




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 Dihlalefi

Outline

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

Problem statement

- To what extent can **achievement** in **1st year** Maths/Physics be used to **understand** students' performance in their **2nd year** Math/Physics modules?
  - ➢ Focus on ***Success*** →
- Which 1st year modules explain 2nd year success?
- Any other student information?

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

Mathematics: Data and variables

Five 2nd year modules

WTW211, WTW218, WTW220, WTW221 and WTW248

Period

2014 – 2016

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

Only *first* attempt marks are used

1st year modules

WTW114, WTW126 and WTW128

Period

2011 - 2015

Additional information

- Grade 12 Maths mark
- Gender
- Language of instruction

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Physics: Data and variables

Two 2nd year modules

PHY253 and PHY263

Period

2014 – 2016

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

Only *first* attempt marks are used

1st year modules

PHY114, PHY124, WTW114, WTW126 and WTW128

Period

2011 - 2015

Additional information

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Statistical techniques

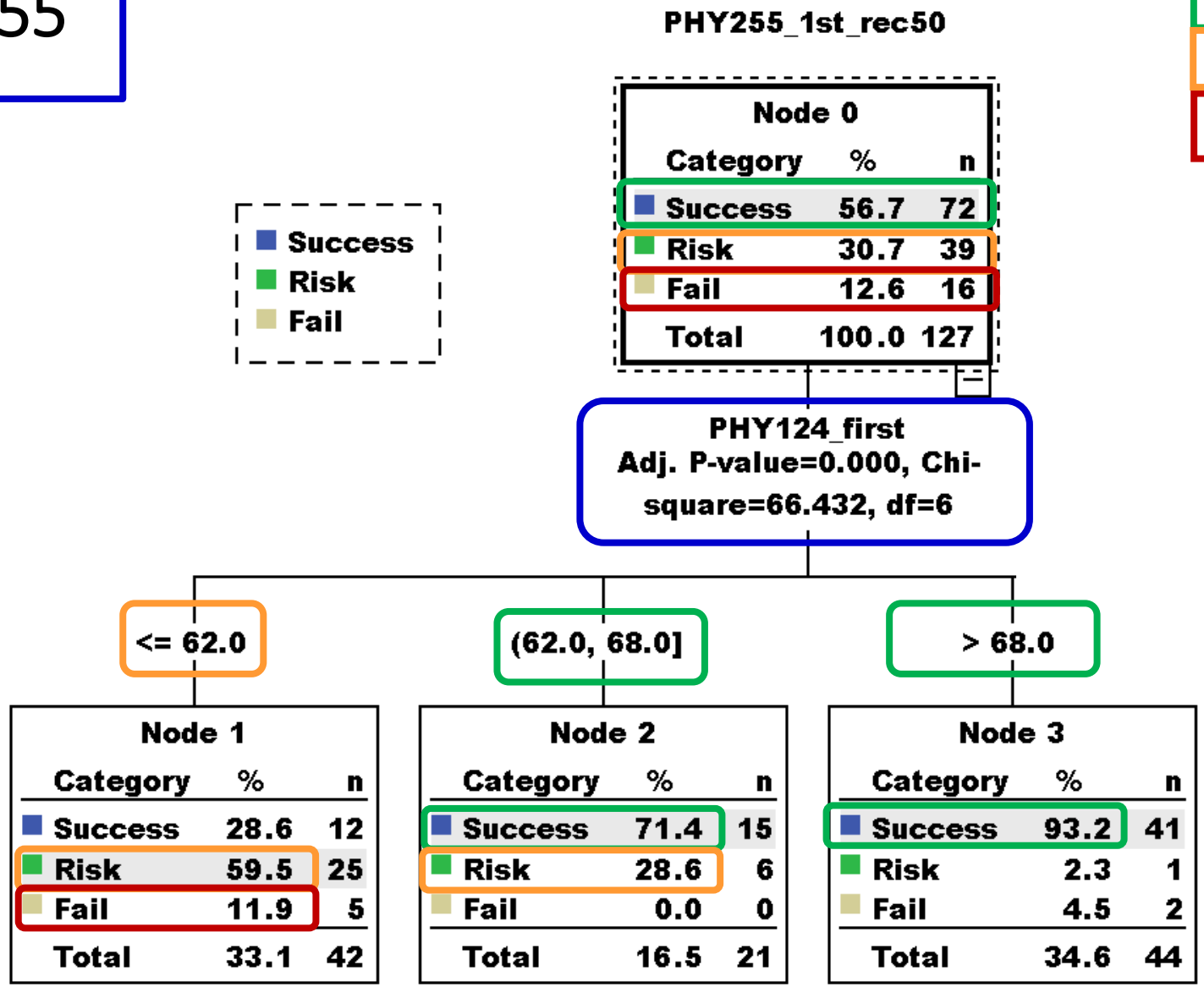
1) CHAID → Data segmentation → Dendogram (tree)

2) Logistic regression models

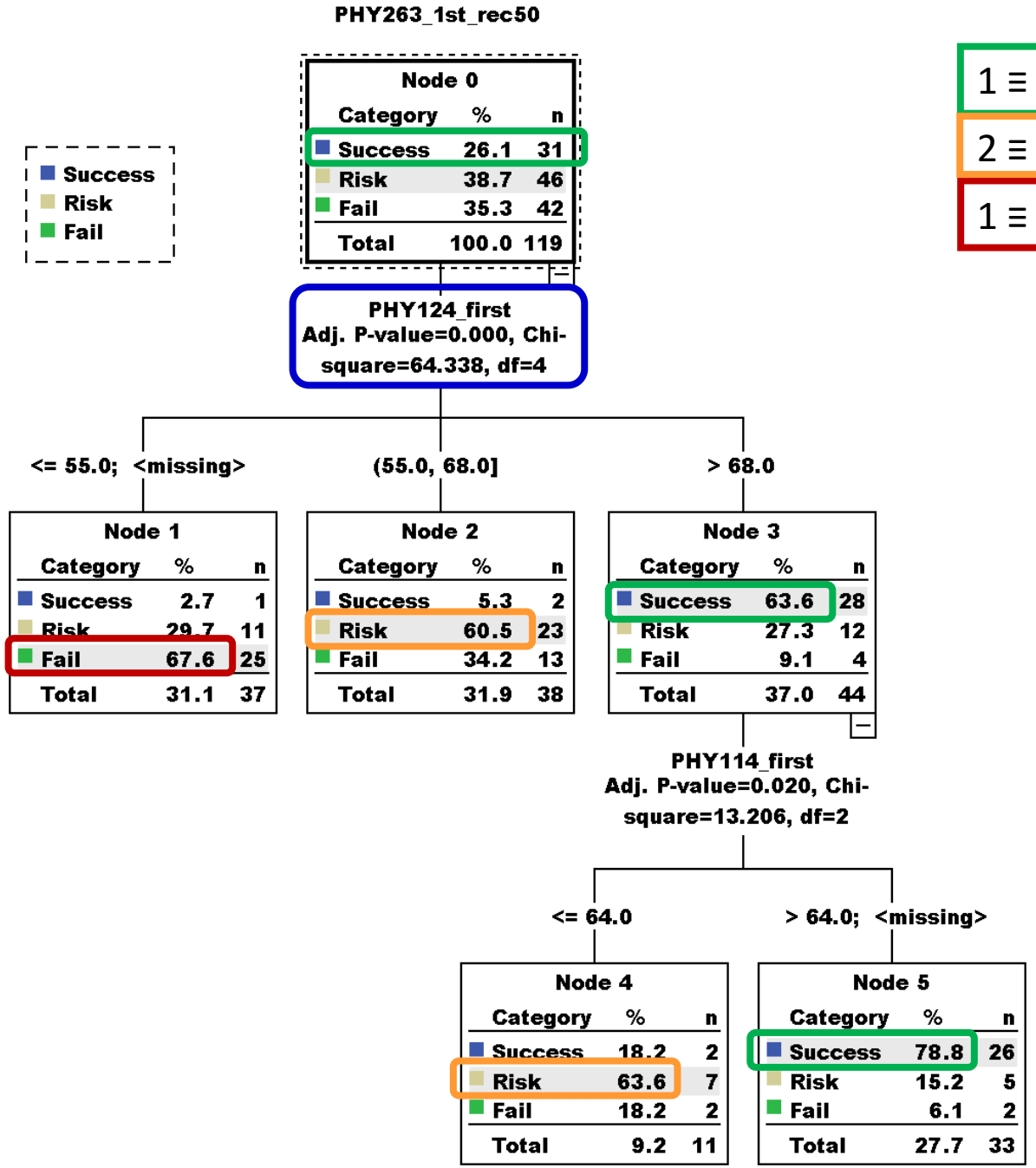
3) Cross-tabulations

PHY 255

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



PHY 263

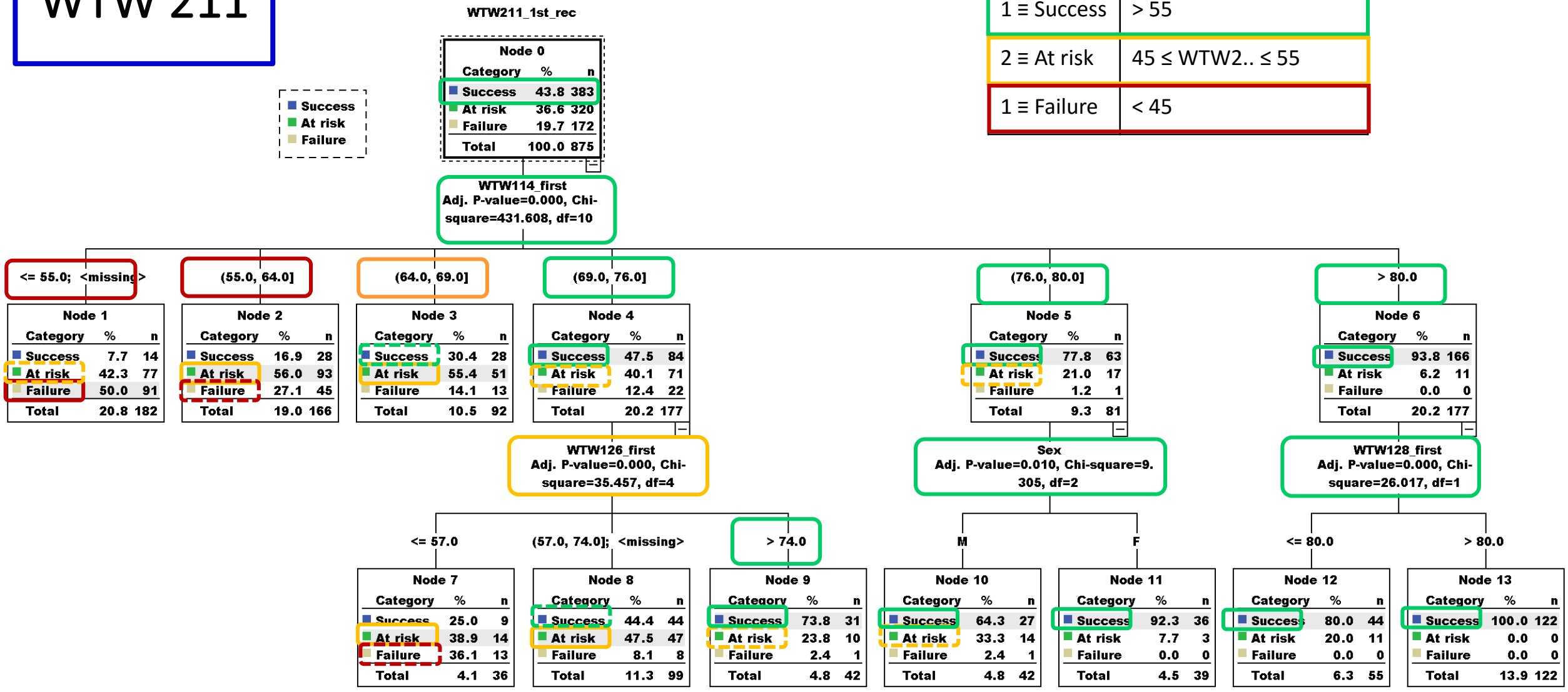


1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

Physics: Data and variables

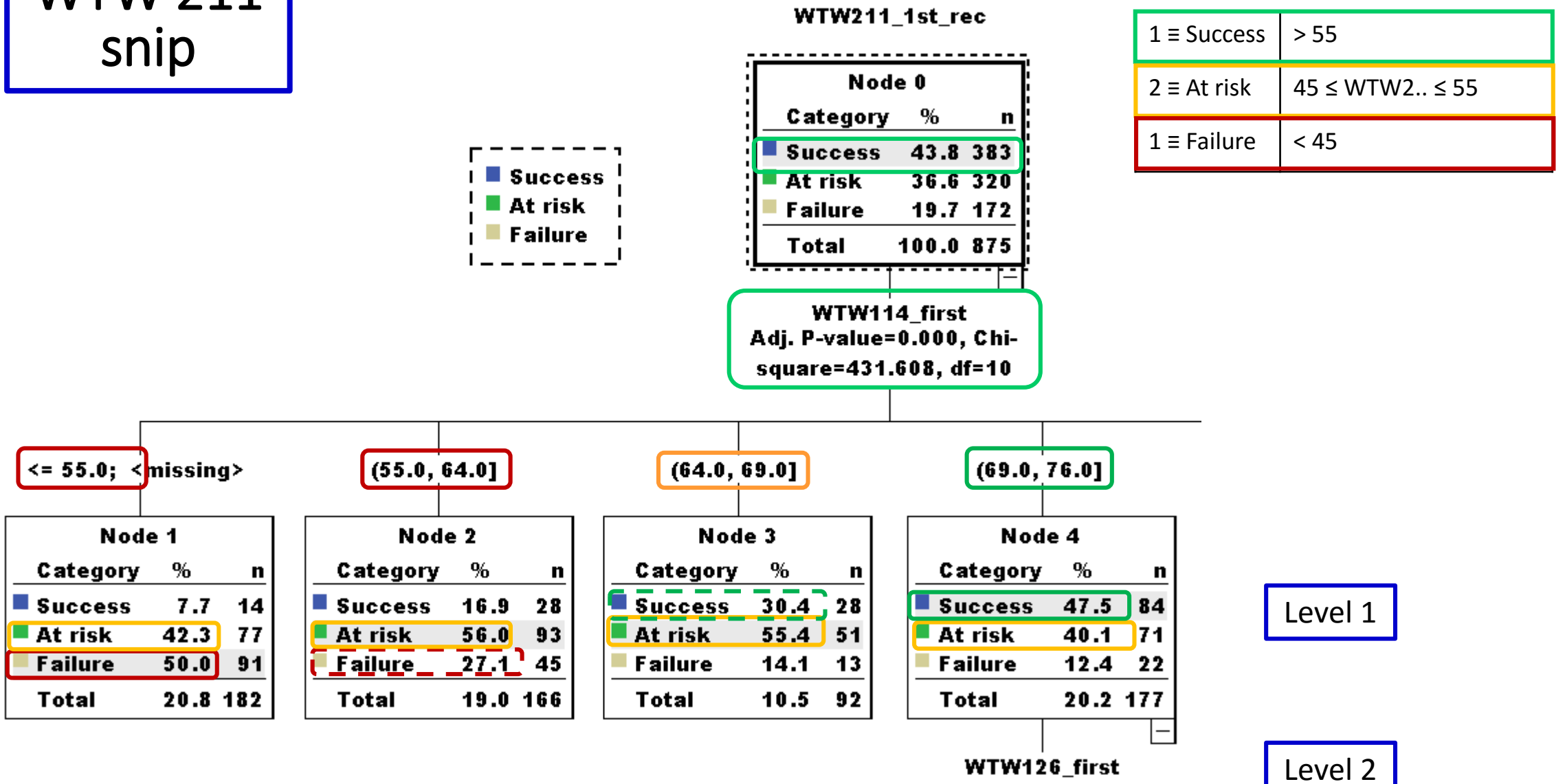
PHY 255	←	PHY 124
PHY 263	←	PHY 124

WTW 211



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

WTW 211 snip



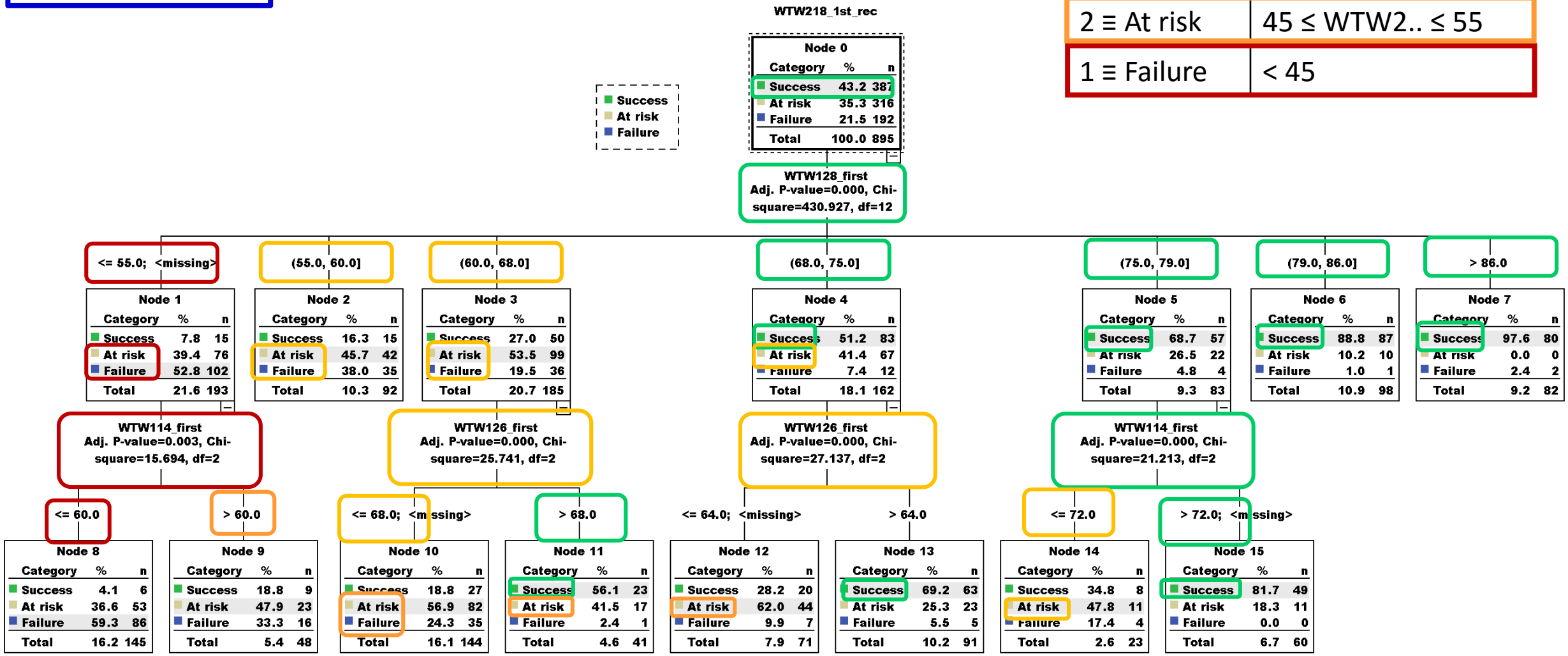
1 ≡ Success	> 55
2 ≡ At risk	45 ≤ WTW2.. ≤ 55
1 ≡ Failure	< 45

Level 1

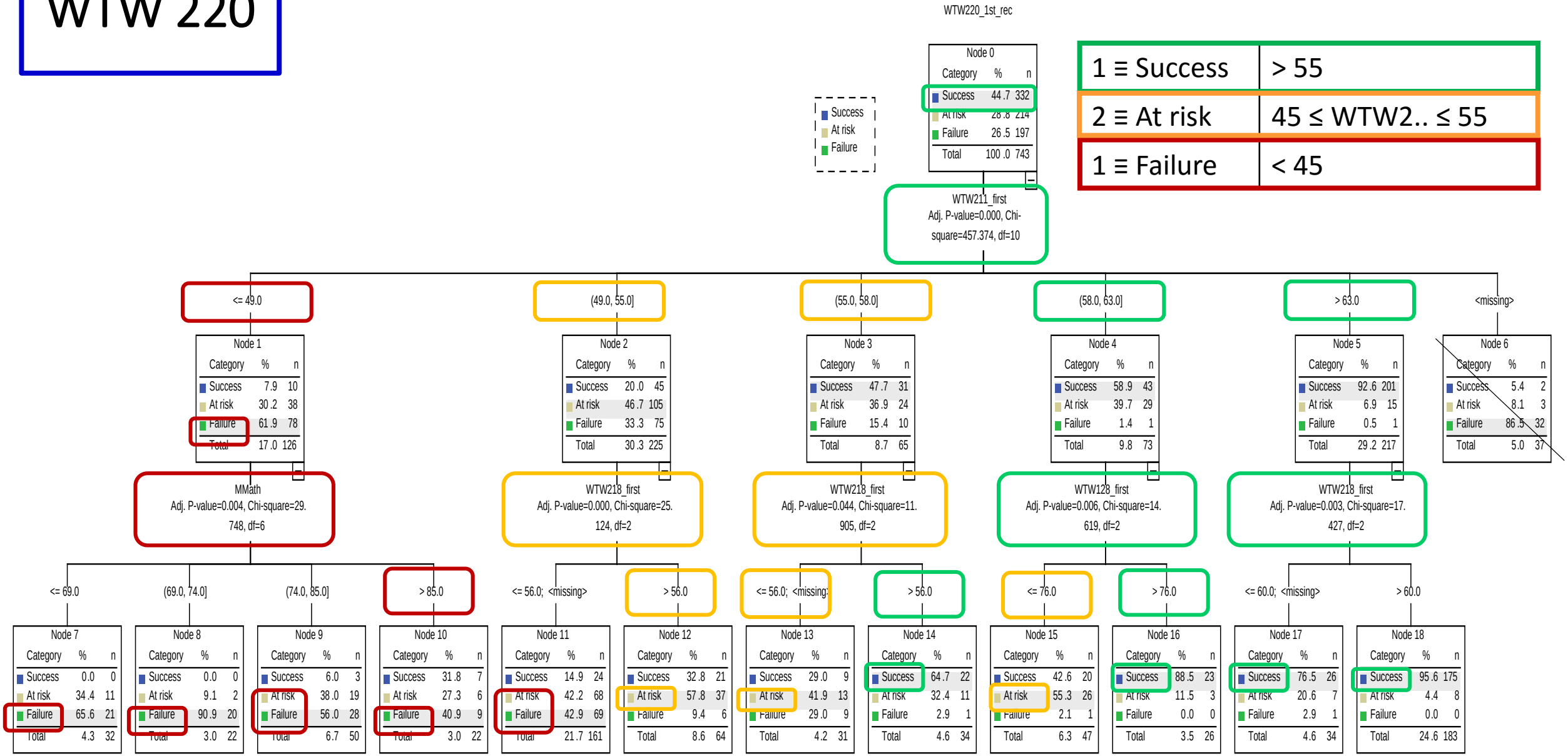
Level 2

WTW 218

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



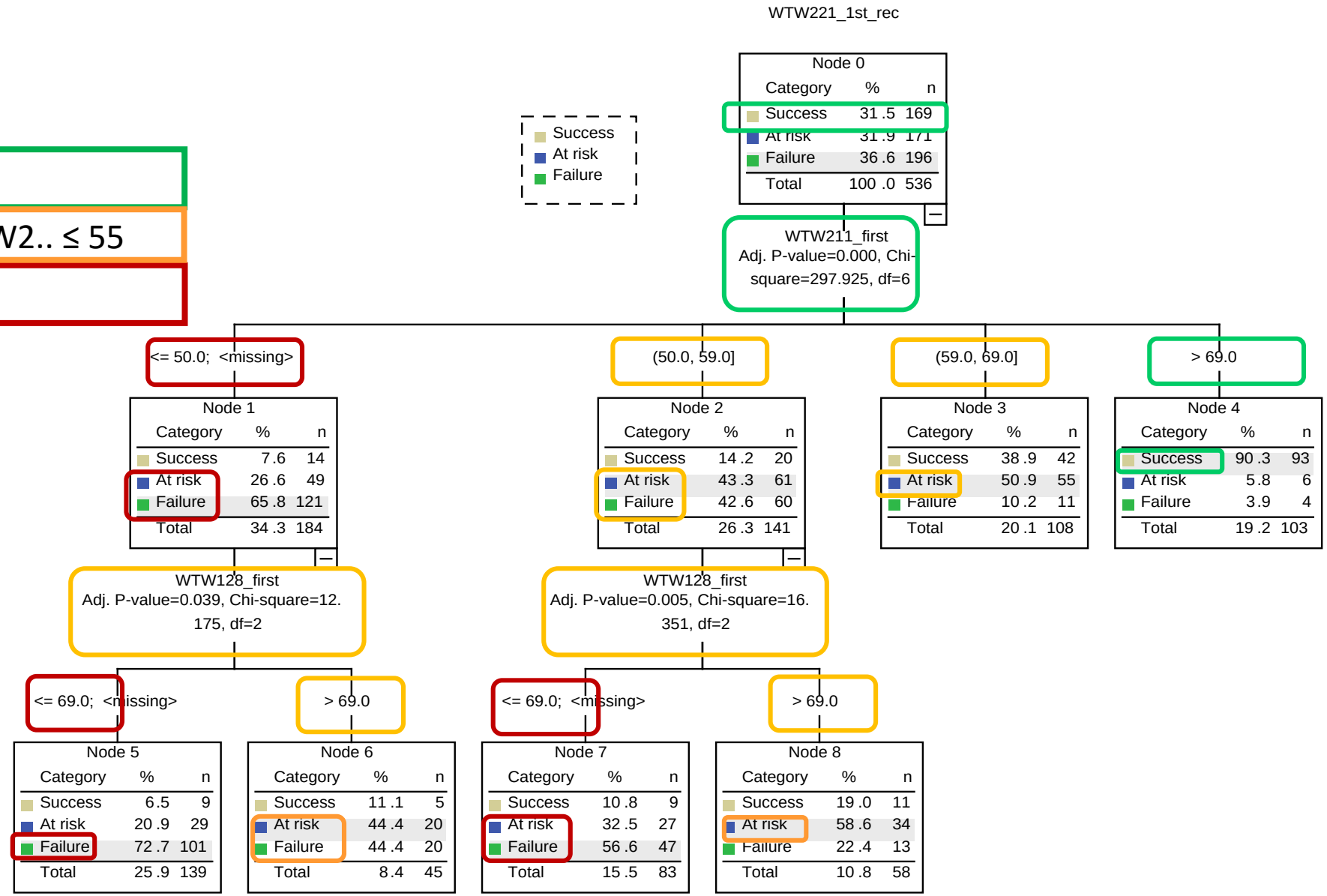
WTW 220



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

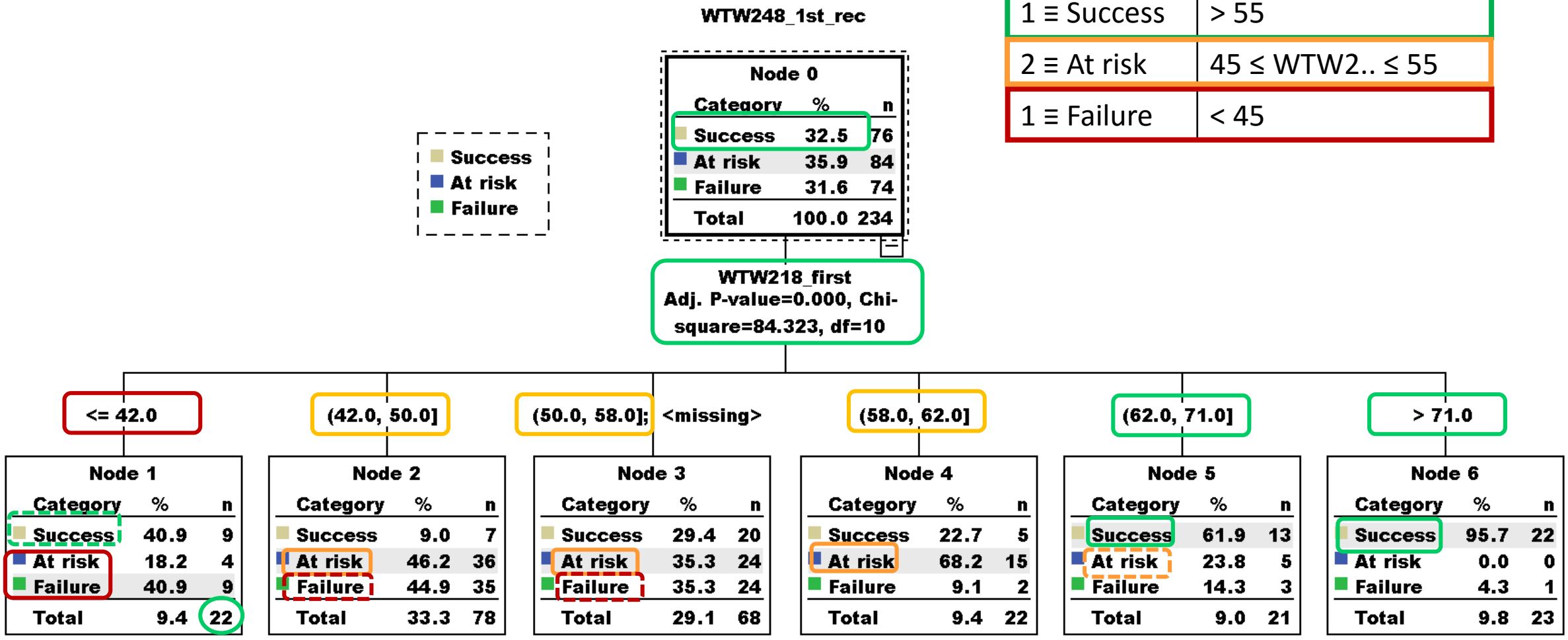
WTW 221

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



WTW 248

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



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: $Y = b_0 + b_1 X$

Multiple regression: $Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k$

Logistic regression: $ln(\frac{p}{1-p}) = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k$

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

Mathematics: multinomial logistic regression

Dependent variable	WTW211_1st_rec n=875	WTW220_1st_rec 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 ≥ 60) & (WTW 126 ≥ 60)
Categories	2 = Not safe	Rest

Crosstabulations

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

WTW220_1st_rec * WTW126_128 Crosstabulation

			WTW126_128		
			1 Safe	2 Not	Total
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		
			1 Safe	2 Not	Total
WTW248_1st_rec	1 Success	Count	71	51	122
		% within WTW126_128	53.8%	26.8%	37.9%
	2 At risk	Count	43	72	115
		% within WTW126_128	32.6%	37.9%	35.7%
	3 Failure	Count	18	67	85
		% within WTW126_128	13.6%	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: -2 Log Likelihood of Reduced Model	Likelihood Ratio Tests: Chi-Square	df	Sig.
Intercept	85.953[a]	.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[a]		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[b]	.	.	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[b]	.	.	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