"Prediction is very difficult, especially if it's about the future."
--Nils Bohr, Nobel laureate in Physics
Training Objective

• Data Analytics is now becoming the game changer in the industry and it is much more relevant for supply chain functions. Predictive Analytics is an advanced Data Analytics that leverages historical data and combines it with forecasting models to predict future outcomes. It can be used across the entire gamut of supply chain such as Plan, Source, Make, Deliver and Return.

• In this session, we will focus on the what, where and how predictive analytics can be used. Some of the techniques covered in this session include: Regression, Time Series Forecasting, Basic Machine Learning Algorithms (Clustering, Cart) and Decision Trees. Through use cases, we will demonstrate application of predictive analytics in supply chain functions.
What if you were able to....

• Predict the outcome of your new product launch?

• Evaluate the impact of your marketing campaigns hourly and make inventory adjustments in real-time?

• Predict potential failure of your critical equipment and plan contingencies?

• Predict the market sentiments and buying behavior of your consumers in real-time?

• Predict the outcome of the US Presidential elections?
Data to Insights: Driving Industry 4.0

Analytics Value Chain

BY 2020

- Average Internet User: 1.5 GB of Traffic / Day
- Autonomous Vehicles: 4 TB of Data / Day
- Connected Airplane: 5 TB of Data / Day
- Smart Factory: 1 PB of Data / Day
- Cloud Video Providers: 750 PB of Video / Day

The Coming Flood of Data in Autonomous Vehicles

- Radar: ~10-100 KB per Second
- Sonar: ~10-100 KB per Second
- GPS: ~50 KB per Second
- Cameras: ~20-40 MB per Second
- Lidar: ~10-70 MB per Second

Domain Knowledge

Soft Skills

Data Scientist

Analytics Skills

Software Skills
Supply Chain Analytics

*Supply Chain Management involves coordination and synchronization of business process, data and decisions among several functions*
Predictive Modeling

- Learn the model based on historical data
- Correlation is used to understand strength of linear relationship. Beware causation versus correlation
- Predictive techniques ($y=f(x)$) such as Regression, Time Series Forecasting, Machine learning, Decision Tree are used to build models

Terms such as data mining and machine learning are often used
Predictive Model Algorithms (y=f(x))

Supervised Learning

- Y = Numeric
  - Regression
    - Linear Regression
      - Best Subsets
      - Stepwise
        - Lasso
        - Ridge
        - Time Series Forecasting
  - Linear Discriminant Analysis

- Y = Categorical
  - Classification
    - Logistic Regression
    - Bayesian Networks
  - Decision Trees
    - Random Forests

- Regression or Classification
  - Support Vector Machines
  - Neural Networks
  - Gradient Boosting
  - K Nearest Neighbors

Unsupervised Learning

- No Y
  - Clustering
    - Hierarchical
    - K Means

- Time Series Forecasting

= we will look at these today
Better Prediction $\Rightarrow$ Better Model!

Three models:
- Under fit: bias
- Over fit: variance
- Just right!

Balance accuracy of fit on training data and generalization of future or test data to get a better model!
Predictive Modeling

• Regression Analysis

• Time Series Forecasting

• Machine Learning & Decision Trees
Regression and Correlation

• Regression analysis is a tool for building statistical models that characterize relationships between a dependent variable and one or more independent variables, all of which are numerical.
  – A regression model that involves a single independent variable is called simple regression. A regression model that involves several independent variables is called multiple regression.

• Correlation is a measure of a linear relationship between two variables, X and Y, and is measured by the (population) correlation coefficient. Correlation coefficients will range from −1 to +1. Scatter plot evaluates the relationship between two variables.
Linear Regression

- Find the line to minimize the squared error of the fit (least squares loss function)

\[
Y = \beta_0 + \beta_1 x
\]

Fit Y = $\beta_0 + \beta_1 x$

(Note - model can extend to any number of x’s)

Error of fit to this point $\Rightarrow$ This is called a residual

Predictive Model

\[
Y = 1 + 0.67x
\]
Watch out for Nonsense Correlations!

- Beware lurking variables!
- Correlation does not mean Causation
Simple Linear Regression

Example: Keith’s Auto Sales

Reed Auto periodically has a special week-long sale. As part of the advertising campaign Reed runs one or more television commercials during the weekend preceding the sale. Based on a sample of 5 previous sales, determine whether number of cars sold is influenced by number of TV Ads.

<table>
<thead>
<tr>
<th>Number of TV Ads (x)</th>
<th>Number of Cars Sold (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

\[ \bar{x} = \frac{\Sigma x}{5} = \frac{10}{5} = 2 \]
\[ \bar{y} = \frac{\Sigma y}{5} = \frac{100}{5} = 20 \]

See Excel output
Simple Linear Regression Excel Output

### Simple Linear Regression - Keith's Auto Sales

<table>
<thead>
<tr>
<th># of Ads</th>
<th>Cars sold</th>
<th>(x_i - x-bar)(y_i - y-bar)</th>
<th>(x_i - x-bar)^2</th>
<th>(y_i - y-bar)^2</th>
<th>(x_i - x-bar)(y_i - y-bar)^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>-1</td>
<td>-1</td>
<td>-6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>-1</td>
<td>-3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>49</td>
</tr>
</tbody>
</table>

### Using Excel Data Analysis Add-in

#### SUMMARY OUTPUT

- **Multiple R:** 0.9366
- **R Square:** 0.8772 (SSR/SST)
- **Adjusted R Sq.:** 0.8363
- **Standard Error:** 2.1602
- **Observations:** 5

#### Regression Statistics

- **R2** indicates the percentage of variation in Y that is explained by the linear function of the variable X.

#### ANOVA (either F test or t test can be used for testing model significance)

- **df:**
  - Regression: 1
  - Residual: 3
  - Total: 4

- **SS:**
  - Regression: 100
  - Residual: 14
  - Total: 114

- **MS:**
  - Regression: 100
  - Residual: 14

- **F:** 21.429

- **Significance F:** 0.018886

#### Equation:

- **Slope ($b_1$):** $5 = F9$G9
- **y intercept ($b_0$):** $10 = C10$C12$B10$

#### Est. Resogr. Eqn.: y = 10 + 5x

### Conclusion

- The correlation coefficient is $0.9366$, indicating a strong positive linear relationship.
- The $p$-value of 0.018886 is less than 0.05, allowing us to reject the null hypothesis $H_0: \beta = 0$. It is significant.

### Scattered Plot

- Scatter Plot shows a positive linear relationship between # of Ads and # of cars sold.

- Correlation Coefficient: $0.9366$
Example: Programmer Salary Survey

A software firm collected data for a sample of 20 computer programmers. A suggestion was made that regression analysis could be used to determine if salary was related to the years of experience and the score on the firm’s programmer aptitude test.

The years of experience, score on the aptitude test, and corresponding annual salary ($1000s) for a sample of 20 programmers is shown in the table.

<table>
<thead>
<tr>
<th>Exper.</th>
<th>Score</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>78</td>
<td>24.0</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>43.0</td>
</tr>
<tr>
<td>1</td>
<td>86</td>
<td>23.7</td>
</tr>
<tr>
<td>5</td>
<td>82</td>
<td>34.3</td>
</tr>
<tr>
<td>8</td>
<td>86</td>
<td>35.8</td>
</tr>
<tr>
<td>10</td>
<td>84</td>
<td>38.0</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>22.2</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>23.1</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
<td>30.0</td>
</tr>
<tr>
<td>6</td>
<td>91</td>
<td>33.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exper.</th>
<th>Score</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>88</td>
<td>38.0</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>26.6</td>
</tr>
<tr>
<td>10</td>
<td>75</td>
<td>36.2</td>
</tr>
<tr>
<td>5</td>
<td>81</td>
<td>31.6</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>29.0</td>
</tr>
<tr>
<td>8</td>
<td>87</td>
<td>34.0</td>
</tr>
<tr>
<td>4</td>
<td>79</td>
<td>30.1</td>
</tr>
<tr>
<td>6</td>
<td>94</td>
<td>33.9</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>28.2</td>
</tr>
<tr>
<td>3</td>
<td>89</td>
<td>30.0</td>
</tr>
</tbody>
</table>

See Excel output
Multiple Linear Regression Excel Output

**Summary Output**

- **R² (R-squared):** 0.9033
- **Adjusted R-squared:** 0.8967
- **Standard Error:** 2.4187
- **Observations:** 20

**ANOVA**

- **Df:** 2
- **SS:** 590.029
- **MS:** 295.014
- **F:** 42.7691
- **Significance:** 0.0000

**Multiple Regression Equation**

\[ \text{Salary} = 3.374 + 1.004 \times \text{Experience} + 0.251 \times \text{Score} \]

**Predicted Salary for given Experience & Score**

<table>
<thead>
<tr>
<th>Experience</th>
<th>Score</th>
<th>Predicted Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>95.34</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>94.60</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>42.16</td>
</tr>
</tbody>
</table>

**Extended Regression Eqn.**

\[ \text{Salary} = 3.174 + 1.004 \times \text{Experience} + 0.251 \times \text{Score} \]

**Residual Output**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted</th>
<th>Residual</th>
<th>Standard Error</th>
<th>Forecasted Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80.358</td>
<td>-4.338</td>
<td>1.675</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>80.975</td>
<td>-4.338</td>
<td>1.675</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>80.975</td>
<td>-4.338</td>
<td>1.675</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>80.975</td>
<td>-4.338</td>
<td>1.675</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>80.975</td>
<td>-4.338</td>
<td>1.675</td>
<td>2.5</td>
</tr>
</tbody>
</table>

**Residual Plots**

- Standardized Residual Plots
- Normal Probability Plot

M. Janakiram Oct 2017
Predictive Modeling

• Regression Analysis

• Time Series Forecasting

• Machine Learning & Decision Trees
Time Series Forecasting?

• Forecasting is an estimate of future performance based on past/current performance
• Qualitative and Quantitative technique can be used for forecasting
• Forecasting can be used to get insight into future performance for better planning and managing the future activities
• Forecasting examples:
  – Revenue forecasting
  – Product demand forecasting
  – Equipment purchase
  – Inventory planning, etc.
Forecasting Inputs

Sources:
- Internal:
  - Performance history
  - Experience
- External:
  - Economy
  - Market
  - Customers
  - Competition

Methods:
- Qualitative:
  - Judgment
  - Intuition
  - Expectations
- Quantitative:
  - Data
  - Statistical analysis
  - Models

Factors:
- Promotions
- Price changes
- Seasonal changes
- Trends / cycles
- Customer tastes
- Gov’t regulations

Forecast model should be better than Naïve models (which use Average or last period values)

→ Forecast !
Combination of qualitative and quantitative techniques can be used for better forecasting.
Qualitative Models

• Field Sales Force
  – Bottom-Up Aggregation of Salesmen’s Forecasts (subjective feel). Surveys are also used very often to get a feel for what the future holds.

• Jury of Executives
  – A forecast developed by combining the subjective opinions of the managers and executives who are most likely to have the best insights. We use it internally (called as Forecast Markets) for product forecast and transition.

• Users’ Expectations
  – Forecast based on users’ inputs (mostly subjective)

• Delphi Method
  – Similar to Jury of Jury of Executives but the process is recursive until consensus among the panel members is reached.
Quantitative Time Series Models

- Naïve Forecast
  - Simple Rules Based upon Prior Experience

- Trend
  - Linear, Exponential, S-Curve and Other Types of Pattern Projection

- Moving Average
  - Un-weighted Linear Combination of Past Actual Values

- Filters
  - Weighted Linear Combination of Past Actual Values

- Exponential Smoothing
  - Forecasts Obtained by Smoothing and Averaging Past Actual Values

- Decomposition
  - Time Series Broken Into Trend, Seasonality, Cyclicality, and Randomness

- Data mining
  - Mostly CART and Neural Network techniques are used for forecasting
Causal/Explanatory Models

- **Regression Models:**
  - Modeling the value of an output based on the values of one or more inputs. Also called causal models (can have multiple inputs).

- **Econometric Models:**
  - Simultaneous Systems of Multiple Regression Equations

- **Life Cycle Models:**
  - Diffusion-based approach, used to predict market size and duration based on product acceptance behavior

- **ARIMA Models:**
  - Class of time series models that take into consideration, various behavior of the response in question based on past time series data.

- **Transfer Function Models:**
  - Class of time series models that take into consideration, various behavior of the response along with causal factors based on past time series data.
Simple Moving Average (or UWMA)

- The Simple Moving Average, or Uniformly Weighted Moving Average (UWMA) for a time period \( t \), with a range span of \( k \), is described as:

\[
SMA_t = \sum_{i=1}^{k} x_{t-i} \quad \frac{k}{k}
\]

- The forecasted value for time \( t+1 \) becomes:

\[
SMA_{t+1} = \sum_{i=1}^{k} x_{(t+1)-i} \quad \frac{k}{k}
\]
Exponentially Weighted Moving Average (EWMA)

- In contrast to Simple Moving Average, Exponentially Weighted Moving Average (EWMA) for a time period $t$, uses all of previously observed data, instead of just a fixed range.

- Smoothed value for a time period, $t$, indicated by $Z_t$.

$$
Z_t = \alpha y_{t-1} + (1-\alpha) Z_{t-1} \quad \text{or} \\
Z_t = Z_{t-1} + \alpha (y_{t-1} - Z_{t-1})
$$

- Each previous observation receives different weight, with more recent observations being weighted more heavily, and weights decreasing exponentially.

- How these weights decrease determined by smoothing parameter, $\alpha$. Normally it ranges from 0.2 to 0.4.
  - When $\alpha$ approaches 0, all prior observations have equal weight
  - When $\alpha$ approaches 1, most weight is given to more recent observations

- Smoothing method should be done on stationary time series without seasonality.
Forecasting Scenario*

• Process: Sets of 10 coin flips

• What’s the naïve model?
• Can you beat it? If so, how?
Forecasting: Common Pitfalls

• Good-fitting model ≠ accurate forecast
  – Standard modeling methods can be (ab)used to **over-fit** the data

Fitting a model to history is easy – good forecasting is not. *(SAS Institute)*
Example: Regression vs. Time Series

Time series: Collection of observations made over equally spaced time intervals. E.g.: cost, inventory levels, shipment times, weekly forecast values.

**Time Series Example:** Monthly CO₂ concentrations from Mauna Loa captured for 14 years

- **Estimate CO₂ data for next 2 years**
- **Intuitively, we could do better than the simple linear regression, right?**

**Time Series Models can yield better predictive models!**
It takes advantage of correlation (s) between the current value and other responses that occurred previously in time.
Components of a Time Series

- The time series data can mainly have the following components:

  **Level**: Absence of trend, seasonality, or noise. Also, check for stability (Stationary vs. Non-stationary)

  **Trend**: Continuing pattern of increase or decrease, either straight-line or curve

  **Cyclical**: A repeating pattern of increases and decreases that occurs with a regular frequency

  **Seasonal**: A repeating pattern of increases and decreases that occurs with a regular frequency

  **Irregular**: Random variations not explained by other components

First plot your data chart to understand any behavior and relationships
Models to Account for Seasonality

- Think of creating a model to account for seasonality as an extension of the exponential smoothing.

- There are 3 terms to estimate in a seasonal model, the Level, the Trend, and the Seasonal components.
Sample Size Guidance

50-100 observations recommended.

More is better.

- 1 to 2 years of weekly data.
- 4 to 8 years of monthly data.
- 12 to 24 years of quarterly data.
Predictive Modeling

• Regression Analysis

• Time Series Forecasting

• Machine Learning & Decision Trees
Machine Learning (ML)/Data Mining (DM)

1. Set clear business goal
2. Get the right data
3. Mine the data
4. Determine what is interesting
5. Presentation

ML/DM follows the problem solving methodology

Pre-processing

Training data

Testing data

Model/Result

New data

Set clear business goal

Get the right data

Mine the data

Determine what is interesting

Presentation
Decision Trees (Classification)

- Finds the best factor split points to minimize the classification error

\[
Y = \begin{cases} 
1 & \text{if } x_1 < 4.5 \text{ and } x_2 < 3 \\
2 & \text{otherwise}
\end{cases}
\]

The model is a piecewise constant function.

Y is a 2 level categorical variable.
Iris data – Clustering Analysis

- Iris data first published by geneticist and statistician Ronald Fisher in 1936 is widely used in machine learning for benchmarking.
- The sepal length, sepal width, petal length, and petal width are measured in mm on fifty Iris specimens from each of three species:
  1. Iris Setosa
  2. Iris Versicolor
  3. Iris Virginica
- Machine learning goal is to classify the species based on the flower characteristics
Iris data

Y: 1=Setosa, 2=Versicolor & 3=Virginica;  X: Sepal Length, Sepal Width, Petal Length and Petal Width

Scatterplot Matrix

<table>
<thead>
<tr>
<th>Name</th>
<th>Data type</th>
<th>Modelling type</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepal length</td>
<td>Number</td>
<td>Numeric</td>
<td>Input</td>
</tr>
<tr>
<td>Sepal width</td>
<td>Number</td>
<td>Numeric</td>
<td>Input</td>
</tr>
<tr>
<td>Petal length</td>
<td>Number</td>
<td>Numeric</td>
<td>Input</td>
</tr>
<tr>
<td>Petal width</td>
<td>Number</td>
<td>Numeric</td>
<td>Input</td>
</tr>
<tr>
<td>Species</td>
<td>String</td>
<td>Categorical</td>
<td>Output</td>
</tr>
<tr>
<td>Species Number</td>
<td>Number</td>
<td>Numeric</td>
<td>Output</td>
</tr>
</tbody>
</table>
Tree building

Scatterplot Matrix

Species
- setosa
- versicolor
- virginica

Sepal length

Sepal width

Petal length

Petal width

setosa 50 33.3333%
versicolor 50 33.3333%
virginica 50 33.3333%
Total = 150

< 0.8

> 0.8

< 1.75

> 1.75

setosa 0 0%
versicolor 49 90.7407%
virginica 5 9.25926%
Total = 54

setosa 0 0%
versicolor 1 2.17391%
virginica 45 97.8261%
Total = 46

OPTIMAL

M. Janakiram Oct 2017
Tree building

Split 1

Split 2
Single tree: Classification vs. regression

**Classification**

Categorical Response: Species

Output: Classification counts, and Misclassification %

**Regression**

Continuous Response: Species Number

Output: mean and variance for each node.
Gartner Magic Quadrant - Data Science

Source: Gartner (February 2017)
Agile is too slow
We need to be “on the bridge”
Consulting = dead zone
We’re changing the world

Data Scientist: The Sexiest Job of the 21st Century
Harvard Business Review October 2012

Business Analytics
Questions?