

9 Cases

9.1 Forecasting Public Transportation Demand

Background

Forecasting transportation demand is important for multiple purposes such as staffing, planning, and inventory control. The public transport system in Santiago de Chile has gone through a major effort of reconstruction. In this context, a business intelligence competition took place in October 2006, which focused on forecasting demand for public transportation. This case is based on the competition, with some modifications.



Figure 9.1: Vicente Valdés metro station, Santiago de Chile

Problem Description

A public transportation company is expecting increased demand for its services and is planning to acquire new buses and extend its terminals. These investments require a reliable forecast of future demand. To create such forecasts, one can use data on historic demand. The company's data warehouse has data on each 15-minute interval between 6:30 and 22:00, on the number of passengers arriving at the terminal. As a forecasting consultant, you have been asked to create a forecasting method that can generate forecasts for the number of passengers arriving at the terminal.

Available Data

Part of the historic information is available in the file *bicup2006.xls*. The file contains the worksheet "Historic Information" with known demand for a 3-week period, separated into 15-minute intervals. The second worksheet ("Future") contains dates and times for a future 3-day period, for which forecasts should be generated (as part of the 2006 competition).

Assignment Goal

Your goal is to create a model/method that produces accurate forecasts. To evaluate your accuracy, partition the given historic data into two periods: a training period (the first two weeks) and a validation period (the last week). Models should be fitted only to the training data and evaluated on the validation data.

Although the competition winning criterion was the lowest Mean Absolute Error (MAE) on the future 3-day data, this is not the goal for this assignment. Instead, if we consider a more realistic business context, our goal is to create a model that generates reasonably good forecasts on any time/day of the week. Consider not only predictive metrics such as MAE, MAPE, and RMSE, but also look at actual and forecasted values, overlaid on a time plot.