Accuracy of bus schedules

---Team E
Tian Wu, Tim Higgins, Matt Belenky, Chao Wang, John Sperger

Introduction

What are we interested in?:

1. How accurate are the bus schedules in Pittsurgh?
Do buses always show up?
Are the schedules useful?

2. How punctual are the Pittsburgh buses?
Can we quantify this information?

Research questions:

Can we identify what factors affect the punctuality of buses?

Motivations

1. Make people's life easier:
How late can you expect a bus to be?
Is going to the bus stop early ever useful?
How often do you have three 61's show up at the same time?

2. Potentially make suggestions to PAT to improve the efficiency of the bus system.

3. Inform students.

Sample Selection

Strata based sampling

We knew that we would need 8 hours of observation from each of our four strata as a prerequisite for significant results

Using an excel spreadsheet all of the possible hours of observation in one stratum during our observation period were entered and assigned a number

Using a random number generator from random.org 8 random numbers were generated and the hours of observation corresponding to those numbers were added to the sample

This process was repeated for each of the stratum in our experiment

Glitches

When we were selecting a sample we assumed that with five group members we would have someone available who could collect data during any given hour

This turned out not to be the case - some hours nobody could make. Solution - resample the hours that nobody could make randomly. This is still only pseudo random but better than pure convenience

Observation errors - some buses will sit and park for a long time. We are measuring the departure time of buses, but if a number of buses are departing at the same time it is easy to not notice the parked bus leaving

What we have done so far...

We randomized the times for observing the buses's so that it's more accurate

Created a table that assigns which times to observe the buses's and with which characteristics

Assigned each person in the group 5 different times (consisting of 1 hour each) to observe the buses's

How successful we've been...

We've been able to observe 2 bus stops simultaneously with one person

Been successfully keeping up with the bus schedule regardless of the weather

The randomized times and the table for what to observe, have both been extremely helpful and easy to follow

Variables measured

Definition of “Being on time for a bus” ([-2.5, 2.5])
Day of the week
Road condition
Weather
Rush hour or not
Bus route number
Inbound / outbound
Scheduled depart time
Actual depart time
Clustering or not
Special Situations (special events, breakdown)
etc.

Analysis & the use of it

Strength of stratification
-More representative sample
-Easier to generalize

Regression Analysis
-Find out main factors that affect the punctuality of buses
-Find the best regression model for all periods, and models for each stratum
-Compare models for different time periods
-Other investigations (What affects punctuality on Monday?)

Give out the our own prediction of arriving time using our best regression models

What's left

1. What's left to do:
   a) Need more sampling
   b) Post-sampling processing
      Time periods
      Weather (Sunny, Rainy, Extreme)
      Bus Lines (which line arrives more precisely)

2. Cut-off Date: March 27
   Since the Port Authority Transportation is changing the schedule

3. Issued Raised:
   Less meaningful of our research

Potential Improvement in the Future

1. More sampling through the year

2. More technology and efficiency (Use video taping, etc.)

3. More factors
   i.e. Gender, Years of Driving, Payment of the drivers

4. Conduct a survey about how long are people willing to wait for a bus rather than take alternatives. If we need improvements in public transportation, such as subway