Government 320: Public Opinion and Public Choice

Spring 2007

Tuesday and Thursday 2:55–4:10 (MG 165)

Professor: Walter R. Mebane, Jr.

Office: 217 White Hall (255-3868); email `wrm1@cornell.edu`

Office hours: M 2–4 or other times by appointment.

Course web page:

`http://macht.arts.cornell.edu/wrm1/gov320.html`

- **election fraud: is fraud (legitimate) political manipulation?**
- **detecting anomalies**
- **distinguishing anomalies from fraud**
- **diagnosing fraud**

- **election fraud: is fraud (legitimate) political manipulation?**
- **detecting anomalies**
- **distinguishing anomalies from fraud**
- **diagnosing fraud**
- **history of fraudulent elections in the United States**

- **election fraud: is fraud (legitimate) political manipulation?**
- **detecting anomalies**
- **distinguishing anomalies from fraud**
- **diagnosing fraud**
- **history of fraudulent elections in the United States**
- **elsewhere (and election monitoring: observers, PVT)**

- **detecting anomalies**
- **Florida 2000: wrong outcome, but why?**
  - **ex-felon lists**
  - **butterfly ballot**
  - **other machines and ballots**

- detecting anomalies
- Florida 2000: wrong outcome, but why?
  - ex-felon lists
  - butterfly ballot
  - other machines and ballots
- Florida 2004: fraud alleged
  - conservative Democrats
  - hacked machines?

- **Election Forensics**
  - **statistically analyzing recorded vote counts to detect anomalies and try to diagnose fraud**
- **regularities and departures from regularities**
  - **using relationships with covariates to detect outliers**
  - **checking whether vote counts match expected distributions**

- **election forensics and recounts**
  - **two kinds of errors (or frauds) in vote counts**
    - **miscounting the ballots that were cast**
    - **counting falsified ballots**

- **election forensics and recounts**
  - **two kinds of errors (or frauds) in vote counts**
    - **miscounting the ballots that were cast**
    - **counting falsified ballots**
- **recounts can detect the first kind but not the second kind**
  - **exception: physically inspecting ballots may spot signs that some or all are fake**
  - **this depends on there being physical ballots to inspect**
- **statistical analysis may be able to detect both kinds of distortions**

- an example from the 2006 Mexican presidential election
  - relationship between presidential votos nulos and senate votos nulos
  - use casilla (ballot box) counts
  - the linear predictor is

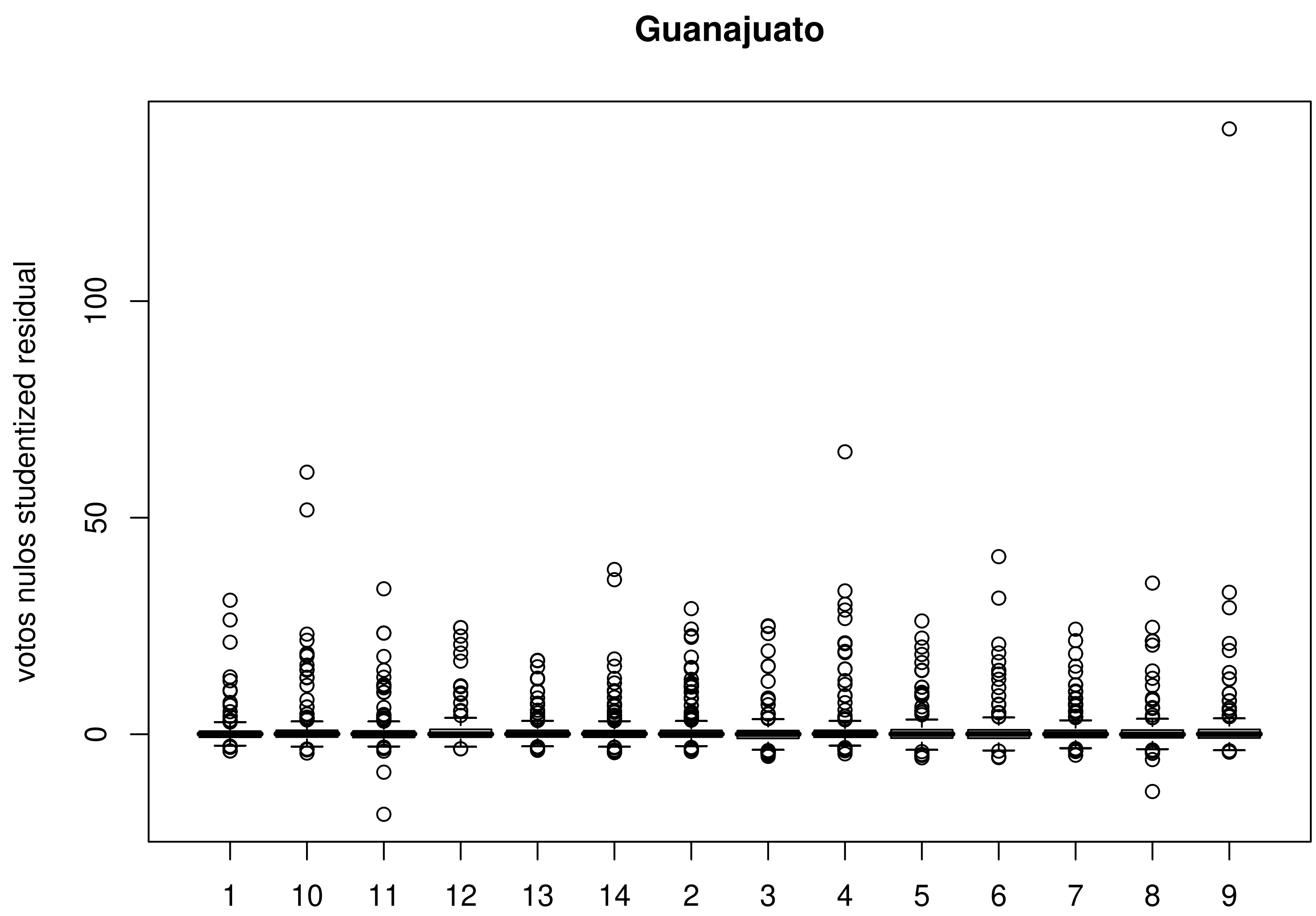
$$Z_i = d_0 + d_1 \textbf{logitz}(\textbf{SenateVN}_i)$$

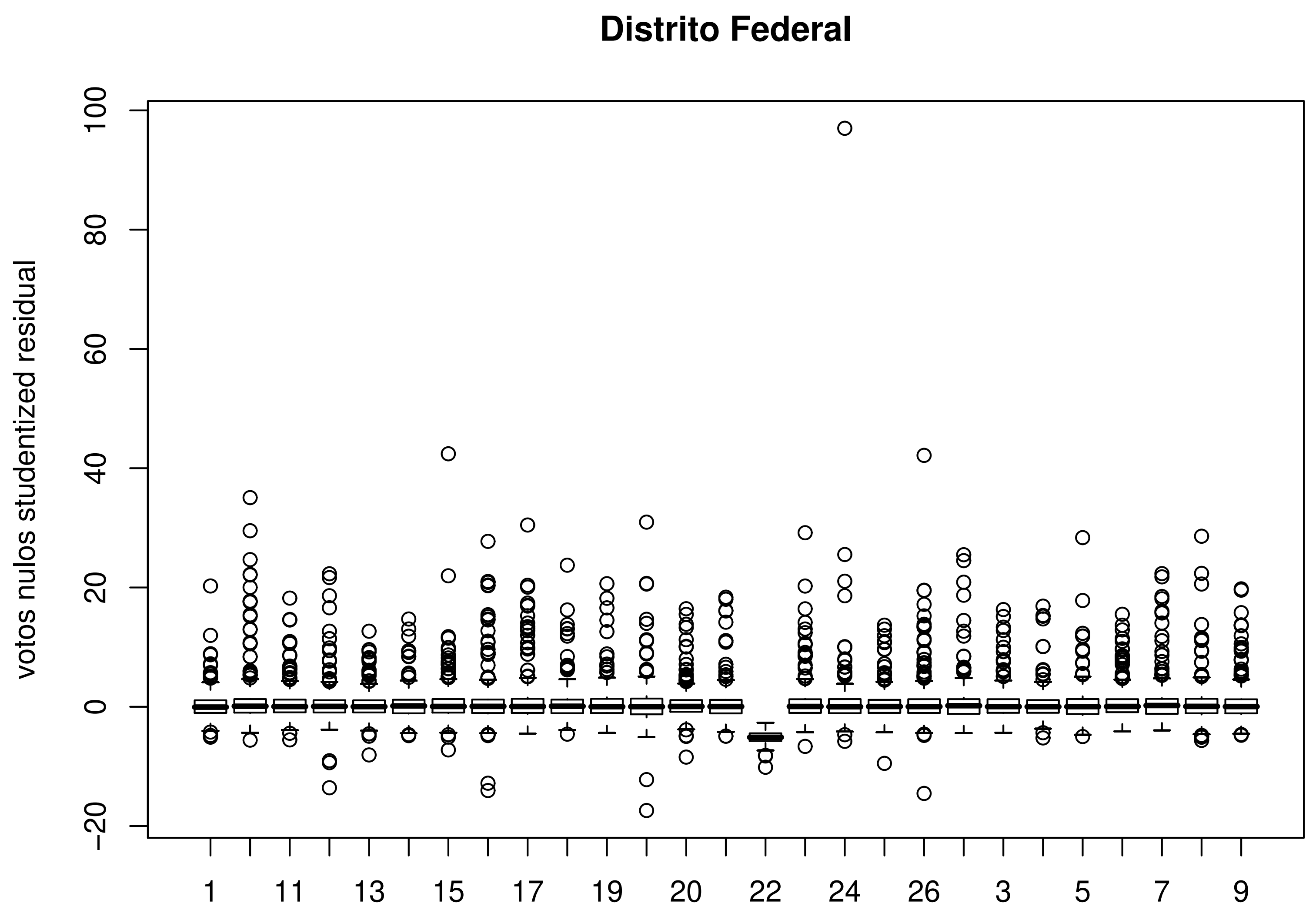
    SenateVN represents the proportion of votos nulos for senate votes at casilla i
    logitz(p) denotes the log-odds function adjusted to handle zero counts (add $1/2$ to each count before computing p)

- an example from the 2006 Mexican presidential election
  - relationship between presidential votos nulos and senate votos nulos
  - use casilla (ballot box) counts
  - the linear predictor is

    $$Z_i = d_0 + d_1 \textbf{logitz}(\textbf{SenateVN}_i)$$

    SenateVN represents the proportion of votos nulos for senate votes at casilla i
    logitz(p) denotes the log-odds function adjusted to handle zero counts (add $1/2$ to each count before computing p)
  - estimate separately for each legislative district
  - outliers are prevalent


Guanajuato
votos nulos studentized residual
0
50
100
1
10
11
12
13
14
2
3
4
5
6
7
8
9



Distrito Federal
votos nulos studentized residual
−20
0
20
40
60
80
100
1
11
13
15
17
19
20
22
24
26
3
5
7
9

- an example from the 2006 Mexican presidential election
  - relationship between presidential votos nulos and senate votos nulos
  - use casilla (ballot box) counts
  - estimate separately for each legislative district
  - outliers are prevalent
    - 130,020 casillas are in the analysis (from 299 districts)

proportion of residuals larger than		
2	3	4
.11	.06	.04

- **checking whether vote counts conform with expected distributions**

- **checking whether vote counts conform with expected distributions**
- **digits of vote counts and Benford's Law**
  - **compare vote counts' second digits to the second digit Benford's Law (2BL)**
  - **there are strong arguments against expecting vote counts' first digits to satisfy Benford's Law for first digits**

Frequency of First and Second Digits according to Benford's Law

digit	0	1	2	3	4	5	6	7	8	9
first	—	.301	.176	.124	.097	.079	.067	.058	.051	.046
second	.120	.114	.109	.104	.100	.097	.093	.090	.088	.085

- **the statistic is**

$$X^2_{B_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2 q_{B_2 i})^2}{d_2 q_{B_2 i}}$$

where

- $q_{B_{2i}}$ **is the expected relative frequency with which the second significant digit is** i **(the values shown in the second line of table of Benford's Law frequencies)**
- d_{2i} **is the number of times the second digit is** i **among the precincts being considered**
- $d_2 = \sum_{i=0}^{9} d_{2i}$

- **the statistic is**

$$X^2_{B_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2 q_{B_2 i})^2}{d_2 q_{B_2 i}}$$

where

  - $q_{B_{2i}}$ **is the expected relative frequency with which the second significant digit is** i **(the values shown in the second line of table of Benford's Law frequencies)**
  - d_{2i} **is the number of times the second digit is** i **among the precincts being considered**
  - $d_2 = \sum_{i=0}^{9} d_{2i}$

- **with one set of counts (for one office in one area), use the critical value of** χ^2_9 **for test level** $\alpha = .05$**, which is 16.9**

- **looking at multiple sets of counts, control for the false discovery rate (FDR)**

- **an example from the 2004 American election: Florida, Miami-Dade County**
  - **vote counts for major party candidates for president (Kerry and Bush) and for the Senate (Castor and Martinez)**
  - **also vote counts for eight proposed constitutional amendments**
  - **with 20 tests, the FDR-controlled critical value for χ^2_9 is 25.5**

Florida Constitutional Amendments on the Ballot in 2004

		Yes	No
1	Parental Notification of a Minor's Termination of Pregnancy	4,639,635	2,534,910
2	Constitutional Amendments Proposed by Initiative	4,574,361	2,109,013
3	The Medical Liability Claimant's Compensation Amendment	4,583,164	2,622,143
4	Authorizes Voters to Approve Slot Machines in Parimutuel Facilities	3,631,261	3,512,181
5	Florida Minimum Wage Amendment	5,198,514	2,097,151
6	Repeal of High Speed Rail Amendment	4,519,423	2,573,280
7	Patients' Right to Know About Adverse Medical Incidents	5,849,125	1,358,183
8	Public Protection from Repeated Medical Malpractice	5,121,841	2,083,864

Miami-Dade Election Day First-digit Benford's Law Tests

item	Benf.	item	Benf.
Bush	29.3	Am. 4 Yes	144.8
Kerry	39.9	Am. 4 No	119.6
Martinez	35.6	Am. 5 Yes	115.4
Castor	22.0	Am. 5 No	27.6
Am. 1 Yes	86.2	Am. 6 Yes	98.8
Am. 1 No	80.5	Am. 6 No	84.0
Am. 2 Yes	95.6	Am. 7 Yes	130.3
Am. 2 No	60.0	Am. 7 No	49.9
Am. 3 Yes	60.5	Am. 8 Yes	123.0
Am. 3 No	51.5	Am. 8 No	102.6

Note: $n = 757$ precincts. Pearson chi-squared statistics, 8 df.

Miami-Dade Election Day Second-digit Benford's Law Tests

item	Benf.	item	Benf.
Bush	7.9	Am. 4 Yes	3.3
Kerry	9.5	Am. 4 No	5.7
Martinez	8.9	Am. 5 Yes	17.9
Castor	12.0	Am. 5 No	5.8
Am. 1 Yes	2.5	Am. 6 Yes	4.3
Am. 1 No	5.5	Am. 6 No	9.1
Am. 2 Yes	16.7	Am. 7 Yes	17.1
Am. 2 No	7.2	Am. 7 No	8.4
Am. 3 Yes	3.3	Am. 8 Yes	12.7
Am. 3 No	12.9	Am. 8 No	6.5

Note: $n = 757$ precincts. Pearson chi-squared statistics, 9 df.

- **why should we expect vote counts to satisfy 2BL?**
- **model vote counts as results of particular mixtures**
- **at least two mechanisms can generate counts that satisfy 2BL (and not 1BL)**
  - **mechA: mix support that varies over precincts with a small random frequency of errors**
  - **mechB: mix support that varies over precincts with varying precinct sizes**

2BL Tests for Simulated Precinct Vote Counts (First Mechanism)

Size	Benf.	Size	Benf.	Size	Benf.	Size	Benf.
500	10.3	1,500	18.6	3,800	11.3	7,100	8.3
600	9.5	1,600	21.6	3,900	9.2	7,200	9.1
700	10.0	1,700	19.9	4,000	12.2	7,300	8.9
800	9.0	1,800	17.5	4,100	10.5	7,400	9.3
900	10.0	1,900	14.0	4,200	10.4	7,500	7.8
1,000	9.7	2,000	14.1	4,300	9.1	7,600	7.9
1,100	10.4	2,100	9.7	4,400	10.2	7,700	9.1
1,200	12.0	2,200	8.7	4,500	12.3	7,800	10.9
1,300	12.3	2,300	11.6	4,600	9.9	7,900	8.7
1,400	13.4	2,400	12.2	4,700	11.2	8,000	9.0

Note: Chi-squared statistics, 9 df, 25 Monte Carlo replications.

- **why should we expect vote counts to satisfy 2BL?**
- **while precinct vote counts should satisfy 2BL, counts on voting machines used in each precinct should not**
  - **voting machine counts are subject to "roughly equal division with leftovers" (REDWL)**
  - **simulations verify the REDWL mechanism**

- why should we expect vote counts to satisfy 2BL?
- while precinct vote counts should satisfy 2BL, counts on voting machines used in each precinct should not
  - voting machine counts are subject to "roughly equal division with leftovers" (REDWL)
  - simulations verify the REDWL mechanism
- and actual machine-level vote counts do not satisfy 2BL

Miami-Dade Election Day Second-digit Benford's Law Tests

item	Benf.	item	Benf.
Bush	17.2	Am. 4 Yes	43.5
Kerry	44.0	Am. 4 No	25.4
Martinez	11.5	Am. 5 Yes	57.6
Castor	12.7	Am. 5 No	25.6
Am. 1 Yes	43.6	Am. 6 Yes	29.7
Am. 1 No	19.8	Am. 6 No	15.3
Am. 2 Yes	38.7	Am. 7 Yes	53.2
Am. 2 No	11.9	Am. 7 No	136.7
Am. 3 Yes	78.0	Am. 8 Yes	54.2
Am. 3 No	25.7	Am. 8 No	23.2

Note: $n = 7,064$ precinct-machines. Pearson chi-squared stats, 9 df.

- **the 2BL test can detect artificial manipulations of vote counts that otherwise satisfy 2BL**
- **simulations show a wide range of ways to manipulate the votes can be detected**
  - **adding votes**
  - **subtracting votes**
  - **switching votes**

Simulated "Repeater" Vote Switching: Receive Votes When Above Expectation

	Receiver (cand. 1)			Donor (cand. 2)		
fraction	500	1000	2000	500	1000	2000
0	9.6	8.7	12.4	11.1	11.9	13.0
0.01	11.2	13.3	15.0	9.3	10.3	11.4
0.02	12.7	17.7	27.1	8.8	12.2	13.2
0.03	15.5	27.2	44.1	10.5	10.7	14.2
0.04	25.6	41.8	68.9	10.9	13.1	16.9
0.05	24.8	38.1	67.2	11.2	13.6	17.1
0.06	23.6	42.2	74.2	12.0	15.1	19.3
0.07	28.2	48.4	89.9	12.9	15.6	22.1
0.08	33.5	58.1	112.8	13.5	17.3	26.5
0.09	32.7	56.5	107.7	12.9	18.0	29.3

Simulated “Repeater” Vote Switching: Receive Votes When Below Expectation

	Receiver (cand. 1)			Donor (cand. 2)		
fraction	500	1000	2000	500	1000	2000
0	9.6	10.3	12.8	9.7	10.3	12.2
0.01	10.0	13.1	15.0	10.4	11.4	14.3
0.02	12.6	18.3	28.0	11.8	12.7	19.9
0.03	18.6	26.8	50.3	13.5	18.3	22.8
0.04	25.9	44.5	80.0	12.4	19.4	26.7
0.05	26.5	45.4	74.8	16.1	21.5	31.4
0.06	28.5	46.6	87.1	14.8	21.5	37.9
0.07	33.1	57.1	102.2	17.0	24.9	42.1
0.08	39.0	71.8	128.4	16.8	26.3	45.4
0.09	38.0	68.1	126.9	19.6	27.0	40.9

- **wider application of the 2BL test: recent American presidential votes**
  - **precinct vote counts in the 2000 and 2004 elections, separately for the precincts in each county**
  - **impose FDR-control using the number of counties in each state**
    - **(see maps [in showmappbenf0004fdr.R])**

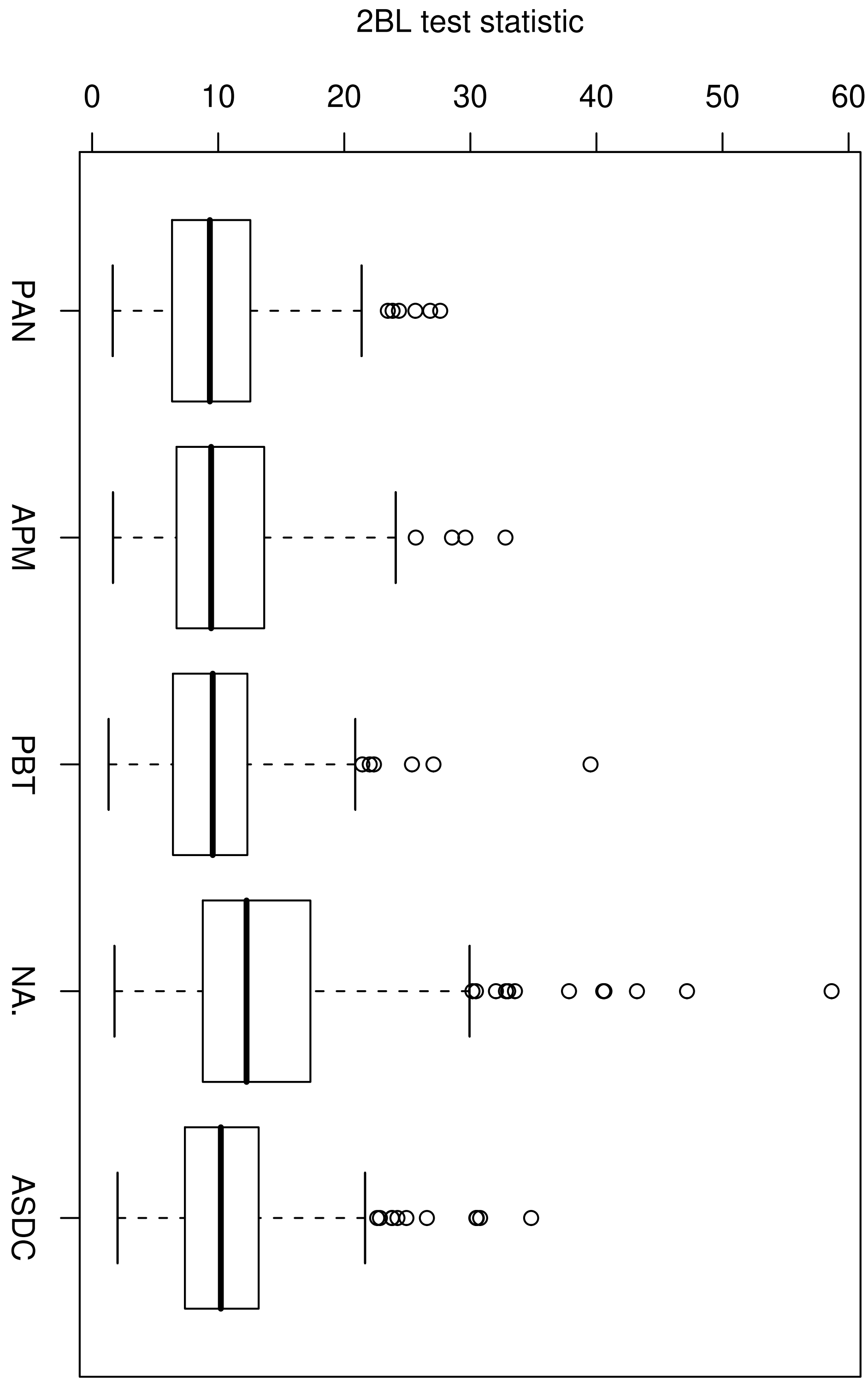
Counties with Signficant 2BL Tests using State-specific FDR Adjustment: 2000

		Gore votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Los Angeles, CA	5,045	5,011	54.8	4,930	20.3
Kent, DE	61	61	9.0	61	22.2
Latah, ID	34	31	36.7	34	3.8
Cook, IL	5,179	5,097	46.7	4,145	24.4
Dupage, IL	714	714	28.0	714	41.6
Lake, IL	403	403	33.7	402	16.1
Passaic, NJ	295	295	27.7	294	5.6
Hamilton, OH	1,025	1,020	48.7	988	8.9
Hancock, OH	67	67	34.3	67	9.9
Summit, OH	624	624	31.6	612	11.6
Philadelphia, PA	1,681	1,680	29.5	1,249	34.7
King, WA	2,683	2,665	27.0	2,641	8.9

Counties with Signficant 2BL Tests using State-specific FDR Adjustment: 2004

		Kerry votes		Bush votes	
County	J	d_2	$X^2_{B_2}$	d_2	$X^2_{B_2}$
Los Angeles, CA	4,984	4,951	70.2	4,929	12.4
Orange, CA	1,985	1,887	26.2	1,904	32.6
Jefferson, CO	324	323	30.0	323	10.4
Kootenai, ID	75	75	30.9	75	12.1
Cook, IL	4,562	4,561	44.5	4,026	27.8
DuPage, IL	732	732	35.2	732	9.1
Clay, MO	76	76	28.4	76	4.0
Summit, OH	475	475	42.7	474	21.0
Davis, UT	213	212	42.6	213	6.0
Utah, UT	247	241	9.2	246	27.6
Benton, WA	177	168	29.2	173	14.8

- **the 2BL test applied to votes for president in the 2006 Mexican election**
  - **seccion vote counts, separately for the secciones in each legislative district**
  - **over all 300 districts, the FDR-controlled critical value for χ^2_9 is 32.4**
  - **over 1500 district-party combinations, the FDR-controlled critical value for χ^2_9 is 36.4**


2BL test statistic
0
10
20
30
40
50
60
PAN
APM
PBT
NA.
ASDC

- the statistical tests and the partial recount done of votes for president in the 2006 Mexican election
  - the original count included 41,791,322 ballots
  - 40,588,729 votes were recorded for one of the parties
  - the original difference between the PAN and PBT vote totals was 243,934 votes, which is 0.58 percent of the ballots cast

- **the statistical tests and the partial recount done of votes for president in the 2006 Mexican election**
  - **the original count included 41,791,322 ballots**
  - **40,588,729 votes were recorded for one of the parties**
  - **the original difference between the PAN and PBT vote totals was 243,934 votes, which is 0.58 percent of the ballots cast**
- **the recount**
  - **about nine percent of the casillas were manually recounted**
  - **I use data from 11,651 recounted casillas (which I think is all of them)**

Net Vote Count Changes in the Mexico 2006 Recount

	PAN	APM	PBT	NA.	ASDC
original	15,000,284	9,301,441	14,756,350	401,804	1,128,850
change	$-13,333$	$-1,885$	-58	$-1,578$	$1,836$

Note: Some of the recounted votes included here are from casillas that were canceled in the final official results.

- **relationship between the 2006 Mexican recount changes and the two kinds of statistical tests**
- **definitions for casilla-level variables**

$$\textbf{CHANGE} = \begin{cases} 1, & \textbf{if the vote count changed for any party} \\ 0, & \textbf{otherwise} \end{cases}$$

$$\textbf{NULOS2} = \begin{cases} 1, & \textbf{if the votos nulos } |\textbf{residual}| \geq 2 \\ 0, & \textbf{otherwise} \end{cases}$$

- **definitions for district-level variable**

$$\textbf{2BL} = \begin{cases} 1, & \textbf{if the 2BL statistic for any party} \geq 16.9 \\ 0, & \textbf{otherwise} \end{cases}$$

Recount Changes and Test Statistics

	CHANGE		
NULOS2	0	1	n
0	0.33	0.67	9,200
1	0.28	0.72	2,215

Pearson chi-squared = 20.1

	CHANGE		
2BL	0	1	n
0	0.29	0.71	5,001
1	0.33	0.67	6,650

Pearson chi-squared = 21.5

- **relationship between the 2006 Mexican recount changes and the two kinds of statistical tests**
  - **unusually large votos nulos counts for a casilla are associated with more vote count changes if that casilla is recounted**
  - **unusually large 2BL test statistics for a district are associated with fewer vote count changes when casillas in that district are recounted**
- **does this mean that the 2BL test is picking up the fact that votes were faked, in ways that the recount did not detect?**

- **relationship between the 2006 Mexican recount changes and the two kinds of statistical tests**
- **is the 2BL test picking up the fact that votes were faked, in ways that the recount did not detect?**
- **consider the possibility of strategic voting (to mw07.pdf)**

- **is election manipulation election fraud?**
- **are either election manipulation or election fraud heresthetic?**
  - **election manipulation as dimension manipulation (unlikely)**
  - **election manipulation as agenda control**
  - **election manipulation as strategic voting**
- **the key issue is dictatorship (or oligarchy), which heresthetic (via Arrow's theorem) is normatively justified to oppose**
- **election fraud seems intuitively to be dictatorial, but why is that?**