come variabile risposta la trasformazione tramite radice quadrata dei voti a Buchanan ( $\sqrt{y_i}$ ). Si nota inoltre che la trasformazione logaritmica é comunque migliore della variabile risposta usata direttamente.

### 4.3 Test di eteroschedasticità

Dopo aver determinato il modello più appropriato, ora possiamo effettuare dei test diagnostici per valutare la presenza o meno di eteroschedasticità nei residui del modello adottato.

Supponiamo che il modello della forma (4.3) sia ben adattato ai dati e dopo aver calcolato la stima  $\hat{\beta}_j$  dei parametri  $\beta_j$  possiamo ricavare i residui:

$$(4.5) \quad e_i = h(y_i) - \sum_j x_{ij} \hat{\beta}_j, \quad 1 \leq i \leq n$$

Ora possiamo considerare i diagrammi di dispersione dei valori assoluti dei residui rispetto alle possibili trasformazioni di  $y_i$  osservando se la variabilità dei residui aumenta o diminuisce nel grafico, indicando quindi la presenza di eteroschedasticità nei residui stessi.

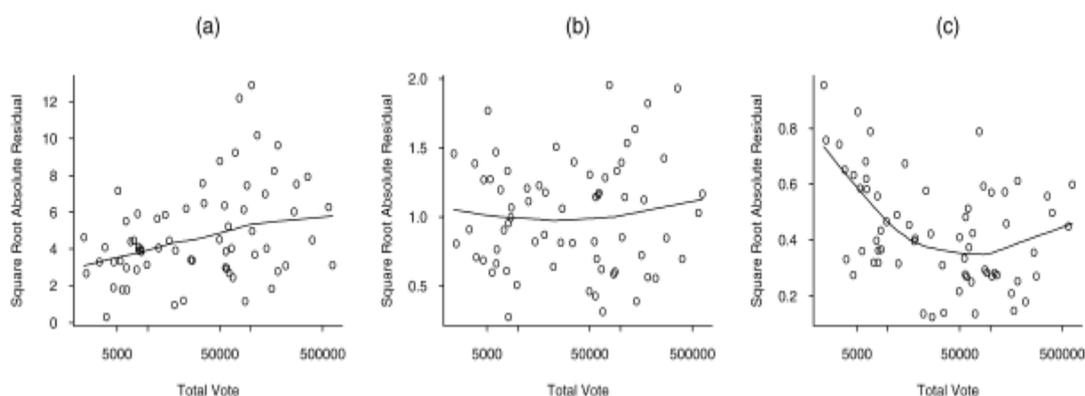


Figura 4.1: Grafici relativi alle radici quadrate dei valori assoluti dei residui rispetto alle tre diverse variabili risposta

Figure a p.450 di *Statistical Science* vol.17 (2002)

Nel grafico (a) della Figura 4.1 c'è la radice quadrata dei valori assoluti dei residui della regressione basata sulla variabile risposta diretta,  $y_i$ , e balza subito all'occhio che i residui tendono a crescere all'aumentare del totale dei voti, quindi è presente eteroschedasticità. Nella parte (b) della medesima figura la radice quadrata dei valori assoluti dei residui è relativa alla regressione della prima trasformazione della variabile risposta,  $\sqrt{y_i}$ . Qui traspare una discreta costanza dei residui all'aumentare dei voti totali, siamo quindi in presenza di omoschedasticità.

Nell'ultima parte della figura la variabilità della radice quadrata dei valori assoluti dei residui, risultanti stavolta dalla regressione della variabile risposta  $\log(y_i / N_i)$ , è ancor più evidente rispetto al primo grafico, quindi è chiara la presenza di eteroschedasticità.

Per avvalorare l'ipotesi di omoschedasticità della trasformazione  $\sqrt{y_i}$  sono stati fatti altri grafici rispetto ad altre variabili e si è arrivati sempre alla medesima conclusione.

#### 4.4 Altre diagnostiche di adattamento del modello

In questo paragrafo si vuole valutare l'influenza del dato riferito alla contea di Palm Beach, rispetto al modello di regressione scelto. Innanzitutto consideriamo due diagnostiche particolari:

- residui standardizzati
- DFFITS (si veda ad esempio Belsley, Kuh e Welsch, 1980).

Per quanto riguarda i primi la definizione è:

$$(4.6) \quad d_i^* = \frac{h(y_i) - \sum_j x_{ij} \hat{\beta}_{j(i)}}{s_{(i)}}$$

dove  $\hat{\beta}_{j(i)}$  è la stima dei coefficienti  $\beta_j$  e  $s_{(i)}$  è la stima dello scarto quadratico medio  $\sigma$ , basato su tutte le osservazioni tolta la  $i$ -esima. Mentre i valori DFFITS sono dati da una standardizzazione delle differenze  $\hat{y}_i - \hat{y}_{(i)}$ , con  $\hat{y}_{(i)}$  pari al valore previsto dal modello stimato senza l' $i$ -esima osservazione.

Per valutare realmente se il numero di voti a Buchanan nella contea di Palm Beach sia un'osservazione influente o anomala il modello di regressione prescelto, quello con la radice quadrata della variabile

risposta,  $\sqrt{y_i}$ , é stato riadattato includendo anche la rilevazione corrispondente a quella contea ed usando i regressori corrispondenti a tale variabile risposta (cfr. tabella 4.2). I valori dei residui standardizzati e DFFITS sono riportati in Figura 4.2, secondo la tecnica suggerita da Atkinson (1985):

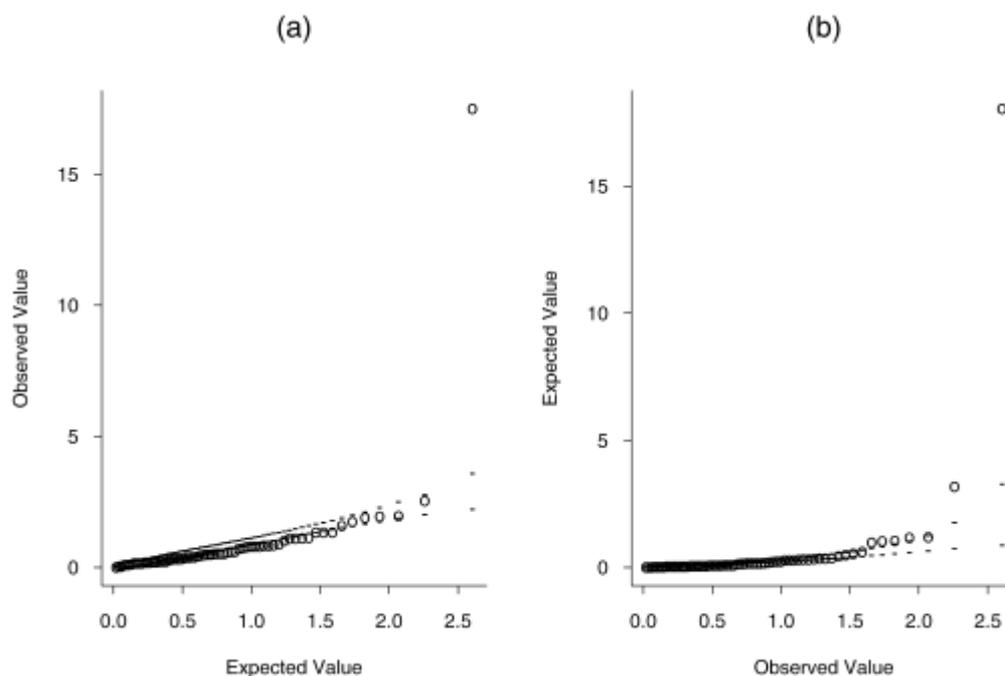


Figura 4.2: Valori attesi ed osservati dei residui standardizzati e DFFITS

Figure a p.452 di *Statistical Science* vol.17 (2002)

Nei grafici, vengono inoltre riportate le “bande di simulazione” (cfr. Atkinson, 1985). In entrambi i casi c’è un’osservazione notevolmente anomala che corrisponde a Palm Beach. Questa é la prima chiara evidenza in questo studio che la contea in esame é davvero estremamente incoerente con il resto dei dati. Ma il dato osservato a Palm Beach era l’unico anomalo?

Dalla Figura 4.2 si nota che ci sono altri valori, specialmente nella parte (b), al di fuori dei confini di simulazione. Non sembra tuttavia chiaro se siano anomali o semplicemente distorti dalla presenza di

Palm Beach, quindi si è ripetuta l'intera analisi per ricavare residui standardizzati e DFFITS togliendo l'osservazione, ora certamente definita anomala, di Palm Beach.

I risultati della nuova analisi sono disegnati in Figura 4.3, nessun valore risulta esterno all'intervallo di simulazione:

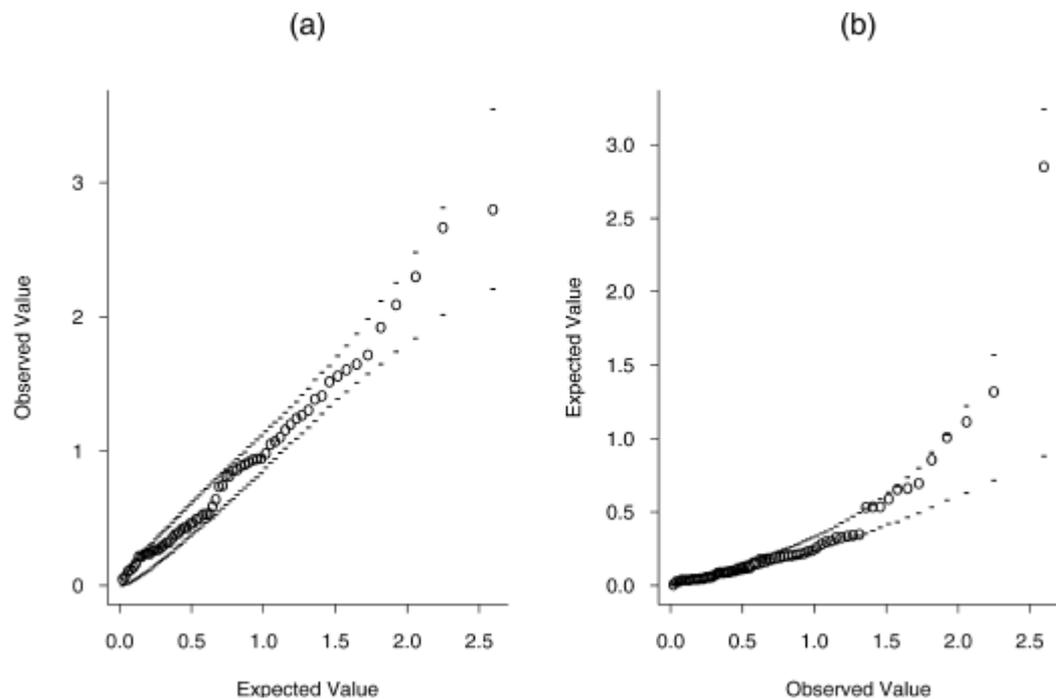


Figura 4.3: Come la figura 4.2 omettendo il dato di Palm Beach

Figure a p.453 di *Statistical Science* vol.17 (2002)

#### 4.5 Test della normalità dei dati

Un altro importante aspetto da considerare quando si adatta un modello di regressione lineare normale riguarda la normalità dei dati. Dai grafici del paragrafo precedente si evince (cfr. Figura 4.2a) che i dati non sembrano assolutamente approssimabili con una normale, infatti l'anomalia di Palm Beach sconvolge l'andamento della curva; d'altro canto, nella Figura 4.3(a) senza la contea in questione, le ipotesi di normalità sembrano soddisfatte.

Comunque, tali discrepanze richiedono dei test formali di normalità:

- la statistica di Looney e Gullede (1985), simile al più conosciuto test di Shapiro-Wilk
  - la statistica di Kolmogorov-Smirnov
  - la statistica di Cramér-von Mises
  - la statistica di Anderson-Darling (cfr. D'Agostino e Stephens, 1986)
- Smith (2002) ha riportato i risultati basati sempre sui residui standardizzati della Figura 4.2(a) usando 1000 simulazioni ed ottenendo i seguenti livelli di significatività osservati relativi ai quattro test suddetti:

1. 0
2. 0.031
3. 0.008
4. 0

che portano a rifiutare l'ipotesi nulla di normalità dei dati in tutti i test effettuati. Per contro gli stessi test eseguiti sui dati della Figura 4.3(a), senza Palm Beach, hanno dato tutti e quattro livelli di significatività osservati superiori a 0.3 portando ad accettare l'ipotesi nulla di normalità dei dati.

I medesimi test sono stati applicati anche al modello dove la variabile risposta era  $\log(y_i / N_i)$  con risultati analoghi, cioè rifiuto dell'ipotesi nulla con il dato di Palm Beach incluso ed ipotesi nulla accettata se tale dato viene escluso. Tale risultato non si ripeteva con la variabile risposta diretta,  $y_i$ , infatti l'ipotesi nulla viene rifiutata anche se il dato di Palm Beach viene escluso. Questo chiaramente rafforza la convinzione che, mentre i due modelli basati su  $\sqrt{y_i}$  e  $\log(y_i / N_i)$  danno un adattamento accettabile ad una distribuzione normale, quello con la variabile risposta  $y_i$  non lo è affatto.

## Capitolo 5

### APPROFONDIMENTI

#### 5.1 Prevedere il voto di Palm Beach

Ritorniamo ora allo scopo precipuo dell'analisi e cioè la previsione del voto di Palm Beach da un modello di regressione lineare dove sono stati considerate solo le osservazioni delle altre 66 contee, e non il dato relativo alla contea di Palm Beach stessa.

Visti i precedenti risultati, lo studio di Smith si é focalizzato sui modelli relativi alle due trasformazioni della variabile risposta,  $\sqrt{y_i}$  e  $\log(y_i / N_i)$  ed ha considerato i corrispondenti modelli di selezione delle variabili esplicative,  $C_p$  di Mallows e la selezione *backward*, come mostrato nella tabella 4.2.

La seguente tabella 5.1 evidenzia i risultati ottenuti, sia in termini di stima puntuale che di intervallo di previsione al 95%.

Variabile risposta	Metodo di selezione	Previsione puntuale	Intervallo di previsione
$\sqrt{y_i}$	$C_p$ o <i>Backward</i>	371	(219,534)
$\log(y_i / N_i)$	$C_p$	363	(180,735)
$\log(y_i / N_i)$	<i>Backward</i>	371	(182,758)

Tabella 5.1: Stime ed intervalli di previsione con tre versioni di modelli lineari

Da questi risultati si nota chiaramente che le stime puntuali per le tre versioni dei modelli considerati sono decisamente vicine ed anche gli intervalli di previsione sono sufficientemente coerenti per sostenere l'ipotesi che il voto di Buchanan nella contea di Palm Beach, che si rammenta essere stato di 3407 unità, sia stato oltremodo anomalo rispetto alle previsioni basate su un qualsiasi modello di regressione ragionevole.

E' impressionante osservare che anche l'estremo superiore dell'intervallo più ampio é meno del 25% del dato osservato.

## Capitolo 6

### CONCLUSIONI

Smith (2002) riporta che analisi precedenti a quella da lui effettuata, relative ai dati delle elezioni in Florida, si sono basate su modelli di regressione aventi come variabili risposta  $y_i$  e  $\log(y_i / N_i)$ . La propensione di parecchi ricercatori era verso la variabile logaritmica, mentre altri, addirittura, attestavano la medesima efficacia dei due modelli usati rispetto all'effetto Palm Beach.

L'indagine di Smith (2002) invece conferma che le analisi effettuate usando  $y_i$  come variabile risposta falliscono per:

- mancanza di omoschedasticità
- impossibilità di approssimazione con una distribuzione normale.

Rispetto agli altri studi eseguiti su modelli con la trasformazione della variabile risposta  $\log(y_i / N_i)$  l'indagine di Smith ha rilevato una significativa eteroschedasticità. L'analisi basata sulla trasformazione  $\sqrt{y_i}$  sembra soddisfacente sotto tutti i punti di vista.

In tutti i casi, il risultato elettorale di Palm Beach é un dato veramente fuori da ogni giustificazione statistica. La ricerca di Smith (2002) accerta in maniera chiara e conclusiva che il voto del candidato di estrema destra, Pat Buchanan, nella contea di Palm Beach non può essere spiegato come una normale variazione statistica e la ragionevole distorsione del voto in quello stato é almeno di 2500 voti.

Per concludere, si vuol far notare che, nonostante i tanti sforzi effettuati da molti statistici per trovare un modello per rendere plausibile il risultato elettorale ottenuto da Pat Buchanan a Palm Beach nessuno c'è riuscito. Questo chiaramente lascia molti dubbi riguardo l'affidabilità delle votazioni stesse, ma soprattutto sul modo di disegnare una scheda elettorale.

Meno importante è sembrata la modalità del conteggio dei voti alla quale era stato dato molto risalto nell'immediatezza dell'evento.

Tornando all'aspetto statistico della questione, appare evidente che, per trovare un modello statistico che si adattasse ai dati l'unico modo ragionevole per farlo era quello di escludere il voto di Palm Beach. Tale considerazione porta a concludere che ci sono state realmente delle situazioni anomale non giustificabili statisticamente e che, come accennato nel primo capitolo, le scienze statistiche mettono a disposizione dei ricercatori ed analisti degli ottimi strumenti per studiare casi reali altrimenti impossibili da studiare.

## Ringraziamenti

Il primo pensiero e ringraziamento più grande va a mia moglie Patrizia, senza il suo aiuto morale ma anche materiale soprattutto nell'organizzazione dei momenti di studio non sarei mai arrivato a questo momento. I miei figli Luca e Giorgio, in particolar modo in questi ultimi mesi, li ho sentiti molto vicini e mi hanno aiutato sopportandomi quando ero nervoso a causa dello studio.

Ricordo inoltre tutti i colleghi ed amici di IBM che hanno ascoltato pazientemente tutte le mie storie su esami e professori. Un altro pensiero particolare va agli amici di Banca Antonveneta che mi hanno sopportato, ma anche *supportato* soprattutto nella stesura di questo elaborato.

Alla memoria di mia madre che mi ha sempre spinto a studiare ed a mio padre che quando ho ripreso l'università mi ha finalmente capito.

Infine vorrei ringraziare i docenti che ho incontrato in questi tre anni, perché grazie ad alcuni di loro mi sono appassionato alla statistica, il prof. Masarotto, ed inoltre alla professoressa Salvan che con grande disponibilità mi ha aiutato in quest'ultima fatica.

## BIBLIOGRAFIA

- ADAMS, G. (2001). Voting irregularities in Palm Beach, Florida. *Chance* **14**(1) 22-24.
- AGRESTI, A. e PRESNELL, B. (2002). Misvotes, Undervotes and Overvotes: The 2000 Presidential Election in Florida. *Statistical Science* **17**(4) 436-440.
- ALDRICH, J. H., NIEMI, R.G., RABINOWITZ, G. e RODHE, D. W. (1982). The measurement of public opinion about public policy: A report on some new issue question formats. *American Journal of Political Science* **26** 391-414.
- ATKINSON, A. C. (1985). *Plots, Transformations and Regression*. Oxford Univ. Press.
- BELSLEY, D., KUH, E. e WELSCH, R. E. (1980). *Regression Diagnostics*. Wiley, New York.
- Science. *Statistical Science* **17**(4) 381-382.
- BOX, G.E.P. e COX, D.R. (1964). An analysis of transformations (with discussion). *J.Roy. Statist. Soc. Ser. B* **26** 211-252.
- CASELLA, G. e GILL, J. (2002). Voting, Elections and Statistical Science. *Statistical Science* **17**(4) 381-382.
- D'AGOSTINO, R. B. e STEPHENS, M. A., eds. (1986). *Goodness-Of-Fit Techniques*. Dekker, New York.
- LOONEY, S. W. E GULLEDGE, T. R. (1985). Use of correlation Coefficient with normal probability plots. *Amer. Statist.* **39** 75-79.
- MALLOWS, C. L. (1973). Some comments on  $C_p$ . *Technometrics* **15** 661-675.
- SMITH, R. L. (2002). A Statistical Assessment of Buchanan's Vote in Palm Beach County. *Statistical Science* **17**(4) 441-457.

[www.skirsch.com/politics/election2000/reallocation.htm](http://www.skirsch.com/politics/election2000/reallocation.htm)

[www.skirsch.com/politics/election2000/palm\\_beach\\_balloting.htm](http://www.skirsch.com/politics/election2000/palm_beach_balloting.htm)