

# Revised Sample Design— Florida Oversample and Pennsylvania National Sample Allocation

Charles H. Franklin  
*University of Wisconsin, Madison*

Saturday, June 26, 2004 at 11:59

## 1 Pennsylvania and Florida

This revision to the BYU-Wisconsin 2004 Election Panel replaces Pennsylvania with Florida as an oversample state and returns Pennsylvania to the national sample.

This change has three consequences:

1. Florida now has 400 cases.
2. Pennsylvania now has its allocation of cases based on the national sample design.
3. Other states change their sample size due to the reallocation. While most states retain their sample size, fourteen increase their sample allocation.
4. Total sample is increased to 3001.

The first two of these changes are straight forward. They affect sample weights and standard errors only slightly. Florida will have 400 cases allocated to it and join Ohio as the two state-level samples. Likewise Pennsylvania's case allocation is reduced to its share based on the national sample design.

This change also affects the allocation to all other states, though most remain unchanged. The reason is that we remove Florida from the national sample and replace it with the smaller state of Pennsylvania. In the national allocation, Florida originally received 201 cases. Under the new national allocation, Pennsylvania now receives 172 cases. The 29

case difference must be reallocated among the other states, resulting in 14 states that change their allocation, all with small increases in sample size.

Due to rounding of sample allocation, the total sample is now 3001.  
The following code

1. Clears the workspace to ensure no previously declared variables inadvertently get used.
2. Reads the data, recodes the battleground as before and subsets to remove Alaska and Hawaii.
3. Introduces Ohio and Florida as new strata which will have sample sizes fixed at 400.
4. Calculates the allocation of cases across all states.
5. Calculates the sample characteristics for all the resulting domains and populations of interest.

```
> rm(list = ls(all = TRUE))
> setwd("c:/data/campaign2004/news1absurvey")
> library(foreign)
> library(xtable)
> pop <- read.dta("statetoandpop.dta")
> attach(pop)
> pop$bgres[state == "CO"] <- 1
> pop$bgres[state == "LA"] <- 1
> attach(pop)
> pop$bg <- bgpres
> pop$bg[bgres == 0 & bgren == 0] <- 0
> pop$bg[bgres == 0 & bgren == 1] <- 1
> pop$bg[bgres == 1] <- 2
> pop$bg[state == "OH"] <- 3
> pop$bg[state == "FL"] <- 4
> attach(pop)
> pop3 <- subset(pop, state != "AK" & state != "HI")
> attach(pop3)
> table(bg)
```

```
bg
 0  1  2  3  4
22  8 17  1  1
```

```
> table(state, bg)
```

```
      bg
state 0 1 2 3 4
AL  1 0 0 0 0
AR  0 0 1 0 0
AZ  0 0 1 0 0
CA  0 1 0 0 0
CO  0 0 1 0 0
CT  1 0 0 0 0
DC  1 0 0 0 0
DE  1 0 0 0 0
FL  0 0 0 0 1
GA  0 1 0 0 0
IA  0 0 1 0 0
ID  1 0 0 0 0
IL  0 1 0 0 0
IN  1 0 0 0 0
KS  1 0 0 0 0
KY  0 1 0 0 0
LA  0 0 1 0 0
MA  1 0 0 0 0
MD  1 0 0 0 0
ME  0 0 1 0 0
MI  0 0 1 0 0
MN  0 0 1 0 0
MO  0 0 1 0 0
MS  1 0 0 0 0
MT  1 0 0 0 0
NC  0 1 0 0 0
ND  1 0 0 0 0
NE  1 0 0 0 0
NH  0 0 1 0 0
NJ  1 0 0 0 0
NM  0 0 1 0 0
NV  0 0 1 0 0
NY  1 0 0 0 0
OH  0 0 0 1 0
OK  0 1 0 0 0
OR  0 0 1 0 0
```

```

PA 0 0 1 0 0
RI 1 0 0 0 0
SC 0 1 0 0 0
SD 0 1 0 0 0
TN 1 0 0 0 0
TX 1 0 0 0 0
UT 1 0 0 0 0
VA 1 0 0 0 0
VT 1 0 0 0 0
WA 0 0 1 0 0
WI 0 0 1 0 0
WV 0 0 1 0 0
WY 1 0 0 0 0

> Npop <- sum(to2000)
> Npoph0 <- sum(to2000[bg == 0])
> Npoph1 <- sum(to2000[bg == 1])
> Npoph2 <- sum(to2000[bg == 2])
> Npoph3 <- sum(to2000[bg == 3])
> Npoph4 <- sum(to2000[bg == 4])
> Npoph <- c(Npoph0, Npoph1, Npoph2, Npoph3, Npoph4)
> n <- 3000
> pq <- c(0.65 * 0.35, 0.25, 0.25, 0.25, 0.25)
> nhnps <- c(600, 500, 1100, 400, 400)
> whnps <- nhnps/((Npoph/sum(Npoph)) * n)
> pop3$ninps[bg == 0] <- whnps[1] * to2000[bg == 0]/Npop *
+   n
> pop3$ninps[bg == 1] <- whnps[2] * to2000[bg == 1]/Npop *
+   n
> pop3$ninps[bg == 2] <- whnps[3] * to2000[bg == 2]/Npop *
+   n
> pop3$ninps[bg == 3] <- whnps[4] * to2000[bg == 3]/Npop *
+   n
> pop3$ninps[bg == 4] <- whnps[5] * to2000[bg == 4]/Npop *
+   n
> attach(pop3)
> sum(ninps)

[1] 3000

> pop3$ninps <- round(pop3$ninps + 0.01, digits = 0)

```

```
> attach(pop3)
```

```
> sum(ninps)
```

```
[1] 3001
```

```
> print(cbind(state, ninps, bgsen, bgpres, bg))
```

	state	ninps	bgsen	bgpres	bg
[1,]	"AL"	"29"	"0"	"0"	"0"
[2,]	"AZ"	"57"	"0"	"1"	"2"
[3,]	"AR"	"32"	"0"	"1"	"2"
[4,]	"CA"	"209"	"1"	"0"	"1"
[5,]	"CO"	"56"	"1"	"1"	"2"
[6,]	"CT"	"20"	"0"	"0"	"0"
[7,]	"DE"	"5"	"0"	"0"	"0"
[8,]	"DC"	"4"	"0"	"0"	"0"
[9,]	"FL"	"400"	"1"	"1"	"4"
[10,]	"GA"	"51"	"1"	"0"	"1"
[11,]	"ID"	"8"	"0"	"0"	"0"
[12,]	"IL"	"92"	"1"	"0"	"1"
[13,]	"IN"	"38"	"0"	"0"	"0"
[14,]	"IA"	"47"	"0"	"1"	"2"
[15,]	"KS"	"17"	"0"	"0"	"0"
[16,]	"KY"	"30"	"1"	"0"	"1"
[17,]	"LA"	"70"	"1"	"1"	"2"
[18,]	"ME"	"23"	"0"	"1"	"2"
[19,]	"MD"	"33"	"0"	"0"	"0"
[20,]	"MA"	"42"	"0"	"0"	"0"
[21,]	"MI"	"150"	"0"	"1"	"2"
[22,]	"MN"	"82"	"0"	"1"	"2"
[23,]	"MS"	"18"	"0"	"0"	"0"
[24,]	"MO"	"92"	"1"	"1"	"2"
[25,]	"MT"	"6"	"0"	"0"	"0"
[26,]	"NE"	"11"	"0"	"0"	"0"
[27,]	"NV"	"22"	"1"	"1"	"2"
[28,]	"NH"	"20"	"0"	"1"	"2"
[29,]	"NJ"	"51"	"0"	"0"	"0"
[30,]	"NM"	"22"	"0"	"1"	"2"
[31,]	"NY"	"105"	"0"	"0"	"0"
[32,]	"NC"	"55"	"1"	"0"	"1"

```

[33,] "ND" "5" "0" "0" "0"
[34,] "OH" "400" "0" "1" "3"
[35,] "OK" "26" "1" "0" "1"
[36,] "OR" "53" "0" "1" "2"
[37,] "PA" "172" "1" "1" "2"
[38,] "RI" "7" "0" "0" "0"
[39,] "SC" "31" "1" "0" "1"
[40,] "SD" "6" "1" "0" "1"
[41,] "TN" "33" "0" "0" "0"
[42,] "TX" "105" "0" "0" "0"
[43,] "UT" "12" "0" "0" "0"
[44,] "VT" "4" "0" "0" "0"
[45,] "VA" "44" "0" "0" "0"
[46,] "WA" "87" "1" "1" "2"
[47,] "WV" "25" "0" "1" "2"
[48,] "WI" "91" "1" "1" "2"
[49,] "WY" "3" "0" "0" "0"

```

```

> anhnps <- c(0, 0, 0, 0, 0)
> anhnps[1] <- sum(ninps[bg == 0])
> anhnps[2] <- sum(ninps[bg == 1])
> anhnps[3] <- sum(ninps[bg == 2])
> anhnps[4] <- sum(ninps[bg == 3])
> anhnps[5] <- sum(ninps[bg == 4])
> print(anhnps)

```

```
[1] 600 500 1101 400 400
```

```

> an <- sum(ninps)
> awhnps <- anhnps/(Npoph/sum(Npoph)) * an)
> nallsenatebg <- sum(ninps * bgsen)
> nallpresbg <- sum(ninps * bgpres)
> print(c(nallsenatebg, nallpresbg))

```

```
[1] 1490 1901
```

```

> senatepoph <- c(sum(to2000[bgsen == 1 & state !=
+ "FL"]), to2000[state == "FL"])
> senatenh <- c(sum(ninps[bgsen == 1 & state != "FL"]),
+ ninps[state == "FL"])
> senatepq <- c(0.25, 0.25)

```

```

> senatese <- sqrt(sum(((senatepoph/sum(senatepoph))^2) *
+ ((senatepoph - senatenh)/senatepoph) * senatepq/senatenh))
> senatemoe <- 2 * senatese
> print(c(senatepoph, senatenh, senatepq))

[1] 4.4563e+07 6.0060e+06 1.0900e+03 4.0000e+02 2.5000e-01
[6] 2.5000e-01

> print(c(senatese, senatemoe))

[1] 0.01367199 0.02734399

> prespoph <- c(sum(to2000[bgpres == 1 & state != "FL" &
+ state != "OH"]), to2000[state == "OH"], to2000[state ==
+ "FL"])
> presnh <- c(sum(ninps[bgpres == 1 & state != "FL" &
+ state != "OH"]), ninps[state == "OH"], ninps[state ==
+ "FL"])
> prespq <- c(0.25, 0.25, 0.25)
> presse <- sqrt(sum(((prespoph/sum(prespoph))^2) *
+ ((prespoph - presnh)/prespoph) * prespq/presnh))
> presmoe <- 2 * presse
> print(c(prespoph, presnh, prespq))

[1] 3.1918e+07 4.8230e+06 6.0060e+06 1.1010e+03 4.0000e+02
[6] 4.0000e+02 2.5000e-01 2.5000e-01 2.5000e-01

> print(c(presse, presmoe))

[1] 0.01211949 0.02423898

> nobgpop <- sum(to2000[bg == 0])
> nobgn <- sum(ninps[bg == 0])
> nobgpq <- 0.65 * 0.35
> nobgse <- sqrt(nobgpq/nobgn)
> nobgmoe <- 2 * nobgse
> print(c(nobgpop, nobgn, nobgpq))

[1] 4.0015e+07 6.0000e+02 2.2750e-01

> print(c(nobgse, nobgmoe))

```

```

[1] 0.01947220 0.03894440

> OHse <- sqrt(0.25/anhnps[4])
> OHmoe <- 2 * OHse
> FLse <- sqrt(0.25/anhnps[5])
> FLmoe <- 2 * FLse
> POPpoph <- c(sum(to2000[bg == 0]), sum(to2000[bg ==
+ 1]), sum(to2000[bg == 2]), to2000[state == "OH"],
+ to2000[state == "FL"])
> POPnh <- c(sum(ninps[bg == 0]), sum(ninps[bg == 1]),
+ sum(ninps[bg == 2]), ninps[state == "OH"], ninps[state ==
+ "FL"])
> POPpq <- c(0.65 * 0.35, 0.25, 0.25, 0.25, 0.25)
> POPse <- sqrt(sum(((POPpoph/sum(POPpoph))^2) * ((POPpoph -
+ POPnh)/POPpoph) * POPpq/POPnh))
> POPmoe <- 2 * POPse
> print(c(POPpoph, POPnh, POPpq))

[1] 4.0015e+07 2.7453e+07 3.1918e+07 4.8230e+06 6.0060e+06
[6] 6.0000e+02 5.0000e+02 1.1010e+03 4.0000e+02 4.0000e+02
[11] 2.2750e-01 2.5000e-01 2.5000e-01 2.5000e-01 2.5000e-01

> print(c(POPse, POPmoe))

[1] 0.0101536 0.0203072

```

## 2 The NEW Bottom Line

To summarize the results explained in detail above:

We will take a sample divided into five strata: non-battleground, Senate only battleground, presidential battleground, and state samples of Ohio and Florida.

The sample sizes for each stratum are given in Table 1.

With this sample design, the standard errors and margins of error within each stratum and for the population are given in Table 2.

The sample can be weighted to provide population inferences while maintaining a weighted  $n$  equal to the actual  $n$  of 3001 (within rounding error!). These weights and the weighted  $n$  are given in Table 3.

The allocation of cases to states is given in Table 4.

Stratum	Sample size
NonBattleground	600
Senate only battleground	500
Presidential battleground	1101
Ohio	400
Florida	400
ALL Senate battleground	1490
Total	3001

Table 1: Sample sizes for each stratum, with oversampling of battleground states and state samples for Ohio and Florida. The sample size for ALL Senate battleground respondents includes those that are also in presidential battleground states and the 400 case oversample in Florida.

Stratum	Std. Error	Margin of Error
NonBattleground	0.0195	0.0389
All Senate battleground	0.0137	0.0273
Presidential battleground	0.0121	0.0242
Ohio	0.025	0.05
Florida	0.025	0.05
Population	0.0102	0.0203

Table 2: Standard errors and margin of error for each stratum. Note that the Senate here is for ALL Senate battleground cases, not those in the Senate ONLY battleground states. Estimates include the oversample cases from Ohio and Florida.

Stratum	Weight	Weighted $n$
Nonbattleground	1.816	1090
Senate only battleground	1.495	748
Presidential battleground	0.789	869
Ohio	0.328	131
Florida	0.409	164
ALL Senate battleground		1529

Table 3: Sample weights and weighted sample size for each stratum. The last line is the sample size for all Senate battleground respondents.

	State	n	State	n
1	AL	29	NE	11
2	AZ	57	NV	22
3	AR	32	NH	20
4	CA	209	NJ	51
5	CO	56	NM	22
6	CT	20	NY	105
7	DE	5	NC	55
8	DC	4	ND	5
9	FL	400	OH	400
10	GA	51	OK	26
11	ID	8	OR	53
12	IL	92	PA	172
13	IN	38	RI	7
14	IA	47	SC	31
15	KS	17	SD	6
16	KY	30	TN	33
17	LA	70	TX	105
18	ME	23	UT	12
19	MD	33	VT	4
20	MA	42	VA	44
21	MI	150	WA	87
22	MN	82	WV	25
23	MS	18	WI	91
24	MO	92	WY	3
25	MT	6		

Table 4: State sample sizes.