



Michel Philipp (Autor)  
**Contributions to Machine Learning and  
Psychometrics**  
Computational, Graphical, and Statistical Methods for  
Assessing Stability

Michel Philipp



**CONTRIBUTIONS TO MACHINE LEARNING  
AND PSYCHOMETRICS**

Computational, Graphical, and Statistical Methods  
for Assessing Stability



Cuvillier Verlag Göttingen  
Internationaler wissenschaftlicher Fachverlag

<https://cuvillier.de/de/shop/publications/7439>

Copyright:  
Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen,  
Germany  
Telefon: +49 (0)551 54724-0, E-Mail: [info@cuvillier.de](mailto:info@cuvillier.de), Website: <https://cuvillier.de>



# Contents

<b>Scope of this work</b>	<b>1</b>
<b>1 A toolkit for stability assessment of tree-based learners</b>	<b>11</b>
1.1 Introduction . . . . .	12
1.2 Instability of trees . . . . .	13
1.3 Measuring variable selection and cutpoint stability . . . . .	15
1.3.1 Variable selection analysis . . . . .	17
1.3.2 Cutpoint analysis . . . . .	19
1.4 Discussion . . . . .	23
<b>2 Measuring the stability of results from supervised statistical learning</b>	<b>25</b>
2.1 Introduction . . . . .	26
2.1.1 Related work . . . . .	27
2.2 Stability measuring framework . . . . .	29
2.2.1 Semantic versus structural similarity . . . . .	30
2.2.2 Measuring similarity based on predictions . . . . .	32
2.2.3 Stability measurement procedure . . . . .	33
2.3 Framework settings . . . . .	34
2.3.1 Similarity and dissimilarity measures . . . . .	34
2.3.2 Resampling and evaluation methods . . . . .	36
2.4 Simulation and benchmark experiments . . . . .	39
2.4.1 Data-generating processes . . . . .	40
2.4.2 Study 1: Impact of the DGP . . . . .	41
2.4.3 Study 2: Impact of resampling and evaluation methods . . . . .	44



---

2.4.4	Benchmark experiment . . . . .	48
2.5	Discussion . . . . .	49
2.5.1	Summary and recommendations . . . . .	49
2.5.2	Limitations and future research . . . . .	50
2.6	Implementation . . . . .	52
2.7	Acknowledgments . . . . .	52
<b>3</b>	<b>On the estimation of standard errors in cognitive diagnosis models</b>	<b>53</b>
3.1	Introduction . . . . .	54
3.2	Cognitive diagnosis models . . . . .	56
3.2.1	Theory and estimation of standard errors . . . . .	57
3.2.2	The G-DINA model . . . . .	60
3.3	Illustrations . . . . .	63
3.3.1	Coverage study . . . . .	63
3.3.2	Empirical example . . . . .	71
3.4	Discussion . . . . .	73
3.5	Computational details . . . . .	75
<b>4</b>	<b>Detecting DIF in cognitive diagnosis models</b>	<b>77</b>
4.1	Introduction . . . . .	78
4.2	Theoretical background . . . . .	80
4.2.1	The DINA model . . . . .	80
4.2.2	DIF in the DINA model . . . . .	80
4.3	Simulation study . . . . .	84
4.3.1	Type I error study . . . . .	86
4.3.2	Power study . . . . .	86
4.4	Discussion . . . . .	89
	<b>Conclusion and outlook</b>	<b>93</b>
<b>A</b>	<b>Supplementary material: Chapter 2</b>	<b>103</b>
A.1	Similarity measures . . . . .	103
A.2	Simulation experiments . . . . .	103



---

<b>B</b>	<b>Supplementary material: Chapter 3</b>	<b>111</b>
B.1	Blockwise matrix inversion . . . . .	111
<b>C</b>	<b>R codes</b>	<b>113</b>
C.1	R package stablelearner . . . . .	113
C.1.1	Installation . . . . .	113
C.1.2	Stability of tree-based methods . . . . .	114
C.1.3	Stability of results from supervised statistical learning . . . . .	116
C.2	R package Rcdm . . . . .	121
C.2.1	Installation . . . . .	122
C.2.2	Data preparation . . . . .	122
C.2.3	Model fitting . . . . .	123
C.2.4	Model diagnosis . . . . .	123
C.2.5	Model comparison . . . . .	125
C.2.6	DIF detection . . . . .	126
C.3	Rasch tree example . . . . .	126
<b>D</b>	<b>Publications</b>	<b>127</b>