

Group 4: Joel, Xiaoyi, Emily, Zhou, and Matt
STAT 8801
Rendahl
27 Jan 2012

Case Study 1: Forest Survey

Our client needs guidance on how to organize crew assignments for conducting a forest survey. This is the ideal situation from a statistical consulting perspective, as the data has not yet been collected, so any potential problems can be addressed before it is too late. Whenever data is being collected for analysis, statistical expertise is appropriate. In particular, when issues of randomization are involved, it is not a bad idea to check with the experts.

Before getting into the details of how to determine the crew assignments, we thought of several questions for the client. From the perspective of the forestry folks, mixing up the crews makes for happy campers. From our standpoint as statisticians we are more concerned with what is involved in the data collection. What information, exactly, are they collecting? Why is the data collection so strenuous? Is the expectation that most of the variability in data will be due to environmental factors as opposed to crew characteristics, or should we anticipate that certain crew members will be significantly more adept at gathering data than others, and therefore certain crew combinations might consistently out-gather the rest? From a whimsically philosophical angle, if a tree grows in the forest, but no one is there to count it, does it count?

The answers to the above questions will largely determine how concerned we need to be regarding the crew assignments. If we learn that there will not really be any difference among the crews in terms of data collection, then the focus should be on keeping the crews happy as explained by the client. In other words, if we could hypothetically send two different crews to the same location for the same week, would they essentially return with the same data? If the answer is an unequivocal “yes”, then the statistical side of this problem is probably not all that important. We can simply provide some randomization tools or ideas and leave behind worries of whether or not Tony and Julia work in the same crew together six times while Louise and Carlos only work together twice, and other such imbalances that would most likely occur without painstaking experimental design methods.

On the other hand, if there is reason to suspect the possibility that some crews will outperform others significantly, we might want to spend a bit more time figuring out how to make the assignments. Perhaps there are several crew members known for their amazing data collecting prowess such that we want to make an effort to ensure they are not always on the same crews. Perhaps there are other concerns that might arise after finding out more about the data collection process. Absent serious issues, randomizing the crew assignments (after assigning data recorders randomly) should work fine to balance out these contingencies.

First, we can assign each employee a number from one to twenty-one. Then we can randomly generate samples in order to assign data recorders. The following R output demonstrates how this could be done:

```
> sample(21) [1] >13 20 1 17 14 19 12 7 9 8 2 10 16 3 18 5 15 6 11 21 4  
> sample(21,6) [1] >10 16 21 5 12 4
```

After assigning recorders, we could randomly assign the rest of the people to the three crews through a similar number generation.

Alternatively, we could just generate completely random assignments and then make sure no one is recorder more than two times, as in the following table. The numbers below in each column are completely (pseudo-)random permutations of the numbers one to twenty-one. For each week the first person in each crew was designated recorder (red) unless they had already done it. Only at the very end did this scheme not work, where number nine is assigned twice before number one is once, but overall it worked out pretty well (and one and nine could easily be switched to amend this).

		Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Wk8	Wk9
Crew	1	21	21	8	2	9	3	12	18	16
Crew	1	8	16	7	6	15	2	18	21	14
Crew	1	17	8	5	9	11	6	1	1	12
Crew	1	20	6	14	8	3	19	17	10	3
Crew	1	6	9	11	21	8	13	20	17	4
Crew	1	3	7	9	1	7	9	3	19	1
Crew	1	13	13	3	16	13	1	14	11	21
Crew	2	14	15	10	7	4	7	16	20	18
Crew	2	10	2	19	17	17	17	21	8	20
Crew	2	2	14	17	14	19	20	6	9	6
Crew	2	12	12	18	19	16	14	19	16	17
Crew	2	1	17	16	20	2	8	11	12	15
Crew	2	7	10	13	18	18	16	2	4	11
Crew	2	15	18	21	13	1	21	13	3	9
Crew	3	11	5	20	12	20	12	9	14	7
Crew	3	5	3	12	3	12	15	8	2	10
Crew	3	4	11	4	5	6	4	4	7	2
Crew	3	18	19	6	15	21	18	10	13	8
Crew	3	19	4	15	4	10	11	15	6	5
Crew	3	9	1	2	10	5	10	5	15	13
Crew	3	16	20	1	11	14	5	7	5	19

In the wilderness, the crews could get their assignments randomly using playing cards or drawing straws or numbers from a hat, but if they are all going to be in the same place anyway, it would save a lot of time to have the assignments predetermined from a chart like the one above. If it is only a matter of choosing the data recorder in the wilderness after the crews have already been assigned, then having the people who have not been recorder draw straws would work. Once they reach the seventh week or so, and everyone has been recorder once in a particular crew, then everyone who has not been recorder twice draws straws. The chance that they would end up with a group in the last week made up of seven people who have all be data recorders twice is zero (since only six people will have to serve twice), so this method would work.