Managing “At risk” business students:
Statistical analysis of student data profiles

Christo Potgieter
Wilfred Greyling
Vida Botes
Overview

• TEC/MoE funding $\rightarrow$ Student Outcomes
  $\rightarrow$ Pro-active management, powerful tools

• McCarthy/Scott (2005), Potgieter/Ferguson/Roberton (2009), Potgieter/Greyling/Ferguson (2010), etc

• THIS: Statistical model re early predicting pass/fail

• Reflection and further research
Planning empirical explorations

1. To what extent do the **first two assessments** in a module allow us to predict final course outcome?

2. How accurately can one predict the final outcome **after each of the first and second assessments**?

3. How significant in size is the number of students whose **final results differ from initial assessment scores**?

4. Is the level of **incorrect predictions** from a specific statistical model acceptably low?

5. To what extent do the initial two assessments discriminate between the selected **two qualifications**?
Population sampling

• **BBS and GradDip**
  – Share some classes
  – GradDip: Most students from overseas (mostly India)
  – GradDip: All have degrees, most in Business
  – BBS +-40% *international students*
  – BBS includes China pathway students

• Therefore:
  – Can compare student cohorts
  – Contributes to research re international students
Processing

- Data extracted from student database for 2009.
- Semesters one and two, as well as Summer School.
- **Remove records** with any empty fields (Missed assessments, Students withdrew, Recording method, etc) and those with only one assessment and the exam ($<2\%$)
- Retained marks for two assessments and final outcome (pass/fail)
## Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ass1_M</td>
<td>61.479</td>
<td>17.4199</td>
</tr>
<tr>
<td>Ass2_M</td>
<td>68.493</td>
<td>14.5925</td>
</tr>
<tr>
<td>Fin_M</td>
<td>64.827</td>
<td>10.4930</td>
</tr>
</tbody>
</table>
## Classification results

(a. 83.3% of original grouped cases correctly classified)
## Model summary: Prediction after first and second assessment

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.609(^a)</td>
<td>.371</td>
<td>.371</td>
<td>8.3244</td>
</tr>
<tr>
<td>2</td>
<td>.725(^b)</td>
<td>.525</td>
<td>.525</td>
<td>7.2351</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Ass1_M

b. Predictors: (Constant), Ass1_M, Ass2_M

c. Dependent Variable: Fin_M
Tentative start?

• F, Beta and T values shows model appears to be ok

• Using the **first assessment** as predictor...the model explains **37.1%** of the variance on the course outcome

• With the first **two assessments** as predictors...the model explains **52.5%** of the variance on the course outcome
Re research questions (1/2)

1. To what extent do the first two assessments in a module allow us to predict final course outcome? 52%

2. How accurately can one predict the final outcome after each of the first and second assessments? 37% → 52%

3. How significant in size is the number of students whose final results differ from initial assessment scores? 17% (only 0.5% Pass → Fail)
4. Is the level of incorrect predictions from a specific statistical model acceptably low? **YES**

5. To what extent do the initial two assessments discriminate between the selected two qualifications? **Insignificant**

Final outcome much same as first assessment…

*Do we really need 2\textsuperscript{nd}/3\textsuperscript{rd} assessments and exams???

(Thinking like cost-saving manager…)*
Possible improvements (1/4)

• Were statistical methods and software used in suitable fashion, considering the complexity of the methods and tools and the implications of conclusions?

• Would conclusions be the same for years other than the 2009 sample?

• Would a module-based analysis be meaningful and what would such an analysis yield? For example, are there any modules where the first two assignments are not useful predictors of the final outcome?
Possible improvements (2/4)

• Do courses with traditional high failure rates, perhaps ones such as Accounting and Law, also display a pattern similar to the overall trend?

• BBS is a multi-year multiple-specialization degree programme – is the pattern the same at each level of year of study?

• Do international students perhaps progress differently from domestic students (for yet unidentified reasons)?
Possible improvements (3/4)

• Is the pattern for international students under **pathway arrangements** (i.e. starting here with second and third year modules) any different?

• Would the statistical findings be any different if we only looked at students **doing a course for the first time** (i.e. excluding students repeating modules after having failed)?

• Is there possibly a (yet unidentified) group of students **causing most failures** and how should they be supported?
Possible improvements (4/4)

• Do GradDip students with an earlier degree in a business field perform different from other students?

• Does the pattern of progression for students change over the study period, especially when doing the multi-year degree?

• How does this compare to other qualifications (analysis done on IT students, 2010) and also other institutions?

Better fit for Business – WHY???
Conclusion

• Concept of using sophisticated statistical tool on unverified historical data in limited time range is **challenging**:
  – Data is often messier than anticipated
  – Statistical analysis is rather complex
  – Organization practices not focussed to add value

• Observation of statistics and experience
  – Enables us to more comprehensively plan future initiatives to improve student outcomes as expected by stakeholders

• Good learning experience to **plan more mature projects**.

• **But is it worth the effort??????????????????? (Considering interventions by tutors during semester...)**
CHEERS!