

رسم جدول فراوانی و محاسبه ی شاخص های آماری

(Eviwes)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

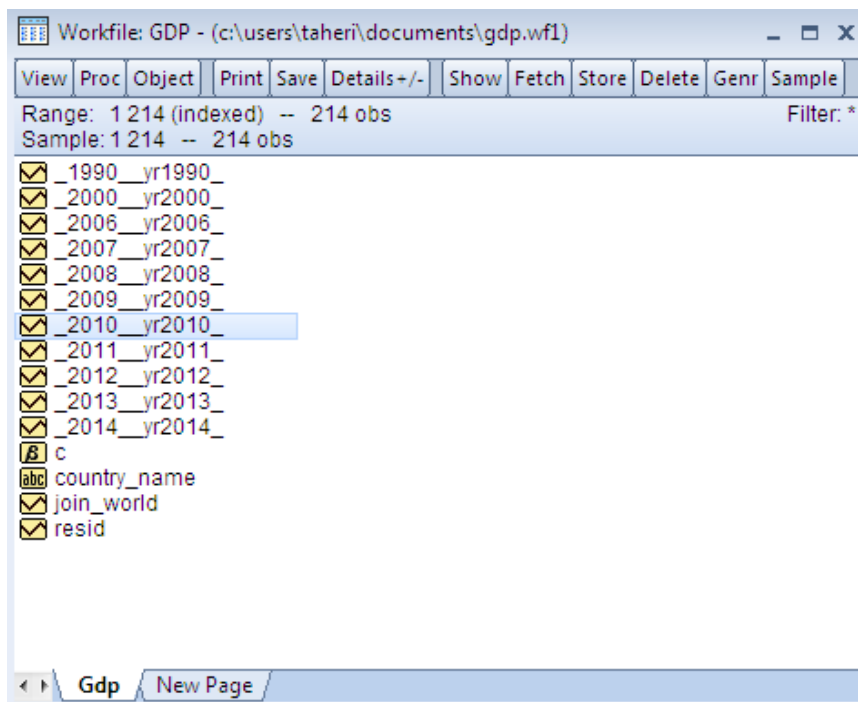
مرکز آماری خوارزمی

مقدمه:

یکی از مواردی که در خروجی های آماری مورد استفاده قرار می گیرد استفاده از جدول توزیع فراوانی داده ها است. در این جدول فراوانی و فراوانی تجمعی هر داده و درصد های آن نمایش داده می شود. در ادامه چگونگی رسم جدول فراوانی داده و بدست آوردن شاخص های مرکزی و پراکندگی داده ها در نرم افزار eviews را آموزش می دهیم.

محاسبه ی شاخص های آماری

برای انجام مراحل که در ادامه بیان میشود از داده های مربوط به GDP کشورهای مختلف از سالهای ۱۹۹۰ تا ۲۰۱۴ استفاده شده است. از سال ۲۰۰۶ به بعد داده ها به صورت سالانه بیان شده است و سال های ۲۰۰۰ و



۱۹۹۰ نیز به داده های مورد بررسی اضافه شده است. تعداد کشورهای مورد بررسی (مشاهدات) ۲۱۴ است و ستون دیگری با عنوان join_world نیز وجود دارد. این ستون دارای دو کارکتر ۱ و ۲ است؛ که ۱ عملکرد کشور متناسب با برنامه های بانک جهانی و ۲ عدم عملکرد مطابق با برنامه های بانک جهانی است (داده های این ستون فرضی است و متناسب با واقعیت نیست).

یکی از روش ها برای محدود کردن داده ها در این نرم افزار نوشتن فرمان زیر است.

```
smpl If join_world=1
```

با نوشتن این فرمان در قسمت پنجره ی فرمان داده هایی که متغیر `join_worl` برابر ۱ است انتخاب می کند و محاسبات را با توجه به محدودیت اعمال شده انجام می دهد؛ و در قسمت `obs` تعداد مشاهدات را نمایش می دهد. برای بازگشت به حال قبلی هم می توان از فرمان زیر استفاده کرد.

`smpl @all`

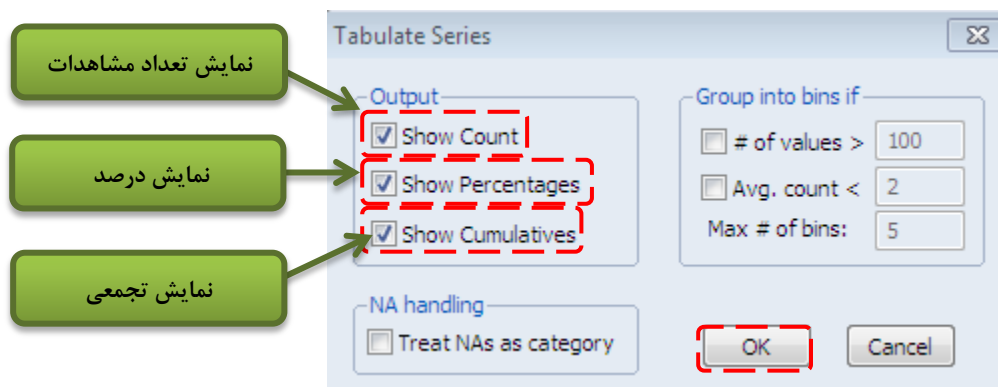
برای بدست آوردن جدول توزیع فراوانی مراحل زیر را انجام بدهید. در ابتدا بر روی متغیری که قصد داریم جدول توزیع آن را رسم کنید دبل کنید تا صفحه ای مشابه زیر ظاهر گردد.

2009 [YR2009]	
Last updated: 12/15/15 - 19:35	
Imported from 'C:\Users\taheri\Desktop\gdp.xlsx'	
Excel name: 2009 [YR2009]	
Afghanistan	1.25E+10
Albania	1.20E+10
Algeria	1.37E+11
American Samoa	NA
Andorra	3.65E+09
Angola	7.55E+10
Antigua and Barbuda	1.21E+09
Argentina	3.79E+11
Armenia	8.65E+09
Aruba	2.50E+09

همانطور که در تصویر روبرو دیده می شود داده های مربوط به سال ۲۰۰۹ انتخاب شده است؛ و داده های مربوط به GDP برای تمام مشاهدات نمایش داده شده است. مسیر زیر را طی نمایید.

View > One_Way Tabulation

پس از انجام این مراحل پنجره ی زیر باز می شود.



گزینه هایی مربوط به نمایش تعداد مشاهدات، نمایش درصد مشاهدات، نمایش تجمعی مشاهدات را انتخاب کرده و بقیه آیکن ها را غیر فعال می کنیم. با زدن دکمه ی `ok` صفحه ی مشابه زیر باز می شود که جدول فراوانی مشاهدات مربوط به سال ۲۰۰۹ را نمایش می دهد.

Series: _2009_YR2009_ Workfile: GDP::Gdp\

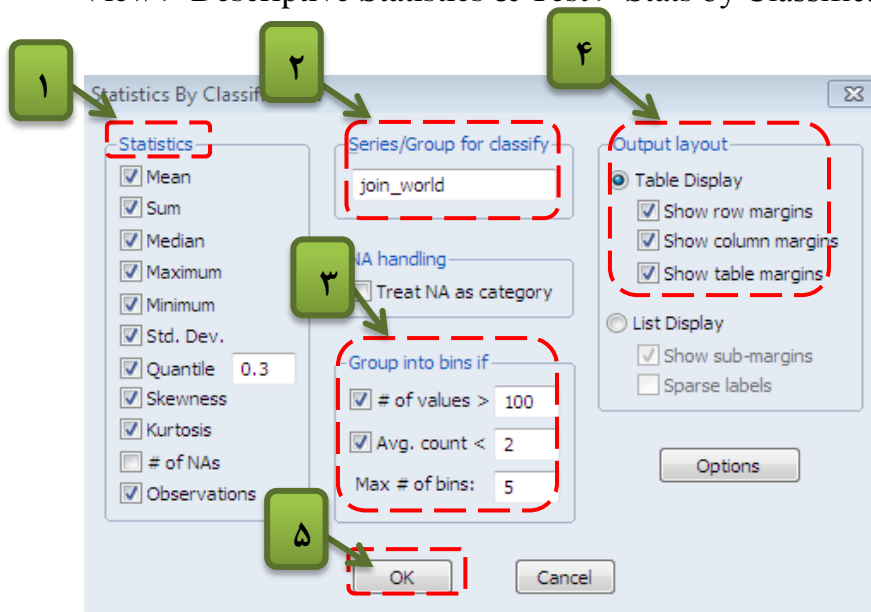
View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stats I

Tabulation of _2009_YR2009
Date: 12/15/15 Time: 20:46
Sample: 1 214
Included observations: 196
Number of categories: 196

Value	Count	Percent	Cumulative Count	Cumulative Percent
27101313	1	0.51	1	0.51
1.27E+08	1	0.51	2	1.02
1.52E+08	1	0.51	3	1.53
1.87E+08	1	0.51	4	2.04
1.96E+08	1	0.51	5	2.55
2.78E+08	1	0.51	6	3.06
3.21E+08	1	0.51	7	3.57
4.99E+08	1	0.51	8	4.08
5.24E+08	1	0.51	9	4.59
5.86E+08	1	0.51	10	5.10
5.98E+08	1	0.51	11	5.61
6.10E+08	1	0.51	12	6.12

در ادامه تحلیل آماری دیگری مبتنی بر دسته بندی مشاهدات را انجام می دهیم. از منوی view در صفحه ی مربوط به داده های سال ۲۰۰۹ استفاده می کنیم. مسیر زیر را طی کنید.

View > Descriptive Statistics & Test > Stats by Classification ...



صفحه ی روبرو باز می شود.

در این صفحه شاخص هایی که برای تحلیل کاربر مورد نیاز است قرار داده شده است. با فعال کردن هر یک از این آماره ها محاسبه توسط نرم افزار انجام می شود.

۱. آماره ها موجود در قسمت statistics به ترتیب به شرح زیر است: میانگین،

مجموع، میانه، بیشینه، کمینه، انحراف استاندارد، چارک، چولگی، کشیدگی، #of NAs، مشاهدات است. در قسمت چارک به صورت پیش فرض عدد ۰.۵ قرار داده شده است که چارک دوم را محاسبه

- می کند و با تغییر عدد داخل مستطیل می توان دهک ها و صدک های مورد نظر را برای محاسبه به نرم افزار داد. در اینجا با زدن عدد ۰.۳ فرمان محاسبه دهک سوم به نرم افزار داده شده است.
۲. عبارت `join_world` در کادر `series/group for classify` نوشته تا داده ها با انتخاب سری فعال وابسته به این سری تشکیل شود.
۳. در قسمت `group info bins if` اعداد ۱۰۰، ۲ و ۵ را قرار می دهیم. این اعداد به این معنی است که در صورت رسیدن تعداد مقادیر بیش از ۱۰۰ یا متوسط فراوانی کمتر از ۲ گروه هایی از مشاهدات تشکیل می شود که تعداد این گروه های نباید از ۵ تجاوز کند.
۴. در بخش `output layout` نمایش خروجی داده ها به صورت جدول و نمایش تمامی حاشیه های ردیفی و ستونی است که آنها را فعال می کنیم.
۵. در پایان با زدن دکمه `ok` جدول زیر که محاسبه تمامی آماره های مدنظر است نمایش داده شده است.

Series: _2009_YR2009_ Workfile: GDP::Gdp\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stats Ident

Descriptive Statistics for _2009_YR2009_
Categorized by values of JOIN_WORLD
Date: 12/15/15 Time: 21:05
Sample: 1 214
Included observations: 196

JOIN_WOR...	Mean	Median	Max	Min.	Quant.*	Sum.	Std. Dev.	Skew.	Kurt.	Obs.
1	2.37E+11	1.28E+10	5.06E+12	27101313	6.61E+09	2.20E+13	7.98E+11	5.143681	30.06057	93
2	3.61E+11	2.85E+10	1.44E+13	1.27E+08	6.86E+09	3.72E+13	1.49E+12	8.329795	77.73755	103
All	3.02E+11	2.04E+10	1.44E+13	27101313	6.79E+09	5.92E+13	1.21E+12	8.890026	97.17649	196

*Quantiles computed for p=0.3, using the Rankit (Cleveland) definition.

نتیجه، جدولی از سال ۲۰۰۹ است که بر اساس تطابق با برنامه ی جهانی و عدم تطابق دسته بندی شده اند. ۱ تطابق و ۲ عدم تطابق با برنامه های بانک جهانی است. در اینجا تعداد نمونه ها با تعداد مشاهدات متفاوت است که حکایت از تعدادی داده ی گم شده (`missing`) است. در اینجا می توان میانگین، میانه، ... داده ها به تفکیک اعداد برای تطابق و عدم تطابق با برنامه های بانک جهانی محاسبه شده است.

بدین ترتیب آماره های مورد نظر محاسبه می شود.

منبع :

www.behkaman.ir موسسه فرهنگی دیجیتال بهگامان