Predictive Modeling in Workers Compensation

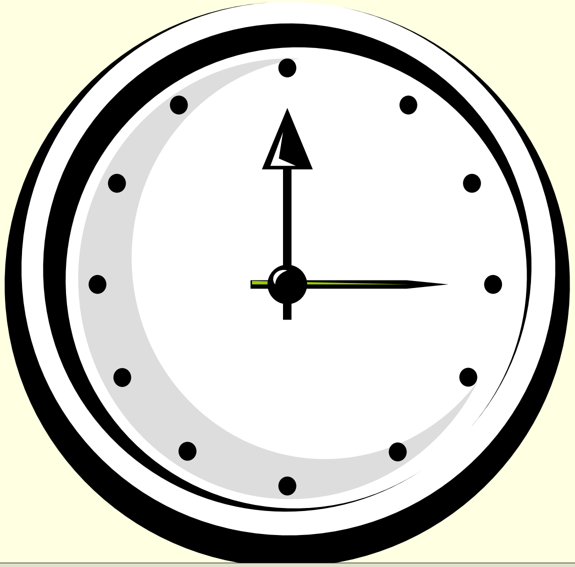
2008 CAS Ratemaking Seminar

Prepared by
Louise Francis, FCAS, MAAA
Francis Analytics and Actuarial Data Mining, Inc.
www.data-mines.com
Louise.francis@data-mines.cm

Objectives

- Introduce predictive modeling and where modeling fits in actuarial practice
- Discuss connection to traditional analytical procedures
- Discuss applications of predictive modeling in Workers Compensation

Timeline of Casualty Actuarial Evolution

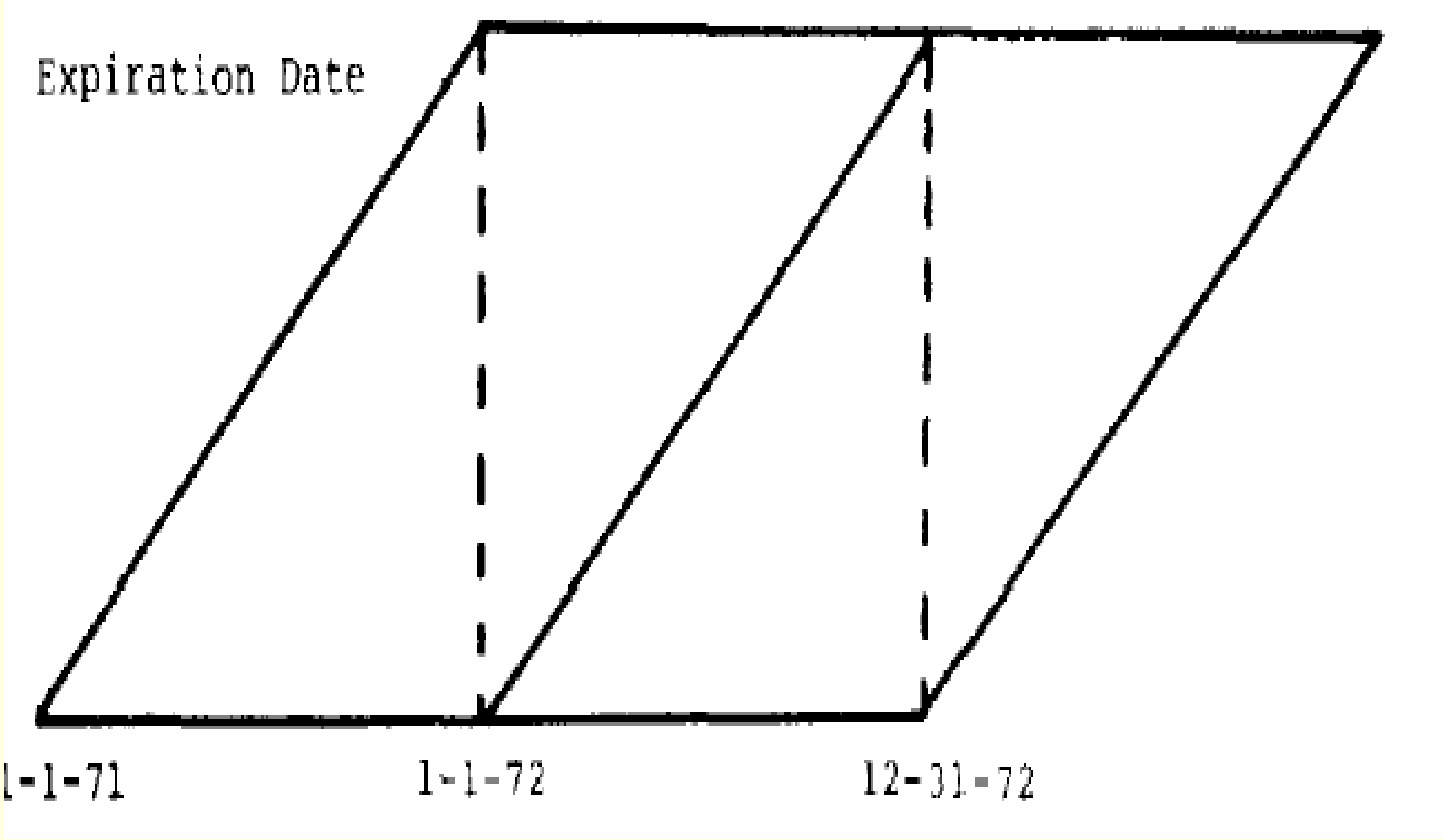


A Casualty Actuary's Perspective on Data Modeling

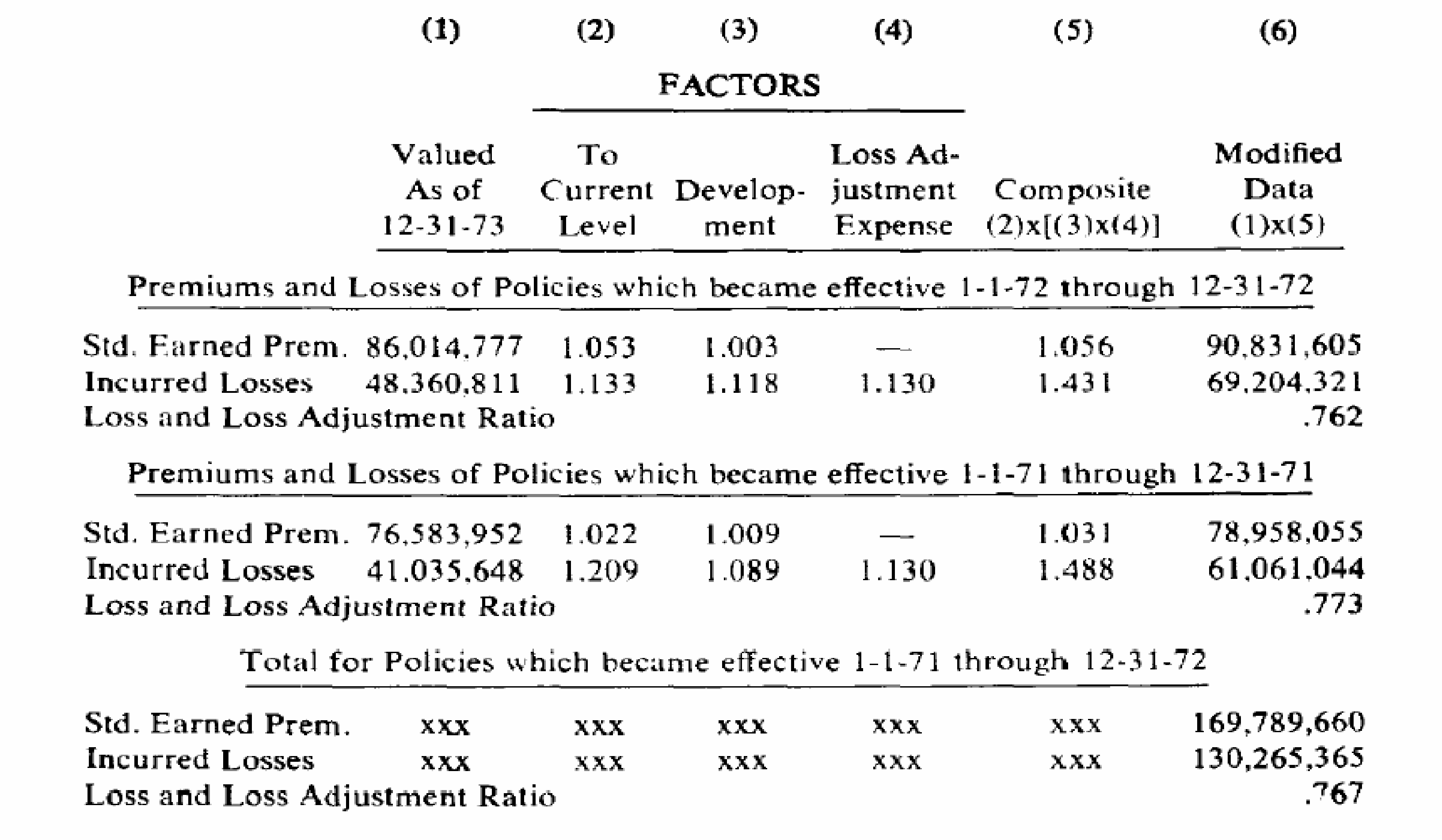
- The Stone Age: 1914 – …
  - Simple deterministic methods: Slice and dice data based on a few categories
    - Compute means or relativities in each cell
    - Ignore interactions and other multivariate relationships
    - Often ad-hoc
    - Based on empirical data – little use of parametric models
- The Pre – Industrial age: 1970 - …
  - Fit probability distributions to severity data
  - Focus is typically on underwriting, not claims
- The Industrial Age – 1985 …
  - Research published on computer catastrophe models
  - Use simulation to quantify variability
- The Computer Age 1990s…
  - European actuaries begin to use GLMs
  - At end of 20st century, large companies and consulting firms start to use data mining
- The Current era
  - In personal lines, modeling the rule rather than the exception
    - Often GLM based, though GLMs evolving to GAMs
  - Commercial lines beginning to embrace modeling for ratemaking and underwriting

Stone Age Example: WC Ratemaking:

Wineman, 1990 Discussion Paper Program

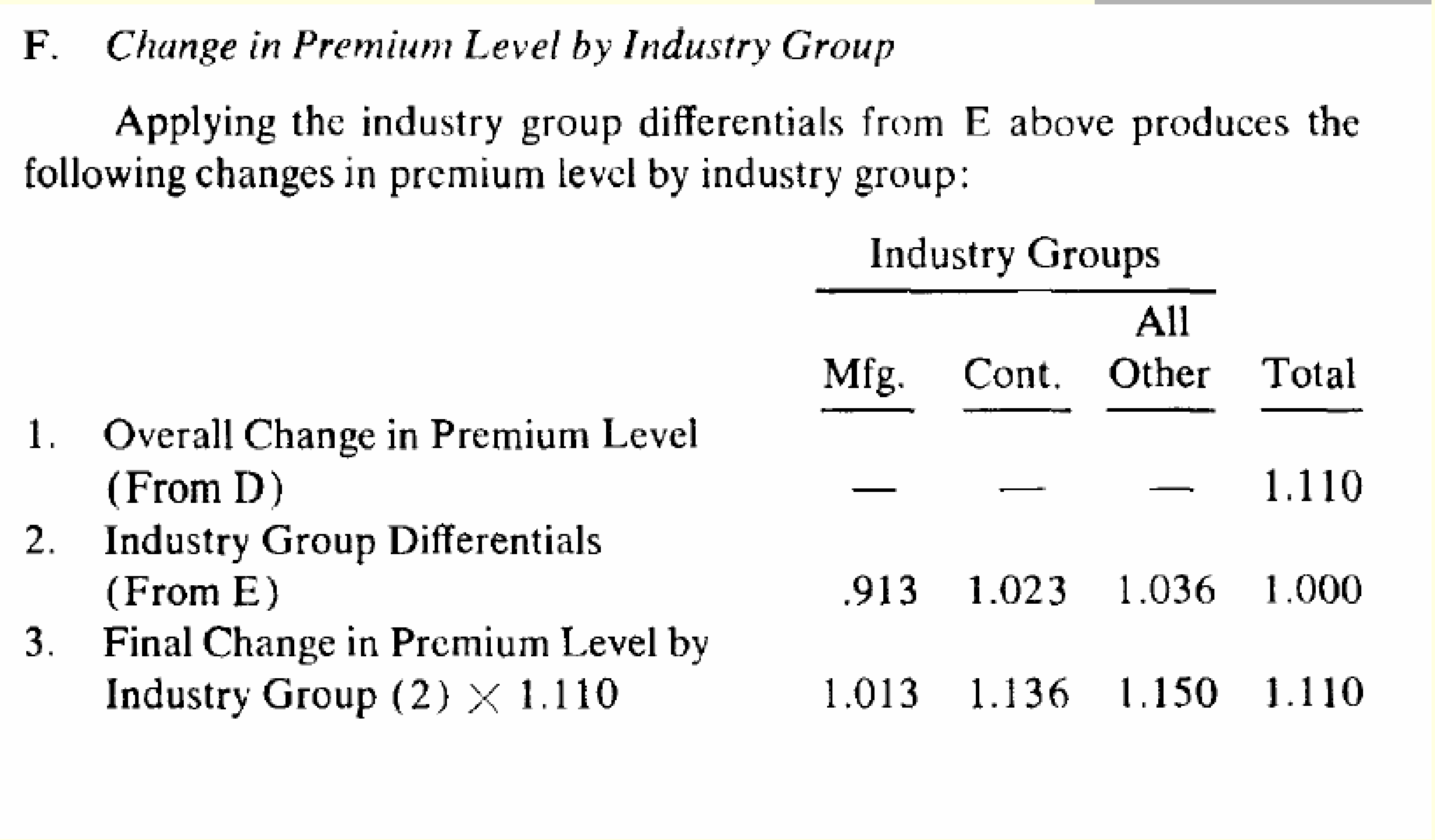


WC Ratemaking (Kallop – 1975)



	(1)	(2)	(3)	(4)	(5)	(6)
		FACTORS				
	Valued As of 12-31-73	To Current Level	Development	Loss Adjustment Expense	Composite (2)x[(3)x(4)]	Modified Data (1)x(5)
Premiums and Losses of Policies which became effective 1-1-72 through 12-31-72						
Std. Earned Prem.	86,014,777	1.053	1.003	—	1.056	90,831,605
Incurred Losses	48,360,811	1.133	1.118	1.130	1.431	69,204,321
Loss and Loss Adjustment Ratio						.762
Premiums and Losses of Policies which became effective 1-1-71 through 12-31-71						
Std. Earned Prem.	76,583,952	1.022	1.009	—	1.031	78,958,055
Incurred Losses	41,035,648	1.209	1.089	1.130	1.488	61,061,044
Loss and Loss Adjustment Ratio						.773
Total for Policies which became effective 1-1-71 through 12-31-72						
Std. Earned Prem.	xxx	xxx	xxx	xxx	xxx	169,789,660
Incurred Losses	xxx	xxx	xxx	xxx	xxx	130,265,365
Loss and Loss Adjustment Ratio						.767

WC Ratemaking, cont.

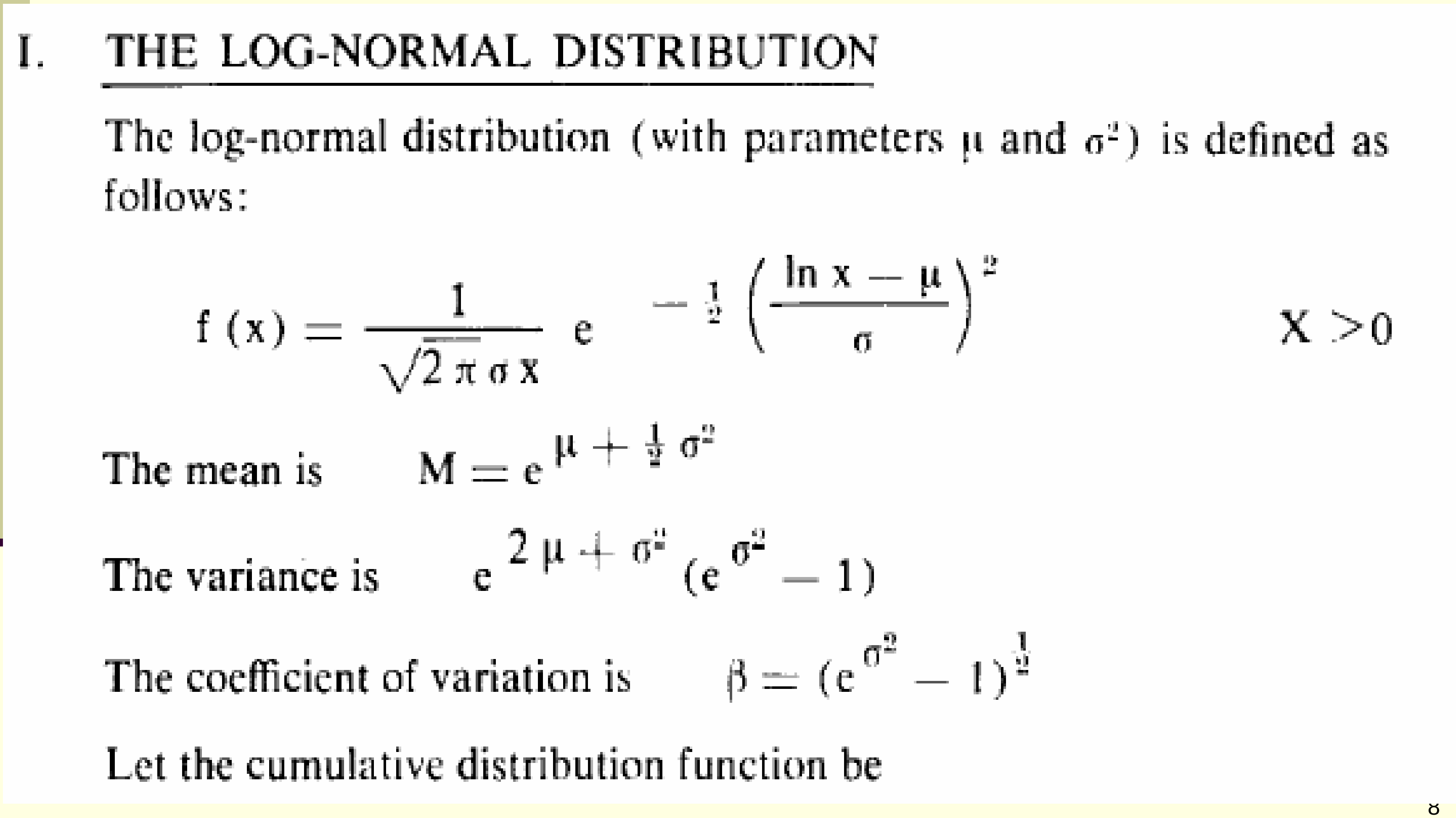


F. *Change in Premium Level by Industry Group*

Applying the industry group differentials from E above produces the following changes in premium level by industry group:

	Industry Groups			
	Mfg.	Cont.	All Other	Total
1. Overall Change in Premium Level (From D)	—	—	—	1.110
2. Industry Group Differentials (From E)	.913	1.023	1.036	1.000
3. Final Change in Premium Level by Industry Group (2) × 1.110	1.013	1.136	1.150	1.110

Pre-Industrial: Model for Increased Limits Factors: Finger, PCAS, 1976



I. THE LOG-NORMAL DISTRIBUTION

The log-normal distribution (with parameters μ and σ^2) is defined as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \qquad X > 0$$

The mean is $\quad M = e^{\mu + \frac{1}{2}\sigma^2}$

The variance is $\quad e^{2\mu + \sigma^2}\,(e^{\sigma^2} - 1)$

The coefficient of variation is $\quad \beta = (e^{\sigma^2} - 1)^{\frac{1}{2}}$

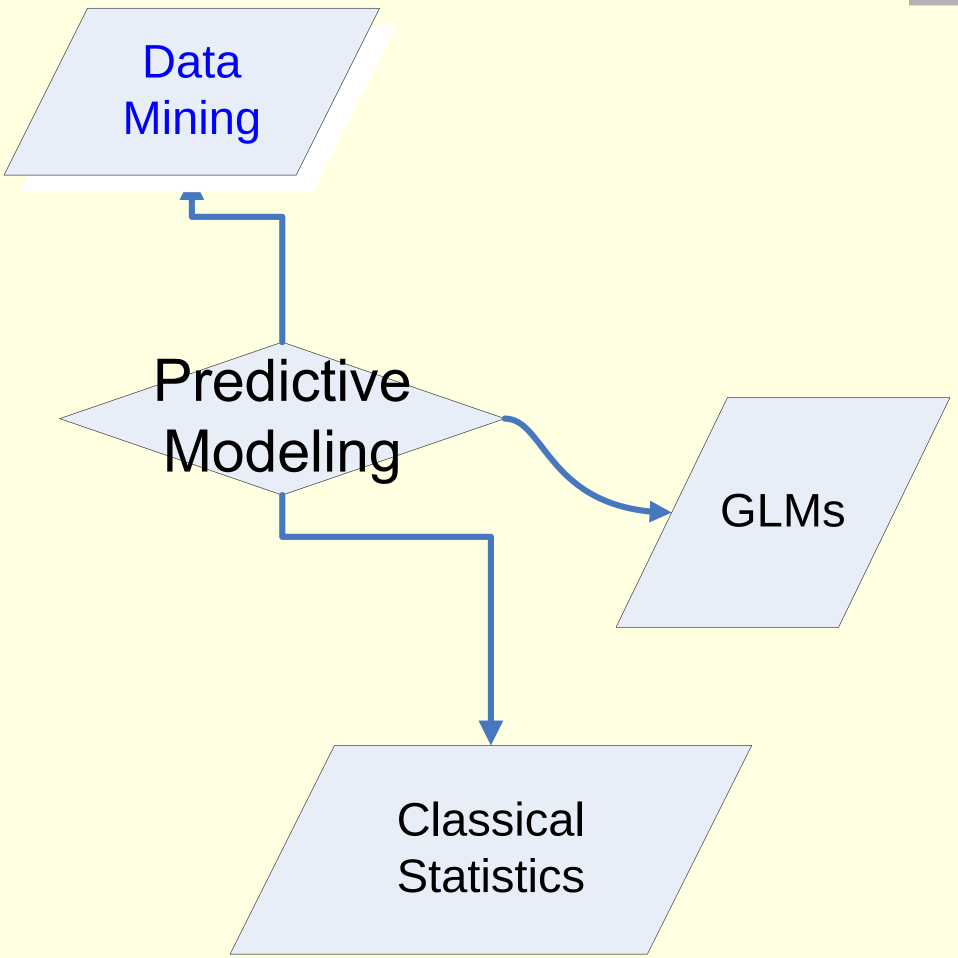
Let the cumulative distribution function be

Predictive Modeling Overview

A Premise about Advanced Modeling

- Advanced data mining and machine learning procedures are fancy versions of more basic procedures that many people already understand
- Predictive modeling software allows users to analyze large databases to solve business decision problems

Predictive Modeling Family



Major Categories of Modeling

- Supervised learning
  - Most common situation
  - A dependent variable
    - Claim Frequency
    - Loss Ratio
    - Renew/non-renew
    - Fraud/Legitimate
  - Some methods
    - Regression and GLMs
    - Trees
    - Some neural networks

- Unsupervised learning
  - No dependent variable
  - Group like records together
    - Territory construction
    - Some fraud prediction
    - Text mining
  - Some Methods
    - K-means clustering
    - Principal components
    - Kohonen neural networks

Kinds of Applications

- Classification
  - Target variable is categorical
- Prediction
  - Target variable is numeric

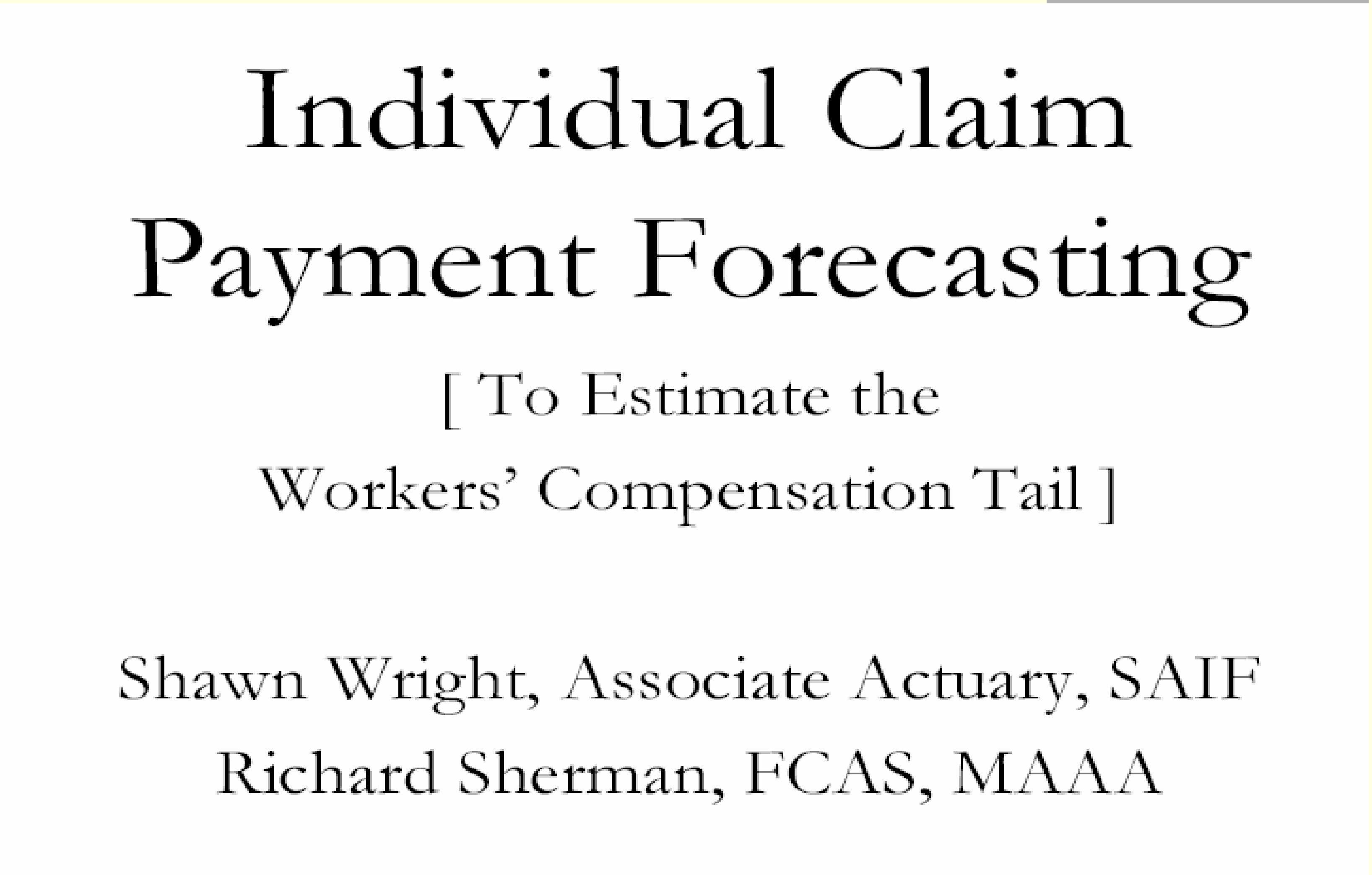
POTENTIAL VALUE OF AN PM SCORING SYSTEM

- Screening to Select Accounts
- Providing Evidence to Support a non-renewal
- Auditing of Canceled Policies to Determine Reasons for Cancellation
- Pricing for Some Accounts (small accounts)
- Provide evidence to regulators to support use of credit information
- Reserving

Underwriting Applications

- Develop model score for policyholders. Use to augment underwriter judgment
- Use to rate accounts
  - More likely to apply to small accounts
- Estimate full lifetime value of account
  - Model liklihood of renewal

WC Reserving



Individual Claim Payment Forecasting

[To Estimate the Workers' Compensation Tail]

Shawn Wright, Associate Actuary, SAIF

Richard Sherman, FCAS, MAAA

TYPES OF FRAUD

- WORKERS' COMPENSATION

- Employee Fraud
- -Working While Collecting

  -Staged Accidents

  -Prior or Non-Work Injuries
- Employer Fraud
- -Misclassification of Employees
- -Understating Payroll
- -Employee Leasing
- -Re-Incorporation to Avoid Mod

Insurance Fraud- The Problem

- ISO/IRC 2001 Study: Auto and Workers Compensation Fraud a Big Problem by 27% of Insurers.
- Mass IFB: 1,500 referrals annually for Auto, WC, and (10%) Other P-L.

Slide provided by Richard Derrig

FRAUD IDENTIFICATION

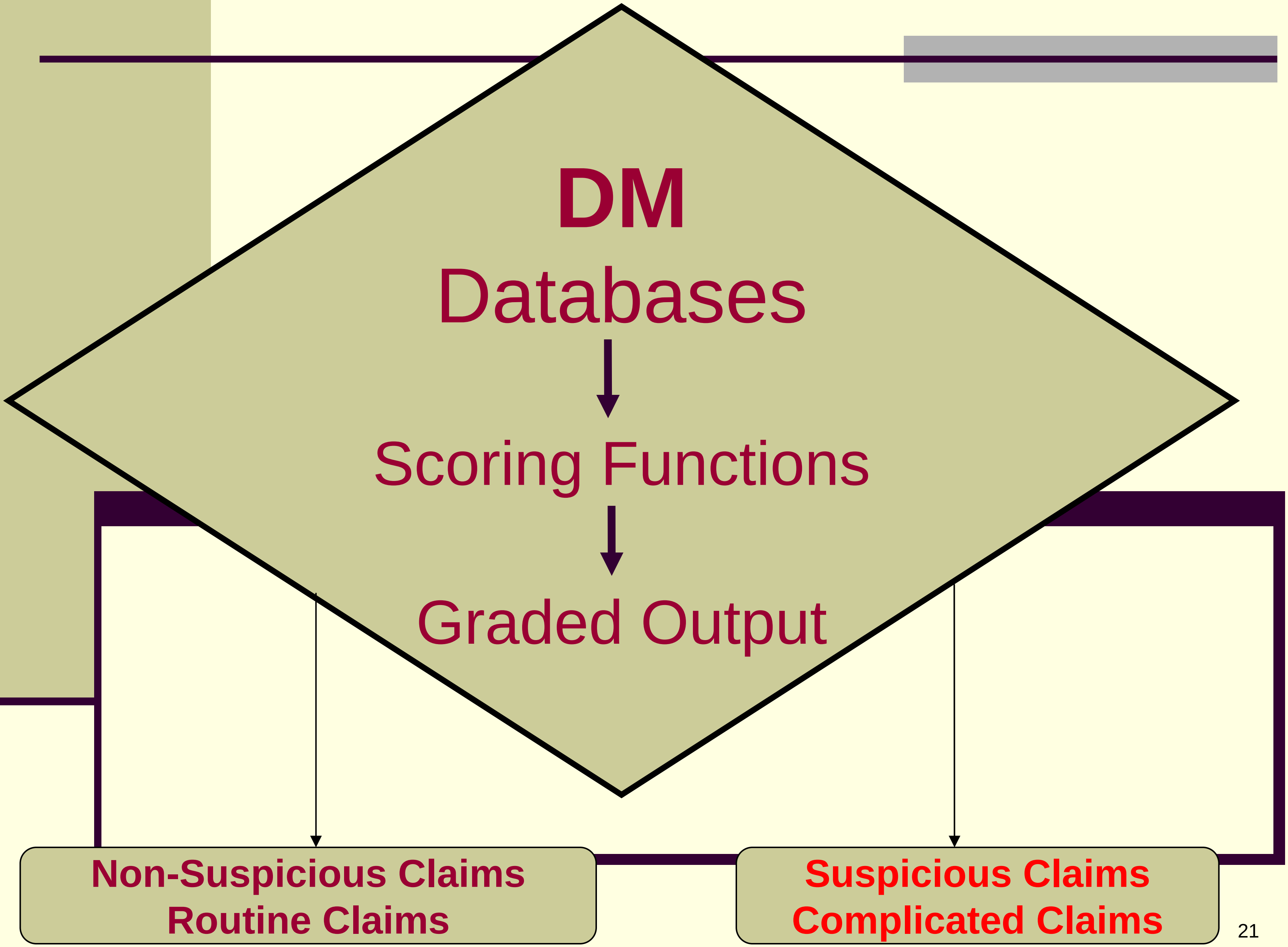
- Experience and Judgment
- Artificial Intelligence Systems
  - **Regression & Tree Models**
  - **Neural Networks**
  - Expert Systems
  - Fuzzy Clusters
  - Genetic Algorithms
  - All of the Above

Slide provided by Richard Derrig

REAL PROBLEM-CLAIM FRAUD

- Classify all claims
- Identify valid classes
  - Pay the claim
  - No hassle
  - Visa Example
- Identify (possible) fraud
  - Investigation needed
- Identify “gray” classes
  - Minimize with “learning” algorithms

Slide provided by Richard Derrig



DM

Databases

Scoring Functions

Graded Output

Non-Suspicious Claims
Routine Claims

Suspicious Claims
Complicated Claims

Slide provided by Richard Derrig

Underwriting Red Flags

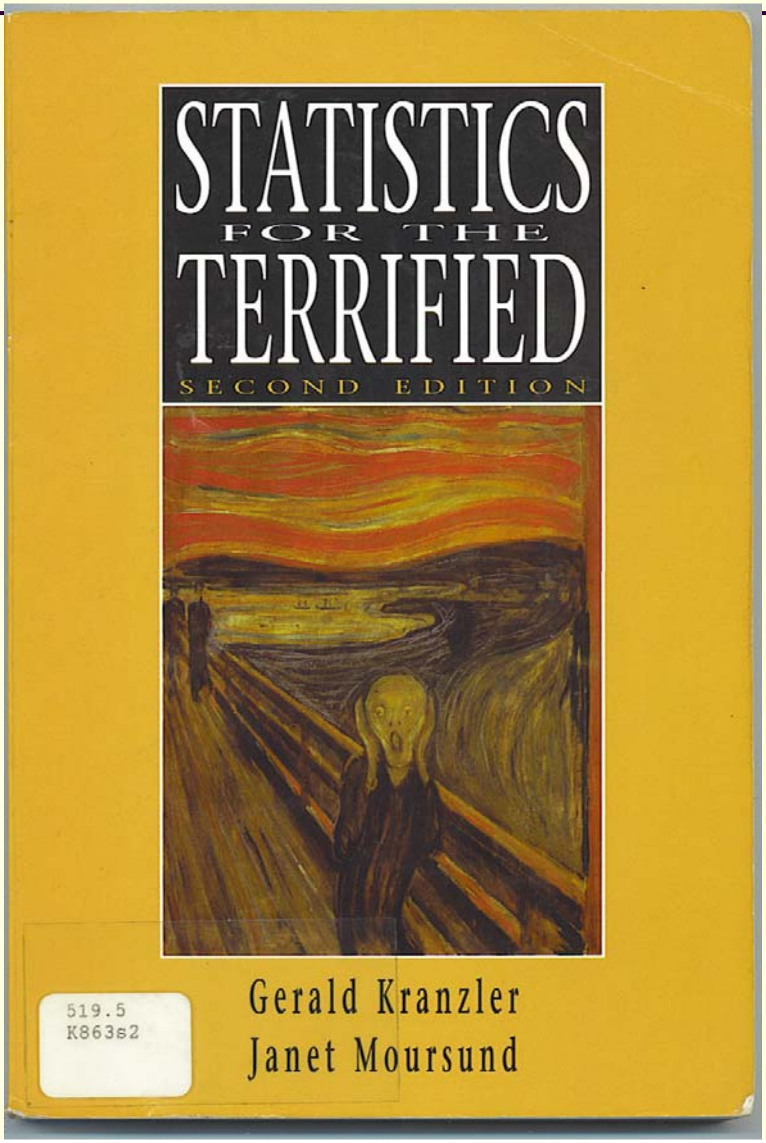
- Prior Claims History (Mod)
- High Mod versus Low Premium
- Increases/Decreases in Payroll
- Changes of Operation
- Loss Prevention Visits
- Preliminary Physical Audits
- Check Websites

Slide provided by Richard Derrig

Core Part of a Business Strategy

Motivation for for Methods

Many of the Methods are Intuitive



The Software Used in This Presentation

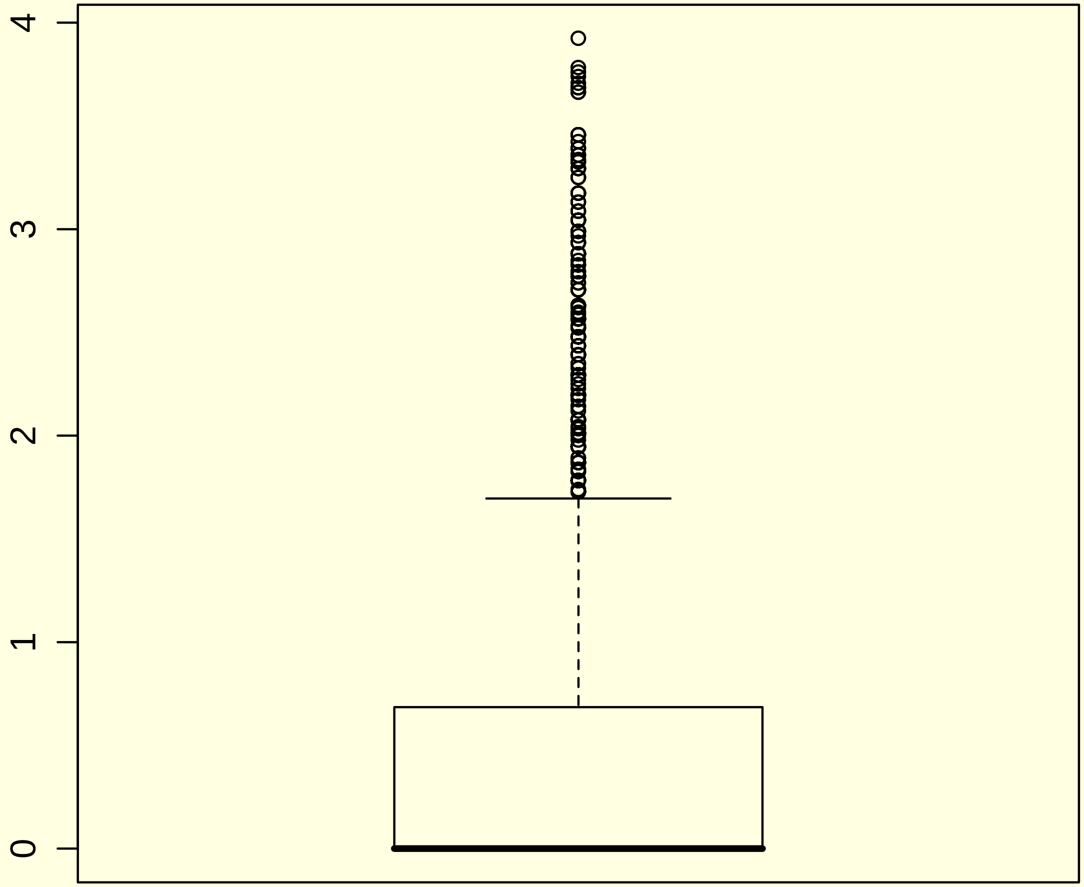
- Microsoft Excel
- R
  - Free statistical software
  - Get a book on using R
    - John Fox, *An R and S-PLUS Companion to Applied Regression*
  - Download from www.r-project .org
    - Install tree and nnet packages for decision trees and neural networks

Data Exploration in Predictive Modeling

Exploratory Data Analysis

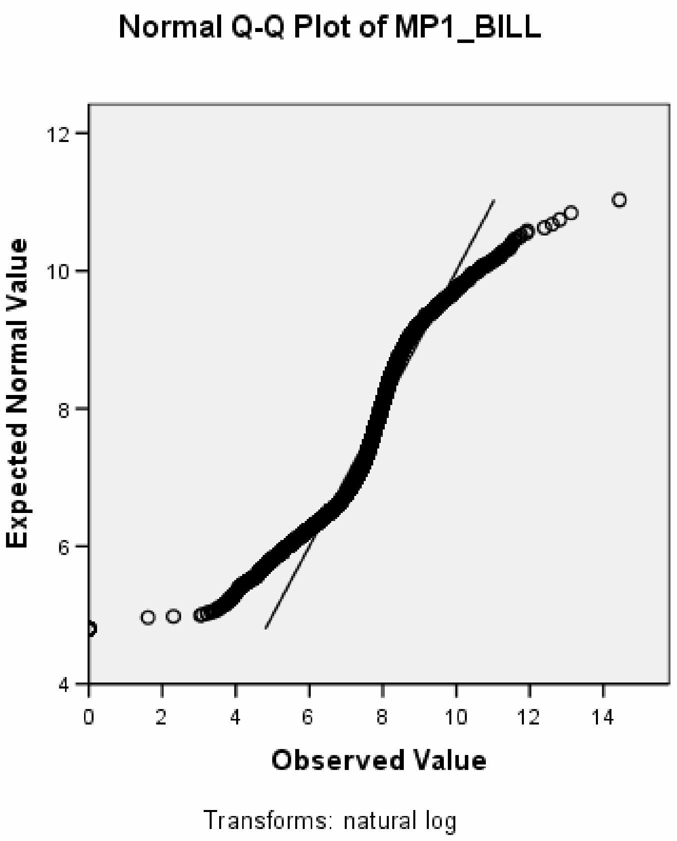
- Typically the first step in analyzing data
- Makes heavy use of graphical techniques
- Also makes use of simple descriptive statistics
- Purpose
  - Find outliers (and errors)
  - Explore structure of the data

Log of Box plot in R

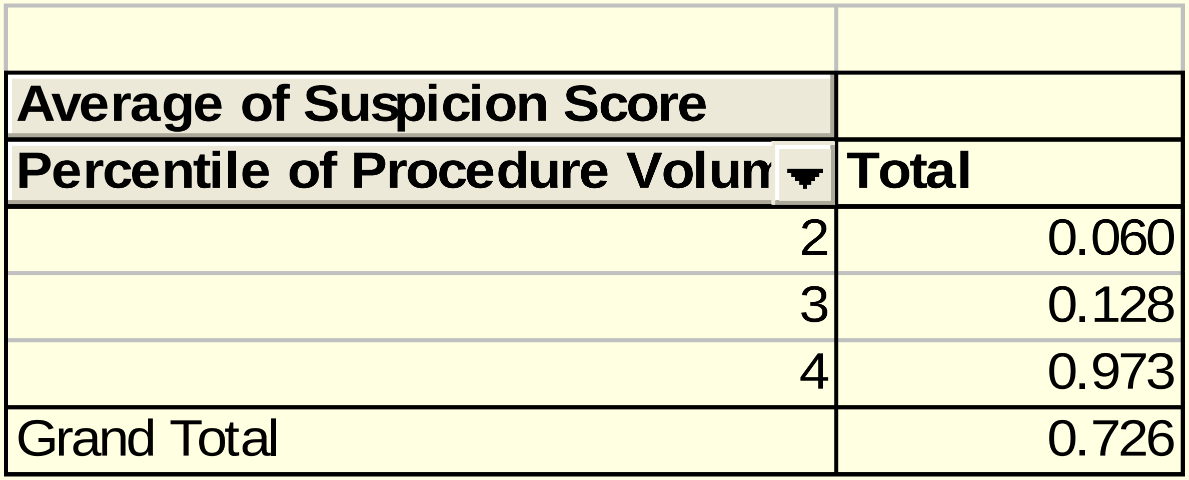


Log of average Procedures

Is the Data Normal? Q-Q Plots



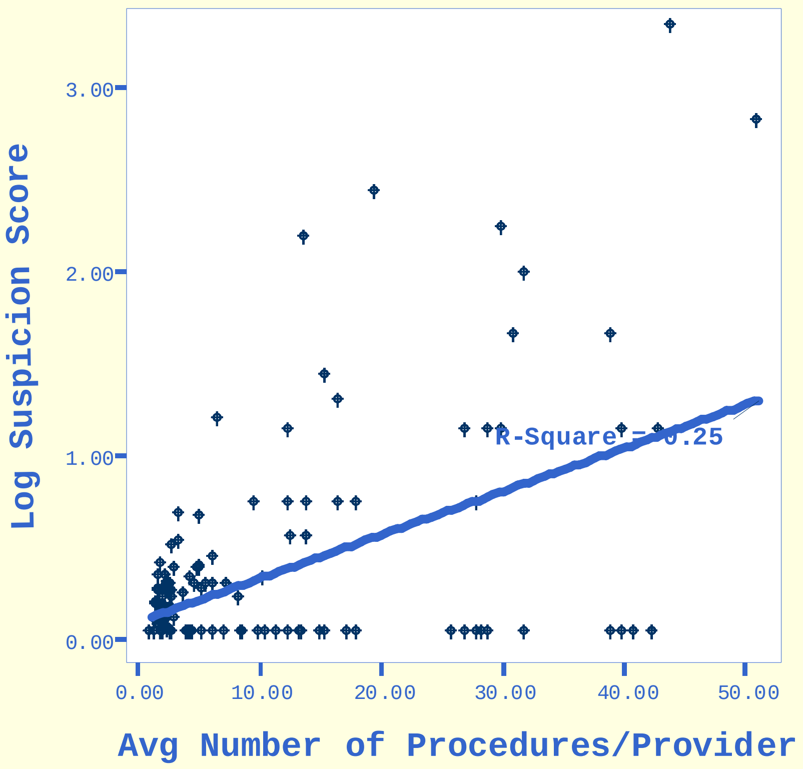
In Excel: Use Pivot Tables to Examine Relationship between Suspicion Indicator and Volume of Procedures for Provider (WC Data)



Average of Suspicion Score	
Percentile of Procedure Volun ▾	**Total**
2	0.060
3	0.128
4	0.973
Grand Total	0.726

Regression

A Model of Relationship Between Suspicion Score and Avg Number of Procedures/Claimant for Provider (WC Data)

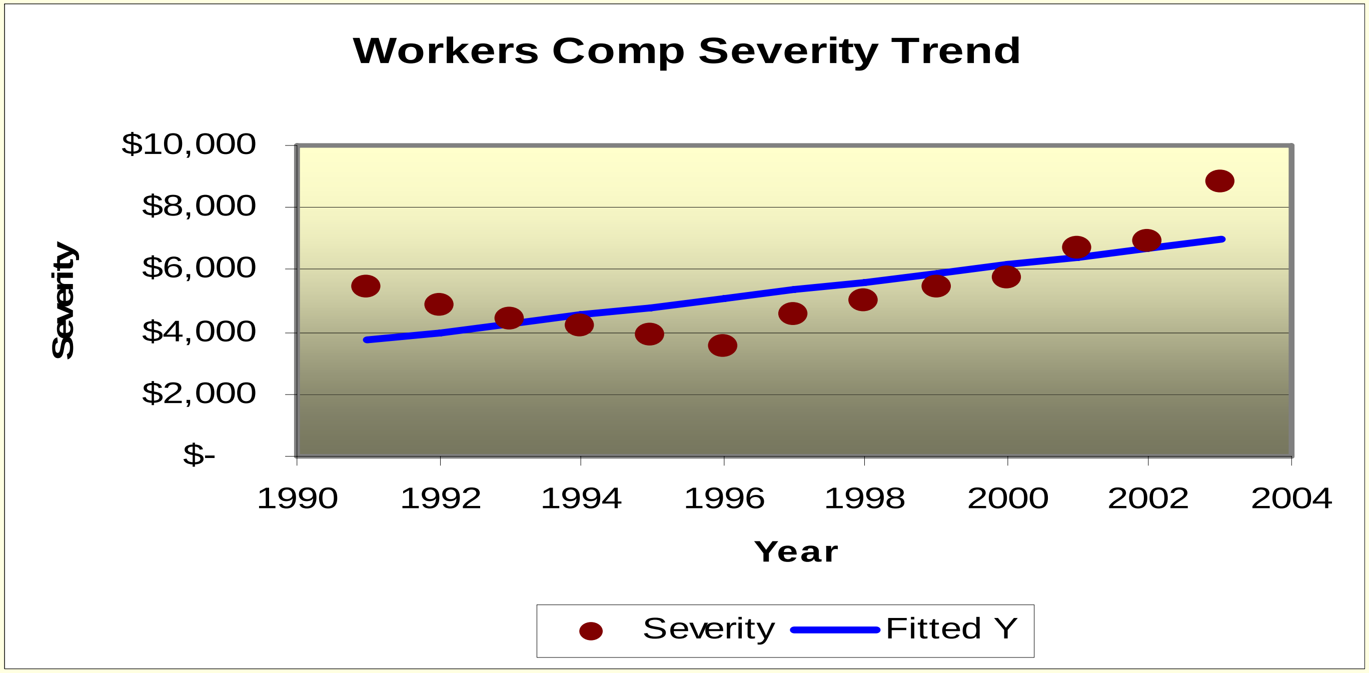


Linear Regression

Classical Statistics: Regression

- Estimation of parameters: Fit line that minimizes deviation between actual and fitted values

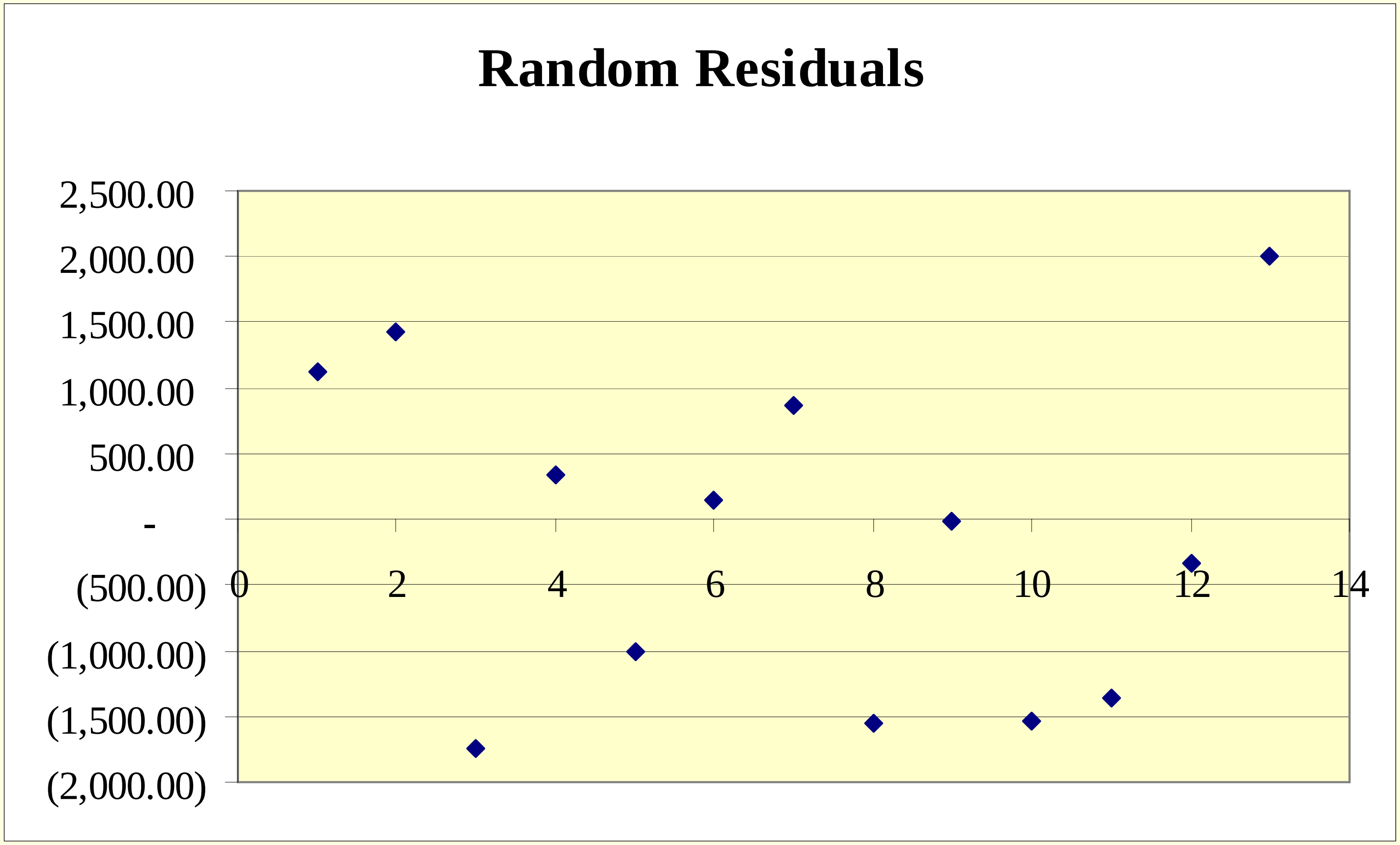
$$\min(\sum(Y_i - \hat{Y})^2)$$



Assumptions of Regression

- Errors independent of value of X
- Errors independent of value of Y
- Errors independent of prior errors
- Errors are from normal distribution
- Linearity

Random Residuals

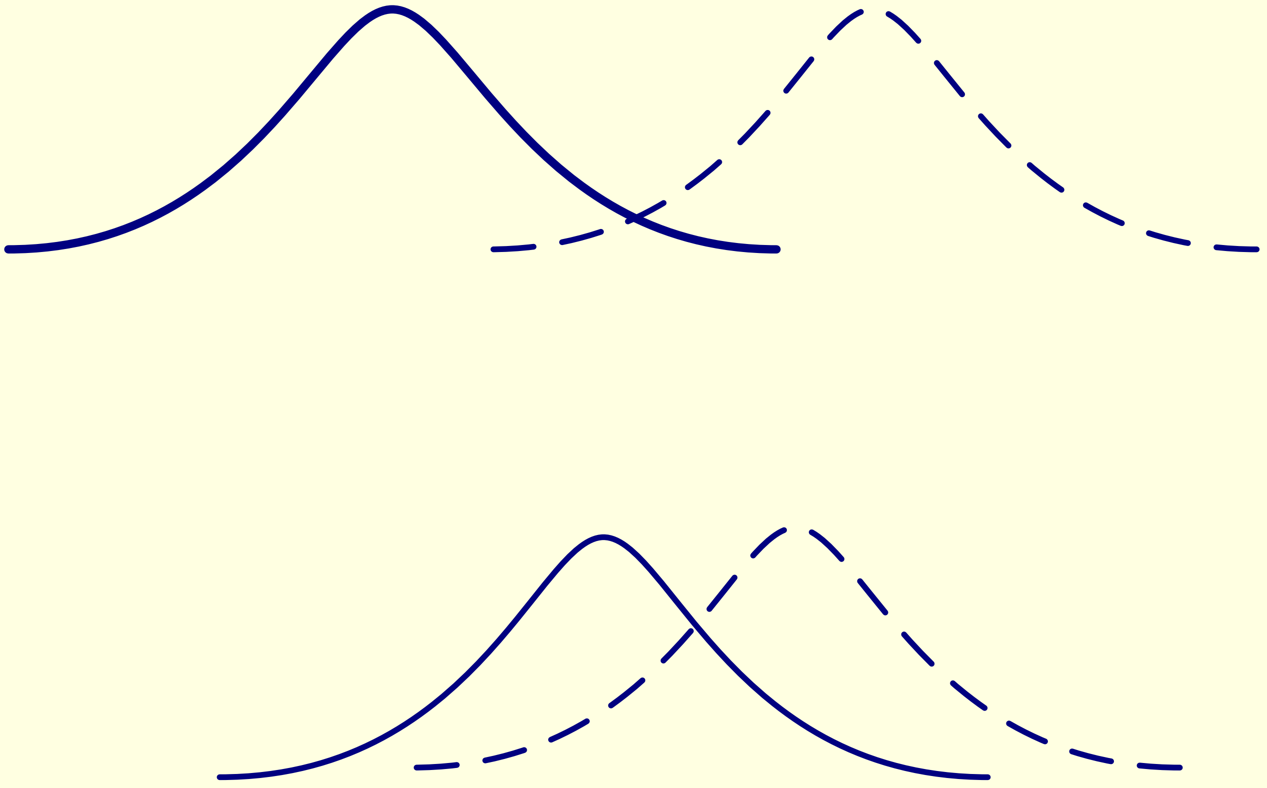


Discriminant Analysis

What is Discriminant Analysis?

- It is a procedure for identifying relationships between qualitative criterion and quantitative predictors.
- It identifies the boundaries between groups of objects
- The method has been used for classification problems for a very long time
- More recently it has been supplanted by logistic regression

Discriminant Analysis Predicts Class By Finding Variables that Separate Two Groups



The Discriminant Function and Its Use

- The function uses a weighted combination of predictor values to distribute objects to one of the criterion groups

$$L = b_1 x_1 + b_2 x_2 + \mathrm{K} + b_k x_k$$

- The various x values represent the predictor variables. The b values represent the weights that are associated with each of the variables.

Function and Use (cont.)

- To decide which values fall under which groups categories, a **cutoff score** is used.
- If the value of the discriminant function is higher than the cutoff score then it falls into one category and into the other if it is lower than the cutoff score.

Discriminant Analysis in Excel

- In some cases Discriminant Analysis can be done in Excel using the Regression function that is a part of the Data Analysis Tools Pack
- This can only be done if the dependent variable is binary

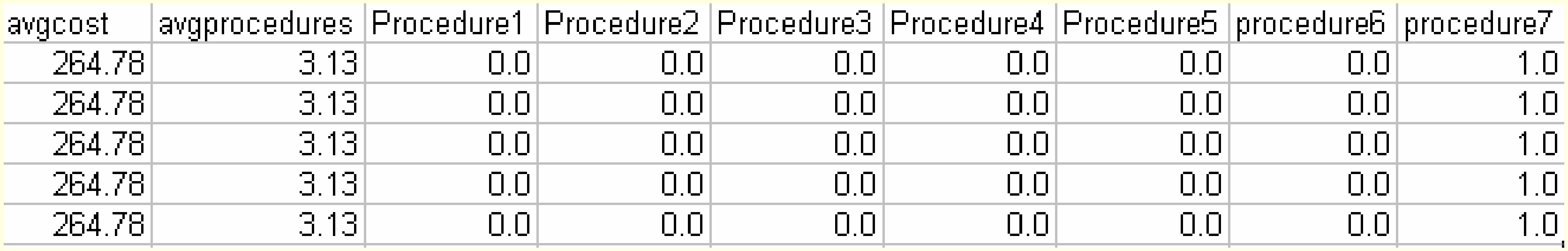
Example of Discriminant Analysis in Excel

- The dependent variable is the original suspicion score which is classified as either a 1 or a 0
  - It receives a 1 if the original score is greater than 0 and a 0 otherwise
- The two independent variables are the average number of procedures per claimant for one provider and the average cost of the procedures

Dummy Variables

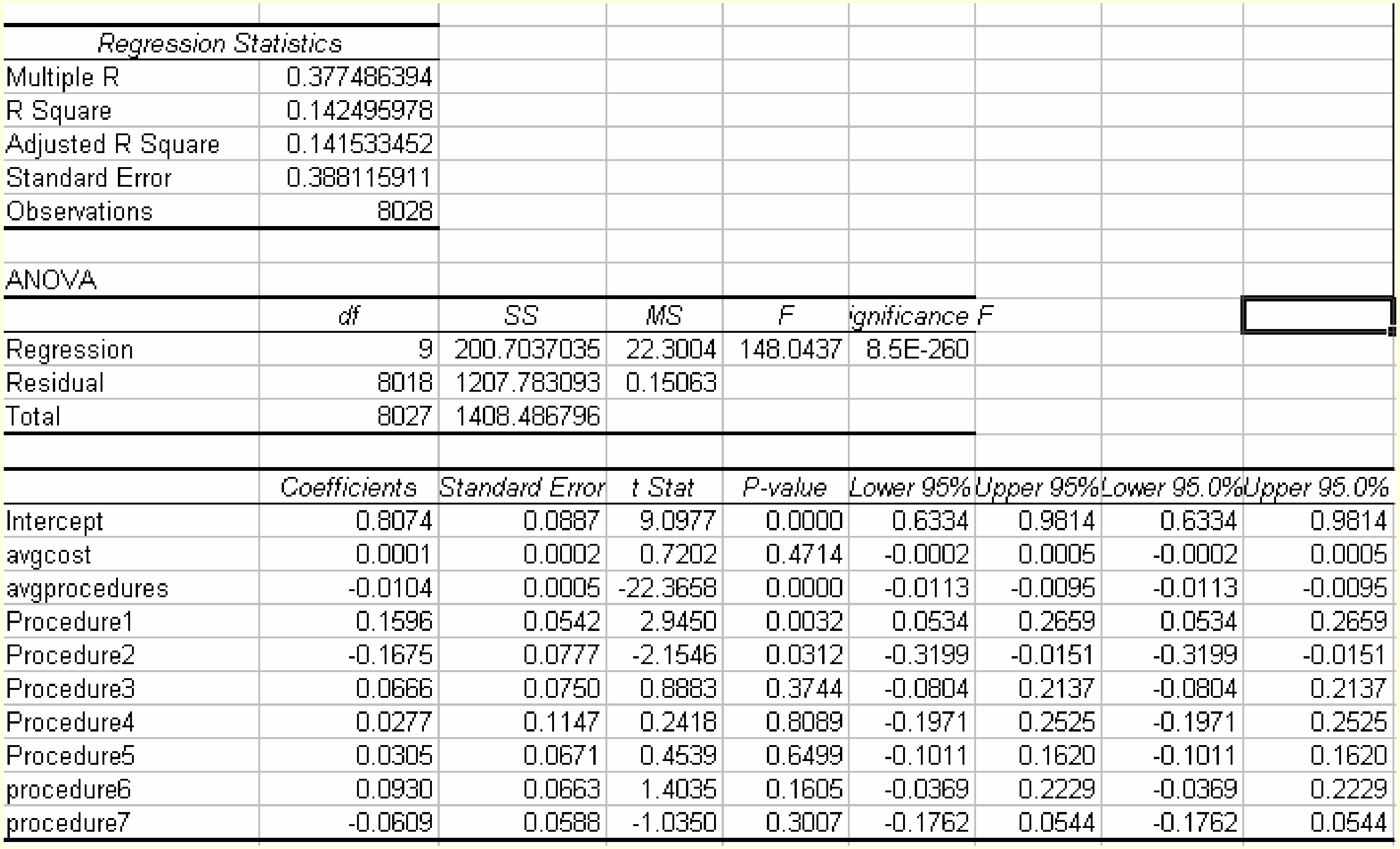
- Dummy Variables are used for coding information about categorical variables
- In our example:
  - Procedure Dummy 1 equals 1 if Procedure equals 1 it equals 0 otherwise
  - Procedure Dummy 2 equals 1 if Procedure equals 2 it equals 0 otherwise
  - Procedure Dummy 3 equals 1 if Procedure equals 1 , it equals 0 otherwise
  - Etc.
- ❑ Usually there is 1 fewer dummy variables than the number of categories.

Design Matrix with Dummy Variables



avgcost	avgprocedures	Procedure1	Procedure2	Procedure3	Procedure4	Procedure5	procedure6	procedure7
264.78	3.13	0.0	0.0	0.0	0.0	0.0	0.0	1.0
264.78	3.13	0.0	0.0	0.0	0.0	0.0	0.0	1.0
264.78	3.13	0.0	0.0	0.0	0.0	0.0	0.0	1.0
264.78	3.13	0.0	0.0	0.0	0.0	0.0	0.0	1.0
264.78	3.13	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Discriminant Analysis Example



Regression Statistics	
Multiple R	0.377486394
R Square	0.142495978
Adjusted R Square	0.141533452
Standard Error	0.388115911
Observations	8028

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	9	200.7037035	22.3004	148.0437	8.5E-260
Residual	8018	1207.783093	0.15063		
Total	8027	1408.486796			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.8074	0.0887	9.0977	0.0000	0.6334	0.9814	0.6334	0.9814
avgcost	0.0001	0.0002	0.7202	0.4714	-0.0002	0.0005	-0.0002	0.0005
avgprocedures	-0.0104	0.0005	-22.3658	0.0000	-0.0113	-0.0095	-0.0113	-0.0095
Procedure1	0.1596	0.0542	2.9450	0.0032	0.0534	0.2659	0.0534	0.2659
Procedure2	-0.1675	0.0777	-2.1546	0.0312	-0.3199	-0.0151	-0.3199	-0.0151
Procedure3	0.0666	0.0750	0.8883	0.3744	-0.0804	0.2137	-0.0804	0.2137
Procedure4	0.0277	0.1147	0.2418	0.8089	-0.1971	0.2525	-0.1971	0.2525
Procedure5	0.0305	0.0671	0.4539	0.6499	-0.1011	0.1620	-0.1011	0.1620
procedure6	0.0930	0.0663	1.4035	0.1605	-0.0369	0.2229	-0.0369	0.2229
procedure7	-0.0609	0.0588	-1.0350	0.3007	-0.1762	0.0544	-0.1762	0.0544

Classification Errors

- However, with this function also comes the possibility that although the calculation is correct the category into which the results is placed is not the right one.
- The smaller the difference between the two groups of the predictor variable, the larger the overlap and misclassification

Errors of Classification

How Good is the Prediction?

- Very sophisticated methods can be ineffective when applied to real-life situations
- We usually hold out a portion of the data to use for testing. This data is not used at all in model fitting.
- The Question: How accurate is the model on the test data?

Testing the Validity of the Prediction

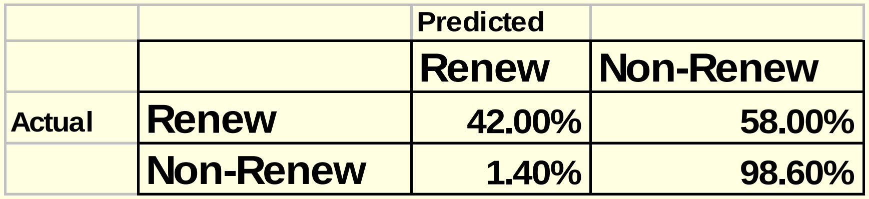
- This can be done by using a ***confusion matrix***. This matrix will show the errors and the accurate predictions

		Predicted	
		Renew	Non-Renew
Actual	Renew	490	10
	Non-Renew	10	90

		Predicted	
		Renew	Non-Renew
Actual	Renew	98%	2%
	Non-Renew	10%	90%

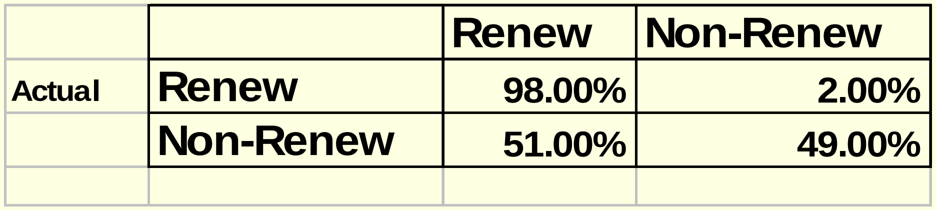
What is the Confusion Matrix telling Us?

- Sensitivity- The percent of true-positives that are accurately predicted
- Specificity- percent of true-negatives that are accurately predicted



		Predicted	
		Renew	Non-Renew
Actual	Renew	42.00%	58.00%
	Non-Renew	1.40%	98.60%

Examples of Bad Prediction



		Renew	Non-Renew
Actual	Renew	98.00%	2.00%
	Non-Renew	51.00%	49.00%

Trees

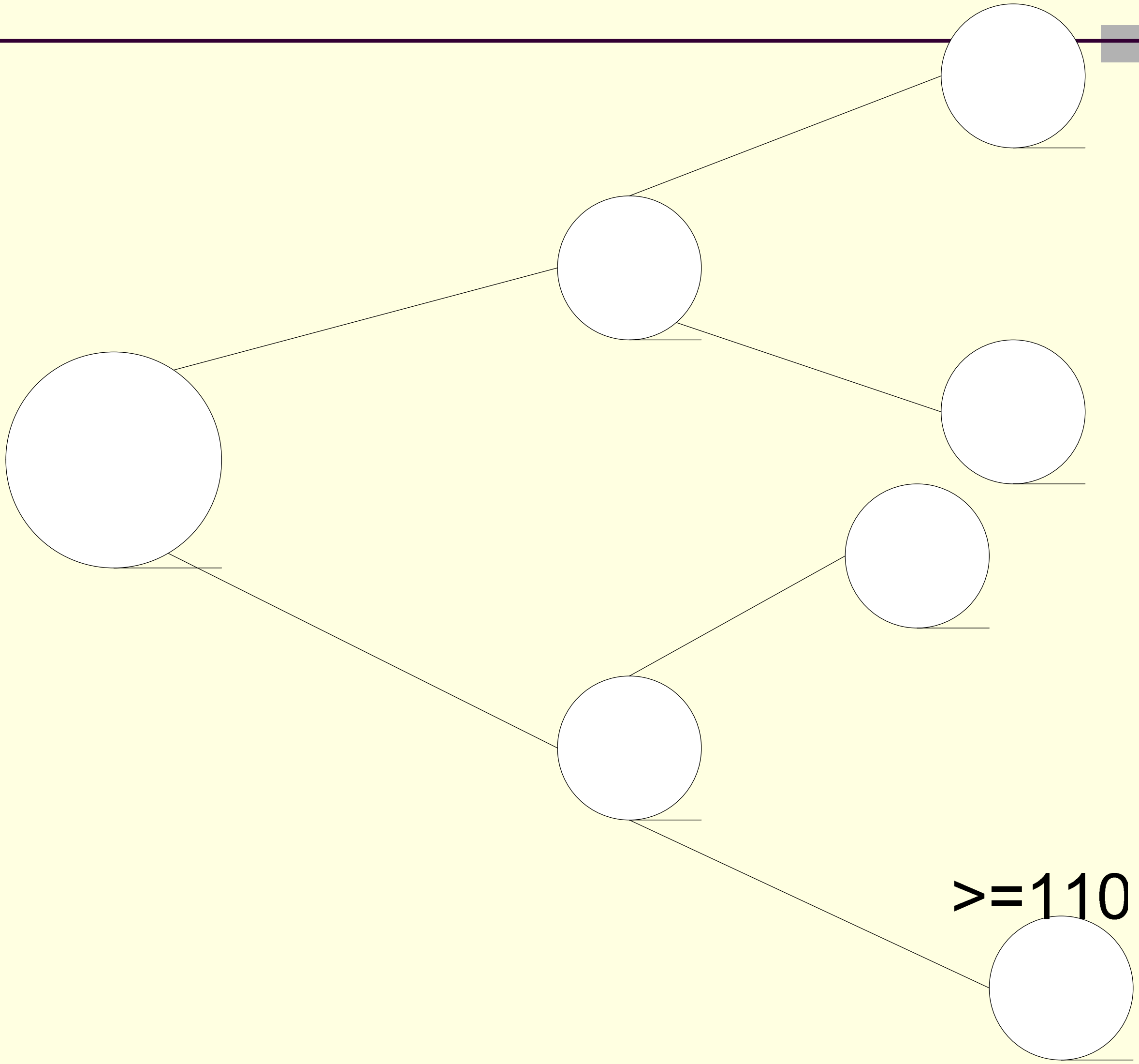
What are Trees?

- They are simple explanations of the data and the relationships within it
- They can be used for classification, prediction or estimation
- Trees divide data into subsets whose data is increasingly more similar.

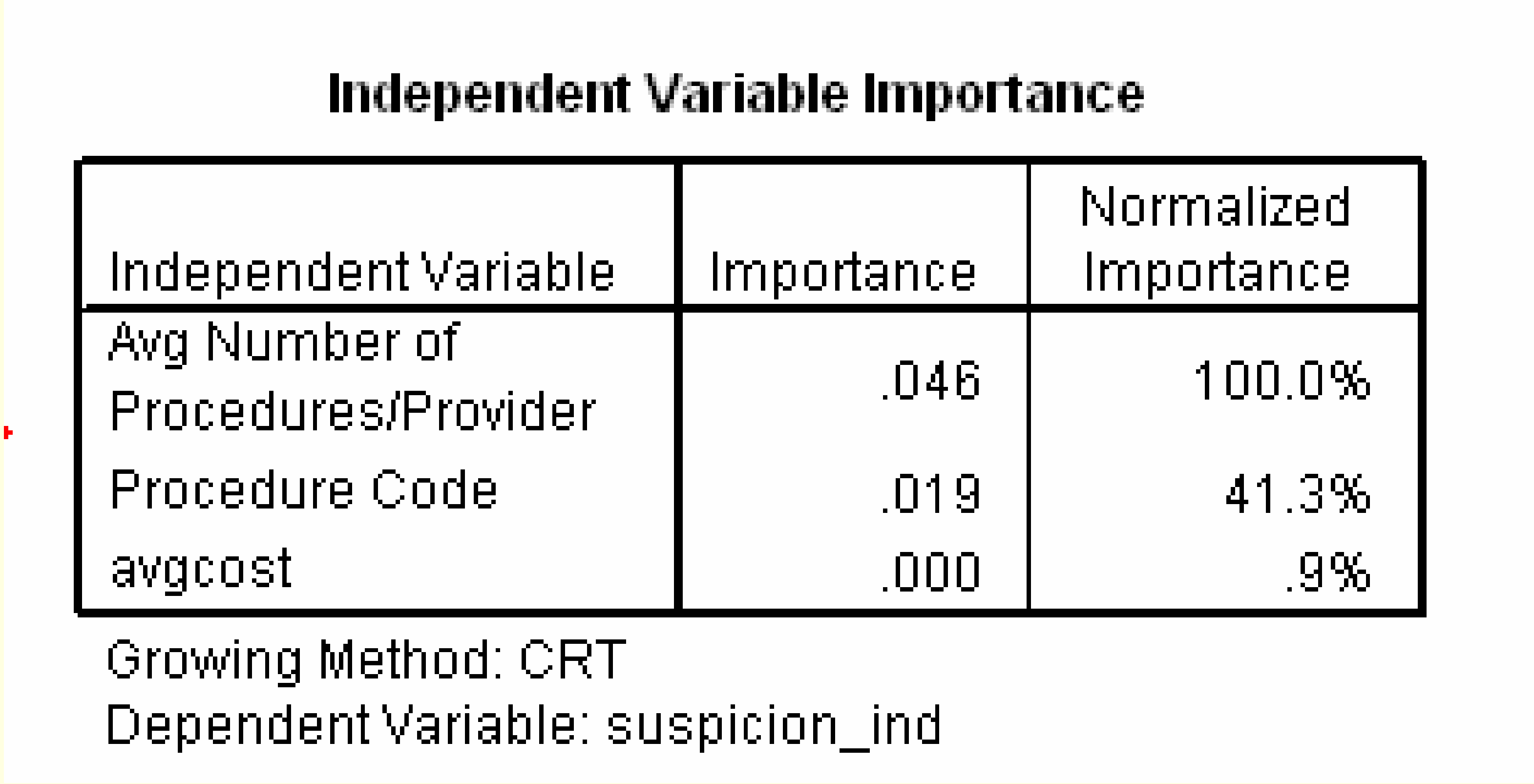
How do they Work?

- The tree function tests all the possible splits on all of the possible independent variables
- Then it decides which gives the largest gains in goodness of fit and chooses this split
- To keep the tree from having useless branches, a full tree is diagrammed but then the branches that increase the error are removed from the tree
- When using categorical data the data is separated according to the answer to the question
- When using continuous data, it is split according to an average value as far away as possible from the other averages.

A Decision Tree



Independent Variable Importance

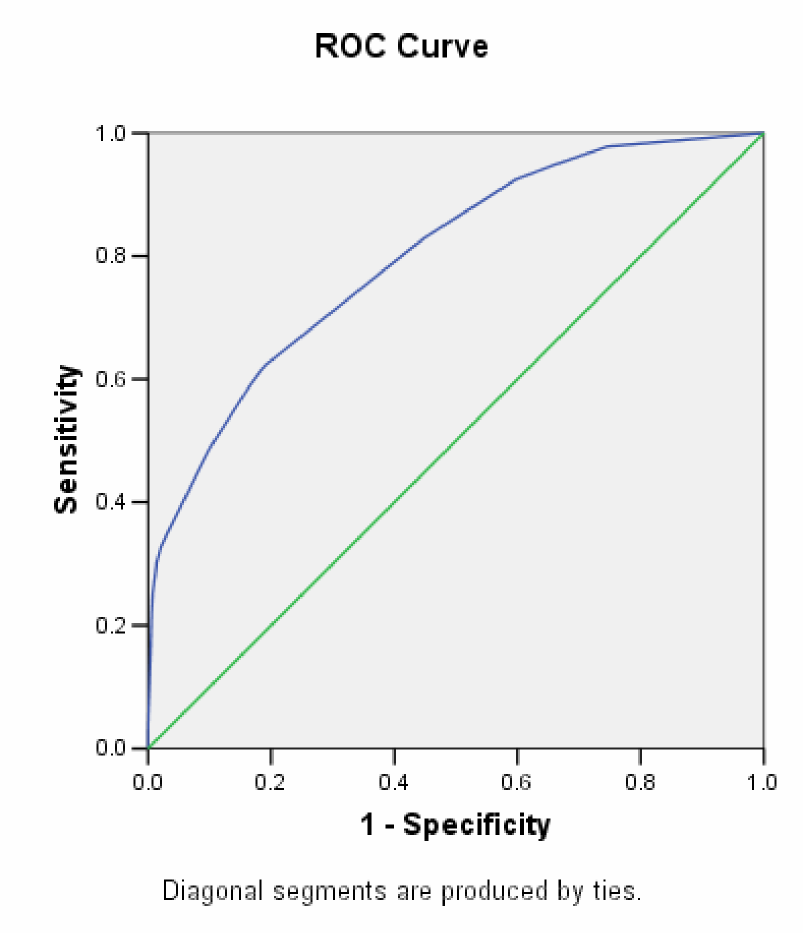


Independent Variable Importance

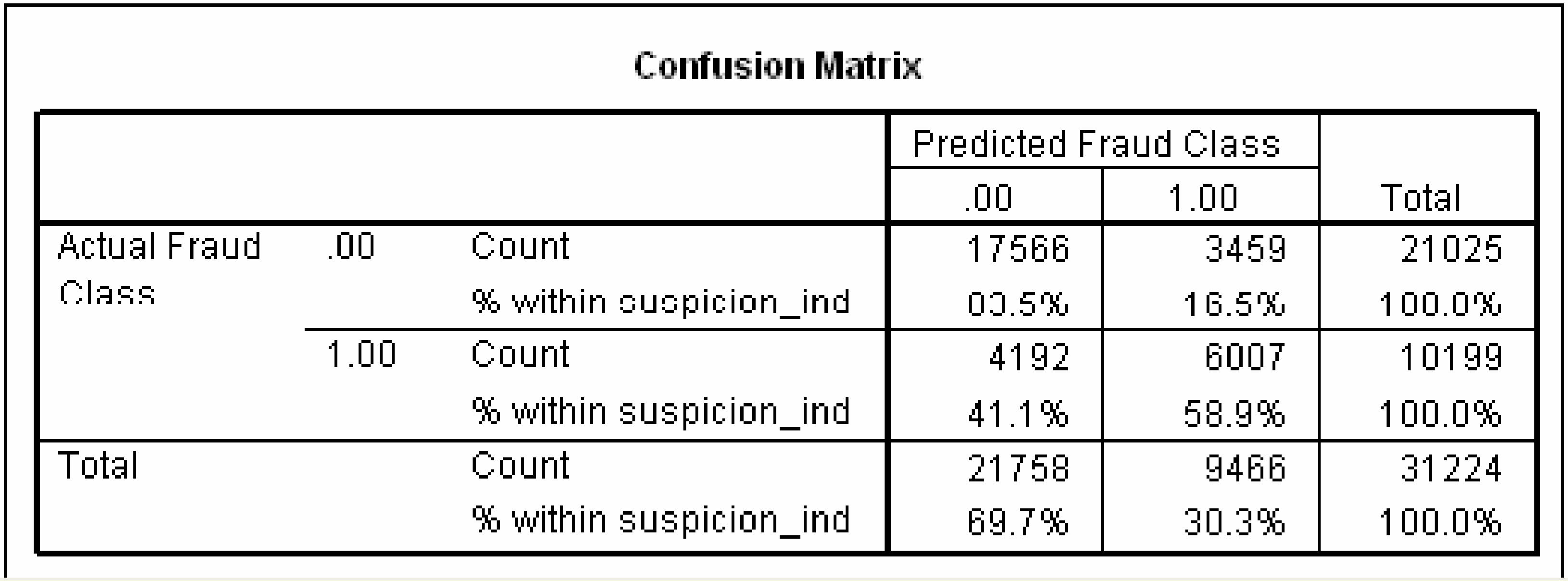
Independent Variable	Importance	Normalized Importance
Avg Number of Procedures/Provider	.046	100.0%
Procedure Code	.019	41.3%
avgcost	.000	.9%

Growing Method: CRT
Dependent Variable: suspicion_ind

ROC Curve


ROC Curve
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity
Diagonal segments are produced by ties.

Confusion Matrix



Confusion Matrix

			Predicted Fraud Class		Total
			.00	1.00	
Actual Fraud Class	.00	Count	17566	3459	21025
		% within suspicion_ind	03.5%	16.5%	100.0%
	1.00	Count	4192	6007	10199
		% within suspicion_ind	41.1%	58.9%	100.0%
Total		Count	21758	9466	31224
		% within suspicion_ind	69.7%	30.3%	100.0%

Trees in Excel

- Trees can also be made in Excel with the help of a program on the following site:

  http://www.geocities.com/adotsaha/CTree/CtreeinExcel.html

Library for Getting Started

- Dahr, V, *Seven Methods for Transforming Corporate into Business Intelligence*, Prentice Hall, 1997
- Berry, Michael J. A., and Linoff, Gordon, *Data Mining Techniques*, John Wiley and Sons, 1997, 2003
- Find a comprehensive book for doing analysis in Excel such as: John Walkebach, *Excel 2003 Formulas* or Jospeh Schmuller, Statistical Analysis With Excel for Dummies
- If you use R, get a book like: Fox, John*, An R and S-PLUS Companion to Applied Regression*, Sage Publications, 2002
- Francis, L.A., Neural Networks Demystified, *Casualty Actuarial Society Forum,* Winter, pp. 254-319, 2001. Found at www.casact.org
- Francis, L.A., “Taming Text: An Introduction to Text Mining”, CAS Winter Forum, March 2006, www.casact.org
- Francis, L.A., Martian Chronicles: Is MARS better than Neural Networks? *Casualty Actuarial Society Forum,* Winter, pp. 253-320, 2003.