The Refugee Center Website Traffic

Huafeng Zhang
An organization to help refugees from all over the world to start their lives in the U.S.

Education, health, career, rights & law, culture, community

needs maintenances (offline).
Set a day or days without conflicting too many of their users
Q #1 Is there a significant difference in the mean # of visits on different days of the week?

Q #2 If so, which day or days of the week have fewest visits?
Sampling scheme? Sampling size? pro? GA R V d c
Stratified random sampling?
Stratified Random Sampling:
1. Focus on Important Subpopulations;
2. Ensures that the estimates can be made with equal accuracy and statistical power in different groups;
3. More manageable because I can use different techniques for different subpopulations.

SRS:
Doesn’t provide Subsamples of the population.

Quota Sampling:
The samples are biased because not all of the observations have a chance of selection.
d = 0.02*yu = 5.62

The refugee center staff told me that their desired precision is that the point estimate of mean sessions of a day in a week is within 2% of the mean sessions in a day.

Variance: \( S^2 = 4756.77 \)

Since I can request the population data,

{\begin{align*}
&\text{Population size} N=183, \text{Population mean } y_u =281.2, \alpha=0.05. \\
&\text{Sample size } n = 139.97 \quad (140) \\
&n = \frac{n_0}{1 + (n_0/N)} \quad \text{where } n_0 = (Z_{\alpha/2}S/d)^2
\end{align*}}
Sampling Process

- **Mon**
  - N1=20

- **Tue**
  - N2=20

- **Wed**
  - N3=20

- **Thu**
  - N4=20

- **Fri**
  - N5=20

- **Sat**
  - N6=20

- **Sun**
  - N7=20

Sample
n=140
Boxplots for the Raw Sample Data
Diagnostic Plots for the One-way ANOVA Test
Tukey Honest Significant Difference Test

95% family-wise confidence level

Differences in mean levels of day
<table>
<thead>
<tr>
<th>Day</th>
<th>Diff</th>
<th>Lwr</th>
<th>Upr</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat/Fri</td>
<td>0.81</td>
<td>0.69</td>
<td>0.95</td>
<td>0.003</td>
</tr>
<tr>
<td>Sat/Sun</td>
<td>0.91</td>
<td>0.77</td>
<td>1.07</td>
<td>0.58</td>
</tr>
<tr>
<td>Fri/Sun</td>
<td>1.12</td>
<td>0.95</td>
<td>1.32</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Conclusions

- Don’t have access to the data for the entire population
- The data are expensive to collect

Scope of Inference

- Cannot infer causal relationship between day & sessions
- Can make inference of the results to the population

Nontrivial Sampling Process

- Don’t have access to the data for the entire population
- The data are expensive to collect
Thank you!