100/100

(a) see comments below

(b) see attached sheets on your sampling plans

Team E I.3
36-303
Project Proposals
Tian Wu, Tim Higgins, Matt Belenky, Chao Wang, John Sperger

Final Proposal Choice
Proposal 2—Accuracy of Pittsburgh Buses around Carnegie Mellon

A.
Why is the topic interesting?

Working efficiently and saving as much time as possible is key to raising one's quality of life, especially to students and teachers of Carnegie Mellon, one of the most intense universities in the world. Many students and teachers are late for classes or important appointments due to the tardiness of public transportation. Waiting for a bus for too long not only disturbs people's plans, but also negatively affects people's health and mood. After measuring the actual lateness of buses we could potentially also attempt to measure individual perception of bus lateness.

B.
What questions do you propose to study?

a) Definition of "Being on time for a bus" ([-2.5, 2.5])
b) How many buses are on time.
c) How much clustering are buses per day? (Three 61's in a row)
d) Factors causing the lateness
    i) Weather
    ii) Special Events (such as Super Bowl etc.)
    iii) Rush Hour (both morning and evening)
    iv) Traffic technical problems (broke down for a bus)
    v) Bus route
e) Suggestions and solutions

C.
What previous research has been done?
Chao Wang

these are fine but I'd like to see them formatted more like a formal list of references or annotated bibliography at the end of a paper.

http://www.nctr.usf.edu/jpt/pdf/JPT%209-2%20Mishalani.pdf
i) This survey is about the passengers wait time perceptions at bus stops. It helps us to clarify the definition of "a late bus"
ii) The survey gives us insight on making suggestions and solutions to deal with lateness of buses.
Author:
Rabi G. Mishalani and Mark M. McCord, Ohio State University
John Wirtz, Edwards and Kelcey, Inc., Chicago, Illinois

Tim:
This study is an analysis of how useful more complete information on public transportation would be. That is, it provides information on how useful our study would be. In particular it shows that without some sort of increased information, individuals are extremely bad at

determining how early or late they should show up for their public transit of choice.
Journal of Public Transportation Volume 12, No. 3, 2009
http://www.nctr.usf.edu/jpt/pdf/JPT12-3Caulfield.pdf

Tian
The UK Deparment of Transport's report publishes on the punctuality of the bus system. It investigates the reason for the delay and suggests what entities and partners involved in the public transportation should do to improve the service.
http://www.dft.gov.uk/pgr/regional/buses/buspunctuality/

Matt
Author Name: Dan Weikel
Date: August 29, 2009
Title: "MTA gets low marks for bus service in survey"
Website (LA times): http://articles.latimes.com/2009/aug/29/local/me-bus-riders29
Group Member: Matt Belenky
Summary: The article mentions how people riding the MTA buses in the LA area were very upset with their overall service, specifically on-time performance, overcrowding, and a shortage of service. Riders gave the agency a D in overall service and people were paying too much and waiting too long for terrible service. We aren't necessarily investigating what the people think of the service but we are researching the specific complaints these LA residents mention: on-time performance, what factors affect buses coming late, is there a shortage in service. In short we are taking the complaints used in LA and seeing how accurate this data is quantitatively.

John: This study investigated how passengers of public transit buses perceived the reliability of buses. Because the study compared subjective measures to outside information, the study provides useful information for what factors of actual reliability are most responsible for increasing user utility.
Daskalakis, Nikolaos G. and Anthony Stathopoulos "Users' Perceptive Evaluation of Bus Arrival Time Deviations in Stochastic Networks*" Journal of Public Transportation, Vol. 11, No. 4*, 2008
http://www.nctr.usf.edu/jpt/pdf/JPT11-4Daskalakis.pdf

D.

The sampling frame will be the list of port authority buses that CMU students use to get to and from school. The sample population will be all the buses that stop at the Forbes and Morewood bus stops. These two stops are adequate for our purposes for several reasons. First, the other Forbes stops (such as Hamburg and Beeler) are very close to the Forbes and Morewood stops. If the bus is late at Hamburg, then it will likely be just as late 100 feet down the road at Forbes and Morewood. Second, the other stops close to campus (such as Fifth and Morewood, and the Wilkins stops) are used by a substantially smaller percentage of the student population and mostly for the purposes of shuttling to non residential locations. The areas in Shadyside in which students live are well serviced by the faster and more reliable loop buses. Buses moving up and down fifth will also be independent from buses moving up and down Forbes because many travel through different neighborhoods and thus face different traffic patterns. This would lead to trouble in data analysis because the independent routes could lead to a bimodal

See attached sheet on sampling for this project

distribution, hurt the accuracy of any inference we would like to make about how late buses usually are, or increase the number of man hours beyond what is feasible for our group.

Easy to fix by doing separate analyses at different stops -- see also attached sheet on sampling

Include realistic estimates of time if you want to make this argument

E.

The target population is all of the buses which stop on campus that CMU students use to get to school. The target population is differs only slightly from the sampling frame for this study because port authority bus routes are set and all of the buses that stop outside of Hamburg Hall or across the street from it stop at the Forbes/Morewood bus stop. Still, there will be some coverage error because some students take buses from the 5th/Morewood bus stops which have different routes and don't stop at Forbes/Morewood.

A possible source of sampling error for this study comes from choosing to randomly select hours to watch buses (e.g. 8-9AM) rather than randomly selecting individual buses (e.g. the 10:25PM 28X). While the hours will be randomly selected, this still means that the sampling method will be a combination of convenience (hour long blocks) and random selection. However, using individual buses as the sampling unit would be completely unfeasible given the manpower we have.

A potentially serious source of non-sampling bias is that the study will be conducted over a relatively brief period (2 weeks) in the winter. Seasonal differences in route deviations are likely due to issues such as weather that vary by season. In addition, special events in Pittsburgh could cause excessive delays that are not typical (e.g. post super bowl riots).

Bias due to seasonal differences – by measuring the weather each day during our sample we hope to control for the differences in delays due to weather. Using data about the frequency of inclement weather in Pittsburgh during other seasons the model we develop could theoretically account for seasonal differences. Alternatively, a follow up study could be run in subsequent months.

Error due to special events – if the events cause serious outliers we can just remove them from the data set as a form of post survey data adjustment.

F.

Our mode of data collection will be observational, we intend to observe the buses and note which buses arrives and when. This can then be used to determine if the buses are arriving on time by comparing it to a time table. For this proposal, an observational method is effectively the only usable one, but it is also the best since you cannot survey buses, etc. In addition to determine which buses to observe we will use the following method. We will assign every hourly increment in a period spanning approximately two weeks a numerical value. Using a random number generator will select a number of those hours. These will be the ones during which we take our measurements.

G.

Our primary variable will be deviation from the listed arrival time in minutes. Secondary variables will be weather, time of day, rush hour (yes or no) and route. We will also observe other conditions and may include them as variables.