

An abstract graphic featuring three sets of concentric circles in shades of orange. One large set is in the top right, a medium set is in the center, and another large set is in the bottom right. Thin orange lines cross the page diagonally from the top left and bottom right towards the center.

STAT 5060 Project

Ho Joon Choi : Brian Sigurdson : Scott Smith

In this project we attempt to estimate the percentage of “badly damaged” books in the Jerome Library’s first and eighth floor stacks. A pilot study will be used to estimate the necessary sample size for the formal study. A one-stage cluster sample with ratio estimation will be employed to estimate the percentage of “badly damaged” books.

4/27/2010

Table of Contents

Original Proposal.....	3
General Description	3
Population Parameter.....	3
Data Gathering	3
Sample Size Determination	4
Time-frame collection.....	4
Figure 1 First Floor.....	5
Figure 2 Eighth Floor	6
Changes Suggested by Instructor	6
Data Gathering.....	7
Findings	7
Changes and Adjustments	8
Conclusion	8
Appendix A: Summary of Stack / Block Layout.....	9
Appendix B: Pilot Data.....	14
Appendix C: SRS Data	15
Ho Joon Choi	15
Scott Smith.....	16
Brian Sigurdson	17

Original Proposal

The following is the original proposal for Group four.

General Description

Group 4 proposes to estimate the total number of “badly damaged” books on the Jerome Library first and eight floor stacks. A One-Stage Cluster Sampling approach will be employed to estimate the total number of “badly damaged” books and its associated confidence interval. A detailed description of “badly damaged” is provided later in this proposal.

The library stacks are numbered 1 – 192 and will serve as the natural clusters, or primary sampling units (psu), for the sample. The library’s maps of the stacks are provided at the end of this proposal.

The secondary sampling units (ssu) will be the books found on a specific section of the cluster’s shelf. Additional details on the exact location of the ssu are provided later in this proposal.

Population Parameter

The population parameter to be estimated will be the total number of “badly damaged” books. The desired bound is 10%. A pilot study of 10 randomly sampled clusters will be conducted to help estimate the between and within variation so that a SRS cluster size, n , can be determined. The bound may need to be reevaluated after the pilot study, if the SRS cluster size is prohibitively large.

Data Gathering

A SRS of clusters to sample will be determined by a random number generator to select the required number of cluster from the 192 clusters available. According to the maps, and personal observation, the library stacks face North and South. The main library entrance faces East and West.

Once a stack has been located, proceed to the North-West corner of the stack and locate the bottom shelf. Depending on the length of the stack, a shelf will be split into one or more sections. Select the most northern section of the bottom shelf. These are the books, the ssu, to be inspected.

A “badly damaged” book is one in which the corners or the spine are in poor condition. Poor condition for the corner is when there is excessive rounding of the corner, often causing the cover material to split, displaying the underlying material. Poor condition for the spine is when is no longer stiff, as when new, but has become rather loose from wear. This is evidenced by being able to apply opposing pressure to the front and back cover of the text with the palms of your hands, while holding the text in front of you, and feeling that the spine does not provide much resistance to the movement of the text’s front and back covers.

This is easier to do in practice than describe on paper. In extreme cases the glue will have begun to crack and pages may already begin to release themselves from the spine.

Sample Size Determination

The formula to estimate the required SRS cluster size is as follows:

$$n = \frac{(Z_{\alpha/2}^2)(v)}{e^2}$$

and

$$v = \left[\frac{MSB}{M} + \left(1 - \frac{m}{M}\right) \frac{MSW}{m} \right] = \left[\frac{MSB}{\bar{M}} \right]$$

Where, the MSB and the MSW are estimated from the pilot study. Since we'll be using a One-Stage Cluster Sampling approach, M is the number of ssu in each cluster, which equals m, the number of ssu to sample in each cluster. Since we have unequal ssu values, and we'll be taking a census of each cluster, \bar{M} will equal the average ssu from the pilot study.

Time-frame collection

The data will be collected as soon as possible. An example of the data collection for follows:

Group 4 Data Collection Form	
Member:	
Cluster #:	
First Call #:	
Last Call #:	
Cluster size:	

Data collection Section		
Book #	Corners	Spine
1		
2		
3		
4		
5		
6		
7		
8		
9		

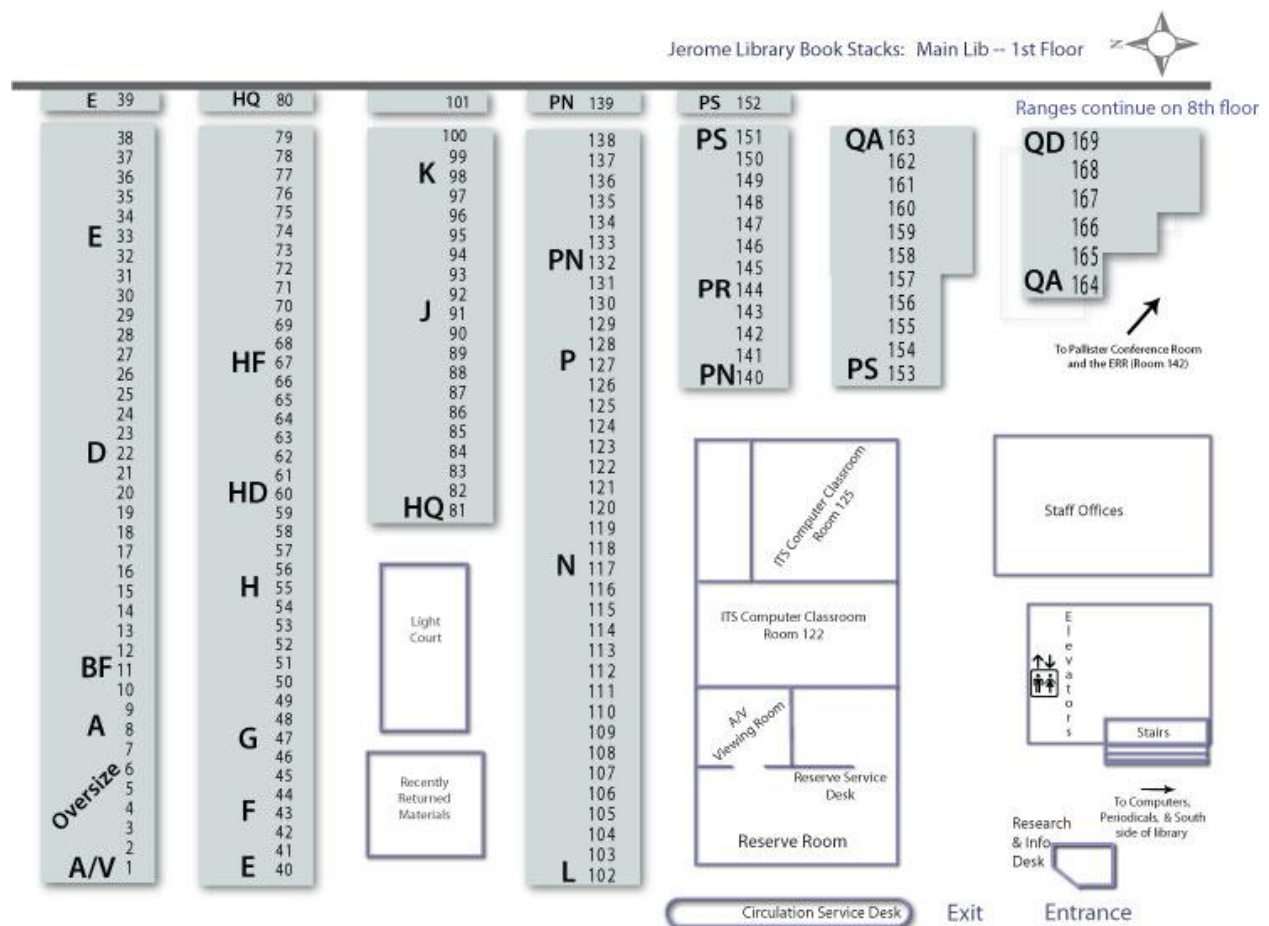


Figure 1 First Floor

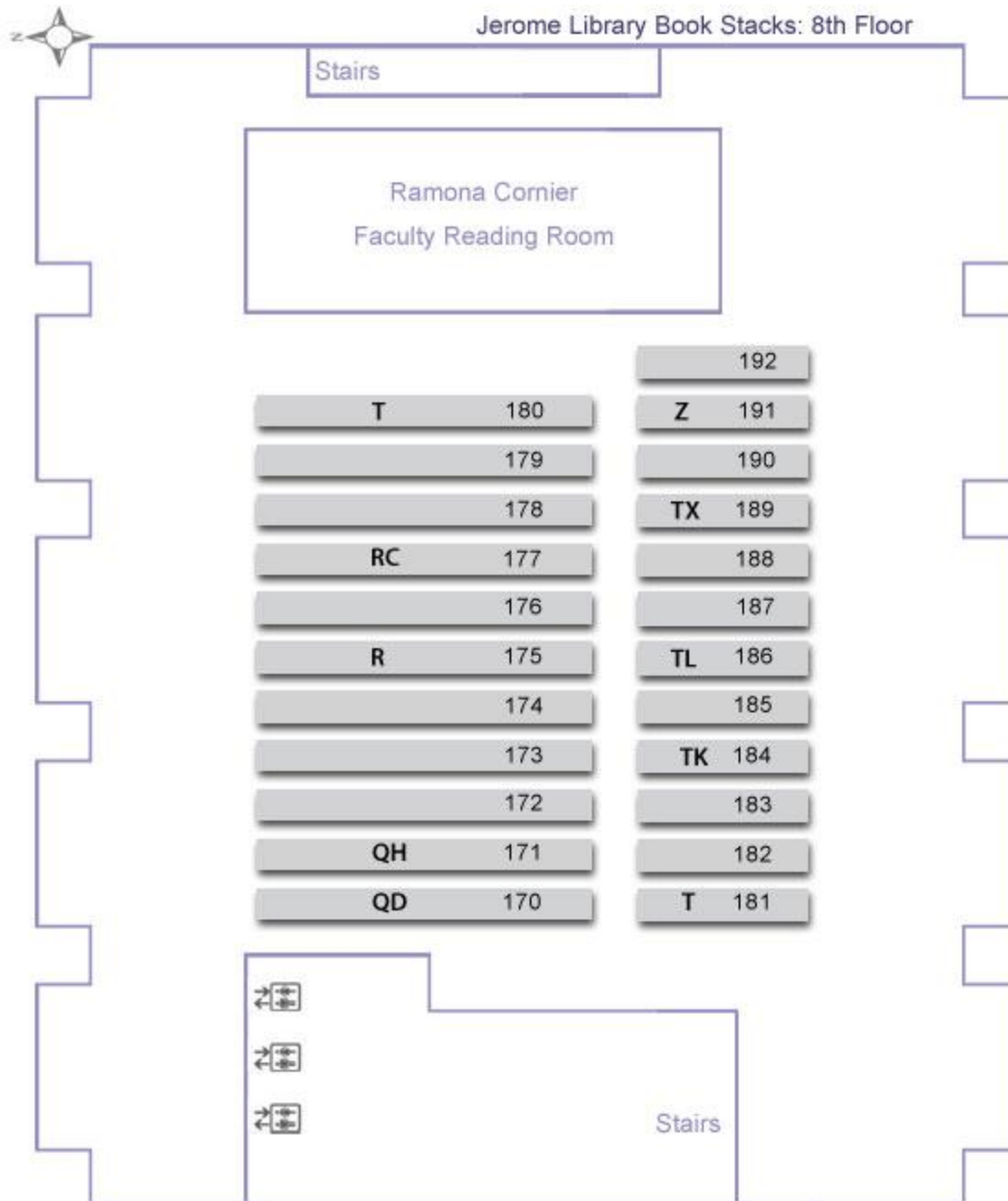


Figure 2 Eighth Floor

Changes Suggested by Instructor

We had several discussions with the instructor about various approaches and modifications to our project. Eventually, we decided that a one-stage cluster sample with ratio estimation was likely the most efficient and effective strategy for our population. This is expanded on in the next section.

Data Gathering

We mapped the first and eighth floor stacks by noting the number of blocks in each stack. A stack's blocks were determined to be the intersection of its rows and columns. Each stack is essentially a two-sided matrix with six rows, but whose columns varied from stack to stack. A column is hard to describe, but easy to see when looking at a stack. This was not as difficult as it may sound, as the stacks had long "runs" in which the number of blocks per stack did not change. We then assigned each block of books a unique number. Our final block count was 12,834, which was a quite a bit more than we expected. The stack summary mapping is found on page 9, Appendix A.

We then conducted a pilot study of 20 clusters to determine the between and within variation to use in deriving the necessary sample size for our project. The pilot study data is located on page 14, Appendix B. The calculations are as follows:

$$N = 12834; \quad n = 18; \quad \bar{M} = 31.263$$

$$\hat{p} = .045; \quad s_p^2 = \frac{\sum_{i=1}^{19} (d_i - \hat{p}M_i)^2}{18} = 2.002066$$

$$n \approx \frac{Ns_p^2}{NB^2 \frac{\bar{M}^2}{z_{\alpha/2}^2} + s_p^2} = \frac{12834 (2.002066)}{\frac{12834 (.0001)(31.263)^2}{1.96^2} + 2.002066} = 78.21 \approx 79$$

We decided to error on the side of conservatism and sample 27 clusters each, for a total sample of 81 clusters.

Now, we were each assigned a count sheet of clusters to sample, with five extra random numbers in case there was a problem with the originally selected block. These sheets and their values are included on page 15, Appendix C. We each used one extra random number due to the selected block being empty.

For our actual SRS count, we inspected each book in the selected block, noting the number of "bad" books. We also counted the total number of books in the cluster to use as our auxiliary variable. We used these values in deriving our estimated proportion of "badly damaged" books on the stacks.

Findings

Out of 2297 sampled books, 62 were found to be in poor condition. Since the study was designed as a one-stage cluster sample with 12834 clusters (each cluster being a section of shelf surveyed), the following calculations result.

First, we estimated proportion of books in poor condition.

$$\hat{p} = \bar{y}_{est} = \frac{\sum_{i=1}^n t_i}{\sum_{i=1}^n M_i} = \frac{62}{2297} = .026992$$

Next, we estimated of the variance of the proportion estimate must be calculated in order to produce a bound on the error of estimation.

$$\hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{1}{nM} s_p^2 \quad \text{where} \quad s_p^2 = \frac{\sum_{i=1}^n (t_i - \hat{p}M_i)^2}{n-1} = .999268$$

Let \bar{M} be estimated by the average cluster size in the sample, 28.358.

$$\hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{1}{n\bar{M}^2} s_p^2 = \left(1 - \frac{81}{12834}\right) \frac{1}{(81)(28.358)^2} .999268 = .0015052$$

$$SE_{\hat{p}} = \sqrt{\hat{V}(\hat{p})} = .0038797$$

In order to place a bound on the error of estimation, we use a z-value of 1.96. Multiplying this by the estimated standard error of the sample proportion, we are 95% confident that the true proportion of books in poor condition is between .01938 and .03460.

Changes and Adjustments

If we redid the project, an area that we would likely expand upon is the use of pilot studies. We would likely use three pilots studies to determine our SRS size. Their results would be averaged together to produce the SRS for the project.

An area that we would not reduce is the amount of pre-sample planning that should go into a project. We didn't feel too inhibited in our project execution and analysis, because he had spent a considerable amount of time discussing our population's structure and the best way to approach it.

Conclusion

The group members were very flexible in the project approach and execution of activities. Everyone willingly accepted responsibility for various tasks, which kept the project steadily moving forward from inception. We generally believe that the one-stage cluster sampling with ratio estimation is still the best approach for this project, given our current level of skill and knowledge of sampling methods.

Appendix A: Summary of Stack / Block Layout

Stack #	# shelves	# blocks	1 or 2 sided	Begin Block	End Block	Comment
1	0	0	2	0	0	DVD's
2	0	0	2	0	0	empty
3	0	0	2	0	0	oversized books
4	0	0	2	0	0	oversized books
5	0	0	2	0	0	oversized books
6	0	0	2	0	0	oversized books
7	6	6	2	1	72	
8	6	6	2	73	144	
9	6	6	2	145	216	
10	6	6	2	217	288	
11	6	6	2	289	360	
12	6	6	2	361	432	
13	6	6	2	433	504	
14	6	6	2	505	576	
15	6	3	2	577	612	
16	6	4	2	613	660	
17	6	6	2	661	732	
18	6	6	2	733	804	
19	6	6	2	805	876	
20	6	6	2	877	948	
21	6	6	2	949	1020	
22	6	6	2	1021	1092	
23	6	6	2	1093	1164	
24	6	6	2	1165	1236	
25	6	6	2	1237	1308	
26	6	6	2	1309	1380	
27	6	6	2	1381	1452	
28	6	6	2	1453	1524	
29	6	6	2	1525	1596	
30	6	6	2	1597	1668	
31	6	6	2	1669	1740	
32	6	6	2	1741	1812	
33	6	6	2	1813	1884	
34	6	6	2	1885	1956	
35	6	6	2	1957	2028	
36	6	6	2	2029	2100	
37	6	6	2	2101	2172	

38	6	6	2	2173	2244	
39	6	5	1	2245	2274	
40	6	5	2	2275	2334	
41	6	5	2	2335	2394	
42	6	5	2	2395	2454	
43	6	5	2	2455	2514	
44	6	5	2	2515	2574	
45	6	5	2	2575	2634	
46	6	5	2	2635	2694	
47	6	5	2	2695	2754	
48	6	5	2	2755	2814	
49	6	5	2	2815	2874	
50	6	5	2	2875	2934	
51	6	5	2	2935	2994	
52	6	5	2	2995	3054	
53	6	5	2	3055	3114	
54	6	5	2	3115	3174	
55	6	5	2	3175	3234	
56	6	5	2	3235	3294	
57	6	5	2	3295	3354	
58	6	5	2	3355	3414	
59	6	5	2	3415	3474	
60	6	5	2	3475	3534	
61	6	5	2	3535	3594	
62	6	5	2	3595	3654	
63	6	5	2	3655	3714	
64	6	5	2	3715	3774	
65	6	5	2	3775	3834	
66	6	5	2	3835	3894	
67	6	5	2	3895	3954	
68	6	5	2	3955	4014	
69	6	5	2	4015	4074	
70	6	5	2	4075	4134	
71	6	5	2	4135	4194	
72	6	5	2	4195	4254	
73	6	5	2	4255	4314	
74	6	5	2	4315	4374	
75	6	5	2	4375	4434	
76	6	5	2	4435	4494	
77	6	5	2	4495	4554	
78	6	5	2	4555	4614	

79	6	5	2	4615	4674	
80	6	5	2	4675	4734	
81	6	6	2	4735	4806	
82	6	6	2	4807	4878	
83	6	6	2	4879	4950	
84	6	6	2	4951	5022	
85	6	6	2	5023	5094	
86	6	6	2	5095	5166	
87	6	6	2	5167	5238	
88	6	6	2	5239	5310	
89	6	6	2	5311	5382	
90	6	6	2	5383	5454	
91	6	6	2	5455	5526	
92	6	6	2	5527	5598	
93	6	6	2	5599	5670	
94	6	6	2	5671	5742	
95	6	6	2	5743	5814	
96	6	6	2	5815	5886	
97	6	6	2	5887	5958	
98	6	6	2	5959	6030	
99	6	6	2	6031	6102	
100	6	6	2	6103	6174	
101	6	6	2	6175	6246	
102	6	3	1	6247	6264	
103	6	5	2	6265	6324	
104	6	5	2	6325	6384	
105	6	5	2	6385	6444	
106	6	5	2	6445	6504	
107	6	5	2	6505	6564	
108	6	5	2	6565	6624	
109	6	4	2	6625	6672	
110	6	4	2	6673	6720	
111	6	5	2	6721	6780	
112	6	5	2	6781	6840	
113	6	5	2	6841	6900	
114	6	5	2	6901	6960	
115	6	5	2	6961	7020	
116	6	5	2	7021	7080	
117	6	5	2	7081	7140	
118	6	5	2	7141	7200	
119	6	5	2	7201	7260	

120	6	5	2	7261	7320	
121	6	5	2	7321	7380	
122	6	5	2	7381	7440	
123	6	5	2	7441	7500	
124	6	5	2	7501	7560	
125	6	5	2	7561	7620	
126	6	5	2	7621	7680	
127	6	5	2	7681	7740	
128	6	6	2	7741	7812	
129	6	6	2	7813	7884	
130	6	6	2	7885	7956	
131	6	6	2	7957	8028	
132	6	6	2	8029	8100	
133	6	6	2	8101	8172	
134	6	6	2	8173	8244	
135	6	6	1	8245	8280	
136	6	0	0	8281	8280	empty
137	6	0	0	8281	8280	empty
138	6	0	0	8281	8280	empty
139	6	0	0	8281	8280	empty
140	6	0	0	8281	8280	empty
141	6	0	0	8281	8280	empty
142	6	0	0	8281	8280	empty
143	6	0	0	8281	8280	empty
144	6	8	2	8281	8376	
145	6	5	2	8377	8436	
146	6	7	2	8437	8520	
147	6	7	2	8521	8604	
148	6	7	2	8605	8688	
149	6	7	2	8689	8772	
150	6	6	2	8773	8844	
151	6	7	2	8845	8928	
152	6	7	2	8929	9012	
153	6	7	2	9013	9096	
154	6	7	1	9097	9138	
155	6	7	2	9139	9222	
156	6	7	2	9223	9306	
157	6	7	2	9307	9390	
158	6	6	2	9391	9462	
159	6	9	2	9463	9570	
160	6	10	2	9571	9690	

161	6	10	2	9691	9810	
162	6	9	2	9811	9918	
163	6	10	2	9919	10038	
164	6	10	2	10039	10158	
165	6	10	2	10159	10278	
166	6	3	2	10279	10314	
167	6	6	2	10315	10386	
168	6	8	2	10387	10482	
169	6	11	2	10483	10614	
170	6	11	2	10615	10746	
171	6	11	2	10747	10878	
172	6	9	2	10879	10986	
173	6	9	2	10987	11094	
174	6	9	2	11095	11202	
175	6	9	2	11203	11310	
176	6	9	2	11311	11418	
177	6	9	2	11419	11526	
178	6	9	2	11527	11634	
179	6	7	2	11635	11718	
180	6	7	2	11719	11802	
181	6	7	2	11803	11886	
182	6	7	2	11887	11970	
183	6	6	2	11971	12042	
184	6	6	2	12043	12114	
185	6	6	2	12115	12186	
186	6	6	2	12187	12258	
187	6	6	2	12259	12330	
188	6	6	2	12331	12402	
189	6	6	2	12403	12474	
190	6	6	2	12475	12546	
191	6	6	2	12547	12618	
192	6	6	2	12619	12690	
193	6	6	2	12691	12762	
194	6	6	2	12763	12834	

Appendix B: Pilot Data

Pilot Sample Random Numbers							
Item #	Block #	Stack #	Row	Column	Side Block Located	total	bad
1	1465	28	1	3	1	26	3
2	2634	45	6	10	2	28	0
3	2733	47	3	7	2	31	4
4	3163	54	1	9	2	34	3
5	4261	73	1	2	1	28	0
6	4419	75	3	8	2	29	3
7	4650	79	6	6	2	21	0
8	5263	88	1	5	1	30	0
9	6157	100	1	10	2	43	3
10	6814	112	4	6	2	33	0
11	7986	131	6	5	1	46	3
12	8206	134	4	6	1	37	1
13	8599	147	1	14	2	27	0
14	9754	161	4	11	2	32	2
15	10096	164	4	10	1	23	0
16	10517	169	5	6	1	31	3
17	11700	179	6	11	2	36	0
18	12605	191	5	10	2		
19	12765	194	3	1	1	30	0

Extra Random Numbers							
20	1489	28	1	7	2		
21	1784	32	2	8	2		
22	3374	58	2	4	1		
23	9190	155	4	9	2		
24	10314	166	6	6	2		
25	11829	181	3	5	1	29	2

Appendix C: SRS Data

Ho Joon Choi

Ho Joon Choi							
Pilot Sample Random Numbers							
Item #	Block #	Stack #	Row	Column	Side Block Located	total	bad
1	275	10	5	10	2	32	0
2	1156	23	4	11	2	25	0
3	1199	24	5	6	1	23	0
4	2042	36	2	3	1	30	0
5	2306	40	2	6	2	27	2
6	2319	40	3	8	2	23	1
7	3029	52	5	6	2	29	3
8	4030	69	4	3	1	29	0
9	4056	69	6	7	2	22	0
10	4066	69	4	9	2	29	0
11	4480	76	4	8	2	22	1
12	4568	78	2	3	1	29	2
13	5219	87	5	9	2	25	0
14	6173	100	5	12	2	40	0
15	6557	107	5	9	2	39	0
16	6640	109	4	3	1	20	0
17	7868	129	2	10	2	27	1
18	8455	146	1	4	1		
19	8804	150	2	6	1	23	0
20	9027	153	3	3	1	31	1
21	9281	156	5	10	2	36	0
22	9329	157	5	4	1	28	0
23	10265	165	5	18	2	24	0
24	11617	178	1	16	2	36	0
25	11751	180	3	6	1	37	1
26	12538	193	4	3	1	32	0
27	12582	194	6	2	1	44	0

Extra Random Numbers							
1	2186	38	2	3	1	23	2
2	11061	173	3	13	2		
3	7513	124	1	3	1		
4	5763	95	3	4	1		
5	4212	72	6	3	1		

Scott Smith

Scott Smith							
Pilot Sample Random Numbers							
Item #	Block #	Stack #	Row	Column	Side Block Located	total	bad
1	597	15	3	4	2	24	2
2	848	19	2	8	2	37	2
3	1735	31	1	12	2	30	3
4	1861	33	1	9	2	26	0
5	2351	41	5	3	1	25	0
6	2542	44	4	5	1	21	2
7	2869	49	1	10	2	28	3
8	3065	53	5	2	1	29	1
9	3355	58	1	1	1	35	2
10	4109	70	5	6	2	25	1
11	4515	77	3	4	1	29	1
12	4686	80	6	2	1	29	0
13	4752	81	6	3	1	32	2
14	5920	97	4	6	1	34	0
15	6048	99	6	3	1	36	2
16	6289	103	1	5	1	57	1
17	6454	106	4	2	1	31	3
18	6623	108	5	10	2	28	0
19	6911	114	5	2	1	29	3
20	8209	134	1	7	2	23	2
21	8572	147	4	9	2		
22	9021	153	3	2	1	33	0
23	9066	153	6	9	2	26	0
24	10449	168	3	11	2	17	1
25	10872	171	6	21	2	16	0
26	11777	180	5	10	2	33	3
27	12259	187	1	5	1	31	0

Extra Random Numbers							
1	4633	79	1	4	1	26	2
2	727	17	1	12	2		
3	9806	161	2	20	2		
4	12488	192	2	3	1		
5	1121	23	5	5	1		

Brian Sigurdson

Brian Sigurdson							
Pilot Sample Random Numbers							
Item #	Block #	Stack #	Row	Column	Side Block Located	total	bad
1	112	8	4	7	2	24	1
2	344	11	2	10	2	36	0
3	429	12	3	12	2	32	3
4	767	18	5	6	1	22	1
5	1196	24	2	6	1	26	0
6	1737	31	3	12	2	30	1
7	2323	40	1	9	2	15	0
8	3179	55	5	1	1	29	0
9	3187	55	1	3	1	29	1
10	3586	61	4	9	2	26	0
11	5036	85	2	3	1	31	0
12	5127	86	3	6	1	30	2
13	5178	87	6	2	1	30	0
14	5537	92	5	2	1	29	0
15	5904	97	6	3	1	25	0
16	6710	110	2	7	2	23	0
17	7364	121	2	8	2	24	0
18	7401	122	3	4	1	24	0
19	7908	130	6	4	1	33	1
20	8626	148	4	4	1	empty	
21	9202	155	4	11	2	27	0
22	10040	164	2	1	1	18	0
23	10178	165	2	4	1	24	1
24	10184	165	2	5	1	28	0
25	10589	169	5	18	2	25	0
26	11078	173	2	16	2	31	1
27	11525	177	5	18	2	26	0

Extra Random Numbers							
1	4305	73	3	9	2	25	0
2	5398	90	4	3	1		
3	465	13	3	6	1		
4	6153	100	3	9	2		
5	4764	81	6	5	1		