2013 CAIR Conference

Using Data Mining to Model Student Success for the Purpose of Refining Nursing Program Admission Criteria

---

by

Shana Ruggenberg
Nursing & Health Sciences

Serhii Kalynovs'kyi
Institutional Research

Pacific Union College

PUC Nursing Program

- Seventh-day Adventist Christian liberal arts college with diverse student body
- AS Degree in Nursing
  - Traditional 2-year program
  - Non-traditional LVN-RN program
- BSN Degree in Nursing
  - Non-traditional RN to BSN program

Pacific Union College

Admission to Nursing Program

- Complete five prerequisite courses (minimum C):
  - Algebra and Chemistry – HS or College
  - College English
  - Human Anatomy (or Physiology)
  - Introduction to Nursing
- Minimum cognate/GE GPA – 2.7
  - Repeats for failure limited to two courses
- Minimum ACT English – 19 or better
- TEAS Score – Proficient level or better
- Institutional Research (IR) score – 0.7 or better

Pacific Union College

Outcomes Evaluation

- Measures of success
  - On-time completion of nursing program
  - Success on NCLEX-RN at first attempt

- Measures important to
  - Student
  - Program/Institution – BRN, ACEN, WASC

Pacific Union College

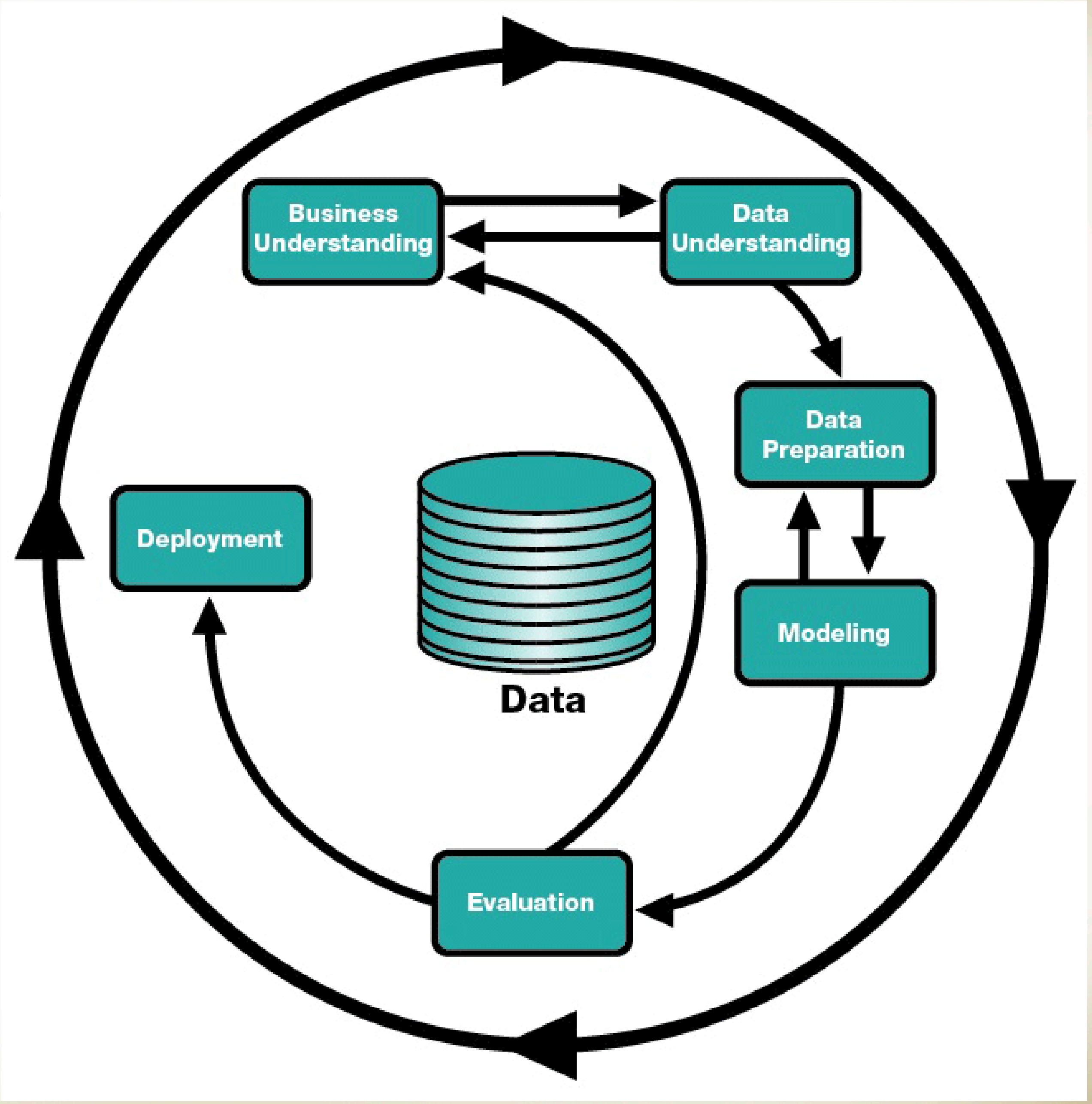
The Study – Challenges to Address

- Data – not enough, too much, or inconsistent
  - Few prerequisites = gaps in data
  - Collection methods change with time
  - Values recorded differently in time

- Sampling issues
  - Admission criteria determines admission
  - Admission required for inclusion in study

Pacific Union College

CRISP-DM Protocol



Source: http://crisp-dm.eu/

Data Understanding and Preparation: What Is Available? (Data Inventory)

- **Two outcome (dependent) variables:**
  - Program Completion:
    - *Completed*
    - *Failed*
    - *Withdrew because of failing*
    - *Withdrew without failing*
  - Passing NCLEX on the First Attempt:
    - *Passed*
    - *Failed*
    - *Never Took*
- **23 predictor (independent) variables** *(next page)*

Pacific Union College

What Is Available?
Independent (Predictor) Variables

- Nursing GPA
- IR Score
- TEAS Total Score
- TEAS Reading Subscore
- TEAS Math Subscore
- TEAS Science Subscore
- TEAS English Subscore
- ACT English Score
- Number of Repeats
- Number of Quarters Applied
- Number of Completed Classes (Max 12)
- Math Grade
- Chemistry/Physics Grade
- Intro to Nursing Grade
- ENGL 101 Grade
- Anatomy Grade
- Physiology Grade
- Microbiology Grade
- Nutrition Grade
- General Psychology Grade
- Human Development Grade
- Sociology Grade
- Speech Grade

Pacific Union College

More Data Understanding: Group Comparisons

- Are there statistically significant differences between averages for passing and failing groups?
  - *Make outcome variables dichotomous*
  - *Do the tests of significance for both outcome variables*
  - *Compare the results – which predictor variables show significant differences for which outcome variable and which do not:*
    - ✓ *Some are significant for both*
    - ✓ *Some are not significant for neither*
    - ✓ *Some are significant for one outcome but not the other*
  - *Do for combined outcome variable (completed the program and passed the NCLEX on the first attempt)*
- Procedure helped to gain better data understanding

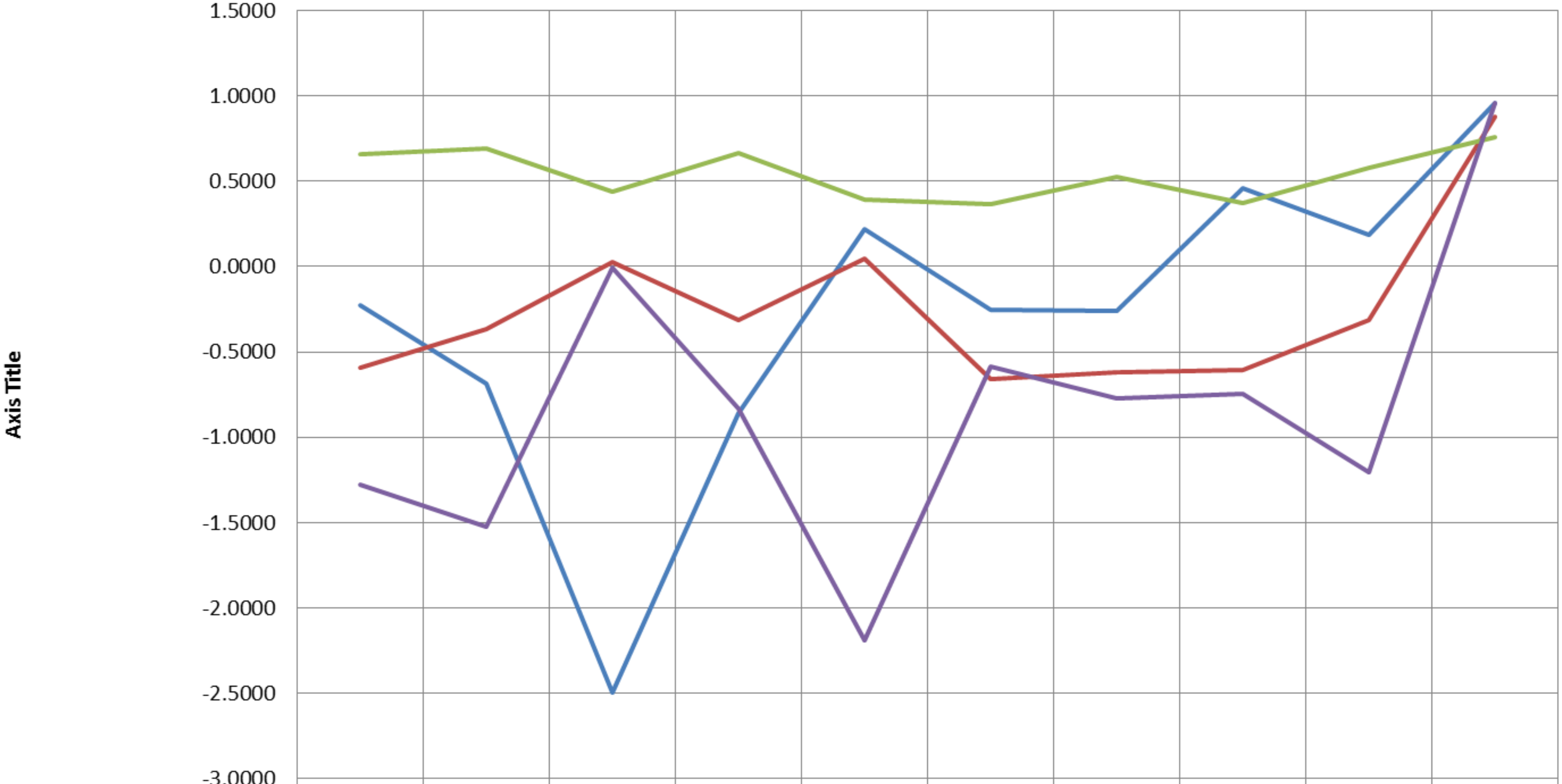
Pacific Union College

More Data Understanding: Classifying Variables and Cases

- Attempt to use *factor analysis* to extract factors and evaluate redundancy in variables
  - *Failed because of the missing data*
- Attempt to use *cluster analysis* to find possible similarities between the cases *(the set of characteristics makes a student's profile; cluster analysis is combining students into clusters)*
  - Do for both outcome variables
  - Use the prior knowledge of which of the variables matter
  - Compare the average passing rates for each cluster
  - See which variables make a bigger difference

Pacific Union College

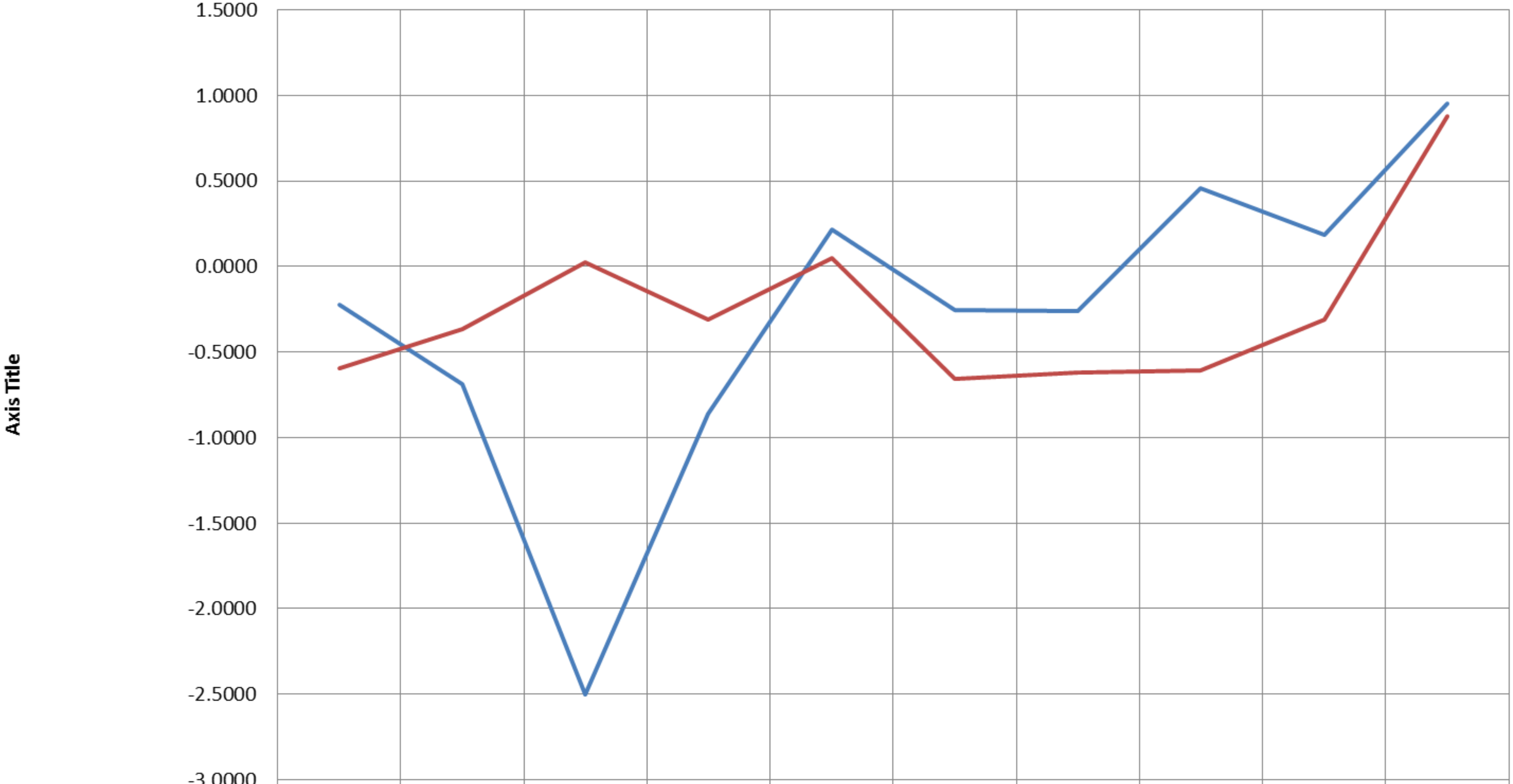
Distribution of the Means of Relevant Standardized Variables for First Four Program Completion Clusters



	Nursing GPA	IR Score	TEAS Total Score	ACT English Score	Intro to Nursing Grade	Anatomy Grade	Physiology Grade	Microbiolo gy Grade	Nutrition Grade	Number of Taken Classes
Cluster 1 (N=11) F=36%	-0.2261	-0.6857	-2.4987	-0.8591	0.2177	-0.2554	-0.2621	0.4556	0.1829	0.9550
Cluster 2 (N=85) F=18%	-0.5948	-0.3664	0.0229	-0.3134	0.0492	-0.6574	-0.6207	-0.6070	-0.3125	0.8767
Cluster 3 (N=63) F=3%	0.6572	0.6915	0.4390	0.6622	0.3936	0.3628	0.5221	0.3727	0.5793	0.7568
Cluster 4 (N=12) F=42%	-1.2801	-1.5268	-0.0105	-0.8307	-2.1878	-0.5846	-0.7736	-0.7478	-1.2028	0.9599

	Nursing GPA	IR Score	TEAS Total Score	ACT English Score	Intro to Nursing Grade	Anatomy Grade	Physiology Grade	Microbiology Grade	Nutrition Grade	Number of Taken Classes
Cluster 3 (N=63) F=3%	0.6572	0.6915	0.4390	0.6622	0.3936	0.3628	0.5221	0.3727	0.5793	0.7568
Cluster 4 (N=12) F=42%	-1.2801	-1.5268	-0.0105	-0.8307	-2.1878	-0.5846	-0.7736	-0.7478	-1.2028	0.9599

Distribution of the Means of Relevant Standardized Variables for Program Completion Clusters 1 and 2



	Nursing GPA	IR Score	TEAS Total Score	ACT English Score	Intro to Nursing Grade	Anatomy Grade	Physiology Grade	Microbiology Grade	Nutrition Grade	Number of Taken Classes
Cluster 1 (N=11) F=36%	-0.2261	-0.6857	-2.4987	-0.8591	0.2177	-0.2554	-0.2621	0.4556	0.1829	0.9550
Cluster 2 (N=85) F=18%	-0.5948	-0.3664	0.0229	-0.3134	0.0492	-0.6574	-0.6207	-0.6070	-0.3125	0.8767

More Data Preparation: Final Data Decisions

- Choice of Outcome (Target) Variable
  - *Combined outcome – completed the program AND passed the NCLEX on the first attempt (409 cases)*
- Choice of Predictor Variables *(possible criteria: significance, missing values). Decided to remove:*
  - IR Score (replacing)
  - Component TEAS (not available)
  - Math and Chemistry (Pass or HS entries)
  - Intro to Nursing (missing values)
  - Human Development (missing values)
- Choice of cases
  - Out of 409 cases, 194 were complete with remaining variables (44 of them were failing cases)
  - Balanced: 44 failing + 44 passing (based upon random selection); the rest of the cases used for validation

Pacific Union College

Data Modeling: Discriminant Analysis

- Why Discriminant Analysis?
  - Classic method which has stood the test of time
  - Often produces models not inferior to modern methods
  - More importantly: provides discriminant scores which are easy to interpret and use independently from analysis software
- Based on simple idea
  - Linear combination of initial variables

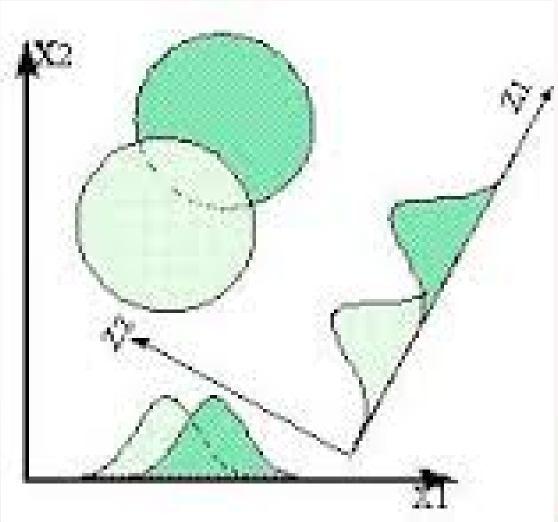


Image Source: http://www.ict-m.com/ictm/public/Applications/Optimization/Multivariate/default.aspx

Discriminant Analysis: Classification Success

Classification Results[a,b]

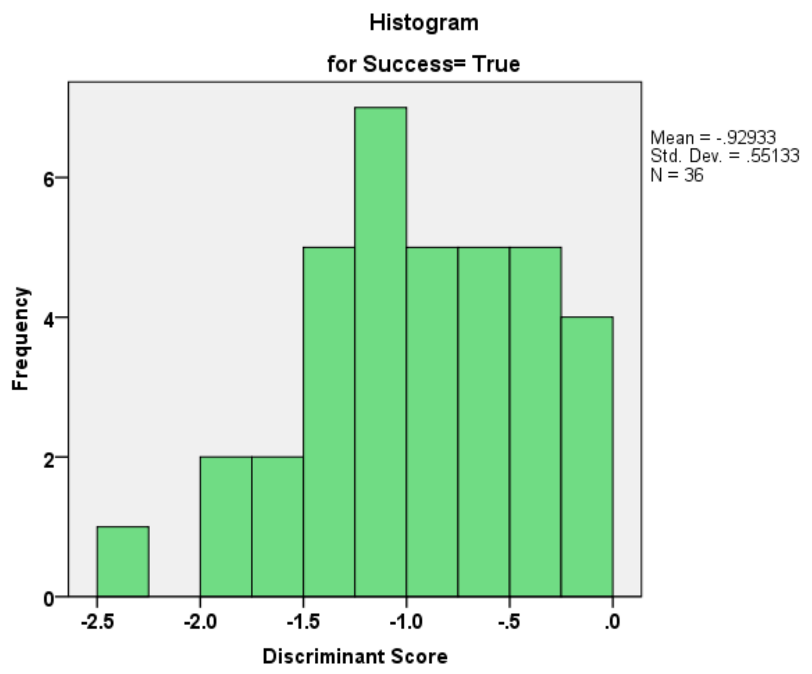
			Combined Outcome	Predicted Group Membership: Fail	Predicted Group Membership: Pass	Total
Cases Selected	Original	Count	Fail	34	10	44
			Pass	13	31	44
		%	Fail	77.3	22.7	100.0
			Pass	29.5	70.5	100.0
Cases Not Selected	Original	Count	Fail	2	1	3
			Pass	41	70	111
		%	Fail	66.7	33.3	100.0
			Pass	36.9	63.1	100.0

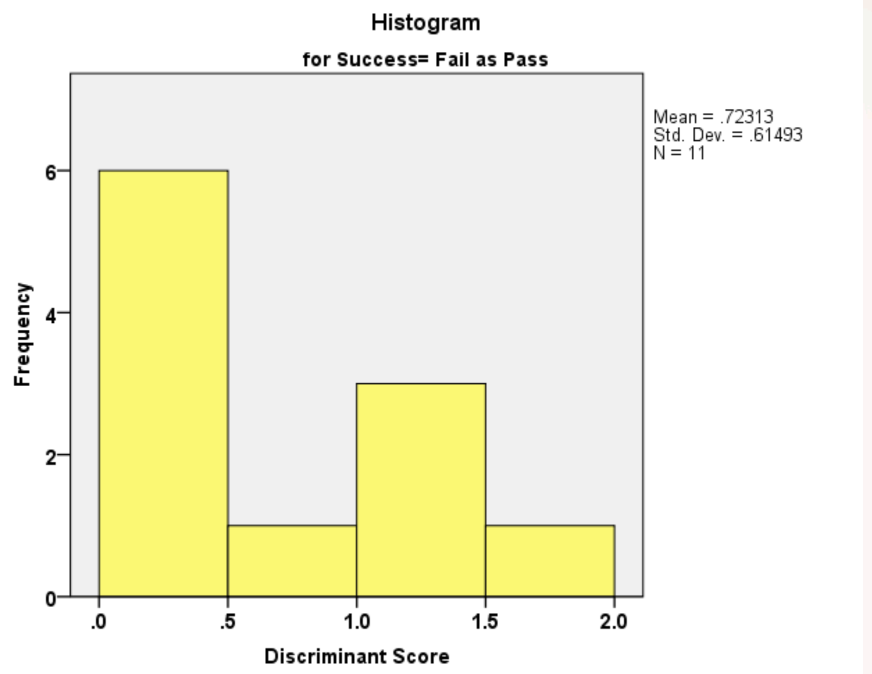
a. 73.9% of selected original grouped cases correctly classified.

b. 63.2% of unselected original grouped cases correctly classified.

Pacific Union College

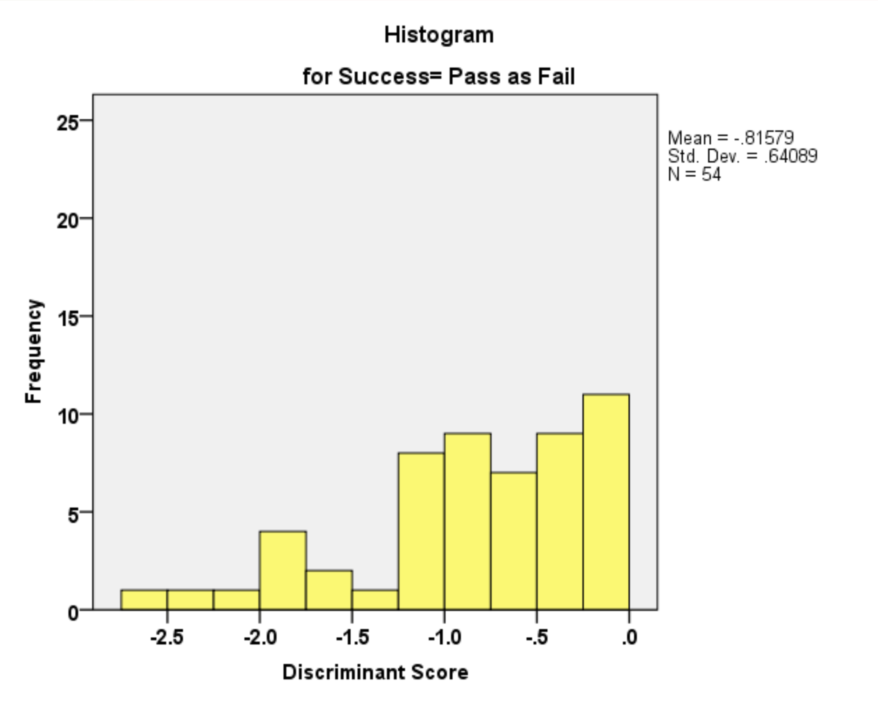
Discriminant Scores Analysis: Failing Cases

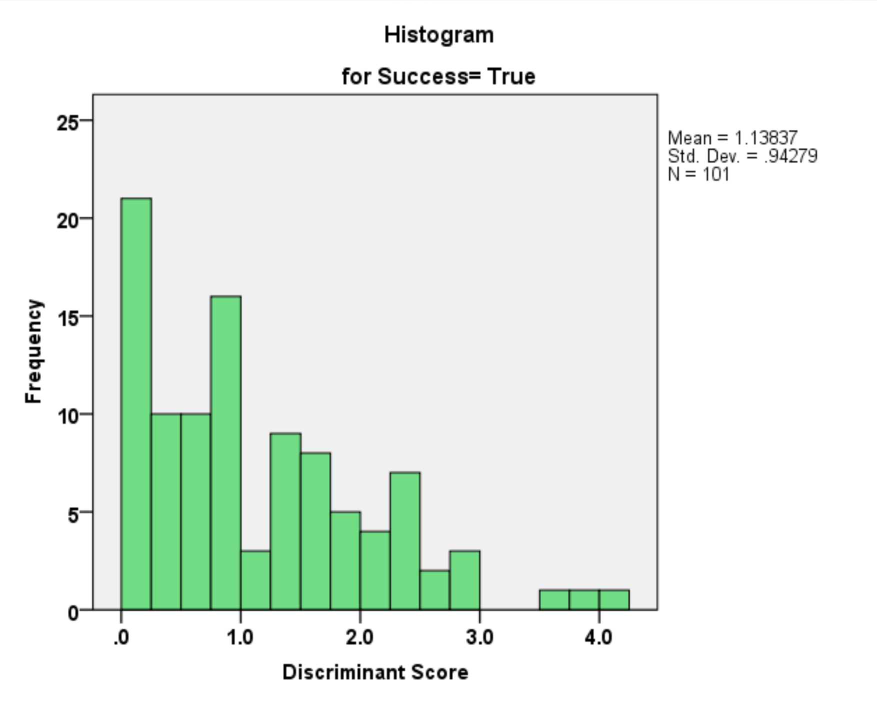




		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fail as Pass	11	2.4	5.4	5.4
	True	137	30.4	67.8	73.3
	Pass as Fail	54	12.0	26.7	100.0
	Total	202	44.8	100.0	
Missing	System	249	55.2		
Total		451	100.0		

Discriminant Scores Analysis: Passing Cases





		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fail as Pass	11	2.4	5.4	5.4
	True	137	30.4	67.8	73.3
	Pass as Fail	54	12.0	26.7	100.0
	Total	202	44.8	100.0	
Missing	System	249	55.2		
Total		451	100.0		

Data Modeling: Single Decision Tree (SDT)

- **What is a Decision Tree?**
  - *Logical model represented as two-way split tree that shows how the value of a target variable can be predicted by a series of splits controlled by the values of predictor variables*
  - **Two decisions are made for each split :**
    - *What would be the "splitting variable?"*
    - *What would be the "split point (value)?"*

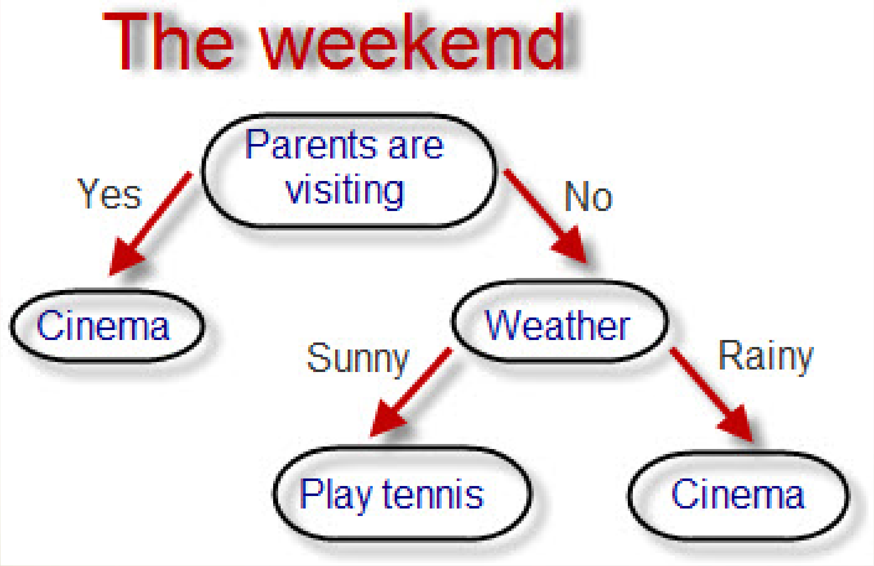


Image Source: http://wiki.bizagi.com/en/index.php?title=Policiy_Rule-Decision_Table-Group_And_Precondition

Pacific Union College

Single Decision Tree: Classification Success*

============ Misclassification Tables ============

--- Training Data ---

Category	Actual Count	Actual Weight	Misclassified Count	Misclassified Weight	Misclassified Percent	Misclassified Cost
Fail	44	44	11	11	25.000	0.250
Pass	44	44	8	8	18.182	0.182
Total	88	88	19	19	21.591	0.216

Overall accuracy = 78.41%

--- Validation Data ---

Category	Actual Count	Actual Weight	Misclassified Count	Misclassified Weight	Misclassified Percent	Misclassified Cost
Fail	44	44	17	17	38.636	0.386
Pass	44	44	12	12	27.273	0.273
Total	88	88	29	29	32.955	0.330

Overall accuracy = 67.05%

* *V-fold cross validation was used*

Pacific Union College

Single Decision Tree: Classification Success For Complete Dataset

Only four variables included:

GPA	100%
ACT English Score	73.4%
Speech Grade	40.2%
TEAS Total Score	30.4%

Count

		Analyze? No	Analyze? Yes	Total
Combined Outcome	Fail	48	44	92
	Pass	273	44	317
Total		321	88	409

Total records = 409

Pass/Fall ratio = 3.45

Accuracy = 66.50%

True Pass (TP) = 209 (51.1%)

True Fail (TF) = 63 (15.4%)

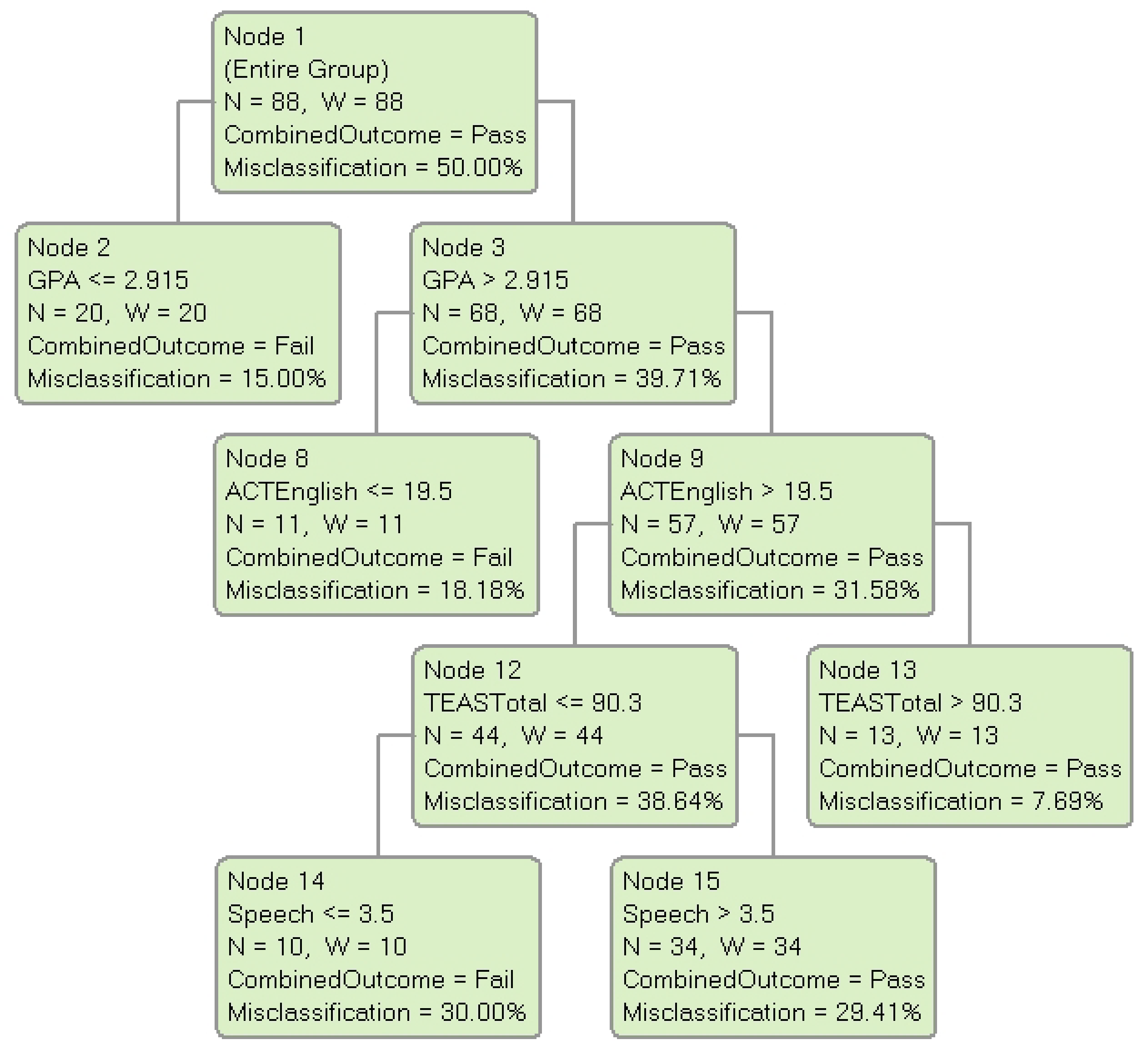
False Pass (FP) = 29 (7.1%)

False Fail (FF) = 108 (26.4%)

Sensitivity = 65.93%

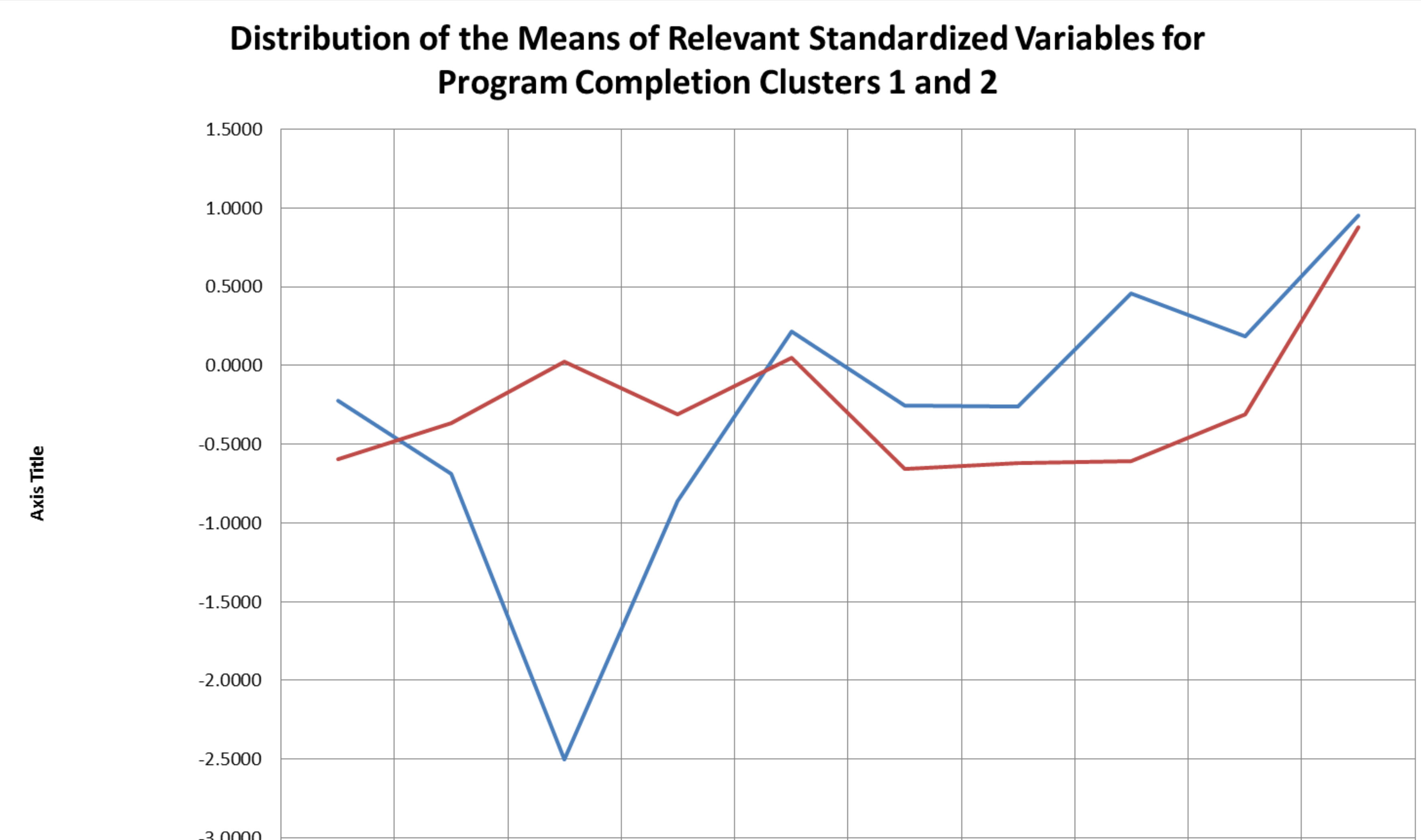
Specificity = 68.48%

Single Decision Tree: A Flowchart



Pacific Union College

SDT: Comparing Results with Cluster Analysis



	Nursing GPA	IR Score	TEAS Total Score	ACT English Score	Intro to Nursing Grade	Anatomy Grade	Physiology Grade	Microbiology Grade	Nutrition Grade	Number of Taken Classes
Cluster 1 (N=11) F=36%	-0.2261	-0.6857	-2.4987	-0.8591	0.2177	-0.2554	-0.2621	0.4556	0.1829	0.9550
Cluster 2 (N=85) F=18%	-0.5948	-0.3664	0.0229	-0.3134	0.0492	-0.6574	-0.6207	-0.6070	-0.3125	0.8767

Models Evaluation: Comparing Misclassifications

- **Do two models misclassify the same cases?**

			SDT Classification Result			
			Fail as Pass	True	Pass as Fail	Total
DA Classification Result	Fail as Pass	Count	8	3	0	11
		% within DA Classification Result	72.7%	27.3%	.0%	100.0%
	True	Count	4	112	21	137
		% within DA Classification Result	2.9%	81.8%	15.3%	100.0%
	Pass as Fail	Count	0	18	36	54
		% within DA Classification Result	.0%	33.3%	66.7%	100.0%
Total		Count	12	133	57	202
		% within DA Classification Result	5.9%	65.8%	28.2%	100.0%

Pacific Union College

Models Evaluation: Comparing Misclassifications

- **Do two models misclassify the same cases?**

			SDT Classification Result			
			Fail as Pass	True	Pass as Fail	Total
DA Classification Result	Fail as Pass	Count	8	3	0	11
		% within DA Classification Result	72.7%	27.3%	.0%	100.0%
	True	Count	4	112	21	137
		% within DA Classification Result	2.9%	81.8%	15.3%	100.0%
	Pass as Fail	Count	0	18	36	54
		% within DA Classification Result	.0%	33.3%	66.7%	100.0%
Total		Count	12	133	57	202
		% within DA Classification Result	5.9%	65.8%	28.2%	100.0%

Agreement between models: (8+112=36)/202 = 77.2%

Pacific Union College

Models Evaluation: Comparing Misclassifications

- **Do two models misclassify the same cases?**

			SDT Classification Result			
			Fail as Pass	True	Pass as Fail	Total
DA Classification Result	Fail as Pass	Count	8	3	0	11
		% within DA Classification Result	72.7%	27.3%	.0%	100.0%
	True	Count	4	112	21	137
		% within DA Classification Result	2.9%	81.8%	15.3%	100.0%
	Pass as Fail	Count	0	18	36	54
		% within DA Classification Result	.0%	33.3%	66.7%	100.0%
Total		Count	12	133	57	202
		% within DA Classification Result	5.9%	65.8%	28.2%	100.0%

Pacific Union College

Models Evaluation: Comparing Misclassifications

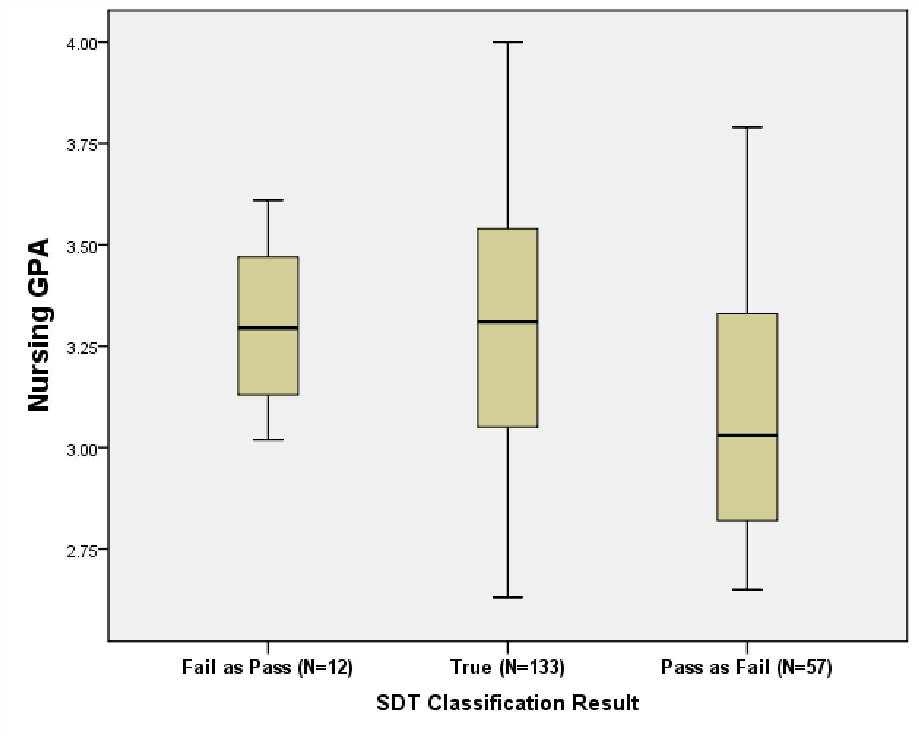
- **Do two models misclassify the same cases?**

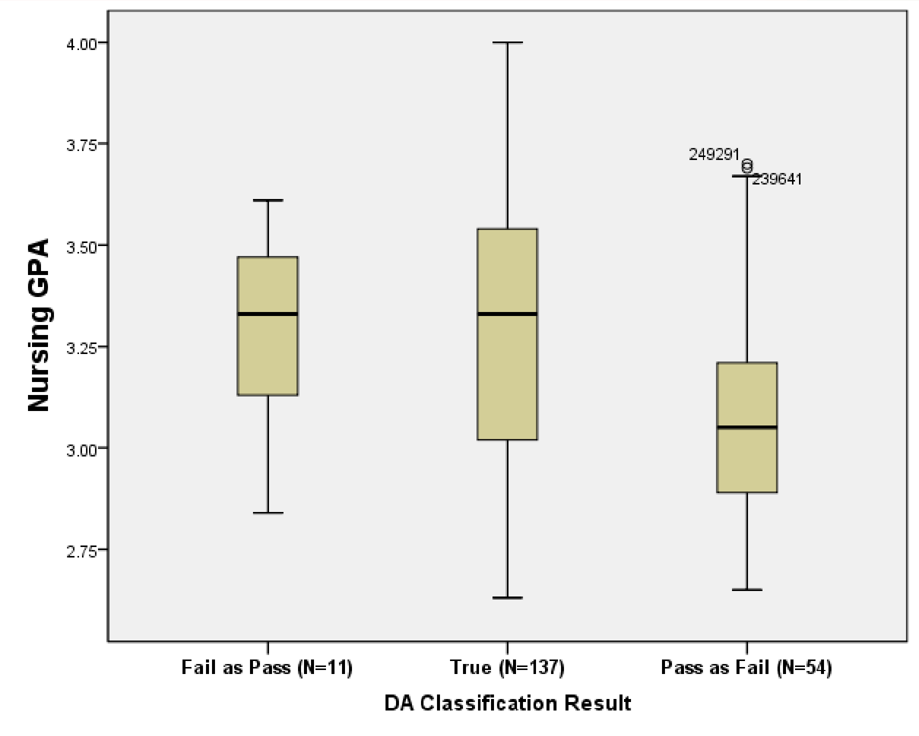
			SDT Classification Result			Total
			Fail as Pass	True	Pass as Fail	
DA Classification Result	Fail as Pass	Count	8	3	0	11
		% within DA Classification Result	72.7%	27.3%	.0%	100.0%
	True	Count	4	112	21	137
		% within DA Classification Result	2.9%	81.8%	15.3%	100.0%
	Pass as Fail	Count	0	18	36	54
		% within DA Classification Result	.0%	33.3%	66.7%	100.0%
Total		Count	12	133	57	202
		% within DA Classification Result	5.9%	65.8%	28.2%	100.0%

Agreement between the models matches or exceeds the agreement between the models and the reality

Pacific Union College

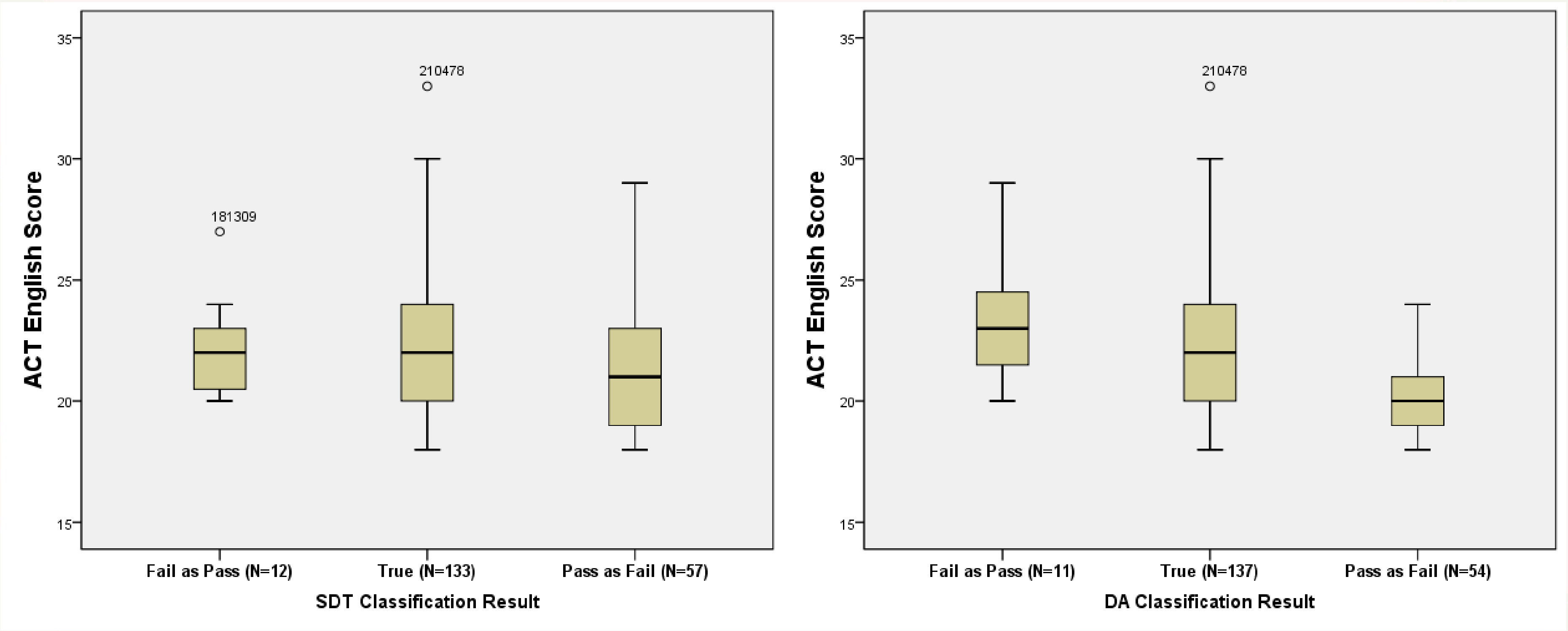
Models Evaluation: Why Misclassifying?





Boxplots for variable Nursing GPA in three classification groups for DA and SDT models

Models Evaluation: Why Misclassifying?



Boxplots for variable ACT English Score in three classification groups for DA and SDT models

Pacific Union College

Models Evaluation: Conclusions

- *Discriminant scores* could be used in place of the outdated IR Score as the objective success predictor scores. However, there is a rather big “grey area” between scores of -1 and 1. In such cases the Admission Committee should use other considerations.
- The *Single Decision Tree* model provides a useful alternative. This model may also suggest some cutout values for published admissions policies.
- *Other data mining procedures* gave similar classification accuracies. This suggests that predictive power is determined in much greater degree by the character of the data than by the choice of a model.
- The *accuracy* would be higher if applied to all applicants; however, this cannot be verified because there would be no completion data for those not accepted.
- *Model deployment* would be the ultimate evaluation if we are to see the higher rates of students’ success several years down the road as the models are used in the admission process.

Pacific Union College

Admission to Nursing – Changes?

- Discriminant Analysis
  - *Could replace Institutional Research (IR) score*
- Single Decision Tree (SDT)
  - *Simple, most elements available*
  - *Suggests changes to criteria*
- Minimum cognate/GE GPA – currently 2.7
  - *Change to 3.0*
- Repeats for failure – currently limited to two
  - *Consider removing as absolute criterion*

Pacific Union College

Pacific Union College

Shana Ruggenberg
Nursing & Health Sciences
sruggenberg@puc.edu

Serhii Kalynovs'kyi
Institutional Research
skalynovskyi@puc.edu