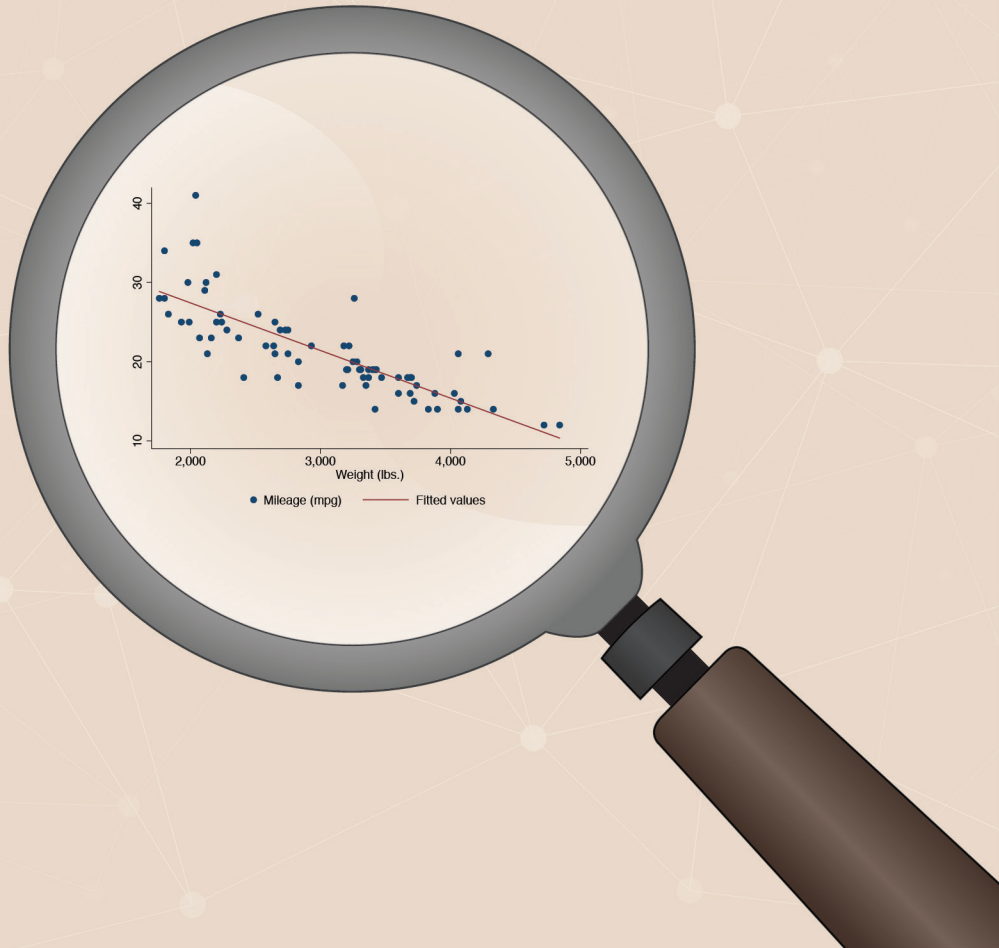


مدیریت، تحلیل و نمایش دیتا

رهنمود عملی با استفاده از سروی مردم افغانستان





The Asia Foundation

مدیریت، تحلیل و نمایش دیتا

رهنمود عملی با استفاده از سروی مردم افغانستان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

مدیریت، تحلیل و نمایش دیتا

نویسنده

فہیم احمد یوسفزی

طرح و دیزاین

سید راشد سادات

چاپ اول: کابل، ۱۳۹۸

صاحب امتیاز: بنیاد آسیا، دفتر افغانستان

(The Asia Foundation)

تمام حقوق این اثر محفوظ است. تکثیر یا تولید مجدد آن به هر صورت بدون اجازه

صاحب امتیاز آن ممنوع میباشد.

فهرست عناوین

فصل اول: مدیریت دیتا.....	۲
1. آشنایی با صفحه ی برنامه ی ستاتا.....	۲
2. وارد کردن دیتا از اکسل به ستاتا (بخش اول).....	۳
3. ذخیره کردن دیتا.....	۵
4. ذخیره کردن فرمان ها.....	۶
5. حذف کردن مشاهدات و متغیرات خاص.....	۷
6. تبدیل کردن نام متغیرات.....	۸
7. ایجاد یک متغیر جدید (بخش اول).....	۹
8. اضافه کردن لیبل و نوت در متغیرات.....	۱۲
9. استفاده از اصطلاح in برای قید کردن مشاهدات.....	۱۴
10. استفاده از اصطلاح if برای قید کردن مشاهدات (بخش اول).....	۱۸
11. گرداننده های عقلانی و منطقی.....	۱۹
12. ایجاد تغییر در یک متغیر.....	۲۳
13. ایجاد یک متغیر جدید (قسمت دوم).....	۲۵
14. استفاده از اصطلاح if برای قید کردن مشاهدات (بخش دوم).....	۲۹
15. تبدیل کردن یک متغیر پیوسته به متغیر کتگوری شده (بخش اول).....	۳۱
16. مشاهده نمودن لیبل های ارزش های یک متغیر.....	۳۴
17. اضافه کردن لیبل بالای ارزش های یک متغیر.....	۳۷
18. تبدیل نمودن متغیر های غیر عددی به متغیر عددی.....	۳۹
19. تبدیل کردن یک متغیر پیوسته به متغیر کتگوری شده (بخش دوم).....	۴۰
20. جدا ساختن متغیر های غیر عددی.....	۴۴
21. از بین بردن فاصله ها در یک متغیر غیر عددی.....	۴۵
22. مقایسه نمودن دو دیتاست.....	۴۷

فصل دوم: تحلیل دیتا ۵۰

23. جدول سازی یک طرفه ۵۰

24. جدول سازی دو طرفه ۵۱

25. جدول سازی برای متغیرات چندین جوابه ۵۶

26. آمار توصیفی احصایوی ۶۲

27. استفاده از فرمان bysort ۶۷

28. پیدا کردن ارزش های خالی در متغیرات ۶۹

29. آزمون T و یا t-test ۷۱

30. آزمون همبستگی ۷۳

فصل سوم: نمایش دیتا ۸۰

31. پای چارت ۸۰

32. بار چارت ۸۳

33. باکس گراف ۸۶

34. گراف پراگندگی ۸۹

35. گراف خطی ۹۱

36. نمایش سوالات چندین جوابه ۹۱

37. موضوعات عمومی مربوط به گراف ۹۶

سپاسگذاری

این کتاب محصول همکاری و زحمات زیاد از کارمندان و همکاران بنیاد آسیا میباشد. این کتاب تحت نظر عبدالله احمدزی، ریس بنیاد آسیا در افغانستان و توسط فهیم احمد یوسفزی، عضو تیم پالیسی و تحقیق بنیاد آسیا، تحت هدایت داکتر تبسم اکسیر تهیه شده است و سید راشد سادات طرح و دیزاین این کتاب را انجام داده است.

در اینجا جا دارد از اعضای تیم پالیسی و تحقیق هر یک مصطفی یادگاری، خدیجه حیات و محمد شریح شیون که وقت گرانبهای خویش را برای مرور و ارایه نظریات و پیشنهادات در غنامندی و بهتر شدن این اثر صرف نموده اند ابراز سپاس و امتنان نمایم.

همچنان بنیاد آسیا از اداره انکشاف بین المللی ایالات متحده امریکا (USAID)، وزارت امور خارجه و تجارت استرالیا (DFAT) و انجمن همکاری های بین المللی آلمان (GIZ) بخاطر حمایت شان از برنامه های ارتقای ظرفیت تحقیقاتی در افغانستان ابراز سپاس و امتنان مینماید.

در مورد سروی مردم افغانستان

برای اینکه در اکثریت از مثال های این کتاب از دیتای سروی مردم افغانستان استفاده شده است بهتر است در مورد این سروی بیشتر بدانیم تا در قسمت تحلیل آن سهولت ایجاد شود. سروی مردم افغانستان منحصیث سروی عمومی سالانه ی بنیاد آسیا یکی از دیرینه ترین سروی سرتاسری در افغانستان میباشد که دربر گیرنده نظریات و دیدگاه های بیشتر از 112000 افغان در موارد همچون اقتصاد، دسترسی به خدمات، مسایل جوانان، نظریات افغانان در مورد نقش زنان در جامعه، مهاجرت و دسترسی به معلومات میباشد.

نمونه سروی سال 2018، شامل 15012 زن و مرد دارای سنین 18 سال و بزرگتر از آن می باشد که از تمام مردم افغانستان در 34 ولایت نماینده گی می کند. این نمونه متشکل از 50,3% مردان و 49,7% زنان است که دربر گیرنده ی 19,4% خانوارهای شهری و 80,6% خانوار های روستایی میباشد. نمونه سروی با استفاده از تازه ترین اطلاعات جمعیتی (2016-2017) منتشره از سوی اداره احصائیه مرکزی افغانستان غرض نمایندگی به سطح ملی و تعادل جنسیتی توزین گردیده است، زیرا این کتاب رهنمود برای مدیریت، تحلیل و نمایش دیتا میباشد و توزین دیتا خارج از موضوع بحث است، بنابراین در مثال های ارایه شده در این کتاب از توزین استفاده نشده است.

برای توزین دیتای سروی مردم افغانستان از متغیر MergeWgt10 استفاده شده میتواند.

هدف بنیاد آسیا از انجام این سروی، تهیه به موقع دیتا و تحلیل آن به خاطر حمایت از تلاش های حکومت و اتباع افغانستان جهت ساختن یک جامعه با ثبات و شگوف، میباشد. گزارش و مجموعه ی دیتا در وبسایت <http://asiafoundation.org/afghansurvey/> قابل دسترس است.

متغیرات که زیاد ترین کاربرد را در این رهنمود داشته است قرار ذیل میباشد.

- z1: جنسیت پاسخ دهندهگان.
- z2: سن پاسخ دهندهگان.
- z55: بلندترین سطح تحصیل پاسخ دهندهگان.
- x4: نظریات پاسخ دهندهگان در مورد اینکه افغانستان به سمت درست در حرکت است و یا سمت اشتباه.
- z20: نظریه دهندهگان در مورد اینکه اگر فرصت مهیا شود افغانستان را ترک میکنند و یا نه.
- x82: سطح رضایت دهندهگان از دموکراسی.
- z47: سطح خوشحالی دهندهگان به صورت عموم.
- x369: نظریات دهندهگان در رابطه به سن مناسب ازدواج برای دختران.
- x370: نظریات دهندهگان در رابطه به سن مناسب ازدواج برای پسران.
- m4: زون های افغانستان.
- m7: ولایات افغانستان.
- m6b: محل زیست پاسخ دهندهگان (شهری و روستایی).
- m8: سال های که سروی انجام شده است.

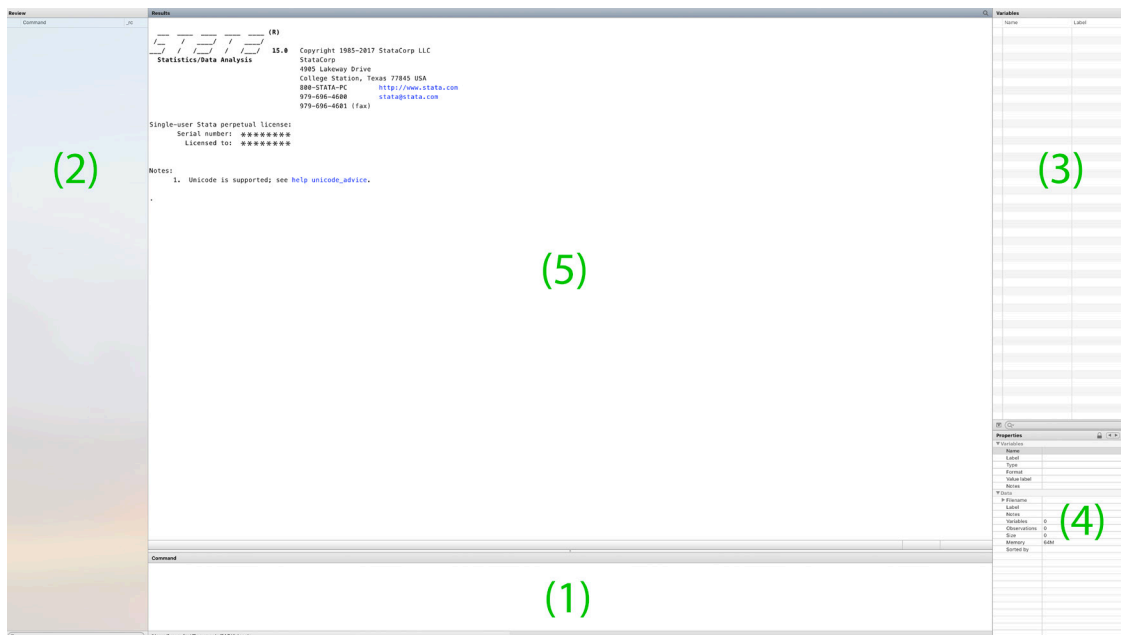
فصل اول

مدیریت دیتا Data Management

امروزه کاربرد آمار و دیتا یکی از ارکان اساسی بسیاری از رشته های علمی به ویژه علوم اجتماعی را تشکیل می دهد و همچنان استفاده از دیتا و آمار برای اخذ تصمیم گیری های آگاهانه یک امر اجتناب ناپذیر می باشد. نخستین مرحله بعد از جمع آوری دیتا، ترتیب دادن آن به یک نظم خاص می باشد تا برای تحلیل آماده شود. بعضی از تحلیلگران دیتا به این باور هستند که در حدود هشتاد فیصد وقت یک تحلیلگر دیتا صرف تنظیم کردن و آماده ساختن دیتا برای تحلیل می شود که خود نمایان گر اهمیت مدیریت دیتا می باشد. در این بخش موضوعات مربوط به اینکه چطور دیتای خام برای تحلیل به بهترین شکل آن تنظیم و آماده شود و اینکه کدام نکات در هنگام مدیریت دیتا مدنظر گرفته شود بحث گردیده است.

1. آشنایی با صفحه ی برنامه ی ستاتا

چون این کتاب رهنمای است که چگونه از برنامه ستاتا (Stata) برای مدیریت، تحلیل و نمایش دیتا استفاده شود در نخست در مورد نحوه کاربرد این نرم افزار و اینکه چگونه از این برنامه به شکل خوبتر آن استفاده شود معلومات داده شده است. برنامه ستاتا یکی از سافت ویر های پر کاربرد در بخش تحلیل دیتا می باشد که از قدرت بالای در قسمت مدیریت دیتا و تحلیل دیتا برخوردار است. بعد از نصب برنامه ی ستاتا زمان که برنامه ستاتا را باز کنید صفحه ی ذیل برای تان نمایش داده می شود که عبارت از صفحه ی اصلی ستاتا می باشد.

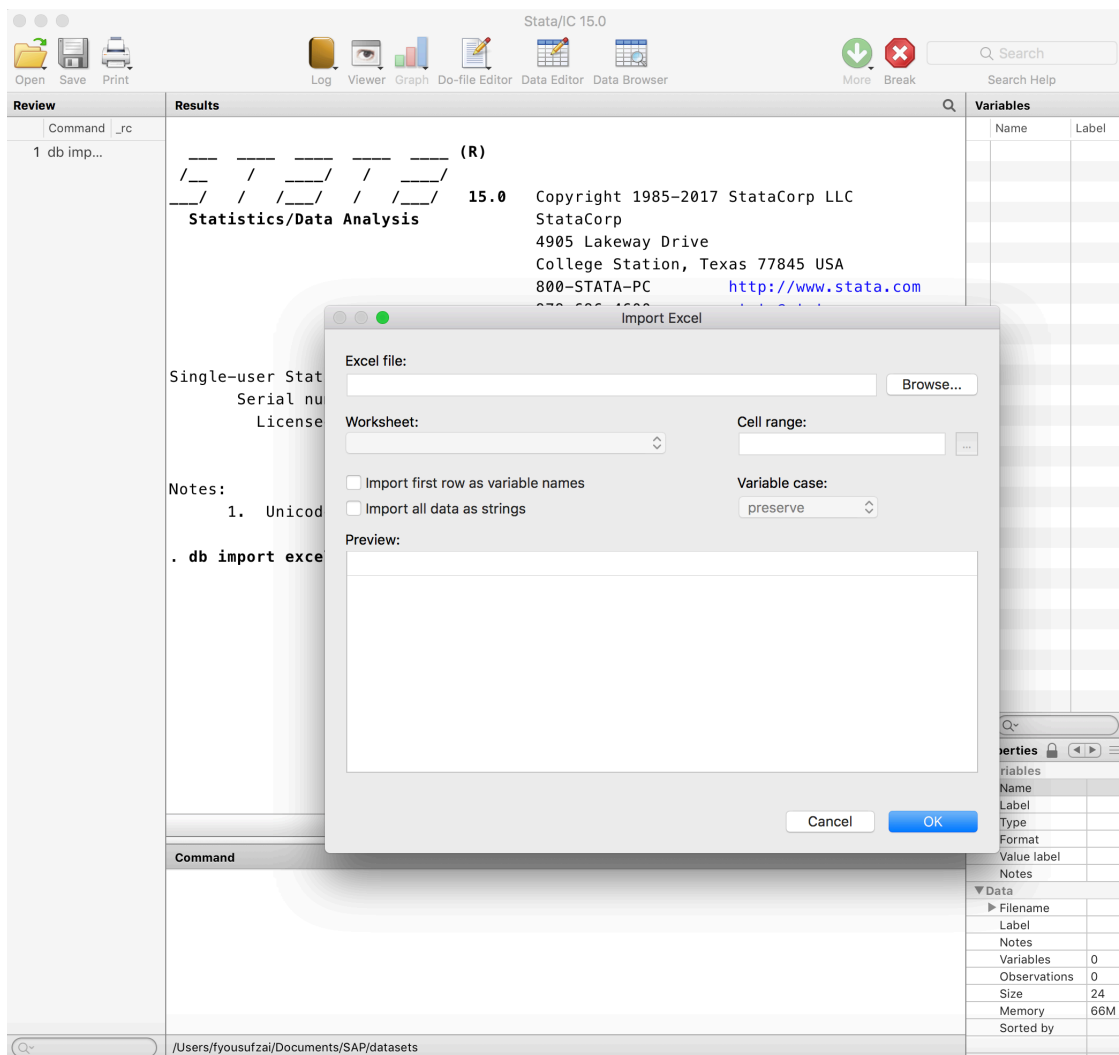


صفحه ی اصلی ستاتا دارای بخش های اساسی ذیل میباشد.

1. قسمت فرمان ها (Command): در این قسمت تمام فرمان های را که میخواهید انجام شود نوشته میشود.
2. بخش بازدید از فرمان ها (Review): در این قسمت تمام فرمان های که قبلاً اجرا شده است برای استفاده ی دوباره ی آن در آینده ذخیره میشود که با یک بار کلیک بالای هر فرمان که در قسمت بازدید فرمان ها عین فرمان در قسمت فرمان ها نمایش داده میشود و با دو بار کلیک بالای هر فرمان همان فرمان اجرا میشود.
3. متغیر ها (Variables): در قسمت متغیر ها تمام متغیرات و یا نام هر ستون دیتا همراه با لیبل آن نمایش داده میشود.
4. مشخصات (Properties): در این قسمت بعضی از مشخصات متغیرات به مانند نوع متغیر، لیبل متغیر و غیره موضوعات نمایش داده میشود.
5. نتایج (Results): در این بخش نتایج تمام فرمان های انجام شده نمایش داده میشود.

2. وارد کردن دیتا از اکسل به ستاتا (بخش اول)

برای وارد کردن دیتا از فایل اکسل به ستاتا در بخش فرمان ها db import excel نوشته بعداً صفحه ی ذیل باز میگردد که از طریق گزینه browse فایل که به فارمت اکسل باشد را برای وارد نمودن به ستاتا انتخاب کرده میتوانید.



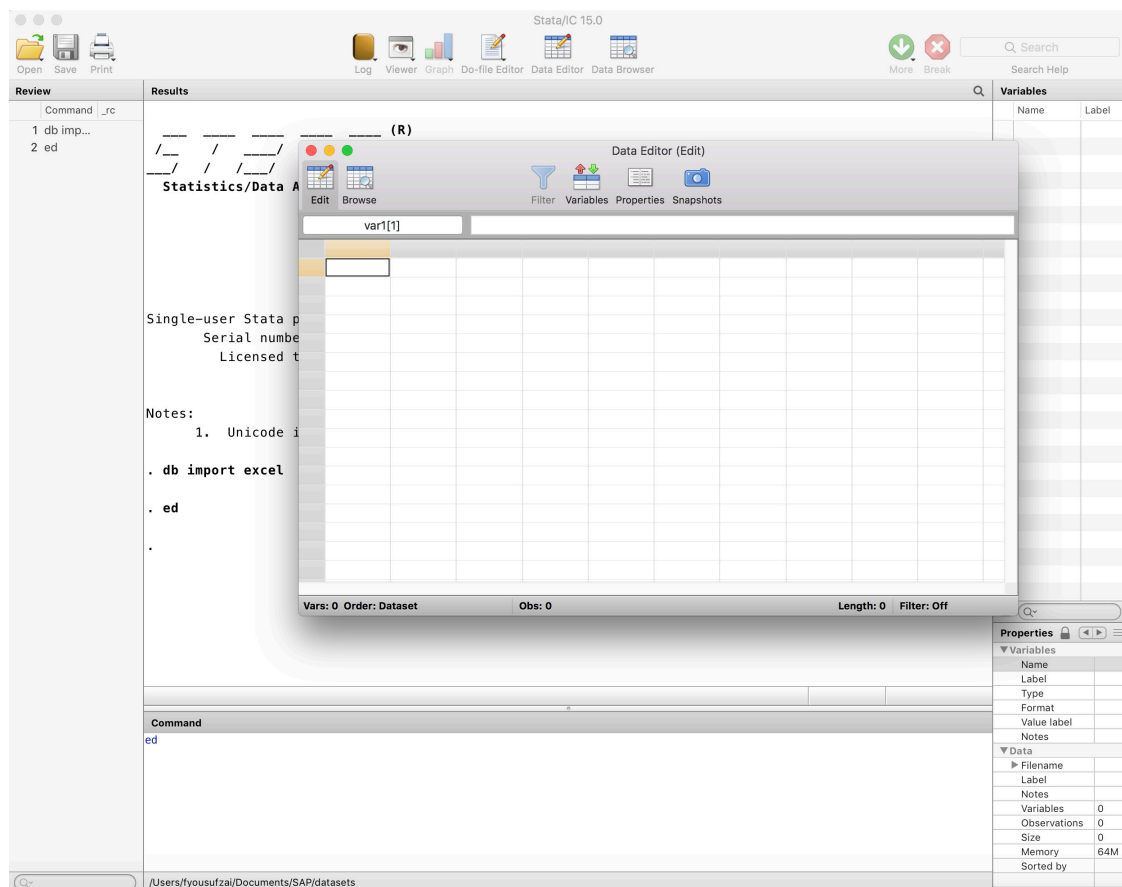
1. Browse: در صفحه ی بالا از بخش browse میتونید فایل مورد نظر تان را انتخاب کنید
2. Worksheet: به یاد داشته باشید که ستاتا به صورت خودکار اولین صفحه ی فایل اکسل را انتخاب میکند، در صورت که فایل شما دارای چندین صفحه باشد میتونید صفحه ی مورد نیاز را از طریق گزینه ی worksheet انتخاب کنید.
3. Import first row as variable names: بعد از اینکه فایل مورد نظر و صفحه ی مشخص را انتخاب کردید باید ببینید که اولین ردیف دیتای شما در فایل اکسل نام متغیرات است و یا خیر. در صورت که اولین ردیف در فایل اکسل نام متغیرات باشد گزینه ی import first row as variable names را انتخاب کنید در غیر آن صورت را به حالت خودش بگذارید.
4. Variable case: از طریق این گزینه میتوانید انتخاب کنید که نام متغیرات به حروف کلان، حروف خورد و یا به همان شکل اصلی آن بدون کدام تغییر نمایش داده شود.

5. Cell range: از طریق این گزینه میتوانید یک بخش مشخص از دیتا را برای وارد کردن به سافتویر ستاتا انتخاب کنید.

قبل از اینکه دیتا را به ستاتا وارد کنید بهتر است که به شکل منظم tidy باشد، یعنی هر ستون نمایانگر یک سوال (متغیر) باشد، هر ردیف نمایانگر یک مشاهده باشد.

3. ذخیره کردن دیتا

برعلاوه از صفحه ی اصلی ستاتا که نتیجه ی تمام فرمان های انجام شده در آنجا نمایش داده میشود، زمان که دیتا را به سافتویر ستاتا وارد کردید تمام دیتای خام در بخش data editor ذخیره میشود که میتوانید با دو بار کلیک در قسمت data editor و یا با نوشتن edit در بخش فرمان ها صفحه data editor را باز کنید.

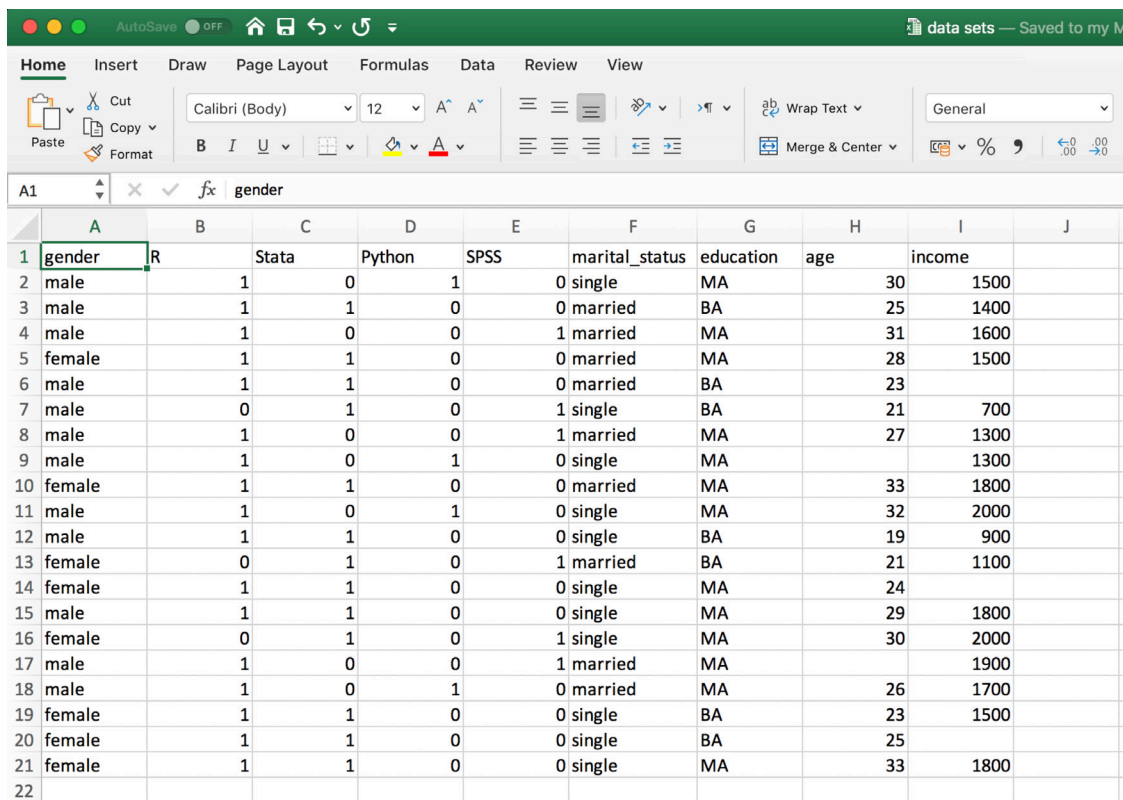


به یاد داشته باشید که با بسته کردن صفحه ی ستاتا تمام دیتا حذف میشود. برای اینکه که دیتای وارد شده را به فارمت ستاتا (dta.) ذخیره کنید میتوانید که از فرمان save کار بگیرید. به طور مثال در صورت که خواسته باشید دیتای وارد شده را به نام first_day ذخیره کنید پس میتوانید از فرمان ذیل کار بگیرید.

save first_day.dta

زمان که دیتا را به فارمت ستاتا ذخیره کردید، بار دیگر میتوانید از طریق مینوی file و گزینه ی open دیتای مورد نظر تان را باز کنید و یا هم میتوانید این کار را با دو بار کلیک بالای دیتای مورد نظر انجام دهید.

برای مثال دیتای ذیل را که در فارمت اکسل میباشد آن را به ستاتا وارد کرده و به نام test.dta ذخیره کنید.



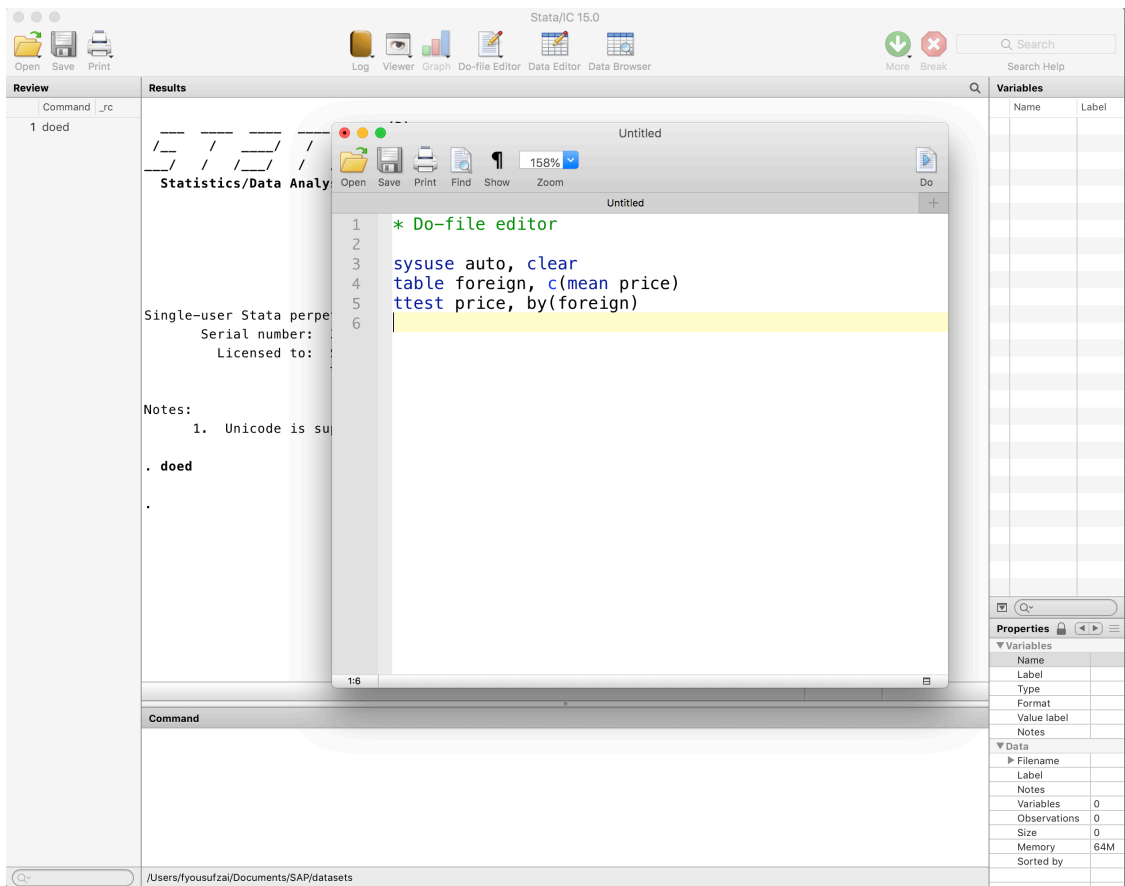
	A	B	C	D	E	F	G	H	I	J
1	gender	R	Stata	Python	SPSS	marital_status	education	age	income	
2	male	1	0	1	0	0 single	MA	30	1500	
3	male	1	1	0	0	0 married	BA	25	1400	
4	male	1	0	0	1	1 married	MA	31	1600	
5	female	1	1	0	0	0 married	MA	28	1500	
6	male	1	1	0	0	0 married	BA	23		
7	male	0	1	0	1	1 single	BA	21	700	
8	male	1	0	0	1	1 married	MA	27	1300	
9	male	1	0	1	0	0 single	MA		1300	
10	female	1	1	0	0	0 married	MA	33	1800	
11	male	1	0	1	0	0 single	MA	32	2000	
12	male	1	1	0	0	0 single	BA	19	900	
13	female	0	1	0	1	1 married	BA	21	1100	
14	female	1	1	0	0	0 single	MA	24		
15	male	1	1	0	0	0 single	MA	29	1800	
16	female	0	1	0	0	1 single	MA	30	2000	
17	male	1	0	0	1	1 married	MA		1900	
18	male	1	0	1	0	0 married	MA	26	1700	
19	female	1	1	0	0	0 single	BA	23	1500	
20	female	1	1	0	0	0 single	BA	25		
21	female	1	1	0	0	0 single	MA	33	1800	
22										

Example.xlsx (<https://github.com/Fahim-Ahmad/Stata-programming>)

4. ذخیره کردن فرمان ها

قبلاً خواندید تمام فرمان های که انجام دهید در بخش بازدید و یا review ذخیره میشود، اما فرمان ها در این بخش موقتاً میباشد و در صورت که صفحه ی ستاتا را بسته کنید تمام فرمان ها از آن بخش حذف میشود.

ستاتا یک بخش دیگر به نام do-file برای ذخیره کردن فرمان ها دارد که شما را قادر میسازد تمام فرمان ها و نوت های مورد نیاز تان را در آنجا ذخیره کنید که میتوانید با کلیک کردن بالای do-file editor و یا با نوشتن فرمان doed در بخش فرمان ها یک do-file جدید باز کنید.



5. حذف کردن مشاهدات و متغیرات خاص

در صورت که خواسته باشید بعضی از متغیرات و یا مشاهدات خاص را از دیتا حذف کنید میتوانید از فرمان drop کار بگیرید. ساختار فرمان drop برای حذف مشاهدات قرار ذیل است.

drop in range

که به عوض range میتوانید یک نمبر و یا یک محدوده از نمبرات را نوشته کنید.

به طور مثال در صورت که خواسته باشید در یک دیتاست مشاهده ی پنجم را حذف کنید میتوانید نوشته کنید *drop in 5*. به همین ترتیب در صورت که خواسته باشید از مشاهده ی پنجم الی دهم را حذف کنید میتوانید *drop in 5/10* نوشته کنید. در صورت که خواسته باشید بعضی از متغیرات را حذف کنید میتوانید با نوشتن *drop* و نام های متغیرات به تعقیب آن این کار را انجام بدهید. ساختار فرمان drop برای حذف متغیرات قرار ذیل است.

drop var_list

به عوض varlist میتوانید نام یک و یا چندین متغیر را نوشته کنید. به طور مثلا در صورت که در دیتا سیت شما متغیرات *gender* و *age* داشته باشید و بخواهید هر دوی آن را حذف کنید باید که *drop gender age* نوشته کنید و در صورت که خواسته باشید تنها متغیر جنسیت را حذف کنید میتوانید *drop gender* نوشته کنید.

مثال: در قدم نخست دیتای سروی مردم افغانستان را باز کنید و با نوشتن فرمان `dis _N` میتوانید ببینید که دیتای سروی مردم افغانستان دربرگیرنده ی بیشتر از 10000 مشاهده و یا observation میباشد. در صورت که خواسته باشد از مشاهده ی اولی الی مشاهده 50000 حذف کنید پس میتوانید از فرمان ذیل استفاده کنید.

```
drop in 1/50000
```

با نوشتن فرمان `ds` نام تمام متغیرات که در دیتا سیت سروی مردم افغانستان ذخیره است در صفحه ی ستاتا لست میشود. در صورت که خواسته باشید متغیرات سن (`z2`)، سطح تحصیل (`z55`) و جنسیت (`z1`) را از دیتای عمومی حذف کنید میتوانید که از فرمان ذیل کار بگیرید.

```
drop z2 z55 z1
```

در صورت که خواسته باشید تعداد زیاد از متغیرات/مشاهدات را حذف کنید و چند تعداد محدود از متغیرات/مشاهدات را در دیتا داشته باشید بهتر است از فرمان `keep` استفاده شود. فرمان `keep` برعکس فرمان `drop` میباشد. یعنی با نوشتن `drop in 1/10` میتوانید مشاهدات اولی الی دهم را حذف کنید ولی با نوشته فرمان `keep in 1/10` تنها مشاهدات اولی الی دهم ذخیره میشود و دیگر تمام مشاهدات حذف میشود به همین ترتیب با نوشتن فرمان `drop z1 z2 z55` متغیرات `z1`، `z2` و `z55` حذف میشود و دیگر متغیرات هنوز هم در دیتاست موجود میباشد ولی با اجرای فرمان `keep z1 z2 z55` تمام متغیرات حذف میشود و تنها همان متغیرات ذکر شده در دیتاست باقی میماند.

به طور مثال در صورت که خواسته باشید در دیتای سروی مردم افغانستان از مشاهده ی اولی الی 50000 در دیتا موجود باشد و دیگر مشاهدات حذف گردد میتوانید از فرمان ذیل استفاده کنید.

```
Keep in 1/50000
```

در صورت که خواسته باشید در دیتای سروی مردم افغانستان تنها متغیرات سن (`z2`)، سطح تحصیل (`z55`) و جنسیت پاسخ دهنده (`z1`) را داشته باشید و دیگر تمام متغیرات حذف گردد میتوانید از فرمان ذیل استفاده کنید.

```
keep z1 z2 z55
```

میتوانید که هر دو فرمان `keep` و `drop` را همراه با اصطلاح `if` یکجا استفاده کنید و بعضی از مشاهدات را به اساس یک مشخصه ی متغیر دیگر حذف کنید. به طور مثال در صورت که خواسته باشید در سروی مردم افغانستان تمام مشاهدات دیگر از بین برود و تنها دیتای سال 2018 را داشته باشید میتوانید از یکی از فرمان های ذیل استفاده کنید.

```
drop if m8!=2018
```

```
keep if m8==2018
```

در دروس دهم و چهاردهم به صورت مفصل در مورد چگونگی استفاده از اصطلاح `if` بحث شده است.

6. تبدیل کردن نام متغیرات

برای تبدیل کردن نام متغیرات میتوانید از فرمان `rename` کار بگیرید. ساختار تبدیل کردن نام یک متغیر قرار ذیل میباشد.

```
rename old_var new_var
```

به طور مثال در صورت که نام متغیر `education` باشد و بخواهید آنرا به `edu` تبدیل کنید میتوانید از فرمان ذیل استفاده کنید.

```
rename education edu
```

همچنان میتوانید نام چندین متغیر را به صورت همزمان نیز تغییر دهید. ساختار فرمان `rename` برای تبدیل نمودن نام چندین متغیر به صورت همزمان قرار ذیل میباشد.

```
(rename (old_varlist) (new_varlist
```

به طور مثال در صورت که شما ۳ متغیر به نام های `education`، `ethnicity` و `Age` داشته باشد و بخواهید که آنها را به نام های `edu`، `ethnic` و `age` تبدیل کنید میتوانید از فرمان ذیل کار بگیرید.

```
(rename (education ethnicity Age) (edu ethnic age
```

مثال: در قدم نخست دیتای سروی مردم افغانستان را باز کنید. در دیتای سروی مردم افغانستان متغیر z1 نشان دهنده ی جنسیت پاسخ دهنده و متغیر z55 نشان دهنده ی سطح تحصیل پاسخ دهنده میباشد که میتوانید آنرا با اجرای فرمان `edit z1 z55` در بخش `data editor` مشاهده کنید.

برای تبدیل کردن نام های متغیرات z1 و z55 به gender و education میتوانید از فرمان ذیل کار بگیرید.
`rename (z1 z55) (gender education)`
و نتیجه ی آنرا در `data editor` میتوانید با اجرای فرمان `edit gender education` ببینیم.

نوت: به یاد داشته باشید که در ستاتا برای تعیین نام متغیرات بعضی از قوانین قرار ذیل وجود دارد:

a. دو متغیر نمیتواند در عین زمان دارای یک نام باشد.

b. نام متغیرات با عدد شروع نمیشود.

c. در نام متغیرات از علامات ریاضیکی مانند جمع، تفریق، ضرب و تقسیم استفاده نمیشود.

همچنان از فرمان `rename` میتوانید برای تغییر دادن نام متغیرات به حروف خورد و یا کلان استفاده کنید. در صورت که خواسته باشید نام یک و یا چندین متغیر به حروف کلان نمایش داده شود میتوانید از فرمان ذیل کار بگیرید.

```
rename var_list, upper
```

در صورت که خواسته باشید نام یک و یا چندین متغیر به حروف خورد نمایش داده شود میتوانید از فرمان ذیل کار بگیرید.
`rename var_list, lower`

7. ایجاد یک متغیر جدید (بخش اول)

برای ایجاد یک متغیر جدید میتوانید از فرمان `generate` و یا مخفف آن `gen` کار بگیرید. ساختار فرمان `generate` برای ایجاد یک متغیر عددی قرار ذیل میباشد.

```
generate var_name=value
```

به طور مثال در صورت که خواسته باشید یک متغیر جدید به نام `year` بسازید که تمام ردیف های آن دربر گیرنده ی عدد 2019 باشد میتوانید از فرمان ذیل کار بگیرید.

```
generate year = 2019
```

ساختار فرمان `generate` برای ایجاد یک متغیر غیر عددی قرار ذیل میباشد.

```
generate var_name="text"
```

به طور مثال در صورت که خواسته باشید یک متغیر جدید به نام `country` بسازید که تمام مشاهدات آن به نام افغانستان باشد میتوانید از فرمان ذیل کار بگیرید.

```
generate country="Afghanistan"
```

نوت: در صورت ایجاد متغیر جدید باید در قسمت نام گذاری متغیرات که قبلاً بیان گردید دقت داشته باشید.

همچنان میتوانید با استفاده از فرمان `generate` یک متغیر جدید بسازید که کاملاً مشابه به یک متغیر از قبل ذخیره شده در دیتاست باشد. در آن صورت ساختار فرمان فوق قرار ذیل میباشد:

```
generate new_var=old_var
```

به طور مثال در دیتا سروی مردم افغانستان متغیر سن به نام z2 ذخیره گردیده است، در صورت که بخواهید یک متغیر دیگر به نام Age که مشابه آن باشد ایجاد کنید میتوانید از فرمان ذیل استفاده کنید:

generate Age = z2

و نتیجه ی آن را میتوانید با اجرای فرمان edit Age z2 ببینید.

Age[1]		18
	Age	z2
1	18	18
2	20	20
3	20	20
4	50	50
5	22	22
6	35	35
7	24	24
8	32	32
9	24	24
10	35	35
11	20	20
12	23	23
13	19	19
14	28	28
15	40	40
16	30	30
17	40	40
18	30	30
19	48	48
20	30	30

همچنان میتوانید که عملیات ریاضیکی را بالای یک متغیر جدید با استفاده از فرمان generate نیز انجام بدهید. به طور مثال در سروی مردم افغانستان متغیر z2 که نشان دهنده ی سن افراد سروی شده میباشد، در صورت که خواسته باشید یک متغیر جدید به نام dob که مخفف date of birth میباشد ایجاد کنید که نشان دهنده ی سال تولد هر شخص باشد میتوانید این کار را با تفریق کردن سن افراد سروی شده از سال فعلی انجام دهید.

generate dob=2019-z2

z2[1]		18
	z2	dob
1	18	2001
2	20	1999
3	20	1999
4	50	1969
5	22	1997
6	35	1984
7	24	1995
8	32	1987
9	24	1995
10	35	1984
11	20	1999
12	23	1996
13	19	2000
14	28	1991
15	40	1979
16	30	1989
17	40	1979
18	30	1989
19	48	1971
20	30	1989

به همین ترتیب در صورت که خواسته باشید یک متغیر جدید به نام age_10 ایجاد کنید که نشان دهنده ی سن 10 سال بعد افراد سروی شده باشد میتوانید از فرمان ذیل استفاده کنیم.

```
generate age_10 = z2+10
```


z2[1]		18
	z2	age_10
1	18	28
2	20	30
3	20	30
4	50	60
5	22	32
6	35	45
7	24	34
8	32	42
9	24	34
10	35	45
11	20	30
12	23	33
13	19	29
14	28	38
15	40	50
16	30	40
17	40	50
18	30	40
19	48	58
20	30	40

۸. اضافه کردن لیبل و نوت در متغیرات

قبلاً بیان گردید که در ستاتا نام متغیرات نمیتواند زیاد تر از یک کلمه باشد. به طور مثال در صورت که خواسته باشید یک متغیر جدید به نام سن بسازید امکان ندارد که به مانند اکسل نام متغیر (و یا ستون مربوطه) “age of respondents” باشد، اما میتوانید که از علامه ی “_” به عوض فاصله کار بگیرید و نام متغیر جدید را “age_of_respondents” بگذارید.

```
generate age of respondents = z2
too many variables specified
r(103);
generate age_of_respondents = z2
```

اما همیشه وقت خویتر این است که در نام متغیرات از مخففیات استفاده شود. به طور مثال به عوض date of birth میتوانید به صورت مخفف dob نوشته کنید و یا به عوض highest education level میتوانید که edu و یا education نوشته کنید. در ستاتا زمان که یک متغیر جدید را ایجاد نمودید، بعداً میتوانید بالای آن لیبل و یا نوت مشخص وضع کنید تا بیان کننده آن باشد که متغیر مربوطه نشاندهنده ی کدام موضوع است. به طور مثال در دیتای سروی مردم افغانستان نام های بعضی از متغیرات عبارت از z2، z55، x4، m7، m8 و امثال آن میباشد. اما برای این که هر متغیر ذکر شده نشان دهنده ی کدام موضوع میباشد بالای آن لیبل و یا نوت های مشخص اضافه شده است که میتوانید آنرا به استفاده از فرمان های describe و یا note مشاهده نمود.

ساختار فرمان describe به شکل ذیل می باشد.

```
describe var_list
```

یعنی اولاً فرمان describe نوشته شده و به تعقیب آن نام متغیرات نوشته شود.

به طور مثال در صورت که خواسته باشید لیبل متغیرات z1، z55، x4، m7 و m8 را مشاهده کنید میتوانید از فرمان ذیل استفاده کنید.

```
describe z1 z55 x4 m7 m8
```

variable name	storage type	display format	value label	variable label
z1	byte	%8.0g	Z1	Gender
z55	byte	%8.0g	Z55	* What is the highest level (grade) of school you have completed, not including sc
x4	int	%8.0g	X4_M	* Generally speaking, do you think things in Afghanistan today are going in the ri
m7	byte	%8.0g	M7	Province
m8	int	%8.0g		Year of Interview

نوت: فرمان describe برعلاوه از لیبل ی یک متغیر (variable label) معلومات دیگر مانند لیبل قیمت ها متغیر (value label) و فارمت متغیر را نیز نشان میدهد که در درس های بعدی به تفصیل در مورد آن بیان میگردد.

با اجرای فرمان describe لیبل متغیرات در قسمت variable label نمایش داده میشود. البته اگر یک متغیر دارای لیبل و نوت در عین زمان باشد پیش روی آن علامه ی * می باشد که نوت آن را میتوانید با استفاده از فرمان note مشاهده کنید. ساختار فرمان note به شکل ذیل می باشد.

```
note var_list
```

به طور مثال در صورت که خواسته باشید نوت متغیرات z55 و x4 را مشاهده کنید میتوانید از فرمان ذیل استفاده کنید.

```
note z55 x4
```

z55:

1. What is the highest level (grade) of school you have completed, not including schooling in Islamic madrasa? (calculate the highest level into years, if none, write down zero)

x4:

1. Generally speaking, do you think things in Afghanistan today are going in the right direction, or do you think they are going in the wrong direction?

طریقه ی اضافه نمودن لیبل بالای یک متغیر به نوعی می باشد که اولاً فرمان label variable را نوشته و به تعقیب آن هر لیبل مشخص که خواسته باشید در داخل « » نوشته کنید.

به طور مثال در صورت که خواسته باشید یک متغیر جدید به نام Age ایجاد کنید که تمام مشاهدات آن خالی و یا missing باشد و بالای آن لیبل ی "age of respondents" را نصب کنید پس میتوانید از فرمان ذیل استفاده کنید.

```
generate Age = .
```

```
label variable age "age of respondents"
```

که البته برای مشاهده ی نتیجه ی آن میتوانید از فرمان describe استفاده کنیم.

همچنان ساختار فرمان note برای اضافه نمودن نوت در یک متغیر قرار ذیل می باشد.

`note varname : This is a note`

به طور مثال در صورت که خواسته باشید بالای متغیر سن نوت (All respondents are +18) را نوشته کنید میتوانید از فرمان ذیل استفاده کنید.

`note Age : All respondents are +18`

و نتیجه ی آنرا میتوانید با استفاده از فرمان `note` مشاهده کنید.

۹. استفاده از اصطلاح `in` برای قید مشاهدات

در برنامه ی ستاتا هر فرمان که انجام میدهید بالای تمام مشاهدات اجرا میشود. به طور مثال در دیتای سروی مردم افغانستان در صورت که خواسته باشید دیتای متغیرات سن `z2`، محل زیست `m6b` و جنسیت پاسخ دهنده `z1` را مشاهده کنید میتوانید از فرمان ذیل استفاده کنید.

`edit z2 m6b z1`

z1[1]		2	
	z2	m6b	z1
1	18	Urban	Female
2	20	Urban	Female
3	20	Urban	Female
4	50	Urban	Female
5	22	Urban	Female
6	35	Urban	Female
7	24	Urban	Female
8	32	Urban	Female
9	24	Urban	Female
10	35	Urban	Female
1220	50	Urban	Female
1221	30	Urban	Female
1222	45	Urban	Female
12011	22	Rural	Female
12012	22	Rural	Female
12013	36	Rural	Female
12014	31	Rural	Female
12015	29	Rural	Female
12016	18	Rural	Female
12017	53	Rural	Female
12018	33	Rural	Female
12019	67	Rural	Female

قسمی که قبلا بیان گردید فرمان فوق بالای تمام مشاهدات اجرا میشود و بخاطر که دیتای سروی مردم افغانستان بیشتر از صد هزار مشاهده را دربر دارد بنابر این تمام مشاهدات متغیرات فوق نشان داده میشود. اما در صورت که خواسته باشید فرمان فوق و یا هر فرمان دیگر را برای تعداد مشخص از مشاهدات قید کنید میتوانید از اصطلاح `in` استفاده کنید.

به طور مثال در صورت که بخواهید متغیرات سن، محل زیست و جنسیت 10 پاسخ دهنده ی اولی را مشاهده کنید میتوانید از فرمان ذیل استفاده کنید.

```
edit z2 m6b z1 in 1/10
```

z2[2]		20			
	z2	m6b	z1		
1	18	Urban	Female		
2	20	Urban	Female		
3	20	Urban	Female		
4	50	Urban	Female		
5	22	Urban	Female		
6	35	Urban	Female		
7	24	Urban	Female		
8	32	Urban	Female		
9	24	Urban	Female		
10	35	Urban	Female		

و یا در صورت که خواسته باشید دیتای متغیرات فوق را تنها در مشاهدات 112010 الی 112022 مشاهده کنید میتوانید از فرمان ذیل کار بگیرید.

```
edit z2 m6b in 112010/112022
```

z2[112010]		25			
	z2	m6b	z1		
112010	25	Urban	Female		
112011	34	Urban	Female		
112012	66	Urban	Male		
112013	30	Urban	Male		
112014	30	Urban	Male		
112015	40	Urban	Male		
112016	28	Urban	Male		
112017	25	Urban	Male		
112018	30	Urban	Female		
112019	47	Urban	Female		
112020	33	Urban	Female		
112021	60	Urban	Female		
112022	34	Urban	Female		

به همین ترتیب قسم که قبلاً بیان گردید در صورت که خواسته باشید یک متغیر جدید را که کاملاً مساوی به متغیر سن z2 باشد ایجاد کنید پس میتوانید از فرمان ذیل استفاده کنید.

`generate Age=z2`

و نتیجه آن را میتوانید از طریق فرمان `edit z2 Age` مشاهده کنیم.

z2[1]		18
	z2	Age
1	18	18
2	20	20
3	20	20
4	50	50
5	22	22
6	35	35
7	24	24
8	32	32
9	24	24
10	35	35
112012	66	66
112013	30	30
112014	30	30
112015	40	40
112016	28	28
112017	25	25
112018	30	30
112019	47	47
112020	33	33
112021	60	60
112022	34	34

قسم که دیده میشود با اجرای فرمان فوق یک متغیر جدید به نام age ایجاد گردیده که کاملاً مساوی به متغیر z2 در تمام مشاهدات میباشد، اما در صورت که بخواهید فرمان فوق تنها در 10 مشاهده ی اولی اجرا شود میتوانید از اصطلاح in استفاده کنید .

```
gen Age = z2 in 1/10
```

z2[1]		18
	z2	Age
1	18	18
2	20	20
3	20	20
4	50	50
5	22	22
6	35	35
7	24	24
8	32	32
9	24	24
10	35	35
11	20	.
12	23	.
13	19	.
14	28	.
15	40	.
16	30	.
17	40	.
18	30	.
19	48	.
20	30	.

10. استفاده از اصطلاح if برای قید کردن مشاهدات (بخش اول)

در پهلوی این که از اصطلاح in برای قید کردن مشاهدات در اجرای یک فرمان استفاده میشود میتوانید که از اصطلاح if نیز برای اینکار استفاده کنید. اما تفاوت اصطلاح in و if در کجا است؟ تفاوت عمده ی آن در این است که از اصطلاح in برای قید کردن مشاهدات به اساس مُبر مشاهدات استفاده میشود ولی از اصطلاح if برای قید کردن مشاهدات به اساس مشخصه ی یک متغیر استفاده میشود.

به طور مثال در سروی مردم افغانستان متغیر z1 نشان دهنده ی جنسیت پاسخ دهنده گان میباشد. برای اینکه بدانید به چی تعداد از افراد سروی شده زن و یا مرد میباشد میتوانید از فرمان table به شکل ذیل کار بگیرید.

table z1

Gender	Freq.
Male	59,994
Female	52,028

نوت: فرمان table در صورت که همراه به یک متغیر اجرا شود جدولی فریکونسی یک طرفه را میسازد. بعداً در مورد کاربرد آن به تفصیل بیان میگردد.

حالا در صورت که خواسته باشید تعداد مرد و زن سروی شده را در میان مشاهدات اولی الا یک هزارم دریابید میتوانید که فرمان table را به همراه اصطلاح in یکجا به شکل ذیل استفاده کنید.

```
table z1 in 1/1000
```

Gender	Freq.
Male	500
Female	500

اما در صورت که خواسته باشید بدانید به چی تعداد از مردان و زنان سروی شده بالاتر از ۵۰ سال بوده است از اصطلاح in استفاده نمیشود زیرا نمیدانیم که افراد بالاتر از ۵۰ سال در کدام مشاهدات قرار گرفته اند. برای انجام اینکار میتوانید از اصطلاح if به همراه فرمان table به شکل ذیل استفاده کنیم.

```
table z1 if z2>50
```

Gender	Freq.
Male	9,013
Female	3,905

و یا به طور مثال در صورت که خواسته باشید تعداد زنان و مردان که کمتر از ۳۰ سال عمر دارند را پیدا کنید میتوانید از فرمان ذیل استفاده کنید.

```
table z1 if z2<30
```

Gender	Freq.
Male	21,952
Female	22,087

در فرمان بالا قسمت اول آن (table z1) یک جدول فرکانسی متغیر جنسیت را میسازد و قسمت دوم آن ($z2 < 30$) if اما در صورت که خواسته باشید تعداد مردان و زنان که بالاتر از ۲۰ سال و کمتر از ۳۰ سال قرار دارند را پیدا کنید چی کار را باید انجام داد؟ برای دانستن این مورد و کاربرد عملی تر اصطلاح if باید گرداننده های عقلانی و منطقی را بدانید.

11. گرداننده های عقلانی و منطقی

• گرداننده های عقلانی

علایم نامساوات ($\neq$)، بزرگتر ($>$)، کوچکتر ($<$)، بزرگتر و مساوی ($\geq$)، کوچکتر و مساوی ($\leq$) و مساوی ($=$) را در برنامه استاتاستیک گرداننده های عقلانی میگویند. از گرداننده های عقلانی فوق میتوانید همراه اصطلاح if برای قید کردن مشاهدات و یا به زبان ساده مشروط کردن فرمان استفاده میشود. به طور مثال در صورت که $z2 < 30$ نوشته کنید به این معنی است که متغیر $z2$ کوچکتر

از 30 باشد (یعنی تمام افراد که کمتر از 30 سال سن دارند) به همین ترتیب در صورت که نوشته کنید $m8==2018$ به این معنی که متغیر $m8$ مساوی به 2018 باشد (یعنی تمام افراد که در سال 2018 سروی شده اند).

شاید متوجه شده باشید که این علایم در ستاتا کاملاً مشابه علایم ریاضیکی نیست بلکه تفاوت های اندکی دارند. برای دانستن تفاوت های آن به جدول ذیل توجه کنید.

ریاضی	برنامه ستاتا	
>	>	بزرگتر از
<	<	کوچکتر از
=	==	مساوی به
≥	>=	بزرگتر و مساوی به
≤	<=	کوچکتر و مساوی به
≠	!=	مساوی نیست به

برای درک بهتر گرداننده های عقلانی همراه با اصطلاح if به مثال های ذیل توجه کنید.
به طور مثال در صورت که خواسته باشید تعداد مردان و زنان را در سال 2018 دریابید.

```
table z1 if m8==2018
```

Gender	Freq.
Male	7,550
Female	7,462

قسمت اول فرمان فوق (table z1) جدول فرکونس متغیر جنسیت را میسازد و قسمت دوم آن (if $m8==2018$) آنرا محدود به همان اشخاص می‌کند که در سال 2018 سروی شده اند.
به همین ترتیب در صورت که خواسته باشید تعداد افراد شهری و دهاتی را در سال 2018 دریابید میتوانید از فرمان ذیل استفاده کنید.

```
table m6b if m8==2018
```

CS0 Geographi c Code	Freq.
Rural	12,105
Urban	2,907

• گرداننده های منطقی

توسط گرداننده های عقلانی می‌توانید یک فرمان را به مشاهدات مخصوص قید کنید. اما شاید پیش شما سوال خلق شود که چگونه یک فرمان را با بیشتر از یک شرط قید سازید. به طور مثال در صورت که خواسته باشید تعداد مردان زنان که در سن پایینتر از 25 ساله را پیدا کنید می‌توانید از نوشته کنید که (table z1 if z2<25) اما برای اینکه تعداد مردان و زنان که پایینتر از 25 سال اند و در سال 2018 سروی شده اند را پیدا کنید چی باید کرد؟ در این موارد باید که فرمان فوق همراه با اصطلاح if با دو شرطیه قید گردد.

برای اینکه همراه با یک فرمان بیشتر از یک شرطیه استفاده شود می‌توانید از گرداننده های منطقی & و | استفاده کنید که هر دو همراه با اصطلاح if استفاده میشوند و در واقع مکمل کننده ی گرداننده های عقلانی اند. علامه & که (and) خوانده میشود و علامه ی | که (or) خوانده میشود هر دو همراه با اصطلاح if یکجا برای استفاده از دو شرطیه در یک فرمان به کار میرود. اما کاربرد آن ها کاملاً متفاوت میباشد.

زمان که دو شرطیه توسط علامه & یکجا شود پس فرمان را محدود به همان مشاهدات میکند که باید هر دو شرطیه صدق کند (یعنی مشاهدات را که یک شرطیه در آن صدق کند اما شرطیه ی دوم در آن صدق نکند شامل نمیشود). یعنی فرمان برای تمام مشاهدات انجام میشود که هر دو شرطیه صدق کند. به طور مثال در صورت که نوشته کنید if z2>20 & z2<25 پس فرمان برای تمام مشاهداتی اجرا میشود که در متغیر z2 بزرگتر از 20 باشد (افراد که بالاتر از 20 ساله اند) و همچنان در متغیر z2 کوچکتر از 25 باشد (افراد که خردتر از 25 ساله اند).

tab z2 if z2>20 & z2<25

How old are you?	Freq.	Percent	Cum.
21	2,759	23.52	23.52
22	3,696	31.51	55.03
23	2,661	22.69	77.72
24	2,613	22.28	100.00
Total	11,729	100.00	

زمان که دو شرطیه توسط علامه | یکجا شود پس فرمان محدود به همان مشاهدات میشود که حد اقل یکی از شرطیه ها در آن صدق کند. به طور مثال در صورت که نوشته کنید if z2<25 | z2>80 پس تنها همان مشاهدات را در نظر میگیرد که در متغیر z2 بالاتر پایینتر از 25 باشد (خردتر از 25 ساله باشند) یا در متغیر z2 بزرگتر از 80 باشد (افراد که بالاتر از 80 ساله اند).

```
tab z2 if z2<25 | z2>80
```

How old are you?	Freq.	Percent	Cum.
18	6,223	22.14	22.14
19	4,081	14.52	36.67
20	6,000	21.35	58.02
21	2,759	9.82	67.84
22	3,696	13.15	80.99
23	2,661	9.47	90.46
24	2,613	9.30	99.75
81	12	0.04	99.80
82	15	0.05	99.85
83	12	0.04	99.89
84	3	0.01	99.90
85	19	0.07	99.97
87	1	0.00	99.98
88	1	0.00	99.98
90	2	0.01	99.99
91	1	0.00	99.99
92	1	0.00	99.99
96	1	0.00	100.00
97	1	0.00	100.00
Total	28,102	100.00	

با استفاده از گرداننده های منطقی می‌توانید بیشتر از چند شرطیه را یکجا ساز. حتی می‌توانید هر دو گرداننده ی & و | را یکجای استفاده کنید.

به طور مثال در صورت که خواسته باشید تعداد مردان و زنان که در سال 2018 سروی شده اند (m8==2018) و بالاتر از 20 سال و پایینتر از 25 سال اند (z2>20 & z2<25) را پیدا کنید می‌توانید از فرمان ذیل استفاده کنید.

```
tab z1 if m8==2018 & z2>20 & z2<25
```

Gender	Freq.	Percent	Cum.
Male	713	46.91	46.91
Female	807	53.09	100.00
Total	1,520	100.00	

و همچنان در صورت که خواسته باشید تعداد مردان و زنان را که در سال 2018 سروی شده اند (m8==2018) و بالاتر پایینتر از 25 سال یا بالاتر از 80 سال اند (z2<25 | z2>80) را دریابید پس می‌توانید از فرمان ذیل استفاده کنید.

```
tab z1 if m8==2018 & (z2<25 | z2>80)
```

Gender	Freq.	Percent	Cum.
Male	1,539	43.90	43.90
Female	1,967	56.10	100.00
Total	3,506	100.00	

12. ایجاد تغییر در یک متغیر

برای ایجاد تغییر در یک متغیر می‌توانید از فرمان `replace` کار بگیرید. ساختار فرمان `replace` برای تغییر در متغیرات عددی قرار ذیل می‌باشد.

```
replace var_name=value
```

به طور مثال در صورتی که خواسته باشید در دیتای سروی مردم افغانستان تمام مشاهدات را در متغیر `z2` به 0 تبدیل کنید پس می‌توانید از فرمان `replace` به شکل ذیل استفاده کنید.

```
replace z2=0
```

z2[1]		0
	z2	
1	0	
2	0	
3	0	
4	0	
5	0	
6	0	
7	0	
8	0	
9	0	
10	0	
11	0	
12	0	
13	0	
14	0	
15	0	
16	0	
17	0	
18	0	
19	0	
20	0	

و در صورت که خواسته باشید تمام مشاهدات را به عدد 1 تبدیل کنید پس میتوانید از فرمان ذیل استفاده کنید.

`replace z2=1`

z2[1]		1
	z2	
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	1	
10	1	
11	1	
12	1	
13	1	
14	1	
15	1	
16	1	
17	1	
18	1	
19	1	
20	1	

طور که قبلاً بیان گردید هر فرمان که در ستاتا انجام دهید بالای تمام مشاهدات اجرا میشود. بنابر این زمان که با استفاده از فرمان `replace` متغیر سن را به 0 تبدیل کنید تمام مشاهدات در متغیر سن تبدیل به عدد 0 میشود و به همین صورت زمان که با استفاده از فرمان فوق متغیر سن را به 1 تبدیل کنید تمام مشاهدات در این متغیر تبدیل به عدد 1 میشود. اما میتوانید اکثریت از فرمان های ستاتا را همراه با اصطلاح `if` (و یا هم اصطلاح `in`) استفاده کنید تا فرمان اجرا شده تنها در بعضی از مشاهدات خاص اجرا شود.

به طور مثال در صورت که خواسته باشید متغیر سن `z2` را به نحوی تغییر بدهید که تمام افراد سروی شده که دارای سن بالاتر از ۳۵ سال باشد به عدد 0 نمایش داده شود و تمام افراد که ۳۵ ساله و سن پایینتر از ۳۵ سال دارند به عدد 1 نمایش داده شود میتوانید از فرمان `replace` و اصطلاح `if` به شکل ذیل استفاده کنید.

```
replace z2=1 if z2<=35
replace z2 = 0 if z2>35
```

z2[1]		1
	z2	
1	1	
2	1	
3	1	
4	0	
5	1	
6	1	
7	1	
8	1	
9	1	
10	1	
11	1	
12	1	
13	1	
14	1	
15	0	
16	1	
17	0	
18	1	
19	0	
20	1	

به این ترتیب تمام ارزش های متغیر z2 به دو عدد 0 و 1 تبدیل شده است که عدد 1 نمایندگی از افراد میکند که 35 ساله و پائینتر از آن هستند و عدد 0 نمایندگی از افراد میکند که بالاتر از 35 سال هستند.

13. ایجاد یک متغیر جدید (قسمت دوم)

قبلاً خواندید که با استفاده از فرمان generate میتوانید یک متغیر جدید بسازید. قسم که در درس هشتم بیان گردید همراه با فرمان generate میتوانید یک متغیر جدید ایجاد کنید که مساوی به یک متغیر دیگر باشد. به طور مثال در سروی مردم افغانستان متغیر m8 نشان دهنده ی سال های که سروی انجام شده میباشد. در صورت که بخواهید یک متغیر جدید به نام year ایجاد کنید که کاملاً مساوی به این متغیر باشد میتوانید از فرمان generate به شکل ذیل استفاده کنیم.

`generate year = m8`

m8[1]		2006		
	m8	year		
1	2006	2006		
2	2006	2006		
3	2006	2006		
4	2006	2006		
5	2006	2006		
6	2006	2006		
7	2006	2006		
8	2006	2006		
9	2006	2006		
10	2006	2006		
11	2006	2006		
12	2006	2006		
13	2006	2006		
14	2006	2006		
15	2006	2006		
16	2006	2006		
17	2006	2006		
18	2006	2006		
19	2006	2006		
20	2006	2006		

همچنان در سروی مردم افغانستان از افراد سروی شده در مورد نظریات شان در باره ی آینده ی کشور پرسان شده است که آیا کشور افغانستان به سمت درست و یا سمت اشتباه در حرکت است که این معلومات در متغیر x4 ذخیره گردیده است که شما میتوانید آنرا با اجرای فرمان ذیل مشاهده کنید.

edit x4

x4[4]		102	
	x4		
1	right direction		
2	wrong direction		
3	right direction		
4	wrong direction		
5	wrong direction		
6	wrong direction		
7	right direction		
8	wrong direction		
9	right direction		
10	right direction		
11	some in right,...		
12	right direction		
13	right direction		
14	wrong direction		
15	wrong direction		
16	right direction		
17	right direction		
18	right direction		
19	right direction		
20	some in right,...		

x4[3]		101	
	x4		
1	right direction		
2	wrong direction		
3	right direction		
4	wrong direction		
5	wrong direction		
6	wrong direction		
7	right direction		
8	wrong direction		
9	right direction		
10	right direction		
11	some in right,...		
12	right direction		
13	right direction		
14	wrong direction		
15	wrong direction		
16	right direction		
17	right direction		
18	right direction		
19	right direction		

قسم که مشاهده میشود متغیر x4 به رنگ آبی نمایش داده میشود که بیان کننده ی این موضوع است که این متغیر در اصل به شکل عددی ذخیره است و اما تفاوت آن در این است که بالای هر عدد یک لیبل مشخص وضع گردیده است. به طور مثال بالای عدد 101 لیبل right direction و بالای عدد 102 لیبل wrong direction وضع گردیده است. و زبان ساده متغیر x4 عبارت از متغیر لیبل دار میباشد.

نوت: در مورد اینکه چگونه لیبل های مکمل یک متغیر را مشاهده کنید و اینکه چگونه یک لیبل مشخص را بالای اعداد یک متغیر نصب کنیم به تفصیل در درس های بعدی بیان میگردد.

در صورت که خواسته باشید یک متغیر جدید به نام direction که مساوی به متغیر x4 باشد ایجاد کنید میتوانید از فرمان generate به شکل ذیل استفاده کنید.

generate direction = x4

و نتیجه ی آن را با اجرای فرمان edit x4 direction مشاهده نموده میتوانید.

x4[10]		101		
	x4	direction		
1	right direction	101		
2	wrong direction	102		
3	right direction	101		
4	wrong direction	102		
5	wrong direction	102		
6	wrong direction	102		
7	right direction	101		
8	wrong direction	102		
9	right direction	101		
10	right direction	101		
11	some in right,...	103		
12	right direction	101		
13	right direction	101		
14	wrong direction	102		
15	wrong direction	102		
16	right direction	101		
17	right direction	101		
18	right direction	101		
19	right direction	101		

قرار که دیده میشود با اجرای فرمان فوق یک متغیر جدید به نام direction ایجاد گردیده است که مساوی به متغیر x4 میباشد. اما تفاوت در این جا است که در متغیر جدید بالای هر عدد لیبل های مذکور و یا value labels وضع نگردیده است. راه حل چی است؟ حد اقل دو راه حل وجود دارد. اول اینکه بالای متغیر جدید به صورت جدا گانه لیبل های اعداد و یا value labels را اضافه کنید و دوم اینکه برای ایجاد متغیر جدید به عوض فرمان generate از فرمان clonevar کار بگیرید. ساختار فرمان clonevar به شکل ذیل میباشد.

```
clonevar new_var=old_var
```

فرمان clonevar به مانند فرمان generate یک متغیر جدید ایجاد میکند اما تفاوت اینجا است که متغیر جدید مشخصات متغیر قبلی از جمله لیبل اعداد و یا value labels را نیز همراه با خود میداشته باشد.

به طور مثال اینبار یک متغیر جدید به نام dir از طریق فرمان clonevar ایجاد کنید.

```
clonevar dir = x4
```

و هر سه متغیر x4, متغیر که از طریق فرمان generate ایجاد کردید به نام direction و متغیر که از طریق فرمان clonevar به نام dir ایجاد شده است را با استفاده از فرمان ذیل مشاهده کرده میتوانید.

```
edit x4 direction dir
```

dir[3]		101	
	x4	direction	dir
1	right direction	101	right direction
2	wrong direction	102	wrong direction
3	right direction	101	right direction
4	wrong direction	102	wrong direction
5	wrong direction	102	wrong direction
6	wrong direction	102	wrong direction
7	right direction	101	right direction
8	wrong direction	102	wrong direction
9	right direction	101	right direction
10	right direction	101	right direction
11	some in right,...	103	some in right,...
12	right direction	101	right direction
13	right direction	101	right direction
14	wrong direction	102	wrong direction
15	wrong direction	102	wrong direction
16	right direction	101	right direction
17	right direction	101	right direction
18	right direction	101	right direction
19	right direction	101	right direction
20	some in right,...	103	some in right,...

14. استفاده از اصطلاح if برای قید کردن مشاهدات (بخش دوم)

قبلاً خواندید هر فرمان که انجام دهید بالای تمام مشاهدات اجرا میشود. برای اینکه یک فرمان را بالای یک تعداد مشخص از مشاهدات انجام دهید میتوانید از اصطلاح in کار بگیرید. به طور مثال در صورت که خواسته باشید جدول فرکونسی متغیر جنسیت را تنها در 100 مشاهده ی اولی به دست بیاورید میتوانید از فرمان ذیل استفاده کنید.

```
table z1 in 1/100
```

Gender	Freq.
Male	49
Female	51

اما زمان که خواسته باشید جدول فرکونسی متغیر جنسیت را تنها برای کسان بدست بیاورید که کوچکتر از 30 ساله باشند در این صورت باید از اصطلاح if استفاده شود.

```
table z1 if z2<30
```

Gender	Freq.
Male	2,584
Female	3,175

به همین ترتیب همراه با اصطلاح if میتوانید بیشتر از یک شرطیه بگذارید. به طور مثال در صورت که خواسته باشید جدول فرکونسی متغیر جنسیت را برای تمام کسان که کوچکتر از 30 ساله هستند (شرطیه اول) و یا افراد که بالاتر از 50 ساله هستند (شرطیه دوم) بدست بیاورید میتوانید از فرمان ذیل استفاده کنید.

```
table z1 if z2<30 | z2>50
```

Gender	Freq.
Male	3,877
Female	3,732

به همت ترتیب در صورت که خواسته باشید یک فرمان را برای مشاهدات خاص انجام دهید میتوانید که از اصطلاح if استفاده کنید. در صورت که شرطیه به اساس یک متغیر عددی (مانند متغیر سن در مثال فوق) تعیین شده باشد پس استفاده از اصطلاح if بسیار ساده میباشد.

اما در صورت که شرطیه به اساس یک متغیر لیبل دار (مانند متغیر x4 که در بر گیرنده ی نظریات مردم افغانستان در رابطه به آینده ی این کشور است) تعیین شده باشد پس چی کار باید کرد؟
یک بار متغیر x4 را مشاهده کنید دیده میشود که به رنگ آبی نمایش داده شده است.

edit x4

x4[3]		101
x4		
1	right direction	
2	wrong direction	
3	right direction	
4	wrong direction	
5	wrong direction	
6	wrong direction	
7	right direction	
8	wrong direction	
9	right direction	
10	right direction	
11	some in right, some...	
12	right direction	
13	right direction	
14	wrong direction	
15	wrong direction	
16	right direction	
17	right direction	
18	right direction	
19	right direction	
20	some in right, some...	

قسم که دیده میشود متغیر x4 در اصل یک متغیر عددی میباشد اما بالای هر عدد یک لیبل مشخص وضع است. به طور مثال بالای عدد 101 لیبل right direction وضع است. حالا در صورت که خواسته باشید یک فرمان خاص تنها بالای کسان که گفته اند کشور به سمت درست در حرکت است اجرا شود پس همان فرمان همراه با اصطلاح if به اساس عدد ذکر شده، نه لیبل آن، باید اجرا شود.

به طور مثال در صورت که خواسته باشید جدول فرکونسی متغیر جنسیت را برای تمام کسان که در متغیر x4 گزینه ی right direction را انتخاب کرده ان بدست بیاورید پس میتوانید از فرمان ذیل استفاده کنید.

```
table z1 if x4==101
```

Gender	Freq.
Male	25,096
Female	21,316

به همین ترتیب اگر خواسته باشید جدول فرکونسی متغیر جنسیت تنها برای کسان که 30 ساله و خردتر از آن باشد و در متغیر x4 گزینه ی right direction را انتخاب کرده باشد بدست بیاورید پس میتوانید از فرمان ذیل استفاده کنید.

```
table z1 if x4==101 & z2<=30
```

Gender	Freq.
Male	10,836
Female	10,942

15. تبدیل کردن یک متغیر پیوسته به متغیر کتگوری شده (بخش اول)

در درس **هفتم** در مورد اینکه چگونه با استفاده از فرمان generate یک متغیر جدید ایجاد کنید و در درس **دوازدهم** روی این موضوع که چطور در یک متغیر از قبل ایجاد شده توسط فرمان replace تغییرات وضع کنید بحث گردید. در این بخش در مورد اینکه چگونه از فرمان های generate و replace به صورت همزمان تغییر یک متغیر پیوسته به متغیر کتگوری شده استفاده کنید بیان میشود.

نوت: برای یاد آوری مجدد توصیه میشود که دروس **هفتم** و **دوازدهم** یک بار دیگر مرور گردد.

به طور مثال در سروی مردم افغانستان سن افراد سروی شده در متغیر z2 ذخیره شده است که به شکل پیوسته میباشد. در صورت که خواسته باشید یک متغیر جدید به نام Age ایجاد کنید که دارای 3 کتگوری به شکل ذیل باشد میتوانید از فرمان های generate و replace استفاده کنید.

- 1 = افراد که دارای سن 35 سال و پاینتر از آن اند
- 2 = افراد که دارای سن 36 سال الی 55 سال اند
- 3 = افراد که دارای سن بالاتر از 55 سال اند

در قدم نخست یک متغیر جدید توسط فرمان generate به نام Age ایجاد کرده که تمام آن خالی و یا missing باشد.
`generate Age = .`

z2[1]		18
	z2	Age
1	18	.
2	20	.
3	20	.
4	50	.
5	22	.
6	35	.
7	24	.
8	32	.
9	24	.
10	35	.
11	20	.
12	23	.
13	19	.
14	28	.
15	40	.
16	30	.
17	40	.
18	30	.
19	48	.
20	30	.

بعداً توسط فرمان replace متغیر Age را به 1 تبدیل کنید در صورت که ارزش های متغیر z2 کمتر از 35 و یا مساوی به 35 باشد.

`replace Age = 1 if z2 <= 35`

z2[1]		18
	z2	Age
1	18	1
2	20	1
3	20	1
4	50	.
5	22	1
6	35	1
7	24	1
8	32	1
9	24	1
10	35	1
11	20	1
12	23	1
13	19	1
14	28	1
15	40	.
16	30	1
17	40	.
18	30	1
19	48	.
20	30	1

در قدم بعدی باز هم توسط فرمان replace متغیر Age را به 2 تبدیل کنید در صورت ارزش های متغیر z2 بیشتر از 35 و کمتر از 56 باشد.

در این بار استفاده از اصطلاح if دو شرطیه استفاده شده است. اول اینکه متغیر z2 بزرگتر از 35 باشد و دوم اینکه متغیر z2 کوچکتر از 56 باشد. بنابر این برای یکجا ساختن این دو شرطیه از گرداننده های عقلانی باید استفاده شود.

`replace Age=2 if z2>35 & z2<56`

20	30	1
21	18	1
22	28	1
23	48	2
24	45	2
25	19	1
26	35	1
27	20	1
28	30	1
29	28	1
30	28	1
31	18	1
32	64	.
33	30	1
34	31	1
35	21	1
36	33	1
37	27	1
38	42	2
39	20	1
40	32	1

و در قدم اخیر متغیر age را به 3 تبدیل کنید در صورت که ارزش های متغیر z2 بزرگتر از 55 باشد.

```
replace Age=3 if z2>55
```

40	32	1
41	60	3
42	32	1
43	55	2
44	60	3
45	36	2
46	29	1
47	40	2
48	30	1
49	44	2
50	44	2
51	58	3
52	27	1
53	74	3
54	34	1
55	24	1
56	55	2
57	32	1
58	42	2
59	24	1
60	28	1

به این صورت متغیر Age را ایجاد گردید که شکل کتگوری شده ی متغیر سن z2 میباشد. و نتیجه ی آن را میتوانید با نوشتن فرمان `edit z2 Age` مشاهده کنید و یا هم با استفاده از فرمان `table` میتوانید که نتیجه ی آنرا به در یک جدول مشاهده کنیم. `table Age`

Age	Freq.
1	64,706
2	39,026
3	8,290

16. مشاهده نمودن لیبل های ارزش های یک متغیر.

بسیاری از متغیرات که در سروی مردم افغانستان وجود دارد به نحوه ی است که نوع ذخیره آن به شکل عددی میباشد اما بالای هر عدد آن یک لیبل وضع گردیده است و یا به زبان ساده متغیر لیبل دار است. به طور مثال متغیر x4 که نشان دهنده ی نظریات افراد سروی شده در مورد آینده ی کشور و متغیر z1 که نشان دهنده ی جنسیت افراد سروی شده میباشد عبارت از متغیر های لیبل دار میباشند.

به همین ترتیب در صورت که `data editor` را باز کنید تمام متغیر های که به رنگ آبی نمایش داده میشوند عبارت از متغیر های لیبل دار است.

x4[4]		102		
	x4	z1		
1	right direction	Female		
2	wrong direction	Female		
3	right direction	Female		
4	wrong direction	Female		
5	wrong direction	Female		
6	wrong direction	Female		
7	right direction	Female		
8	wrong direction	Female		
9	right direction	Female		
10	right direction	Female		
11	some in right, some...	Male		
12	right direction	Male		
13	right direction	Male		
14	wrong direction	Male		
15	wrong direction	Male		
16	right direction	Male		
17	right direction	Male		
18	right direction	Male		
19	right direction	Male		
20	some in right, some...	Male		

برای اینکه بدانید کدام لیبل بالای یک متغیر خاص نصب گردیده، در بخش فرمان ها describe نوشته و به تعقیب آن نام متغیر متذکره را نوشته کنید. به طور مثال در دیتای سروی مردم افغانستان در صورت که خواسته باشید بدانید که بالای متغیر های x4 و z1 کدام لیبل ها نصب گردیده میتوانیم از فرمان ذیل استفاده کنید.

describe x4 z1

```
. describe x4 z1
```

variable name	storage type	display format	value label	variable label
x4	int	%20.0g	X4_M	* Generally speaking, do you think things in Afghanistan today are going in the ri
z1	byte	%8.0g	Z1	Gender

بعد از اجرای فرمان فوق در قسمت value label لیبل های مشخص هر متغیر نمایش داده میشود و در صورت که بالای یک متغیر هیچ لیبل وضع نباشد پس این قسمت خالی میباشد.

از نتیجه ی فرمان فوق واضح است که بالای متغیر x4 لیبل X4_M و بالا ی متغیر z1 لیبل Z1 وضع است. در قدم بعدی برای دانستن لیبل های مشخص هر متغیر میتوانید از فرمان label list استفاده کنید ساختار فرمان label list به شک ذیل میباشد

label list value_label

یعنی اولاً فرمان label list را نوشته و به تعقیب آن نام لیبل باید نوشته شود.

به طور مثال در سروی مردم افغانستان در صورت که خواسته باشید لیبل های متغیر x4 را مشاهده کنید پس میتوانید از فرمان ذیل استفاده کنیم.

```
label list X4_M
X4_M:
101 right direction
102 wrong direction
103 some in right, some in wrong direction
998 refused
999 don't know
```

قسم که دیده میشود متغیر x4 دارای 5 کتگوری میباشد و بالای هر کتگوری آن یک لیبل مشخص وضع گردیده است. به همین ترتیب با اجرای فرمان label list Z1 میتوانید لیبل های مشخص متغیر جنسیت را بدانید.

```
label list Z1
Z1:
1 Male
2 Female
```

طریقه ی دیگر مشاهده کردن لیبل هر متغیر استفاده از فرمان labellist است و ساختار فرمان labellist به شکل ذیل میباشد.

```
labellist var_list
```

یعنی در صورت که خواسته باشید لیبل های مشخص یک و یا چندین متغیر را بدانید اولاً labellist نوشته و به تعقیب آن نام متغیرات مشخص را نوشته کنید. به طور مثال در صورت که خواسته باشید لیبل های متغیر x4 و z1 را مشاهده کنید میتوانید از فرمان ذیل استفاده کنید.

```
labellist x4 z1
X4_M:
101 right direction
102 wrong direction
103 some in right, some in wrong direction
998 refused
999 don't know
Z1:
1 Male
2 Female
```

قسم که دیده میشود فرمان labellist نه تنها لیبل های هر متغیر را بلکه نام لیبل هر متغیر را نیز نشان میدهد.

نوت: از آنجای که فرمان labellist همراه با خود ستاتا نصب نمیشد باید اولاً آنرا در ستاتا نصب و یا install کنید. طریقه ی انستال کردن یک فرمان در ستاتا به شکل ذیل میباشد.

```
ssc install command_name
```

یعنی برای انستال کردن یک فرمان اولاً ssc install نوشته و به تعقیب آن نام فرمان مشخص را نوشته کنید. به طور مثال فرمان labellist را میتوانید که به شکل ذیل نصب و یا انستال کنید.

```
ssc install labellist
```

```
. ssc install labellist
```

```
checking labellist consistency and verifying not already installed...
```

```
installing into /Users/fahimyousufzai/Library/Application Support/Stata/ado/plus/...
```

```
installation complete.
```

17. اضافه کردن لیبل بالای ارزش های یک متغیر

قسم که در درس قبلی مشاهده نمودید ستاتا شما را قادر میسازد تا بالای متغیرات که به شکل عددی ذخیره گردیده است لیبل های مشخص وضع کنیم.

البته برای این کار دو مرحله نیاز است:

a. اولاً ساختن یک لیبل مشخص

b. دوم نصب کردن آن بالای یک متغیر مشخص

به طور مثال در دیتای ذیل متغیر gender_1 را در نظر بگیرید که دارای دو ارزش 0 و 1 میباشد.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_status	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

Example.dta (<https://github.com/Fahim-Ahmad/Stata-programming>)

در صورت که خواسته باشید در متغیر فوق بالای عدد یک لیبل male و بالای عدد 0 لیبل female نوشته شود اولاً باید یک لیبل مشخص ایجاد کنید بعداً لیبل ایجاد شده را بالای متغیر فوق نصب کنید.

با استفاده از فرمان label define میتونید یک لیبل مشخص بسازید. البته ساختار فرمان label define به شکل ذیل میباشد.

```
label define label_name value_1"label_1" value_2"label_2" ... value_n"label_n"
```

یعنی در صورت که خواسته باشید یک لیبل، به طور مثال به نام sex بسازید که دارای دو کتگوری 0 و 1 باشد که به ترتیب نشان دهنده female و male باشد میتوانید از فرمان label define به شکل ذیل استفاده کنید.

```
label define sex 1"Male" 0"Female"
```

تا اینجا کار یک لیبل به نام sex ساخته شده است اما هنوز بالای کدام متغیر خاص نصب نگردیده است. برای اینکه یک لیبل را بالای یک متغیر خاص وضع کنید میتوانید از فرمان label value استفاده کنید که ساختار آن به شکل ذیل میباشد.

```
label value var_name label_name
```

یعنی در قدم اول فرمان label value نوشته به تعقیب آن نام متغیر و در ادامه نام لیبل را نوشته کنید. به طور مثال در دیتای که قبلاً نمایش داده شد نام متغیر gender_1 میباشد و نام لیبل که ساختیم sex میباشد. حالا برای نصب این لیبل بالای متغیر ذکر شده میتوانید از فرمان label value به شکل ذیل استفاده کنید.

```
label value gender_1 sex
```

و نتیجه آن را با اجرای فرمان edit gender_1 مشاهده کنیم.

gender_1[4]		
	gender_1	
1	Male	
2	Male	
3	Male	
4	Female	
5	Male	
6	Male	
7	Male	
8	Male	
9	Female	
10	Male	
11	Male	
12	Female	
13	Female	
14	Male	
15	Female	
16	Male	
17	Male	
18	Female	
19	Female	
20	Female	

نوت: همان طور که دو متغیر به یک نام بوده نمیتواند امکان ندارد که دو لیبل نیز به یک نام باشد.

18. تبدیل نمودن متغیر های غیر عددی به متغیر عددی

در درس های قبلی خواندید که چگونه یک متغیر که به شکل عددی باشد بالای آن لیبل های مشخص وضع کنید، اما امکان دارد که بعضی اوقات متغیرات به شکل غیر عددی جمع آوری شده میباشد.

به طور مثال جنست افراد را میتوانیم به شکل عدد 1 و 2 ذخیره کرده و دوباره بالای آن لیبل های مشخص را وضع کنید و یا هم میتوانیم آنرا از آغاز به شکل male و female ذخیره کنید که در این صورت این متغیر به شکل غیر عددی و یا string ذخیره میشود.

به طور مثال دیتای ذیل را در نظر بگیرید که در آن متغیر gender_2 نشان دهنده ی جنسیت افراد میباشد و به شکل غیر عددی ذخیره شده است.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_status	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

از آنجای که متغیر gender_2 یک متغیر غیر عددی و یا string میباشد بنابر آن به رنگ سرخ نشان داده میشود. در صورت که خواسته باشید این متغیر را به متغیری عددی تبدیل کنید میتوانید که از فرمان encode استفاده کنید. ساختار فرمان encode به شکل ذیل میباشد.

```
encode old_var, generate (new_var)
```

یعنی برای اینکه یک متغیر را از شکل غیر عددی به متغیر عددی تبدیل کنید اولاً فرمان encode را نوشته کرده و به تعقیب آن نام متغیر که به شکل غیر عددی است را نوشته کنید و بعد از کامه generate نوشته و داخل قوس نام متغیر جدید را نوشته کنید.

به طور مثال در صورت که در دیتای قبلی متغیر gender_2 را که به شکل غیر عددی است بخواهید به متغیر عددی تبدیل کنید و نام آنرا GENDER بگذارید، میتوانید از فرمان encode به شکل ذیل استفاده کنید.

```
encode gender_2, generate (GENDER)
```

و نتیجه ی آنرا با اجرای فرمان edit gender_2 GENDER میتوانید مشاهده کنید.

GENDER[4]			1
	gender_2	GENDER	
1	male	male	
2	male	male	
3	male	male	
4	female	female	
5	male	male	
6	male	male	
7	male	male	
8	male	male	
9	female	female	
10	male	male	
11	male	male	
12	female	female	
13	female	female	
14	male	male	
15	female	female	
16	male	male	
17	male	male	
18	female	female	
19	female	female	
20	female	female	

19. تبدیل کردن یک متغیر پیوسته به متغیر کتگوری شده (بخش دوم)

در دروس قبلی خواندید که چگونه از فرمان generate برای ایجاد یک متغیر جدید استفاده کنید (درس هفتم)، چطور در یک متغیر با استفاده از فرمان replace تغییرات وضع کنید (درس دوازدهم) و اینکه چگونه از اصطلاح if استفاده کنید (درس دهم و چهاردهم). همچنان در درس پانزدهم روی این موضوع که چطور با استفاده از فرمان های نام برده یک متغیر پیوسته را به متغیر کتگوری شده تبدیل کنیم بحث شد.

نوت: برای یاد آوری مجدد توصیه میشود تا دروس هفتم، دهم، دوازدهم و پانزدهم را مرور کنید.

یکی از طریقه های که میتوانید در یک متغیر عددی تغییرات وضع کنید عبارت از فرمان recode میباشد. به طور مثال دیتای ذیل را در نظر بگیرید.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_st-s	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

در دیتای فوق متغیرات age، gender_1 و income عبارت از متغیرات عددی میباشد که میتوانید برای ایجاد تغییرات در این متغیرات از فرمان recode استفاده کنید. به طور مثال در صورت که خواسته باشید در متغیر gender_1 تمام مشاهدات که عدد 0 است به عدد 2 تبدیل کنید میتوانید از فرمان ذیل استفاده کنید.

recode gender_1 (0=2)

و نتیجه ی آنرا میتوانید با استفاده از فرمان edit gender_1 مشاهده کنید.

gender_1[24]	
	gender_1
1	1
2	1
3	1
4	2
5	1
6	1
7	1
8	1
9	2
10	1
11	1
12	2
13	2
14	1
15	2
16	1
17	1
18	2
19	2
20	2

دیده میشود که در متغیر فوق تمام مشاهدات که عدد 0 داشته به عدد 2 تبدیل شده است. به همین ترتیب با استفاده از فرمان recode نه تنها یک عدد را بلکه میتوانید چندین عدد را به صورت همزمان تغییر بدهید. به طو مثال به متغیر age در دیتای فوق توجه کنید، در صورت که خواسته باشید تمام افراد که 18 الی 25 ساله اند به 1 تبدیل گردد و کسان که 26 الی 33 ساله اند به عدد 2 تبدیل شود و مشاهدات که خالی اند به حالت خودش باقی بماند میتوانید از فرمان ذیل استفاده کنید.

`recode age (18/25 = 1) (26/33 = 2) (. = .)`

و نتیجه ی آن را میتوانید با استفاده از فرمان edit age ببینید.

age[24]	
	age
1	2
2	1
3	2
4	2
5	1
6	1
7	2
8	.
9	2
10	2
11	1
12	1
13	1
14	2
15	2
16	.
17	2
18	1
19	1
20	2

به همین ترتیب با استفاده از فرمان فوق میتوانید یک متغیر را به چندین کتگوری تقسیم بندی کنید. به طور مثال در دیتای فوق به متغیر income توجه کنید که کمترین عدد در آن 700 و بزرگترین عدد 2000 میباشد. در صورت که خواسته باشید متغیر فوق را به 3 کتگوری به نحوی تبدیل کنید که کتگوری اول از 700 الی 1000 باشد، کتگوری دوم از 1001 الی 1500 باشد و کتگوری سوم از 1501 الی 2000 باشد و مشاهدات که دارای ارزش خالی (missing values) هستند بدون کدام تغییر باشند پس میتوانید از فرمان ذیل استفاده کنید.

`recode income (700/1000 = 1) (1001/1500 = 2) (1501/2000 = 3) (. = .)`

و نتیجه ی آن را میتوانید با اجرای فرمان edit income ببینید.

income[24]		
	income	
1	2	
2	2	
3	3	
4	2	
5	.	
6	1	
7	2	
8	2	
9	3	
10	3	
11	1	
12	2	
13	.	
14	3	
15	3	
16	3	
17	3	
18	2	
19	.	
20	3	

به همین ترتیب در سروی مردم افغانستان متغیر سن z2 را میتوانید با استفاده از فرمان recode به کتگوری های مختلف تقسیم بندی کنید.

متغیر سن در سروی مردم افغانستان یک متغیر عددی میباشد که خورده ترین ارزش در آن 18 و بزرگترین ارزش در آن 97 میباشد. در صورت که خواسته باشید این متغیر را به 4 کتگوری ذیل تقسیم کنید.

- 1 = افراد که میان 18 الی 25 سال اند
- 2 = افراد که میان 26 الی 35 سال اند
- 3 = افراد که میان 36 الی 45 سال اند
- 4 = افراد که بالاتر از 45 سال اند

در این صورت میتوانید از فرمان recode به شکل ذیل استفاده کنید.

recode z2 (18/25 = 1) (26/35 = 2) (36/45 = 3) (46/97 = 4)

و نتیجه ی آنرا میتوانید میتوانید با استفاده از فرمان table z2 به شکل یک جدول ببینید.

table z2

How old are you?	Freq.
1	32,863
2	31,843
3	25,360
4	21,956

طور که دیده میشود متغیر سن (z2) به یک متغیر کتگوری شده تبدیل شده است که دارای 4 کتگوری میباشد.

20. جدا ساختن متغیرهای غیر عددی

در دروس قبلی خواندید که دیتای خام در دیتا سیت باید به منظم به فارمت tidy ذخیره گردد. یعنی هر ستون نمایانگر یک سوال باشد و هر ردیف نشان دهنده یک مشاهده باشد. به طور مثال زمان که یک سروی را راه اندازی میکنید و یکی از سوالات به شکل ذیل باشد:

سوال: دو سافتویر که برای تحلیل دیتا استفاده میکنید کدام ها اند؟

جواب اول: _____

جواب دوم: _____

در این صورت دیتای این سوال نباید در یک ستون ذخیره شود زیرا یک پاسخ دهنده حق دارد تا حد اکثر دو جواب به این سوال ارایه کند. اما ممکن بعضی از اوقات شما با دیتا سیت رو برو شوید که هر دو پاسخ را در یک ستون ذخیره کرده است. به طور مثال به متغیر software در دیتای ذیل توجه کنید .

gender_1[28]													
gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_st=s	education	age	income	
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	MA	31	1600	
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R	MA	.	1300	
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R	MA	29	1800	
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R	BA	23	1500	
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

در این صورت برای این که جوابات در دو سوال جداگانه ذخیره شود میتوانید از فرمان split به شکل ذیل استفاده کنید.

split software, parse(,)

software1[5]		Stata	
	software	software1	software2
1	R, Python	R	Python
2	R, Stata	R	Stata
3	R, SPSS	R	SPSS
4	Stata, R	Stata	R
5	Stata, R	Stata	R
6	SPSS, Stata	SPSS	Stata
7	R, SPSS	R	SPSS
8	Python, R	Python	R
9	Stata, R	Stata	R
10	R, Python	R	Python
11	R, Stata	R	Stata
12	Stata, SPSS	Stata	SPSS
13	Stata, R	Stata	R
14	Stata, R	Stata	R
15	SPSS, Stata	SPSS	Stata
16	R, SPSS	R	SPSS
17	Python, R	Python	R
18	Stata, R	Stata	R
19	Stata, R	Stata	R
20	Stata, R	Stata	R

نوت: برای اینکه جوابات در متغیر فوق به اساس کامه جدا شده اند بنابر این در داخل گزینه ی () parse کامه نوشته شده است. در صورت که به اساس یک نشانه ی دیگر، فرض مثال در صورت که جوابات به اساس نقطه جدا شده باشد پس در آن صورت در داخل () parse باید نقطه نوشته شود. در صورت که اگر به اساس فاصله جدا شده باشد پس ضرورت به استفاده از () parse نمیباشد و میتوانید مستقیماً فرمان split را نوشته و به تعقیب آن نام متغیر را نوشته کنید.

21. از بین بردن فاصله ها در یک متغیر غیر عددی

سافتویر ستاتا case sensitive و یا حساس با حروف میباشد. یعنی برای ستاتا کلمه ی «male» با کلمه ی «Male» تفاوت دارد و فرق آن این است که اولی آن با حرف خورد نوشته شده است و دومی آن با حرف کلان نوشته شده است. به همین ترتیب کلمه ی «female» با کلمه ی « female» فرق دارد زیرا اولی آن بدون فاصله است و دومی آن همراه با فاصله است. همیشه توصیه این است که زمان ذخیره نمودن دیتا باید فاصله های اضافی همراه با کلمات نوشته نشود. به طور مثال در دیتا ست ذیل به متغیر education دقت کنید که نشان دهنده ی سطح تحصیل میباشد.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_st-s	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

در صورت که خواسته باشید بدانید به چی تعداد از اشخاص در یک کتگوری مشخص تحصیلی هستند میتوانید از فرمان table استفاده کنید.

table education

education	Freq.
BA	6
BA	1
MA	13

قسم که در جدول فوق دیده میشود در جدول فوق کتگوری BA در دو ردیف نشان داده شده است که یکی آن همراه با فاصله و دیگر آن بدون فاصله میباشد بنابر این در صورت که خواسته باشید بدانید به چی تعداد از اشخاص در این کتگوری وجود دارند باید فرکونسی های هر دو کتگوری را جمع کنید و در صورت که تعداد کتگوری های متغیر زیاد تر باشد و این اشتباه در چندین کتگوری صورت گرفته باشد بنأ این کار درد سر ساز میشود. بهترین راه حل این است که تمام فاصله ها حذف گردد. برای از بین بردن فاصله ها از یک متغیر میتوانید که فرمان generate را همراه با strtrim به شکل ذیل استفاده کنید.

```
generate edu=strtrim(education)
```

edu[24]		
	education	edu
1	MA	MA
2	BA	BA
3	MA	MA
4	MA	MA
5	BA	BA
6	BA	BA
7	MA	MA
8	MA	MA
9	MA	MA
10	MA	MA
11	BA	BA
12	BA	BA
13	MA	MA
14	MA	MA
15	MA	MA
16	MA	MA
17	MA	MA
18	BA	BA
19	BA	BA
20	MA	MA

```
table edu
```

edu	Freq.
BA	7
MA	13

22. مقایسه نمودن دو دیتاست

اولین مرحله بعد از ختم سروی های که توسط قلم و کاغذ صورت گرفته باشد این است که تمام دیتا در یک سافتویر ذخیره گردد. البته ذخیره کردن دیتا یکی از مهمترین مراحل است که اکثریت اوقات امکان خطا نیز در آن میباشد. به طور مثال بعضی اوقات به نسبت غفلت در زمان ذخیره کردن دیتا امکان دارد یک شخص که 18 ساله باشد 81 ساله ذخیره گردد به همین ترتیب زمان که در ذخیره کردن دیتا دقت صورت نگیرد ممکن میزان اشتباه بسیار بلند برود. به این لحاظ توصیه میشود که بعد از یک بار ذخیره کردن دیتا یک فیصدی خاص دیتا دوباره در یک دیتا سیت جداگانه ذخیره شود

(double data entry) و دوباره با دیتاست اولی مقایسه گردد که آیا تغییرات وجود دارد یا نه. برای مقایسه کردن دو دیتا سیت می‌توانید از فرمان cf استفاده کنید. ساختار فرمان cf به شکل ذیل است.

```
cf varlist using data.dta
cf _all using data.dta
```

یعنی اولاً فرمان cf را نوشته به تعقیب آن متغیرات را که می‌خواهید با دیتاسیت دیگر مقایسه شود نوشته کرده و بعداً کلمه ی using را نوشته و به تعقیب آن نام دیتا سیت دومی را نوشته کنید. در صورت که خواسته باشید تمام متغیرات دو دیتا ست را با هم مقایسه کنید پس باید بعد از فرمان cf بدون ذکر نام متغیرات کلمه ی _all را نوشته کنید و متباقی ساختار فرمان یکسان است.

فرض مثال دو دیتاسیت ذیل را مد نظر بگیرید که معلومات در مورد عین اشخاص در دو دیتاسیت جداگانه ذخیره شده است و یک دیتاست معلومات در مورد سن افراد برای بعضی از افراد متفاوت میباشد.

id[20]						
	id	gender	age	software	education	
1	1	male	25	Stata, R	MA	
2	2	male	22	SPSS, Stata	BA	
3	3	male	23	R, SPSS	BA	
4	4	male	26	Python, R	MA	
5	5	female	25	Stata, R	MA	
6	6	male	19	R, Python	MA	
7	7	male	20	R, Stata	BA	
8	8	female	22	Stata, SPSS	BA	
9	9	female	26	Stata, R	MA	
10	10	female	23	Stata, R	BA	

compare1.dta (<https://github.com/Fahim-Ahmad/Stata-programming>)

id[1]		1				
	id	gender	age	software	education	
1	1	male	24	Stata, R	MA	
2	2	male	21	SPSS, Stata	BA	
3	3	male	22	R, SPSS	BA	
4	4	male	26	Python, R	MA	
5	5	female	25	Stata, R	MA	
6	6	male	19	R, Python	MA	
7	7	male	20	R, Stata	BA	
8	8	female	21	Stata, SPSS	BA	
9	9	female	27	Stata, R	MA	
10	10	female	23	Stata, R	BA	

compare2.dta (<https://github.com/Fahim-Ahmad/Stata-programming>)

برای مقایسه ی این دو دیتاست اولاً یکی از این دیتا ها را در ستاتا وارد کرده و بعداً دیتای وارد شده را با دیتاست دومی مقایسه کنید. البته ناگفته نباید گذاشت دیتای که در ستاتا وارد شده است به نام master dataset و دیتای دومی که می‌خواهید با آن مقایسه شود به نام using data set یاد میشود.

فرض مثال دیتاست اولی را که به نام compare1.dta است در ستاتا وارد کنید و برای اینکه با دیتاست دومی compare2.dta مقایسه کنید می‌توانید از فرمان ذیل استفاده کنید.

```
use compare1.dta, clear
cf _all using compare2.dta
```

```
. cf _all using compare2.dta
      age:  5 mismatches

r(9);
```

همچنان برای معلومات بیشتر در مورد تفاوت ها می‌توانید در اخیر فرمان verbose نوشته کنید.

```
cf _all using compare2, verbose
```

```
. cf _all using compare2, verbose
      age: 5 mismatches
           obs 1. 24 in master; 25 in using
           obs 2. 21 in master; 22 in using
           obs 3. 22 in master; 23 in using
           obs 8. 21 in master; 22 in using
           obs 9. 27 in master; 26 in using

r(9);
```

برعلاوه ی آن برای این که متغیرات که از همدیگر تغیر هم ندارند مشاهده کنید میتوانیم در اخیر فرمان فوق all نوشته کنید.

```
cf _all using compare2, verbose all
```

```
. cf _all using compare2, verbose all
      id: match
      gender: match
      age: 5 mismatches
           obs 1. 24 in master; 25 in using
           obs 2. 21 in master; 22 in using
           obs 3. 22 in master; 23 in using
           obs 8. 21 in master; 22 in using
           obs 9. 27 in master; 26 in using
      software: match
      education: match

r(9);
```

فصل دوم

تحلیل دیتا Data Analysis

در بخش اول موضوعات مدیریت دیتا مخصوصاً این که چگونه دیتای خام را برای تحلیل آماده بسازید تحت پوشش قرار گرفت. در این بخش روی این موضوع بحث میشود که چگونه دیتای خام را تحلیل کنید و کدام موضوعات را در تحلیل دیتا در نظر داشته باشید که البته شامل موضوعات از قبیل جدول سازی های یک طرفه و دو طرفه، جدول سازی برای متغیر های چندین جوابه، احصایه ی توصیفی و همچنان احصایه ی استنباطی مانند آزمون T و همبستگی میشود.

قابل یاد آوری است که در اکثریت مثال های این بخش کوشش صورت گرفته است که از دیتای سروی مردم افغانستان استفاده شود و نظریات تنها کسان که در سال 2018 سروی شده اند را مد نظر گرفته ایم. بنابر این بعد از این که دیتای سروی مردم افغانستان را باز کردید از طریق فرمان drop و یا keep دیتای سال های دیگر را حذف کنید و تنها دیتای سال 2018 را نگهدارید.

و با
`drop if m8 != 2018`

برای اینکه چطور از فرمان keep و یا drop برای فیلتر کردن دیتا استفاده کنیم مراجعه شود به درس پنجم. در صورت که در مرحله ی اول این کار را انجام ندهید پس مجبور هستید همراه با هر فرمان که در این بخش تدریس میشود از اصطلاح if کار بگیرید تا دیتای یک سال مشخص را بدست بیاورید.

23. جدول سازی یک طرفه

با استفاده از فرمان table میتوانید ابتدایی ترین معلومات احصایوی یک متغیر که عبارت از تعداد فریکونسی هر کتگوری میباشد را به دست بیاورید. به طور مثال در دیتای سروی مردم افغانستان برای اینکه بدانید به چی تعداد از افراد سروی شده مرد و یا زن میباشد میتوانید از فرمان ذیل استفاده کنید.

`table z1`

`. table z1`

Gender	Freq.
Male	7,550
Female	7,462

.

برعلاوه از فریکونسی هر کتگوری در یک متغیر، فیصدی مربوطه ی هر کتگوری نیز از جمله معلومات مهم یک متغیر میباشد که میتوانید آنرا با تقسیم نمودن فریکونسی یک کتگوری به نسبت تمام فریکونسی ها به دست بیاورید. به طور مثال در صورت که در جدول بالا خواسته باشید فیصدی مردان سروی شده را به دست بیاورید باید که تعداد مجموعی مردان را (7,550) تقسیم مجموع تعداد مردان و زنان (7,550+7462) کنید.

البته در صورت که تعداد کتگوری های یک متغیر زیادتر باشد این کار وقت بیشتر را نیاز دارد و استفاده از این روش موثر نمیشد. با استفاده از فرمان tabulate و یا مخفف آن tab میتوانید که تعداد فریکونسی و فیصدی مربوطه ی هر کتگوری یک متغیر خاص را به دست بیاوریم.

ساختار فرمان tabulate برای جدول سازی های یک طرفه به نحوه است که اولاً فرمان tabulate را نوشته و به تعقیب آن نام متغیر را نوشته میکنیم.

```
tabulate var_name
tabulate z1
```

Gender	Freq.	Percent	Cum.
Male	7,550	50.29	50.29
Female	7,462	49.71	100.00
Total	15,012	100.00	

به طور مثال با اجرای فرمان ذیل میتوانیم فیصدی افراد که در هر زون مشخص سروی شده اند را دریابید.

```
tabulate m4
```

Region	Freq.	Percent	Cum.
Central/Kabul	2,461	16.39	16.39
East	1,423	9.48	25.87
South East	1,122	7.47	33.35
South West	1,831	12.20	45.54
West	1,134	7.55	53.10
North East	3,216	21.42	74.52
Central/Hazarajat	729	4.86	79.38
North West	3,096	20.62	100.00
Total	15,012	100.00	

24. جدول سازی دو طرفه

سروی مردم افغانستان از افراد سروی شده در مورد پلان آینده شان در باره ی اینکه اگر فرصت مهیا شود آیا افغانستان را ترک میکنند و یا خیر نیز پرسیده است که البته دیتای این سوال در متغیر z20 ذخیره شده است. در قدم نخست جدول یک طرفه ای آن را با استفاده از فرمان tabulate به دست میاوریم.

tabulate z20

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Freq.	Percent	Cum.
Yes	5,487	36.55	36.55
No	8,746	58.26	94.81
Refused (vol.)	83	0.55	95.36
Don't know (vol.)	696	4.64	100.00
Total	15,012	100.00	

قسم که دیده میشود 36.55 فیصد افراد سروی شده گفته اند در صورت که موقع مهیا شود افغانستان را ترک میکنند. اما در صورت که خواسته باشید نظریات افراد سروی شده را در قسمت این سوال به اساس کتگوری ها مختلف مانند زن و مرد، شهری و دهاتی و حالت معدنی. و امثال آن بدانیم با که از جدولی های دو طرفه استفاده کنیم. جدول های دو طرفه را نیز میتوانید با استفاده از فرمان tabulate بدست بیاورید که ساختار آن قرار ذیل میباشد.

tabulate var_1 var_2

یعنی اولاً فرمان tabulate را نوشته و به تعقیب آن نام دو متغیر را نوشته کنید. به طور مثال در دیتایی سروی مردم افغانستان در صورت که خواسته باشید نظریات مردان و زنان را در مورد سوال z20 اینکه آیا کشور را ترک میکنند و یا نه، میتوانید از فرمان ذیل استفاده کنید.

tabulate z20 z1

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883	2,604	5,487
No	4,342	4,404	8,746
Refused (vol.)	42	41	83
Don't know (vol.)	283	413	696
Total	7,550	7,462	15,012

در جدول فوق دیده میتوانید که از میان 7,550 نفر مردان سروی شده به تعداد 2,883 نفر آن در جواب سوال فوق گزینه ی بلی Yes را انتخاب نموده اند و از میان 7,462 زنان سروی شده به تعداد 2,604 نفر این گزینه را انتخاب نموده اند. به یاد داشته باشید زمان که فرمان tabulate را همراه با دو متغیر انجام دهید، کتگوری های متغیر که در اول نوشته کنید در قسمت ردیف (row) های جدول و کتگوری های متغیر که دوم نوشته کنید در قست بالایی و یا ستون (column) جدول نمایش داده میشود. به نتیجه ی دو فرمان ذیل دقت کنید:

tabulate z1 z20

Gender	Tell me, if given opportunity, would you leave Afghanistan and live somewhere el				Total
	Yes	No	Refused (	Don't kno	
Male	2,883	4,342	42	283	7,550
Female	2,604	4,404	41	413	7,462
Total	5,487	8,746	83	696	15,012

tabulate z20 z1

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883	2,604	5,487
No	4,342	4,404	8,746
Refused (vol.)	42	41	83
Don't know (vol.)	283	413	696
Total	7,550	7,462	15,012

قسم که میبینید زمان که فرمان tabulate همراه با دو متغیر اجرا شود تنها فریکونسی های هر کتگوری متغیرات را نشان میدهد، اما زمان که خواسته باشید فیصدی های مربوطه ی هر کتگوری نمایش داده شود میتوانید از گزینه های row و column استفاده کنید.

زمان که از گزینه های row و یا column استفاده شود باید در قسمت تحلیل آن توجه ی زیاد صورت گیرد زیرا در صورت استفاده از گزینه ی column فیصدی نسبی کتگوری های هر ستون نشان میدهد و در صورت استفاده از row فیصدی نسبی کتگوری های هر ردیف نشان میدهد.

به طور مثال زمان که فرمان tabulate z20 z1 را اجرا کنید کتگوری های متغیر z20 در قسمت ردیف نشان داده میشود و کتگوری های متغیر جنسیت در قسمت ستون جدول نمایش داده میشود. به شکل ذیل توجه کنید:

tabulate z20 z1

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883	2,604	5,487
No	4,342	4,404	8,746
Refused (vol.)	42	41	83
Don't know (vol.)	283	413	696
Total	7,550	7,462	15,012

حالا در صورت که از گزینه ی column استفاده شود به این معنی است که فیصدی های نسبی به اساس متغیر z1 (جنسیت پاسخ دهنده) نمایش داده میشود.

Key
<i>frequency</i> <i>column percentage</i>

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883 38.19	2,604 34.90	5,487 36.55
No	4,342 57.51	4,404 59.02	8,746 58.26
Refused (vol.)	42 0.56	41 0.55	83 0.55
Don't know (vol.)	283 3.75	413 5.53	696 4.64
Total	7,550 100.00	7,462 100.00	15,012 100.00

یعنی در قسمت تحلیل آن باید گفت که از مجموع مردان سروی شده، 37% آنها به سوال z20 جواب بلی را انتخاب کرده اند. یعنی 37% مردان میگویند در صورت که فرصت مهیا شود افغانستان را ترک میکنند و به همین ترتیب باید گفت که از مجموع خانم های سروی شده 33.58% آنها گفته اند که در صورت مهیا شدن فرصت افغانستان را ترک میکنند. به همین ترتیب زمان که از گزینه ی row همراه با فرمان tabulate استفاده شود در آن صورت فیصدی های نسبی به اساس متغیر میباشد که کتگوری های آن در قسمت ردیف جدول نوشته شده است.

tabulate z20 z1, row

Key
<i>frequency</i> <i>row percentage</i>

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883 52.54	2,604 47.46	5,487 100.00
No	4,342 49.65	4,404 50.35	8,746 100.00
Refused (vol.)	42 50.60	41 49.40	83 100.00
Don't know (vol.)	283 40.66	413 59.34	696 100.00
Total	7,550 50.29	7,462 49.71	15,012 100.00

در این صورت در قسمت تحلیل آن باید گفت که از مجموع کسان که به سوال z20 جواب بلی گفته اند، 55.46% آن ها مردان میباشد و از مجموع کسان مه به سوال z20 جواب بلی گفته اند 44.54% آن ها زنان میباشد. همراه با فرمان tabulate میتوانید که هر دو گزینه ی column و row را به صورت همزمان نیز استفاده کنید. در صورت استفاده از گزینه های column و row به صورت همزمان باید به کلید رهنمایی جوابات دقت داشته باشید.

tabulate z20 z1, column row

Key
<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

Tell me, if given opportunity, would you leave Afghanistan and live somewhere el	Gender		Total
	Male	Female	
Yes	2,883 52.54 38.19	2,604 47.46 34.90	5,487 100.00 36.55
No	4,342 49.65 57.51	4,404 50.35 59.02	8,746 100.00 58.26
Refused (vol.)	42 50.60 0.56	41 49.40 0.55	83 100.00 0.55
Don't know (vol.)	283 40.66 3.75	413 59.34 5.53	696 100.00 4.64
Total	7,550 50.29 100.00	7,462 49.71 100.00	15,012 100.00 100.00

طور که در کلید جوابات میبینید ردیف اول نشان دهنده ی فریکوانسی هر کتگوری و ردیف دوم نشان دهنده فیصدی به اساس کتگوری های که در ردیف جدول (در اینجا کتگوری های متغیر z20) و ردیف سومی نشان دهنده ی فیصدی ها به اساس کتگوری های متغیر که در ستون جدول (در اینجا کتگوری های متغیر z1) میباشد.

25. جدول سازی برای متغیرات چندین جوابه

در اکثریت سروی های اجتماعی سوالات که دارای بیشتر از دو جواب هستند معمول میباشد. نوع تحلیل این نوع سوالات بستگی به این دارد که چگونه دیتا ذخیره شده است، به این اساس قبل از این که برویم به طرف این که چگونه این سوالات را تحلیل کنید بهتر است تا میتود های ذخیره کردن این نوع سوالات را یاد بگیرید.

برای ذخیره کردن متغیرات که دارای بیشتر از یک جواب میباشد حد اقل دو طریقه ی اساسی وجود دارد. **Indicator mode**: در این طریقه برای هر گزینه یک متغیر جداگانه ترتیب داده میشود. به طور مثال به سوال ذیل توجه کنید.

کدام سافتویر ذیل را برای تحلیل دیتا استفاده میکنید؟

Stata .a

R .b

SPSS .c

Python .d

حالا یک فرد ممکن زیاد تر از یک سافتویر را برای تحلیل دیتا استفاده کند به این ترتیب ممکن زیادتر از دو گزینه را انتخاب کند. یکی از راه های ذخیره کردن دیتای سوال فوق به نحوی میباشد که برای هر جواب مشخص یک متغیر جداگانه ایجاد شود. به طور مثال به دیتای ذیل توجه کنید متغیرات stata, r, python, spss به این شکل ذخیره گردیده است.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_st-s	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

نوت: در این میتود تعدا متغیرات مساوی به تعداد جوابات ارایه شده میباشد. به طور مثال در صورت که تعداد جوابات یک سوال ۱۰ گزینه باشد باید ۱۰ متغیر و در صورت که تعداد جوابات ۱۵ گزینه باشد باید ۱۵ متغیر ساخته شود. در صورت که دیتا به شکل فوق ذخیره شده باشد آسان ترین طریقه این است که با استفاده از فرمان tabulate برای هر متغیر یک جدول جداگانه ترتیب گردد.

. tabulate r

R	Freq.	Percent	Cum.
No	3	15.00	15.00
Yes	17	85.00	100.00
Total	20	100.00	

. tabulate stata

Stata	Freq.	Percent	Cum.
No	7	35.00	35.00
Yes	13	65.00	100.00
Total	20	100.00	

. tabulate python

Python	Freq.	Percent	Cum.
No	16	80.00	80.00
Yes	4	20.00	100.00
Total	20	100.00	

. tabulate spss

SPSS	Freq.	Percent	Cum.
No	14	70.00	70.00
Yes	6	30.00	100.00
Total	20	100.00	

2. Polytomous mode: طریقه ی فوق در صورت خوب است که تعداد جوابات از قبل معلوم و تعیین شده باشد. اما بسیار از سوالات به نوع است که پیشینی تعداد تمام جوابات از قبل مشکل و یا حتی ناممکن میباشد. و یا حتی در صورت که تعداد جوابات از قبل تعیین شده باشد خوبر است تا جوابات به ترتیب ذکر آن در دیتا ذخیره گردد. به طور مثال سوال فوق را میتوانید به شکل ذیل نیز بیان کنیم.

دو سافتویر که برای تحلیل دیتا استفاده میکنید کدام ها اند؟

جواب اول: _____

جواب دوم: _____

در این صورت تعداد جوابات مشخص نیست و در صورت ذکر یک گزینه جدید در هنگام جمع آوری دیتا، لیست جوابات نیز آپدیت گردد. در این صورت میتوانید دیتای جمع آوری شده را در دو متغیر ذخیره کنید. به طور مثال به متغیرات response_1 و response_2 در دیتای ذیل توجه کنید.

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_status	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

در صورت که دیتا به شکل فوق ذخیره شده باشد برای تحلیل آن باید از فرمان `mrtab` کار گرفته شود. ساختار این فرمان به نوع است که اولاً `mrtab` نوشته و به تعقیب آن نام های متغیرات را نوشته کرده و در اخیر فرمان `poly` نوشته کنید.

`mrtab varlist, poly`

فرض مثال اگر بخواهید که جدول فریکونسی و فیصدی دیتای فوق را به دست بیاورید پس میتوانید از فرمان ذیل استفاده کنیم.

`mrtab response_1 response_2, poly`

	Frequency	Percent of responses	Percent of cases
Python	4	10.00	20.00
R	17	42.50	85.00
SPSS	6	15.00	30.00
Stata	13	32.50	65.00
Total	40	100.00	200.00

Valid cases: 20

Missing cases: 0

همچنان برای ترتیب جوابات از کمترین به بیشترین می‌توانید در اخیر فرمان فوق باید sort نوشته شود.

```
mrtab response_1 response_2, poly sort
```

	Frequency	Percent of responses	Percent of cases
Python	4	10.00	20.00
SPSS	6	15.00	30.00
Stata	13	32.50	65.00
R	17	42.50	85.00
Total	40	100.00	200.00

Valid cases: 20

Missing cases: 0

همچنان برای ترتیب جوابات از بیشترین به کمترین باید در اخیر فرمان sort des نوشته شود.

```
mrtab response_1 response_2, poly sort des
```


	Frequency	Percent of responses	Percent of cases
R	17	42.50	85.00
Stata	13	32.50	65.00
SPSS	6	15.00	30.00
Python	4	10.00	20.00
Total	40	100.00	200.00
Valid cases:	20		
Missing cases:	0		

نتیجه ی فرمان فوق به نحوه است که در قسمت ردیف های جدول کتگوری های متغیرات ذکر شده نوشته شده و در قسمت ستون جدول تعداد فریکونسی هر کتگوری ذکر گردیده به همین ترتیب فیصدی نسبی هر کتگوری به اساس مجموع دفعات ذکر شده ی تمام جوابات ها (percent of responses) و فیصدی نسبی هر کتگوری به اساس مجوع افراد جواب دهنده و یا تعداد کلی مشاهدات (percent of cases) نیز بیان گردیده است.

در جدول فوق دیده میشود که مجموع دفعات که هر تمام جوابات ذکر گردیده است 40 میباشد و مجموع مشاهدات 20 میباشد. به این ترتیب برای به دست آوردن فیصدی نسبی هر کتگوری به اساس مجموع دفعات ذکر شده ی تمام جوابات (percent of responses) فریکونسی همان کتگوری مشخص تقسیم 40 میشود و در صورت که خواسته باشید فیصدی نسبی هر کتگوری را به اساس مجموع افراد جواب دهنده (percent of cases) به دست بیاورید فریکونسی مربوطه ی همان کتگوری باید تقسیم 20 شود.

به طور مثال 13 نفر گزینه ی Stata را حد اقل در یکی از جوابات انتخاب کرده اند. برای به دست آوردن این که چند فیصد مجموعی تمام افراد گزینه Stata را انتخاب کرده اند باید که 13 تقسیم 20 شود.

$$\text{Frequency} = 13$$

$$\text{Percent of responses} = (13/20) * 100 = 65.00\%$$

$$\text{Percent of responses} = (13/40) * 100 = 32.50\%$$

یعنی گفته میتوانیم که 65.00 فیصد افراد سروی شده گزینه ی Stata را انتخاب کرده است.

در سروی مردم افغانستان از افراد سروی شده پرسیده شده است که آیا فکر میکنند افغانستان به سمت درست در حرکت است و یا سمت اشتباه (x4) به تعقیب آن از افراد که گزینه ی "سمت درست" را انتخاب کرده اند پرسیده شده است تا دو دلیل را که فکر میکنند افغانستان به سمت درست در حرکت است کدام ها میباشد. دیتای آن در متغیرات x5a و x5b ذخیره گردیده است.

زمان که دیتای سروی مردم افغانستان را باز کنید میتوانید این متغیر را با اجرای فرمان ذیل مشاهده کنید.

x4[31]		102	
	x4	x5a	x5b
1	wrong direction	.	.
2	wrong direction	.	.
3	right direction	reconstruction / rebuilding	prevention/elimination of suicide attac...
4	wrong direction	.	.
5	don't know	.	.
6	wrong direction	.	.
7	wrong direction	.	.
8	wrong direction	.	.
9	right direction	good security	don't know
10	wrong direction	.	.
11	wrong direction	.	.
12	wrong direction	.	.
13	wrong direction	.	.
14	wrong direction	.	.
15	wrong direction	.	.
16	wrong direction	.	.
17	wrong direction	.	.
18	wrong direction	.	.
19	right direction	peace / end of the war	reconstruction / rebuilding
20	right direction	reconstruction / rebuilding	having better relations with foreign co...

قسم که دیده میشود در صورت که در متغیر x4 گزینه ی right direction انتخاب شده باشد پس سوال x5a و x5b از آنان پرسیده شده است و در صورت که در متغیر x4 دیگر گزینه ها انتخاب شده باشد متغیرات x5a و x5b خالی میباشد (و یا ممکن در آن نوشته باشد).

در این صورت برای تحلیل این متغیرات از فرمان mrtab به همراه اصطلاح if استفاده به شکل ذیل باید استفاده شود.
mrtab x5a x5b if x4==101, poly sort des

		Frequency	Percent of responses	Percent of cases
104	reconstruction / rebuilding	1498	15.54	30.85
103	good security	1252	12.98	25.79
999	don't know	798	8.28	16.44
101	peace / end of the war	751	7.79	15.47
115	reduction in the level of administrative corruption	442	4.58	9.10
132	improvement in education system	430	4.46	8.86
106	economic revival	399	4.14	8.22
109	schools for girls have opened	380	3.94	7.83
113	good government	293	3.04	6.04
121	road reconstruction	289	3.00	5.95
108	democracy / elections	220	2.28	4.53
117	having active ana and anp	220	2.28	4.53
130	removing the taliban	177	1.84	3.65
127	more electricity supply now than before	169	1.75	3.48
114	international assistance	154	1.60	3.17
110	women can now work	126	1.31	2.60
123	development in agriculture system	116	1.20	2.39
124	more job opportunities available	111	1.15	2.29
119	having legal constitution	108	1.12	2.22
154	presidential elections	98	1.02	2.02
151	improved justice	96	1.00	1.98
128	national unity	95	0.99	1.96
136	prevention/elimination of crimes	92	0.95	1.89
102	disarmament	77	0.80	1.59

998	refused	4	0.04	0.08
168	having active airports	3	0.03	0.06
171	organizations are active	3	0.03	0.06
181	youth are involved in politics	3	0.03	0.06
183	distribution of new id cards	3	0.03	0.06
193	Removing Daish/ISIS	3	0.03	0.06
141	decrease number of accidents	2	0.02	0.04
148	decrease in arbitrary attacks by the us	2	0.02	0.04
156	access to products	2	0.02	0.04
158	creation of arbakies	2	0.02	0.04
165	nothing is going to right direction	2	0.02	0.04
159	detecting and defusing mines	1	0.01	0.02
163	banking services	1	0.01	0.02
166	better and more media	1	0.01	0.02
186	building sports stadiums	1	0.01	0.02
189	establishment of provincial council	1	0.01	0.02
Total		9642	100.00	198.60
Valid cases:		4855		
Missing cases:		0		

26. آمار توصیفی احصایوی

برای تحلیل متغیرات کیفی و یا متغیرات که به شکل کتگوری باشد می‌توانید از جداول استفاده کنید. اما سوالات که به شکل کمی می‌باشد مانند سن، معاش، تعداد ساعات که یک فرد در یک روز کار میکند و امثال آن عبارت از دیتای کمی می‌باشد که خلاصه ساختن این متغیرات به شکل جداول مطلوب نمی‌باشد. در سروی مردم افغانستان متغیر z2 نشان دهنده ی سن افراد سروی شده می‌باشد. در صورت که بخواهید آنرا به شکل جدول نمایش بدهید نتیجه ی آن به شکل ذیل می‌باشد.

tabulate z2

How old are you?	Freq.	Percent	Cum.
18	743	4.95	4.95
19	530	3.53	8.48
20	706	4.70	13.18
21	328	2.18	15.37
22	427	2.84	18.21
23	372	2.48	20.69
24	393	2.62	23.31
25	646	4.30	27.61
26	387	2.58	30.19
27	385	2.56	32.75
28	514	3.42	36.18
29	328	2.18	38.36
30	991	6.60	44.96
31	358	2.38	47.35
32	384	2.56	49.91

67	32	0.21	97.93
68	30	0.20	98.13
69	13	0.09	98.21
70	102	0.68	98.89
71	15	0.10	98.99
72	17	0.11	99.11
73	17	0.11	99.22
74	7	0.05	99.27
75	41	0.27	99.54
76	14	0.09	99.63
77	9	0.06	99.69
78	5	0.03	99.73
79	2	0.01	99.74
80	32	0.21	99.95
81	2	0.01	99.97
82	4	0.03	99.99
83	1	0.01	100.00
Total		15,012	100.00

در این صورت بهترین روش این است که اینچنین متغیرات به کتگوری های جداگانه تقسیم شده و بعداً جدول فریکونسی و فیصدی آن ساخته شود. برای این که چطور میتوانیم یک متغیر پیوسته را به متغیر کتگوری شده تبدیل کنید مراجعه شود به درس 15.

همچنان ما میتوانیم از آمار توصیفی مانند اوسط mean میدیان median انحراف معیاری و امثال آن استفاده کنیم. در ستاتا با استفاده از فرمان summarize و یا مخفف آن sum میتوانیم آمار توصیفی احصایوی یک متغیر کمی را بدست بیاوریم.

به طور مثال در صورت که خواسته باشیم اوسط سن افراد سروی شده را به دست بیاورید پس میتوانیم از فرمان ذیل استفاده کنید.

```
summarize z2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	15,012	35.16354	13.00086	18	83

طور که دیده میشود فرمان summarize نه تنها اوسط یک متغیر را نشان میدهد بلکه انحراف معیاری، خوردترین ارزش، بالاترین ارزش و مجموع تعداد مشاهدات را نیز نشان میدهد.

با نوشتن کلمه ی detail در اخیر فرمان فوق ما میتوانیم معلومات دیگر را مانند پراکندگی و میدیان را نیز بدست بیاورید.

```
summarize z2, detail
```

How old are you?

Percentiles		Smallest		
1%	18	18		
5%	19	18		
10%	20	18	Obs	15,012
25%	25	18	Sum of Wgt.	15,012
50%	33		Mean	35.16354
		Largest	Std. Dev.	13.00086
75%	43	82		
90%	54	82	Variance	169.0223
95%	60	82	Skewness	.8233967
99%	72	83	Kurtosis	3.254772

همچنان در صورت که خواسته باشیم آمار توصیفی سن افراد سروی شده را تنها در زون غرب بدست بیاورید پس میتوانید از فرمان های ذیل استفاده کنید.

des m4

variable name	storage type	display format	value label	variable label
m4	byte	%17.0g	M4	Region

label list M4

M4:

- 1 Central/Kabul**
- 2 East**
- 3 South East**
- 4 South West**
- 5 West**
- 6 North East**
- 7 Central/Hazarajat**
- 8 North West**

```
summarize z2 if m4==5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,134	34.6455	12.62339	18	80

همچنان میتوانیم معلومات احصایوی متغیر سن را با استفاده از فرمان فوق به زون های مختلف دیگر نیز به دست بیاورید.

```
. summarize z2 if m4==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	2,461	35.42097	13.48989	18	80

```
. summarize z2 if m4==2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,423	32.63528	11.30649	18	80

```
. summarize z2 if m4==3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,122	34.34581	12.0795	18	81

```
. summarize z2 if m4==4
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,831	33.38613	10.47798	18	80

```
. summarize z2 if m4==6
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	3,216	35.64956	13.64789	18	82

```
. summarize z2 if m4==7
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	729	37.63923	14.68946	18	82

```
. summarize z2 if m4==8
```

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	3,096	36.57041	13.68155	18	83

برعلاوه طریقه ی فوق، می‌توانید فرمان summarize و یا مخفف آن sum را همراه فرمان tabulate یکجا نوشته کرده تا معلومات احصایوی یک متغیر خاص را به اساس هر کتگوری متغیر دیگر بدست بیاورید.

فرض مثال در صورت که خواسته باشید اوسط و انحراف معیاری سن افراد سروی شده در سروی مردم افغانستان را در هر یک از زون ها به صورت ساده به دست بیاورید پس می‌توانید از فرمان ذیل کار بگیرید.

```
tabulate m4, summarize(z2)
```

Region	Summary of How old are you?		
	Mean	Std. Dev.	Freq.
Central/K	35.420967	13.489892	2,461
East	32.635278	11.306487	1,423
South Eas	34.345811	12.079495	1,122
South Wes	33.386128	10.477976	1,831
West	34.645503	12.623389	1,134
North Eas	35.649565	13.647895	3,216
Central/H	37.639232	14.689465	729
North Wes	36.570413	13.681546	3,096
Total	35.163536	13.000857	15,012

و یا در صورت که خواسته باشید اوسط سن را به صورت جداگانه برای مردان و زنان در زون های مختلف بدست بیاورید پس
میتوانید از فرمان ذیل استفاده کنید.

tabulate m4 z1, summarize(z2)

Means, Standard Deviations and Frequencies of How old are you?

Region	Gender		Total
	Male	Female	
Central/K	36.547697	34.320482	35.420967
	14.728077	12.064241	13.489892
	1216	1245	2461
East	33.951841	31.338912	32.635278
	12.035194	10.386299	11.306487
	706	717	1423
South Eas	35.879218	32.801431	34.345811
	13.214073	10.607105	12.079495
	563	559	1122
South Wes	34.284783	32.478595	33.386128
	10.886265	9.9728936	10.477976
	920	911	1831
West	37.800699	31.434164	34.645503
	13.352508	10.94876	12.623389
	572	562	1134
North Eas	38.351572	33.00738	35.649565
	14.820904	11.815086	13.647895
	1590	1626	3216
Central/H	40.121884	35.203804	37.639232
	15.499878	13.429776	14.689465
	361	368	729
North Wes	39.421702	33.432836	36.570413
	14.583706	11.851017	13.681546
	1622	1474	3096
Total	37.242517	33.060038	35.163536
	14.025321	11.499218	13.000857
	7550	7462	15012

طور که دیده میشود فرمان فوق اوسط سن افراد سروی، انحراف معیاری و مجموع افراد سروی شده را به هر کتگوری مرد و

زن به اساس زون های مختلف نمایش داده است.

27. استفاده از فرمان bysort

فرمان bysort میتواند از جمله یکی از پر کار آمد ترین فرمان های ستاتا به شمار برود. در صورت که بخواهید یک فرمان را انجام بدهید و عین فرمان را بار بار به اساس کتگوری های متغیر دیگر انجام دهید ضرورت به استفاده ی فرمان مورد نظر همراه با اصطلاح if میباشد. به طور مثال در صورت که خواسته باشید دریابید در سروی مردم افغانستان به چی تعداد از افراد در مورد آینده ی کشور خوش بین و یا بد بین هستند میتوانید از فرمان ذیل استفاده کنید.

tabulate x4

Generally speaking, do you think things in Afghanistan today are going in the ri	Freq.	Percent	Cum.
right direction	4,855	32.34	32.34
wrong direction	9,180	61.15	93.49
refused	74	0.49	93.98
don't know	903	6.02	100.00
Total	15,012	100.00	

اما در صورت که خواسته باشید معلومات فوق را برای هر کتگوری متغیر جنسیت بدست بیاورید می توانید از فرمان bysort استفاده کنید. ساختار این فرمان به نوع میباشد که اولاً bysort نوشته کرده و به تعقیب آن نام یک و یا چندین متغیر را نوشته و بعد از آن فرمان اصلی خود را نوشته کنید.

bysort z1: tabulate x4

- Male			
Generally speaking, do you think gs in Afghanistan today are going in the ri	Freq.	Percent	Cum
right direction	2,486	32.93	32.9
wrong direction	4,609	61.05	93.9
refused	46	0.61	94.5
don't know	409	5.42	100.0
Total	7,550	100.00	
= Female			
Generally speaking, do you think gs in Afghanistan today are going in the ri	Freq.	Percent	Cum
right direction	2,369	31.75	31.7
wrong direction	4,571	61.26	93.0
refused	28	0.38	93.3
don't know	494	6.62	100.0

همچنان در صورت که خواسته باشیم معلومات احصایوی سن افراد سروی شده را در زون های مختلف بدست بیاورید نیز

میتوانیم از فرمان bysort و summarize به صورت یکجایی به شکل ذیل استفاده کنید.

bysort m4: summarize z2

-> m4 = Central/Kabul

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	2,461	35.42097	13.48989	18	80

-> m4 = East

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,423	32.63528	11.30649	18	80

-> m4 = South East

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,122	34.34581	12.0795	18	81

-> m4 = South West

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,831	33.38613	10.47798	18	80

-> m4 = West

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	1,134	34.6455	12.62339	18	80

-> m4 = North East

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	3,216	35.64956	13.64789	18	82

-> m4 = Central/Hazarajat

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	729	37.63923	14.68946	18	82

-> m4 = North West

Variable	Obs	Mean	Std. Dev.	Min	Max
z2	3,096	36.57041	13.68155	18	83

28. پیدا کردن ارزش های خالی در متغیرات

در سروی های اجتماعی ارزش های خالی و یا missing values میتواند چندین علت داشته باشد. دو نوع معمول آن این است که اولاً یک شخص اصلاً حاضر به سروی نمی شود که به نام unit-nonresponse نیز یاد میگردد و دوم این که یک شخص به یک و یا چند سوال مشخص جواب نمی دهد که به نام item-nonresponse نیز یاد میگردد.

برای اینکه بدانید به چی تعداد از ارزش های خالی در یک متغیر موجود است راه های مختلف وجود دارد. یکی از راه های آسان آن استفاده از فرمان tabulate میباشد. در صورت که در اخیر این فرمان , missing نوشته کنید بر علاوه از فرکونسی و فیصدی دیگر کتگوری های آن متغیر، تعداد مشاهدات خالی را نیز در آن نشان میدهد.

به طور مثال به دیتای ذیل دقت کنید. در متغیرات age, marital_status, income بعضی از ارزش های خالی و یا missing values وجود دارد.

gender_1[28]													
gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_status	education	age	income	
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

برای پیدا نمودن تعداد ارزش های خالی در هر یک از متغیرات میتوانیم از فرمان های ذیل استفاده کنیم.

```
tabulate marital_status, missing
```

marital_status	Freq.	Percent	Cum.
married	4	20.00	20.00
single	8	40.00	60.00
single	8	40.00	100.00
Total	20	100.00	

```
tabulate age, missing
```

age	Freq.	Percent	Cum.
19	1	5.00	5.00
21	2	10.00	15.00
23	2	10.00	25.00
24	1	5.00	30.00
25	2	10.00	40.00
26	1	5.00	45.00
27	1	5.00	50.00
28	1	5.00	55.00
29	1	5.00	60.00
30	2	10.00	70.00
31	1	5.00	75.00
32	1	5.00	80.00
33	2	10.00	90.00
.	2	10.00	100.00
Total	20	100.00	

tabulate income, missing

income	Freq.	Percent	Cum.
700	1	5.00	5.00
900	1	5.00	10.00
1100	1	5.00	15.00
1300	2	10.00	25.00
1400	1	5.00	30.00
1500	3	15.00	45.00
1600	1	5.00	50.00
1700	1	5.00	55.00
1800	3	15.00	70.00
1900	1	5.00	75.00
2000	2	10.00	85.00
.	3	15.00	100.00
Total	20	100.00	

همچنان می‌توانید برای پیدا نمودن ارزش های خالی از فرمان count همراه با اصطلاح if استفاده کنید. به طور مثال در صورت که خواسته باشید ارزش های خالی را در متغیر های فوق دریابید پس می‌توانید از فرمان ذیل استفاده کنید.

```
count if missing(marital_status)
count if missing(age)
count if missing(income)
```

```
. count if missing(marital_status)
4

. count if missing(age)
2

. count if missing(income)
3
```

در صورت که خواسته باشید ارزش های خالی را در چندین متغیر به صورت همزمان به دست بیاورید می‌توانید از فرمان `mdesc` استفاده کنید. ساختار آن به نوع می‌باشد که اولاً `mdesc` نوشته و به تعقیب آن نام متغیر و یا متغیرات که می‌خواهید ارزش های خالی آن را به دست بیاورید باید نوشته شود.

```
mdesc marital_status age income
```

Variable	Missing	Total	Percent Missing
marital_status	4	20	20.00
age	2	20	10.00
income	3	20	15.00

29. آزمون T و یا t-test

t-test از جمله آزمون های احصایوی برای مقایسه ی اوسط دو گروه است که آیا اوسط دو گروه از هم تفاوت معنی دار دارد و یا خیر. به طور مثال در صورت که خواسته باشید دریابید که آیا مردان معاش زیاد به دست می‌آورند و یا زنان، منطقی به نظر نمی‌آید که شما تمام مردان و زنان یک جمعیت را سروی کنید و در مورد معاش آنان پرسید. بلکه شما می‌توانید به صورت نمونه به طور مثال همراه با 300 مرد و 300 خانم سروی کنید و فرض مثال شما دریابید که مردان به صورت اوسط 4000 اما خانم ها به صورت اوسط 6000 ماهانه معاش دارند. برای دانستن اینکه این تفاوت در اوسط معاشات نمونه ی سروی شده معنی دار می‌باشد و در سطح تمام جامعه صدق میکند و یا اینکه این تفاوتی معنی دار میان اوسط هر دو گروه وجود ندارد می‌توانید از آزمون T استفاده کنید.

تفاوت اوسط ها امکان زیاد دارد در صورت های ذیل معنی دارد باشد.

1. در صورت که تفاوت اوسط دو گروه بزرگتر باشد.
2. در صورت که تعداد نمونه ی انتخاب شده زیادتر باشد.
3. جوابات به صورت متداوم نزدیک به اوسط بوده و پراکنده گی کمتر داشته باشد.

t-test انواع مختلف دارد که از پر کاربرد ترین آن t-test دو طرفه می‌باشد که خود آن به دو نوع دیگر می‌باشد.

1. T-test برای مقایسه نمودن اوسط دو متغیر جداگانه. فرض مثال شما نمرات شاگردان را قبل و بعد

از انجام یک تریننگ در دو متغیر جداگانه به نام های `pre_test` و `post_test` ذخیره کرده اید. برای این که بدانید اوسط نمرات قبل و بعد از برگزاری تریننگ یکسان بوده و یا متفاوت میباشد میتواند از فرمان `ttest` در ستاتا به شکل ذیل استفاده کنید.

```
ttest pre_test == post_test
```

2. T-test برای مقایسه اوسط دو گروه. فرض مثال شما در مورد معاش ماهانه ی که خانم ها و مردان در یک روز مصرف میکنند معلومات جمع آوری کرده باشید و آن را در یک متغیر به نام `income` و متغیر جنسیت را به نام `gender` ذخیره کرده باشید برای این که بدانید آیا خانم ها معاش زیادتر بدست میاورند و یا مردان، و یا به زبان ساده تر اینکه تفاوت میان اوسط معاش ماهوار میان زنان و مردان معنی دارد است و یا خایر میتوانید از فرمان `ttest` به شکل ذیل استفاده کنید.

```
ttest income, by(gender)
```

به طور مثال با استفاده از دیتای سروی مردم افغانستان در صورت که خواسته باشید بدانید آیا اوسط سن زنان با مردان تفاوت دارد و یا نه و آیا این تفاوت معنی دار میباشد و یا به صورت تصادفی اتفاق افتاده است میتوانید از فرما ذیل استفاده کنید.

```
ttest z2, by(z1)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Male	7,550	37.24252	.1614133	14.02532	36.9261	37.55893
Female	7,462	33.06004	.1331192	11.49922	32.79909	33.32099
combined	15,012	35.16354	.1061091	13.00086	34.95555	35.37152
diff		4.182479	.2094653		3.771902	4.593057

diff = mean(Male) - mean(Female)

t = **19.9674**

Ho: diff = 0

degrees of freedom = **15010**

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = **1.0000**

Pr(|T| > |t|) = **0.0000**

Pr(T > t) = **0.0000**

در فرمان بالا فرضیه ی صفر این است که تفاوت میان اوسط ها وجود ندارد و یا اینکه تفاوت میان اوسط سن و مرد صفر است ($H_0: \text{diff} = 0$) و فرضیه a این است که تفاوت میان اوسط ها وجود دارد و یا اینکه مساوی به صفر نیست ($H_a: \text{diff} \neq 0$) در صورت که قیمت pr کوچکتر از 0.05 باشد پس گفته میتوانید که اوسط هر دو گروه از هم تفات دارد و این تفاوت معنی

دار است در غیر آن گفته می‌توانید که تفاوت میان اوسط هر گروه معنی دار نه بلکه به صورت اتفاقی می‌باشد.

30. آزمون همبستگی

آزمون همبستگی طور که از نام آن پیدا است رابطه بین دو متغیر را نشان می‌دهد. به طور مثال شما از افراد مختلف در مورد وزن و میزان مصرف قند (شرینی) آنها پرسان میکنید، برای اینکه دریابید آیا میان مصرف قند و وزن افراد رابطه‌ی معنی دار (مستقیم و یا غیر مستقیم) وجود دارد و یا خیر پس می‌توانید از آزمون همبستگی استفاده کنید.

آزمون همبستگی برای پیدا نمودن این که دو متغیر با هم چی نوع رابطه (مستقیم و یا معکوس) دارد و درجه این رابطه تا کدام اندازه (قوی، متوسط، ضعیف) است به کار میرود. ضریب همبستگی که به حرف r نشان داده میشود میتواند که بین ارزش های 1 الی -1 قیمت بگیرد. در صورت که ضریب همبستگی منفی باشد نشان دهنده ی رابطه معکوس بین دو متغیر و در صورت که ضریب همبستگی قیمت مثبت داشته باشد رابطه مستقیم بین دو متغیر را نشان میدهد.

گذشته از علامه ی ضریب همبستگی (مثبت و یا منفی) قیمت مطلقه ی آن نشان دهنده ی درجه ی همبستگی میان دو متغیر میباشد، یعنی ضریب همبستگی 6- رابطه ی قوی تر را نسبت به ضریب همبستگی 3+ نشان میدهد. در صورت که قیمت مطلقه ضریب همبستگی 1 باشد نشان دهنده ی رابطه ی مکمل میباشد و در صورت که قیمت ضریب همبستگی 0 باشد نشان دهنده ی این موضوع است که میان دو متغیر اصلاً هیچ رابطه وجود ندارد. در صورت که قیمت مطلقه ی ضریب همبستگی میان 0 الی 0.3 باشد نشان دهنده ی همبستگی ضعیف، در صورت که قیمت آن میان 0.4 الی 0.7 باشد نشان دهنده ی همبستگی متوسط و در صورت که قیمت آن بزرگتر از 0.7 باشد نشان دهنده ی همبستگی قوی میباشد.

البته در صورت که خواسته باشید رابطه بین متغیرات پیوسته را پیدا کنید پس می‌توانید از آزمون همبستگی پیرسون Pearson استفاده کنید و در صورت که خواسته باشید رابطه بین متغیرات که به شکل کتگوری ترتیبی باشد را دریابید پس می‌توانید از آزمون همبستگی سپرمن Spearman استفاده کنید.

برای اجرای آزمون همبستگی پیرسون در ستاتا می‌توانید از فرمان `pwcorr` استفاده کنید. بطور مثال در صورت که خواسته باشید در دیتای سروی مردم افغانستان رابطه میان سن افراد سروی شده ($z2$) و سطح تحصیل آنان ($z55$) را پیدا کنید، زیرا هردو متغیر به شکل پیوسته اند پس می‌توانید از آزمون همبستگی Pearson استفاده کنید.

در قدم نخست متغیر که نشان دهنده ی سطح تحصیل است را میبینیم که دارای کدام ارزش ها است.

`tabulate z55`

What is the highest level (grade) of school you have completed, not including sc	Freq.	Percent	Cum.
0	7,553	50.31	50.31
1	308	2.05	52.36
2	391	2.60	54.97
3	406	2.70	57.67
4	382	2.54	60.22
5	353	2.35	62.57
6	642	4.28	66.85
7	294	1.96	68.80
8	406	2.70	71.51
9	335	2.23	73.74
10	450	3.00	76.74
11	283	1.89	78.62
12	1,713	11.41	90.03
13	56	0.37	90.41
14	722	4.81	95.22
15	31	0.21	95.42
16	412	2.74	98.17
17	1	0.01	98.17
19	1	0.01	98.18
Informal schooling at home or at a lite	253	1.69	99.87
Refused	19	0.13	99.99
Don't know	1	0.01	100.00
Total	15,012	100.00	

چون در آزمون همبستگی از وریانس و یا تغییرات قیمت استفاده میشود بنابر این از متغیرات باید قیمت های refused و don't know و امثال آنها پاک شود. برای اینکه این جوابات به کدام ارزش ها در یک متغیر ذخیره شده است میتوانید از فرمان labellist استفاده کنید.

labellist z55

Z55:

97 Informal schooling at home or at a literacy class

98 Refused

99 Don't know

طور که دیده میشود این کتگوری ها در متغیر z55 به قیمت های 98,97 و 99 ذخیره شده. برای اینکه این قیمت ها در زمان انجام آزمون همبستگی شامل تحلیل نشود میتوانید که هنگام اجرای فرمان کورلیشن از اصطلاح if استفاده کنید و یا اینکه یک متغیر جدید بسازید که این قیمت ها شامل آن نباشد و متغیر جدید را در آزمون همبستگی استفاده کنید و یا اینکه این قیمت ها را در داخل خود متغیر به ارزش های خالی تبدیل کنید و دوباره آزمون کورلیشن را انجام دهید.

انجام فرمان کورلیشن همراه با اصطلاح if.

```
pwcorr z55 z2 if z55<97
```

	z55	z2
z55	1.0000	
z2	-0.2302	1.0000

ساختن یک متغیر جدید، به نام education که قیمت های 98, 97 و 99 را شامل نباشد.

```
gen education=z55 if z55<97
```

```
pwcorr education z2
```

	educat~n	z2
education	1.0000	
z2	-0.2302	1.0000

تبدیل کردن قیمت های 98, 97 و 99 به ارزش های خالی در عین متغیر.

```
recode z55 (97 98 99 = .)
```

```
pwcorr z55 z2
```

	z55	z2
z55	1.0000	
z2	-0.2302	1.0000

در هر سه حالت فوق دیده میشود که ضریب همبستگی منفی است (-0.2321). به این معنی که رابطه میان سن (z2) و سطح تحصیل (z55) معکوس است. یعنی افراد که دارای سن بالاتر اند، سطح تحصیل کمتر را گذارش داده اند. اما از آنجای که قیمت مطلقه ی ضریب همبستگی 0.2321 است پس گفته میتوانیم که رابطه ی ذکر شده یک رابطه چندان قوی نیست. در صورت که خواسته باشید بدانید آیا رابطه ی بدست آمده از لحاظ احصایوی معنی دار است و یا خیر باید که قیمت p و یا p-value را به دست بیاورید.

برای بدست آوردن قیمت P پس باید در اخیر فرمان فوق sig نوشته کنید.

pwcorr z55 z2, sig

	z55	z2
z55	1.0000	
z2	-0.2302 0.0000	1.0000

هر زمان که قیمت P کوچکتر از 0.05 باشد پس گفته می‌توانید که رابطه ی بدست آمده معنی دار است. در مثال فوق نتیجه گیری کرده می‌توانید که در سروی مردم افغانستان میان سن و سطح تحصیل افراد رابطه ی منفی وجود دارد که این رابطه از لحاظ احصایوی معنی دارد و است.

نوع دیگر از آزمون همبستگی عبارت از همبستگی پیرسون Pearson میباشد. تفاوت عمده ی که همبستگی Pearson و همبستگی Spearman دارد در این میباشد که آزمون همبستگی Pearson برای دریافت رابطه و شدت همبستگی میان متغیرات که به شکل پیوسته باشد به کار میرود اما آزمون همبستگی Spearman برای دریافت رابطه و شدت همبستگی میان متغیرات که به شکل کتگوری باشد به کار میرود. در هر نوع آزمون همبستگی علامت ضریب همبستگی (مثبت و یا منفی) نشان‌دهنده ی جهت همبستگی (مستقیم و یا معکوس) و قیمت مطلقه ی ضریب همبستگی نشان دهنده ی شدت همبستگی میباشد.

به طور مثال در سروی مردم افغانستان در مورد میزان رضایت افراد دیموکراسی (x82) و میزان کلی خوشحالی آنها (z47) پرسیده شده است که با استفاده از فرمان tabulate می‌توانید ببینید که این متغیرات به شکل پیوسته است و یا به شکل کتگوری شده اند.

tabulate x82

On the whole, how satisfied or dissatisfied are you with the way democracy works	Freq.	Percent	Cum.
very satisfied,	2,451	16.33	16.33
somewhat satisfied,	6,899	45.96	62.28
somewhat dissatisfied	3,668	24.43	86.72
very dissatisfied?	1,875	12.49	99.21
refused (vol.)	9	0.06	99.27
don't know (vol.)	110	0.73	100.00
Total	15,012	100.00	

tabulate z47

(Ask All) In general in your life, would you say you are very happy, somewhat happy	Freq.	Percent	Cum.
Very happy	5,300	35.31	35.31
Somewhat happy	6,765	45.06	80.37
Not very happy	2,319	15.45	95.82
Not at all happy	575	3.83	99.65
Refused	2	0.01	99.66
Don't know	51	0.34	100.00
Total	15,012	100.00	

دیده میشود که هر دو متغیر فوق به شکل کتگوری میباشد بنابر این باید از آزمون همبستگی Spearman استفاده گردد. قبلاً نیز ذکر گردید که آزمون همبستگی با وریانس و یا تغییرات قیمت ها سر و کار دارد بنابر این ارزش های Refused و Don't know باید از متغیرات حذف گردد. در قدم نخست با استفاده از فرمان labellist ببینید که هر کتگوری دو متغیر فوق درای کدام ارزش ها میباشد.

labellist x82 z47

X82_M:

101 very satisfied,
102 somewhat satisfied,
103 somewhat dissatisfied
104 very dissatisfied?
998 refused (vol.)
999 don't know (vol.)

Z47:

1 Very happy
2 Somewhat happy
3 Not very happy
4 Not at all happy
98 Refused
99 Don't know

طور که میبینید کتگوری های Refused و Don't know در متغیر x82 (که نشاندهنده میزان رضایت از دموکراس است) به قیمت های 998 و 999 میباشد و در متغیر z47 (که نشاندهنده ی میزان کلی خوشحالی افراد سروی شده است) به قیمت های 98 و 99 ذخیره شده اند. بنأ این قیمت ها باید که به ارزش های خالی تبدیل شود که میتوانید این کار را با استفاده از فرمان replace و یا هم فرمان recode انجام دهید.

• با استفاده از فرمان recode

recode x82 (998 999=.)

recode z47 (98 9 = .)

• با استفاده از فرمان `replace`

```
replace x82=. if x82==998 | x82==999  
replace z47=. if z47==98 | z47==99
```

بعد از اینکه این قیمت های فوق را به ارزش های خالی تبدیل کردید میتوانید از فرمان `spearman` برای آزمون همبستگی به شکل ذیل استفاده کنیم.

```
spearman z47 x82
```

```
Number of obs =    14841  
Spearman's rho =    0.2083  
  
Test of Ho: z47 and x82 are independent  
Prob > |t| =    0.0000
```

قسم که دیده میشود قیمت p و یا p -value کوچکتر از 0.05 است بنابر آن گفته میتوانید که یک رابطه ی معنی دار میان میزان خوشحالی و سطح رضایت از دموکراسی افراد سروی شده وجود دارد و همچنان برای این که ضریب همبستگی (ρ) دارای علامت مثبت است گفته میتوانید که یک رابطه ی مستقیم میان سطح رضایت از دموکراسی و سطح خوشحالی وجود دارد. یعنی کسان که خوشحال هستند دارای میزان رضایت زیادتر از دموکراسی اند نسبت که به کسان که به صورت عموم در زندگی خوشحال نیستند و یا کمتر خوشحال هستند. از آنجای که ضریب همبستگی 0.2013 میباشد گفته میتوانید که این همبستگی نسبتاً ضعیف است.

فصل سوم

نمایش دیتا Data Visualization

در دو بخش قبلی موضوعات از قبیل اینکه چطور دیتای خام را برای تحلیل آماده کنید و همچنان اینکه کدام نکات را در هنگام تحلیل دیتا مدنظر بگیرید مورد بحث قرار گرفت. در دنیای امروزی نه تنها آماده ساختن دیتا و تحلیل دیتا از اهمیت والای برخوردار است بلکه این موضوع که چطور دیتا نمایش داده شود و کدام میتود ها برای نمایش دیتا انتخاب گردد از جمله موضوعات مهم میباشد.

اشکال احصایوی می توانند اطلاعات موجود را به سرعت در معرض دید قرار دهند، شنیده‌اید که می‌گویند: یک تصویر خوب ارزش هزار کلمه را دارد. با دیدن یک گراف احصایوی در یک نگاه می‌توان به بسیاری از ویژگی‌های مجموعه‌ای از اطلاعات پی برد در این جا کوشش میشود تا مورد استفاده از اشکال احصایوی که بیشترین کاربرد را دارند و نوع ترسیم آن ها در سافتویر ستاتا بصورت ساده و عام فهم بیان گردد و هدف عمده و اساسی این بخش آشنایی ابتدایی با نمایش دیتا و این که از کدام میتود برای نمایش کدام نوع دیتا استفاده شود میباشد.

قابل یاد آوری است که در اکثریت مثال های این بخش نیز کوشش صورت گرفته است که از دیتای سروی مردم افغانستان استفاده شود و نظریات تنها کسان که در سال 2018 سروی شده اند را مد نظر گرفته ایم. بنابر این بعد از این که دیتای سروی مردم افغانستان را باز کردید از طریق فرمان drop و یا keep دیتای سال های دیگر را حذف کنید.

```
drop if m8 != 2018
```

و یا

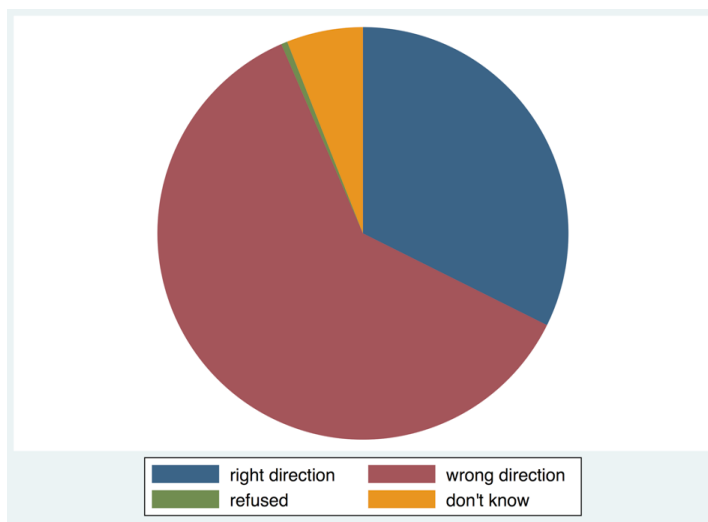
```
keep if m8 == 2018
```

برای اینکه چطور از فرمان keep و یا drop برای فلتر کردن دیتا استفاده شود به درس پنجم مراجعه کنید.

31. پای چارت

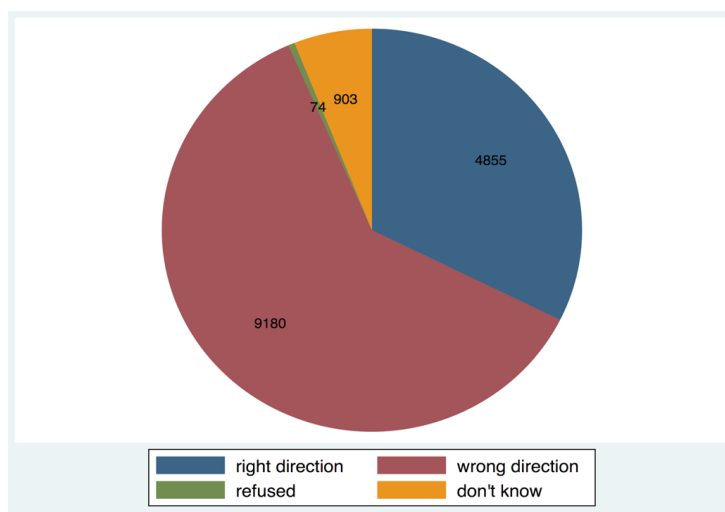
در صورت که خواسته باشید معلومات یک متغیر که به شکل کتگوری باشد را در روی گراف های احصایوی نشان بدهید میتوانید از پای چارت استفاده کنید که برای ترسیم آن در ستاتا میتوانید از فرمان graph pie استفاده کنید. به طور مثال در سروی مردم افغانستان در مورد نظریه ی سروی شده گان در مورد آینده کشور پرسان شده است اینکه افغانستان به سمت درست در حرکت است و یا سمت اشتباه که در متغیر x4 ذخیره شده است. برای نمایش آن در گراف میتوانید از فرمان ذیل استفاده کنید.

```
graph pie, over(x4)
```



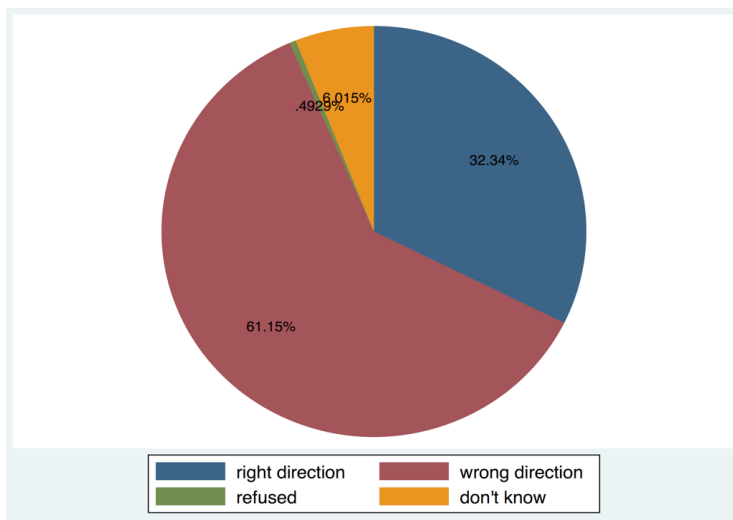
از شکل فوق به وضاحت معلوم میشود تعداد زیاد افراد که کتگوری wrong direction را انتخاب کرده اند بیشتر از دیگر کتگوری ها است. اما برای دانستن آمار دقیق هر کتگوری باید که فرکونسی مشخص هر کتگوری بالای آن وضع شده باشد. برای اینکه فرکونسی هر کتگوری در شکل فوق نمایش داده شود میتوانید که از گزینه ی plabel استفاده کنیم.

`graph pie, over(x4) plabel(_all sum)`



در شکل فوق دیده میشود که 4855 نفر گزینه ی right direction و 9180 نفر گزینه ی wrong direction را انتخاب کرده اند. اکثریت اوقات خوبتر این است که به عوض فرکونسی، هر کتگوری به فیصدی نمایش داده شود. برای انجام این کار در قسمت plabel به عوض _all sum باید _all percent نوشته شود.

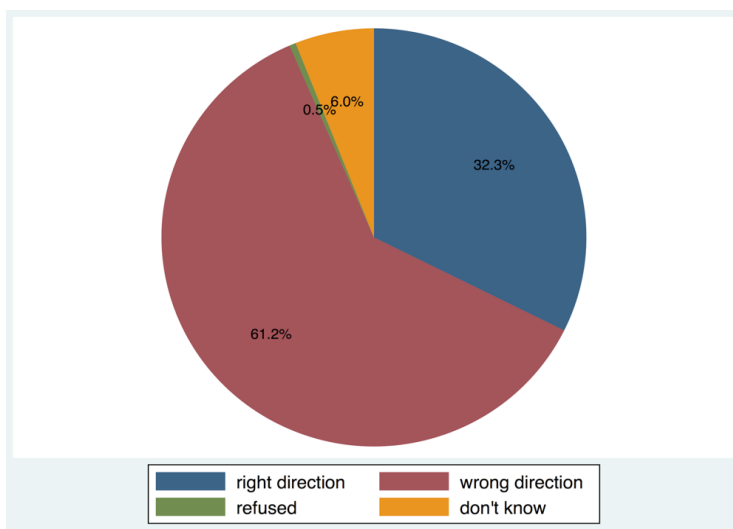
```
graph pie, over(x4) plabel(_all percent)
```



در شکل فوق دیده میشود که فیصدی مربوطه ی هر کتگوری بالای آن وضع شده است. طور که دیده میشود فیصدی مربوطه ی بعضی از کتگوری ها دارای 2 خانه اعشاری، در بعضی از کتگوری های دارای 3 خانه اعشاری و بعضی از کتگوری ها دارای 4 خانه اعشاری میباشد.

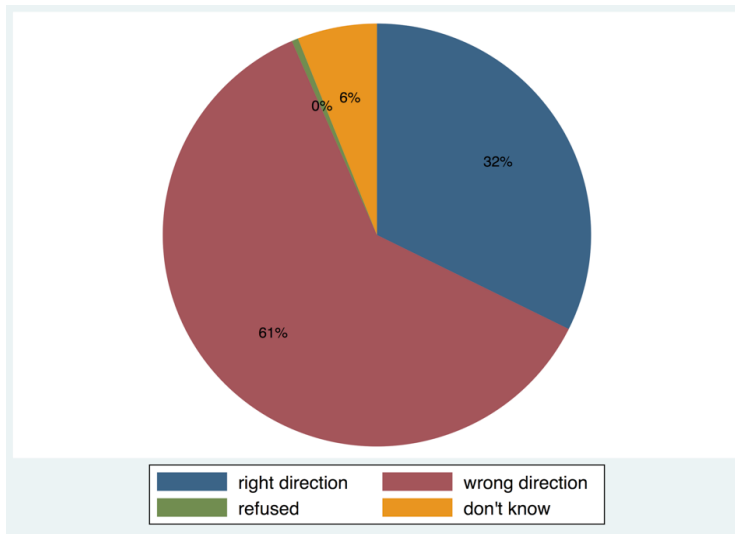
برای اینکه تعداد خانه های اعشاری را به دلخواه خودتان تغییر بدهید پس میتوانید که در قسمت plabel فارمت آن را تغییر دهید. به طور مثال در صورت که خواسته باشید تنها یک خانه ی اعشاری نمایش داده شود میتوانید از فرمان ذیل استفاده کنید.

```
graph pie, over(x4) plabel(_all percent, format(%9.1f))
```



در قسمت فارمت به یاد داشته باشید در صورت که 9.1f% نوشته کنید به این معنی که تمام فیصدی ها الی یک خانه ی اعشاری نمایش داده شود، در صورت که 9.2f% نوشته به این معنی که تمام فیصدی ها الی دو خانه فیصدی نمایش داده شود و همین ترتیب در صورت که 9.8f% نوشته کنید به این است که الی 8 خانه اعشاری نمایش داده شود. و در صورت که خواسته باشید تمام خانه های اعشاری حذف شود میتوانید 9.0f% نوشته کنید.

```
graph pie, over(x4) plabel(_all percent, format(%9.0f))
```

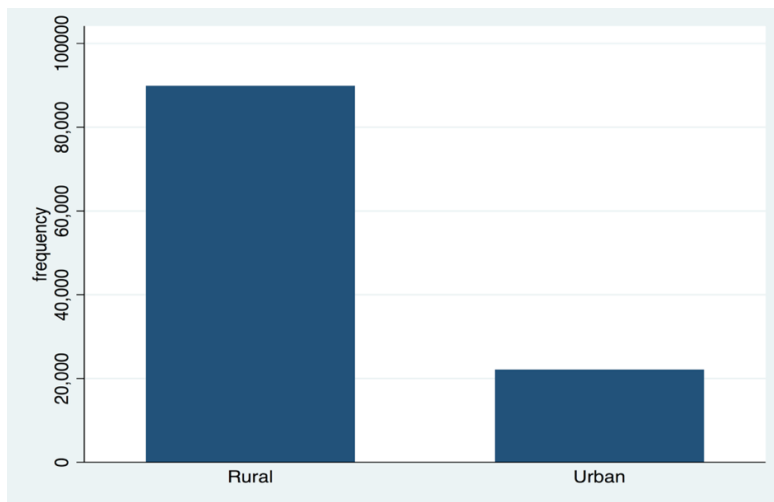


32. بار چارت

بار چارت نیز یکی از اشکال احصایوی میباشد که برای نشان دادن فریکونسی و یا فیصدی هر یک از کتگوری های یک متغیر به کار میرود. در این نوع گراف احصایوی تعداد بار ها نشان دهنده تعداد کتگوری های یک متغیر است و بلندی هر بار نشان دهنده فریکونسی و یا فیصدی همان کتگوری میباشد.

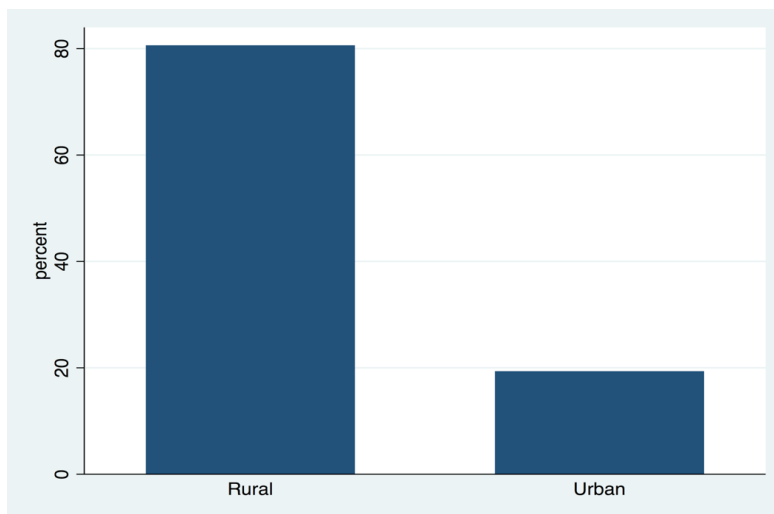
به طور مثال در سروی مردم افغانستان معلومات در مورد این که افراد سروی شده در روستا ها زندگی میکنند و یا در شهرها نیز ذخیره گردیده است (m6b). برای نمایش این دیتا در بار چارت میتوانید از فرمان graph bar استفاده کنید.


```
graph bar (count), over(m6b)
```



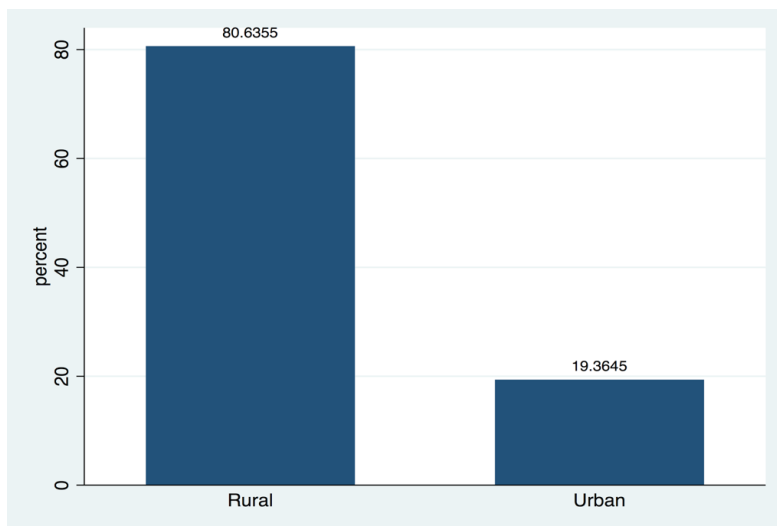
طور که در شکل فوق دیده میشود بلندی هر بار نشان دهنده ی فرکانسی مربوطه همان کتگوری میباشد. برای تبدیل کردن آن به فیصدی میتوانید که از فرمان فوق کلمه ی (count) را حذف کنید.

```
graph bar, over(m6b)
```

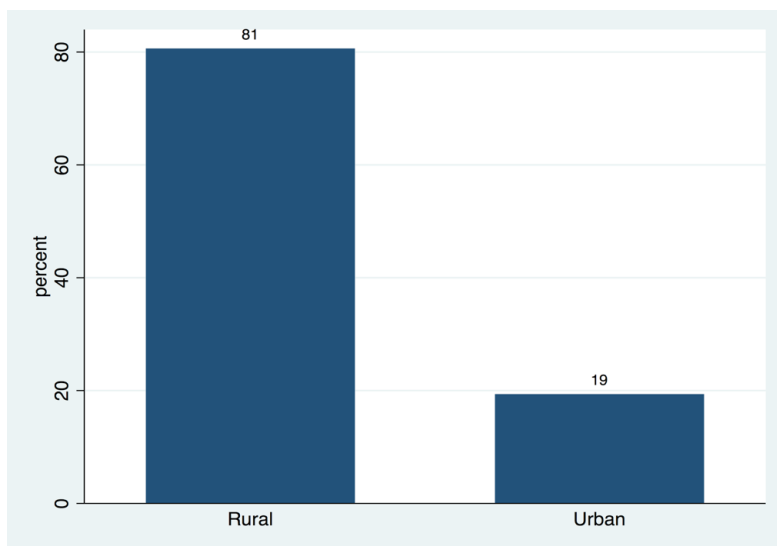


همچنان برای اینکه فیصدی هر کتگوری در هر بار نمایش داده شود میتوانید از گزینه ی blabel به شکل ذیل استفاده کنید.

```
graph bar, over(m6b) blabel(bar)
```

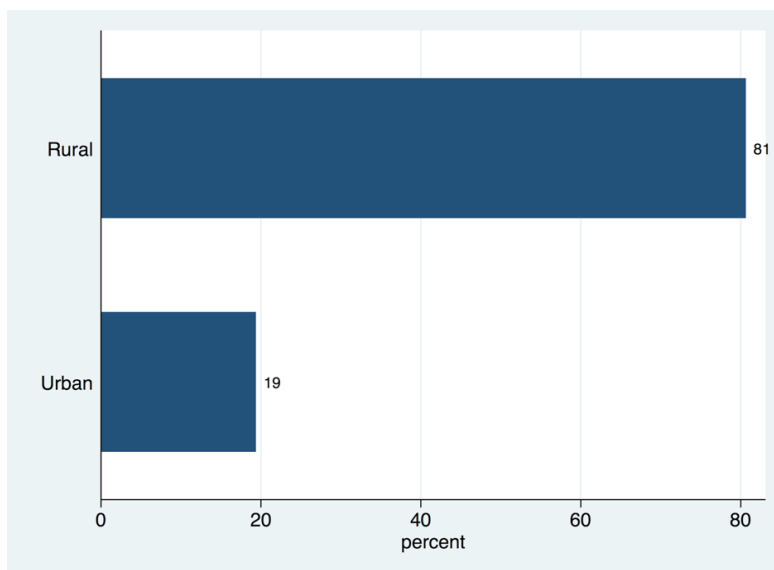


همچنان برای تعیین تعداد خانه های اعشاری می‌توانید از گزینه ی format استفاده کنیم.
`graph bar, over(m6b) blabel(bar, format(%9.0f))`



در تمام گراف های فوق تعداد بار ها که نشان دهنده ی تعداد کتگوری های متغیر میباشد به صورت عمودی نشان داده شده است، در صورت که خواسته باشید بار ها به صورت افقی نشان داده شود می‌توانید که از فرمان `graph hbar` استفاده کنید و دیگر ساختار فرمان به مانند `graph bar` میباشد. به طور مثال در صورت که خواسته باشید در گراف فوق هر یک از بار ها به صورت افقی نمایش داده شود می‌توانید از فرمان ذیل استفاده کنید.

```
graph hbar, over(m6b) blabel(bar, format(%9.0f))
```



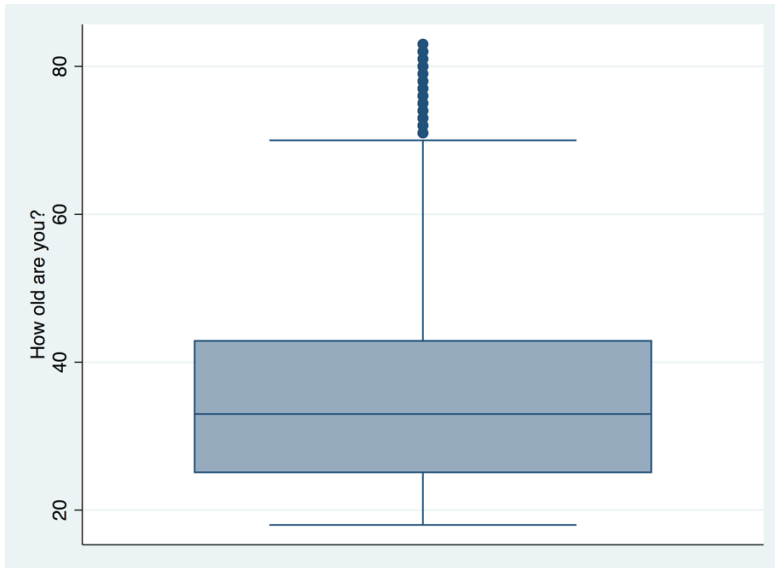
33. باکس گراف

برای نمایش دیتای متغیر که به شکل پیوسته باشد از بار گراف و یا پای گراف استفاده نمیشود مگر اینکه آن متغیر به شکل کتگوری شده تبدیل شود. در صورت که بخواهید معلومات احصایوی یک متغیر که به شکل پیوسته است را در اشکال احصایوی نشان دهید بهترین گزینه استفاده از باکس گراف است. یک باکس گراف معلومات احصایوی ذیل را در مورد یک متغیر نشان میدهد.

- خوردترین ارزش minimum value
- کوارتایل اول first quartile
- وسط median
- کوارتایل سوم third quartile
- بزرگترین ارزش maximum value

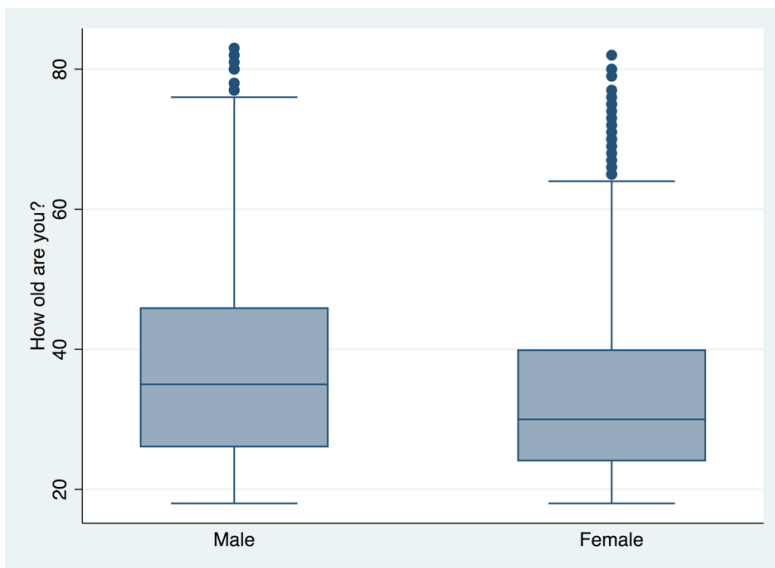
برای ترسیم باکس گراف در ستاتا باید graph box نوشته و به تعقیب آن نام متغیر را نوشته کنید. به طور مثال در صورت که خواسته باشید معلومات احصایوی فوق را در متغیر سن z2 در دیتای سروی مردم افغانستان نشان دریابید پس میتوانید از فرمان ذیل استفاده کنید.

```
graph box z2
```



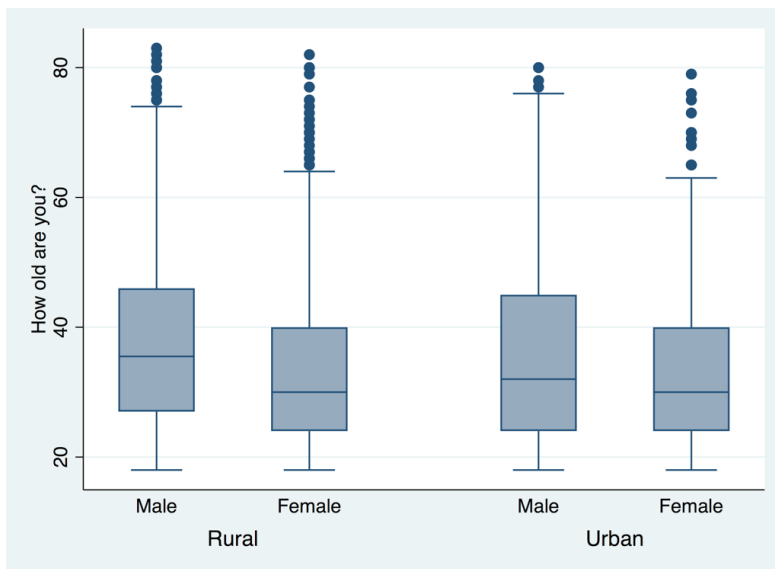
به همین ترتیب در صورت که خواسته باشید معلومات احصایوی فوق را به اساس جنسیت پاسخ دهندهگان دریابید پس میتوانید از فرمان ذیل استفاده کنید.

`graph box z2, over(z1)`



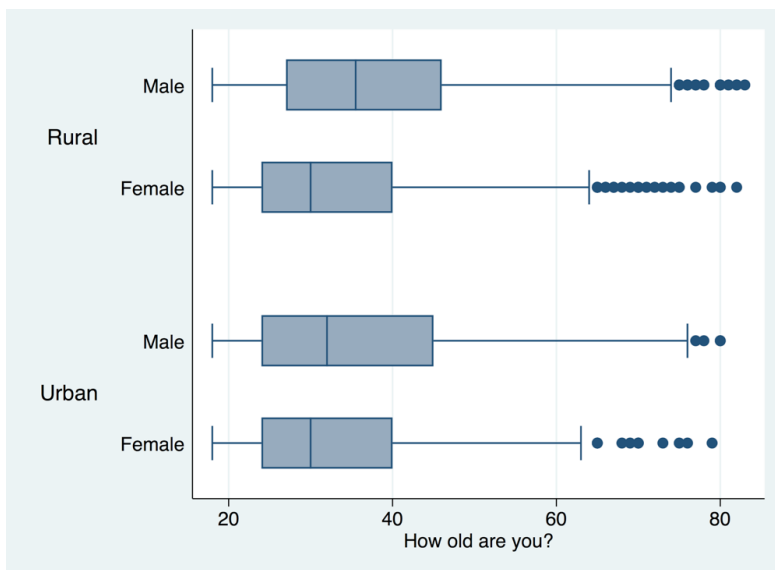
به همین ترتیب نه تنها معلومات فوق را به اساس یک متغیر بلکه به اساس چندین متغیر نیز نمایش داده میتوانید. به طور مثال برای اینکه خواسته باشید معلومات فوق را به اساس جنسیت z1 و محل زیست m6b پاسخ دهندهگان به دست بیاورید پس میتوانید از فرمان ذیل استفاده کنید.

```
graph box z2, over(z1) over(m6b)
```



همچنان در صورت که خواسته باشید این گراف به شکل افقی نشان داده شود میتواند به عوض `graph box` فرمان `graph hbox` نوشته کنید و متباقی ساختار فرمان یک سان میباشد.

```
graph hbox z2, over(z1) over(m6b)
```



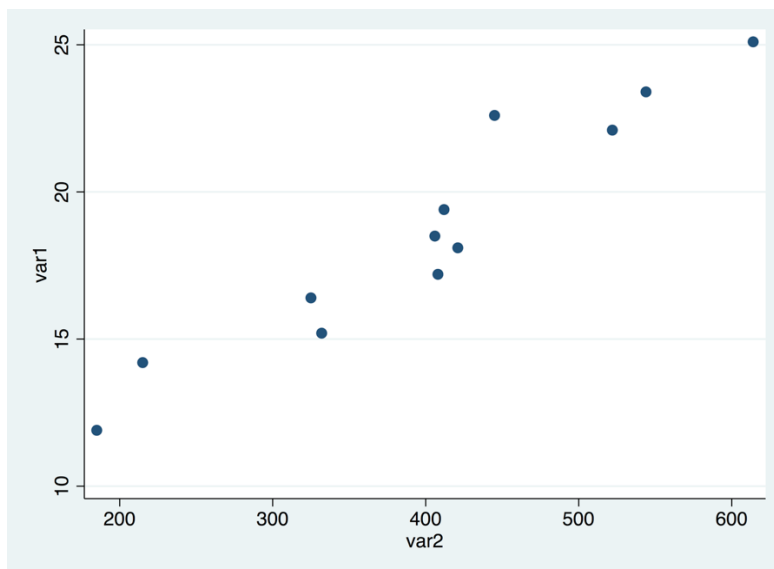
34. گراف پراگندگی

گراف پراگندگی بیشتر برای نشان رابطه بین دو متغیر که به شکل پیوسته بوده باشد به کار میرود. به طور مثال دیتای ذیل را در نظر بگیرید که نشان دهنده ی درجه ی حرارت در روز های مختلف (var1) و مجموعه ی فروشات یکی از شرکت های آیسکریم (var2) میباشد.

var1[1]			14.2
	var1	var2	
1	14.2	215	
2	16.4	325	
3	11.9	185	
4	15.2	332	
5	18.5	406	
6	22.1	522	
7	19.4	412	
8	25.1	614	
9	23.4	544	
10	18.1	421	
11	22.6	445	
12	17.2	408	

برای ترسیم گراف پراگندگی میتوانید از فرمان scatter استفاده کنید و ساختار آن طوری است که اولاً scatter نوشته کرده بعداً به تعقیب آن نام های متغیرات را نوشته کنید. البته باید یاد آور شد متغیر که در اول نوشته میشود در قسمت y-axis گراف و متغیر که دوم نوشته میشود در قسمت x-axis گراف نشان داده میشود.

```
scatter var1 var2
```



در دیتا ست سروی مردم افغانستان از افراد سروی شده در مورد اینکه سن مناسب ازدواج برای خانم ها (x369) و سن مناسب ازدواج برای مردان (x370) چند سالگی است پرسیده شده است. برای اینکه این معلومات را در نمودار پراگندگی نشان بدهید میتوانید از فرمان scatter استفاده کنید اما قبل از آن جوابات refused و don't know را به ارزش های خالی باید تبدیل کنید. برای دانستن اینکه جوابات refused و don't know در این متغیرات دارای کدام ارزش ها میباشد میتوانید از فرمان labellist به شکل ذیل استفاده کنید.

```
labellist x369 x370
```

X369:

98 Refused (vol.)
99 Don't Know (vol.)

X370:

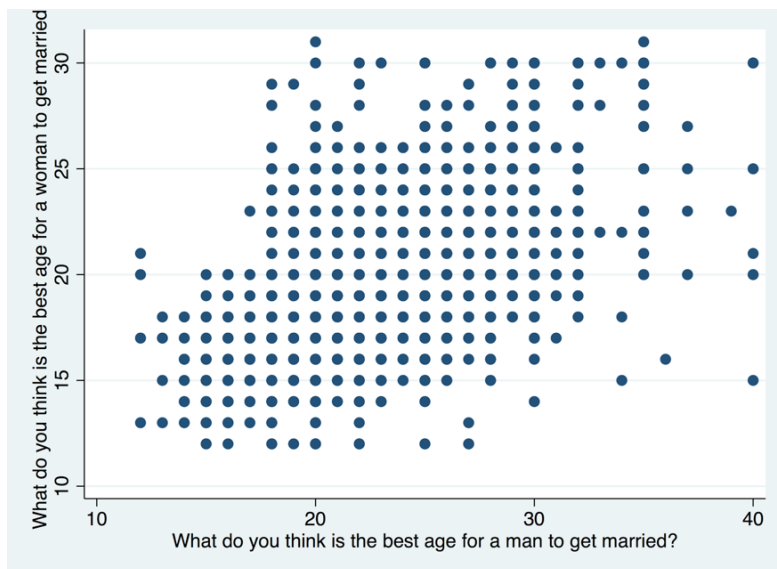
98 Refused (vol.)
99 Don't Know (vol.)

بعد از آن برای تبدیل کردن این قیمت ها به قیمت های خالی (missing values) میتوانید از فرمان ذیل کار بگیرید.

```
recode x369 (98 99=.)  
recode x370 (98 99=.)
```

بعد از آن برای ترسیم گراف پراگندگی میتوانید از فرمان ذیل استفاده کنید.

```
scatter x369 x370
```



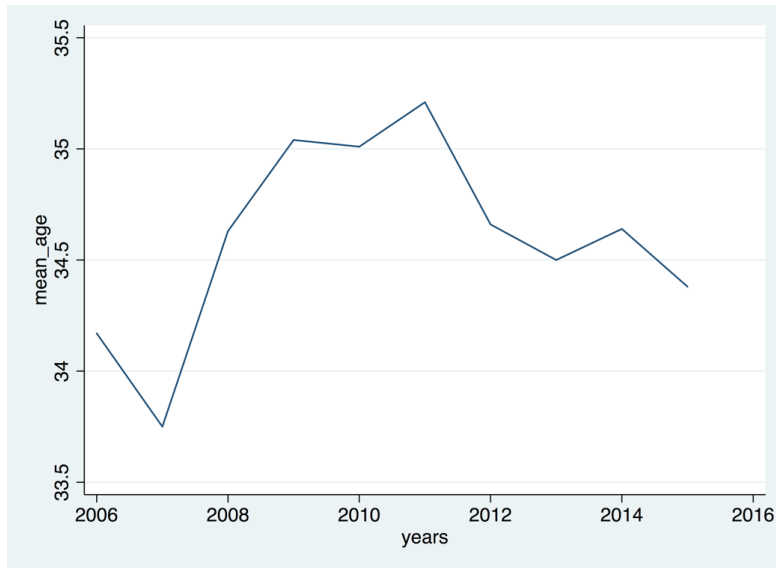
35. گراف خطی

در صورت که دیتا در مورد یک متغیر در طول زمان مختلف داشته باشید و بخواهید آنرا بر روی گراف نشان دهید یکی از معمولترین روش ها استفاده از گراف خطی میباشد. به طور مثال دیتای ذیل را مد نظر بگیرید که نشان دهنده اوسط سن در سال های مختلف میباشد.

years[1]			2006
	years	mean_age	
1	2006	34.17	
2	2007	33.75	
3	2008	34.63	
4	2009	35.04	
5	2010	35.01	
6	2011	35.21	
7	2012	34.66	
8	2013	34.5	
9	2014	34.64	
10	2015	34.38	

برای اینکه دیتای فوق را بر روی گراف خطی نشان دهید می‌توانید از فرمان `line` استفاده کنید. ساختار این فرمان به نوع می‌باشد که اولاً `line` نوشته کرده به تعقیب آن متغیر که نشانده ی سال ها می‌باشد و بعد از آن نام متغیر دومی را نوشته کنید.

```
line mean_age year
```



36. گراف سوالات چندین جوابه

سوالات چندین جوابه در سروی های اجتماعی بسیار معمول است. این که سوالات چندین جوابه به کدام شکل تحلیل شود بر می‌گردد به اینکه این جوابات آن به کدام طریق ذخیره گردیده است. به دو طریق می‌تواند دیتای سوالات چندین جوابه را ذخیره گردد.

- اول این که برای هر جواب یک متغیر جداگانه ساخته شود (indicator mode)
- دوم این که جوابات به ترتیب که پاسخ دهندگان ذکر کرده اند ذخیره شود. (Polytomous mode)

به طور مثال به سول ذیل توجه کنید.
 دو سافتویر که برای تحلیل دیتا استفاده می‌کنید کدام ها اند؟

جواب اول: _____
 جواب دوم: _____

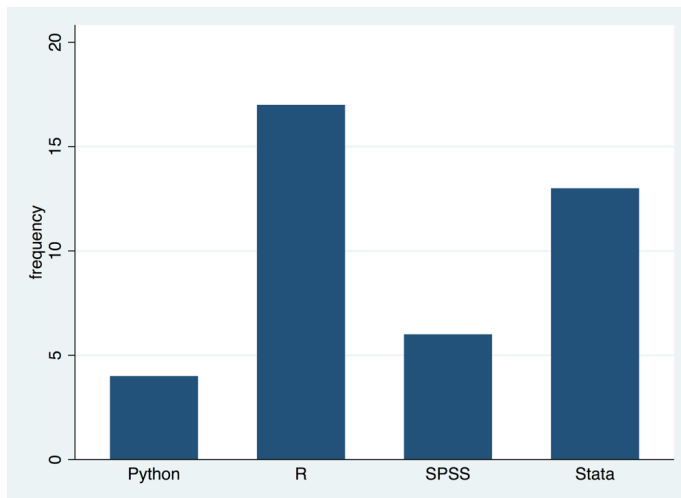
حالا به دیتا سیت ذیل توجه کنید:

gender_1[28]													
	gender_1	gender_2	software	r	stata	python	spss	response_1	response_2	marital_status	education	age	income
1	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	30	1500
2	1	male	R, Stata	Yes	Yes	No	No	R	Stata	married	BA	25	1400
3	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS		MA	31	1600
4	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	28	1500
5	1	male	Stata, R	Yes	Yes	No	No	Stata	R	married	BA	23	.
6	1	male	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	BA	21	700
7	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	27	1300
8	1	male	Python, R	Yes	No	Yes	No	Python	R		MA	.	1300
9	0	female	Stata, R	Yes	Yes	No	No	Stata	R	married	MA	33	1800
10	1	male	R, Python	Yes	No	Yes	No	R	Python	single	MA	32	2000
11	1	male	R, Stata	Yes	Yes	No	No	R	Stata	single	BA	19	900
12	0	female	Stata, SPSS	No	Yes	No	Yes	Stata	SPSS	married	BA	21	1100
13	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	24	.
14	1	male	Stata, R	Yes	Yes	No	No	Stata	R		MA	29	1800
15	0	female	SPSS, Stata	No	Yes	No	Yes	SPSS	Stata	single	MA	30	2000
16	1	male	R, SPSS	Yes	No	No	Yes	R	SPSS	married	MA	.	1900
17	1	male	Python, R	Yes	No	Yes	No	Python	R	married	MA	26	1700
18	0	female	Stata, R	Yes	Yes	No	No	Stata	R		BA	23	1500
19	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	BA	25	.
20	0	female	Stata, R	Yes	Yes	No	No	Stata	R	single	MA	33	1800

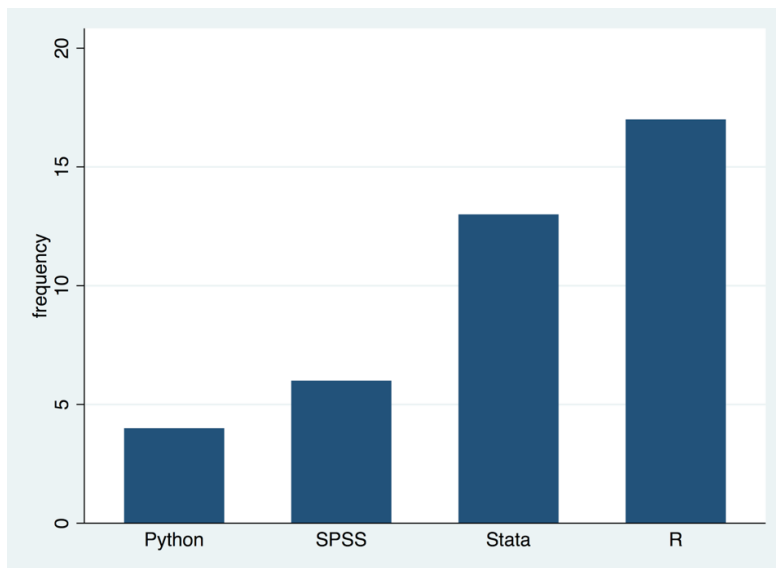
حالا میتوانید که برای هر جواب یک متغیر جداگانه ایجاد کنید. مانند متغیرات stata, r, python و spss و یا هم میتوانید آنرا به ترتیب ذکر هر پاسخ ذخیره کنید. مانند متغیرات response_1 و response_2.

در صورت که دیتا به شکل اولی ذخیره شده باشد پس به سادگی میتوانید هر متغیر را بر روی بار گراف توسط فرمان graph bar نشان دهید اما در صورت که دیتای خام به شکل دومی ذخیره شده باشد برای نمایش آن در بارگراف باید از فرمان mr-graph استفاده کنید. ساختار فرمان ذیل به نحوی است که اولاً bar mrgraph نوشته کرده به تعقیب آن نام های متغیرات را نوشته کنید و در اخیر فرمان کلمه ی , poly را نوشته کنید. به طور مثال در صورت که خواسته باشید دیتای فوق را در بارگراف نشان دهید پس میتوانید از فرمان ذیل استفاده کنید.

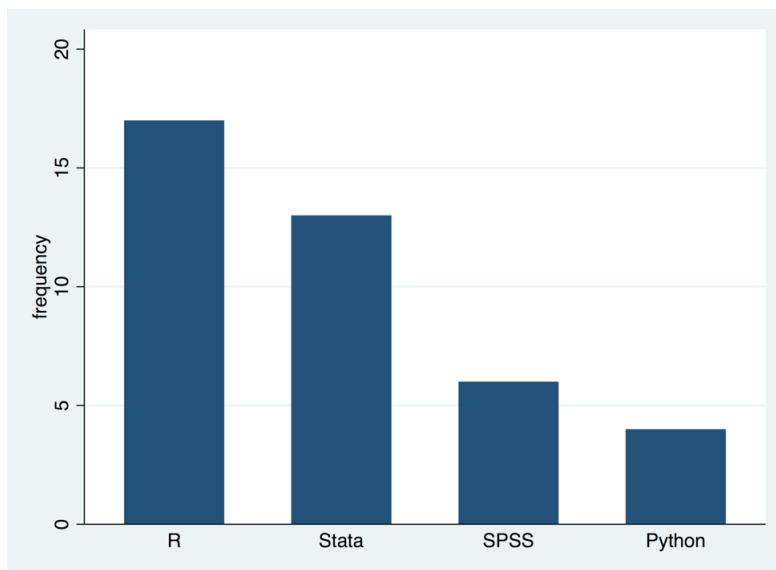
mrgraph bar response_1 response_2, poly



برای اینکه کتگوری های از کمترین به بزرگترین ترتیب شود میتوانیم در اخیر فرمان فوق کلمه ی sort را اضافه کنیم.
`mrgraph bar response_1 response_2, poly sort`

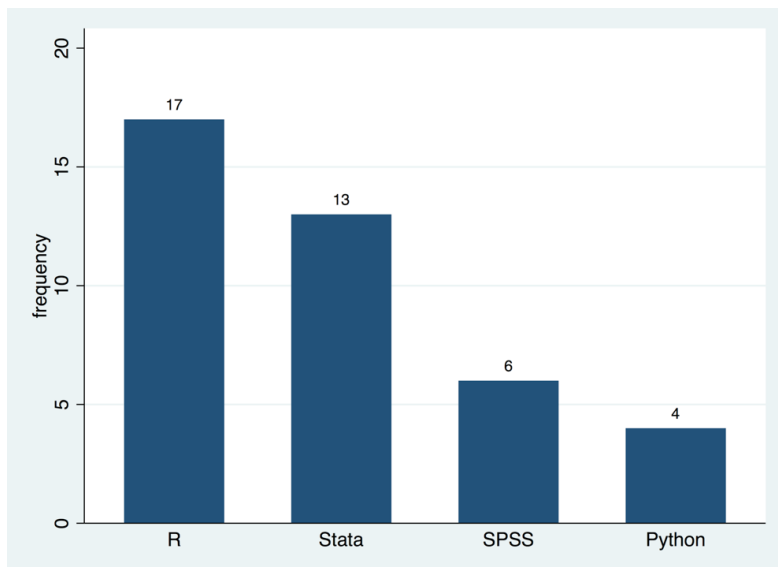


در صورت که خواسته باشید از بزرگترین به کوچکترین ترتیب گردد میتوانید در اخیر فرمان فوق sort des نوشته کنید.
`mrgraph bar response_1 response_2, poly sort des`



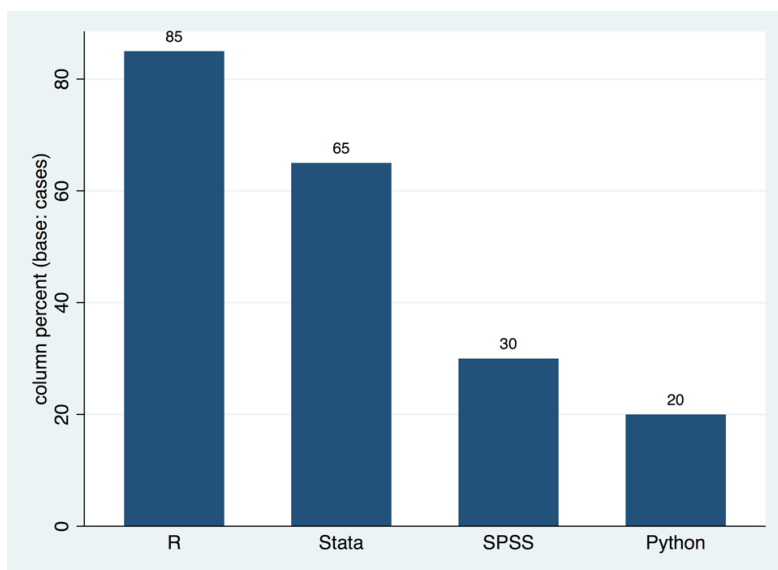
همچنان برای اینکه فرکونسی مربوطه هر کتگوری نمایش داده شود میتوانید که از گزینه balebel استفاده کنید.

```
mrgraph bar response_1 response_2, poly sort des blabel(bar, format(%9.0f))
```



در صورت که خواسته باشید فیصدی مربوطه ی هر کتگوری نشان داده شود می‌توانید در اخیر فرمان فوق stat(column) نوشته کنید.

```
mrgraph bar response_1 response_2, poly sort des blabel(bar, format(%9.0f)) stat(column)
```



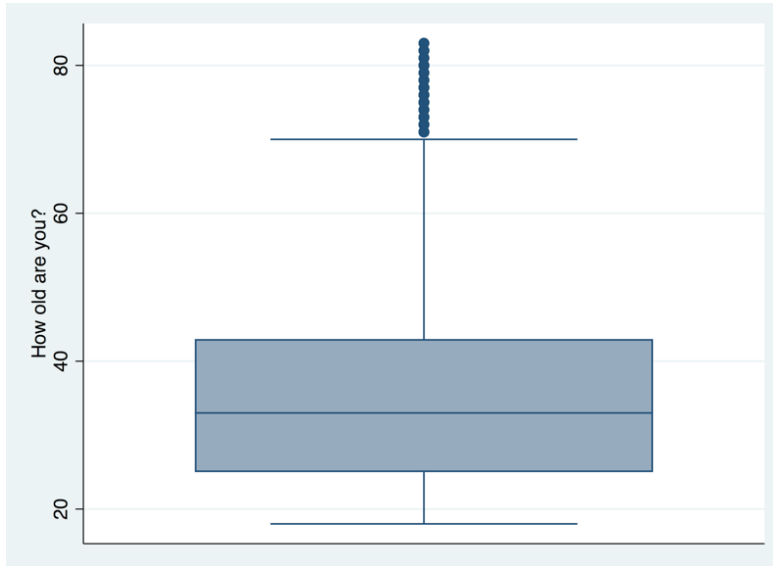
37. موضوعات عمومی مربوط به گراف

در دروس قبلی در مورد اینکه از کدام گراف برای نشان داده کدام نوع دیتا استفاده شود بحث صورت گرفت که ترسیم نمودن هر گراف همراه با فرمان های مشخص آن بیان گردید. البته همانطور که برای هر گراف فرمان مشخص خودش وجود دارد به همان ترتیب برای تغییرات در هر گراف گزینه های مشخص نیز وجود دارد. به طور مثال برای این که دیتای مربوطه ی هر کتگوری نشان داده شود میتوانید از گزینه ی blabel () استفاده کنید اما برای انجام این کار در شکل احصایوی پای چارت باید که از گزینه ی plabel () استفاده شود.

برعلاوه اینکه برای هر گراف گزینه ها مشخص خودش وجود دارد به همین ترتیب بعضی از گزینه های وجود دارد که در اکثریت گراف ها کاربرد دارد مانند اضافه نمودن عنوان در گراف و اضافه نمودن نوت در گراف و یا هم ذخیره نمودن گراف و امثال آن که در این درس روی آن بحث میگردد.

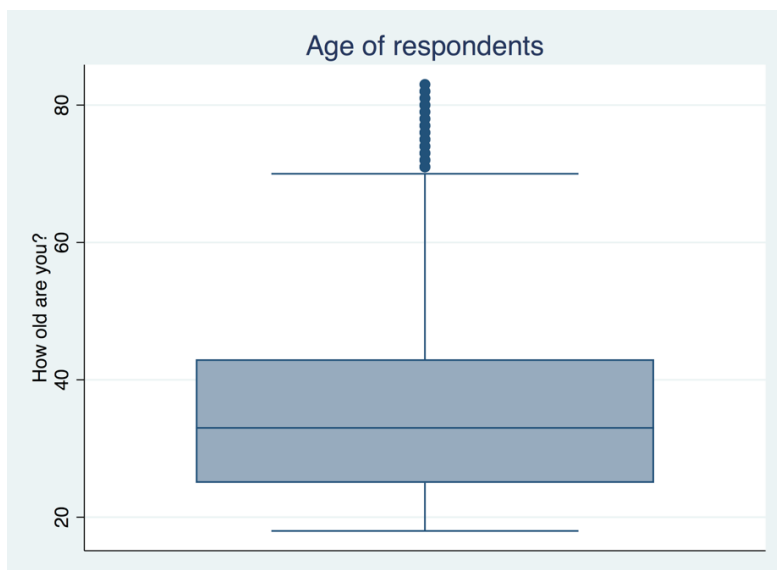
برای اضافه نمودن عنوان در گراف میتوانید از گزینه ی title () استفاده کنید. به طور مثال در سروی مردم افغانستان برای نمایش دیتای سن افراد سروی شده z2 بر روی باکس پلات میتوانید از فرمان ذیل کار بگیریم.

graph box z2



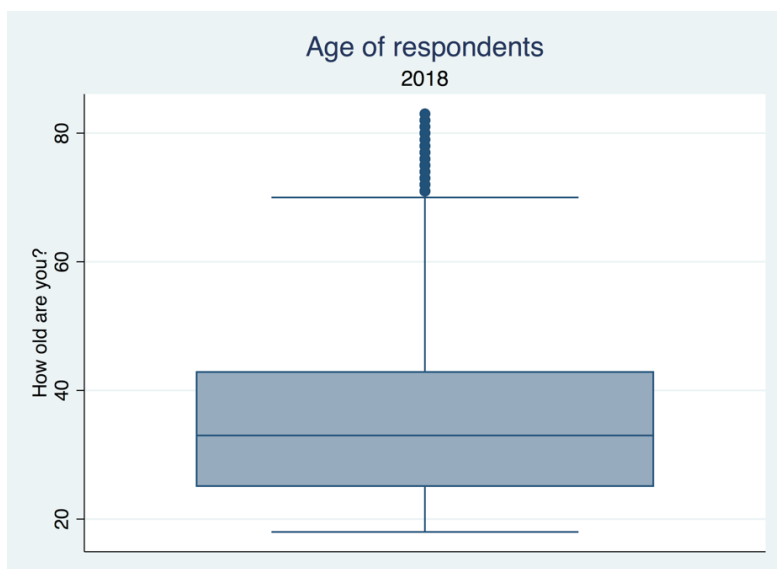
حالا برای اضافه نموده عنوان در گراف فوق میتوانید از فرمان ذیل کار بگیریم.

```
graph box z2, title("Age of respondents")
```



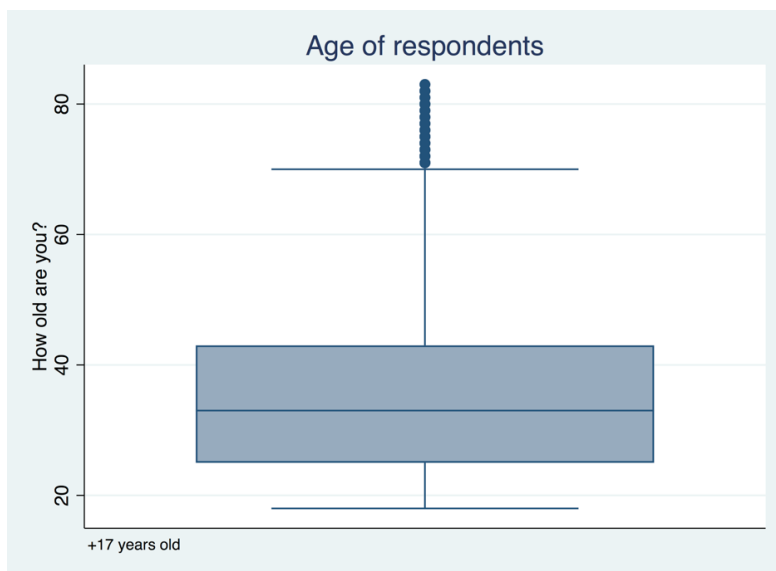
همچان برای اضافه نموده عنوان فرعی میتوانید از گزینه ی subtitle () استفاده کنید.

```
graph box z2, title("Age of respondents") subtitle("2018")
```



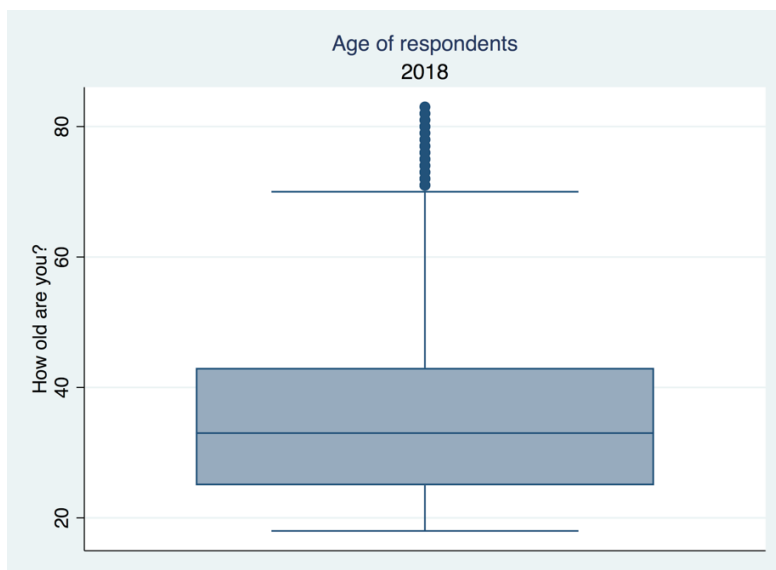
به همین ترتیب برای اضافه کردن یک نوت مشخص بر روی گراف میتوانید از گزینه ی note () استفاده کنید.

```
graph box z2, title("Age of respondents") note("+17 years old")
```



همچنان میتوانید از طریق گزینه ی size () اندازه عنوان اصلی گراف و عنوان فرعی و همچنان اندازه ی نوت را تغییر دهید. به طور مثال در صورت که خواسته باشید در گراف فوق اندازه عنوان عمومی و عنوان فرعی یک برابر باشد میتوانید از فرمان ذیل استفاده کنید.

```
graph box z2, title("Age of respondents", size(medium)) subtitle("2018", size(medium))
```

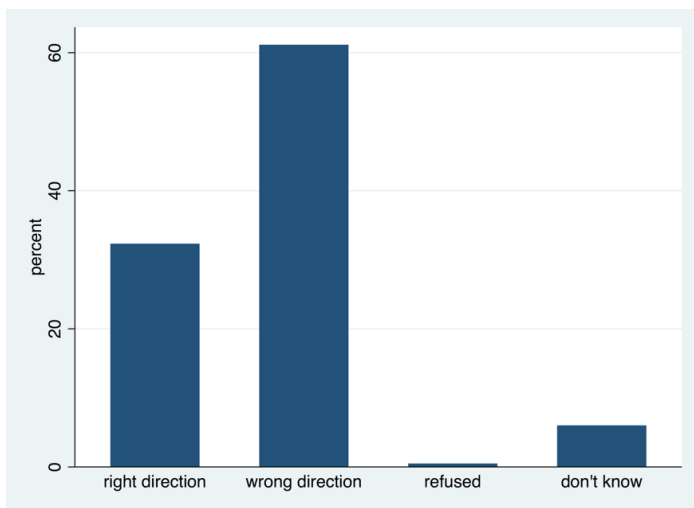


برای معلومات زیاد در مورد تعیین سایز یا اندازه نوشته های فوق میتوانید در قسمت فرمان ها `help textsizestyle` نوشته کنید و دیگر معلومات اضافی را در آنجا به خوانش بگیرید.

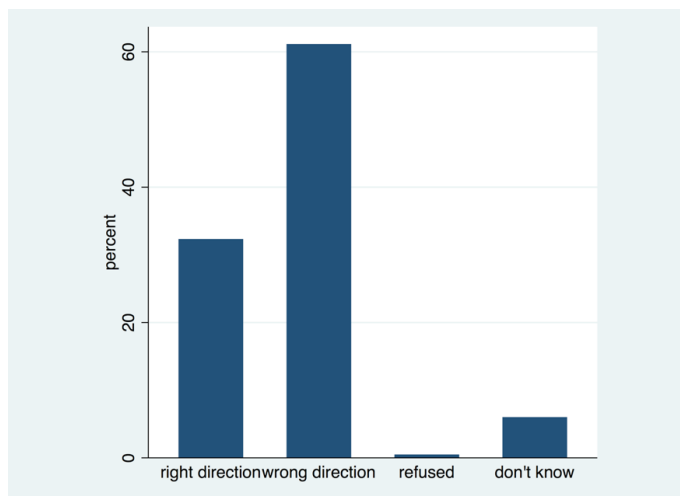
گزینه ی دیگر که در تمام فرمان های که برای ترسیم گراف های احصایوی به کار میرود قابلیت استفاده را دارد عبارت از `aspectratio` () میباشد.

در صورت که در داخل گزینه فوق عدد 1 را نوشته کنید به این معنی است که عرض و طول گراف به یک اندازه باشد. زمان که یک عدد بزرگتر از 1 را نوشته کنید به این معنی است که عرض گراف بزرگتر از طول آن باشد و در صورت که یک عدد کوچکتر از 1 را نوشته کنید به این معنی است که طول گراف بزرگتر از عرض آن میباشد.

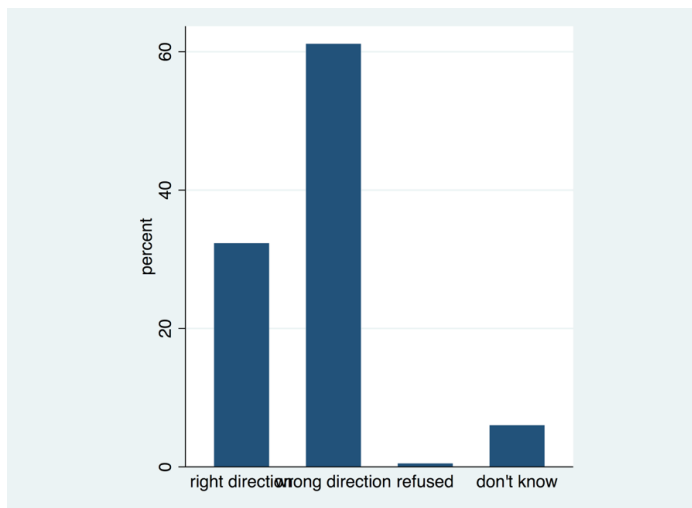
`graph bar, over(x4)`



`graph bar, over(x4) aspectratio(1)`



`graph bar, over(x4) aspectratio(0.5)`



بعد از این که یک گراف را ترسیم کردید شاید اولین سوال که در ذهن تان بیاید این است که چطور آنرا ذخیره (save) کنید. برای ذخیره کردن گراف می‌توانید از فرمان `graph export` استفاده کنید. طریقه ی استفاده از این فرمان به نحوی است که اولاً `graph export` نوشته و به تعقیب آن نام گراف مربوطه را همراه با فارمت دلخواه تان می‌توانید نوشته کنید. به طور مثال در صورت که خواسته باشید گراف فوق را به نام `direction` به فارمت `pdf` و `png` ذخیره کنید پس می‌توانید از فرمان های ذیل استفاده کنید.

`graph export direction.pdf`

`graph export direction.png`

برای اینکه بدانید گراف های ترسیم شده در ستاتا به کدام فارمت ها ذخیره شده می‌تواند پس می‌توانید در بخش فرمان ها `help graph export` نوشته کنید.



The Asia Foundation

www.asiafoundation.org