The Stata Journal (2023)
**23**, Number 2, pp. 545–577 DOI: 10.1177/1536867X231175334

# Reporting empirical results to .docx files

Yuan Xue
School of Management
Huazhong University of Science and Technology
Wuhan, China
xueyuan19920310@163.com

Chuntao Li
School of Economics
Henan University
Kaifeng, China
chtl@henu.edu.cn

Haitao Si
School of Economics and Management
Wuhan University
Wuhan, China
sihaitao0114@163.com

**Abstract.** Reporting empirical results to automatically generate structured tables is important but time consuming for empirical researchers. Because of the lack of commands that can effectively create and edit Office Open XML documents (`.docx` documents), neither official commands nor community-contributed commands could tabulate results to this regularly used document type until `putdocx` was launched in Stata 15. In this article, we introduce four new commands: `sum2docx`, `corr2docx`, `t2docx`, and `reg2docx`. These new commands are all based on `putdocx`. They can be coalesced and can report summary statistics, correlation coefficient matrices, split-sample $t$ tests, and regression results automatically in one `.docx` file. The commands are user friendly and can provide researchers with new options for reporting empirical results.

**Keywords:** st0719, sum2docx, corr2docx, t2docx, reg2docx, putdocx, Office Open XML documents, empirical results

## 1 Introduction

Official and community-contributed commands such as `outreg` (Gallup 1998, 1999, 2000, 2001, 2012), `outreg2` (Wada 2005), `estout` (Jann 2005, 2007; Jann and Long 2010), and `xml_tab` (Lokshin and Sajaia 2008) can produce empirical results in rich text format (`.rtf`), LaTeX, HTML, XML, and other document types with formatted tables. However, they cannot report empirical results in Office Open XML documents (`.docx` documents), because it was impossible to directly create `.docx` documents using versions of Stata prior to 15. In this article, we introduce four new commands that we have developed: `sum2docx`, `corr2docx`, `t2docx`, and `reg2docx`. These commands report summary statistics, correlation coefficient matrices, mean-comparison $t$ tests between subsamples, and regression tables that are used in empirical research. They can produce the tables in one `.docx` document.

Office Open XML is a zipped, XML-based file format developed by Microsoft. It is used as the default choice by many Windows users since Microsoft Office 2007. An

 

Office Open XML document (`.docx` document) is a compressed file format based on the Office Open XML standard that occupies less memory yet has a good execution effect on complex editing tasks compared with previous `.doc` documents. Thus, `.docx` has gradually become the most widely used document format.

Since Stata 15, Stata has had a powerful command, `putdocx`, that can create `.docx` documents containing text paragraphs, graphics, and tables with a few simple lines of commands. The four commands we are introducing are all based on this command. If these four commands are combined with `putdocx`, then the whole article, not just the empirical results, can be generated. Also, there is an additional feature, Mata programming, that is used to ensure the speed of program execution and calculate the statistics in the latest versions of these commands.

We will first present the syntax and basic usage of these four commands. Then, we will provide specific examples to illustrate how to use them to report empirical results to `.docx` documents.

# 2 Syntax and basic usage

## 2.1 sum2docx

The summary statistics table is usually among the first tables reported in empirical research. It shows the sample mean, median, and standard deviation, for example. Thus, it gives a logical picture about the distribution and characteristics of the data. The summary statistics work with the `sum2docx` command, which produces a formatted table that contains any summary statistic generated by the command `summarize`. The summary statistics are stored in the return list, and then `sum2docx` produces a table to a given `.docx` document. The basic syntax of `sum2docx` is

`sum2docx` *varlist* [ *if* ] [ *in* ] [ *weight* ] `using` *filename* [ `,` *options* ]

where *varlist* is a list of nonstring variables. The `*` and `?` wildcards can also be used while specifying the *varlist*, or `_all` can be used to indicate that all the variables should be included in the list. *filename* is a `.docx` document. It is used to store the table. A file path can also be specified to indicate the folder in which to store the file. `aweight`s, `fweight`s, and `iweight`s are allowed; see [U] **11.1.6 weight**. The command has the following options:

| *options* | Description |
|---|---|
| `replace` | overwrite an existing `.docx` file |
| `append`[ (*apopts*) ] | append the output to an existing file and change style definitions along with page break, header, and footer settings |
| `landscape` | change the document orientation from portrait to landscape |
| `margin(`*type*`, #`[ *unit* ]`)` | set page margins for the document |
| `title(`*string*`)` | specify the title of the table |
| `pagesize(`*psize*`)` | set the page size of the document; *psize* may be `letter`, `legal`, `A3`, `A4`, or `B4JIS`; default is `pagesize(A4)` |
| `font(`*fontname*[ , *size*[ , *color* ] ]`)` | set the font, font size, and font color for the document; note that the font setting option is the same as in `putdocx`, where font, font size, and color can be left blank to allow for default settings |
| `note(`*string*[ , *cell_fmt_options* ]`)` | add notes under the table |
| `layout(`*layouttype*`)` | adjust column width; *layouttype* could be `fixed`, `autofitwindow`, or `autofitcontents` |
| `varname` | output variables' names in the first column of the table, which is the default choice |
| `varlabel` | output variables' labels instead of names in the first column of the table |
| `stats(`*stats*`)` | specify the statistics to be output, which are the following: `N`, `mean`[ (*fmt*) ], `var`[ (*fmt*) ], `sd`[ (*fmt*) ], `skewness`[ (*fmt*) ], `kurtosis`[ (*fmt*) ], `sum`[ (*fmt*) ], `sum_w`[ (*fmt*) ], `min`[ (*fmt*) ], `median`[ (*fmt*) ], `max`[ (*fmt*) ], `p1`[ (*fmt*) ], `p5`[ (*fmt*) ], `p10`[ (*fmt*) ], `p25`[ (*fmt*) ], `p75`[ (*fmt*) ], `p90`[ (*fmt*) ], `p95`[ (*fmt*) ], `p99`[ (*fmt*) ]; default format is `%9.3f` |

Note that among the above listed options, `replace`, `append()`, `landscape`, `margin()`, `title()`, `pagesize()`, `font()`, `note()`, `layout()`, `varname`, and `varlabel` are the same in all four commands being discussed in this article. Thus, these options will not be relisted when introducing the options for other commands later in this article.

When we use `sum2docx`, the statistics output should be defined in option `stats()`. The order of statistics in the output table is according to the order that we list them in option `stats()`. For instance, statistics may include sample size (`N`), sample mean (`mean`), sample median (`median`), and standard deviation (`sd`). Return values can be accessed via the `summarize` command for a complete list of statistics that can be listed in option `stats()`. For example, with `mus208psid.dta` from chapter 8 of *Microeconometrics Using Stata, Second Edition* (Cameron and Trivedi 2022),[1] the following command can import the data and output the number of observations, mean, standard deviation, minimum, median, and maximum into a formatted table, as shown in table 1.

```
. net get mus2, from("http://www.stata-press.com/data/mus2")
checking mus2 consistency and verifying not already installed...
  (output omitted)
. use mus208psid
(A.C.Cameron & P.K.Trivedi (2022): Microeconometrics Using Stata, 2e)
. sum2docx exp - lwage using "summary.docx", stats(N mean sd min median max)
> replace title("Table 1. Summary statistics")
Summary statistics table has been written to file summary.docx.
```

Table 1. Summary statistics

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.854 | 10.966 | 1.000 | 18.000 | 51.000 |
| wks | 4165 | 46.812 | 5.129 | 5.000 | 48.000 | 52.000 |
| occ | 4165 | 0.511 | 0.500 | 0.000 | 1.000 | 1.000 |
| ind | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 1.000 |
| south | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 1.000 |
| smsa | 4165 | 0.654 | 0.476 | 0.000 | 1.000 | 1.000 |
| ms | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 |
| fem | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 1.000 |
| union | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 1.000 |
| ed | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 17.000 |
| blk | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 1.000 |
| lwage | 4165 | 6.676 | 0.462 | 4.605 | 6.685 | 8.537 |

You can open the `.docx` file by either clicking on the filename link in the Results window or using a command like `shellout summary.docx`. The default format for each statistic number is `%9.3f`. The format can also be defined by the option `stats()`. For instance, the format could be defined via `stats()`, and then you could append the new table into a previous document using the option `append`:

1. According to Cameron and Trivedi (2022), the data were drawn from the Panel Study of Income Dynamics and could be used to show the relationship between wages and the years of full-time work experience. The variable labels are self-explaining.

```
. sum2docx exp - lwage using "summary.docx", append
>  stats(N mean(%9.2f) sd(%9.3f) min(%9.3g) median(%9.0g) max(%9.2f))
>  title("Table 2. Summary statistics with specified formats of statistics")
Summary statistics table has been written to file summary.docx.
```

Table 2. Summary statistics with specified formats of statistics

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.85 | 10.966 | 1 | 18 | 51.00 |
| wks | 4165 | 46.81 | 5.129 | 5 | 48 | 52.00 |
| occ | 4165 | 0.51 | 0.500 | 0 | 1 | 1.00 |
| ind | 4165 | 0.40 | 0.489 | 0 | 0 | 1.00 |
| south | 4165 | 0.29 | 0.454 | 0 | 0 | 1.00 |
| smsa | 4165 | 0.65 | 0.476 | 0 | 1 | 1.00 |
| ms | 4165 | 0.81 | 0.389 | 0 | 1 | 1.00 |
| fem | 4165 | 0.11 | 0.316 | 0 | 0 | 1.00 |
| union | 4165 | 0.36 | 0.481 | 0 | 0 | 1.00 |
| ed | 4165 | 12.85 | 2.788 | 4 | 12 | 17.00 |
| blk | 4165 | 0.07 | 0.259 | 0 | 0 | 1.00 |
| lwage | 4165 | 6.68 | 0.462 | 4.61 | 6.68461 | 8.54 |

The table's font can be specified with the option `font()`. However, in Stata 17 the suboption `stylesrc(own)` is needed in the option `append()` so that the newly appended section can follow its own font. Otherwise, `font()` will be ignored, and the style definitions will keep consistent with the previous document.

```
. sum2docx exp - lwage using "summary.docx", append(stylesrc(own))
>   stats(N mean sd min median max) font("Consolas", 10, black)
>   title("Table 3. Summary statistics with specified fonts")
Summary statistics table has been written to file summary.docx.
```

Table 3. Summary statistics with specified fonts

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.854 | 10.966 | 1.000 | 18.000 | 51.000 |
| wks | 4165 | 46.812 | 5.129 | 5.000 | 48.000 | 52.000 |
| occ | 4165 | 0.511 | 0.500 | 0.000 | 1.000 | 1.000 |
| ind | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 1.000 |
| south | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 1.000 |
| smsa | 4165 | 0.654 | 0.476 | 0.000 | 1.000 | 1.000 |
| ms | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 |
| fem | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 1.000 |
| union | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 1.000 |
| ed | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 17.000 |
| blk | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 1.000 |
| lwage | 4165 | 6.676 | 0.462 | 4.605 | 6.685 | 8.537 |

Notes to the table can be added with the option `note()`. *cell_fmt_options* such as `font()` can be added as suboptions in Stata 17.

```
. sum2docx exp - lwage using "summary.docx", append
>  stats(N mean sd min median max) title("Table 4. Summary statistics with notes")
>  note("Data source: Chapter 8 of Microeconometrics Using Stata, Second Edition",
>  font("Consolas", 7, black))
Summary statistics table has been written to file summary.docx.
```

Table 4. Summary statistics with notes

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.854 | 10.966 | 1.000 | 18.000 | 51.000 |
| wks | 4165 | 46.812 | 5.129 | 5.000 | 48.000 | 52.000 |
| occ | 4165 | 0.511 | 0.500 | 0.000 | 1.000 | 1.000 |
| ind | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 1.000 |
| south | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 1.000 |
| smsa | 4165 | 0.654 | 0.476 | 0.000 | 1.000 | 1.000 |
| ms | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 |
| fem | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 1.000 |
| union | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 1.000 |
| ed | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 17.000 |
| blk | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 1.000 |
| lwage | 4165 | 6.676 | 0.462 | 4.605 | 6.685 | 8.537 |

Data source: Chapter 8 of Microeconometrics Using Stata, Second Edition

Column width can be adjusted with the option `layout()`. The layout type may be `fixed`, `autofitwindow`, or `autofitcontents`; the default is `layout(autofitwindow)`. The following commands report a table with the column width resized to fit the contents.

```
. sum2docx exp - lwage using "summary.docx", append
>  stats(N mean sd min median max)
>  title("Table 5. Summary statistics - autofitcontent")
>  layout(autofitcontent)
Summary statistics table has been written to file summary.docx.
```

Table 5. Summary statistics - autofitcontent

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.854 | 10.966 | 1.000 | 18.000 | 51.000 |
| wks | 4165 | 46.812 | 5.129 | 5.000 | 48.000 | 52.000 |
| occ | 4165 | 0.511 | 0.500 | 0.000 | 1.000 | 1.000 |
| ind | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 1.000 |
| south | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 1.000 |
| smsa | 4165 | 0.654 | 0.476 | 0.000 | 1.000 | 1.000 |
| ms | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 |
| fem | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 1.000 |
| union | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 1.000 |
| ed | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 17.000 |
| blk | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 1.000 |
| lwage | 4165 | 6.676 | 0.462 | 4.605 | 6.685 | 8.537 |

The option `varlabel` can be used to output variables' labels rather than variables' names in the first column of the table. If the table is a bit wide, the option `landscape` can change the document orientation to landscape.

```
. sum2docx exp - lwage using "summary.docx", append(pagebreak) landscape
>   stats(N mean sd min p25 median p75 max) varlabel
>   title("Table 6. Summary statistics with variables' labels")
Summary statistics table has been written to file summary.docx.
```

Table 6. Summary statistics with variables' labels

| VarName | Obs | Mean | SD | Min | P25 | Median | P75 | Max |
|---|---|---|---|---|---|---|---|---|
| Years of full-time work experience | 4165 | 19.854 | 10.966 | 1.000 | 11.000 | 18.000 | 29.000 | 51.000 |
| Weeks worked | 4165 | 46.812 | 5.129 | 5.000 | 46.000 | 48.000 | 50.000 | 52.000 |
| Occupation; occ==1 if in a blue-collar occupation | 4165 | 0.511 | 0.500 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| Industry; ind==1 if working in a manufacturing industry | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Residence; south==1 if in the South area | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| smsa==1 if in the standard metropolitan statistical area | 4165 | 0.654 | 0.476 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| Marital status | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Female or male | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| If wage set be a union contract | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Years of education | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 12.000 | 16.000 | 17.000 |
| Black | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| log wage | 4165 | 6.676 | 0.462 | 4.605 | 6.395 | 6.685 | 6.953 | 8.537 |

We can also change page size to fit a wide table.

```
. sum2docx exp - lwage using "summary.docx", append(pagebreak) pagesize(A3)
>   stats(N mean sd min p25 median p75 max) varlabel
>   title("Table 7. Summary statistics with specified page size")
Summary statistics table has been written to file summary.docx.
```

Table 7. Summary statistics with specified page size

| VarName | Obs | Mean | SD | Min | P25 | Median | P75 | Max |
|---|---|---|---|---|---|---|---|---|
| Years of full-time work experience | 4165 | 19.854 | 10.966 | 1.000 | 11.000 | 18.000 | 29.000 | 51.000 |
| Weeks worked | 4165 | 46.812 | 5.129 | 5.000 | 46.000 | 48.000 | 50.000 | 52.000 |
| Occupation; occ==1 if in a blue-collar occupation | 4165 | 0.511 | 0.500 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| Industry; ind==1 if working in a manufacturing industry | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Residence; south==1 if in the South area | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| smsa==1 if in the standard metropolitan statistical area | 4165 | 0.654 | 0.476 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| Marital status | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Female or male | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| If wage set be a union contract | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Years of education | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 12.000 | 16.000 | 17.000 |
| Black | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| log wage | 4165 | 6.676 | 0.462 | 4.605 | 6.395 | 6.685 | 6.953 | 8.537 |

Finally, page margins can be set via the option `margin()`. The `margin()` option can be specified multiple times in one command to account for different margin settings.

```
. sum2docx exp - lwage using "summary.docx", append(pagebreak)
>   stats(N mean sd min median max) margin(top, 1cm) margin(bottom, 1in)
>   title("Table 8. Summary statistics with specified page margins")
Summary statistics table has been written to file summary.docx.
```

Table 8. Summary statistics with specified page margins

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.854 | 10.966 | 1.000 | 18.000 | 51.000 |
| wks | 4165 | 46.812 | 5.129 | 5.000 | 48.000 | 52.000 |
| occ | 4165 | 0.511 | 0.500 | 0.000 | 1.000 | 1.000 |
| ind | 4165 | 0.395 | 0.489 | 0.000 | 0.000 | 1.000 |
| south | 4165 | 0.290 | 0.454 | 0.000 | 0.000 | 1.000 |
| smsa | 4165 | 0.654 | 0.476 | 0.000 | 1.000 | 1.000 |
| ms | 4165 | 0.814 | 0.389 | 0.000 | 1.000 | 1.000 |
| fem | 4165 | 0.113 | 0.316 | 0.000 | 0.000 | 1.000 |
| union | 4165 | 0.364 | 0.481 | 0.000 | 0.000 | 1.000 |
| ed | 4165 | 12.845 | 2.788 | 4.000 | 12.000 | 17.000 |
| blk | 4165 | 0.072 | 0.259 | 0.000 | 0.000 | 1.000 |
| lwage | 4165 | 6.676 | 0.462 | 4.605 | 6.685 | 8.537 |

## 2.2 corr2docx

Usually, a correlation table is needed in empirical research. The correlation coefficient table reports the linear correlation between each pair of variables. All the pairwise correlations need to be reported with a certain formatted style, such as the number of decimals to be reported and whether using asterisks is needed to indicate the significance level. In addition, some journals may also need both the Spearman and the Pearson correlations. The command `corr2docx` can be used to achieve such tasks with one line of commands. It can report both the Spearman and the Pearson correlation coefficient matrices individually or together in a formatted table, which it then produces in a `.docx` document. The Spearman correlation coefficient matrix is the result of the command `spearman`, and the Pearson correlation coefficient matrix comes from the command `correlate`. The syntax for `corr2docx` is as follows:

`corr2docx` *varlist* [ *if* ] [ *in* ] [ *weight* ] `using` *filename* [ `,` *options* ]

The definitions of *varlist* and *filename* are the same as those for `sum2docx`. `aweight`s and `fweight`s are allowed; see [U] **11.1.6 weight**. The options are as shown below:

| *options* | Description |
|---|---|
| `fmt(`*fmt*`)` | display format of the correlation coefficients; default is `fmt(%9.3f)` |
| `star[(`*symbol_level* `[...])]` | specify the significance level of the coefficient of correlation from higher to lower; by default, it is `*** p < 0.01, ** p < 0.05, * p < 0.1` |
| `starsps` | set the format of symbols to be superscript |
| `nodiagonal` | omit the diagonal cells, which should be 1 by definition |
| `pearson(`*pearson_opts*`)` | suboptions for Pearson's correlation coefficient matrix |
| `spearman(`*spearman_opts*`)` | suboptions for Spearman's correlation coefficient matrix |

| *pearson_opts* | Description |
|---|---|
| `pw` | display all pairwise correlation coefficients |
| `bonferroni` | use Bonferroni-adjusted significance level |
| `sidak` | use Sidak-adjusted significance level |
| `ignore` | do not report Pearson's correlation coefficients |

| *spearman_opts* | Description |
|---|---|
| `pw` | calculate all pairwise correlation coefficients by using all available data |
| `bonferroni` | use Bonferroni-adjusted significance level |
| `sidak` | use Sidak-adjusted significance level |
| `ignore` | do not report Spearman's correlation coefficients |

`corr2docx` can now be used to put a correlation coefficient table into a `.docx` document with the following command:

```
. corr2docx exp - lwage using "correlation.docx", replace landscape
>  title("Table 9. Correlation coefficient")
>  note("The upper triangle is a Spearman correlation. The lower triangle is a
>  Pearson correlation.")
Correlation matrix has been written to file correlation.docx.
```

Table 9. Correlation coefficient

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | -0.143 | 0.078 | 0.163 | -0.053 | 0.048 | 0.164 | -0.091 | 0.064 | -0.214 | 0.038 | 0.226 |
| wks | -0.034 | 1 | 0.033 | -0.021 | 0.082 | -0.039 | 0.028 | -0.058 | -0.198 | -0.030 | -0.011 | -0.100 |
| occ | 0.082 | -0.004 | 1 | 0.226 | 0.041 | -0.202 | 0.071 | -0.085 | 0.378 | -0.631 | 0.084 | -0.304 |
| ind | 0.159 | 0.040 | 0.226 | 1 | -0.077 | -0.069 | 0.170 | -0.178 | 0.146 | -0.238 | -0.047 | 0.023 |
| south | -0.053 | 0.029 | 0.041 | -0.077 | 1 | -0.135 | -0.040 | 0.052 | -0.163 | -0.119 | 0.122 | -0.181 |
| smsa | 0.053 | 0.018 | -0.202 | -0.069 | -0.135 | 1 | -0.106 | 0.104 | 0.027 | 0.167 | 0.115 | 0.222 |
| ms | 0.161 | 0.063 | 0.071 | 0.170 | -0.040 | -0.106 | 1 | -0.723 | 0.115 | -0.004 | -0.215 | 0.266 |
| fem | -0.092 | -0.083 | -0.085 | -0.178 | 0.052 | 0.104 | -0.723 | 1 | -0.113 | -0.014 | 0.209 | -0.310 |
| union | 0.059 | -0.155 | 0.378 | 0.146 | -0.163 | 0.027 | 0.115 | -0.113 | 1 | -0.281 | 0.047 | 0.009 |
| ed | -0.218 | -0.007 | -0.619 | -0.236 | -0.122 | 0.184 | -0.008 | -0.001 | -0.270 | 1 | -0.128 | 0.374 |
| blk | 0.040 | -0.032 | 0.084 | -0.047 | 0.122 | 0.115 | -0.215 | 0.209 | 0.047 | -0.120 | 1 | -0.178 |
| lwage | 0.209 | 0.059 | -0.318 | 0.046 | -0.180 | 0.224 | 0.288 | -0.325 | 0.009 | 0.394 | -0.190 | 1 |

The upper triangle is a Spearman correlation. The lower triangle is a Pearson correlation.

The table contains both Spearman and Pearson correlation coefficients, and the format of the coefficients is `%9.3f` by default. These two matrices can be modified by using the options `spearman()` and `pearson()`. For example, Pearson correlation coefficients can be reported individually in the table and their formats changed to `%9.4f` via the following command:

```
. corr2docx exp - lwage using "correlation.docx", append fmt(%9.4f)
> spearman(ignore) title("Table 10. Pearson correlation coefficient")
Correlation matrix has been written to file correlation.docx.
```

Table 10. Pearson correlation coefficient

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | | | | | | | | | | | |
| wks | -0.0337 | 1 | | | | | | | | | | |
| occ | 0.0822 | -0.0038 | 1 | | | | | | | | | |
| ind | 0.1591 | 0.0404 | 0.2260 | 1 | | | | | | | | |
| south | -0.0527 | 0.0292 | 0.0413 | -0.0769 | 1 | | | | | | | |
| smsa | 0.0526 | 0.0180 | -0.2018 | -0.0689 | -0.1350 | 1 | | | | | | |
| ms | 0.1606 | 0.0625 | 0.0706 | 0.1701 | -0.0403 | -0.1060 | 1 | | | | | |
| fem | -0.0922 | -0.0833 | -0.0847 | -0.1778 | 0.0516 | 0.1044 | -0.7228 | 1 | | | | |
| union | 0.0591 | -0.1548 | 0.3784 | 0.1465 | -0.1628 | 0.0271 | 0.1147 | -0.1132 | 1 | | | |
| ed | -0.2182 | -0.0067 | -0.6194 | -0.2365 | -0.1216 | 0.1843 | -0.0083 | -0.0012 | -0.2695 | 1 | | |
| blk | 0.0404 | -0.0319 | 0.0837 | -0.0475 | 0.1218 | 0.1154 | -0.2150 | 0.2086 | 0.0471 | -0.1196 | 1 | |
| lwage | 0.2093 | 0.0585 | -0.3176 | 0.0458 | -0.1804 | 0.2240 | 0.2875 | -0.3250 | 0.0087 | 0.3939 | -0.1895 | 1 |

The significance level of the correlation coefficients can also be specified from higher to lower using the option `star()`. The default is `*** p < 0.01, ** p < 0.05, * p < 0.1` if this option is specified without any symbol and significance level.

```
. corr2docx exp - lwage using "correlation.docx", append(stylesrc(own))
>  font("times new roman", 9) fmt(%9.4f) star
>  title("Table 11. Correlation coefficient with significance level")
>  note("The upper triangle is a Spearman correlation. The lower triangle is a
>  Pearson correlation.")
Correlation matrix has been written to file correlation.docx.
```

Table 11. Correlation coefficient with significance level

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | -0.1432*** | 0.0776*** | 0.1631*** | -0.0532*** | 0.0480*** | 0.1637*** | -0.0913*** | 0.0635*** | -0.2135*** | 0.0383** | 0.2262*** |
| wks | -0.0337** | 1 | 0.0334** | -0.0209 | 0.0823*** | -0.0388** | 0.0285* | -0.0580*** | -0.1978*** | -0.0300* | -0.0113 | -0.1002*** |
| occ | 0.0822*** | -0.0038 | 1 | 0.2260*** | 0.0413*** | -0.2018*** | 0.0706*** | -0.0847*** | 0.3784*** | -0.6309*** | 0.0837*** | -0.3043*** |
| ind | 0.1591*** | 0.0404*** | 0.2260*** | 1 | -0.0769*** | -0.0689*** | 0.1701*** | -0.1778*** | 0.1465*** | -0.2381*** | -0.0475*** | 0.0230 |
| south | -0.0527*** | 0.0292* | 0.0413*** | -0.0769*** | 1 | -0.1350*** | -0.0403*** | 0.0516*** | -0.1628*** | -0.1191*** | 0.1218*** | -0.1810*** |
| smsa | 0.0526*** | 0.0180 | -0.2018*** | -0.0689*** | -0.1350*** | 1 | -0.1060*** | 0.1044*** | 0.0271* | 0.1670*** | 0.1154*** | 0.2216*** |
| ms | 0.1606*** | 0.0625*** | 0.0706*** | 0.1701*** | -0.0403*** | -0.1060*** | 1 | -0.7228*** | 0.1147*** | -0.0044 | -0.2150*** | 0.2660*** |
| fem | -0.0922*** | -0.0833*** | -0.0847*** | -0.1778*** | 0.0516*** | 0.1044*** | -0.7228*** | 1 | -0.1132*** | -0.0141 | 0.2086*** | -0.3103*** |
| union | 0.0591*** | -0.1548*** | 0.3784*** | 0.1465*** | -0.1628*** | 0.0271* | 0.1147*** | -0.1132*** | 1 | -0.2813*** | 0.0471*** | 0.0094 |
| ed | -0.2182*** | -0.0067 | -0.6194*** | -0.2365*** | -0.1216*** | 0.1843*** | -0.0083 | -0.0012 | -0.2695*** | 1 | -0.1276*** | 0.3743*** |
| blk | 0.0404*** | -0.0319** | 0.0837*** | -0.0475*** | 0.1218*** | 0.1154*** | -0.2150*** | 0.2086*** | 0.0471*** | -0.1196*** | 1 | -0.1783*** |
| lwage | 0.2093*** | 0.0585*** | -0.3176*** | 0.0458*** | -0.1804*** | 0.2240*** | 0.2875*** | -0.3250*** | 0.0087 | 0.3939*** | -0.1895*** | 1 |

The upper triangle is a Spearman correlation. The lower triangle is a Pearson correlation.

If a single star is preferred to indicate the 5% significance level, this `star()` option can be specified as `star(* 0.05)`.

```
. corr2docx exp - lwage using "correlation.docx", append fmt(%9.4f)
>  star(* 0.05)
>  title("Table 12. Correlation coefficient with significance level")
>  note("The upper triangle is a Spearman correlation. The lower triangle is a
>  Pearson correlation.")
Correlation matrix has been written to file correlation.docx.
```

Table 12. Correlation coefficient with significance level

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | -0.1432* | 0.0776* | 0.1631* | -0.0532* | 0.0480* | 0.1637* | -0.0913* | 0.0635* | -0.2135* | 0.0383* | 0.2262* |
| wks | -0.0337* | 1 | 0.0334* | -0.0209 | 0.0823* | -0.0388* | 0.0285 | -0.0580* | -0.1978* | -0.0300 | -0.0113 | -0.1002* |
| occ | 0.0822* | -0.0038 | 1 | 0.2260* | 0.0413* | -0.2018* | 0.0706* | -0.0847* | 0.3784* | -0.6309* | 0.0837* | -0.3043* |
| ind | 0.1591* | 0.0404* | 0.2260* | 1 | -0.0769* | -0.0689* | 0.1701* | -0.1778* | 0.1465* | -0.2381* | -0.0475* | 0.0230 |
| south | -0.0527* | 0.0292 | 0.0413* | -0.0769* | 1 | -0.1350* | -0.0403* | 0.0516* | -0.1628* | -0.1191* | 0.1218* | -0.1810* |
| smsa | 0.0526* | 0.0180 | -0.2018* | -0.0689* | -0.1350* | 1 | -0.1060* | 0.1044* | 0.0271 | 0.1670* | 0.1154* | 0.2216* |
| ms | 0.1606* | 0.0625* | 0.0706* | 0.1701* | -0.0403* | -0.1060* | 1 | -0.7228* | 0.1147* | -0.0044 | -0.2150* | 0.2660* |
| fem | -0.0922* | -0.0833* | -0.0847* | -0.1778* | 0.0516* | 0.1044* | -0.7228* | 1 | -0.1132* | -0.0141 | 0.2086* | -0.3103* |
| union | 0.0591* | -0.1548* | 0.3784* | 0.1465* | -0.1628* | 0.0271 | 0.1147* | -0.1132* | 1 | -0.2813* | 0.0471* | 0.0094 |
| ed | -0.2182* | -0.0067 | -0.6194* | -0.2365* | -0.1216* | 0.1843* | -0.0083 | -0.0012 | -0.2695* | 1 | -0.1276* | 0.3743* |
| blk | 0.0404* | -0.0319* | 0.0837* | -0.0475* | 0.1218* | 0.1154* | -0.2150* | 0.2086* | 0.0471* | -0.1196* | 1 | -0.1783* |
| lwage | 0.2093* | 0.0585* | -0.3176* | 0.0458* | -0.1804* | 0.2240* | 0.2875* | -0.3250* | 0.0087 | 0.3939* | -0.1895* | 1 |

The upper triangle is a Spearman correlation. The lower triangle is a Pearson correlation.

Bonferroni-adjusted or Sidak-adjusted significance levels could also be used in both matrices.

```
. corr2docx exp - lwage using "correlation.docx", star fmt(%9.4f)
>   append(stylesrc(own))
>   title("Table 13. Correlation coefficient with adjusted significance level")
>   spearman(pw bonferroni) pearson(pw sidak) font("times new roman", 9)
>   note("The upper triangle is a Spearman correlation. The lower triangle is a
>   Pearson correlation.")
Correlation matrix has been written to file correlation.docx.
```

Table 13. Correlation coefficient with adjusted significance level

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | -0.1432*** | 0.0776*** | 0.1631*** | -0.0532** | 0.0480 | 0.1637*** | -0.0913*** | 0.0635*** | -0.2135*** | 0.0383 | 0.2262*** |
| wks | -0.0337 | 1 | 0.0334 | -0.0209 | 0.0823*** | -0.0388 | 0.0285 | -0.0580** | -0.1978*** | -0.0300 | -0.0113 | -0.1002*** |
| occ | 0.0822*** | -0.0038 | 1 | 0.2260*** | 0.0413 | -0.2018*** | 0.0706*** | -0.0847*** | 0.3784*** | -0.6309*** | 0.0837*** | -0.3043*** |
| ind | 0.1591*** | 0.0404 | 0.2260*** | 1 | -0.0769*** | -0.0689*** | 0.1701*** | -0.1778*** | 0.1465*** | -0.2381*** | -0.0475 | 0.0230 |
| south | -0.0527** | 0.0292 | 0.0413 | -0.0769*** | 1 | -0.1350*** | -0.0403 | 0.0516* | -0.1628*** | -0.1191*** | 0.1218*** | -0.1810*** |
| smsa | 0.0526** | 0.0180 | -0.2018*** | -0.0689*** | -0.1350*** | 1 | -0.1060*** | 0.1044*** | 0.0271 | 0.1670*** | 0.1154*** | 0.2216*** |
| rms | 0.1606*** | 0.0625*** | 0.0706*** | 0.1701*** | -0.0403 | -0.1060*** | 1 | -0.7228*** | 0.1147*** | -0.0044 | -0.2150*** | 0.2660*** |
| fem | -0.0922*** | -0.0833*** | -0.0847*** | -0.1778*** | 0.0516* | 0.1044*** | -0.7228*** | 1 | -0.1132*** | -0.0141 | 0.2086*** | -0.3103*** |
| union | 0.0591*** | -0.1548*** | 0.3784*** | 0.1465*** | -0.1628*** | 0.0271 | 0.1147*** | -0.1132*** | 1 | -0.2813*** | 0.0471 | 0.0094 |
| ed | -0.2182*** | -0.0067 | -0.6194*** | -0.2365*** | -0.1216*** | 0.1843*** | -0.0083 | -0.0012 | -0.2695*** | 1 | -0.1276*** | 0.3743*** |
| blk | 0.0404 | -0.0319 | 0.0837*** | -0.0475 | 0.1218*** | 0.1154*** | -0.2150*** | 0.2086*** | 0.0471 | -0.1196*** | 1 | -0.1783*** |
| lwage | 0.2093*** | 0.0585** | -0.3176*** | 0.0458 | -0.1804*** | 0.2240*** | 0.2875*** | -0.3250*** | 0.0087 | 0.3939*** | -0.1895*** | 1 |

The upper triangle is a Spearman correlation. The lower triangle is a Pearson correlation.

## 2.3 t2docx

Suppose that evaluating the differences of a list of variables between two groups (for example, the treat and control groups before an experiment) is required by a mean-comparison $t$ test. `t2docx` can be used to achieve this task. This command first uses `ttest` on every individual variable and then reports the most important statistics to one table. Those statistics include the sample size, mean value of each variable in both subsamples, the differences of the mean values, and the corresponding $t$ statistics for each $t$ test. The command also adds stars to indicate the significance level. The basic syntax for this command is written below:

`t2docx` *varlist* [ *if* ] [ *in* ] `using` *filename*`, by(`*groupvar*`)` [ *options* ]

The definitions of *varlist* and *filename* are the same as those for the two commands above. The following options are also provided:

| *options* | Description |
|---|---|
| `by(`*groupvar*`)` | specify a variable that defines the groups |
| `diff(`*value1 value2*`)` | select two values of the *groupvar* for mean-comparison $t$ test if *groupvar* has more than two values; if the *groupvar* can define only two subgroups along with the *if* and *in* conditions, this option can be ignored |
| `fmt(`*fmt*`)` | specify the display formats for group means and their differences, $t$ statistics, and $p$-values, as well as their standard deviations if needed; default is `fmt(%9.3f)` |
| `not` | do not output $t$ statistics |
| `p` | output $p$-value instead of $t$ statistics |
| `se` | output standard error instead of $t$ statistics |
| `nostar` | do not output significance stars |
| `star[(`*symbol_level* `[...])]` | use symbols to indicate significance level |
| `staraux` | the significance symbols will be printed next to the $t$ statistics (or standard errors, etc.) instead of the coefficient |
| `starsps` | set the format of symbols to be superscript |
| `unequal` | allow unpaired data to have unequal variances |
| `welch` | use Welch's approximation |

The option `by()` is required to define the split samples. The variable in the `by()` option must take at least two different values to define several different subsamples. For example, the sample could be divided into two groups by log wage (`lwage`). The differences of other variables between the two groups could be compared as follows:

```
. egen medianwage = median(lwage)
. generate group = lwage >= medianwage
. t2docx exp - blk using "t-test.docx", replace by(group)
> title("Table 14. t-test results")
t-test table has been written to file t-test.docx.
```

Table 14. t-test results

| varname | obs(0) | mean(0) | obs(1) | mean(1) | mean-diff | t |
|---|---|---|---|---|---|---|
| exp | 2024 | 17.997 | 2141 | 21.609 | -3.612*** | -10.770 |
| wks | 2024 | 46.695 | 2141 | 46.922 | -0.227 | -1.430 |
| occ | 2024 | 0.635 | 2141 | 0.394 | 0.241*** | 15.996 |
| ind | 2024 | 0.398 | 2141 | 0.393 | 0.004 | 0.294 |
| south | 2024 | 0.366 | 2141 | 0.219 | 0.147*** | 10.551 |
| smsa | 2024 | 0.560 | 2141 | 0.742 | -0.182*** | -12.561 |
| ms | 2024 | 0.736 | 2141 | 0.888 | -0.152*** | -12.874 |
| fem | 2024 | 0.192 | 2141 | 0.037 | 0.155*** | 16.290 |
| union | 2024 | 0.361 | 2141 | 0.367 | -0.006 | -0.432 |
| ed | 2024 | 11.964 | 2141 | 13.678 | -1.714*** | -20.833 |
| blk | 2024 | 0.111 | 2141 | 0.036 | 0.075*** | 9.402 |

The results include the number of observations, the mean of selected variables in each group, the mean difference between two groups, and the $t$ statistics of the mean difference. Symbols to show the significance levels are printed next to the mean differences. The format of means, mean differences, and $t$ statistics is `%9.3f` by default, which could be changed by the option `fmt()`. The $t$ statistics can also be replaced by $p$-values (standard errors) using the option `p` (`se`). The significance levels are then printed next to the $p$-values.

```
. t2docx exp - blk using "t-test.docx", append by(group) p staraux fmt(%9.2f)
> title("Table 15. t-test results with p-value")
t-test table has been written to file t-test.docx.
```

Table 15. t-test results with p-value

| varname | obs(0) | mean(0) | obs(1) | mean(1) | mean-diff | p |
|---|---|---|---|---|---|---|
| exp | 2024 | 18.00 | 2141 | 21.61 | -3.61 | 0.00*** |
| wks | 2024 | 46.69 | 2141 | 46.92 | -0.23 | 0.15 |
| occ | 2024 | 0.63 | 2141 | 0.39 | 0.24 | 0.00*** |
| ind | 2024 | 0.40 | 2141 | 0.39 | 0.00 | 0.77 |
| south | 2024 | 0.37 | 2141 | 0.22 | 0.15 | 0.00*** |
| smsa | 2024 | 0.56 | 2141 | 0.74 | -0.18 | 0.00*** |
| ms | 2024 | 0.74 | 2141 | 0.89 | -0.15 | 0.00*** |
| fem | 2024 | 0.19 | 2141 | 0.04 | 0.15 | 0.00*** |
| union | 2024 | 0.36 | 2141 | 0.37 | -0.01 | 0.67 |
| ed | 2024 | 11.96 | 2141 | 13.68 | -1.71 | 0.00*** |
| blk | 2024 | 0.11 | 2141 | 0.04 | 0.07 | 0.00*** |

The format of symbols to show the significance levels can be set to be superscript using the option `starsps`:

```
. t2docx exp - blk using "t-test.docx", append by(group) p starsps fmt(%9.2f)
> title("Table 16. t-test results with superscript symbol")
t-test table has been written to file t-test.docx.
```

Table 16. t-test results with superscript symbol

| varname | obs(0) | mean(0) | obs(1) | mean(1) | mean-diff | p |
|---|---|---|---|---|---|---|
| exp | 2024 | 18.00 | 2141 | 21.61 | -3.61*** | 0.00 |
| wks | 2024 | 46.69 | 2141 | 46.92 | -0.23 | 0.15 |
| occ | 2024 | 0.63 | 2141 | 0.39 | 0.24*** | 0.00 |
| ind | 2024 | 0.40 | 2141 | 0.39 | 0.00 | 0.77 |
| south | 2024 | 0.37 | 2141 | 0.22 | 0.15*** | 0.00 |
| smsa | 2024 | 0.56 | 2141 | 0.74 | -0.18*** | 0.00 |
| ms | 2024 | 0.74 | 2141 | 0.89 | -0.15*** | 0.00 |
| fem | 2024 | 0.19 | 2141 | 0.04 | 0.15*** | 0.00 |
| union | 2024 | 0.36 | 2141 | 0.37 | -0.01 | 0.67 |
| ed | 2024 | 11.96 | 2141 | 13.68 | -1.71*** | 0.00 |
| blk | 2024 | 0.11 | 2141 | 0.04 | 0.07*** | 0.00 |

When the *groupvar* has more than two different values, `t2docx` can report the number of observations and the mean of selected variables of each group and the mean-comparison results between two selected groups. The two groups can be defined by the option `diff()`. For example, the sample can be divided into three or more groups by log wage (`lwage`) using the function `xtile()` in `egenmore` (Cox 2000), and then the differences of other variables between the first and third groups can be compared:

```
. egen newgroup1 = xtile(lwage), nq(3)
. t2docx exp - blk using "t-test.docx", append by(newgroup1) diff(1 3)
> fmt(%9.2f) title("Table 17. Three groups")
t-test table has been written to file t-test.docx.
```

Table 17. Three groups

| varname | obs(1) | mean(1) | obs(2) | mean(2) | obs(3) | mean(3) | mean-diff | t |
|---|---|---|---|---|---|---|---|---|
| exp | 1391 | 17.69 | 1391 | 19.22 | 1383 | 22.67 | -4.98*** | -12.22 |
| wks | 1391 | 46.67 | 1391 | 46.67 | 1383 | 47.09 | -0.42** | -2.11 |
| occ | 1391 | 0.66 | 1391 | 0.55 | 1383 | 0.32 | 0.34*** | 19.15 |
| ind | 1391 | 0.37 | 1391 | 0.43 | 1383 | 0.38 | -0.02 | -0.90 |
| south | 1391 | 0.40 | 1391 | 0.26 | 1383 | 0.21 | 0.18*** | 10.67 |
| smsa | 1391 | 0.53 | 1391 | 0.66 | 1383 | 0.77 | -0.24*** | -13.64 |
| ms | 1391 | 0.69 | 1391 | 0.84 | 1383 | 0.92 | -0.23*** | -15.96 |
| fem | 1391 | 0.24 | 1391 | 0.08 | 1383 | 0.02 | 0.22*** | 18.26 |
| union | 1391 | 0.30 | 1391 | 0.48 | 1383 | 0.31 | -0.00 | -0.18 |
| ed | 1391 | 11.69 | 1391 | 12.79 | 1383 | 14.06 | -2.38*** | -23.63 |
| blk | 1391 | 0.13 | 1391 | 0.06 | 1383 | 0.03 | 0.10*** | 10.30 |

If there are $k$ groups, the output table should have $2 \times k + 3$ columns, which makes the table a bit wide. In this case, the `landscape` option can be used to produce a wide table, as presented below. Here four groups are used as an example.

```
. egen newgroup2 = xtile(lwage), nq(4)
. t2docx exp - blk using "t-test.docx", append(pagebreak) landscape
> by(newgroup2) diff(1 4) fmt(%9.2f) title("Table 18. Four groups")
t-test table has been written to file t-test.docx.
```

Table 18. Four groups

| varname | obs(1) | mean(1) | obs(2) | mean(2) | obs(3) | mean(3) | obs(4) | mean(4) | mean-diff | t |
|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1042 | 17.15 | 1045 | 18.86 | 1037 | 20.62 | 1041 | 22.79 | -5.64*** | -11.94 |
| wks | 1042 | 46.52 | 1045 | 46.90 | 1037 | 46.71 | 1041 | 47.12 | -0.60** | -2.55 |
| occ | 1042 | 0.69 | 1045 | 0.57 | 1037 | 0.46 | 1041 | 0.32 | 0.37*** | 18.12 |
| ind | 1042 | 0.35 | 1045 | 0.45 | 1037 | 0.41 | 1041 | 0.38 | -0.03 | -1.57 |
| south | 1042 | 0.42 | 1045 | 0.30 | 1037 | 0.23 | 1041 | 0.21 | 0.20*** | 10.29 |
| smsa | 1042 | 0.51 | 1045 | 0.62 | 1037 | 0.73 | 1041 | 0.76 | -0.25*** | -12.28 |
| ms | 1042 | 0.64 | 1045 | 0.84 | 1037 | 0.86 | 1041 | 0.92 | -0.28*** | -16.01 |
| fem | 1042 | 0.28 | 1045 | 0.10 | 1037 | 0.06 | 1041 | 0.02 | 0.26*** | 17.98 |
| union | 1042 | 0.27 | 1045 | 0.46 | 1037 | 0.43 | 1041 | 0.30 | -0.04* | -1.91 |
| ed | 1042 | 11.51 | 1045 | 12.50 | 1037 | 13.29 | 1041 | 14.08 | -2.57*** | -22.21 |
| blk | 1042 | 0.14 | 1045 | 0.08 | 1037 | 0.04 | 1041 | 0.03 | 0.11*** | 9.59 |

## 2.4 reg2docx

Regression is an important aspect in empirical research. Although various official and community-contributed statistical packages are good at fitting regression models, the output is not in a standard table form and ready for publication. The command `reg2docx` can report all kinds of regression results in a table and store them in a `.docx` document. This command first uses regression commands to fit one or more regression models, then, after each regression command uses the `estimate store` command to save the regression results to memory. Last, `reg2docx` is called to design a formatted output table for all the saved models and then place them into a `.docx` file. The syntax of this command is

`reg2docx` *modellist* `using` *filename* [ `,` *options* ]

where *modellist* is a list of results stored using the `estimate store` command. The command `reg2docx` has many options, which makes it powerful and flexible. In designing the function of these options, we owe a lot to authors of existing commands such as `estout` and `esttab` (Jann 2005, 2007; Jann and Long 2010). Those familiar with `estout` and `esttab` will find it easy to use `reg2docx`. The options are as follows:

| *options* | Description |
|---|---|
| `b`[ (*fmt*) ] | specify the format for coefficient |
| `t`\|`z`\|`p`\|`se`\|`ci`[ (*fmt*) ] | output $t$ statistics, $z$ statistics, $p$-values, standard error, or confidence interval, and specify the format |
| `scalars(`*scalarlist*`)` | specify the scalars to be output; the format of the scalar is defined in parentheses |
| `noconstant` | do not output intercept |
| `constant` | output intercept; the default |
| `noobs` | do not output the number of observations |
| `nostar` | do not output significance stars |
| `star`[ (*symbol_level* [ `...` ]) ] | output significance of the coefficients |
| `staraux` | print the significance stars next to the $t$ statistics (or standard errors) instead of the coefficient |
| `starsps` | set the format of symbols to be superscript |
| `mtitles(`*titlelist*`)` | specify model's title in the table header |
| `nomtitle` | suppress printing of model titles |
| `depvars` | print the name of the dependent variable of a model as the model's title in the table header |
| `order(`*list*`)` | change order of coefficients |
| `indicate(`*groups*`)` | indicate presence of parameters |
| `addfe(`*groups*`)` | specify user-added text about the presence of fixed effects |
| `drop(`*droplist*`)` | drop individual coefficients |
| `keep(`*keeplist*`)` | keep individual coefficients |
| `noparentheses` | do not print parentheses around $t$ statistics |
| `parentheses` | print parentheses around $t$ statistics |
| `brackets` | use brackets instead of parentheses |

When you use `reg2docx`, the first step is to fit regression models using commands such as `regress`, `xtreg`, `logit`, `ivreg2`, or `reghdfe`. Then, the results for regression models are stored via the `estimate store` command. Finally, `reg2docx` creates a table that is ready for publication because it contains all the saved models and presents them in a `.docx` document. Below is an example of a program that shows how this command works. The example includes `addest`, a community-contributed command developed by Over (2012).

```
. regress lwage exp exp2 wks occ ind south smsa ms union, vce(robust)
  (output omitted)
. test exp exp2
  (output omitted)
. addest, name("JointSignificance") value(`r(F)')
  (output omitted)
. estimates store m1
. xtreg lwage exp exp2 wks occ ind south smsa ms union, fe vce(cluster id)
  (output omitted)
. test exp exp2
  (output omitted)
. addest, name("JointSignificance") value(`r(F)')
  (output omitted)
. estimates store m2
. reg2docx m1 m2 using "regression.docx", replace
> title("Table 19. Regression results")
Regression table has been written to file regression.docx.
```

Table 19. Regression results

| | (1) lwage | (2) lwage |
|---|---|---|
| exp | 0.036*** | 0.113*** |
| | (15.545) | (27.953) |
| exp2 | -0.001*** | -0.000*** |
| | (-12.700) | (-5.075) |
| wks | 0.004*** | 0.001 |
| | (3.483) | (0.966) |
| occ | -0.318*** | -0.021 |
| | (-22.792) | (-1.131) |
| ind | 0.032** | 0.019 |
| | (2.507) | (0.847) |
| south | -0.114*** | -0.002 |
| | (-8.284) | (-0.021) |
| smsa | 0.159*** | -0.042 |
| | (12.233) | (-1.440) |
| ms | 0.320*** | -0.030 |
| | (18.930) | (-1.106) |
| union | 0.070*** | 0.033 |
| | (5.061) | (1.308) |
| _cons | 5.880*** | 4.649*** |
| | (91.509) | (59.595) |
| N | 4165 | 4165 |

If dropping or keeping some coefficients of variables is required, the option `drop()` or `keep()` is used. Another option, `order()`, can also be used to change the order of coefficients because it can arrange the most preferred coefficients at the top of the table. The following two commands generate the same table:

```
. reg2docx m1 m2 using "regression.docx", append keep(exp exp2) order(exp2 exp)
> title("Table 20. Regression results of selected variables")
Regression table has been written to file regression.docx.
. reg2docx m1 m2 using "regression.docx", append
> drop(wks occ ind south smsa ms union) order(exp2 exp)
> title("Table 20. Regression results of selected variables")
Regression table has been written to file regression.docx.
```

Table 20. Regression results of selected variables

| | (1) lwage | (2) lwage |
|---|---|---|
| exp2 | -0.001*** | -0.000*** |
| | (-12.700) | (-5.075) |
| exp | 0.036*** | 0.113*** |
| | (15.545) | (27.953) |
| _cons | 5.880*** | 4.649*** |
| | (91.509) | (59.595) |
| N | 4165 | 4165 |

The option `scalars()` can be used to add any `ereturn` scalars of the regression command to the table, such as $F$ statistics, $R^2$, adjusted $R^2$, and scalars added by the command `addest`.[2] Suppose it is necessary to add the number of observations, $R^2$, and $F$ statistics for the joint significance of variables `exp` and `exp2` because those numbers are already stored as e-class return scalars `e(N)`, `e(r2)`, and `e(JointSignificance)`, which are added by the command `addest` (Over 2012). The `scalars(N r2(%9.2f) JointSignificance(%9.2f))` option can be used to add those numbers to each model. The format of this option is shown below.

```
. reg2docx m1 m2 using "regression.docx", append keep(exp exp2)
> scalars(N r2(%9.2f) JointSignificance(%9.2f))
> title("Table 21. Regression results with scalars")
Regression table has been written to file regression.docx.
```

Table 21. Regression results with scalars

| | (1) | (2) |
|---|---|---|
| | lwage | lwage |
| exp | 0.036*** | 0.113*** |
| | (15.545) | (27.953) |
| exp2 | -0.001*** | -0.000*** |
| | (-12.700) | (-5.075) |
| _cons | 5.880*** | 4.649*** |
| | (91.509) | (59.595) |
| N | 4165 | 4165 |
| $R^2$ | 0.32 | 0.66 |
| JointSignificance | 168.95 | 1591.44 |

The $p$-value (`p`), $z$ statistics (`z`), standard error (`se`), and confidence interval (`ci`) could also be reported instead of the $t$ statistics under the coefficients.

```
. reg2docx m1 m2 using "regression.docx", append scalars(N r2(%9.2f))
>  ci(%9.4f) keep(exp exp2)
>  title("Table 22. Regression results with confidence intervals")
Regression table has been written to file regression.docx.
```

Table 22. Regression results with confidence intervals

| | (1) | (2) |
|---|---|---|
| | lwage | lwage |
| exp | 0.036*** | 0.113*** |
| | [0.0316,0.0407] | [0.1053,0.1212] |
| exp2 | -0.001*** | -0.000*** |
| | [-0.0008,-0.0006] | [-0.0006,-0.0003] |
| _cons | 5.880*** | 4.649*** |
| | [5.7543,6.0062] | [4.4956,4.8020] |
| N | 4165 | 4165 |
| $R^2$ | 0.32 | 0.66 |

2. You could use `search addest` to find and install it.

To indicate the presence of parameters, use the option `indicate()`. This option is useful when controlling fixed effects using a set of dummy variables and when it is necessary to report that the effects are controlled instead of their parameters. The presence of fixed effects can be specified by using the option `addfe()`, especially when the fixed-effects estimates are not stored.

```
. regress lwage exp exp2 wks occ ind south smsa ms union, vce(robust)
  (output omitted)
. estimates store m1
. xtreg lwage exp exp2 wks occ ind south smsa ms union i.t, fe vce(cluster id)
  (output omitted)
. estimates store m2
. reg2docx m1 m2 using "regression.docx", append scalars(N r2(%9.2f)) t(%9.2f)
>  keep(exp exp2) indicate("Time FE = *t")
>  addfe("Individual FE = No Yes")
>  title("Table 23. Regression results with fixed effects")
Regression table has been written to file regression.docx.
```

Table 23. Regression results with fixed effects

| | (1) | (2) |
|---|---|---|
| | lwage | lwage |
| exp | 0.036*** | 0.111*** |
| | (15.54) | (26.95) |
| exp2 | -0.001*** | -0.000*** |
| | (-12.70) | (-4.78) |
| _cons | 5.880*** | 4.667*** |
| | (91.51) | (60.30) |
| Time FE | No | Yes |
| Individual FE | No | Yes |
| N | 4165 | 4165 |
| $R^2$ | 0.32 | 0.66 |

Finally, variables' labels are generated instead of variables' names[3] and notes are added using the option `note()`.

```
. reg2docx m1 m2 using "regression.docx", append scalars(N r2(%9.2f)) t(%9.2f)
>   keep(exp exp2) indicate("Time FE = *t") varlabel
>   addfe("Individual FE = No Yes") note("* p<0.1, ** p<0.05, *** p<0.01")
>   title("Table 24. Regression results with labels and notes")
Regression table has been written to file regression.docx.
```

Table 24. Regression results with labels and notes

| | (1) lwage | (2) lwage |
|---|---|---|
| Years of full-time work experience | 0.036*** | 0.111*** |
| | (15.54) | (26.95) |
| Years of experience squared | -0.001*** | -0.000*** |
| | (-12.70) | (-4.78) |
| _cons | 5.880*** | 4.667*** |
| | (91.51) | (60.30) |
| Time FE | No | Yes |
| Individual FE | No | Yes |
| N | 4165 | 4165 |
| $R^2$ | 0.32 | 0.66 |

* p<0.1, ** p<0.05, *** p<0.01

## 2.5 Append individual documents

The four commands discussed above have reported the tables of summary statistics, correlation coefficient matrices, mean-comparison $t$ tests, and regression results to four individual files. When you use these four commands, the tables could be appended to one file with the option `append`. Or the command `putdocx append` could be used to append the contents of the four generated `.docx` documents:

```
. putdocx append "summary.docx" "correlation.docx" "t-test.docx"
> "regression.docx", saving("empirical_results.docx", replace) pagebreak
```

# 3 Example

This section shows how to generate an article by coalescing the text, tables, and graphs into one `.docx` document using `putdocx` and the commands introduced above. Like previous sections, `mus208psid.dta` will be used as the dataset. Note that the article to be generated is just an illustration on how to use the newly introduced commands. Hence, some illustrative words resulting from usage of the command `putdocx` are texts sourced from *Microeconometrics Using Stata, Second Edition*. This exercise goes through some steps. First, `mus208psid.dta` is imported into Stata.

3. Variables' names will be output for the variables without labels.

```
. use mus208psid.dta, clear
(A.C.Cameron & P.K.Trivedi (2022): Microeconometrics Using Stata, 2e)
```

Second, the textual parts of an article, such as the title, abstract, introduction, etc., are generated by `putdocx`. However, some simple sentences for illustrative purposes will be used, as shown below:

```
. putdocx clear
. putdocx begin, font("Times New Roman", 16, black)
. putdocx paragraph, halign(center)
. putdocx text ("The relationship between wages and work experience"), bold
> font("Times New Roman", 30, black)
. putdocx paragraph, halign(center)
. putdocx text ("Author's name"), font("Times New Roman", 16, black) bold
>  linebreak
. putdocx text ("Email: "), font("Times New Roman", 16, black) bold
. putdocx text ("author@email.com"), font("Times New Roman", 16, black)
. putdocx pagebreak
. putdocx paragraph, halign(left)
. putdocx text ("Abstract"), font("Times New Roman", 16, black) bold
. putdocx textblock begin
This article investigates the relationship between work experience and
 wages.  Based on the data from chapter 8 of
 <<dd_docx_display italic: "Microeconometrics Using Stata, Second Edition">>,
 we find that wages increase with experience until a peak and then decline.
putdocx textblock end
. putdocx paragraph, halign(left)
. putdocx text ("Introduction"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
A growing amount of empirical research studies the influential factors of
 wages.  In this study, we examine the relationship between wages and work
 experience.  Based on the data from the Panel Study of Income Dynamics
 (Cameron and Trivedi 2022), we find that wages increase with
 experience until a peak and then decline.
putdocx textblock end
. putdocx save "empirical_results.docx", replace
successfully replaced "C:/mypath/empirical_results.docx"
```

# The relationship between wages and work experience

**Author's name**
**Email:** author@email.com

**Abstract**

This article investigates the relationship between work experience and wages. Based on the data from chapter 8 of *Microeconometrics Using Stata, Second Edition*, we find that wages increase with experience until a peak and then decline.

## Introduction

A growing amount of empirical research studies the influential factors of wages. In this study, we examine the relationship between wages and work experience. Based on the data from the Panel Study of Income Dynamics (Cameron and Trivedi 2022), we find that wages increase with experience until a peak and then decline.

The next step is to add data to the article and report a table of summary statistics and a table of correlation matrices to `empirical_results.docx` using the commands `putdocx`, `sum2docx`, and `corr2docx`. Below is an illustration for the exercise.

```
. putdocx begin, font("Times New Roman", 16, black)
. putdocx paragraph, halign(left)
. putdocx text ("Data"), font("Times New Roman", 24, black) bold
. putdocx textblock begin
Data for the main analysis come from the Panel Study of Income Dynamics
 (Cameron and Trivedi 2022).  The sample consists of 4,165
 individual-year pair observations.
putdocx textblock end
. putdocx paragraph, halign(left)
. putdocx text ("Summary statistics"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
Table 1 displays summary statistics; the variables take on values that
 are within the expected ranges.  The average log-transformed hourly wage
 in cents in our sample is 6.68, which translates into around $7.96 per
 hour.  The average years of full-time work experience is 19.85.
putdocx textblock end
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
. sum2docx exp - lwage using "empirical_results.docx",
> title("Table 1. Summary statistics")
> stats(N mean(%9.2f) sd(%9.2f) min(%9.2f) median(%9.2f) max(%9.2f))
> append
Summary statistics table has been written to file empirical_results.docx.
```

```
. putdocx begin, font("Times New Roman", 16, black)
. putdocx paragraph, halign(left)
. putdocx text ("Correlation matrix"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
Table 2 shows a correlation matrix.  The upper triangle is a Spearman
 correlation matrix, and the lower triangle reports Pearson correlation
 coefficients.  As we can see, there is a positive correlation between
 wages and work experience.
putdocx textblock end
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
. corr2docx exp - lwage using "empirical_results.docx",
> append(pagebreak stylesrc(own)) landscape
> fmt(%9.4f) star title("Table 2. Correlation coefficient") spearman(pw)
> pearson(pw) font("Times New Roman", 9, black)
Correlation matrix has been written to file empirical_results.docx.
```

# Data

Data for the main analysis come from the Panel Study of Income Dynamics (Cameron and Trivedi 2022). The sample consists of 4,165 individual–year pair observations.

## Summary statistics

Table 1 displays summary statistics; the variables take on values that are within the expected ranges. The average log-transformed hourly wage in cents in our sample is 6.68, which translates into around $7.96 per hour. The average years of full-time work experience is 19.85.

Table 1. Summary statistics

| VarName | Obs | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|---|
| exp | 4165 | 19.85 | 10.97 | 1.00 | 18.00 | 51.00 |
| wks | 4165 | 46.81 | 5.13 | 5.00 | 48.00 | 52.00 |
| occ | 4165 | 0.51 | 0.50 | 0.00 | 1.00 | 1.00 |
| ind | 4165 | 0.40 | 0.49 | 0.00 | 0.00 | 1.00 |
| south | 4165 | 0.29 | 0.45 | 0.00 | 0.00 | 1.00 |
| smsa | 4165 | 0.65 | 0.48 | 0.00 | 1.00 | 1.00 |
| ms | 4165 | 0.81 | 0.39 | 0.00 | 1.00 | 1.00 |
| fem | 4165 | 0.11 | 0.32 | 0.00 | 0.00 | 1.00 |
| union | 4165 | 0.36 | 0.48 | 0.00 | 0.00 | 1.00 |
| ed | 4165 | 12.85 | 2.79 | 4.00 | 12.00 | 17.00 |
| blk | 4165 | 0.07 | 0.26 | 0.00 | 0.00 | 1.00 |
| lwage | 4165 | 6.68 | 0.46 | 4.61 | 6.68 | 8.54 |

## Correlation matrix

Table 2 shows a correlation matrix. The upper triangle is a Spearman correlation matrix, and the lower triangle reports Pearson correlation coefficients. As we can see, there is a positive correlation between wages and work experience.

Table 2. Correlation coefficient

| | exp | wks | occ | ind | south | smsa | ms | fem | union | ed | blk | lwage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exp | 1 | -0.1432*** | 0.0776*** | 0.1631*** | -0.0532*** | 0.0480*** | 0.1637*** | -0.0913*** | 0.0635*** | -0.2135*** | 0.0383** | 0.2262*** |
| wks | -0.0337** | 1 | 0.0334** | -0.0209 | 0.0823*** | -0.0388** | 0.0285* | -0.0580*** | -0.1978*** | -0.0300* | -0.0113 | -0.1002*** |
| occ | 0.0822*** | -0.0038 | 1 | 0.2260*** | 0.0413*** | -0.2018*** | 0.0706*** | -0.0847*** | 0.3784*** | -0.6309*** | 0.0837*** | -0.3043*** |
| ind | 0.1591*** | 0.0404*** | 0.2260*** | 1 | -0.0769*** | -0.0689*** | 0.1701*** | -0.1778*** | 0.1465*** | -0.2381*** | -0.0475*** | 0.0230 |
| south | -0.0527*** | 0.0292* | 0.0413*** | -0.0769*** | 1 | -0.1350*** | -0.0403*** | 0.0516*** | -0.1628*** | -0.1191*** | 0.1218*** | -0.1810*** |
| smsa | 0.0526*** | 0.0180 | -0.2018*** | -0.0689*** | -0.1350*** | 1 | -0.1060*** | 0.1044*** | 0.0271* | 0.1670*** | 0.1154*** | 0.2216*** |
| ms | 0.1606*** | 0.0625*** | 0.0706*** | 0.1701*** | -0.0403*** | -0.1060*** | 1 | -0.7228*** | 0.1147*** | -0.0044 | -0.2150*** | 0.2660*** |
| fem | -0.0922*** | -0.0833*** | -0.0847*** | -0.1778*** | 0.0516*** | 0.1044*** | -0.7228*** | 1 | -0.1132*** | -0.0141 | 0.2086*** | -0.3103*** |
| union | 0.0591*** | -0.1548*** | 0.3784*** | 0.1465*** | -0.1628*** | 0.0271* | 0.1147*** | -0.1132*** | 1 | -0.2813*** | 0.0471*** | 0.0094 |
| ed | -0.2182*** | -0.0067 | -0.6194*** | -0.2365*** | -0.1216*** | 0.1843*** | -0.0083 | -0.0012 | -0.2695*** | 1 | -0.1276*** | 0.3743*** |
| blk | 0.0404*** | -0.0319** | 0.0837*** | -0.0475*** | 0.1218*** | 0.1154*** | -0.2150*** | 0.2086*** | 0.0471*** | -0.1196*** | 1 | -0.1783*** |
| lwage | 0.2093*** | 0.0585*** | -0.3176*** | 0.0458*** | -0.1804*** | 0.2240*** | 0.2875*** | -0.3250*** | 0.0087 | 0.3939*** | -0.1895*** | 1 |

Then the sample is divided into two groups by log wage (`lwage`), and the differences of other variables are compared between the two groups. Below is an illustration of this exercise.

```
. egen medianwage = median(lwage)
. generate group = lwage >= medianwage
. putdocx begin, font("Times New Roman", 16, black)
. putdocx sectionbreak
. putdocx paragraph, halign(left)
. putdocx text ("t test"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
We then divide our sample into high-wage and low-wage groups using the
 median of the hourly wages as a benchmark.  The differences in the
 identical of other variables between the groups are compared.  Table 3
 reports the <<dd_docx_display italic: "t">>-test result.  It is shown
 that employees with higher wages have more years of work experience
 and education and are less likely to be females and blue-collar workers
 or live in the South.  The results also show that married people and
 those who live in the standard metropolitan statistical areas earn more.
putdocx textblock end
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
. t2docx exp - blk using "empirical_results.docx",  append by(group) fmt(%9.2f)
>  title("Table 3. t-test table")
t-test table has been written to file empirical_results.docx.
```

## t test

We then divide our sample into high-wage and low-wage groups using the median of the hourly wages as a benchmark. The differences in the identical of other variables between the groups are compared. Table 3 reports the *t*-test result. It is shown that employees with higher wages have more years of work experience and education and are less likely to be females and blue-collar workers or live in the South. The results also show that married people and those who live in the standard metropolitan statistical areas earn more.

Table 3. t-test table

| varname | obs(0) | mean(0) | obs(1) | mean(1) | mean-diff | t |
|---|---|---|---|---|---|---|
| exp | 2024 | 18.00 | 2141 | 21.61 | -3.61*** | -10.77 |
| wks | 2024 | 46.69 | 2141 | 46.92 | -0.23 | -1.43 |
| occ | 2024 | 0.63 | 2141 | 0.39 | 0.24*** | 16.00 |
| ind | 2024 | 0.40 | 2141 | 0.39 | 0.00 | 0.29 |
| south | 2024 | 0.37 | 2141 | 0.22 | 0.15*** | 10.55 |
| smsa | 2024 | 0.56 | 2141 | 0.74 | -0.18*** | -12.56 |
| ms | 2024 | 0.74 | 2141 | 0.89 | -0.15*** | -12.87 |
| fem | 2024 | 0.19 | 2141 | 0.04 | 0.15*** | 16.29 |
| union | 2024 | 0.36 | 2141 | 0.37 | -0.01 | -0.43 |
| ed | 2024 | 11.96 | 2141 | 13.68 | -1.71*** | -20.83 |
| blk | 2024 | 0.11 | 2141 | 0.04 | 0.07*** | 9.40 |

To understand the relationship between wages and years of experience, we run the following program to create figure 8.2 in the book. We then add the picture into `empirical_results.docx`.

```
. putdocx begin, font("Times New Roman", 16, black)
. putdocx paragraph, halign(left)
. putdocx text ("Empirical results"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
Figure 1 shows the quadratic relation between wages and work experience.
 Each point represents an individual-year pair.  The dashed smooth curve
 line is fit by an OLS of lwage on a quadratic in exp.  Wages increase
 until 30 or so years of experience and then decline.
putdocx textblock end
. putdocx save "empirical_results.docx", append(pagebreak stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
. graph twoway (scatter lwage exp, msize(small) msymbol(o))
> (qfit lwage exp, clstyle(p3) lwidth(medthick))
> (lowess lwage exp, bwidth(0.4) clstyle(p1) lwidth(medthick)),
> plotregion(style(none)) title("Overall variation: Log wage versus experience")
> xtitle("Years of experience", size(medlarge)) xscale(titlegap(*5))
> ytitle("Log wage", size(medlarge)) yscale(titlegap(*5)) legend(size(small))
> legend(pos(4) ring(0) col(1)) legend(label(1 "Actual data")
> label(2 "Quadratic fit") label(3 "Lowess")) scheme(sj)
```

```
. graph export "overall_variation.png", replace
(file overall_variation.png not found)
file overall_variation.png saved as PNG format
. putdocx begin
. putdocx paragraph, halign(center)
. putdocx image "overall_variation.png", width(4) linebreak
. putdocx text ("Figure 1"), font("Times New Roman", 12, black)
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
```

## Empirical results

Figure 1 shows the quadratic relation between wages and work experience. Each point represents an individual–year pair. The dashed smooth curve line is fit by an OLS of lwage on a quadratic in exp. Wages increase until 30 or so years of experience and then decline.

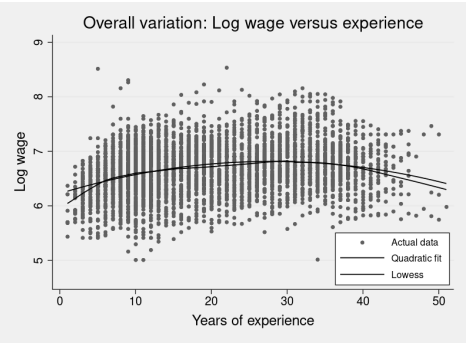


Figure 1

Then the quadratic relationship between wages and years of full-time work experience can be tested. After we run the regressions, all the results are reported to `empirical_results.docx` without the control variables' coefficients. Below are the commands for this exercise.

```
. putdocx begin, font("Times New Roman", 16, black)
. putdocx textblock begin
We test the quadratic relation using regression analysis.  We regress log
 wages(lwage) on years of full-time work experience(exp) and the
 quadratic term of exp(exp2) while controlling other variables.  The
 regression results were reported in table 4.  We report pooled OLS
 results in the first two columns and use cluster-robust standard errors
 in the second column.  In the last two columns, we use fixed-effects and
 cluster-robust standard errors.  In the first and third columns, we also
 control the time fixed effect.  The estimates imply an inverted U-shaped
 relationship between wages and work experience.
putdocx textblock end
```

```
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "C:/mypath/empirical_results.docx"
. xtset id t
  (output omitted)
. regress lwage exp exp2 wks occ ind south smsa ms union fem ed blk, vce(robust)
  (output omitted)
. estimates store m1
. regress lwage exp exp2 wks occ ind south smsa ms union fem ed blk i.t,
>  vce(cluster id)
  (output omitted)
. estimates store m2
. xtreg lwage exp exp2 wks occ ind south smsa ms union, fe vce(cluster id)
  (output omitted)
. estimates store m3
. xtreg lwage exp exp2 wks occ ind south smsa ms union i.t, fe vce(cluster id)
  (output omitted)
. estimates store m4
. reg2docx m1 m2 m3 m4 using "empirical_results.docx", append
> scalars(N r2_a(%9.2f)) drop(wks occ ind south smsa ms union fem ed blk)
> t(%9.2f) indicate("Time FE = *t") addfe("Individual FE = No No Yes Yes")
> title("Table 4. Regression results")
> note("* p<0.1, ** p<0.05, *** p<0.01", font(, 9))
Regression table has been written to file empirical_results.docx.
```

We test the quadratic relation using regression analysis. We regress log wages(lwage) on years of full-time work experience(exp) and the quadratic term of exp(exp2) while controlling other variables. The regression results were reported in table 4. We report pooled OLS results in the first two columns and use cluster–robust standard errors in the second column. In the last two columns, we use fixed-effects and cluster–robust standard errors. In the first and third columns, we also control the time fixed effect. The estimates imply an inverted U-shaped relationship between wages and work experience.

Table 4. Regression results

| | (1) lwage | (2) lwage | (3) lwage | (4) lwage |
|---|---|---|---|---|
| exp | 0.040*** | 0.031*** | 0.113*** | 0.111*** |
| | (18.56) | (7.74) | (27.95) | (26.95) |
| exp2 | -0.001*** | -0.001*** | -0.000*** | -0.000*** |
| | (-14.04) | (-6.10) | (-5.07) | (-4.78) |
| _cons | 5.251*** | 5.108*** | 4.649*** | 4.667*** |
| | (70.52) | (40.92) | (59.60) | (60.30) |
| Time FE | No | Yes | No | Yes |
| Individual FE | No | No | Yes | Yes |
| N | 4165 | 4165 | 4165 | 4165 |
| Adj. $R^2$ | 0.43 | 0.56 | 0.66 | 0.66 |

* p<0.1, ** p<0.05, *** p<0.01

So the commands above could be used to generate more tables in the article. Finally, the conclusion section is added into the document by the command below.

```
. putdocx begin, font("Times New Roman", 16, black)
. putdocx paragraph, halign(left)
. putdocx text ("Conclusion"), font("Times New Roman", 20, black) bold
. putdocx textblock begin
This article investigates the relationship between work experiences and wages.
 Using the data from chapter 8 of
 <<dd_docx_display italic: "Microeconometrics Using Stata, Second Edition">>,
 we find a quadratic relationship between wages and work experience.  Wages
 increase with experience until a peak and then decline.
putdocx textblock end
. putdocx save "empirical_results.docx", append(stylesrc(own))
successfully appended to "/home/sls/sj/software/sum2docx/empirical_results.docx"
```

### Conclusion

This article investigates the relationship between work experiences and wages. Using the data from chapter 8 of *Microeconometrics Using Stata, Second Edition*, we find a quadratic relationship between wages and work experience. Wages increase with experience until a peak and then decline.

## 4 Conclusion

Reporting empirical results can be a tedious and error-prone task. This article has introduced four commands: `sum2docx`, `corr2docx`, `t2docx`, and `reg2docx`. Using them is not only efficient but also possibly helpful in saving a fraction of time that is spent on tasks without them. Summary statistics, correlation coefficient matrices, mean-comparison $t$ tests between two groups, and regression results commonly used in empirical research can be stored into structured tables, and these tables can be coalesced into one `.docx` document. These commands are powerful and user friendly. We believe that these commands can simplify the work of empirical research and attract more researchers to become members of the Stata family.

## 5 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 23-2
. net install st0719     (to install program files, if available)
. net get st0719         (to install ancillary files, if available)
```

## 6 References

Cameron, A. C., and P. K. Trivedi. 2022. *Microeconometrics Using Stata*. 2nd ed. College Station, TX: Stata Press.

Cox, N. J. 2000. egenmore: Stata modules to extend the generate function. Statistical Software Components S386401, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s386401.html.

Gallup, J. L. 1998. sg97: Formatting regression output for published tables. *Stata Technical Bulletin* 46: 28–30. Reprinted in *Stata Technical Bulletin Reprints*. Vol. 8, pp. 200–202. College Station, TX: Stata Press.

———. 1999. sg97.1: Revision of outreg. *Stata Technical Bulletin* 49: 23. Reprinted in *Stata Technical Bulletin Reprints*. Vol. 9, pp. 170–171. College Station, TX: Stata Press.

———. 2000. sg97.2: Update to formatting regression output. *Stata Technical Bulletin* 58: 9–13. Reprinted in *Stata Technical Bulletin Reprints*. Vol. 10, pp. 137–143. College Station, TX: Stata Press.

———. 2001. sg97.3: Update to formatting regression output. *Stata Technical Bulletin* 59: 23. Reprinted in *Stata Technical Bulletin Reprints*. Vol. 10, p. 143. College Station, TX: Stata Press.

———. 2012. A new system for formatting estimation tables. *Stata Journal* 12: 3–28. https://doi.org/10.1177/1536867X1201200102.

Jann, B. 2005. Making regression tables from stored estimates. *Stata Journal* 5: 288–308. https://doi.org/10.1177/1536867X0500500302.

———. 2007. Making regression tables simplified. *Stata Journal* 7: 227–244. https://doi.org/10.1177/1536867X0700700207.

Jann, B., and J. S. Long. 2010. Tabulating SPost results using estout and esttab. *Stata Journal* 10: 46–60. https://doi.org/10.1177/1536867X1001000106.

Lokshin, M., and Z. Sajaia. 2008. Creating print-ready tables in Stata. *Stata Journal* 8: 374–389. https://doi.org/10.1177/1536867X0800800304.

Over, M. 2012. addest: Add scalar, vector, matrix or text information to estimation results. http://digital.cgdev.org/doc/stata/MO/Misc.

Wada, R. 2005. outreg2: Stata module to arrange regression outputs into an illustrative table. Statistical Software Components S456416, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456416.html.

**About the authors**

Yuan Xue is a PhD student in accounting at the Huazhong University of Science and Technology in Wuhan, China.

Chuntao Li is a professor of finance at the Henan University in Kaifeng, as well as a professor of finance at Zhongnan University of Economics and Law in Wuhan, China.

Haitao Si is a PhD student in finance at Wuhan University in Wuhan, China.