

# Using statistical models to predict second year mathematics and physics success

Dr Lizelle Fletcher

Department of Statistics

University of Pretoria



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

Denkleiers • Leading Minds • Dikgopolo tša Dihalefi

# Outline

- Problem statement
- Data and variables
- Statistical techniques
- Mathematics
- Physics
- Conclusion
- Acknowledgements

# Problem statement

- To what extent can **achievement** in **1<sup>st</sup> year** Maths/Physics be used to **understand** students' performance in their **2<sup>nd</sup> year** Math/Physics modules?

➤ Focus on *Success* →

- Which 1<sup>st</sup> year modules explain 2<sup>nd</sup> year success?
- Any other student information?

1 ≡ Success	> 55
2 ≡ At risk	$45 \leq \text{WTW2..} \leq 55$
1 ≡ Failure	< 45

# Mathematics: Data and variables

## Five 2nd year modules

WTW211, WTW218, WTW220, WTW221 and WTW248

### Period

2014 – 2016

## 1<sup>st</sup> year modules

WTW114, WTW126 and WTW128

### Period

2011 - 2015

1 ≡ Success	> 55
2 ≡ At risk	$45 \leq \text{WTW2..} \leq 55$
1 ≡ Failure	< 45

Only *first* attempt marks are used

### Additional information

- Grade 12 Maths mark
- Gender
- Language of instruction

# Physics: Data and variables

## Two 2nd year modules

PHY253 and PHY263

### Period

2014 – 2016

## 1<sup>st</sup> year modules

PHY114, PHY124, WTW114, WTW126 and WTW128

### Period

2011 - 2015

1 ≡ Success	> 55
2 ≡ At risk	$45 \leq \text{WTW2..} \leq 55$
1 ≡ Failure	< 45

Only *first* attempt marks are used

## Additional information

- Grade 12 Maths mark
- Grade 12 Physical Science mark
- Gender
- Language of instruction

# Statistical techniques

- 1) CHAID → Data segmentation → Dendogram (tree)
- 2) Logistic regression models
- 3) Cross-tabulations

# PHY 255

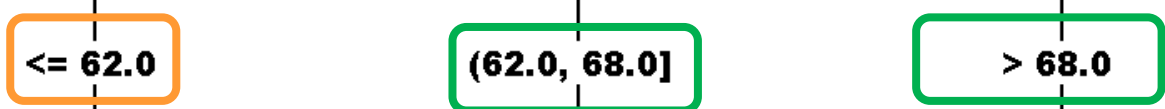
## PHY255\_1st\_rec50

■	Success
■	Risk
■	Fail

Node 0		
Category	%	n
Success	56.7	72
Risk	30.7	39
Fail	12.6	16
<b>Total</b>	<b>100.0</b>	<b>127</b>

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

PHY124\_first  
Adj. P-value=0.000, Chi-square=66.432, df=6



**<= 62.0**

Node 1		
Category	%	n
Success	28.6	12
Risk	59.5	25
Fail	11.9	5
<b>Total</b>	<b>33.1</b>	<b>42</b>

**(62.0, 68.0]**

Node 2		
Category	%	n
Success	71.4	15
Risk	28.6	6
Fail	0.0	0
<b>Total</b>	<b>16.5</b>	<b>21</b>

**> 68.0**

Node 3		
Category	%	n
Success	93.2	41
Risk	2.3	1
Fail	4.5	2
<b>Total</b>	<b>34.6</b>	<b>44</b>

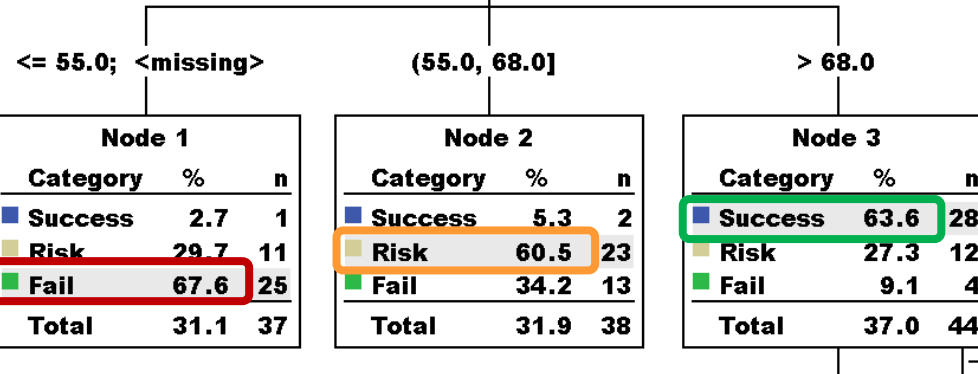
# PHY 263

PHY263\_1st\_rec50

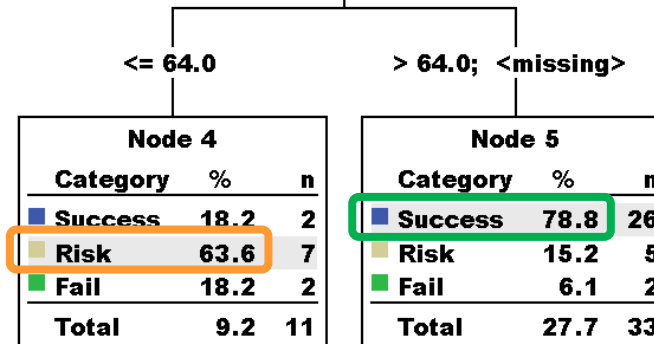
■ Success
■ Risk
■ Fail

Node 0		
Category	%	n
Success	26.1	31
Risk	38.7	46
Fail	35.3	42
<b>Total</b>	<b>100.0</b>	<b>119</b>

PHY124\_first  
Adj. P-value=0.000, Chi-square=64.338, df=4





PHY114 first  
Adj. P-value=0.020, Chi-square=13.206, df=2



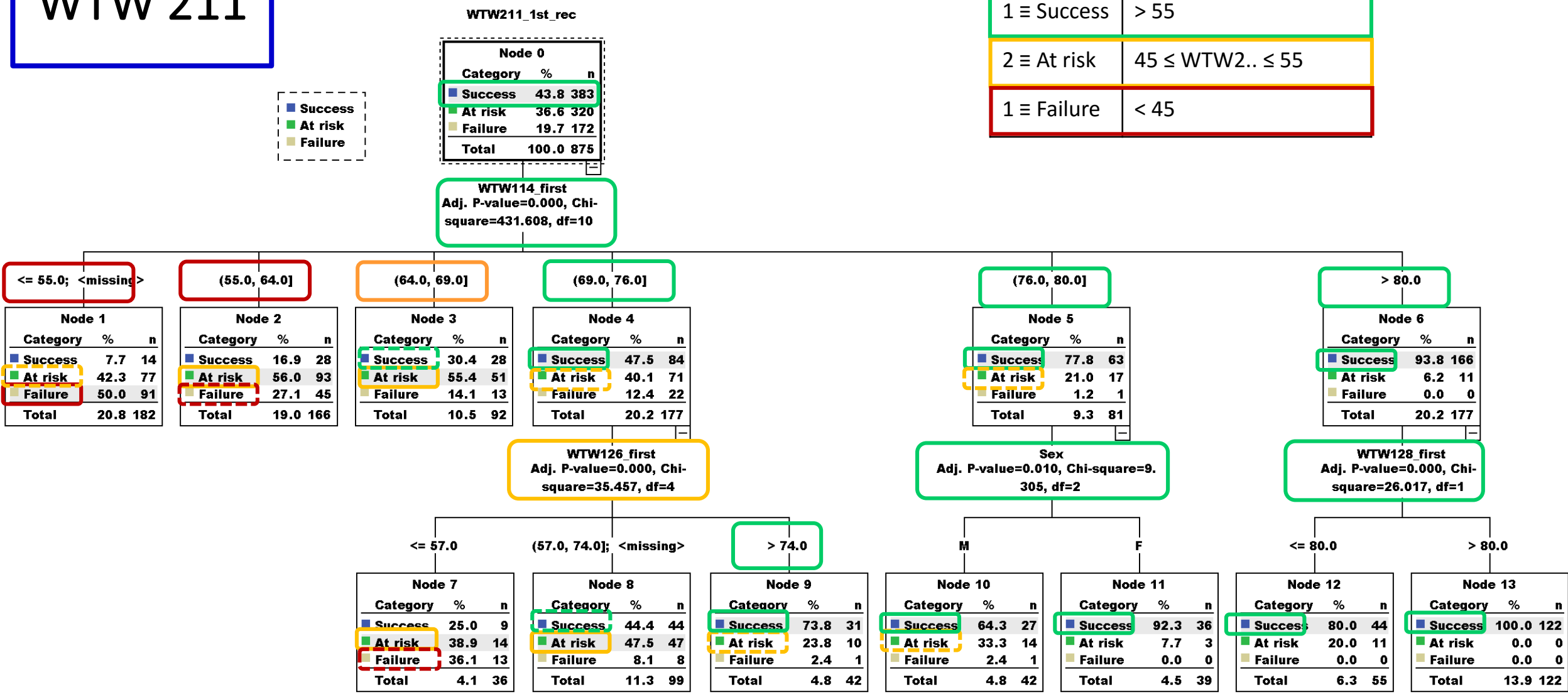
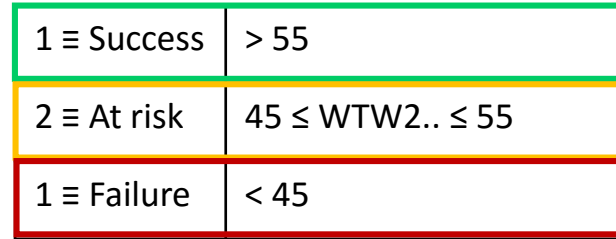
1 $\equiv$ Success	$> 55$
2 $\equiv$ At risk	$45 \leq \text{WTW2..} \leq 55$
1 $\equiv$ Failure	$< 45$



# Physics: Data and variables

PHY 255		PHY 124
PHY 263		PHY 124

# WTW 211



# WTW 211 snip

WTW211\_1st\_rec

■	Success
■	At risk
■	Failure

Node 0		
Category	%	n
Success	43.8	383
At risk	36.6	320
Failure	19.7	172
<b>Total</b>	<b>100.0</b>	<b>875</b>

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

WTW114\_first  
Adj. P-value=0.000, Chi-square=431.608, df=10

<= 55.0; <missing>

(55.0, 64.0]

(64.0, 69.0]

(69.0, 76.0]

Node 1		
Category	%	n
Success	7.7	14
At risk	42.3	77
Failure	50.0	91
<b>Total</b>	<b>20.8</b>	<b>182</b>

Node 2		
Category	%	n
Success	16.9	28
At risk	56.0	93
Failure	27.1	45
<b>Total</b>	<b>19.0</b>	<b>166</b>

Node 3		
Category	%	n
Success	30.4	28
At risk	55.4	51
Failure	14.1	13
<b>Total</b>	<b>10.5</b>	<b>92</b>

Node 4		
Category	%	n
Success	47.5	84
At risk	40.1	71
Failure	12.4	22
<b>Total</b>	<b>20.2</b>	<b>177</b>

WTW126\_first

Level 1

Level 2

# WTW 218

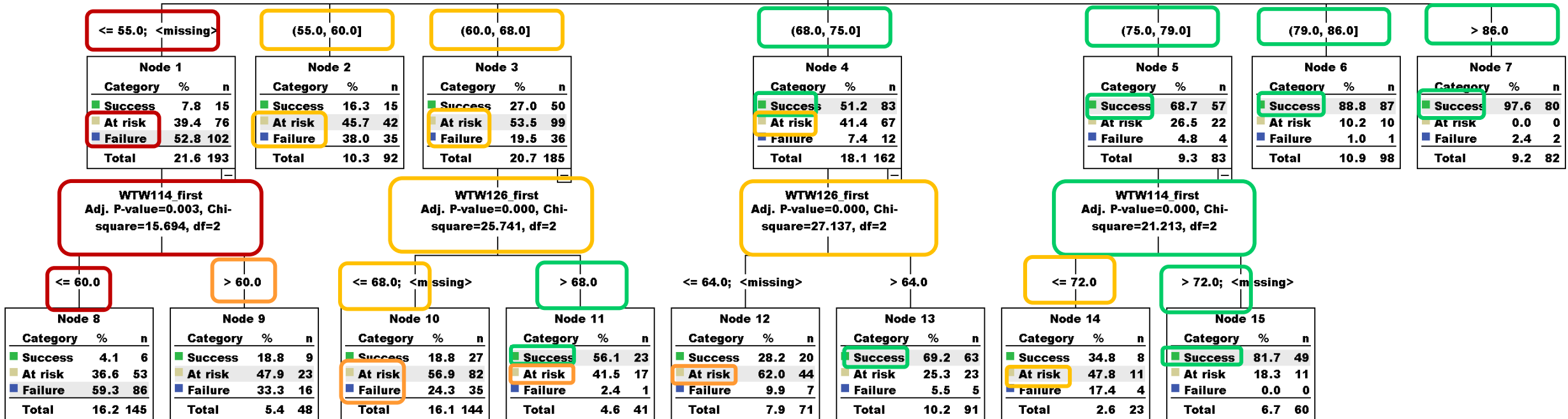
1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

WTW218\_1st\_rec

Node 0		
Category	%	n
Success	43.2	387
At risk	35.3	316
Failure	21.5	192
Total	100.0	895

■ Success
■ At risk
■ Failure

WTW128\_first  
Adj. P-value=0.000, Chi-square=430.927, df=12



# WTW 220

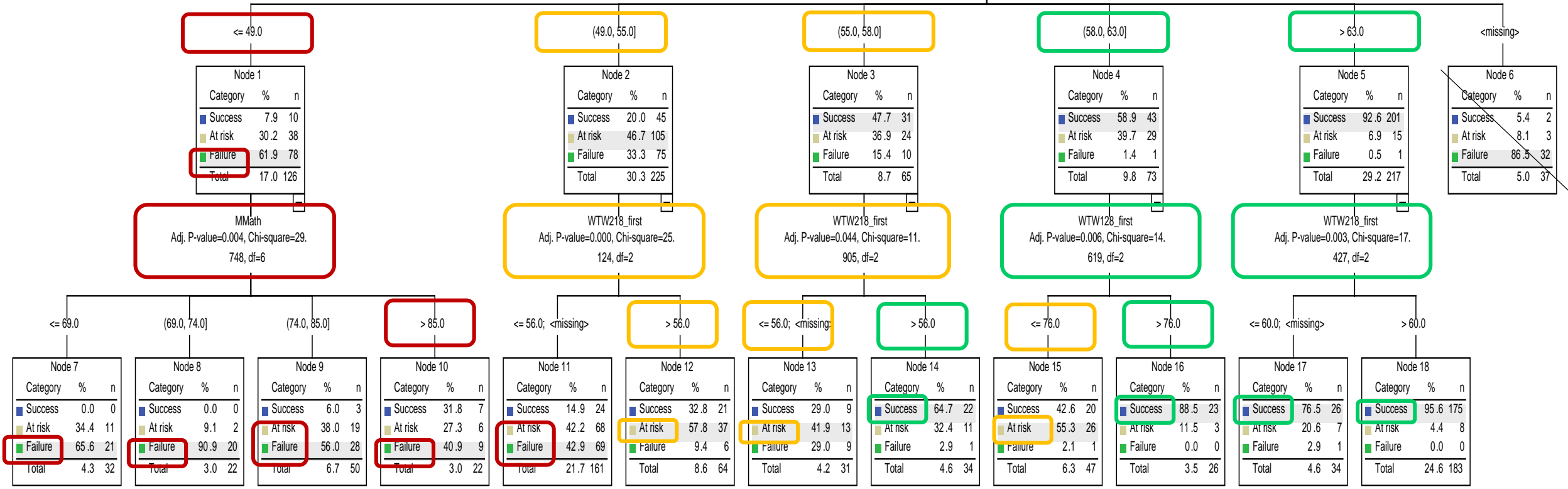
WTW220\_1st\_rec

Node 0		
Category	%	n
Success	44.7	332
At risk	26.8	214
Failure	26.5	197
Total	100.0	743

- Success
- At risk
- Failure

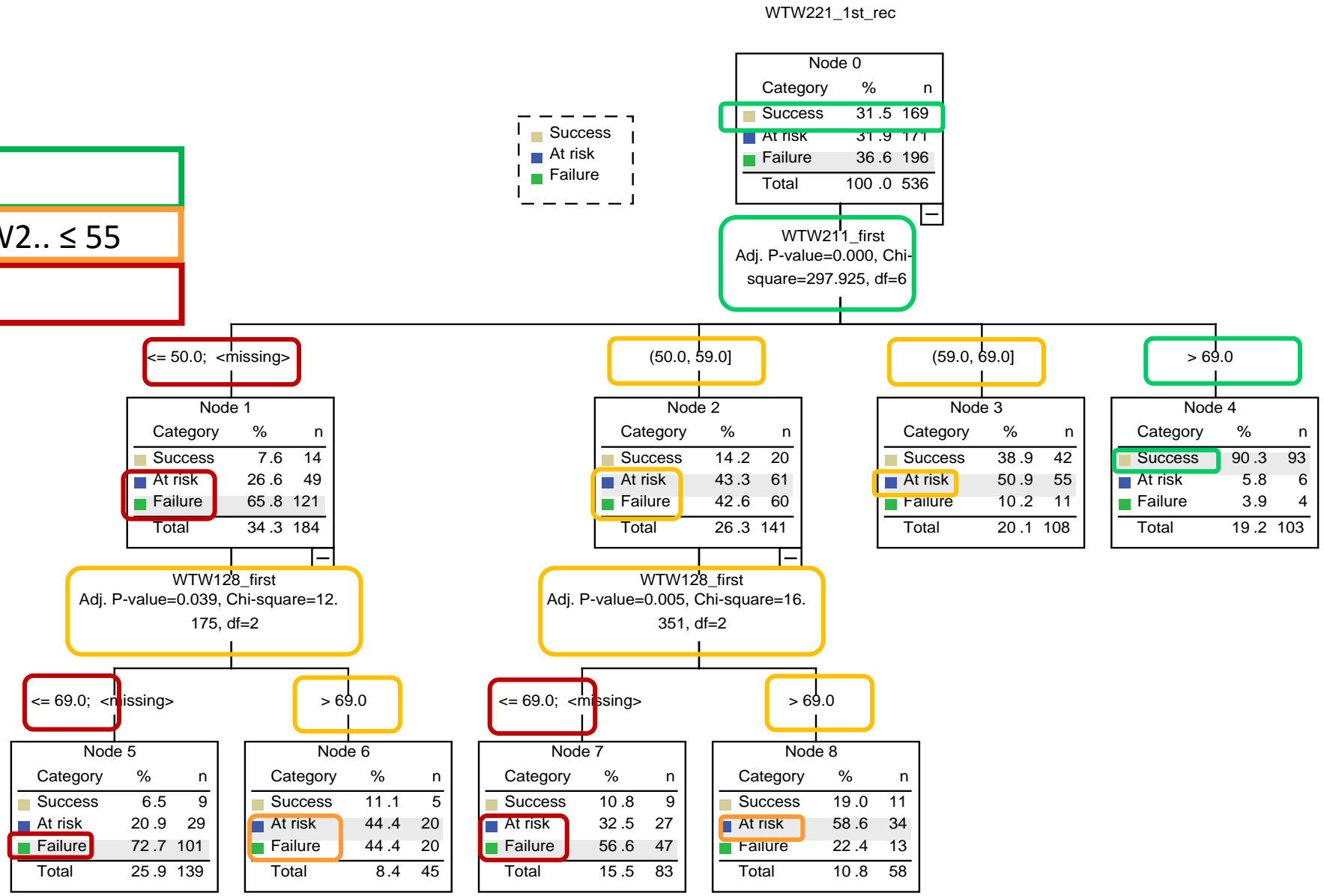
1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

WTW211\_first  
Adj. P-value=0.000, Chi-square=457.374, df=10



# WTW 221

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45



# WTW 248

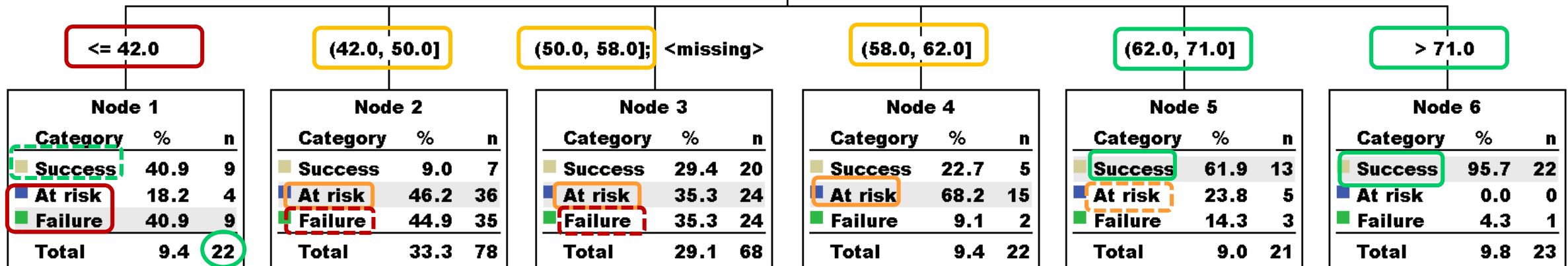
WTW248\_1st\_rec

■	Success
■	At risk
■	Failure






Node 0		
Category	%	n
Success	32.5	76
At risk	35.9	84
Failure	31.6	74
<b>Total</b>	<b>100.0</b>	<b>234</b>

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

WTW218\_first  
Adj. P-value=0.000, Chi-square=84.323, df=10



# Mathematics: Data and variables

WTW 211		WTW 114
WTW 218		WTW 128
WTW 220		WTW 211
WTW 221		WTW 211
WTW 248		WTW 218



# (Very) brief regression lesson

Simple regression:  $Y = b_0 + b_1X$

Multiple regression:  $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$

Logistic regression:  $\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$

$p = P(\text{Success})$  i. e.  $\frac{p}{1-p} = \text{Odds of a success}$

# Mathematics: multinomial logistic regression

Dependent variable	WTW211_1st_rec	WTW220_1st_rec			
	n=875	n=743			
Predictors	n=795				
Sex	0.021				
Language	0.074				
Mmath	0.097				
WTW114_first	<0.00001				
WTW126_first	<0.001				
WTW128_first	<0.0001				
WTW211_first					
WTW218_first					

# Descriptive statistics

		WTW126_128	
		1 Safe	2 Not
WTW220_first	N	619	425
	Minimum	22	14
	Maximum	98	83
	Mean	64.58	48.69
	Std. Deviation	13.933	12.731
WTW248_first	N	131	181
	Minimum	14	11
	Maximum	93	92
	Mean	59.06	48.89
	Std. Deviation	14.938	12.312

WTW 126-128	1 = Safe	(WTW 126 $\geq$ 60) & (WTW 126 $\geq$ 60)
Categories	2 = Not safe	Rest

# Crosstabulations

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

			WTW126_128		Total
			1 Safe	2 Not	
WTW220_1st_rec	1 Success	Count	441	120	561
		% within WTW126_128	70.3%	25.5%	51.1%
	2 At risk	Count	138	162	300
		% within WTW126_128	22.0%	34.4%	27.3%
	3 Failure	Count	48	189	237
		% within WTW126_128	7.7%	40.1%	21.6%
Total	Count	627	471	1098	
	% within WTW126_128	100.0%	100.0%	100.0%	

# Crosstabulations

1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

WTW248_1st_rec * WTW126_128 Crosstabulation			WTW126_128		Total
			1 Safe	2 Not	
WTW248_1st_rec	1 Success	Count	71	51	122
		% within WTW126_128	<u>53.8%</u>	26.8%	37.9%
	2 At risk	Count	43	72	115
		% within WTW126_128	<u>32.6%</u>	37.9%	35.7%
	3 Failure	Count	18	67	85
		% within WTW126_128	<u>13.6%</u>	35.3%	26.4%
Total	Count	132	190	322	
	% within WTW126_128	100.0%	100.0%	100.0%	

# PHY 255: multinomial logistic regression

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	85.953 <sup>a</sup>	.000	0	.
MMath	87.652	1.699	2	.428
MPhysSci	91.763	5.810	2	.055
PHY114_first	89.158	3.205	2	.201
PHY124_first	105.084	19.131	2	.000
WTW114_first	87.538	1.585	2	.453
Sex	86.348	.395	2	.821

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

# Physics:

## Parameter Estimates

PHY255_1st_rec50 <sup>a</sup>		B	Std. Error	Wald	df	Sig.	Exp(B)
Risk	Intercept	16.503	4.107	16.149	1	.000	
	MMath	-.006	.050	.017	1	.896	.994
	MPhysSci	-.111	.050	4.977	1	.026	.895
	PHY114_first	.068	.050	1.851	1	.174	1.070
	PHY124_first	-.190	.054	12.114	1	.001	.827
	WTW114_first	-.008	.047	.032	1	.858	.992
	[Sex=0]	-.403	.742	.295	1	.587	.668
	[Sex=1]	0 <sup>b</sup>	.	.	0	.	.
Fail	Intercept	9.605	5.869	2.678	1	.102	
	MMath	-.106	.085	1.540	1	.215	.899
	MPhysSci	-.080	.095	.719	1	.396	.923
	PHY114_first	.127	.099	1.646	1	.200	1.136
	PHY124_first	-.188	.102	3.423	1	.064	.828
	WTW114_first	.100	.090	1.244	1	.265	1.105
	[Sex=0]	-.555	1.316	.178	1	.673	.574
	[Sex=1]	0 <sup>b</sup>	.	.	0	.	.

a. The reference category is: Success.

b. This parameter is set to zero because it is redundant.

THANK YOU

Siyaphumelela at UP for funding to attend the conference