SAS EM NODES

TORONTO DATA MINING USER GROUP

§sas
THE POWER TO KNOW.

AGENDA | **WHAT'S NEW IN ENTERPRISE MINER**

- Survival Data Mining
- Incremental Response Modelling
- Time Series Similarity Analysis

§sas | THE POWER TO KNOW.

SURVIVAL DATA MINING

SURVIVAL DATA MINING ANALYSIS

PREDICTION OF WHEN AN EVENT WILL HAPPEN AND NOT JUST IF IT WILL HAPPEN VERY POPULAR FOR CUSTOMER BEHAVIOR MODELLING, SUCH AS

- When will a customer churn?
- When will a customer cancel a service or product?

PREDICTS EVENT PROBABILITY FOR TIME INTERVALS FOR EACH CUSTOMER

- i.e. Customer has 50% chance to cancel next month but 75% chance to cancel the month afterward.
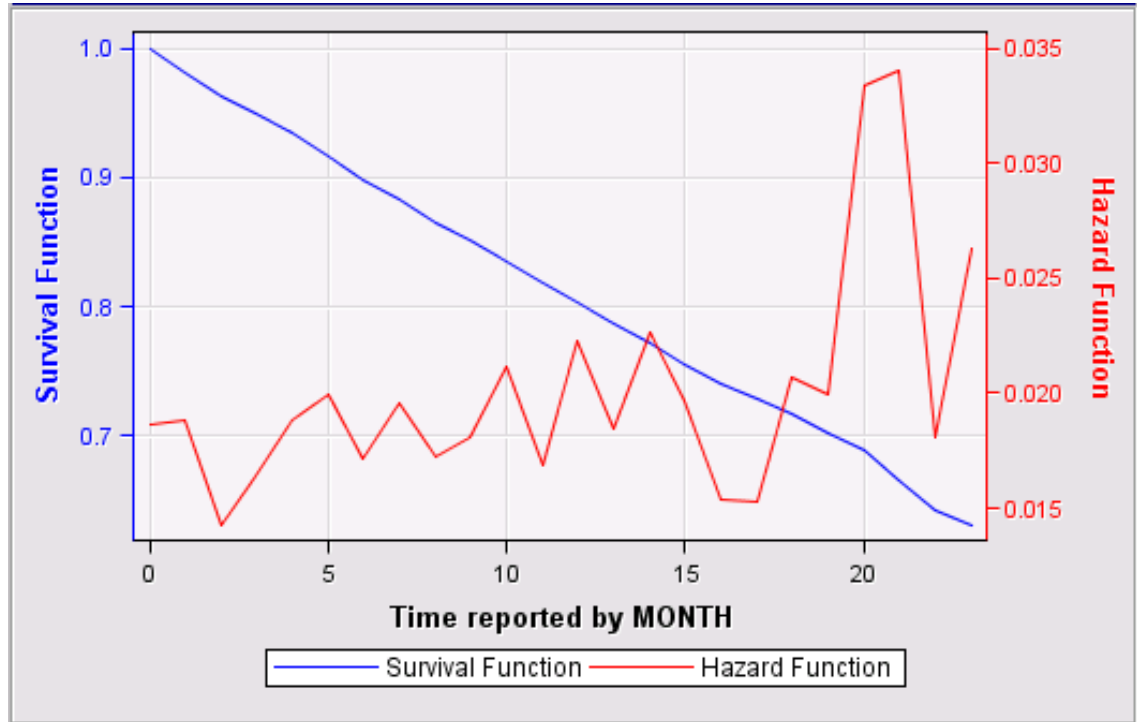
CAN TAKE EXTERNAL FACTORS INTO ACCOUNT

- i.e. customers with more than 2 products tend to stay longer

§sas | THE POWER TO KNOW.

SURVIVAL DATA MINING ANALYSIS

APPROACH

- Look at probability of hazard (event) at discrete time points
- Compared to survival probability (tenure of customers overt time)

- Mean Residual Lifetime: expected time till event occurs based on projecting hazard function into the future
  - Projection based on constant hazard function
  - Projection based on continuing trend of hazard function

INCREMENTAL RESPONSE MODELLING

§sas
THE POWER TO KNOW®

Target:
customers who will buy the product if and only if it is promoted directly to them

Common Issues:

- Negative effects by annoying customers who would be less likely to respond if targeted (**Do Not Disturb**)
- **Waste on** customers who would have responded regardless of the incentive
- **Waste on** customers who would NOT respond whether they are targeted or not

- Traditional response modeling
  - Identifies customers who are likely to respond
  - Targets <u>all</u> the respondents
  - Still wastes resources on customers who will respond no matter what

- Incremental response (IR) modeling
  - Identifies the additional purchases that would not have taken place without the campaign
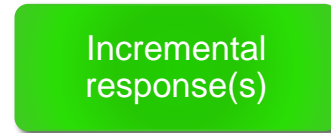 - Measures the true effect of a marketing campaign

Responses

Incremental response(s)

§sas | THE POWER TO KNOW.
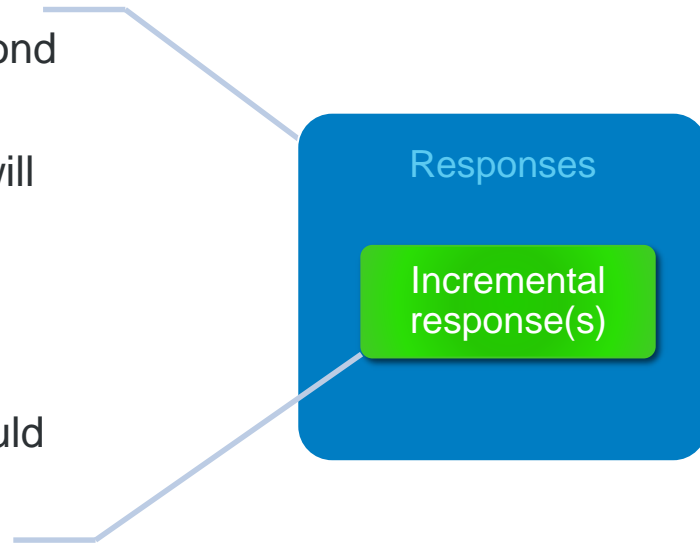
MODELING DETAILS

- True-lift / Net-lift / Uplift Modeling
- Target only customers who produce incremental responses
- **Treatment** (promotion) group: Assume that incremental responses exist, but do not know which observations are incremental responses
- **Control** (no offer) group: No incremental response

Treatment	Control
Response = Yes	Response = Yes
Incremental response(s)	
Response = No	Response = No
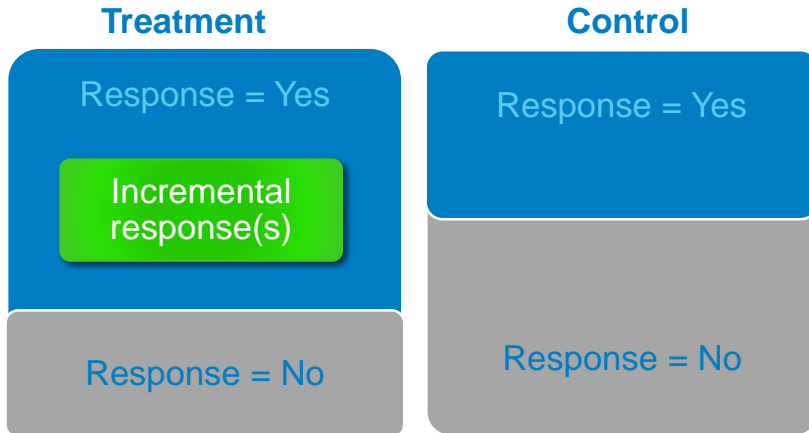
- Build predictive model to identify

Incremental response(s)

§sas | THE POWER TO KNOW.

TIME SERIES SIMILARITY

TIME SERIES SIMILARITY

TRADITIONALLY, DATA MINING AND TIME SERIES ANALYSIS HAVE BEEN SEEN AS SEPARATE APPROACHES TO ANALYZING ENTERPRISE DATA.

HOWEVER, MUCH OF THE DATA GENERATED BY BUSINESS PROCESSES IS TIME-STAMPED.

TIME SERIES DATA MINING IS A MARRIAGE OF FORECASTING AND TRADITIONAL DATA MINING TECHNIQUES THAT USES TIME DIMENSIONS AND PREDICTIVE ANALYTICS TO MAKE BETTER BUSINESS DECISIONS.

§sas | THE POWER TO KNOW.
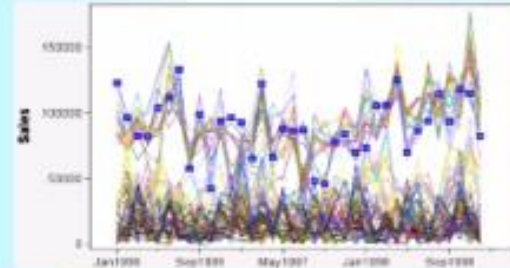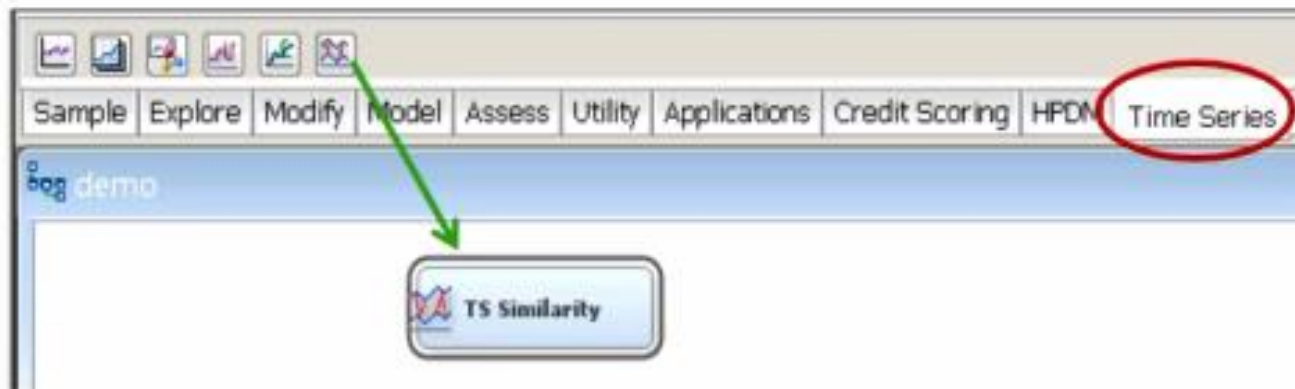
SIMILARITY SEARCH OR INDEXING

Given a measure and a target (query) time series:

- find the most similar or dissimilar time series
- index the time series database

Some real-world examples: Stock price movement, sales patterns, medical device data

ENTERPRISE MINER NODE

- Provides similarity search and clustering
  - Similarity search: when target time series are specified
  - Clustering: no target and all time series are inputs

USEFUL LINKS

- Videos: http://support.sas.com/rnd/app/video/index.html#txtmine
  - Time Series Similarity Analysis
  - Incremental Response
  - Survival Data Mining
- Papers:
  - Time Series Data Mining with SAS® Enterprise Miner™
  - It's About Time: Discrete Time Survival Analysis Using SAS® Enterprise Miner™