INTRODUCTION TO PRODUCING AUTOMATED TABLES IN STATA USING `collect`

Training session
2025 Belgium Stata Conference

NICK DESCHACHT
nick.deschacht@kuleuven.be

Table 1. Descriptive statistics

	N	Mean	SD	Min	Max
wage	2246	7.77	5.76	1.00	40.75
age	2246	39.15	3.06	34	46
grade	2244	13.10	2.52	0	18
race: white	2246	0.73	0.44	0	1
race: black	2246	0.26	0.44	0	1
race: other	2246	0.01	0.11	0	1

Table 3. Regression models

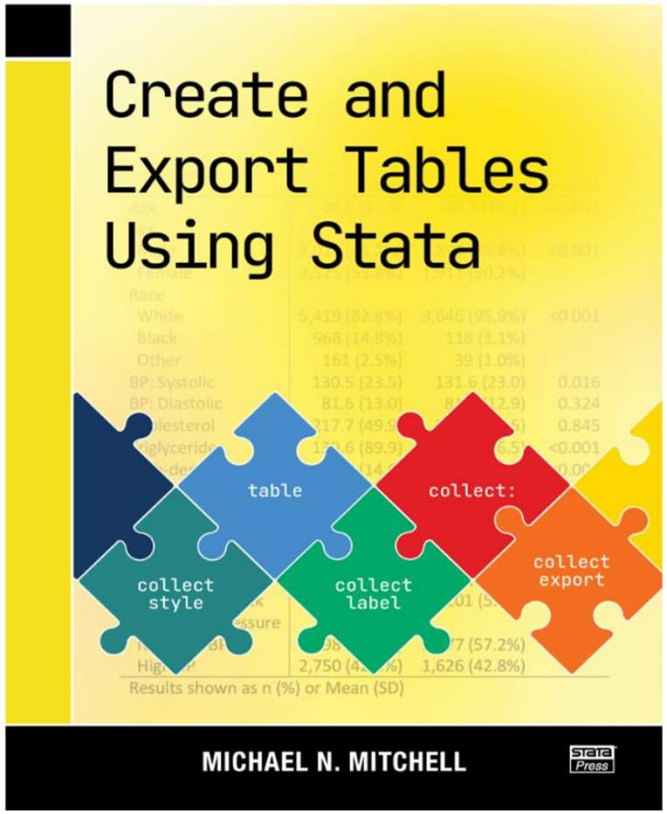
	Model 1	Model 2
Grade	0.741***	0.724***
	(0.046)	(0.046)
Age	-0.048	-0.054
	(0.038)	(0.038)
Race		
Black		-0.638*
		(0.266)
Other		0.507
		(1.075)
Constant	-0.075	0.566
	(1.610)	(1.631)
N	2244	2244
R-squared	0.107	0.109

*** p<.001, ** p<.01, * p<.05

Introduction

- `collect` was introduced in Stata 17 to create tables
- Useful if you want to:
  - Present tables in Word, PDF, LaTeX, …
  - Customize your tables
  - Combine results from difference commands in one table
  - Program table making in a reproducible way
- Stata commands `dtable` (descriptives) and `etable` (regression) are more user-friendly commands built on `collect`.
- `collect` seems more flexible than popular user-written commands such as `esttab`, `asdoc`, …
- Aim of this presentation is to introduce you to `collect` by demonstrating how to make a few tables (do file available online).

Learn more?



Mitchell, M. N. (2025). *Create and export tables using Stata*. Stata Press.

Example 1

Table. Mean wages by race and married

	Single	Married
White	8.93	7.72
Black	6.73	6.97

- **Collections** contain values (results from commands), labels and styles.
- Each value in a collection has **tags**: e.g. the value 8.93 has tags race[White], married[single] and result[mean]
- A tag is a level of a dimension. e.g. the **dimension** married has **levels** single and married
  Not only categorical variables can be a dimension, but also a set of regression coefficients, or various regression commands you executed.

Workflow

collect is not a single command but a set of commands to build tables incrementally

Basic programming workflow:

- Step 1: Collect results from Stata commands
- Step 2: Choose which results to display (collect layout)
- Step 3: Customize your table (labels, number formats, italics, lines, colors, …)
- Step 4: Export your table

collect commands

- The core command `collect layout` creates the table:

```
collect layout (row tags) (column tags) (table tags)
```

- Super-rows (combinations or interactions of dimensions):

```
collect layout (race#married) (result[mean sd])
```

Common collect commands:

- List dimensions of a collection: `collect dims`
- List the levels of a dimension: `collect levelsof`
- Relabel levels: `collect label levels`
- Add a title: `collect title`
- Format style: `collect style cell`
- Export to docx, pdf, ...: `collect export`

Example 2

Table. Mean characteristics of single and married respondents

	Single	Married	p value
wage	8.1	7.6	0.054
age	39.2	39.1	0.447
grade	13.1	13.1	0.571

- `collect layout` loops over the values in a collection. If exactly one value matches the tags for a given cell, it gets placed in the table.
- A style can be customized for specific values or levels: e.g. show 3 decimals for the p values by defining the style for the 'level p' of dimension 'result'

```
collect style cell result[p], nformat(%6.3f)
```

Example 3

Table 1. Descriptive statistics

	N	Mean	SD	Min	Max
wage	2246	7.77	5.76	1.00	40.75
age	2246	39.15	3.06	34	46
grade	2244	13.10	2.52	0	18
race: white	2246	0.73	0.44	0	1
race: black	2246	0.26	0.44	0	1
race: other	2246	0.01	0.11	0	1

- The command `dtable` uses collect to create a table with descriptive statistics in a user-friendly way (with many options). I will use `collect`.
- min/max for wage have 2 decimals → `variable[wage]#result[min max]`
- To set the width of the columns:

```
collect style putdocx, width(10cm) width(25, 15, 15, 15, 15, 15)
```

Example 4. Regression tables

Table 3. Regression models

	Model 1	Model 2
Grade	0.741***	0.724***
	(0.046)	(0.046)
Age	-0.048	-0.054
	(0.038)	(0.038)
Race		
Black		-0.638*
		(0.266)
Other		0.507
		(1.075)
Constant	-0.075	0.566
	(1.610)	(1.631)
N	2244	2244
R-squared	0.107	0.109

*** p<.001, ** p<.01, * p<.05

- Add N and R^2 rows: `collect layout (colname#result[_r_b] result[N r2]) (cmdset)`
- Add significance start: `collect stars`
- Omit reference categories: `collect style showbase off`