پروژه پایان ترم درس:

آمار و احتمال مهندسی

استاد راهنما: دکتر امیری

اعضای گروه:

علی نادری

سامان بیک محمدی

صادق مومنی نژاد

بسم الله الرحمن الرحيم

با سپاس فراوان از دکتر

امیری و مهندس الله یاری که

در آماده سازی این پروژه

از هیچ کمکی دریغ نکردند...

فهرست مطالب:

مقدمه..... ۱

نمونه های تصادفی..... ۲

آمار توصیفی..... ٤

نمودار های نقطه ای..... ٥

نمودار های هیستوگرام..... ٦

کشیدگی و چولگی..... ۸

نمودار های جعبه ای(box plot) ۹

آمار استنباطی(استنتاجی) ۱۱

آزمون نرمال(normal test) ۱۱

آزمون فرض..... ۱۳

. آزمون فرض بر روی میانگین جوامع..... ۱۳

آزمون فرض یک طرفه روی میانگین..... ۱۳

آزمون فرض دو طرفه روی میانگین..... ۱٤

آزمون فرض برابری میانگین دو جامعه..... ۱٥

آزمون فرض دو طرفه بر روی میانگین دو جامعه..... ۱٦

آزمون فرض یک طرفه بر روی میانگین دو جامعه ۱۷

آزمون برابری واریانس های دو جامعه..... ۱۸

نمودار **Interval Plot**..... ۱۹

نمودار **Interval Plot**..... ۲۰

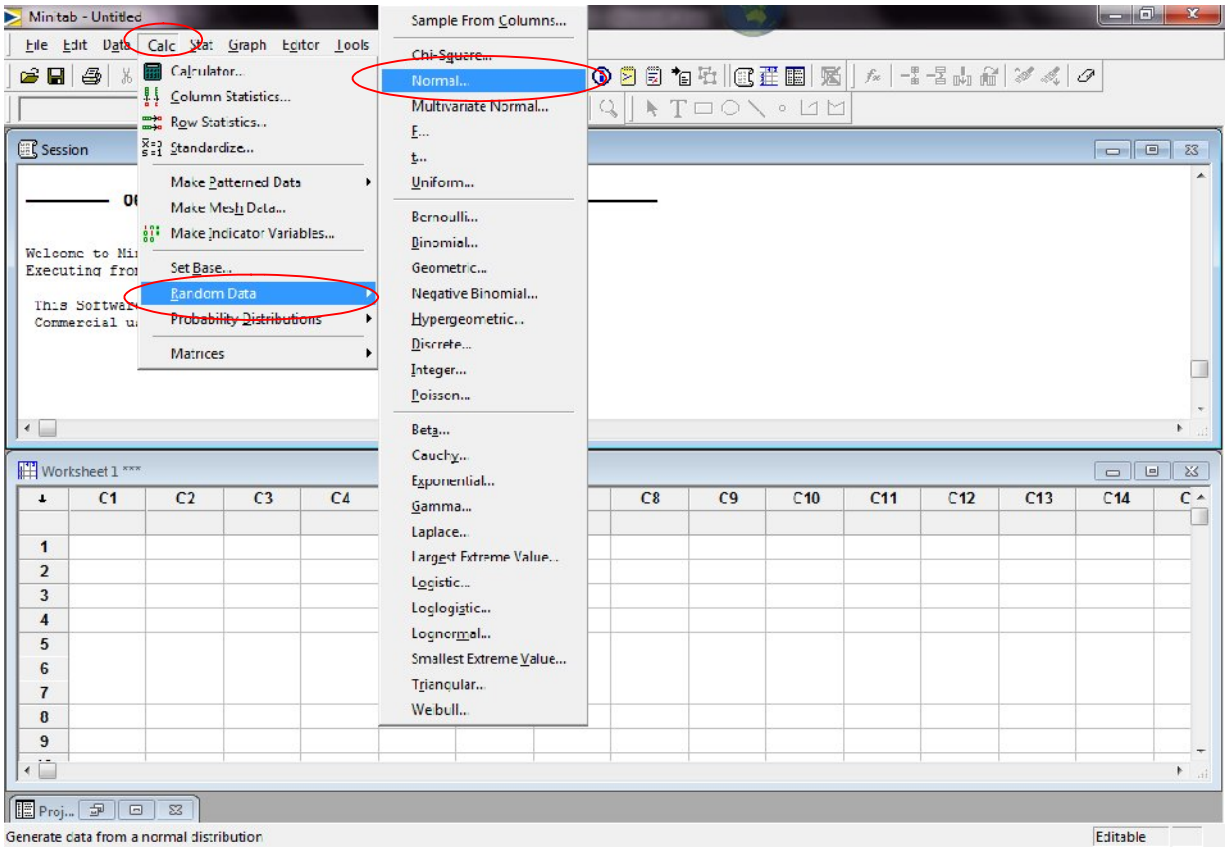
مقدمه

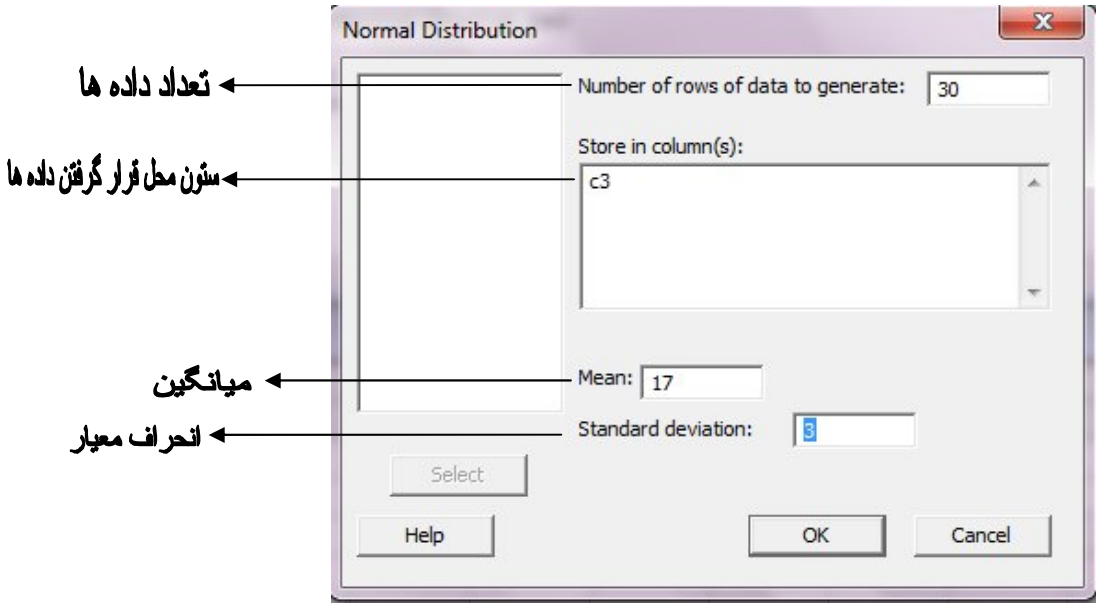
محتویات این پروژه به دو سر فصل کلی آمار توصیفی و آمار استنباطی تقسیم می شود.آمار توصیفی شاخه ای از آمار است که به سازماندهی ، تجزیه و تحلیل ،خلاصه سازی و نمایش داده ها می پردازد ؛ بطوریکه از این طریق بتوان اطلاعات مفید و قابل استفاده ای از مشاهدات کسب نمود.قدمت این دانش از آمار به ۲۰۰ سال پیش باز می گردد و بدلیل سادگی و عدم نیاز به فرضیات محدود کنده کاربرد های فراوانی در علوم مختلف دارد.از کاربرد این شاخه از آمار در این مقال می توان به نمودار های گرافیکی مانند هیستوگرام ، نقطه ای و جعبه ای و همچنین به بررسی معیار های شکل تابع اشاره کرد.در مقابل آمار توصیفی ، شاخه ی دیگری از آمار وجود دارد که به آمار استنباطی یا استنتاجی مشهور است.این شاخه از آمار به استنباط ، نتیجه گیری و قضاوت درباره یک جامعه آماری از طریق مطالعه و بررسی یک نمونه کوچکتر از آن می پردازد.از مباحث مورد استفاده از آمار استنباطی در این پروژه می توان به آزمون های فرض یک طرفه و دو طرفه و آزمون نرمال بودن جوامع ، همچنین مقایسه واریانس و میانگین دو جامعه اشاره کرد.

تذکر: با توجه به اینکه داده های نمونه تنها اطلاعات ناقصی در مورد جامعه ارائه می دهند ، در بسیاری از مراجع پیشنهاد می شود از بکار بردن عبارت" پذیرش فرض صفر" خودداری شود و در صورت عدم رد فرض صفر ، بیان شود که دلیلی برای رد فرض صفر وجود ندارد.لذا در این پروژه نیز همین عبارت بکار برده شده است.

در گام اول به تعریف سه جامعه آماری بصورت تصادفی می پردازیم.برای این منظور سه جامعه ۳۰ تایی را مطابق تصویر تعریف کرده و در ستون های C۱ و C۲ و C۳ قرار می دهیم.میانگین این داده ها به ترتیب عبارت است از ۱۵ و ۱۶ و ۱۷ و انحراف معیار آن ها عبارت است از ۲ و ۳ و ٤ :

calc\Random Data\Normal





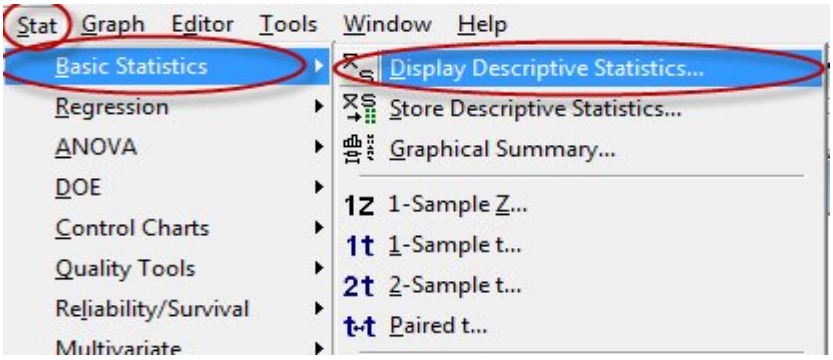
داده ها به قرار زیر است:

↓	C1	C2	C3
1	11.2388	17.5534	25.5288
2	16.6933	17.5782	16.4936
3	17.3296	14.7675	10.4075
4	13.7083	14.8586	16.4427
5	18.8693	13.8422	15.6050
6	17.1097	17.5955	19.3627
7	11.4144	16.9961	16.6141
8	16.2192	16.0307	20.5658
9	11.3738	21.3704	18.5992
10	13.1936	18.3103	17.3577
11	13.4527	20.9081	17.6953
12	14.4070	15.9959	16.8701
13	15.9115	14.7636	15.9773
14	15.9701	11.4594	17.9013
15	14.6668	18.1456	18.4740
16	16.4110	15.8223	15.8830
17	18.3484	19.2660	15.0117
18	14.4006	14.3482	20.9204
19	15.9904	20.2253	14.3102
20	10.8010	12.6747	17.8353
21	12.0492	16.0568	16.6405
22	16.4243	12.2830	14.5236
23	16.8316	18.0918	19.2828
24	11.3434	14.4513	22.6819
25	10.7474	13.4302	17.0231
26	12.6469	19.9353	22.1583
27	17.2869	13.6497	16.5067
28	12.6065	15.8311	18.1336
29	17.7456	8.3204	17.4765
30	15.6843	19.2739	14.6693

آمار توصیفی:

همانطور که در مقدمه اشاره شد در این قسمت به بررسی جامعه از طریق تعدادی از نمودار های آماری می پردازیم.اما ابتدا اطلاعات مربوط به آمار توصیفی را از مسیر زیر بدست می آوریم:

Stat\Basic Statistics\ Display Descriptive Statistics



اطلاعات بدست آمده به شرح زیر می باشد:

Descriptive Statistics: C۱; C۲; C۳

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q۱	Median	Q۳
C۱	۳۰	۰	۱٤.٦۹٦	۰.٤٥۳	۲.٤۸۰	۱۰.۷٤۷	۱۲.٤٦۷	۱٥.۱۷٦	۱٦.۷۲۸
C۲	۳۰	۰	۱٦.۱۲۸	۰.٥٤۷	۲.۹۹۸	۸.۳۲۰	۱٤.۲۲۲	۱٦.۰۱۳	۱۸.۱۸۷
C۳	۳۰	۰	۱۷.٥٦٥	۰.٥۲٦	۲.۸۸۲	۱۰.٤۰۸	۱٥.۹٥٤	۱۷.۱۹۰	۱۸.۷۷۰

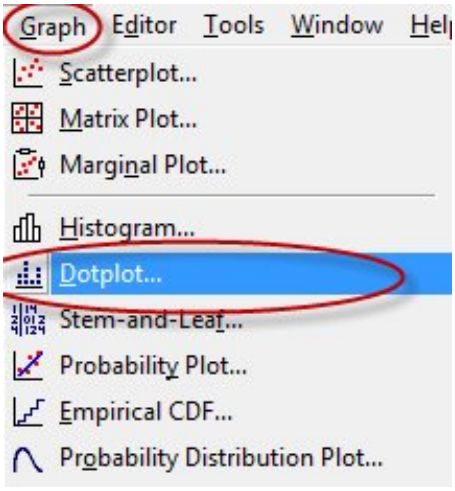
Variable	Maximum	Skewness	Kurtosis
C۱	۱۸.۸٦۹	-۰.۱۸	-۱.۲۸
C۲	۲۱.۳۷۰	-۰.۳٦	۰.۱۹
C۳	۲٥.٥۲۹	۰.٥۰	۱.۸۰

اطلاعات بدست آمده شامل میانگین(mean) انحراف معیار(StDev) کمترین داده ، چارک اول(Q۱) ، چارک سوم(Q۳) ، میانه(Median) و همچنین معیار های چولگی و کشیدگی می باشد.

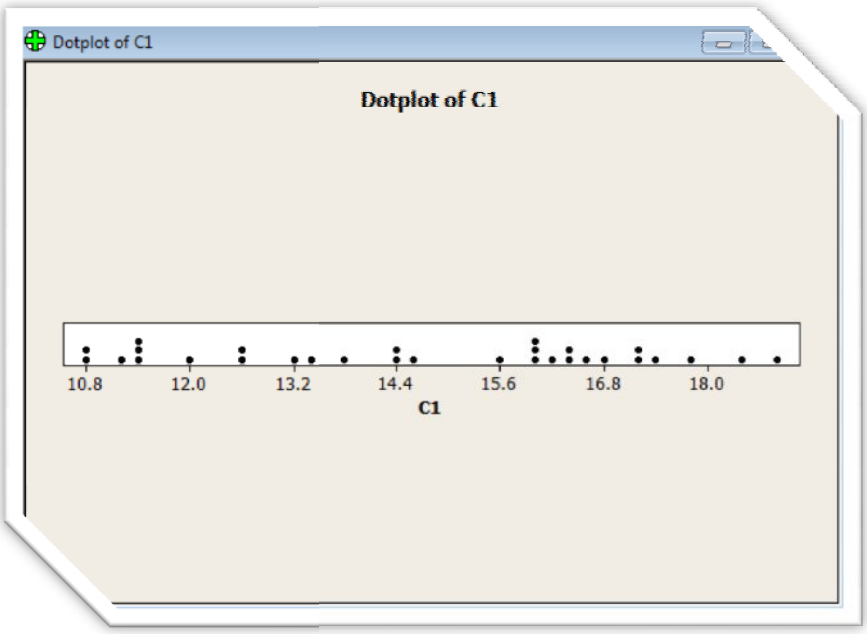
نمودار نقطه ای

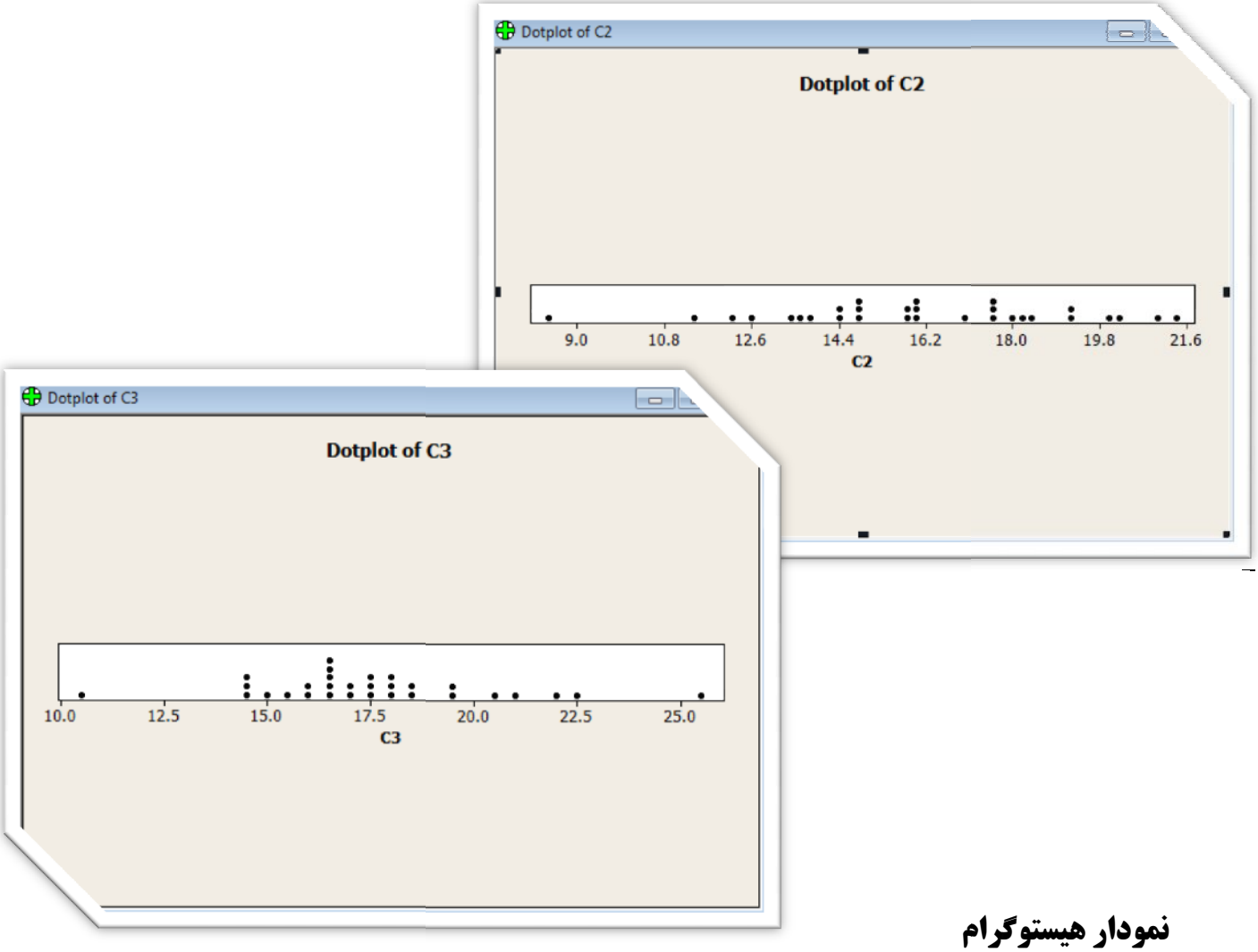
این نمودار داده ها را بصورت نقاط نمایش می دهد و برای مشاهده سریع تجمع داده ها بصورت تقریبی می تواند بسیار مناسب باشد.برای رسم مسیر زیر را مطابق تصویر دنبال می کنیم.

Graph\Dotplot



در اینجا نمودار نقطه ای سه جامعه **C۳ , C۲ , C۱** را مشاهده می کنیم:

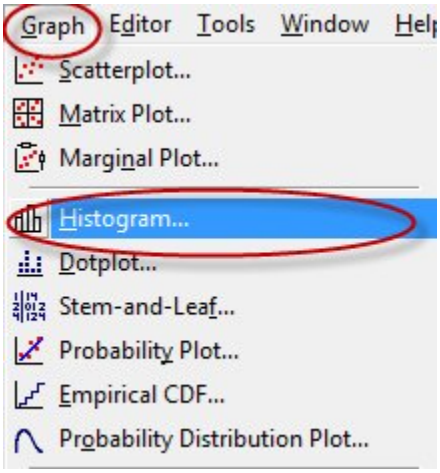




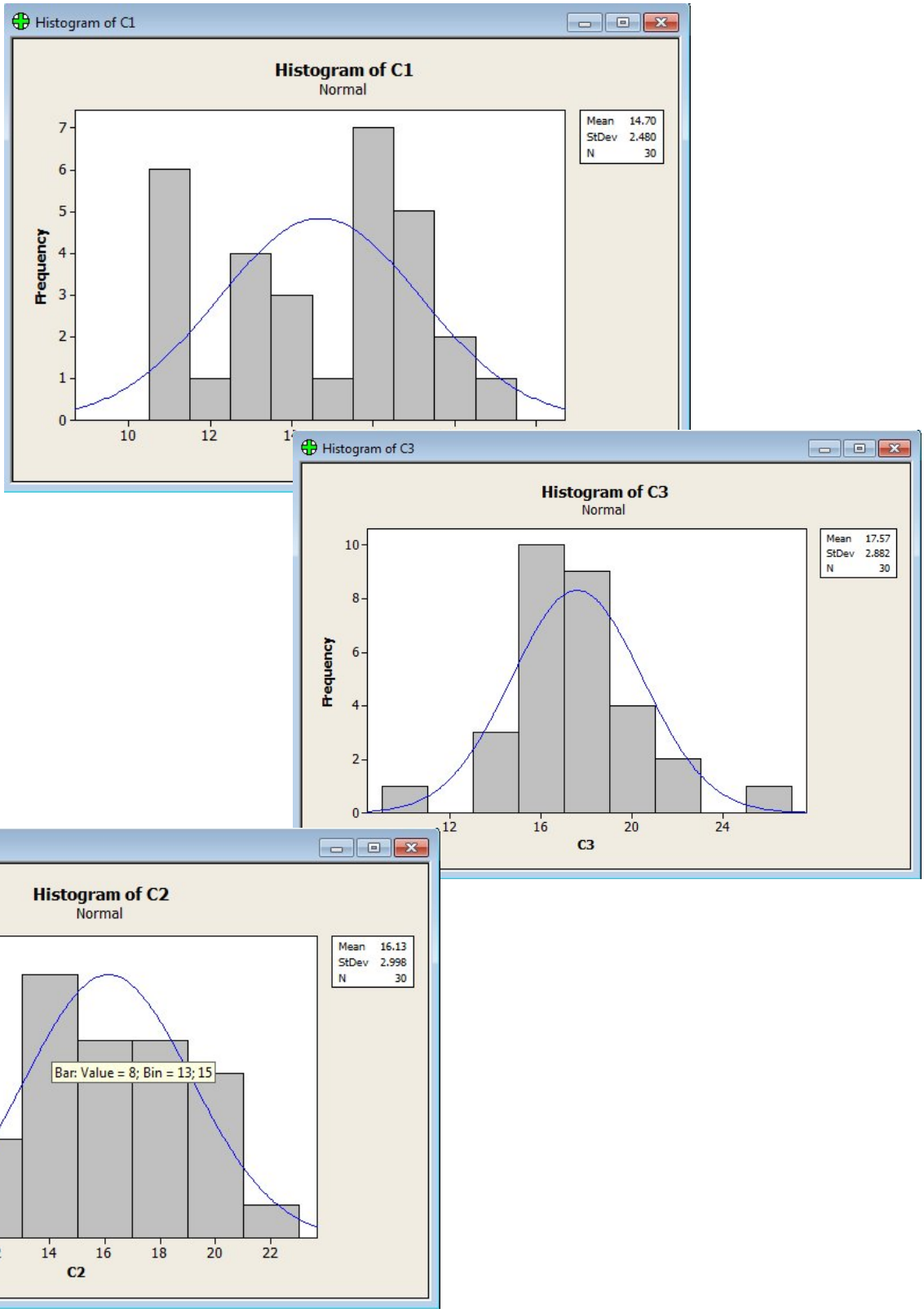
نمودار هیستوگرام

نمودار هیستوگرام یک ابزار بسیار مهم برای نمایش فراوانی داده هاست.زیرا همه ی معیار های فراوانی و پراکندگی و شکل توزیع در آن به خوبی مشخص است.برای رسم این نمودار مسیر زیر را مطابق تصویر دنبال می کنیم:

Graph\Histogram



در اینجا نمودار هیستوگرام سه جامعه C۳ , C۲ , C۱ را مشاهده می کنیم:

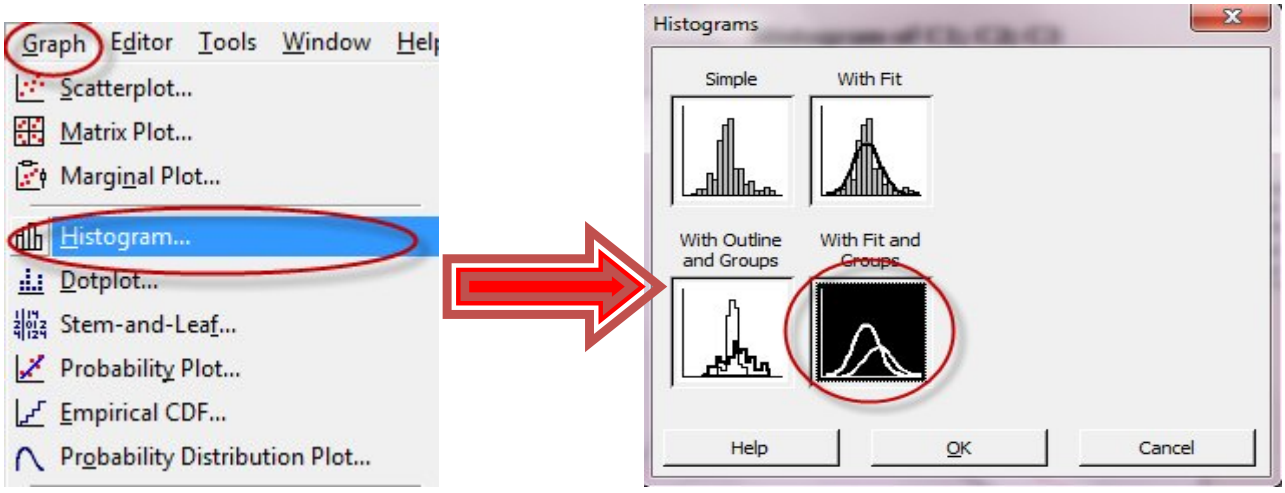


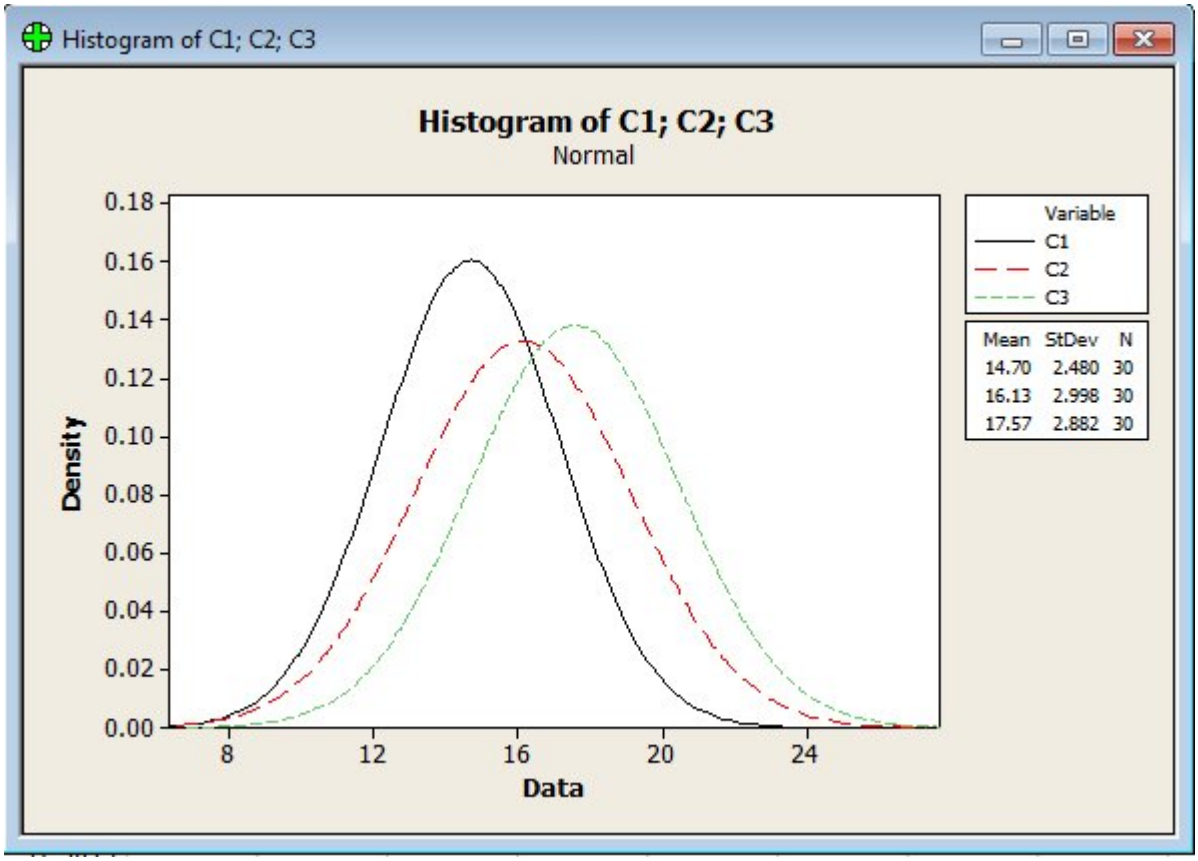
با توجه به شکل نمودار ها ، جوامع ما همانطوری که توقع داشتیم تقریبا نرمال است.در نمودار هیستوگرام ستونی که بیشترین ارتفاع را دارد نشان دهنده ی بیشترین فراوانی است.لذا می توان چنین تصور نمود که مد جامعه ما داده ی همان ستون است.برای مثال در جامعه **C۳** مد نزدیک به عدد ۱٤ است.

کشیدگی و چولگی:

چولگی و کشیدگی از مهمترین معیار های شکل تابع می باشند.برای مثال چولگی یک جامعه معیار بسیار مناسبی برای مشاهده ی قرینگی شکل جامعه است.در جامعه های نرمال چولگی برابر صفر و کشیدگی برابر ۳ است و کشیدگی سایر جوامع با نمودار نرمال مقایسه می شود.برای بررسی این دو معیار هر سه جامعه را در یک نمودار رسم می کنیم.برای این کار مسیر زیر را مانند تصویر دنبال می کنیم:

Graph\Histogram



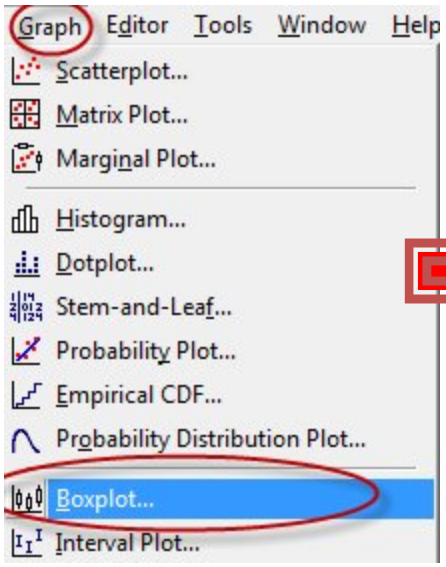


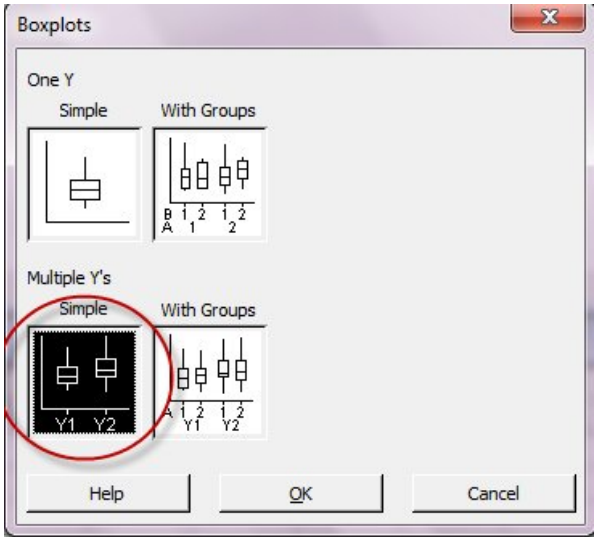
همانطور که مشاهده می شود جامعه C۱ از دو جامعه دیگر کشیده تر است و با اینکه هر سه جامعه تقریبا نرمال هستند اما چولگی C۱ کمی به راست است.

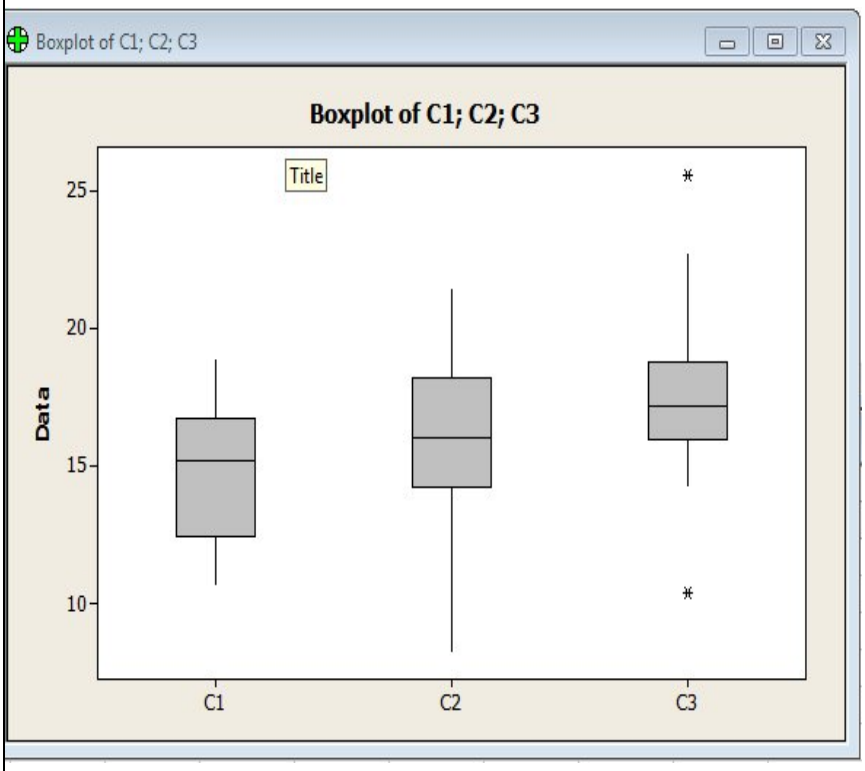
نمودار جعبه ای(box plot)

نمودار جعبه ای ابزار گرافیکی قدرت مندی است که بطور همزمان برخی از خواص مهم یک مجموعه از مشاهدات را مانند تمرکز ، پراکندگی ، میزان تقارن و حضور مشاهدات دورافتاده را بخوبی نشان می دهد .نمودار جعبه ای چارک ها و کوچکترین و بزرگترین مشاهده را در قالب یک مستطیل و دوخط کناری نمایش می دهد.در این جا برای راحتی کار و مقایسه جوامع هر سه نمودار جعبه ای را بر روی یک نمودار رسم می کنیم.برای رسم این نمودار مسیر زیر را مانند تصویر دنبال می کنیم:

Graph\Boxplot







نتیجه بصورت زیر قابل رویت است:

در نمودار مربوط به جامعه C۳ نقاط دور افتاده(outlier) به سادگی قابل مشاهده است.همچنین از این نمودار می توان چنین برداشت کرد که هر سه جامعه دارای میانه های نزدیک به هم هستند.

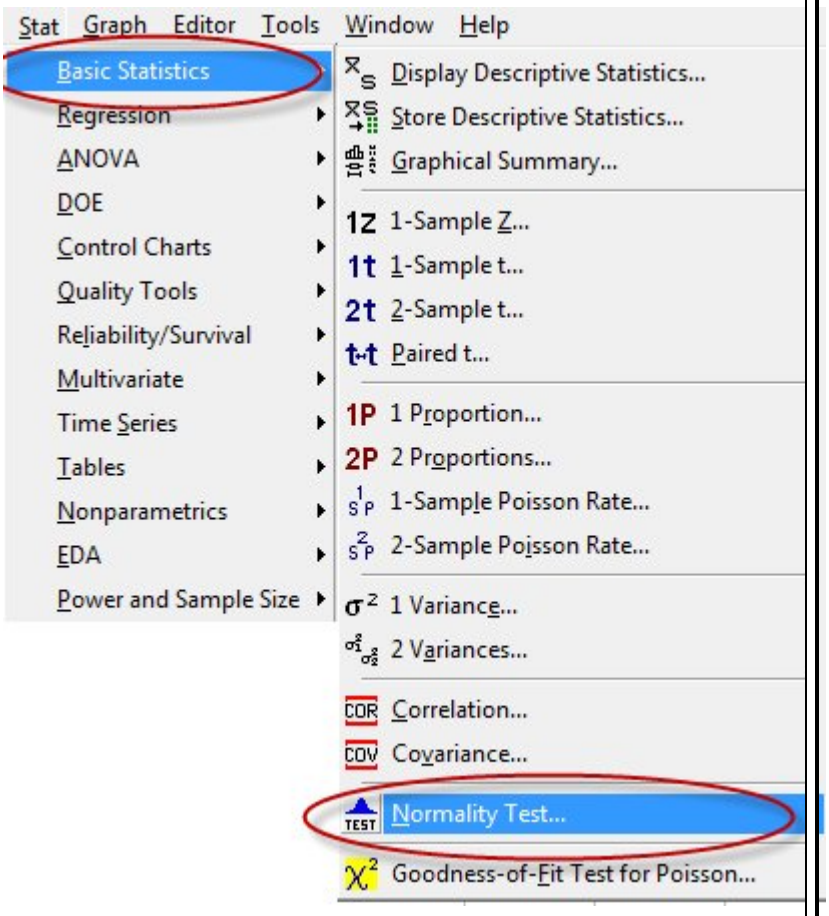
آمار استنباطی(استنتاجی)

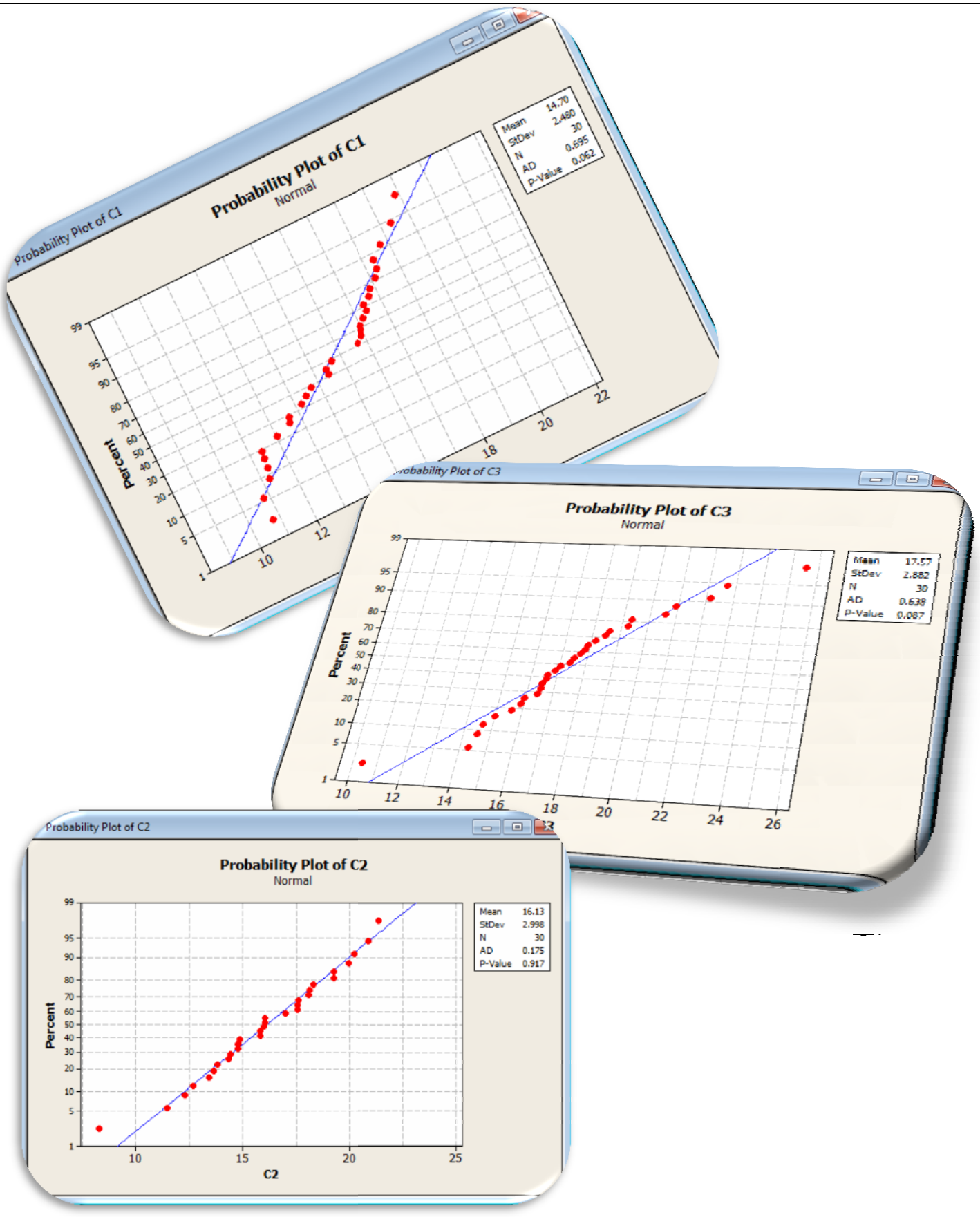
در این قسمت همانطور که در مقدمه اشاره شد با استفاده از مواردی مانند آزمون ها می خواهیم به مطالعه و نتیجه گیری روی جوامع بپردازیم.این قسمت را با مقایسه جوامع، آزمون نرمال ،آزمون فرض بر روی میانگین و آزمون فرض بر روی واریانس جوامع دنبال می کنیم.

آزمون نرمال(normal test)

با استفاده از نتایج این تست چنان که در ادامه خواهد آمد می توانیم بفهمیم که یک جامعه چقدر به جامعه ی نرمال نزدیک است.برای این تست مسیر زیر را پی می گیریم:

Stat \ Basic Statistics \ Normality Test




Probability Plot of C1
Probability Plot of C1
Normal
Mean 14.70
StDev 2.480
N 30
AD 0.695
P-Value 0.062
Percent
Probability Plot of C3
Probability Plot of C3
Normal
Mean 17.57
StDev 2.882
N 30
AD 0.638
P-Value 0.087
Percent
Probability Plot of C2
Probability Plot of C2
Normal
Mean 16.13
StDev 2.998
N 30
AD 0.175
P-Value 0.917
Percent
C2

با توجه به اینکه این نرم افزار به طور پیش فرض آزمون را در سطح اطمینان ۹۵ درصد انجام میدهد بنابراین a برابر با ۵ درصد است. و با توجه به اینکه مقدار p-value در هر سه جامعه از میزان aبیشتر شده است بنابر این دلیلی برای رد فرض صفر وجود ندارد و جوامع همانطوری که توقع داشتیم نرمال هستند.این موضوع از طریق مشاهده تصویر نیز قابل دریافت است چون نقاط تقریبا حول خط هستند.

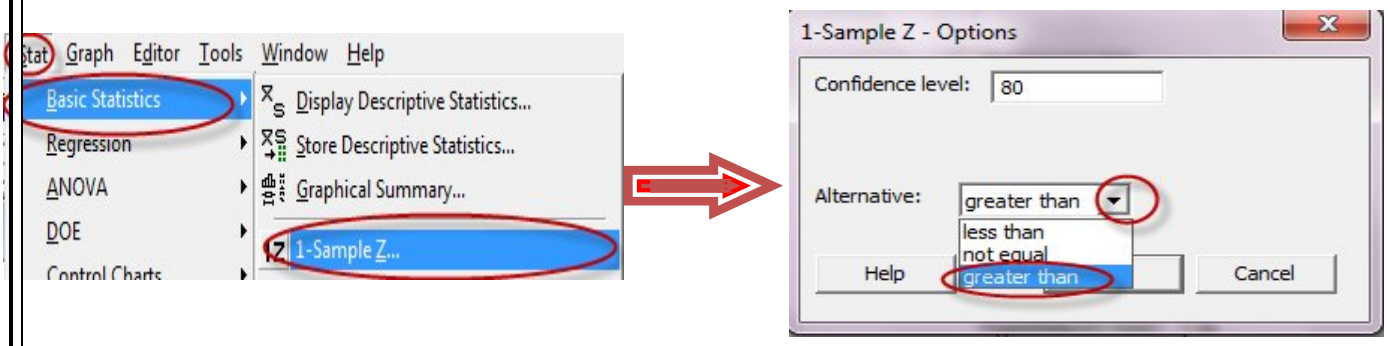
آزمون فرض بر روی میانگین جوامع

این آزمون را برای هر سه جامعه و برای دو حالت فرض یکطرفه و فرض دو طرفه در سطح اطمینان ۸۰ درصد بررسی می کنیم.برای این فاصله اطمینان α برابر با ۰.۲ می باشد.در اینجا می خواهیم بدانیم که آیا میانگین جوامع برابر عدد ۱۶ است یا خیر.

آزمون فرض یک طرفه:

مسیر زیر را پی می گیریم:

Stat \ Basic Statistics\ 1-Sample Z\options\alternative\greater than (or less than)



نتیجه آزمون به شرح زیر می باشد:

One-Sample Z: C۱; C۲; C۳

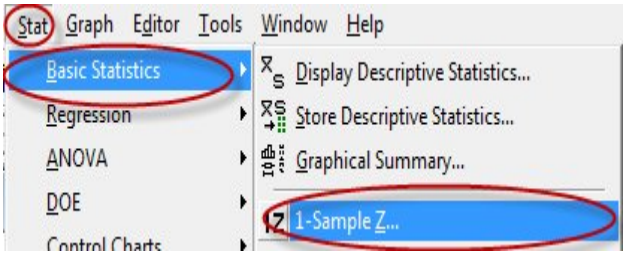
```
Test of mu = ۱۶ vs > ۱۶
The assumed standard deviation = ۲

                                                  ۸۰% Lower
Variable   N     Mean   StDev  SE Mean      Bound       Z      P
C۱        ۳۰  ۱۴.۶۹۶  ۲.۴۸۰   ۰.۳۶۵     ۱۴.۳۸۹  -۳.۰۵۷  ۱.۰۰۰
C۲        ۳۰  ۱۶.۱۲۸  ۲.۹۹۸   ۰.۳۶۵     ۱۵.۸۲۱   ۰.۳۵۰  ۰.۳۶۳
C۳        ۳۰  ۱۷.۵۶۵  ۲.۸۸۲   ۰.۳۶۵     ۱۷.۲۵۸   ۴.۲۹۰  ۰.۰۰۰
```

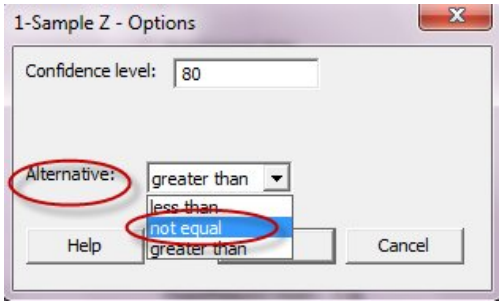
با مقایسه مقدار p-value و a می توانیم از نتیجه آزمون مطلع شویم.در جامعه اول و دوم با توجه به اینکه مقدار p-value از مقدار a بیشتر شده است دلیلی برای رد فرض صفر وجود ندارد اما در جامعه سوم چون مقدار p-value از a کمتر شده است فرض صفر رد می شود.

آزمون فرض دو طرفه:

مسیر زیر را پی می گیریم:

Stat \ Basic Statistics\ 1-Sample Z\options\Alternative\not equal





نتیجه آزمون به شکل زیر می باشد:

One-Sample Z: C۱; C۲; C۳

```
Test of mu = ١٦ vs not = ١٦
The assumed standard deviation = ٢
```

Variable	N	Mean	StDev	SE Mean	٨٠% CI	Z	P
C۱	٣٠	١٤.٦٩٦	٢.٤٨٠	٠.٣٦٥	(١٤.٢٢٨; ١٥.١٦٤)	-٣.٥٧	٠.٠٠٠
C۲	٣٠	١٦.١٢٨	٢.٩٩٨	٠.٣٦٥	(١٥.٦٦٠; ١٦.٥٩٦)	٠.٣٥	٠.٧٢٦
C۳	٣٠	١٧.٥٦٥	٢.٨٨٢	٠.٣٦٥	(١٧.٠٩٧; ١٨.٠٣٣)	٤.٢٩	٠.٠٠٠

در اینجا نیز مانند قسمت قبل عمل می کنیم و مقدار p-value را با a مقایسه می کنیم.نتیجه مقایسه این است که دلیلی برای رد فرض صفر برای جامه دوم وجود ندارد اما در مورد جامعه اول و سوم فرض صفر رد می شود.

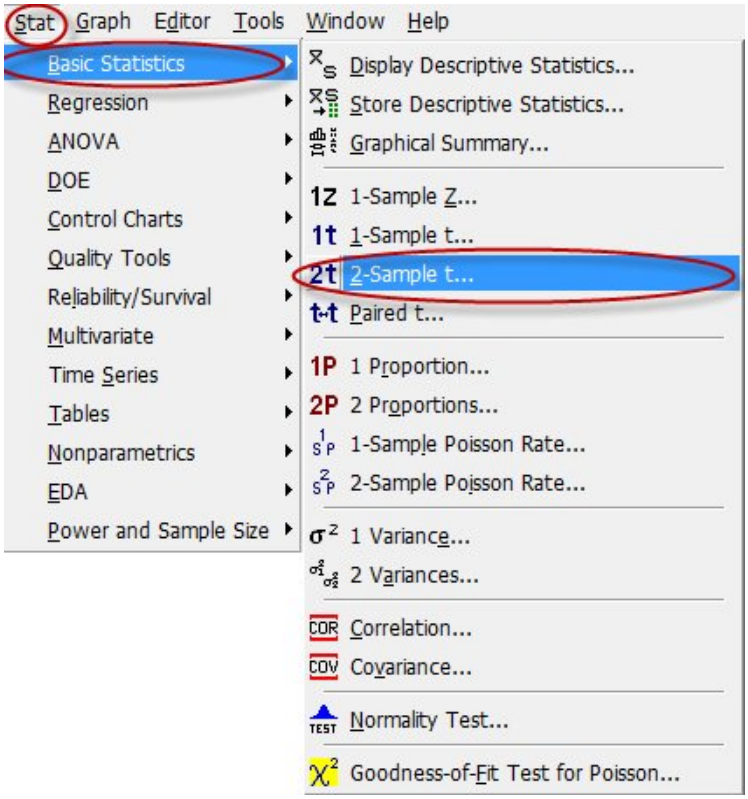
آزمون فرض برابری میانگین دو جامعه

در اینجا می خواهیم میانگین دو جامعه را با هم مقایسه کرده و ببینیم که آیا با هم برابر هستند یا خیر.این آزمون را هم بصورت دو طرفه و هم بصورت یک طرفه انجام خواهیم داد:

آزمون فرض دو طرفه بر روی میانگین دو جامعه

مسیر زیر را دنبال می کنیم:

Stat \ Basic Statistics \ 2-Sample t



نتیجه بصورت زیر خواهد بود:

Two-Sample T-Test and CI: C۱; C۲

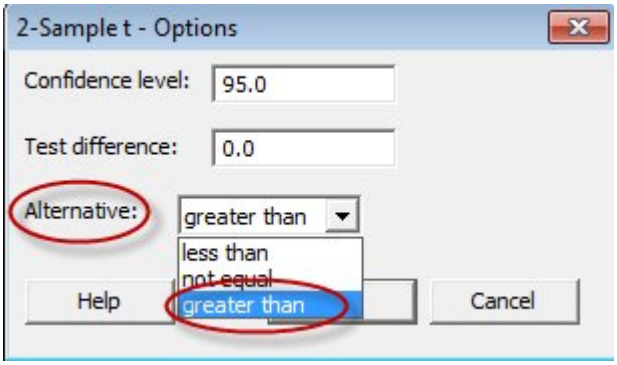
```
Two-sample T for C۱ vs C۲

     N   Mean  StDev  SE Mean
C۱   ۳۰  ۱٤.۷۰  ۲.٤۸    ۰.٤٥
C۲   ۳۰  ۱٦.۱۳  ۳.۰۰    ۰.٥٥


Difference = mu (C۱) - mu (C۲)
Estimate for difference:  -۱.٤۳۲
۹٥% CI for difference:  (-۲.۸٥٥; -۰.۰۰۹)
T-Test of difference = ۰ (vs not =): T-Value = -۲.۰۲  P-
Value = ۰.۰٤۹  DF = ٥٦
```

با تجه به اینکه مقدار p-value از مقدار a کمتر شده و همچنین همانطور که معلوم است فاصله اطمینان در

بازه منفی قرار دارد بنابراین فرض صفر رد شده و جامعه , c۱ و c۲ و میانگین های برابری ندارند.

آزمون فرض یک طرفه بر روی میانگین دو جامعه

در آزمون فرض یکطرفه مقادیر بیشتر و یا مساوی را برای برابری فرض صفر یعنی برابری میانگین دو جامعه می پذیریم همانند قسمت قبل عمل می کنیم اما در قسمت options تنظیمات زیر را اعمال می کنیم:



نتیجه بصورت زیر خواهد بود:

Two-Sample T-Test and CI: C۱; C۲

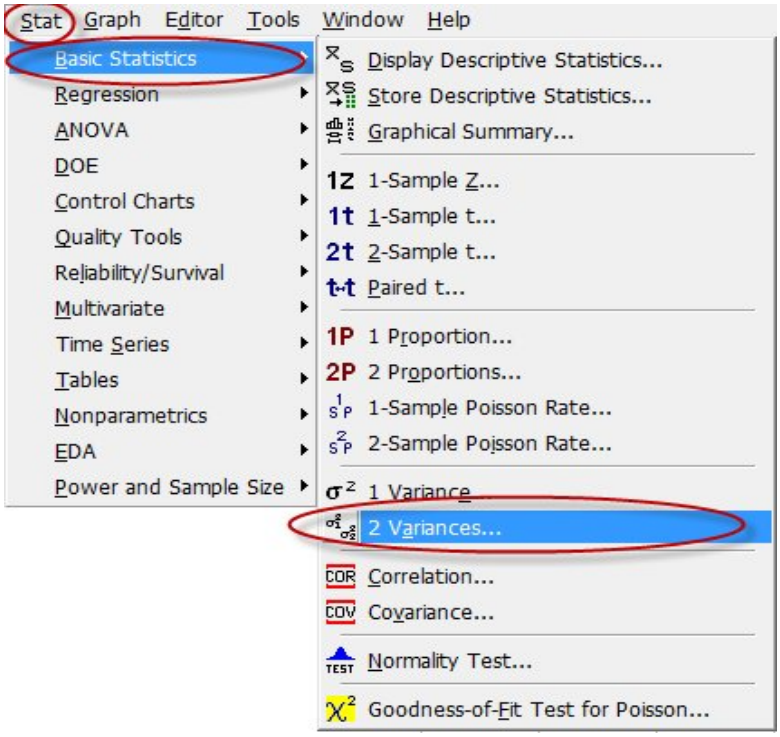
```
Two-sample T for C۱ vs C۲

     N   Mean  StDev  SE Mean
C۱   ۳۰  ۱٤.۷۰  ۲.٤۸     ۰.٤٥
C۲   ۳۰  ۱٦.۱۳  ۳.۰۰     ۰.٥٥


Difference = mu (C۱) - mu (C۲)
Estimate for difference:  -۱.٤۳۲
۹٥% lower bound for difference:  -۲.٦۲۰
T-Test of difference = ۰ (vs >): T-Value = -۲.۰۲  P-Value =
۰.۹۷٦  DF = ٥٦
```

با توجه به اینکه مقدار p-value از مقدار a بیشتر شده است بنابراین دلیلی برای رد فرض صفر وجود ندارد.

آزمون برابری واریانس های دو جامعه

این آزمون را تنها در حالت دو طرفه و برای دو جامعه C۲ و C۳ بررسی می کنیم.برای این کار مسیر زیر را پی می گیریم:

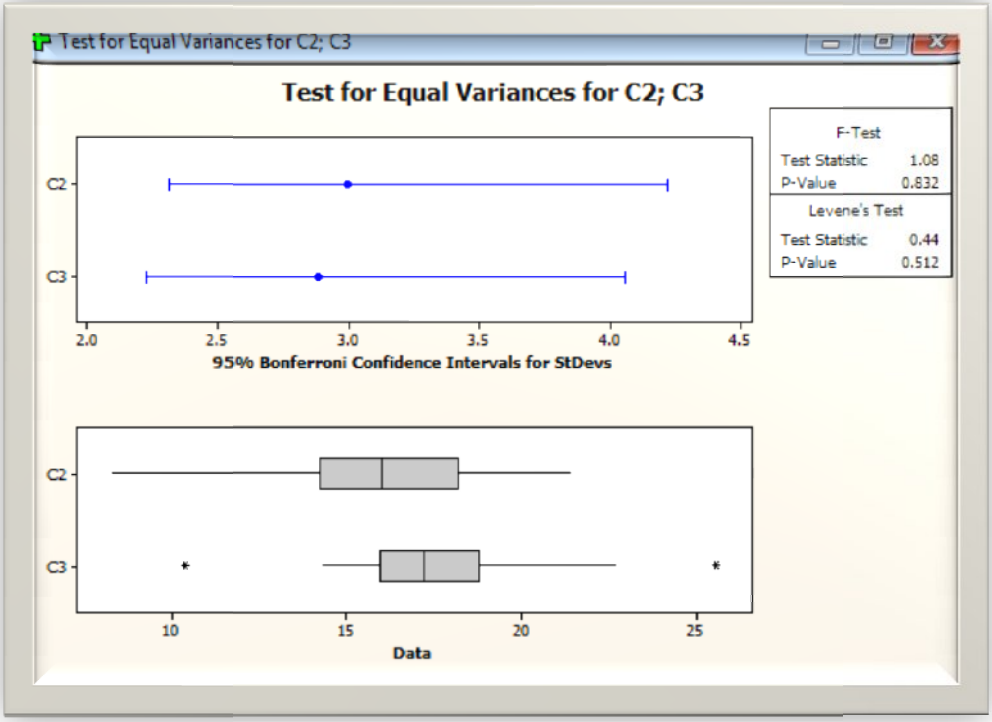
Stat\ Basic Statistics \2Variances…



نتیجه کار بصورت زیر خواهد بود:

Test for Equal Variances: C۲; C۳

```
۹۵% Bonferroni confidence intervals for standard deviations

     N    Lower     StDev    Upper
C۲   ۳۰   ۲.۳۱٤۳۹   ۲.۹۹۸۳۷   ٤.۲۱۷۳۷
C۳   ۳۰   ۲.۲۲٤٤۹   ۲.۸۸۱۹۰   ٤.۰٥۳٥٤


F-Test (Normal Distribution)
Test statistic = ۱.۰۸; p-value = ۰.۸۳۲


Levene's Test (Any Continuous Distribution)
Test statistic = ۰.٤٤; p-value = ۰.٥۱۲
```

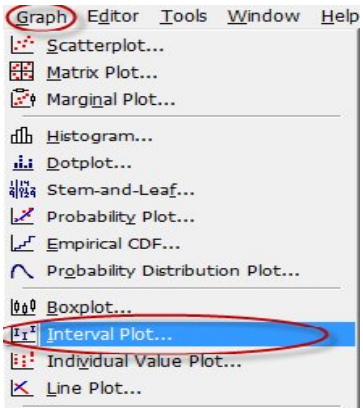


چنانکه می دانیم این توزیع از توزیع F پیروی می کند.همانند قسمت های قبل با مقایسه p-value و a ، از آنجایی که مقدار p-value از a بیشتر شده است بنابراین دلیلی برای رد فرض صفر در فاصله اطمینان ۹۵٪ وجود ندارد.

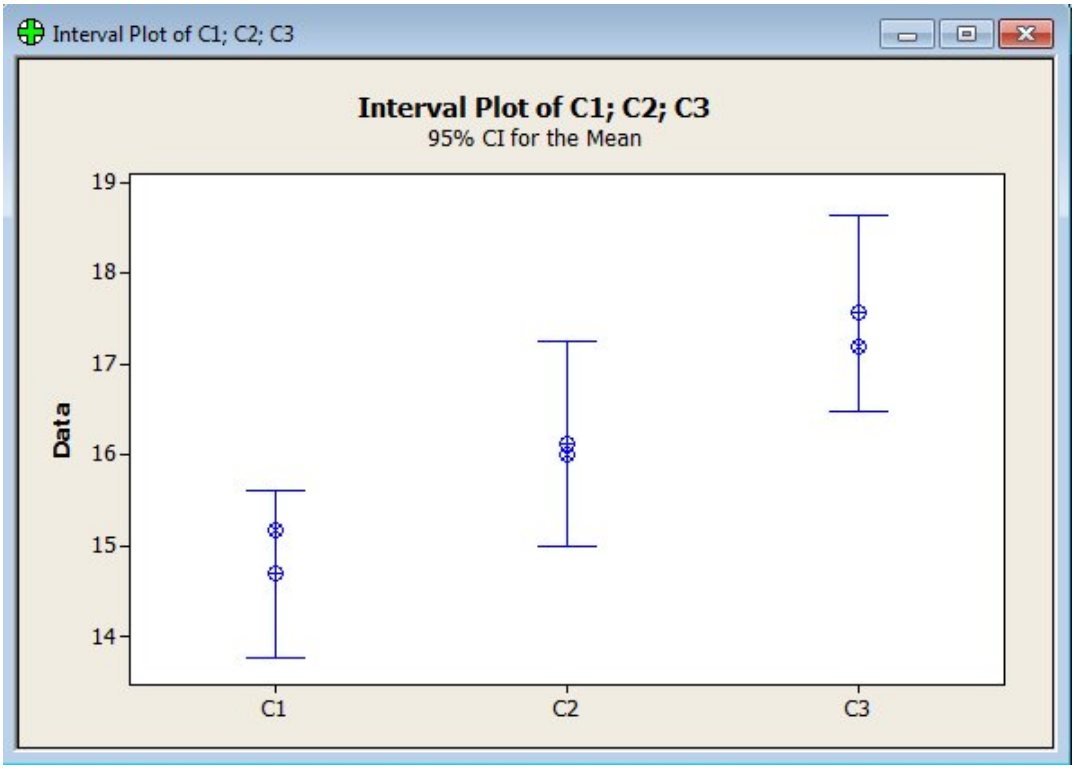
نمودار Interval Plot

این نمودار نیز از نمودارتر های پرکاربرد در آمار است و برای مشاهده همزمان سه جامعه و مقایسه پارامتر هایی مانند میانگین(پیش فرض) و میانه(دلخواه) از آن استفاده خواهیم کرد.نرم افزار بصورت پیش فرض این نمودار را در فاصله اطمینان ۹۵٪ رسم می کند.برای رسم این نمودار بصورت زیر عمل می کنیم:

Graph\Interval Plot



نتیجه کار بصورت زیر خواهد بود:



در این نمودار هم میانه و هم میانگین داده ها قابل مشاهده است.