A

SAMPLING PLAN FOR 1984 GENERAL ELECTION IN UTAH

Sumbitted by:

Alan Austin
Steve Carlson
Janusz Czejdo
Jennifer Goodfellow
David Hatch
Dan Jensen
Kendra Lindsay
John Parker
Janine Perry
Welcome Sauer
Tom Teerlink
Richard Thurman

SAMPLING PLAN FOR 1984 GENERAL ELECTION IN UTAH

INTRODUCTION

A sampling plan is the methodology used for choosing a sample from a population. The following sampling plan was designed to be used in the 1984 general election for the state of Utah. The plan is a ~~3-stage cluster, 2-stage stratified, systematic sample~~. This report will include a general description of the sampling plan, i.e., formulas used to estimate the voter population, how and why we divided our population into the various strata, and the variances associated with these estimates.

ESTIMATION OF VOTER TURNOUT

The first step in selecting a sample of voters in the state of Utah was to estimate the total voter turnout for the upcoming election. Using the data from the 1968, 1972, 1976, 1980 general election years (data obtained from 1983 Utah Statistical Abstracts, Table 6), we did a linear regression in order to predict the estimated voter turnout for the 1984 election. The prediction of estimated turnout for each Utah county is listed below:

	Estimated	Actual
Davis	65,366	61671
Weber	66,838	60776
Tooele	10,326	8853
Beaver	2,253	2219
Juab	2,863	2825
Wayne	1,173	1153
Piute	765	753
Garfield	2,174	1885
Cache	29,338	26456
Iron	8,793	8157
Millard	3,656	5453
Rich	994	923
Kane	2,014	1988
Morgan	2,681	2418
Box Elder	16,089	15161
Washington	13,915	13573
Salt Lake	283,189	266223
Utah	98,163	86746
Carbon	9,954	8849
Summit	5,465	5511
Emery	5,232	4410
Grand	3,556	3341
Wasatch	4,221	3794
Daggett	494	414
San Juan	4,218	3483
San Pete	6,973	6736
Duchesne	5,380	5163
Sevier	7,868	6814
Uintah	7,913	7350
Total	671,785	623098

MSE = ?

These estimated figures were then used in selecting the counties according to the procedure below.

SELECTION OF COUNTIES

The first step in selecting the counties was to divide all the counties in Utah into the three congressional districts, The counties in each district are as follows:

District I		District II	District III	
*Davis	* Weber	Salt Lake	Utah	Wasatch
Tooele	Beaver		Summit	Emery
Juab	Wayne		Grand	San Juan
Piute	Garfield		Daggett	Uintah
Cache	Iron		Sanpete	Carbon
Millard	Rich		Sevier	Duchesne
Kane	Morgan		Part of Salt Lake	
Box Elder	Washington			

The next step was to select certainty counties from each district. Certainty counties were selected on the following three conditions:

1) Large population ~~centers~~ counties: Salt Lake, Davis, Weber, and Utah.
2) Voter behavior different from the state norm: Carbon which had a 47.8% republican vote compared to the state's 72.8% republican vote.
3) Accomodation of student availability: Box Elder and Washington.

With these seven certainty counties selected, the remaining 22 counties (12 in district I and 10 in district III) were stratified according to the percentage of Republican vote in the specific county in the 1980 presidential election. The following strata were created:

District I

Stratum I- Tooele, Beaver, Juab (less than 70% republican vote)
Stratum II- Wayne, Piute (> 70% & < 77% republican vote)
Stratum III- Garfield, Cache (> 77% & < 79% republican vote)
Stratum IV - Iron, Millard (> 79% & < 81% republican vote)
Stratum V- Rich, Kane, Morgan (≥ 81% republican vote).

District III

Stratum I- Summit, Emery, Grand, Wasatch, Daggett (< 70% republican vote)
Stratum II- San Juan, Sanpete, Duchesne, Sevier, Uintah (> 70% republican vote).

One county was selected by PPS sampling from each of the five stratum in district I, and two counties were selected by PPS sampling from each of the two strata in district III. The process of selecting counties by PPS involves the expected turnouts (shown on page 2). The expected turnout for each county was used to obtained a cummlative sum for each stratum. For example, stratum one in district III has five counties and their turnouts are:

County	Expected turnout	Cummulative sum
Summit	5465	5465
Emery	5232	10697
Grand	3556	14253
Wasatch	4220	18473
Daggett	444	18917

Since two counties were to be selected from this strata, the cumulative total was divided by two, resulting in 9458. A random number between 1 and 9458 was then selected. The random number generated was 3016. The cumulative sum that contains 3016 is associated with Summit county, and thus Summit county is the first county selected from this stratum. The second county was selected by adding 9458 to 3016, which adds to 12474 and selects Grand as the other county from this stratum. Thus, Summitt and Grand counties are the two random county selections from stratum I in district III. The same procedure was used in district I and in the other stratum in district III. The only variation in the above procedure for district I is that the cummulative sum was not divided by two, since only one county was being selected per strata. The method outlined above allowed

each county to have a probability of selection that is proportional to the size of the county's expected voter turnout. The selected counties were as follows:

District I	District II	District III
Davis (C) *	Salt Lake (C)	Utah (C)
Weber (C)		Part of Salt Lake (C)
(C)		
Box Elder (C)		Carbon (C)
Washington (C)		Summit
Tooele		Grand
Piute		Sanpete
Cache		Duchesne
Iron		
Kane		

* (C) stands for certainty county.

In all, 16 of the 29 counties in the state were selected. The next step was to decide how many polling places would be sampled in each of the selected counties. This decision was also made considering the proportional size of the expected voter turnout. It was decided that a total of 120 polling places were to be sampled. This decision was made based upon manpower availibility. The total expected voter turnout for the state of Utah was 671,785 and this figure was divided by the 120 polling places resulting in approximately 5600. Using this figure, each stratum (note: certainty counties were considered a strata) was alloted one sample precinct for every 5600 expected voters in the stratum. The breakdown of each county and their number of sample precincts are as follows: -

District I	
Davis	11 precincts
Weber	12 precincts
Box Elder	3 precincts
Washington	3 precincts
Tooele	3 precincts
Piute	1 precinct
Cache	6 precincts
Iron	2 precincts
Total	42 precincts

District II	
Salt Lake	40 precincts
Total	40 precincts

District III	
Utah	17 precincts
Salt Lake	9 precincts
Carbon	2 precincts
Summit	2 precincts
Grand	2 precincts
Sanpete	3 precincts
Duchesne	3 precincts
Total	38 precincts

SELECTION OF POLLING PLACES AND VOTERS TO BE SAMPLED

After selecting the counties that were to be included in the sample, the next step was to determine precisely which polling places and voters would be included in the sample. (In case of any unforeseen difficulties arising on election day, convenient alternate polling places were chosen and sampling intervals determined).

In order to draw a two-stage cluster sample of specific voters, it was first necessary to determine the expected voter turnout for each polling place (note: polling places are clusters, and voters are listing units and elementary units).

The expected voter turnout was estimated by multiplying each polling place's proportion of a previous year's county vote by the 1984 estimated voter turnout for that county (see Appendix A).

Since some counties had added polling places or precincts since the 1980 general election, it was necessary to use either the 1982 election results or the 1984 primary election results in order to estimate the expected voter turnout. Specifically, these counties were: Salt Lake, Weber, Washington, Kane, and Davis counties. Utah county was treated in a similar manner (see Appendix B for details). Once the expected voter turnout for each polling place in the sample was estimated, it became possible to determine which polling places would be sampled. The procedure used for selecting a polling place was a systematic PPS sampling technique, without replacement, as was demonstrated on page 5.

After the sample of polling places was determined, the final step in designing the sampling plan, that is, selecting which voters will be sampled, was possible. The first step in choosing which voters to include in the sample was determining the number of voters to be sampled from each county and each polling place. To determine the number of voters to be sampled in each county, the total expected statewide voter turnout (671,785) was divided into the desired sample size of 2100. The resulting fraction (1/320) represents the fraction of voters who would be sampled statewide. This fraction was then multiplied by the number of expected voters in each sample county and divided by the number of sample polling places from the county. The result was the number of voters to be sampled from each polling place.

The second step in choosing which voters to sample from each polling place was to design a systematic sample strategy for

each polling place by dividing the expected turnout in each polling place by the number of voters to be sampled from that polling place. This division resulted in a constant (K) which represents the sampling interval for that polling place. Then, a random number between one and K was selected for each polling place to decide the first voter to be sampled from each polling place. Every K^{th} voter leaving the polling place thereafter was also sampled.

The following table lists which polling places in each sample county were selected. The estimated turnouts, sampling intervals, and number of voters to be sampled from each polling place are also shown.

The reason for doing things as described above is to get a self-weighting sample — You have not observed that this is a reason for doing what you have done and you should have!

You have also failed to mention the problem of multiple precinct voting places and the problems it created.

	ESTIMATED TURNOUT FOR POLLING PLACE	SAMPLING RATIO FOR POLLING PLACE	EXPECTED NUMBER OF VOTERS TO BE SAMPLED AT POLLING PLACE
LEGISLATIVE DISTRICT I			
DAVIS COUNTY (11)			
Bountiful #2,8,12	1188	1/66	18
Bountiful #23,25,27,40	1710	1/95	18
Bountiful #6,36	1242	1/69	18
Centerville #6,7,8	1566	1/87	18
Keysville #1,5,10	1386	1/77	18
Clinton #1,3,4,6	1512	1/84	18
No. Salt Lake #1,2,3 DC #1	2214	1/123	18
Woodscross #1,3	810	1/45	18
Layton #16,19,20	756	1/42	18
Layton #1,6,7,23	1386	1/77	18
Clearfield #4,6	756	1/42	18
WEBER COUNTY (12)			
Unita #1	263	1/15	17
Roy #4	226	1/13	17
Roy #27	321	1/18	17
Washington Terrace #3	151	1/8	17
Ogden #90	361	1/21	17
Riverdale #8	268	1/15	17
Ogden #3	224	1/13	17
Odgen #65	310	1/18	17
No. Odgen #5	1386	1/81	17
Odgen #26	1302	1/76	17
Plesant View #3	215	1/12	17
Odgen #24	240	1/14	17
TOOELE COUNTY (3)			
Tooele #7	261	1/16	16
Tooele #18	511	1/32	16
Stockton	265	1/17	16
PIUTE COUNTY (1)			
Marysvale	262	1/43	6
IRON COUNTY (2)			
Cedar City #4	749	1/39	19
Cedar City #16	495	1/39	19

	ESTIMATED TURNOUT FOR POLLING PLACE	SAMPLING RATIO FOR POLLING PLACE	EXPECTED NUMBER OF VOTERS TO BE SAMPLED AT POLLING PLACE
LEGISLATIVE DISTRICT I			
CACHE COUNTY (6)			
Logan #13,21	1410	1/88	16
Logan #22	531	1/33	16
Newton	381	1/17	16
Providence #2	668	1/41	16
Richmond #2	505	1/31	16
Smithfield #2	745	1/46	16
KANE COUNTY (1)			
Orderville	258	1/14	18
BOX ELDER COUNTY (3)			
Brigham #1	339	1/21	16
Mantua	273	1/17	16
Brigham #14	336	1/21	16
WASHINGTON COUNTY (3)			
St. George #6	652	1/46	14
Bloomington 1st	399	1/28	14
Springdale	238	1/17	14
LEGISLATIVE DISTRICT II			
SALT LAKE COUNTY (40)			
SLC #4138,4140,4142	1469	1/81	18
SLC #4164	289	1/16	18
SLC #4204	599	1/33	18
SLC #4234	294	1/16	18
SLC #4258,4266	1056	1/58	18
SLC #4294	726	1/40	18
SLC #4312,4314,4342	1394	1/77	18
SLC #4358,4344	1213	1/67	18
SLC #4386,4378,4380,4382	1869	1/103	18
SLC #4496	378	1/21	18
SLC #4446	443	1/24	18
SLC #4486	737	1/40	18
SLC #4604,4612	1956	1/58	18
SLC #1108	462	1/25	18
SLC #1202,1234,1222	1215	1/67	18

	ESTIMATED TURNOUT FOR POLLING PLACE	SAMPLING RATIO FOR POLLING PLACE	EXPECTED NUMBER OF VOTERS TO BE SAMPLED POLLING PLACE
SALT LAKE COUNTY (cont.)			
SLC #1217,1218	332	1/18	18
SLC #1316	746	1/42	18
SLC #2018	310	1/17	18
SLC #2056	513	1/28	18
SLC #2136	502	1/27	18
SLC #2178	275	1/15	18
SLC #2228	394	1/21	18
SLC #2320	278	1/15	18
SLC #2370,2402,4104,4106	1302	1/72	18
SLC #2438	429	1/23	18
SLC #2482	427	1/23	18
SLC #2582,2536	1284	1/71	18
SLC #2546	1742	1/96	18
SLC #2584	524	1/29	18
SLC #2612	237	1/31	18
SLC #2642,2652	867	1/48	18
SLC #2674	559	1/31	18
SLC #2722	326	1/18	18
SLC #2842	172	1/9	18
SLC #2870,2872	1018	1/56	18
SLC #2990, 2974	1123	1/62	18
SLC #2968,2970	1037	1/57	18
SLC #4652,4732,4662	1248	1/69	18
SLC #4663,4664,4668	1042	1/57	18
SLC #4790	175	1/9	18
LEGISLATIVE DISTRICT III			
SALT LAKE COUNTY (9)			
SLC #3134,3136	1161	1/64	18
SLC #3214,3210	996	1/55	18
SLC #3226	567	1/31	18
SLC #3250	362	1/20	18
SLC #3406,3412	880	1/48	18
SLC #3502,3510	659	1/36	18
SLC #3626	289	1/16	18
SLC #1350	499	1/27	18
SLC #1432,1434	1375	1/76	18
CARBON COUNTY (2)			
East Helper	426	1/28	15
No. Dragerton	521	1/35	15

	ESTIMATED TURNOUT FOR POLLING PLACE	SAMPLING RATIO FOR POLLING PLACE	EXPECTED NUMBER OF VOTERS TO BE SAMPLED AT POLLING PLACE
UTAH COUNTY (17)			
American Fork#5	390	1/21	18
Lehi #12	355	1/19	18
Highland #1	411	1/22	18
Orem #2,49,53,61	1668	1/92	18
Orem #13	559	1/31	18
Orem #20	923	1/51	18
Orem #31,32	600	1/33	18
Payson # 2,9	836	1/46	18
Pleasant Grove #3,13	774	1/43	18
Provo #8,9,11	1601	1/88	18
Provo #18,19	842	1/46	18
Provo #33, 71	697	1/37	18
Provo #46,48,75	1414	1/78	18
Provo #62	492	1/27	18
Salem #1,2,4,5	1698	1/94	18
Spanish Fork #9,14,17	856	1/47	18
Springville #9,12,16	1208	1/67	18
SUMMIT COUNTY (2)			
Park City #3,4,5,6	654	1/44	14
#18 Hoytsville	453	1.30	14
GRAND COUNTY (2)			
Moab #3	494	1/35	14
Moab #10	292	1/21	14
SAN PETE COUNTY (3)			
Mt. Pleasant #1	346	1/20	16
Ephraim #1,2,3	1215	1/71	16
Sterling	123	1/7	16
DUCHESNE COUNTY (3)			
Duchesne #2	420	1/24	16
Altamount	629	1/37	16
Roosevelt #1	533	1/31	16

VARIANCES

The formulas used to compute the variances are given below.

1) Estimators (general Notation):

Let x_{ijkl} = the response of person l, in precinct k, in county j, in stratum i.

i = stratum index (1, 2, 3, ...L)

ii = certainty county

j = county index (1, 2, = c_i)

k = precinct index (1,....p_{ij})

l = person index (1,....n'_{ijk})

$N_{i..}$ = actual number of voters in stratum i

$N_{...}$ = actual number of voters in the state, $\sum_i N_{i..}$

C_i = number of counties selected from stratum i

P_{ij} = number of precincts selected from stratum i, county j.

$\hat{p}$ = estimated proportion of voters who favor a certain candidate or proposal

n_{ijk} = number of voters po led in precinct K, in county j, in stratum i.

2) Variance formulas for districts I and III:

$$V(p) = \frac{(320)^2}{(N_{...})^2} \left[\sum C_i \, S^2 d_i + \sum p_{ii} \, S^2 d_{ii} \right]$$

where $S^2 d_i = \sum \frac{(d_{ij} - \bar{d}_i)^2}{(C_i - 1)}$

$$S^2 d_{ii} = \sum \frac{(d_{iik} - \bar{d}_{ii})^2}{p_{ii}(p_{ii} - 1)}$$

$$d_{iik} = (x_{iik} - \hat{p}\, n_{iik})$$

$$d_{ij} = (x_{ij..} - \hat{p}\, n_{ij..})$$

$$\hat{p} = \frac{\hat{x}}{\hat{N}} = \frac{\sum \hat{x}_{ic} + \sum \hat{x}_{inc}}{\sum N_{ic} + \sum N_{inc}}$$

where c = certainty, nc = non-certainity

$\hat{p}$ = est. proportion for pop^{n} or sub population as case may be.

$$\text{also } \hat{p} = \frac{\sum_i \sum_k x_{iik}}{\sum_i \sum_k n_{iik}} + \frac{\sum_i \sum_j x_{ij..}}{\sum_i \sum_j n_{ij..}}$$

3) Variance Formulas for District II:

$$\hat{V}(\hat{p}_2) = \frac{1}{p_{22}\,(p_{22}-1)\,N_{2..}^2} \sum \frac{(x_{22K} - \hat{p}_2\, n_{22K})^2}{\frac{n'_{2}}{N'_{2..}}2k}$$

$$= \frac{(320\ p_{22})^2}{p_{22}\,(p_{22}-1)N_{2..}^2} \sum (x_{22K} - \hat{p}_2\, n_{22K})^2$$

where $\hat{p}_2 = \frac{\hat{x}_2}{\hat{N}_2} = \frac{\sum x_{22K}}{\sum n_{22K}}$ (a ratio estimate)

$$\hat{N}_2 = \frac{1}{p_{22}\,\frac{n'}{N'}22K_{2..}}$$

(an estimate of total voter turnour based on results observed in each of the K precincts sampled in District II)

n_{22K} = actual sample size in precinct K.

4) Actual Variances:

RACE	ESTIMATE	S. D.	CV	95% CI
Mondale	0,2362	0,00997	.04	0,21067, 0,24973
Reagan	0,7566	0,03328	.04	0,71617, 0,79703
Bangerter	0,5751	0,01284	.02	0,54994, 0,60026
Owens	0,4214	0,01288		0,39616, 0,44664
Initiative A (For)	0,3637	0,1270		0,33881, 0,02489
Initiative A (Against)	0,5821	0,01222		0,55816, 0,60604

Congressional Races

District I	Estimate	S.D.	CV	95% C.I
Abrams	0.283308	0.01761	.06	0.25480, 0.32382
Hansen	0.610063	0.01817		0.57445, 0.64568

* DISTRICT II

Farley	0.48276	0.02145	0.44072, 0.52480
Monson	0.51724	0.2324	0.47169, 0.56279

* Due to non response, these estimates were calculated somewhat differently than those of the other two districts. How so?

DISTRICT III

Baird	0.218075	0.03095	0.18713, 0.24902
Neilson	0.679764	0.03770	0.64206, 0.71747

Design Effect $\dfrac{\hat{V}ar(\hat{p}_{SRS})}{\hat{V}ar(\hat{p})}$

Reagan: $\dfrac{0.0001161}{0.00110745} = 1.04835$

Bangerter: $\dfrac{0.0001536}{0.00164814} = 0.93196$

No major differences occur between the variances of the simple random sample and the variances of the multi-stage cluster sample.

SUMMARY AND CONCLUSION

The object of this report was to provide a general description of the KBYU Exit Poll plan, along with the formula used to estimate voter turnout, and variances associated with voter behavior.

A description of the rationale and the procedures involved in creating a 3-stage cluster, 2-stage stratified, systematic sample were given. During this description, an estimation of predicted voter turnout for the state of Utah and 120 selected precincts was presented. Finally the techniques used to derive variance estimators were shown and the estimators were given.

As can be seen from table 1, this sampling plan provided a more acurate prediction of voting practices in the state of Utah than did other statewide exit polls. In all races but one, this poll predicted voter behavior better than KSL or KUTV.

In addition, it should be noted that actual voter behavior was well within our margin of error, lending support to the adequacy of the sampling plan.

TABLE 1

PERCENTAGES OF VOTERS IN UTAH

Candiates	KBYU	KSL	KUTV	ACTUAL VOTE
Regan/Mondale	76/23	78/22		75/25
Bangerter/Owens	57/43	58/42		56/44
Abrams/Hansen	32/67	23/77		23/72
Monsen/Farley	50.4/49.6*	52/48	43/55	50/50
Nielson/Baird	75/25	77/23		75/25
Initiative A	38/60	41/59		39/61

*These figures were calculted on the first, incomplete data set, and the updated figures are 51.7/48.3. It should be noted that the second figures were used to calculate the variances.

What about the non-response rate? We talked about getting this several times.

Appexdix A: Expected voter turnout per polling place

$$N'_{ijk} = \frac{N_{ijk}\ (1980,1982,or1984)}{N_{ij.}} \times N'_{ij.}$$

where: N'_{ijj} = expected turnout per polling place

N_{ijk} = previous year's actual turnout per polling place

$N_{ij.}$ = previous years actual turnout per county

$N'_{ij.}$ = expected turnout per county

n'_{ijk} = expected sample size per polling place

$n'_{...}$ = expected statewide sample size (2100)

$N'...$ = expected statewide voter turnout (681,785)

m_{ijk} = number of clusters (polling places to be sampled from each county (determined in previous section).

Appendix B: Because of the great number of deleted and newly created precincts in Utah County, we modified the usual method of projecting the 1984 voter turnout for each precinct. Since much of the growth between 1980 and 1984 went into the creation of new districts, we had to decrease the growth factor in this manner:

$$\frac{\text{1984 projected county votes}}{\text{1980 county votes}}$$

became

$$\frac{\text{1984 projected county votes}}{\left(\text{1980 county votes} + \text{number of new precincts} \times \frac{\text{1980 county votes}}{\text{\# of precincts 1980}}\right)}$$

For the unchanged precincts, we multipled the modified growth factor by each precinct's 1980 turnout. For the new districts, we multipled the modified growth by the following expression, (identical for each district): $\frac{\text{1980 county votes}}{\text{\# of precincts in1980.}}$

We found information telling us which remaining precincts the deleted precincts had been absorbed into, which allowed us to divide the 1980 votes from each deleted precinct by the number of precincts which absorbed it, then add that figure to the 1980 votes of each of the precincts which absorbed that deleted precinct.