Permutation Test Results

1. Neighborhood/District Permutation Test:
   - Purpose: To test whether or not respondent had a tendency to recruit respondents in the same neighborhood/district as them.
   - Steps:
     a. Test statistic: The number of times a respondent recruited another respondent from his own neighborhood/district, directly from dataset.
        - For Neighborhoods: Test statistic = 6, which means that there were 6 instances of respondents recruiting from their own neighborhood in the dataset.
        - For Districts: Test statistic = 17, which means that there were 17 instances of respondents recruiting from respondents from their own district in the dataset.
   - 2.) Simulation: For the simulation I generated a random sample of neighborhood/district recruitment ties for 10000 trials from the dataset and test statistic for each trial:
      - After that I generated the frequency of each test statistic in the simulation and created a histogram of them, with the red line as the test statistic.
      - From the frequencies I got the p-value:
        - For Neighborhoods: p-value = sum(sim_stats)/10000 = 0.5655
        - For Districts: p-value = sum(sim_stats)/117/10000 = 0.5751
      - Both of these p-values indicate that there wasn’t a significant preferential recruitment for respondents who lived in the neighborhood/district.
   - Conclusion: The results of all the permutations indicate that there wasn’t any significant preferential recruitment in the dataset. These results make sense since there wasn’t a dominant recruitment relation amongst the neighborhoods. Even though the test statistic for districts was higher for the neighborhoods, it still yielded a high p-value, which indicates that even the district recruitment relation weren’t stronger than neighborhood recruitment relations. Even though the neighborhood distance test had the lowest p-value, it wasn’t significant enough to show that there was a strong preference for recruiting respondents that lived close to a respondent’s neighborhood. But nonetheless, comparing the p-values it is seen that respondents had a higher preference recruiting respondents that lived close to them rather than in the same neighborhood/district as them. These results aren’t surprising since more half the data was missing, therefore any conclusions about the recruitment relation would have been tentative.

2. Neighborhood Distance Permutation Test:
   - Purpose: To test whether or not respondents had a tendency to recruit respondents that lived close to their neighborhoods.
   - Steps:
     a. Created a distance matrix whose entries were the distance between all the neighborhoods of Kiev, regardless whether or not they had a recruitment tie in the dataset.
     b. Test statistic: The sum of all the distances between all the recruitment pairs in the dataset.
        - Test statistic = 901.53 km, which is the sum of all distances of the recruitment ties.
     - Simulation: For the simulation I generated a random sample of neighborhood recruitment ties for 10000 trials from the dataset:
       - From the simulated ties I used the test statistic function to sum up the distances of the simulated ties(sim_stats):
         - The simulated ties got their distance from the distance matrix and I added up these distances to get the test statistic for each trial.
       - After that I generated the frequency of each test statistic in the simulation and created a histogram of them.
     - From the frequencies I got the p-value: p-value = sum(sim_stats<901.53/10000 which was 0.092, which means that there wasn’t a significant preferential recruitment for respondents who lived near each other.

Conclusions

The results of all the permutations indicate that there wasn’t any significant preferential recruitment in the dataset. These results make sense since there wasn’t a dominant recruitment relation amongst the neighborhoods. Even though the test statistic for districts was higher for the neighborhoods, it still yielded a high p-value, which indicates that even the district recruitment relation weren’t stronger than neighborhood recruitment relations. Even though the neighborhood distance test had the lowest p-value, it wasn’t significant enough to show that there was a strong preference for recruiting respondents that lived close to a respondent’s neighborhood. But nonetheless, comparing the p-values it is seen that respondents had a higher preference recruiting respondents that lived close to them rather than in the same neighborhood/district as them. These results aren’t surprising since more half the data was missing, therefore any conclusions about the recruitment relation would have been tentative.

Acknowledgments:

Thank you to my advisor Professor Krista Gile for helping me and providing me with guidance through my project I would also like to thank Lisa Johnston for providing me the RDS dataset for Kiev. And Miles Ott(Brown) for his useful comments and suggestions for my permutation tests. As well as Isabelle Beaudry(UMass Amherst) and Kerry Ann Spitzer(MIT), for their feedback on my poster. Finally I would like to thank Mrs. Joanna Barkafeal for her generous donations for my REU program.