The Stata–to–LaTeXGuide*

Asjad Naqvi†

Last updated: 08 December, 2021

Abstract

This article introduces various wrappers for the Stata `estout` package that allows one to generate various summary statistic and regression tables in LaTeX. Rather than going into the details of the estout package options, which are fairly extensive, the aim of this document is to provide "templates" for easy use and replication. This is an evolving document so please check regularly for updates. Since LaTeXrequirements change over time, please see the change log for details, or message me to fix errors if they occur. And please feel free to request or suggest other table formats.

Keywords: Stata, LaTeX, estout, tables, regressions

JEL: 1337

*See the article on Medium for details.

†For comments, suggestions, bugs please email at: asjadnaqvi@gmail.com

Contents

List of Tables

1 Introduction

The document is divided into two sections. Section 2 covers summary statistic tables and Section 3 covers regressions.

Summary statistic tables are a bit more challenging than regression tables due to how the formatting is defined. This is an evolving document where these sections will be filled up over time. Please refer to the Medium article for the Stata part.

Change log:

- 08 Dec 2021: Quick fix to `estwide` and `estauto` to make them compatible with TexLive 2021. This has to do with adding lines (hline or midrules) in tables. The requirements have changed slightly. Please see lines 68 and 80 above for modifications to the file. The guide will get another update at the end of the year so various minor syntax issues will be corrected.
- 20 Jul 2021: Added `aux/sd` table with formatting option (Table 2.2). Added table on summary statistics by overlapping categories (Table 2.10). Section 2 table numbering corrected and homogenized with the Medium article. Minor corrections to the syntax and descriptions.
- 20 Apr 2021: Added Section 4 on writing and formatting equations in LaTeXas supplementary material.
- 04 Mar 2021: Three tables added: Table 2.8 on t-tests, Table 3.5 on stacking standard errors, and Table 3.6 on rotating large tables.
- 27 Feb 2021: Due to several requests, a basic regression table (Table 3.1) with a slow introduction has been added at the beginning of Section 3. The labeling of the other tables is moved down by one step. All figures have been fixed in the Medium document which were showing the wrong Table numbers and significance levels. Several other minor fixes.
- 26 Feb 2021: Minor typos and code fixes on Medium and here on Overleaf. Table numbers resets at each section to fix the correspondence to the Guide.
- 25 Feb 2021: Stata-to-LaTeX guide posted on Medium with the first set of summary statistic tables (Tables 2.1–2.7) and regression tables (Tables 3.2–3.4).

2 Summary statistics

A table can be generated using three different methods. The first two tables call the custom written `estauto` and `estwide` scripts defined above [Weber, 2012]. The last table shows how this can be done directly in LaTeX using the `estout` package written by Ben Jann [Jann, 2005, 2007].

Table 2.1: Summary statistics - `estauto` vs `estwide` vs direct compilation

	Sum	Mean	SD	Min	Max	N
Daily cases	185,166,728	2,081.41	11,601.53	0	414,188	88,962
New deaths	4,010,561	49.77	217.42	0	7,374	80,582
New tests	2,200,722,267	49,545.73	180,385.57	1	3,740,296	44,418
New vaccinations	3,119,147,132	243,360.16	1,323,551.77	1	24,741,000	12,817

	Sum	Mean	SD	Min	Max	N
Daily cases	185,166,728	2,081.41	11,601.53	0	414,188	88,962
New deaths	4,010,561	49.77	217.42	0	7,374	80,582
New tests	2,200,722,267	49,545.73	180,385.57	1	3,740,296	44,418
New vaccinations	3,119,147,132	243,360.16	1,323,551.77	1	24,741,000	12,817

	Sum	Mean	SD	Min	Max	N
Daily cases	185,166,728	2,081.41	11,601.53	0	414,188	88,962
New deaths	4,010,561	49.77	217.42	0	7,374	80,582
New tests	2,200,722,267	49,545.73	180,385.57	1	3,740,296	44,418
New vaccinations	3,119,147,132	243,360.16	1,323,551.77	1	24,741,000	12,817

Note: `estwide` spans the whole length of the page while `estauto` displays the table as it is.

Table 2.2 uses the simpler `mean` and `aux` options with some basic formatting defined, while Table 2.3 calls the more advanced `cells` option which allows for full customization of all the elements. The final output is exactly the same across the two.

Table 2.2: Summary statistics - using `mean` and `aux` options

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.45	2,209.30	2,567.50	1,111.85	189.26	8,338.41	32,939.00
	(2,011.80)	(5,862.08)	(8,812.25)	(2,230.43)	(1,040.40)	(36,281.03)	(54,155.46)
New deaths	15.64	48.86	96.89	20.81	4.97	135.84	638.66
	(52.20)	(128.08)	(302.38)	(48.32)	(31.18)	(498.63)	(894.62)
New tests	19,350.75	55,973.43	10,855.22	28,008.36	4,109.91	140,438.03	515,830.71
	(23,658.54)	(134,072.94)	(16,564.34)	(44,770.56)	(8,676.09)	(401,724.79)	(592,581.80)
New vaccinations	833,270.18	97,570.25	106,165.06	51,843.68	12,154.95	657,632.75	937,416.82
	(3,274,142.89)	(182,922.50)	(228,911.01)	(65,945.36)	(22,522.04)	(1,401,720.57)	(1,028,118.60)

Table 2.3: Summary statistics - using `cells` option

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.45	2,209.30	2,567.50	1,111.85	189.26	8,338.41	32,939.00
	(2,011.80)	(5,862.08)	(8,812.25)	(2,230.43)	(1,040.40)	(36,281.03)	(54,155.46)
New deaths	15.64	48.86	96.89	20.81	4.97	135.84	638.66
	(52.20)	(128.08)	(302.38)	(48.32)	(31.18)	(498.63)	(894.62)
New tests	19,350.75	55,973.43	10,855.22	28,008.36	4,109.91	140,438.03	515,830.71
	(23,658.54)	(134,072.94)	(16,564.34)	(44,770.56)	(8,676.09)	(401,724.79)	(592,581.80)
New vaccinations	833,270.18	97,570.25	106,165.06	51,843.68	12,154.95	657,632.75	937,416.82
	(3,274,142.89)	(182,922.50)	(228,911.01)	(65,945.36)	(22,522.04)	(1,401,720.57)	(1,028,118.60)

Table 2.4 uses a different formatting option for correcting the decimal values in cells.

Table 2.4: Summary statistics - `cells` formatting

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.45	2209.30	2567.50	1111.85	189.26	8338.41	32939.00
	(2011.80)	(5862.08)	(8812.25)	(2230.43)	(1040.40)	(36281.03)	(54155.46)
New deaths	15.64	48.86	96.89	20.81	4.97	135.84	638.66
	(52.20)	(128.08)	(302.38)	(48.32)	(31.18)	(498.63)	(894.62)
New tests	19350.75	55973.43	10855.22	28008.36	4109.91	140438.03	515830.71
	(23658.54)	(134072.94)	(16564.34)	(44770.56)	(8676.09)	(401724.79)	(592581.80)
New vaccinations	833270.18	97570.25	106165.06	51843.68	12154.95	657632.75	937416.82
	(3274142.89)	(182922.50)	(228911.01)	(65945.36)	(22522.04)	(1401720.57)	(1028118.60)

Table 2.5 shows custom square brackets that can be specified using the `par` option in the `cells` command.

Table 2.5: Summary statistics - Square brackets

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.4468	2209.299	2567.504	1111.851	189.2609	8338.406	32939
	[2011.801]	[5862.081]	[8812.246]	[2230.43]	[1040.4]	[36281.03]	[54155.46]
New deaths	15.63801	48.85879	96.89424	20.81333	4.96738	135.8436	638.6556
	[52.20031]	[128.0849]	[302.3825]	[48.32278]	[31.17567]	[498.6288]	[894.6165]
New tests	19350.75	55973.43	10855.22	28008.36	4109.906	140438	515830.7
	[23658.54]	[134072.9]	[16564.34]	[44770.56]	[8676.091]	[401724.8]	[592581.8]
New vaccinations	833270.2	97570.25	106165.1	51843.68	12154.95	657632.7	937416.8
	[3274143]	[182922.5]	[228911]	[65945.36]	[22522.04]	[1401721]	[1028119]

Tables 2.6 and 2.7 show how the decimal points can be customized. The first table displays the mean values with two decimal points and the standard deviations (SD) with three decimal points. The means in the second table have decimal points of 1, 2, 3, and 4 while the SDs have decimal points of 3, 2, 1, and 0 as one goes down. While it is highly desirable to keep decimal points the same, this option can be used to optimize the display, for example, if some variables are in decimals while others are in integers.

Table 2.6: Summary statistics - Mean 2 d.p. and SD 3 d.p.

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.45	2209.30	2567.50	1111.85	189.26	8338.41	32939.00
	(2011.801)	(5862.081)	(8812.246)	(2230.430)	(1040.400)	(36281.034)	(54155.460)
New deaths	15.64	48.86	96.89	20.81	4.97	135.84	638.66
	(52.200)	(128.085)	(302.383)	(48.323)	(31.176)	(498.629)	(894.616)
New tests	19350.75	55973.43	10855.22	28008.36	4109.91	140438.03	515830.71
	(23658.536)	(134072.939)	(16564.342)	(44770.562)	(8676.091)	(401724.787)	(592581.800)
New vaccinations	833270.18	97570.25	106165.06	51843.68	12154.95	657632.75	937416.82
	(3274142.888)	(182922.499)	(228911.012)	(65945.360)	(22522.036)	(1401720.572)	(1028118.597)

Table 2.7: Summary statistics - All decimals customized

	EAP	ECA	LAC	MENA	SSA	S. Asia	N. America
Daily cases	613.4	2209.3	2567.5	1111.9	189.3	8338.4	32939.0
	(2011.801)	(5862.081)	(8812.246)	(2230.430)	(1040.400)	(36281.034)	(54155.460)
New deaths	15.64	48.86	96.89	20.81	4.97	135.84	638.66
	(52.20)	(128.08)	(302.38)	(48.32)	(31.18)	(498.63)	(894.62)
New tests	19350.750	55973.426	10855.218	28008.358	4109.906	140438.028	515830.708
	(23658.5)	(134072.9)	(16564.3)	(44770.6)	(8676.1)	(401724.8)	(592581.8)
New vaccinations	833270.1755	97570.2496	106165.0622	51843.6781	12154.9487	657632.7479	937416.8170
	(3274143)	(182922)	(228911)	(65945)	(22522)	(1401721)	(1028119)

Table 2.8 generates a neat summary statistics of variables by grouping them in categories. The spacing in the variables is generated within Stata by modifying the variable label.

Table 2.8: Summary statistics - Grouping variables

	Mean	SD	Min	Max	N
COVID-19 indicators					
Total cases (units)	1,009,549	3,758,295	2	33,838,746	182
Total deaths (units)	22,470	73,845	1	606,993	177
Total tests (units)	31,817,347	86,292,430	53,441	474,252,139	60
Total vaccinations (units)	28,226,360	141,065,982	22,837	1,365,463,000	99
Socio-economics indicators					
Median age (years)	30.40	9.21	15.10	48.20	178
Age 65+ (years)	8.68	6.21	1.14	27.05	176
Life expectancy (years)	73.13	7.76	53.28	86.75	193
Extreme poverty (%)	14.10	20.60	0.10	77.60	119
HDI (Index)	72.21	15.34	39.40	95.70	177
GDP per capita (USD)	19582.68	20843.09	661.24	116935.60	180
Population density (per sq km)	468.60	2172.45	0.14	20546.77	191

Note: Yes, there are countries with just one case and one death reported to date!

Table 2.9 gives a T-test table by two groups Middle East and North Africa (MENA) and Sub-Saharan Africa (SSA). The first two columns give the averages while the next two provide the mean and standard errors together with significance levels for the two groups. Here MENA clearly outperforms SSA in all the indicators by a significant margin. Such a table can be used to show, for example, the covariates between the treatment and control groups in the (pre-treatment) baseline which should ideally be statistically similar.

Table 2.9: T-tests across two groups by several variables

	MENA	SSA	Diff. (MENA - SSA)	S.E.	Obs.
Age 65+ (years)	5.440	3.383	2.057***	(0.045)	32272
GDP per capita (USD)	30289.436	4790.638	25498.798***	(290.617)	31787
Life expectancy (years)	75.657	62.719	12.938***	(0.053)	32751
Extreme poverty (%)	4.289	37.920	-33.630***	(0.194)	22908
Population density (per sq km)	298.756	108.561	190.195***	(5.288)	31807

Note: *** p<0.01, ** p<0.05, * p<0.10.

Table 2.10 gives summary statistics by different groups that have overlap with each other.

Table 2.10: Summary statistics by different control groups

	All	ECA	Europe	High HDI
Daily cases	2,081.41	2,209.30	2,218.51	2,822.63
	(11,601.53)	(5,862.08)	(5,773.87)	(12,044.77)
New deaths	49.77	48.86	52.19	57.15
	(217.42)	(128.08)	(134.80)	(217.22)
New tests	49,545.73	55,973.43	53,751.62	67,161.36
	(180,385.57)	(134,072.94)	(135,954.51)	(190,200.04)
New vaccinations	243,360.16	97,570.25	89,748.95	130,917.78
	(1,323,551.77)	(182,922.50)	(173,087.90)	(330,622.81)

Note: *** p<0.01, ** p<0.05, * p<0.10.

3 Regressions

Table 3.1: My very first basic regression table

	(1)	(2)
Cases (norm.)	0.011^{***}	0.011^{***}
	(0.001)	(0.001)
Tests (norm.)	0.080^{***}	0.105^{***}
	(0.030)	(0.029)
Stringency Index	0.024	0.024
	(0.016)	(0.016)
Age 65+ (years)		0.124^{***}
		(0.046)
GDP per capita		-2.531^{***}
		(0.550)
Life expectancy (years)		-0.005
		(0.040)
Extreme poverty (%)		-0.001
		(0.004)
Pop. density (pop/km^2)		-0.001^{**}
		(0.000)
Constant	-1.046	-1.208
	(0.815)	(2.066)
Observations	29181	22609

Standard errors in parentheses

Dependent variable: Deaths (norm.). A lot of endogeneity in this specification.

Data source: Our World in Data COVID-19 database.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table generated using the `estauto` option

	(1)	(2)
Cases (norm.)	0.011***	0.011***
	(0.001)	(0.001)
Tests (norm.)	0.080***	0.105***
	(0.030)	(0.029)
Stringency Index	0.024	0.024
	(0.016)	(0.016)
Age 65+ (years)		0.124***
		(0.046)
GDP per capita		-2.531***
		(0.550)
Life expectancy (years)		-0.005
		(0.040)
Extreme poverty (%)		-0.001
		(0.004)
Pop. density (pop/km^2)		-0.001**
		(0.000)
Constant	-1.046	-1.208
	(0.815)	(2.066)
Observations	29181	22609

Note: *** p<0.01, ** p<0.05, * p<0.10. Dependant variable: Deaths (norm.)

Table 3.2: Regressions by different specifications, lags, and `estadd`

	(1)	(2)	(3)	(4)	(5)	(6)
Cases (norm.) (t-10)	0.011***	0.011***	0.011***	0.009***	0.011***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.001)
Tests (norm.) (t-10)	0.074	-0.010	0.077	0.036	0.074***	-0.010
	(0.075)	(0.075)	(0.075)	(0.085)	(0.025)	(0.046)
Stringency Index (t-10)	0.020***	0.020***	0.020***	0.034***	0.020	0.020**
	(0.005)	(0.005)	(0.006)	(0.007)	(0.012)	(0.008)
Obs.	29231	29231	29231	29231	29231	29231
ρ	0.101	0.000	0.132	0.153	0.101	0.000
Time FE	No	Yes	No	Yes	No	Yes
Country FE	No	No	Yes	Yes		
Region FE					Yes	Yes

Note: *** p<0.01, ** p<0.05, * p<0.10. Dependant variable: Deaths (norm.)

Table 3.3: Regressions by groups

	Deaths		Cases		Tests		Vaccines	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stringency Index	0.022***	0.022	0.703***	0.703*	0.009*	0.009	-0.014	-0.014
	(0.007)	(0.017)	(0.187)	(0.392)	(0.005)	(0.007)	(0.201)	(0.219)
Age 65+ (years)	0.172***	0.172***	5.467***	5.467***	0.034	0.034	-0.164	-0.164
	(0.028)	(0.051)	(1.326)	(1.678)	(0.034)	(0.024)	(0.257)	(0.282)
GDP per capita	-0.490	-0.490	90.602	90.602***	7.803***	7.803***	4.934	4.934
	(1.142)	(0.595)	(64.016)	(23.052)	(1.823)	(0.865)	(6.281)	(7.436)
Life expectancy (years)	0.014	0.014	0.471	0.471	-0.069**	-0.069	0.114	0.114
	(0.024)	(0.049)	(1.059)	(1.792)	(0.027)	(0.046)	(0.612)	(0.629)
Extreme poverty (%)	0.002	0.002	0.057	0.057	-0.005	-0.005	-0.127	-0.127
	(0.004)	(0.004)	(0.172)	(0.119)	(0.006)	(0.007)	(0.176)	(0.118)
Pop. density (pop/km^2)	-0.001***	-0.001	-0.017	-0.017	0.001**	0.001	0.003	0.003
	(0.000)	(0.001)	(0.016)	(0.021)	(0.000)	(0.001)	(0.002)	(0.002)
Constant	-2.382	-2.382	-80.215	-80.215	3.809**	3.809	-4.429	-4.429
	(1.459)	(2.391)	(68.412)	(98.313)	(1.863)	(2.904)	(30.761)	(30.600)
Region FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: *** p<0.01, ** p<0.05, * p<0.10

Typically in difference-in-difference (D-in-D) designs, the baseline setting is of the following nature:

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 Z_{it} + \beta_3 (T \times Z)_{it} + \theta_{it} + \epsilon_{it} \tag{1}$$

where Y_{it} is the dependent variable for some location i at time t, T is a time index, Z is the treatment, and θ are control variables. The coefficient of interest is β_3 which implies that if the treatment is successful, then β_3 should be significant. The logic can be extended to multiple treatments and more complex D-in-D designs as well. One comes across papers that just show treatment coefficients for a battery of different specifications, robustness checks, and falsification tests. Without going too much into the details, Table 3.3 produces a 4×4 table of β_3 coefficients. Here the coefficients are collected for the top part, the middle part, and the bottom part separately, and then patched together in Stata. This table can be combined with a host of other elements including grouping regressions, as in Table 3.2, or displaying multiple coefficients per row.

Table 3.4: Table of coefficient of interest from multiple regressions

	(1) Deaths	(2) Cases	(3) Tests	(4) Vaccines
Baseline	0.003**	0.088	0.005*	0.033
	(0.002)	(0.078)	(0.003)	(0.035)
Region FE	0.003***	0.088	0.005*	0.033*
	(0.001)	(0.095)	(0.003)	(0.017)
Controls	0.003	0.159*	0.004	0.028
	(0.002)	(0.095)	(0.003)	(0.043)
Controls + Region FE	0.003**	0.159***	0.004**	0.028
	(0.001)	(0.053)	(0.002)	(0.034)
Observations	37452	40876	24227	1583

Note: *** p<0.01, ** p<0.05, * p<0.10.

Table 3.5 displays different regressions that return the same coefficient but different standard errors, and stacks them up. This idea was inspired by Melissa Dell's *Mining Mita* paper [Dell, 2010] where she shows different standard errors stacked on top of each other together with the significance levels. This table can be extended to include additional variables as well using the routines defined earlier. Reporting different standard errors for exactly the same coefficients is not unusual. Different spatial regressions or difference-in-difference estimators actually do this, and stacking SEs helps minimize precious table space.

Table 3.5: Stacking standard errors from different specifications

	(1) Deaths	(2) Cases	(3) Tests	(4) Vaccines
Policy × Post	0.002	0.149	0.004	0.029
	(0.000)***	(0.013)***	(0.000)***	(0.013)**
	[0.001]**	[0.054]***	[0.002]**	[0.034]
Obs.	37682	41112	24227	1615

Note: *** p<0.01, ** p<0.05, * p<0.10. Random effect S.E.s in curly brackets and clustered S.E. in square brackets.

Table 3.6: Throwing everything in the table and rotating it to fit the page

	Deaths		Cases		Tests		Vaccines	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stringency Index	0.034***	0.034***	2.406**	2.111**	0.026***	0.025**	-0.137	-0.112
	(0.011)	(0.012)	(1.017)	(0.887)	(0.009)	(0.010)	(0.121)	(0.111)
L4.Stringency Index	0.006	-0.000	0.081	0.009	-0.002	-0.001	-0.084	-0.141
	(0.005)	(0.004)	(0.082)	(0.167)	(0.004)	(0.003)	(0.082)	(0.127)
L8.Stringency Index	0.004***	0.006**	-0.136	-0.068	0.001	0.000	0.035	0.059
	(0.001)	(0.003)	(0.084)	(0.165)	(0.003)	(0.002)	(0.044)	(0.055)
L12.Stringency Index	0.010*	0.005***	0.106	0.028	0.003	0.005**	0.113***	0.087**
	(0.006)	(0.001)	(0.083)	(0.117)	(0.002)	(0.002)	(0.041)	(0.034)
L16.Stringency Index	0.004	0.008**	-0.152	-0.129	0.005	0.002	0.288*	0.356*
	(0.003)	(0.004)	(0.294)	(0.285)	(0.007)	(0.004)	(0.171)	(0.193)
Age 65+ (years)		0.202***		6.660***		0.060**		0.024
		(0.048)		(1.772)		(0.026)		(0.188)
GDP per capita		-0.278		112.903***		8.233***		4.055
		(0.530)		(19.723)		(1.043)		(6.260)
Life expectancy (years)		-0.001		-0.529		-0.086*		-0.120
		(0.042)		(1.821)		(0.048)		(0.435)
Extreme poverty (%)		0.009*		0.201*		-0.001		-0.255***
		(0.005)		(0.114)		(0.005)		(0.078)
Pop. density (pop/km^2)		-0.001		-0.014		0.001		0.005
		(0.001)		(0.023)		(0.001)		(0.004)
Constant	-0.863***	-2.094	-4.649	-26.041	0.578	4.572	-12.715	-7.872
	(0.162)	(2.287)	(15.788)	(100.542)	(0.412)	(3.001)	(9.217)	(23.173)
Obs.	52174	35214	56603	38113	30588	23812	1918	1463
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Note: *** p<0.01, ** p<0.05, * p<0.10. All regressions control for month fixed effects to capture seasonal variations.

4 Symbols and Equations

This section will cover some basics of equations and math symbols commonly used in economics. This is to supplement your Stata tables and general paper writing in LaTeX.

First differentiate between inline maths $Y_t = \beta_0 + \beta_1 X_t + \epsilon$ verses maths equations:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon \tag{2}$$

Equations can be labeled and referenced. For example, see the TeX for Equation 2, which shows a simple regression model.

Equation numbers can be turned off as well:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon$$

One can write multiple equations and align them as well. Eqnarray is one way of doing this:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon \tag{3}$$
$$Z_t = \alpha_0 + \alpha_1 Y_t + \epsilon \tag{4}$$

One can also just label the whole equation block once as follows:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \epsilon \\ Z_t &= \alpha_0 + \alpha_1 Y_t + \epsilon \end{aligned} \tag{5}$$

This can also be used to split long equations across lines:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \dots \\ &\quad \cdots + \alpha_0 + \alpha_1 Y_t + \epsilon \end{aligned} \tag{6}$$

Superscripts and subscripts need to be carefully captured using curly brackets:

$$\begin{aligned} \texttt{Wrong} &= \beta^i_0 t + \beta^i j_1 t X^i_t + \epsilon \\ \texttt{Correct} &= \beta^i_{0t} + \beta^{ij}_{1t} X^i_t + \epsilon \end{aligned} \tag{7}$$

If there is only one item in the superscript or the subscript, then it doesn't really matter, but preferably they should have curly brackets around them.

Brackets are also often wrongly specified:

$$
\begin{aligned}
\texttt{Wrong} &= (n^2) + \{\frac{1}{n}\sum_{j=1}^{J}(1+n^j)\} + [\frac{1}{n}] \qquad (8)\\
\texttt{Correct} &= \left(n^2\right) + \left\{\frac{1}{n}\sum_{j=1}^{J}\left(1+n^j\right)\right\} + \left[\frac{1}{n}\right]
\end{aligned}
$$

Using left and right to open and close brackets tells LaTeXto use the maths symbols for brackets rather than the bracket characters. Also note the use of curly brackets. They need a qualifier since standard curly brackets are used for writing syntax in LaTeX.

The remaining maths is just layering on these principles including finding the right symbols. But be careful when combining elements:

$$
\begin{aligned}
\texttt{Wrong} &= \hat{Y_{it}} = Y_{it} - \bar{Y_{it}} \qquad (9)\\
\texttt{Correct} &= \hat{Y}_{it} = Y_{it} - \bar{Y}_{it}
\end{aligned}
$$

Note above how the hats and bars are shifted to the right and are not really over the symbols.

Some other math symbol examples:

$$\lim_{x\rightarrow+\infty} f(x) = f'(x) = \frac{\partial f}{\partial x} + \frac{\partial^2 f}{\partial x^2} \qquad (10)$$

or more complex stuff like:

$$\sum_{j=1}^{J} x_j \leq \prod_{j=1}^{J} x_j \neq \int_a^b f(x)dx \qquad (11)$$

which we can take to the next level:

$$\sum_{k=1}^{K}\sum_{j=1}^{J} x_j y_k \leq \prod_{k=1}^{K}\prod_{j=1}^{J} x_j y_k \neq \int_a^b\int_c^d f(x,y)dxdy \qquad (12)$$

Another feature of math mode is spacing, which you deal with when you have nothing better to do and/or are obsessed with formatting:

$$\begin{aligned}
\mathtt{Negative} &= x\!y\!z \qquad (13)\\
\mathtt{Normal} &= xyz\\
\mathtt{Thin} &= x\,y\,z\\
\mathtt{Medium} &= x\:y\:z\\
\mathtt{Thick} &= x\;y\;z
\end{aligned}$$

Another feature that is helpful are those large curly brackets that define conditions, and are hence called conditional equations:

$$y = \begin{cases} a + bx + cx^2, & \text{if } x \geq 1 \\ \frac{1}{x^2}, & \text{if } 0 < x \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Notice how the text field is used to control the spacing of the text. Within the text field we can write the equations in math form by using the dollar sign.

Similar to the conditional bracket are underscores that can be used to highlight parts of an equation:

$$y_{it} = \beta_0 + \beta_1 P_i + \beta_2 T_t + \underbrace{\beta_3 \left(P \times T\right)_{it}}_{\text{Coef. of interest}} + \epsilon_{it} \tag{15}$$

Basically underscores and overscores can be used on multiple segments of equations as long as the brackets are balanced:

$$y_{it} = \overbrace{\underbrace{\beta_0 + \beta_1 P_i + \beta_2 T_t}_{\text{Parallel trends}} + \underbrace{\beta_3 \left(P \times T_{it}\right)}_{\text{Coef. of interest}} + \epsilon_{it}}^{2\times 2 \text{ Diff-in-Diff model}} \tag{16}$$

Matrices are also simply to construct:

$$\begin{matrix} a & b \\ c & d \end{matrix} \tag{17}$$

or with brackets and equations:

$$
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{18}
$$

Let's make the whole setup more complex:

$$
\underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\boldsymbol{Ax}} + \underbrace{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}_{\boldsymbol{b}} \tag{19}
$$

from which we get

$$
\begin{aligned}
\mathbf{x} &= \mathbf{Ax} + \mathbf{b} \qquad (20) \\
\mathbf{x} - \mathbf{Ax} &= \mathbf{b} \\
(\mathbf{I} - \mathbf{A})\mathbf{x} &= \mathbf{b} \\
\mathbf{x} &= (\mathbf{I} - \mathbf{A})^{\mathbf{-1}}\mathbf{b}
\end{aligned}
$$

Quite a pain to do this copy pasting. It is OK with a few equations but if you are regularly dealing with matrix algebra, better to write some script to make this easier. But this advanced stuff I won't touch here. In the equation above bm or mathbf basically stands for 'bold maths' or 'bold font'. Note that this is different from converting a symbol in a letter and defining it as bold text. So be careful when differentiating between **A,B,**α, Γ, Θ versus $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \boldsymbol{\Theta}$ versus $\mathbf{A}, \mathbf{B}, \alpha, \mathbf{\Gamma}, \mathbf{\Theta}$.

Another under-utilized tool in LaTeXare phantom spaces, that are mostly used to align stuff. These are ghost spaces that are defined by whatever is given in the phantom option:

$$
\begin{aligned}
\texttt{Basic} &= \Pi_t^{ij} \qquad (21) \\
\texttt{Without phantom} &= \Pi_t^{j} \\
\texttt{With phantom} &= \Pi_t^{j}
\end{aligned}
$$

Notice the empty i space in the last equation. Where is this useful? For example, in aligning matrices:

$$
\begin{aligned}
\texttt{Without phantom} &= \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix} \qquad (22) \\
\texttt{With phantom} &= \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix}
\end{aligned}
$$

Here we basically add a 'fake' empty negative sign to align the matrix entries.

Another place where I have personally used phantom spaces, is to explain more complex equations with fractions:

$$\texttt{Without vphantom} = \underbrace{\left(a + bx + cx^2\right)}_{\text{First part}} + \underbrace{\left(\frac{1}{x} + \frac{1}{\sqrt{x}}\right)}_{\text{Second part}} \tag{23}$$

Above we can see above that the underbraces are not aligned. This can be corrected by adding a 'vertical' phantom space or vphantom as follows:

$$\texttt{With vphantom} = \underbrace{\vphantom{\left(\frac{1}{x}\right)}\left(a + bx + cx^2\right)}_{\text{First part}} + \underbrace{\left(\frac{1}{x} + \frac{1}{\sqrt{x}}\right)}_{\text{Second part}} \tag{24}$$

Note above that we phantom space not only some random fraction but also the brackets that go with it since they also add to the vertical space. We do this outside of the main equation of the 'First part' to prevent the 'First part' brackets to be larger than they should be.

If you have any equation questions or some suggestions then please let me know and I will add them here.

References

Melissa Dell. The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6):1863–1903, 2010. ISSN 0012-9682. doi: 10.3982/ECTA8121. URL http://doi.wiley.com/10.3982/ECTA8121.

Ben Jann. Making regression tables from stored estimates. *Stata Journal*, 5(3):288–308(21), 2005. URL https://www.stata-journal.com/article.html?article=st0085.

Ben Jann. Making regression tables simplified. *Stata Journal*, 7(2):227–244(18), 2007. URL https://www.stata-journal.com/article.html?article=st0085{_}1.

Jörg Weber. Automated Table generation in Stata and integration into LaTeX, 2012. URL https://www.jwe.cc/2012/03/stata-latex-tables-estout/.