SESSION-6: STATA DATA ANALYSIS AND VISUALIZATION

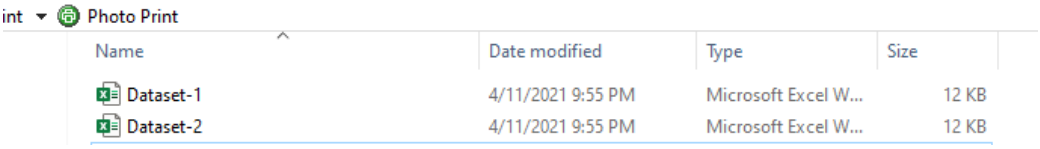
Course detail: http://julhas.com/jsedutech/stata-level-one.html
Mentor: Julhas Sujan

Session Outline:

- Append, Merge, Substring, Concat commands
- Example: how to make age groups from age
- Basic and advance statistical graphs – bar, column, histogram, line, box

Lession-1: Append, Merge, Substring and Concat commands

- Append: Appending two datasets require that both have variables with exactly the same name. **Step-1:** Copy our example dataset and keep the variables name as it is. Then add some data in the new dataset and save as Dataset-2.



We want to append Dataset-2 with Dataset-1. So, the columns/ variables name should be exactly same as Dataset-1.

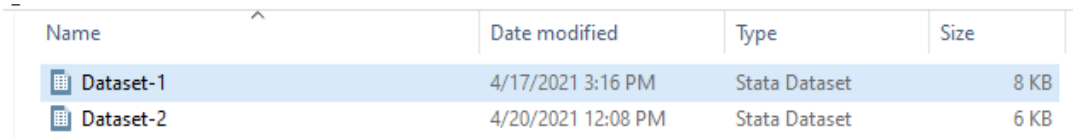
Excel Dataset-1: (Data from 1-30)

	A	B	C	D	E	F	G	H
1	S/N	Gender	Age	Case definition	Inicial We	Current V	Site of Disease	Outcome
2	1	Female	32	Clinically diagnosed case	58.0	58.0	Extrapulmonary	Successful
3	2	Female	19	Clinically diagnosed case	40.0	40.0	Pulmonary	Unsuccessful
4	3	Male	58	Clinically diagnosed case	64.0	64.0	Extrapulmonary	Successful
5	4	Female	28	Clinically diagnosed case	25.0	25.0	Extrapulmonary	Unsuccessful

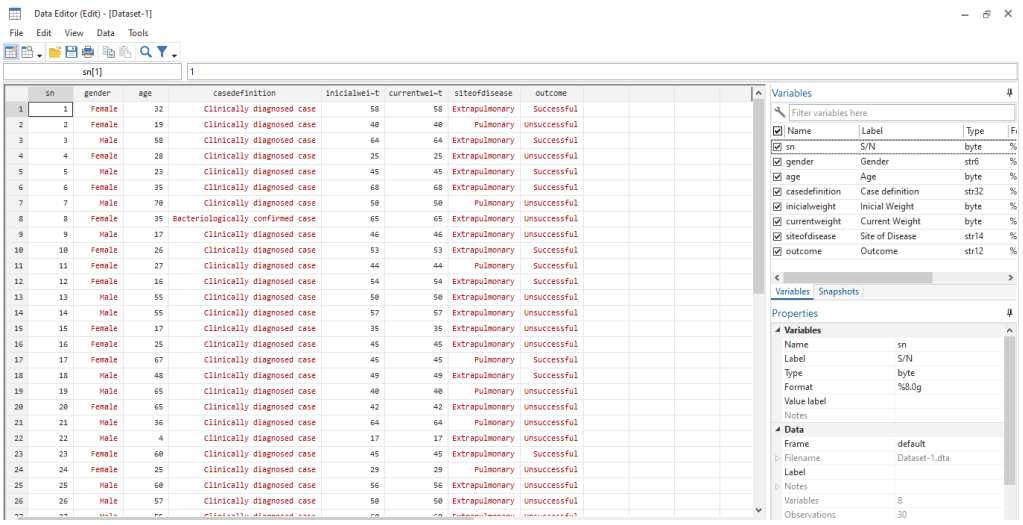
Excel Dataset-2: (New data from 31-35)

	A	B	C	D	E	F	G	H
1	S/N	Gender	Age	Case definition	Inicial We	Current V	Site of Disease	Outcome
2	31	Female	25	Clinically diagnosed case	29.0	29.0	Pulmonary	Unsuccessful
3	32	Male	60	Clinically diagnosed case	56.0	56.0	Extrapulmonary	Unsuccessful
4	33	Male	57	Clinically diagnosed case	50.0	50.0	Extrapulmonary	Unsuccessful
5	34	Male	55	Clinically diagnosed case	60.0	60.0	Extrapulmonary	Unsuccessful
6	35	Female	30	Bacteriologically confirme	44.0	44.0	Extrapulmonary	Successful
7								

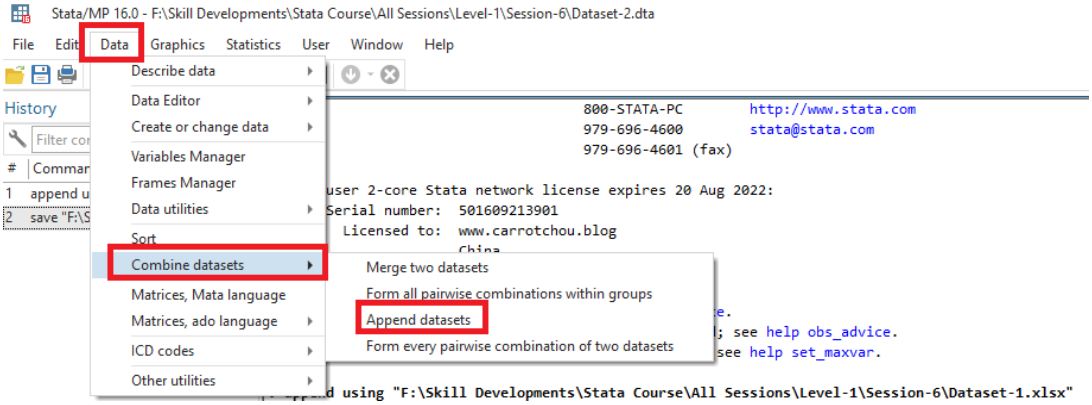
Step-2: Now you need to convert the two datasets in Stata .dta format (you can find the data import process in our session-4).



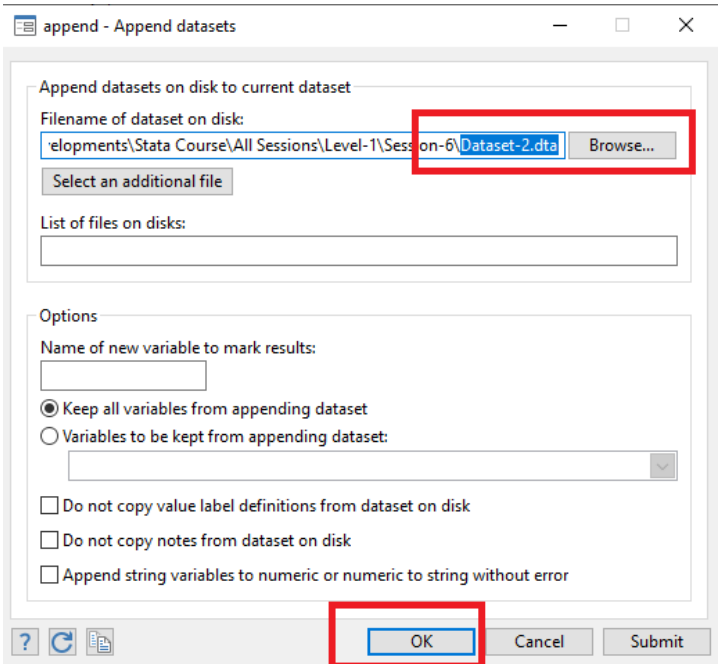
Step-3: Open the Dataset-1 and you can see the following data with 30 records



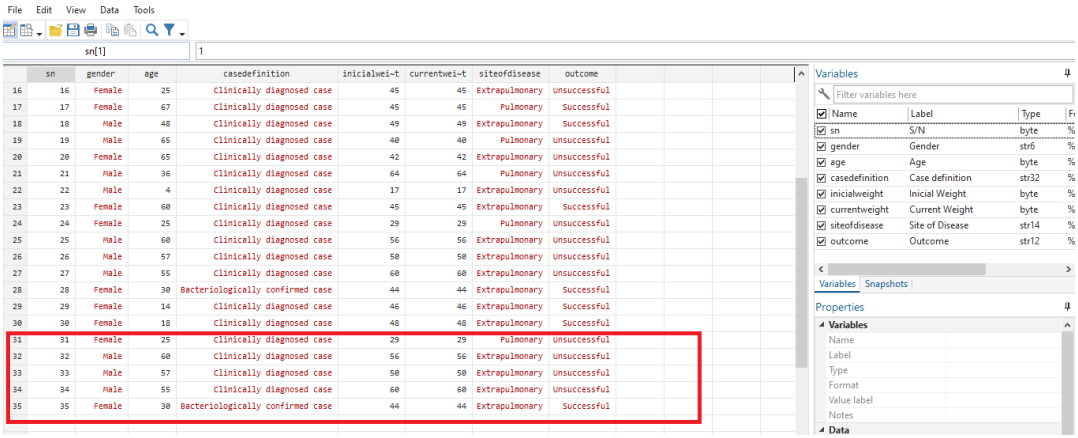
Step-04: Go to Stata >> Data >> Combine Datasets >> Append Datasets



Step-5: Click on Append datasets and you can see the following screen and browse the second dataset and click on Ok button



Step-06: Results



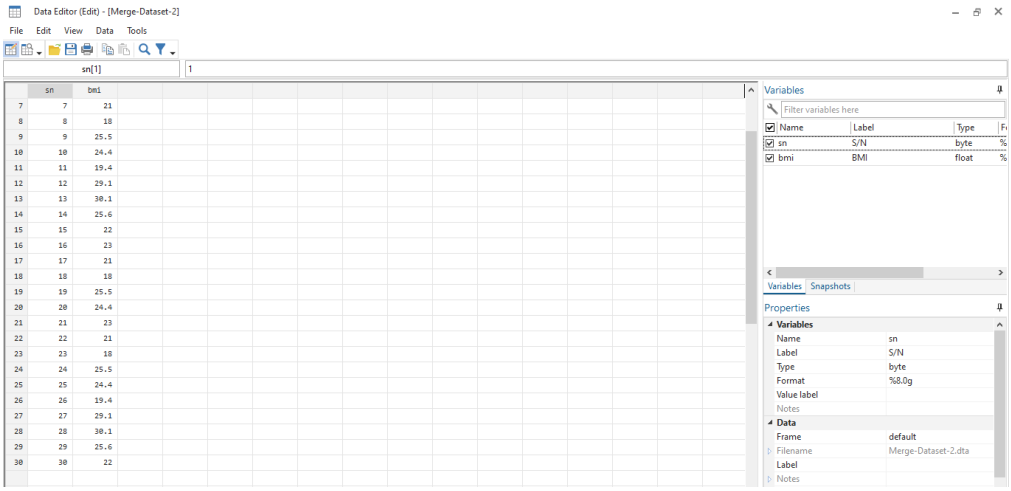
Stata Commands:

use "F:\Skill Developments\Stata Course\All Sessions\Level-1\Session-6\Dataset-1.dta"
append using "F:\Skill Developments\Stata Course\All Sessions\Level-1\Session-6\Dataset-2.dta"

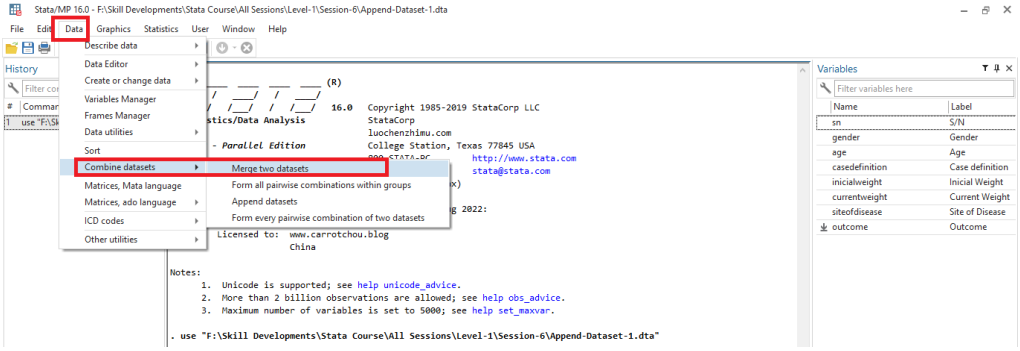
- Merge
  Merging two datasets require that both have at least one variable in common (either string or numeric). If string make sure the categories have the same spelling (i.e. country names, etc.). The common variables must have the same name.

Example: We want to add 30 BMI results for each of the patients in our first Dataset. So we need to keep the one unique ID in both datasets.

Step-1: Keep the Dataset-1 as it is. Prepare sencond dataset as:

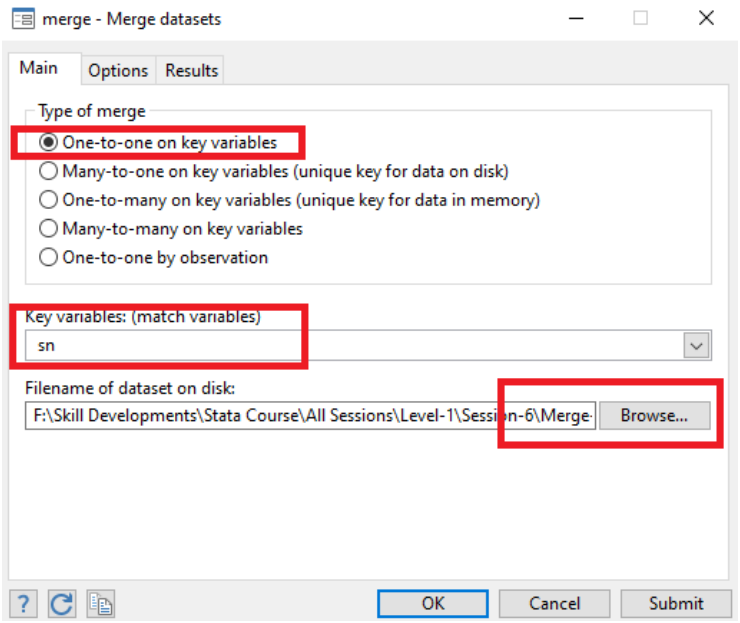


Step-2: Open dataset-1and then go Data >> Combine datasets >> Mege two datasets

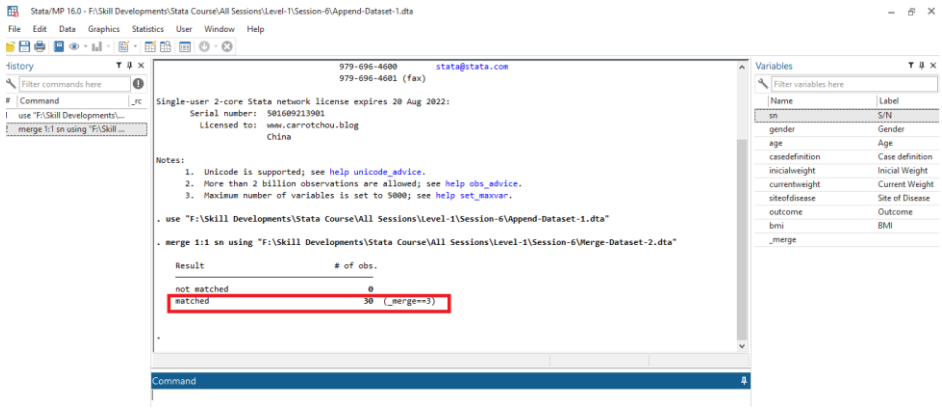


Step-3: Follow the below steps:

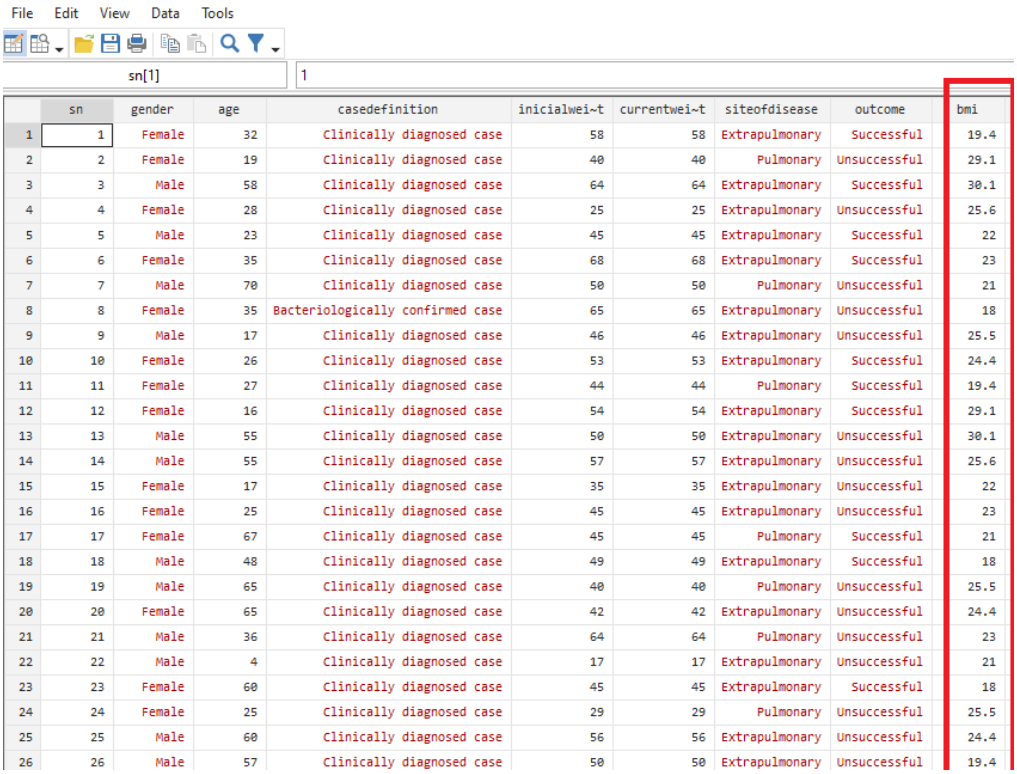
- Select One-one on key variables
- Key variables as “sn”
- Browse the Dataset-2



Click on the Ok button and you can see the following results



Step-04: results



	sn	gender	age	casedefinition	inicialwei~t	currentwei~t	siteofdisease	outcome	bmi
1	1	Female	32	Clinically diagnosed case	58	58	Extrapulmonary	Successful	19.4
2	2	Female	19	Clinically diagnosed case	40	40	Pulmonary	Unsuccessful	29.1
3	3	Male	58	Clinically diagnosed case	64	64	Extrapulmonary	Successful	30.1
4	4	Female	28	Clinically diagnosed case	25	25	Extrapulmonary	Unsuccessful	25.6
5	5	Male	23	Clinically diagnosed case	45	45	Extrapulmonary	Successful	22
6	6	Female	35	Clinically diagnosed case	68	68	Extrapulmonary	Successful	23
7	7	Male	70	Clinically diagnosed case	50	50	Pulmonary	Unsuccessful	21
8	8	Female	35	Bacteriologically confirmed case	65	65	Extrapulmonary	Unsuccessful	18
9	9	Male	17	Clinically diagnosed case	46	46	Extrapulmonary	Unsuccessful	25.5
10	10	Female	26	Clinically diagnosed case	53	53	Extrapulmonary	Successful	24.4
11	11	Female	27	Clinically diagnosed case	44	44	Pulmonary	Successful	19.4
12	12	Female	16	Clinically diagnosed case	54	54	Extrapulmonary	Successful	29.1
13	13	Male	55	Clinically diagnosed case	50	50	Extrapulmonary	Unsuccessful	30.1
14	14	Male	55	Clinically diagnosed case	57	57	Extrapulmonary	Unsuccessful	25.6
15	15	Female	17	Clinically diagnosed case	35	35	Extrapulmonary	Unsuccessful	22
16	16	Female	25	Clinically diagnosed case	45	45	Extrapulmonary	Unsuccessful	23
17	17	Female	67	Clinically diagnosed case	45	45	Pulmonary	Successful	21
18	18	Male	48	Clinically diagnosed case	49	49	Extrapulmonary	Successful	18
19	19	Male	65	Clinically diagnosed case	40	40	Pulmonary	Unsuccessful	25.5
20	20	Female	65	Clinically diagnosed case	42	42	Extrapulmonary	Unsuccessful	24.4
21	21	Male	36	Clinically diagnosed case	64	64	Pulmonary	Unsuccessful	23
22	22	Male	4	Clinically diagnosed case	17	17	Extrapulmonary	Unsuccessful	21
23	23	Female	60	Clinically diagnosed case	45	45	Extrapulmonary	Successful	18
24	24	Female	25	Clinically diagnosed case	29	29	Pulmonary	Unsuccessful	25.5
25	25	Male	60	Clinically diagnosed case	56	56	Extrapulmonary	Unsuccessful	24.4
26	26	Male	57	Clinically diagnosed case	50	50	Extrapulmonary	Unsuccessful	19.4

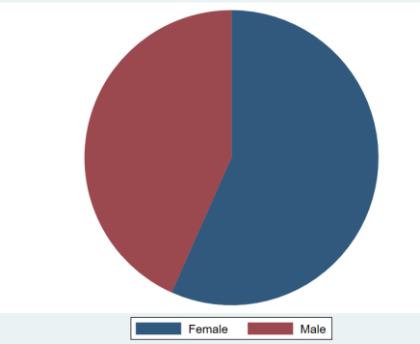
Lession-2: Variable groups - age

Example:

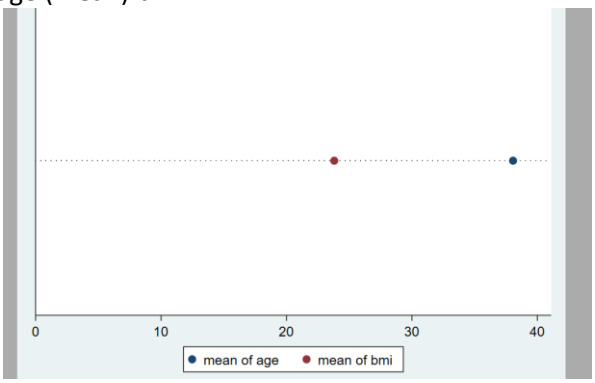
```
gen ageGroup = age
recode ageGroup (0/17 = 0) (18/34 = 1) (35/51 = 2) (52/64 = 3) (65/140 = 4)
label define ageGroup 0 "0-17 Years" 1 "18-34 Years" 2 "35-51 Years" 3 "52-64 Years" 4 ">= 65 Years"
label values ageGroup ageGroup
tab ageGroup
```

Lession-3: Statistical graph

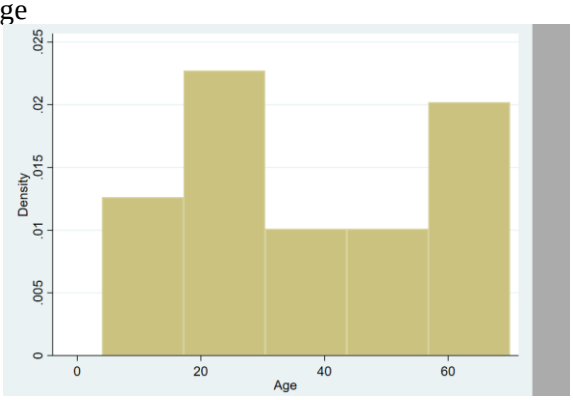
- **Bar:** graph bar (mean) ageGroup
- **Column:** graph hbar ageGroup
- **Pie**: graph pie, over(gender)



- **Dot:** graph dot (mean) age (mean) bmi



- **Line:** twoway (line age bmi), line bmi ageGroup
- **Histogram:** histogram age



Next Session:

- Article writing style and Stata result input to your article
- Descriptive data analysis

Good Luck!

If you need any support, don't heasitate to let me know.