Comparing Techniques for Aggregating Interrelated Replications in Software Engineering

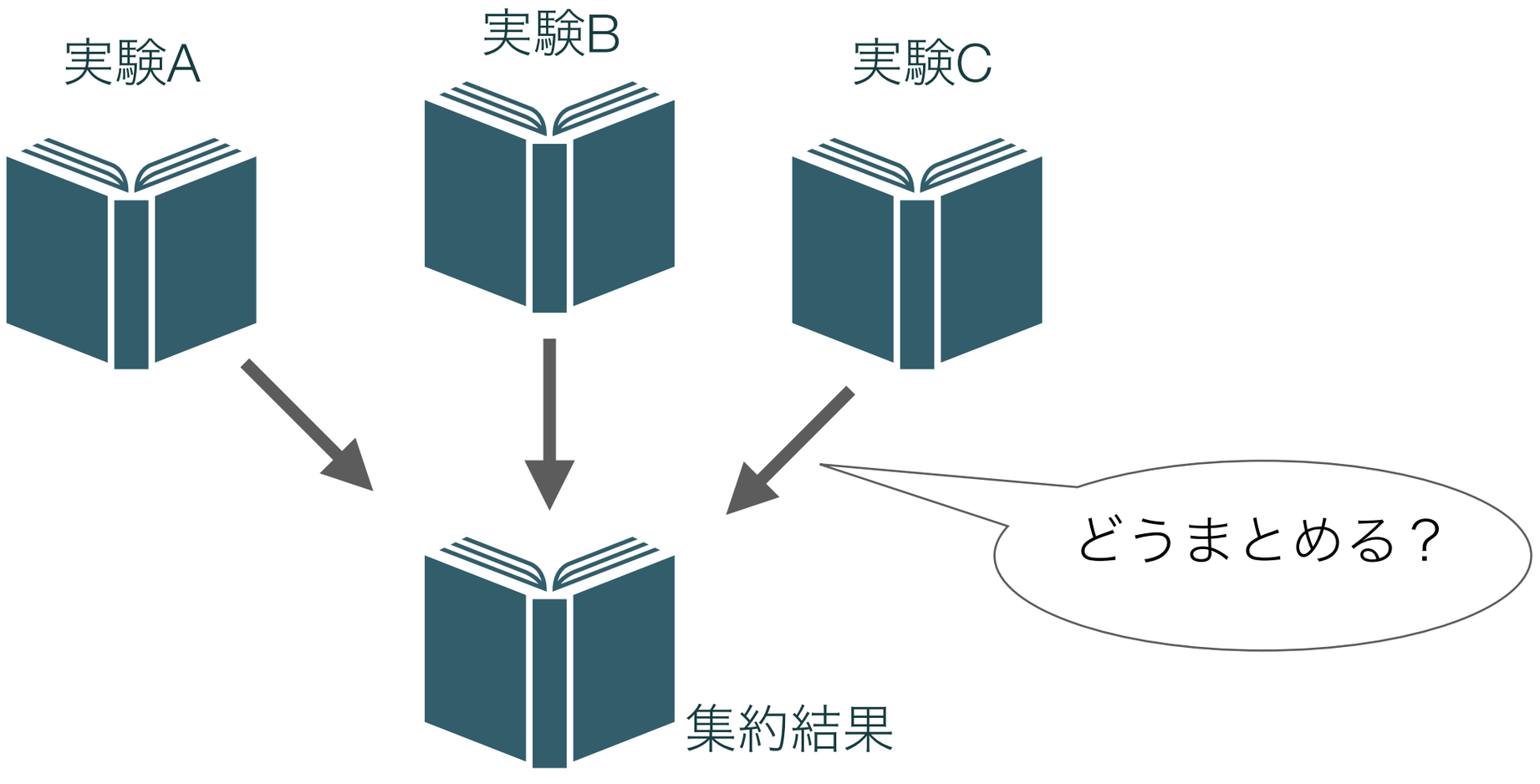
Adrian Santos(M3S-ITEE University of Oulu, Finland),
Natalia Juristo(Universidad Politécnica de Madrid, Spain)

ソフトウェア工学における再現実験結果の集約手法の比較

紹介担当：西川 諒真

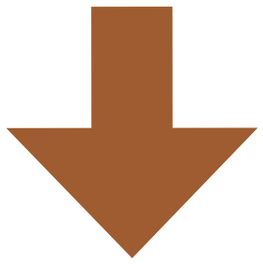
概要

ソフトウェア工学における，同じ実験計画による複数の実験の結果を集約するにはどんな手法が適しているかを調べ，特にAD，IPDといった手法が優れているということを示した.



背景

複数のレプリケーションの実験結果を集約するための手法は
統一されておらずいくつか存在するが，もし不適切な手法を適
用すると集約結果の信頼性に影響を与えるかもしれない.



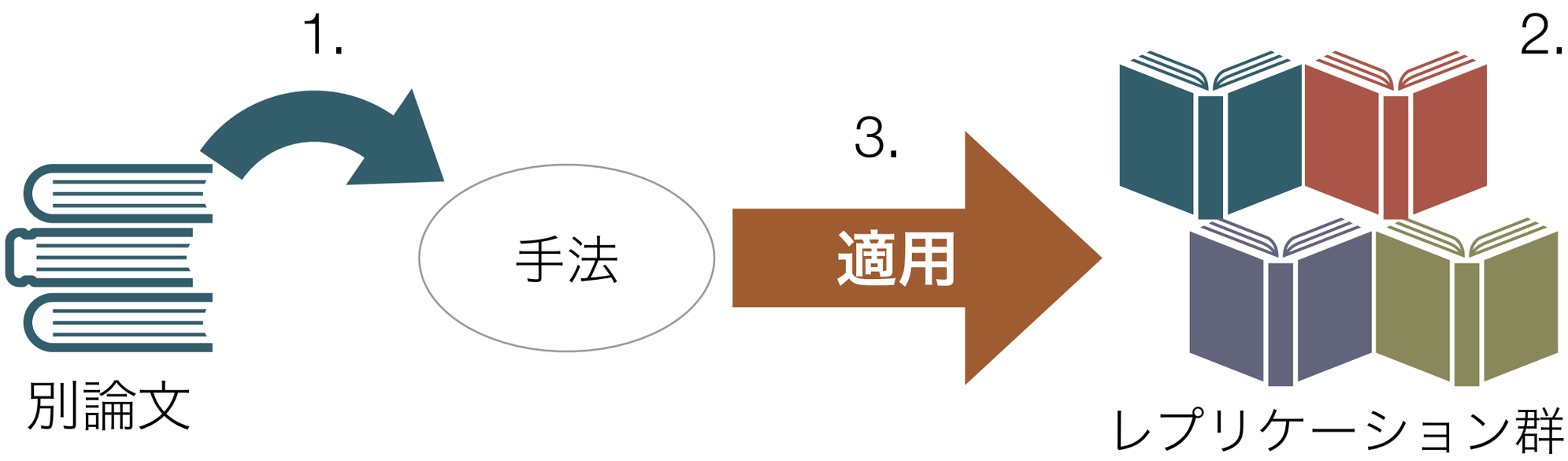
ソフトウェア工学における実験結果の集約手法をまとめ，
実際に適用してどのような長所，短所が発生するかを調べた.

手法

1.ソフトウェア工学で適用された集約手法を特定

2.同じ実験計画を持つレプリケーションのグループを選択

3.各集約手法でレプリケーションを分析し，結果を比較

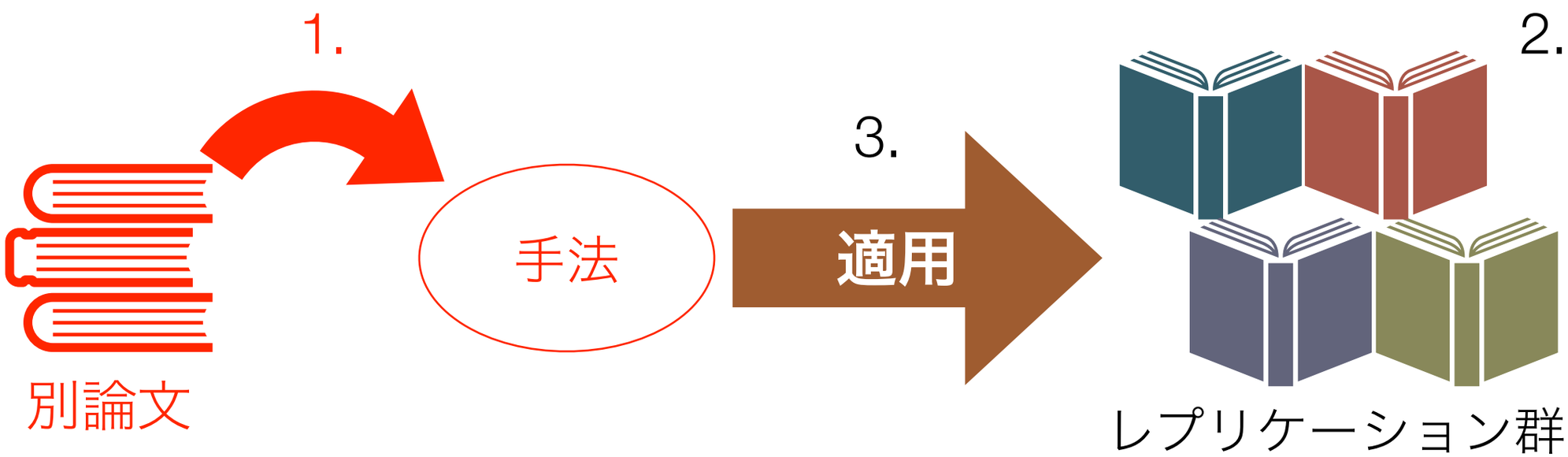


手法

1.ソフトウェア工学で適用された集約手法を特定

2.同じ実験計画を持つレプリケーションのグループを選択

3.各集約手法でレプリケーションを分析し，結果を比較



集約手法

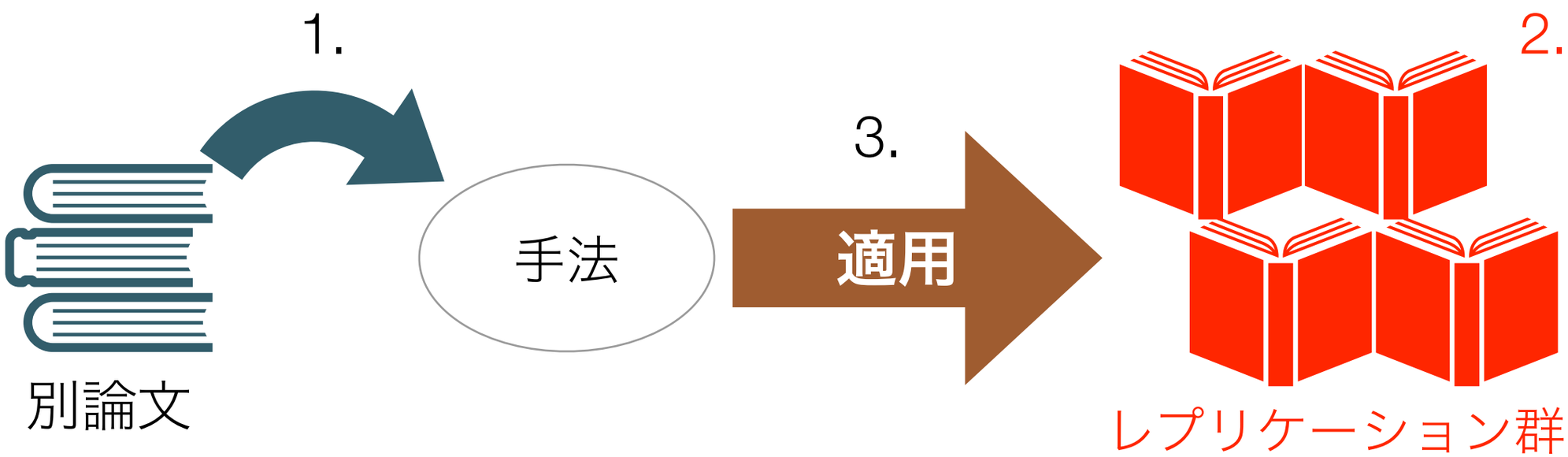
手法	長所	短所
Narrative Synthesis	集約結果を出しやすい	結果に 主観的な判断が伴う
AD (Aggregated Data)	結果が見やすい	効果量に強く依存する
IPD (Individual Participant Data)	データの欠損に強い	実験計画が異なると 集約しづらい
Aggregation of p-values	実験計画が異なっても 集約に影響しづらい	効果量を算出できない

手法

1.ソフトウェア工学で適用された集約手法を特定

2.同じ実験計画を持つレプリケーションのグループを選択

3.各集約手法でレプリケーションを分析し，結果を比較

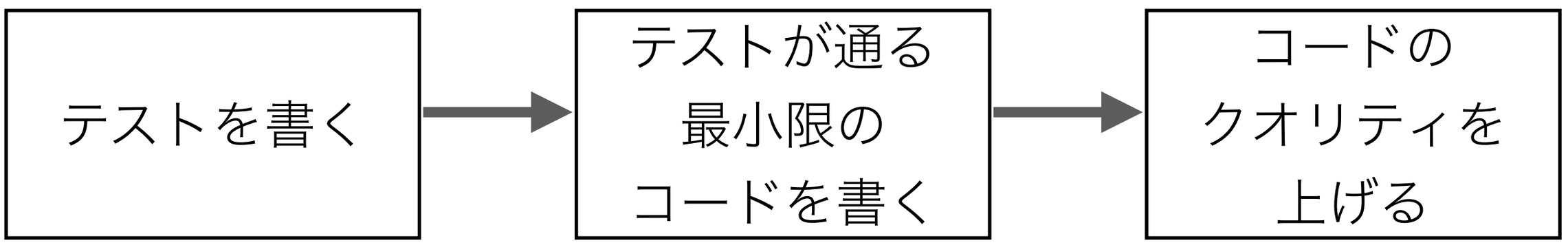


使用したレプリケーション

- TDD（テスト駆動開発）がプログラムの品質に与える影響を評価するために行われた論文の実験に対する４つのレプリケーション.

- ITL，TDDのそれぞれのプログラム開発手法で開発を行い，テストケースの通過率（FC）を比較している.
  （ITLはTDDの逆順に開発を行う手法）

TDD



各レプリケーションでの結果

実験	開発手法	人数	FCの平均	FCの中央値
F-Secure H	ITL	6	30.71	24.16
	TDD	6	40.23	35.34
F-Secure K	ITL	11	22.17	17.98
	TDD	11	35.42	22.41
F-Secure O	ITL	7	16.05	7.87
	TDD	7	68.97	81.03
UPV	ITL	31	33.38	6.74
	TDD	29	77.16	83.92

- 全体的にTDDの方がITLより優れている.

参加者のプログラム経験についての調査

実験	人数	プログラミング	Java	Unit	JUnit
F-Secure H	6	3.67	2.33	2.17	2.17
F-Secure K	11	2.91	1.82	1.64	1.27
F-Secure O	7	3.29	2.71	2.71	2
UPV	33	2.36	1.88	1.04	1

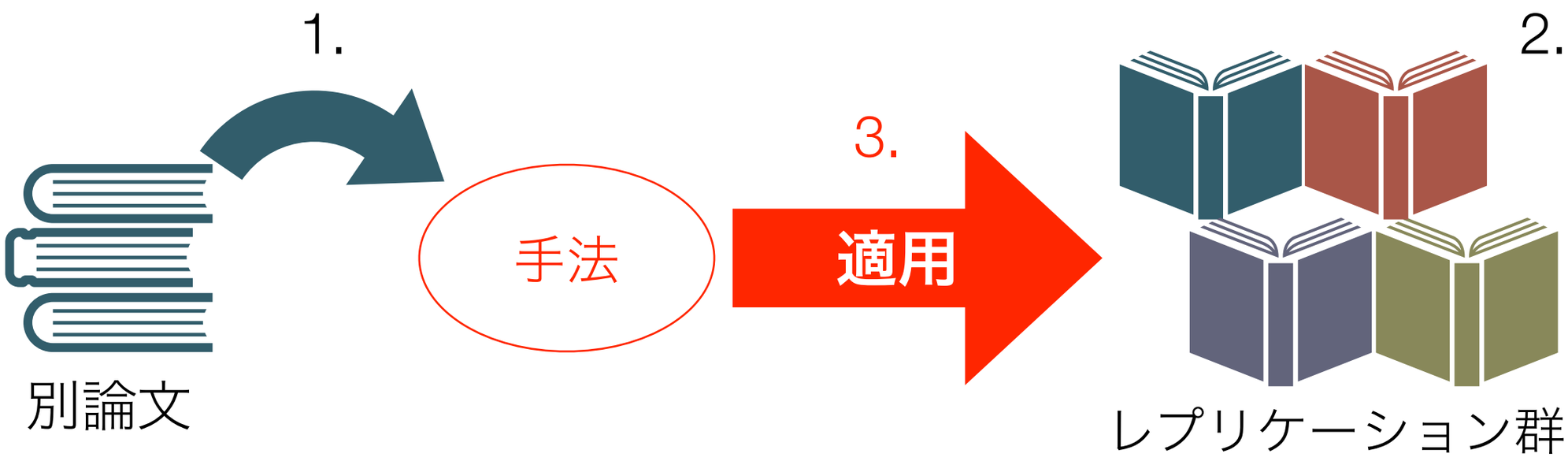
- それぞれ，1=未経験，2=初心者，3=中級者，4=エキスパートとした平均値を示す.
- 結果へ影響を与えた要因を調べる際に使用する.

手法

1.ソフトウェア工学で適用された集約手法を特定

2.同じ実験計画を持つレプリケーションのグループを選択

3.各集約手法でレプリケーションを分析し，結果を比較



各集約手法の結果の比較

集約手法	結果	集約した効果量	p値の集約	結果への影響要因
Narrative synthesis	TDD>ITL	不明	出来なかった	不明
AD	TDD>ITL	大きい	出来た	おそらくJava経験
IPD	TDD>ITL	大きい	出来た	プログラム経験
Aggregation of p-values	TDD>ITL	不明	出来た	不明

結論

- 結果への影響要因を調べる目的においてはNarrative synthesisやAggregation of p-valuesよりもAD, IPDが優れている.

- ADは図を用いた結果の視覚的な集約に優れており,
IPDは欠損データの処理や参加者単位での結果への影響要因の解明に優れている.

所感

- 各手法の長所，短所の比較が分かりやすかった.

- 統計関連の知識が無いと一部の結果がわかりにくいように感じた.