




# Probability Sampling: An Application

Yuhang Liu  
Amy Nussbaum



# Introduction

- General Probability Sampling
  - Definition of a Probability Sample
  - Cost Function and Constraints
  - Sample Design
  - Objective Function
  - Tools
- Application
  - Background Information
  - Development of Cost Function, Constraints
  - Development of Sample Design
  - Development of Objective Function
  - Implementation
- Morals of the Story





# Probability Sample

- Definition: A sample selected in such a way that every member of the sample has a known chance of selection (and every member of the population has a nonzero chance of selection)
- This definition does not require EQUAL probability of selection, as long as we use some sort of randomness to make our decisions.
- Sampling Units
  - Primary Sampling Unit (PSU)
  - Secondary Sampling Unit (SSU)

# Probability Sample

- Advantages
  - Reduces risk of selection bias (allows for unbiased estimators of population parameters)
  - Produces results that are generalizable to the U.S. population (allow for known margins of error from population parameters)





# Cost Function and Constraints

- Cost Function Formula for a Balanced Survey:

$$C_{total} = C_1n + C_2mn$$

$n$  is the total number of PSUs

$m$  is the total number of SSUs

- When the survey is unbalanced, replace  $m$  with  $\bar{m}$
- Constraints can be
  - Financial
  - Time
  - Resources



# Sample Design

- Depends on what information you have and what information you want...
  - Simple Random Sample
  - Cluster Sample
- Requires the use of auxiliary variable:
  - Stratified Sample
  - Probability Proportional to Size Sampling





# Objective Function

- Helps us choose optimal sample size with two goals:
  - Minimize the cost with respect to a certain variance
  - Minimize the variance with respect to a certain cost
  - Both reduce to the same answer!
- Either way, we need to know how to calculate the variance of our statistic of interest.
  - Total
  - Mean
  - Proportion
  - Regression

# Tools for Design

- Excel
  - Solver
    - Allows for customizable objective function
    - Allows for additional constraints
    - Extremely useful for choosing sample sizes
    - Can solve problems numerically
  - “Using Spreadsheet Solvers in Sample Design”
    - (Stokes & Plummer, 2004).



# Tools for Analysis

- SAS
  - Selection: SURVEYSELECT
  - Analysis of Means and Totals: SURVEYMEANS
  - Analysis of Regression: SURVEYREG
  - Analysis of Generalized Regression: SURVEYLOGISTIC
  - Analysis of Frequency Tables: SURVEYFREQ
  - Easily allows use of sampling unit weights
- Do not ignore weights when analyzing data!

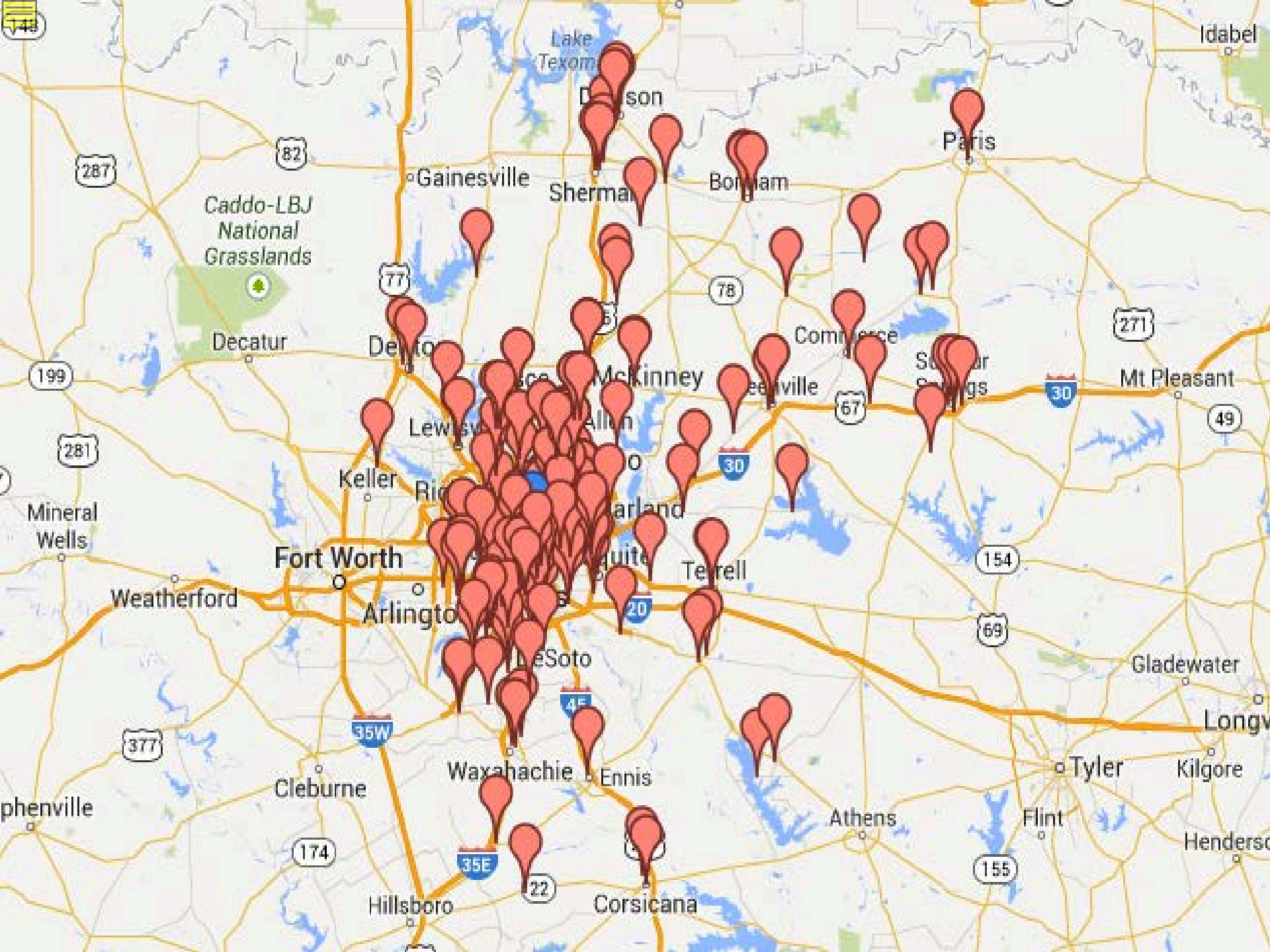
# Application

- Background Information:
  - Predict food-insecurity in the Greater North Texas Area
  - The grant, from the Hunger Center of Dallas, Subsidiary of North Texas Food Bank, focuses on
    - Financial Literacy
    - Social Network
    - Educational Level
  - Reaches beyond the unidirectional definition of poverty
- Sampling Units
  - 172 Food Banks (PSU)
  - Approximately 1000 Clients (SSU)



# Application

- Additional Information on Food Banks
  - ID Number
  - Name
  - Number of clients served in the last three months
  - Mileage
  - Housing in America Survey Info
  - County
  - City
  - Zip Code
  - IBSCO Map
  - Unemployment Rate by County – Proxy Variable





# Development of Cost Function

- Constraints:
  - Financial
  - Cost per survey: \$50 per respondent
    - \$20 incentive, given to the respondent
    - \$15 paycheck, given to the interviewer
    - \$10 paycheck, given for data entry
    - \$5 reimbursement for gas and travel time
  - For  $N=1000$ , total cost is \$50,000

# Development of Cost Function

- Constraints
  - Time
    - Time per survey: 40 minutes
    - 35 minutes for survey
    - 5 minutes for travel time
  - For  $N=1000$ , total time is 660 hours
- Interviewers
  - 4 surveyors, contributing 165 hours each
  - If they spend 10 hours per week interviewing, the survey will take 16 weeks to complete.
  - Plan: conduct surveys November – February (4 months total)



# Development of Cost Function

- Other practical issues:
  - Interviewers could work for a maximum in one day.
  - Interviews could potentially take longer than 30 minutes.
  - Interviewers were sent in pairs for safety reasons.

# Development of Cost Function

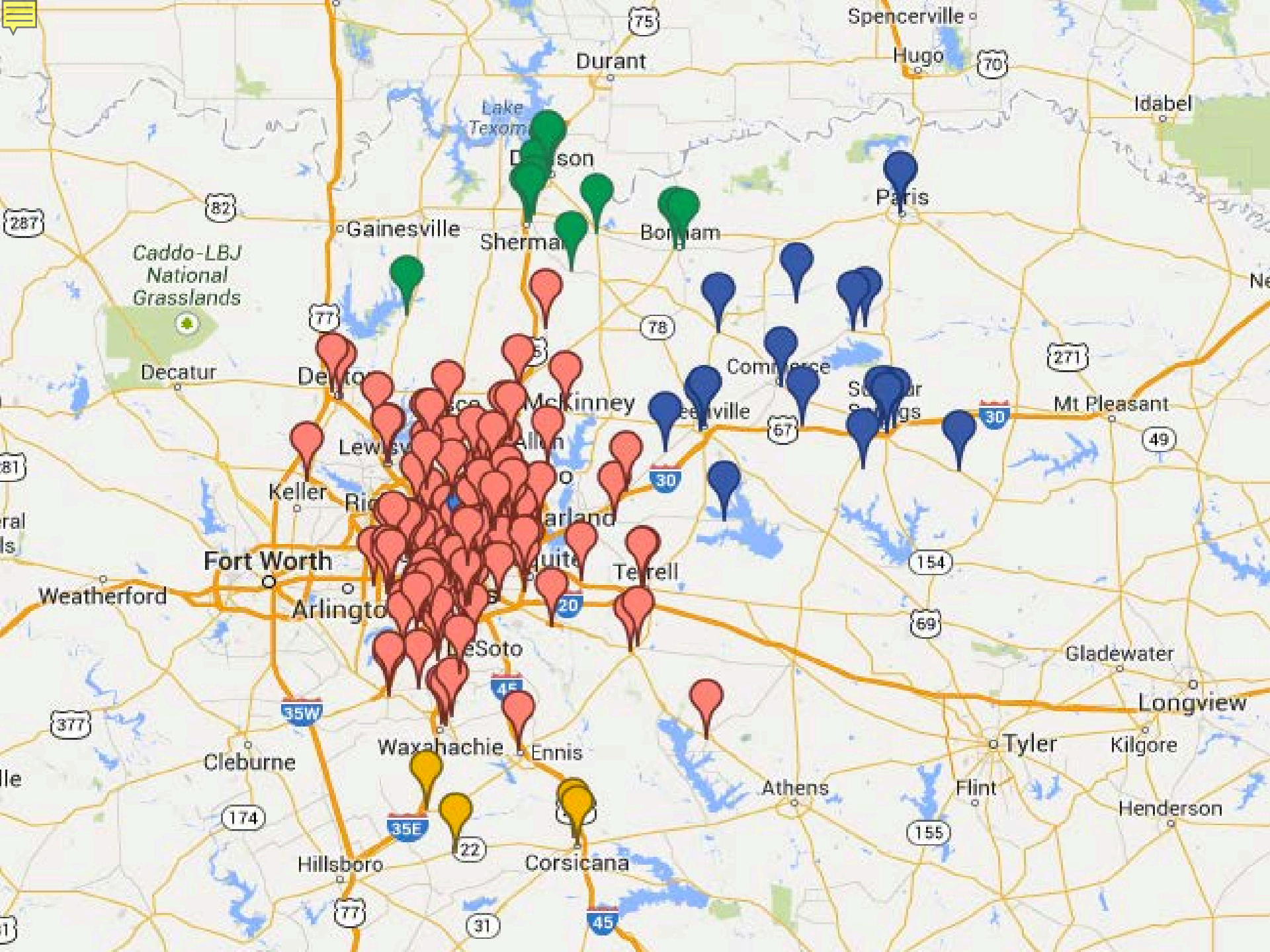
- Assume that interviewers could conduct a constant number of interviews per day, and that the financial cost was constant up to travel reimbursement.
- Travel reimbursement is proportional to the miles the interviewers drive.

$$\text{Cost} = \text{Miles} \times (\text{Reimbursement} + \text{Hourly Wage} / \text{Average Speed})$$



# Development of Sample Design

- Stratification
  - Ensures representation of both urban and rural areas
  - No objective definition of urban and rural
  - Instead, we ensured that food banks from the northern, southern, and eastern regions would be represented.
- Four Geographical Strata
  - Metropolitan Region
  - North Region
  - South Region
  - East Region





# Development of Sample Design

- PPS Sampling
  - Assigned probabilities to each bank proportional to the number of clients served
  - Larger banks have a higher probability of selection
  - Sampling the same number of clients from each bank ensures approximately equal probability of selection for clients (makes analysis easier)

# Development of Objective Function

- Goal: Predict food insecurity using decision trees
  - Closest Statistic: Proportion
- Objective: Minimize variance with respect to cost
- Variance:

$$Var_{total} = \sum [Var_{within} + Var_{between}]$$



# Objective Function

$$Var_{between} = N(1 - \frac{n}{N}) \left( \frac{\sum m^2 (\pi(2\alpha - 1) - E(\alpha))^2}{n - 1} \right)$$

$$Var_{within} = \frac{N}{n} \sum m^2 \left( 1 - \frac{\min(m, N/n)}{m} \right) V(\alpha)$$

where  $\alpha$  represents the proxy variable, unemployment rate,  
 $\pi$  represents the sampling weight based on the number of clients per bank,  
and  $m$  represents the number of clients served at the bank.

# Excel Solver

foodbankssample design - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	clients	mileage	hia	county	city	zip	map	name	latitude	longitude	M^2	strata
2	026797P	324	37.1	YES	COLLIN	PRINCETON	75407	155	CHRIST CARES FOR PRINCETON	33.14572	-96.49664	104976	
3	026687P	549	21.6	YES	COLLIN	ALLEN	75002	1	ALLEN COMMUNITY OUTREACH	33.092846	-96.62447	301401	
4	026357P	571	28.5	YES	COLLIN	WYLIE	75098	50	AMAZING GRACE FELLOWSHIP	33.011975	-96.53607	326041	
5	026568P	1952	14.6	YES	COLLIN	PLANO	75093	49	SEVEN LOAVES FOOD PANTRY	33.03505	-96.80492	3810304	
6	026637P	2079	15.6	YES	COLLIN	PLANO	75074	35	CHRISTIAN BENEVOLENT OUTREACH	33.028921	-96.68102	4322241	
7	026769P	2553	14.8	YES	COLLIN	PLANO	75074	34	A CHRISTIAN FOOD PANTRY	33.028921	-96.68102	6517809	
8	026431P2	204	14.2		COLLIN	PLANO	75075	40	SVPD ST MARKS CONFERENCE	33.024721	-96.74038	41616	
9	026077P	250	20.7		COLLIN	ALLEN	75002	2	FBC ALLEN FOOD PANTRY	33.092846	-96.62447	62500	
10	008146P	396	12		COLLIN	RICHARDSON	75081	42	RICHARDSON EAST COC	32.948974	-96.70916	156816	
11	026501P	461	15.8		COLLIN	PLANO	75074	36	PLANO FOOD PANTRY	33.028921	-96.68102	212521	
12	002499P	736	16.4		COLLIN	PLANO	75074	37	SALVATION ARMY PLANO CORPS	33.028921	-96.68102	541696	
13	002588P	792	24.1		COLLIN	FRISCO	75034	10	FRISCO FAMILY SERVICES CENTER	33.143792	-96.83938	627264	
14	002133P	1716	9.1		COLLIN	RICHARDSON	75081	41	NETWORK OF COMMUNITY MINISTRY	32.948974	-96.70916	2944656	
15	026550P	40	3.7	YES	DALLAS	DALLAS	75209	97	GREATER ZION OUTREACH MIN.	32.845978	-96.82552	1600	
16	002141P	130	13.3	YES	DALLAS	GARLAND	75044	16	EXTENDED FAITH FOOD PANTRY	32.960375	-96.66188	16900	
17	002125P	150	9.2	YES	DALLAS	DALLAS	75223	123	METHODISM BREADBASKET 1	32.792879	-96.74918	22500	
18	026431P	280	6.1	YES	DALLAS	DALLAS	75238	144	ST PATRICKS SVPD	32.873926	-96.70922	78400	
19	026031P	424	14.6	YES	DALLAS	DALLAS	75211	103	MOUNTAIN VIEW COC	32.736931	-96.88253	179776	
20	026803P	567	19.8	YES	DALLAS	DESOTO	75115	56	CHOCOLATE MINT FOUNDATION	32.599286	-96.85882	321489	
21	026787P	603	21.3	YES	DALLAS	GRAND PRAIRIE	75051	19	GRAND PRAIRIE FAMILY CHURCH	32.728982	-97.00428	363609	
22	026575P	607	16.3	YES	DALLAS	DALLAS	75237	142	HARMONY COMMUNITY DEVELOPMENT CORP	32.666984	-96.87633	368449	
23	002166P	721	19.9	YES	DALLAS	GRAND PRAIRIE	75050	17	GRAND PRAIRIE CO-OP	32.759922	-97.01216	519841	
24	026738P	788	8.9	YES	DALLAS	DALLAS	75212	106	UNITED UNIVERSE	32.78238	-96.86945	620944	
25	026065P2	836	5.3	YES	DALLAS	DALLAS	75246	148	CITY SQUARE	32.791878	-96.77365	698896	
26	026202P	1024	16.4	YES	DALLAS	DALLAS	75232	134	TURN AROUND AGENDA	32.664034	-96.83886	1048576	
27	002127P	1057	16.8	YES	DALLAS	FARMERS BRANCH	75234	137	METROCREST SOCIAL SERVICES	32.925975	-96.88322	1117249	
28	002103P	1063	14.4	YES	DALLAS	IRVING	75060	26	IRVING CARES, INC	32.80268	-96.95499	1129969	
29	026604P	1215	12.2	YES	DALLAS	DESOTO	75115	57	DESOTO FOOD PANTRY	32.599286	-96.85882	1476225	
30	026795P	1292	20.4	YES	DALLAS	LANCASTER	75134	64	CC BOYCE MINISTRIES	32.620385	-96.7812	1669264	
31	026538P	1293	8.9	YES	DALLAS	DALLAS	75210	99	COMMUNITY OUTREACH BAPTIST	32.77103	-96.74732	1671849	
32	002086P1	1341	20	YES	DALLAS	GRAND PRAIRIE	75051	20	GRAND PRAIRIE UNITED CHARITIES	32.728982	-97.00428	1798281	
33	026695P	1382	14	YES	DALLAS	DALLAS	75217	117	PLEASANT GROVE FOOD PANTRY	32.72238	-96.67582	1909924	
34	026095P	1478	17.9	YES	DALLAS	MESQUITE	75149	74	MESQUITE SOCIAL SERVICES	32.767329	-96.60759	2184484	
35	002187P2	1594	12.7	YES	DALLAS	DALLAS	75228	127	WILKINSON CENTER	32.825227	-96.67955	2540836	
36	002487P	1601	16.4	YES	DALLAS	DALLAS	75232	131	COMMUNITY OUTREACH CONNECTION	32.664034	-96.83886	2563201	
37	026285P	1703	24.2	YES	DALLAS	SEAGOVILLE	75159	79	COMMUNITY BREADBASKET	32.628858	-96.53811	2900209	
38	003405P	2264	8.4	YES	DALLAS	DALLAS	75203	88	GOLDEN GATE BAPTIST CHURCH	32.745831	-96.80672	5125696	
39	002177P	2829	17.2	YES	DALLAS	MESQUITE	75150	75	SHARING LIFE COMM OUTREACH	32.818392	-96.63355	8003241	
40	026099P2	2980	11.4	YES	DALLAS	MESQUITE	75149	72	BUCKNER CRISIS RELIEF CENTER	32.767329	-96.60759	8880400	

Ready | foodbanks | Cost | variance | Variance2 | sampling sheet | sample of foodbanks | 100%



# Excel Solver

- Solver is an available add-in to MS Excel that handles sampling design
  - Objective Variable
  - Changing Variables
  - Constraints
- Obtains an optimal solution when it exists

# Excel Solver

foodbankssample design - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View SAS

Clipboard Font Alignment Number Styles Cells Editing

	K	L	M	N	O	P	Q	R	S	T
151	-95.55958	33582025	4	0.197		0	-0.197	0.158191	2720023.501	158454.5067
152	-95.55958	3025	4	0.197		0	-0.197	0.158191	245.0141434	5.65532825
153	-95.55958	1923769	4	0.197		0	-0.197	0.158191	155818.3847	8910.277339
154						N=	152	n=		30
155						dbar_U=	0.1	m=		33.33333333
156	ave.mileage	ni	miles		variance bt(before strata)	V1=	44,437,453.60	Sr^2=		71889.75453
157	17.87384615	22.30263158	2391.803		variance w(before strata)	V2=	7,877,640.23	sum of within variance comp		1554797.413
158	56.28333333	0.986842105	333.2566			var=	52,315,093.83			
159	62.12857143	2.368421053	882.8797							
160	79.85909091	4.342105263	2080.539							
161		Total miles:	5688.479							
162						N1=	113.00	n1=		22.30263158
163						N2=	5.00	n2=		0.986842105
164		Total Travel Cost:	5110.15			N3=	12.00	n3=		2.368421053
165		Total Travel Time:	189.616 hours			N4=	22.00	n4=		4.342105263
166										
167										
168										
169					between	v11=	5,752,005.70	Sr^2=		12517.05868
170					within	v12=	6,001,863.76	dbar_U1=		0.176212389
171						v1	11,753,869.46			
172										
173					between	v21=	3,027.11	Sr^2=		148.8741365
174					within	v22=	227,179.61	dbar_U2=		-0.1846
175						v2=	230,206.72			
176										
177					between	v31=	285,388.48	Sr^2=		5848.124531
178					within	v32=	640,221.63	dbar_U3=		-0.14725
179						v3=	925,610.11			
180										
181					between	v41=	57,553.27	Sr^2=		643.292902
182					within	v42=	1,008,375.22	dbar_U4=		-0.177590909
183						v4=	1,065,928.49			
184										
185						variance of stra	13,975,614.78			
186						Note: This variance is computed using alpha's as urban/rural, i.e. 0.1 case, it is almost 1/3 of the variance for the unst				
187						In the fourth sheet ("variance2"), the alpha's are unemployment rate, variance is different				
188						I allocated the sample sizes (n1---n4 here) proportionally according to the strata size. ---Bingche				
189										

foodbanks Cost variance Variance2 sampling sheet sample of foodbanks

Ready Average: 22.75 Count: 16 Sum: 182 100%



# Excel Solver - Variance

foodbankssample design - Microsoft Excel

FileHomeInsertPage LayoutFormulasDataReviewViewSAS

Cut

Copy

Paste

Format Painter

Clipboard

Calibri

11

B

I

U

Font

Alignment

Wrap Text

Merge & Center

Number

General

\$

%

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

AutoSum

Fill

Clear

Sort & Filter

Find & Select

Editing

O156variance bt(before strata)

	K	L	M	N	O	P	Q	R	S	T
148	-95.90886	431649	4	0.181	0	-0.181	0.148239	31141.42638	1822.22346	
149	-96.11289	1723969	4	0.181	0	-0.181	0.148239	124376.1799	7472.145411	
150	-96.11289	4048144	4	0.181	0	-0.181	0.148239	292054.3737	17704.52768	
151	-95.55958	33582025	4	0.197	0	-0.197	0.158191	2720023.501	158454.5067	
152	-95.55958	3025	4	0.197	0	-0.197	0.158191	245.0141434	5.65532825	
153	-95.55958	1923769	4	0.197	0	-0.197	0.158191	155818.3847	8910.277339	
154					N=		152	n=	30	
155					dbar U=		0.1	m=	33.33333333	
156	ave.mileage	ni	miles		variance bt(before strata)	V1=	44,437,453.60	Sr^2=	71889.75453	
157	17.87384615	22.30263158	2391.803		variance w(before strata)	V2=	7,877,640.23	sum of within variance comp	1554797.413	
158	56.28333333	0.986842105	333.2566			var=	52,315,093.83			
159	62.12857143	2.368421053	882.8797							
160	79.85909091	4.342105263	2080.539							
161		Total miles:	5688.479							
162					N1=		113.00	n1=	22.30263158	
163					N2=		5.00	n2=	0.986842105	
164		Total Travel Cost:	5110.15		N3=		12.00	n3=	2.368421053	
165		Total Travel Time:	189.616	hours	N4=		22.00	n4=	4.342105263	
166										
167										
168					between	v11=	5,752,005.70	Sr^2=	12517.05868	
169					within	v12=	6,001,863.76	dbar_U1=	0.176212389	
170						v1	11,753,869.46			
171										
172					between	v21=	3,027.11	Sr^2=	148.8741365	
173					within	v22=	227,179.61	dbar_U2=	-0.1846	
174						v2=	230,206.72			
175										
176					between	v31=	285,388.48	Sr^2=	5848.124531	
177					within	v32=	640,221.63	dbar_U3=	-0.14725	
178						v3=	925,610.11			
179										
180					between	v41=	57,553.27	Sr^2=	643.292902	
181					within	v42=	1,008,375.22	dbar_U4=	-0.177590909	
182						v4=	1,065,928.49			
183										
184										
185					variance of strata		13,975,614.78			
186					Note: This variance is computed using alpha's as urban/rural, i.e. 0,1 case, it is almost 1/3 of the variance for the unstratified					

foodbanksCostvarianceVariance2sampling sheetsample of foodbanks

Average: 21251374.97Count: 12Sum: 106256874.8100%

# Excel Solver - Variance

foodbankssample design - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View SAS

Clipboard Font Alignment Number Styles Cells Editing

O169 between

	K	L	M	N	O	P	Q	R	S	T
151	-95.55958	33582025	4	0.197	0	-0.197	0.158191	2720023.501	158454.5067	
152	-95.55958	3025	4	0.197	0	-0.197	0.158191	245.0141434	5.65532825	
153	-95.55958	1923769	4	0.197	0	-0.197	0.158191	155818.3847	8910.277339	
154						N=	152	n=	30	
155						dbar_U=	0.1	m=	33.33333333	
156	ave.mileage	ni	miles		variance bt(before strata)	V1=	44,437,453.60	Sr^2=	71889.75453	
157	17.87384615	22.30263158	2391.803		variance w(before strata)	V2=	7,877,640.23	sum of within variance comp	1554797.413	
158	56.28333333	0.986842105	333.2566			var=	52,315,093.83			
159	62.12857143	2.368421053	882.8797							
160	79.85909091	4.342105263	2080.539							
161		Total miles:	5688.479							
162						N1=	113.00	n1=	22.30263158	
163						N2=	5.00	n2=	0.986842105	
164		Total Travel Cost:	5110.15			N3=	12.00	n3=	2.368421053	
165		Total Travel Time:	189.616 hours			N4=	22.00	n4=	4.342105263	
166										
167										
168										
169					between	v11=	5,752,005.70	Sr^2=	12517.05868	
170					within	v12=	6,001,863.76	dbar_U1=	0.176212389	
171						v1	11,753,869.46			
172										
173					between	v21=	3,027.11	Sr^2=	148.8741365	
174					within	v22=	227,179.61	dbar_U2=	-0.1846	
175						v2=	230,206.72			
176										
177					between	v31=	285,388.48	Sr^2=	5848.124531	
178					within	v32=	640,221.63	dbar_U3=	-0.14725	
179						v3=	925,610.11			
180										
181					between	v41=	57,553.27	Sr^2=	643.292902	
182					within	v42=	1,008,375.22	dbar_U4=	-0.177590909	
183						v4=	1,065,928.49			
184										
185					variance of strata		13,975,614.78			
186										
187										
188										
189										

Note: This variance is computed using alpha's as urban/rural, i.e. 0,1 case, it is almost 1/3 of the variance for the unstratified case. In the fourth sheet ("variance2"), the alpha's are unemployment rate, variance is different. I allocated the sample sizes (n1---n4 here) proportionally according to the strata size. ---Bingche

foodbanks Cost variance Variance2 sampling sheet sample of foodbanks

Ready Average: 1398519.329 Count: 48 Sum: 27970386.58 100%



# Excel Solver – Minimize Sample Size Based on Variance

foodbanksample design - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View SAS

Get External Data Existing Connections Refresh All Edit Links Connections Sort & Filter Filter Clear Reapply Advanced Test to Columns Remove Duplicates Data Validation Consolidate What-If Analysis Group Ungroup Subtotal Show Detail Hide Detail Data Analysis Solver

S154  $=((Q171+Q175)+Q179)+Q183$

	K	L	M	N	O	P	Q	R	S	T
152	-95.55958	3025	4	0.197		0	-0.197	0.158191	245.0141434	20.79691704
153	-95.55958	1923769	4	0.197		0	-0.197	0.158191	155818.3847	18539.67209
154						N=	152	n=		
155						dbar_U=	0.1	m=		
156	ave.mileage	ni	miles		variance bt(before strata)	V1=	16,017,700.06			
157	17.87384615	45.82598061	3276.346		variance w(before strata)	V2=	7,937,878.33			
158	56.28333333	2.027698257	456.5025			var=	23,955,578.39			
159	62.12857143	4.866475817	1209.389							
160	79.85909091	8.921872331	2849.97							
161		Total miles:	7792.208							
162						N1=	113.00			
163						N2=	5.00			
164		Total Travel Cost:				N3=	12.00			
165		Total Travel Time:	259.7403	hours		N4=	22.00			
166										
167										
168										
169					between	v11=	2,073,338.92			
170					within	v12=	6,050,457.88			
171						v1	8,123,796.80			
172										
173					between	v21=	1,091.14			
174					within	v22=	229,117.29			
175						v2=	230,208.43			
176										
177					between	v31=	102,869.69			
178					within	v32=	644,046.29			
179						v3=	746,915.97			
180										
181					between	v41=	20,745.36			
182					within	v42=	1,014,256.87			
183						v4=	1,035,002.24			
184										
185						variance of strata	10,135,923.43			
186										
187										
188										
189										
190										

Solver Parameters

Set Objective:  $S$154$

To: ☐ Max ☒ Min ☐ Value Of: 0

By Changing Variable Cells:  $S$154$

Subject to the Constraints:  $S$164 <= 7000$

☒ Make Unconstrained Variables Non-Negative

Select a Solving Method: GRG Nonlinear

Solving Method  
Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simple engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.

Help Solve Close

Note: This variance is computed using alpha's as urban/rural, i.e. 0,1 case, it is almost 1/3 of the variance for the unst...  
In the fourth sheet ("variance2"), the alpha's are unemployment rate, variance is different  
I allocated the sample sizes (n1---n4 here) proportionally according to the strata size. ---Bingche

foodbanks Cost variance Variance2 sampling sheet sample of foodbanks

9:58 AM 3/6/2014

# Excel Solver – Final Sample

foodbankssample design - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	id	id	clients	mileage	hia	county	city	zip	map	name	latitude	longitude	psu	cumsize	Sample	Stokes, Lynne: These were sampled with seed 1357 using uniform random number generator	ample id		
1																			
2	1	026012P	956	14.2		DALLAS	DALLAS	75217	119	WAY OF T	32.72238	-96.6758	1	956	18109.3182		6800P	18109.32	
3	4	002052P	35	11.3		DALLAS	DALLAS	75208	95	SALVATIC	32.75146	-96.8386	1	991	46780.3797		2137P	46780.38	
4	5	026550P	40	3.7	YES	DALLAS	DALLAS	75209	97	GREATER	32.84598	-96.8255	1	1031	59144.9523		2341P	59144.95	
5	7	002145P	50	13.4		DALLAS	DALLAS	75227	126	PATHWAY	32.77003	-96.69	1	1081	78696.5339		6769P	78696.53	
6	9	026587P	61	7.2		DALLAS	DALLAS	75243	147	LIFENET S	32.91223	-96.7369	1	1142	55891.7571		6285P	55891.76	
7	10	020083P	62	12		DALLAS	DALLAS	75212	105	GOOD SH	32.78238	-96.8695	1	1204	129451.24		6431P5	129451.2	
8	11	026284P	70	16.1		DALLAS	DALLAS	75233	135	DIVINE IN	32.7044	-96.8722	1	1274	19806.8136		6226P	19806.81	
9	15	002138P	80	23.6		DALLAS	CEDAR HI	75104	53	NEW HOF	32.58854	-96.9495	1	1354	100881.462		152 002074P	100881.5	
10	17	002013P	93	15.2		DALLAS	DALLAS	75211	101	BAPTIST E	32.73693	-96.8825	1	1447	13636.6812		81 026050P	13636.68	
11	18	002597P	102	27		ELLIS	RED OAK	75154	78	NORTH EL	32.52612	-96.8233	1	1549	43575.8002		123 002051P	43575.8	
12	19	002719P	105	19.9		DENTON	CARROLL	75007	4	HOLY CO	33.00542	-96.8977	1	1654	7097.88043		58 002164P	7097.88	
13	21	026056P2	120	14.6		DALLAS	DALLAS	75211	102	WEST OA	32.73693	-96.8825	1	1774	90489.8729		149 026564P	90489.87	
14	22	026088P	128	19.5		DALLAS	IRVING	75061	27	DAYSPIR	32.82673	-96.9614	1	1902	113878.037		156 026588P	113878	
15	23	002141P	130	13.3	YES	DALLAS	GARLAND	75044	16	EXTENDE	32.96038	-96.6619	1	2032	30648.0967		111 002030P	30648.1	
16	24	026055P	131	4.1		DALLAS	DALLAS	75209	98	SERVICE	32.84598	-96.8255	1	2163	87143.4977		148 026099P2	87143.5	
17	25	002350P	137	38.9		KAUFMAN	KAUFMAN	75142	67	KAUFMAN	32.5743	-96.3038	1	2300	65979.4981		139 026637P	65979.5	
18	26	002669P	141	38	YES	KAUFMAN	KAUFMAN	75142	68	SALVATIC	32.5743	-96.3038	1	2441	115385.121		156 026588P	115385.1	
19	28	002125P	150	9.2	YES	DALLAS	DALLAS	75223	123	METHOD	32.79288	-96.7492	1	2591	110722.073		155 002218P	110722.1	
20	29	002497P	152	30.9		ELLIS	MIDLOTHI	76065	183	MANNA H	32.48409	-96.9867	1	2743	120254.786		157 008036P	120254.8	
21	31	002621P	160	18.9		DALLAS	DUNCANV	75116	60	RISEN CH	32.65838	-96.9127	1	2903	87609.3974		148 026099P2	87609.4	
22	33	026810P	162	42.1		DENTON	DENTON	76201	185	FIRST BA	33.2266	-97.1306	1	3065	50851.9379		129 001041P	50851.94	
23	34	002093P	165	14		DALLAS	DALLAS	75216	114	HOLY CRC	32.71008	-96.7972	1	3230	9998.61238		71 002408P	9998.612	
24	35	026508P	166	15.1	DECLINE	COLLIN	PLANO	75075	38	LIFESOUR	33.02472	-96.7404	1	3396	61089.577		135 002680P	61089.58	
25	36	002136P	168	14.6		DALLAS	DALLAS	75217	116	NEWMAN	32.72238	-96.6758	1	3564	37381.36		117 002117P	37381.36	
26	37	026027P	173	17.9		DALLAS	IRVING	75061	28	GOOD SH	32.82673	-96.9614	1	3737	40727.7351		121 002086P1	40727.74	
27	39	002614P	182	15.7		DALLAS	DALLAS	75241	145	CLIFF VIE	32.66938	-96.7744	1	3919	1397.69906		17 002013P	1397.699	
28	40	026799P	201	9.9		DALLAS	GARLAND	75041	13	HEARTS /	32.88153	-96.646	1	4120	42279.3836		122 022586P	42279.38	
29	41	026431P2	204	14.2		COLLIN	PLANO	75075	40	SVDP ST	33.02472	-96.7404	1	4324	48485.9777		127 026095P	48485.98	
30	42	026072P	220	10.8		DALLAS	DALLAS	75212	107	VOICE OF	32.78238	-96.8695	1	4544	17108.6467		88 002360P	17108.65	
31	43	002149P	220	36.6		DENTON	ROANOKE	76262	188	NORTHWE	33.00018	-97.2187	1	4764	23708.217		99 002499P	23708.22	
32	44	003250P	222	36.5	YES	ELLIS	WAXAHAC	75165	83	SALVATIC	32.39702	-96.8328	1	4986	46573.7634		126 002137P	46573.76	
33	45	002505P	224	35.4		ELLIS	WAXAHAC	75165	82	BROWN S	32.39702	-96.8328	1	5210				45176.06	
34	47	026077P	250	20.7		COLLIN	ALLEN	75002	2	FBC ALLE	33.09285	-96.6245	1	5460				125120.4	
35	49	026431P	280	6.1	YES	DALLAS	DALLAS	75208	144	ST PATRIC	32.87393	-96.7092	1	5740				87066.52	
36	50	026011P	287	4.5		DALLAS	DALLAS	75206	93	KINGDOM	32.82613	-96.7712	1	6027				100930.7	
37	51	026766P	288	9.5		DALLAS	DALLAS	75210	100	NORTH TE	32.77103	-96.7473	1	6315				106350.7	
38	52	002375P	293	39.3		ELLIS	ENNIS	75119	62	HELPING	32.33124	-96.6196	1	6608				113363.5	
39	55	026797P	324	37.1	YES	COLLIN	PRINCET	75407	155	CHRIST C	33.14572	-96.4966	1	6932					
40	58	002164P	353	9.1		DALLAS	GARLAND	75042	14	METRO FA	32.91563	-96.674	1	7285					
41	60	026068P	377	7.5		DALLAS	DALLAS	75215	111	ST PHILIP	32.76103	-96.7704	1	7662					
42	63	008146P	396	12		COLLIN	RICHARD	75081	42	RICHARD	32.94897	-96.7092	1	8058					
43	64	026590P	397	30.7	YES	ELLIS	MIDLOTHI	76065	184	SENIOR C	32.48409	-96.9867	1	8455					
44	65	026812P	400	8.7		DALLAS	DALLAS	75229	129	TEENA'S	32.89538	-96.8599	1	8855					
45	68	026031P	424	14.6	YES	DALLAS	DALLAS	75211	103	MOUNTAI	32.73693	-96.8825	1	9279					

Ready | foodbanks | Cost | variance | Variance2 | sampling sheet | sample of foodbanks | 100%



# Excel Solver – Final Sample

foodbankssample design - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
130	170	026851P	956	15.8		DALLAS MESQUITE	75150	76	TCW OUT	32.81839	-96.6336	1	131793						
131	172	026867P	956	37.2		ROCKWAI ROYSE CI	75189	NA	FUMC RO	32.96051	-96.3128	1	132749						
132	48	026515P	267	41.9		ELLIS ITALY	76651	191	ITALY MIN	32.17578	-96.8802	2	267	1433.40556					
133	53	026682P	315	60		NAVARRC FROST	76641	190	FROST CC	32.04427	-96.7862	2	582	4509.08536					
134	67	005164P	423	59.4		NAVARRC CORSICA	75151	77	SALVATIC	32.0624	-96.4735	2	1005						
135	110	002554P	898	58.5		NAVARRC CORSICA	75110	55	NORTHSIC	32.07823	-96.4461	2	1903						
136	147	002491P	2941	59.5	YES	NAVARRC CORSICA	75110	54	NAVARRC	32.07823	-96.4461	2	4844						
137	171	002074P5	956	58.4		KAUFMAN MABANK	75147	71	CCS- FUM	32.31426	-96.1107	2	5800						
138	2	026040P	26	59.4	DECLINE	GRAYSON SHERMAN	75090	47	SUBSTAN	33.63281	-96.5911	3	26	3973.67791					
139	14	002161P	77	61.9		GRAYSON SHERMAN	75090	44	HARMONY	33.63281	-96.5911	3	103	2248.11353					
140	16	002616P	90	58.1		GRAYSON TOM BEA	75489	180	FIRST BA	33.52158	-96.4837	3	193	9061.58513					
141	46	026086P	227	66	YES	GRAYSON SHERMAN	75090	46	SALVATIC	33.63281	-96.5911	3	420						
142	59	003109P	368	72		GRAYSON DENISON	75020	6	JACK & JII	33.75405	-96.5787	3	788						
143	61	001037P	378	62.4	YES	GRAYSON BELLS	75414	157	COMMUNI	33.61563	-96.419	3	1166						
144	76	002111P	513	60.3		GRAYSON SHERMAN	75090	45	KEY CAR	33.63281	-96.5911	3	1679						
145	77	026001P	539	65.7		FANNIN BONHAM	75418	158	BASIC FO	33.58377	-96.1818	3	2218						
146	83	002380P	605	50.2	YES	DENTON PILOT POI	76258	187	SHEPHER	33.38565	-96.9521	3	2823						
147	90	002204P	632	70.1		GRAYSON DENISON	75020	5	DENISON	33.75405	-96.5787	3	3455						
148	101	002200P	760	65.8	YES	FANNIN BONHAM	75418	160	MANNA H	33.58377	-96.1818	3	4215						
149	109	002397P	843	69.6		GRAYSON DENISON	75020	7	NEW BEG	33.75405	-96.5787	3	5058						
150	158	026735P	5090	66.9	YES	FANNIN BONHAM	75418	159	FANNIN C	33.58377	-96.1818	3	10148						
151	160	026852P	956	41.4		COLLIN ANNA	75409	156	ANNA FO	33.35609	-96.5191	3	11104						
152	3	026623P	26	86.3		HOPKINS SULPHUR	75482	178	TRINITY H	33.14375	-95.601	4	26	6966.38627					
153	6	026627P	43	88.6		HOPKINS SULPHUR	75482	176	SOUTH LI	33.14375	-95.601	4	69	5174.8869					
154	8	005170P	55	105		LAMAR PARIS	75460	168	SALVATIC	33.66027	-95.5596	4	124	738.425123					
155	12	026644P	71	88.9		DELTA COOPER	75432	165	NEW LIFE	33.37701	-95.6857	4	195	12684.634					
156	13	026685P	74	46.8		HUNT CADDO M	75135	65	TWELVE C	33.07565	-96.2377	4	269	6274.34007					
157	20	026628P	118	85.9	YES	DELTA PECAN G	75469	170	PECAN G	33.43474	-95.8306	4	387	2748.63292					
158	27	026640P	148	85		HOPKINS SULPHUR	75483	179	FRUITFUL	33.16861	-95.5855	4	535						
159	30	026646P	153	83.2		HOPKINS SULPHUR	75482	173	CENTRAL	33.14375	-95.601	4	688						
160	32	002364P2	160	55.2		HUNT GREENVIL	75401	151	BECAUSE	33.15023	-96.1129	4	848						
161	38	026625P	178	70.6		HOPKINS CUMBY	75433	166	CUMBY FI	33.13307	-95.8177	4	1026						
162	54	026632P	317	84.2		HOPKINS SULPHUR	75482	174	COMMUNI	33.14375	-95.601	4	1343						
163	56	026631P	338	84.2		HOPKINS SULPHUR	75482	175	MINISTER	33.14375	-95.601	4	1681						
164	57	026636P	345	47.7	YES	HUNT QUINLAN	75474	171	LAKE ARE	32.8963	-96.07	4	2026						
165	62	026624P	385	103	YES	HOPKINS PICKTON	75471	192	PINE FOR	33.04158	-95.3953	4	2411						
166	66	026639P	400	83.9	YES	HOPKINS SULPHUR	75482	172	CALVARY	33.14375	-95.601	4	2811						
167	69	026622P	428	86.6		DELTA COOPER	75432	164	FUMC CO	33.37701	-95.6857	4	3239						
168	73	026096P	496	72.3	YES	HUNT WOLFE C	75496	181	WOLFE C	33.35054	-96.0571	4	3735						
169	93	002562P	657	71.3		HUNT COMMERC	75429	161	WORD OF	33.23772	-95.9089	4	4392						
170	120	002364P	1313	56.1		HUNT GREENVIL	75401	152	HANDS OI	33.15023	-96.1129	4	5705						
171	125	026271P2	1387	111		LAMAR PARIS	75460	169	SVDP OUI	33.66027	-95.5596	4	7092						
172	138	002400P	2012	56.1		HUNT GREENVIL	75401	153	HUNT COL	33.15023	-96.1129	4	9104						
173	159	026551P	5795	105	YES	LAMAR PARIS	75460	167	DOWNTON	33.66027	-95.5596	4	14899						
174																			

Stokes, Lynne: These were sampled using seed 2468 on uniform random number generator

Stokes, Lynne: These were sampled using seed 3579 on uniform random number generator

Stokes, Lynne: These were sampled using 4680 on random number generator

125 026271P2  
66 026639P

foodbanks Cost variance Variance2 sampling sheet sample of foodbanks

Ready 100%

# Final Sample

- Eight Interviews per Day
  - 42 Food Banks
  - 24 Clients per Food Bank
- Twelve Interviews per Day
  - 30 Food Banks
  - 33 Clients per Food Bank



# Other Practical Issues

- How do we select clients once we're actually inside the food banks?
  - As randomly as possible
  - No possible frame
  - Choose them systematically, i.e., every tenth person
- Non-response: Clients may choose not to participate

# Concluding Remarks

- When in doubt, it's helpful to think of what we would do if we had more data.
- It's also helpful to start from a simple scenario where you know all the formulas and analysis and work your way up to a more complicated scenario.



# Concluding Remarks

- This is not a math problem! We rarely have enough information to find the optimal solution, and if we do, it may be too complicated. Our goal is to do something reasonable.
- Part of doing something reasonable is to take a look at different scenarios and see how the results may be affected.
- If you spend time making a complicated design, you need to make sure the analysis is also done the right way.

# References

- Stokes, L., & Plummer, J. (2004). Using Spreadsheet Solvers in Sample Design. *Computational Statistics & Data Analysis* , 44 (3), 527-546.

Special thanks to the STAT 6380 students, especially Chelsea Allen and Charles South, for helping finish this project and providing materials for this presentation.