K

We decide to use a stratified random sampling over a 2-week period.

We will stratify the observation time into 4 periods each day:

7am-10am, 10am-1pm, 1pm-4pm, 4pm-7pm

We will sample in 1hr period from these strata.

Reason:

A SRS of hour or half hour periods would increase the likelihood of the time of day negatively impacting the validity of our results, for this reason stratification allows us to compensate for certain problems (like rush hour) and ensure the robustness of our results.

L

Observational protocol:

Weather condition (sunny, windy, rainy, snowy,cloudy…)

Light/dark level (Day, Night, Dawn/Dusk

Road condition (dry, wet, covered with ice…)

Temperature (at beginning of measuring period as recorded by weather.com)

Inbound/Outbound

Date

Day of the week

bus number/route

When the bus is supposed to leave the bus stop

When the bus actually leaves the bus stop

Level of lateness (value of the difference between the scheduled and actual departure time)

Show/No show status (No show is defined as being so late that it’s arrival time is within five minutes of the next scheduled bus of that type)

Highly unusual conditions (these will be written down and are meant to include notable and unusual circumstances like a broken water main, major traffic accident, etc. This may be used to throw data out as an outlier).

Special events and other planned conditions.

Rush hour (yes/no)

Clustering (number of other buses of the same number that arrive within 2 minutes of each other).

M

We set ME= 0.05

SD=5 min

(From the selective research, 5min seems to be a good starting point for the standard deviation of the bus lateness. )

Z95%=1.96

N=12(hrs/day)\*7(days/wk)\*2(wk)\*60(min/hr)=168 \* 60 = 10080 min

n0=Z^2\*SD^2/ME^2=1.96^2\* 5^2/ 0.05= 1920.8 min

n>= Nn0/(N+n0) = 10080\*1920.8 / (10080+1920.8) = 1613.364 min = 26.88941 hrs

So we need to sample about 27 hours out of 168 hours over the 2-week period.