نرم افزار SAS

فصل اول مقدمه
دستورات محاسباتی و اماری نرم افزار به کمک منوهای نرم افزار
برنامه نویسی با استفاده از نرم افزار SAS
معرفی داده ها به نرم افزار SAS
۱-شیوه معرفی داده ها با استفاده از صفحه کلید
نکاتی در مورد تعریف داده ها در قسمت DATA LINES
۲-شیوه معرفی داده ها با استفاده از فایل ورودی
چند نکته در مورد فراخوانی داده ها با استفاده از فایل ورودی
شیوه ادغام دو یا چند DATA STEP
طریقه ایجاد DATA SET توسط عملگرهای ریاضی و توابع SAS
توابع ریاضی در SAS
سایر عبارات لازم در DATA STEP
تجزیه و تحلیل اطلاعات در نرم افزار (PROC STEP)
فصل دوم رویه های SAS
۱)رویه PRINT
۲-رویه SORT
۳-رویه MEANS
٤-رویه FREQ
عبارت EXACT
عبارت TEST
عبارت TABLES
ازمون جداول فراوانی دوطرفه
٥-رویه TTEST
روشهای توصیفی
روشهای مورد استفاده برای برازش توزیع به مشاهدات
روشهای استنباطی
توصیف متغیرهای کمی
٦-رویه UNVARIATE
زیر دستورهای رویه UNIVARIATE
۱- BY VARIABLE(S);
۲- FREQ VARIABLE;
۳- HISTOGRAM VARIABLE(S) / OPTION;
٤- INSET KEYWORD(S);
٥- نام OUTPUT OUT= DATA SET
KEYWORD(S)=NAME(S);
٦- PROB PLOT VARIABLE(S) / OPTION;
۷- QQ PLOT VARIABLE(S) / OPTION;
۸- VAR VARIABLE(S);
۹- WEIGHT VARIABLE;
انالیز واریانس یکطرفه
۷-رویه ANOVA
زیر دستورهای رویه ANOVA
۱- زیر دستور CLASS
۲- زیر دستور MODEL
۳- زیر دستور TEST
٤- زیر دستور MEANS
۸- رویه CORR
زیر دستورهای رویه CORR
۱- زیر دستور PARTIAL

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۲- زیر دستور WITH
۹- رویه GLM
زیر دستورهای رویه GLM
۱- CLASS
۲- FREQ
۳- ID
٤- WEIGHT
٥- CONTRAST
٦-OUTPUT
۷- RANDOM
۸- ESTIMAT
۹- MEANS
۱۰- MODEL
۱۰-رویه NPAR1WAY
زیر دستورهای رویه NPAR1WAY
۱-BY
۲- CLASS
۳-EXACT
٤-FREQ
٥-OUTPUT
٦-VAR
۱۱-رویه REG
زیر دستورهای رویه REG
۱- ID
۲- MODEL
۳- OUTPUT
٤- PLOT
٥- RESTRICT
٦- TEST
۱۲-رویه CATMOD
زیر دستورهای رویه CATMOD
۱- DIRECT
۲- MODEL
۳- CONTRAST
٤- FACTORS
٥- LOGIN
٦- RESPONSE
۷- RESTRICT
۱۳-رویه PLOT
زیر دستورهای رویه PLOT
۱- BLOCK
۲- BY
۳- HBAR
٤- PIE
٥- STAR
٦- VBAR
فصل سوم-تمرینات

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

فصل اول:مقدمه

این نرم افزار از جمله تخصصی ترین نرم افزارهای اماری است.سرعت پردازش اطلاعات به نحو قابل توجه ای در این نرم افزار نسبت به سایر نرم افزارهای اماری بالاست.بنابراین در کارهای تخصصی اماری که شامل اطلاعات پیچیده می باشند بهتر است از این نرم افزار استفاده شود.
نگارشهای متفاوتی از این نرم افزار در دسترس است. اولین نگارش موجود این نرم افزار نگارش تحت سیستم DOS است.بعد از ایجاد سیستم عاملهای ویندوز, لینوکس و یونکس,این نرم افزار قابلیت نصب تحت این سیستم عاملها را در نگارشهای بعدی یافت.
اولین نگارش تحت ویندوز این نرم افزار نگارش ۶,۱۲ می باشد.نگارشهای بعدی تحت سیستم عامل ویندوز عبارتند از: ۷,۰-۸,۰-۸,۰۲-۹,۱
نکته:در این جزو نرم افزار SAS 9.01 آموزش داده می شود.
این نرم افزار از نگارش اول دارای سه پنجره می باشد که عبارتند از:
الف)پنجره program editor : برای برنامه نویسی
ب) پنجره out put : مختص خروجی نر افزار
ج) پنجره log : برای نمایش پیغام ها و شیوه اجرا
نکته: توجه کنید نرم افزار SAS به کوچکی یا بزرگی حروف حساس نمی باشد.

دستورات محاسباتی و اماری نرم افزار به کمک منوهای نرم افزار

STATISTICS→ANALYSIS→ANALYST

برنامه نویسی با استفاده از نرم افزار SAS

برنامه نویسی با استفاده از نرم افزار SAS با یک سری جملات اجباری و اختیاری امکان پذیر است.
شیوه برنامه نویسی در محیط این نرم افزار تقریبا شبیه برنامه نویسی به زبان C می باشد.به این معنی که تمامی عبارات نرم افزار بایستی به ; ختم شود.
برای برنامه نویسی در محیط این نرم افزار بایستی در دو مرحله برنامه نویسی را طراحی نمود.بنابراین بایستی در یک روند لازم ابتدا اطلاعات به نرم افزار معرفی شده و سپس دستورات لازم برای تجزیه و تحلیل اطلاعات لیست شود.
به مرحله ورود اطلاعات و معرفی انها به نرم افزار DATA STEP گوئیم.
تجزیه و تحلیل اطلاعات توسط قسمت PROC STEP امکان پذیر است.
از نظر تعداد معرفی DATA STEP , PROC STEP در یک پنجره برنامه نویسی محدودیتی وجود ندارد.
نرم افزار SAS از نظر باز کردن همزمان چند پنجره برنامه نویسی هیچگونه محدودیتی ندارد.
پنجره خروجی (OUT PUT) شامل نتیجه اعمال دستورات و عبارات PROC STEP های نرم افزار را نشان می دهد.
پنجره LOG برای مشاهده پیغام ها, هشدارها و نتیجه درستی یا عدم صحت انجام دستورات نرم افزار می باشد.

معرفی داده ها به نرم افزار SAS

۱-شیوه معرفی داده ها با استفاده از صفحه کلید

شکل کلی یک DATA STEP برای معرفی داده ها با استفاده از صفحه کلید به صورت زیر است.

```
DATA STATEMENT;
INPUT STATEMENT;
CARDS;(OR DATALINES;)
.....
.....
.....
RUN;(OR ;)
```

بنابراین یک DATA STEP با عبارت اجباری DATA شروع شده و به عبارت RUN ختم می شود.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

به محض معرفی یک DATA STEP نرم افزار پس از اجرای برنامه اطلاعات را در یک فایل غیر اسکی بنام DATA SET ذخیره خواهد کرد.
اگر بخواهیم در یک پنجره برنامه نویسی چند DATA STEP معرفی نمائیم بهتر است نام DATA SET را پس از عبارت DATA به نرم افزار معرفی نمائیم.
عبارت اجباری INPUT برای مشخص نمودن لیست نام متغیرهای لازم در مورد داده هاست.

```
INPUT X Y ;
```

نرم افزار SAS قادر است اطلاعات را بصورت متغیرهای عددی(NUMERIC) و متغیرهای حرفی (STRING) دریافت کند.
چنانچه متغیر یا متغیرها عددی باشند نام انها را حداکثر با ۸ کاراکتر در جلوی عبارت INPUT یادداشت می کنیم ولی اگر برخی از متغیرها از نوع حرفی باشند بایستی حتما بعد از نام انها از علامت $ حداقل با یک فاصله استفاده نمود.
مقادیر متغیرهای حرفی می توانند شامل هر نوع کاراکتر باشند. اگر بین مقدار یک مشاهده متغیر حرفی فاصله وجود داشته باشد.نرم افزار نمی تواند مقدار کامل ان را دریافت کند.مقادیر متغیرهای حرفی بایستی حداکثر با ۸ کاراکتر باشد.

نکاتی در مورد تعریف داده ها در قسمت DATA LINES

۱-مشاهدات در مورد هر متغیر بایستی حداقل با یک فاصله از یکدیگر جدا شوند.

Example no1:

```
INPUT X Y;
CARDS;
15   18
10   9
..    ..
..    ..
..    ..
7     12
;
```

۲-اگر بخواهیم مشاهدات را بصورت سطری برای متغیرها در نظر بگیریم. بایستی قبل از ; در دستور INPUT از علامت @@ استفاده شود.

Example no2:

```
INPUT X Y @@;
CARDS;
15 18 10 9…. 7 12
RUN;
```

۳-اگر متغیر یا متغیرهای حرفی داشته باشیم بایستی مقادیر انها حداکثر با ۸ کاراکتر باشند و فاصله بین کاراکتر های هر مشاهده وجود نداشته باشد.
۴-اگر بخواهیم مقادیر مشاهدات حرفی را که مقادیر انها دارای یک فاصله می باشد به نرم افزار معرفی نمائیم بایستی بعد از علامت $ در عبارت INPUT در مورد ان متغیر از علامت & استفاده کنیم. همچنین پس از تایپ علامت فوق در قسمت DATA LINES بایستی مقادیر مشاهدات با حداقل دو کاراکتر از یکدیگر جدا شوند.

Example no3:

```
DATA;
INPUT X NAME$&;
CARDS;
12  J A
13  J
14  B
;
```

۵-اگر مقادیر متغیرهای حرفی بیش از ۸ کاراکتر باشند می توان تعداد کاراکترهای ان را در عبارت LENGTH قبل از عبارت INPUT معرفی کنیم.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

Example no4:

```
DATA;
LENGTH NAME $ 10;
INPUT NAME$;
CARDS;
ABCDEFGHI
A
B
C
;
```

نکته : طول تعداد متغیرهای حرفی را نی توان عددی بین ۸ تا ۲۰۰ در نظر گرفت.

۲-شیوه معرفی داده ها با استفاده از فایل ورودی

گاهی اوقات مشاهدات را در فایل متنی یا بانک اطلاعاتی دیگر ذخیره کرده و می خواهیم همان اطلاعات را در نرم افزار فراخوانی کنیم.بدین منظور بایستی فرمت فایل داده ها بصورت فرمت DAT یا TXT یا PRN در نظر گرفته و طوری ان را مشخص کنیم که مشاهدات از سطر اول فایل متنی شروع شده باشد.
بهتر است ابتدا اطلاعات را توسط نرم افزار EXCEL به فرمت PRN(formatted text) تبدیل کنید.
شکل کلی DATA STEP برای فراخوانی مشاهدات از یک فایل متنی ورودی بصورت زیر است.

```
DATA DATASET;
INFILE 'توسعه. نام فایل \: درایو' ;
INPUT نام متغیرها ;
RUN;
```

چند نکته در مورد فراخوانی داده ها با استفاده از فایل ورودی

۱-اگر بخواهیم برخی اطلاعات موجود در بانک اطلاعاتی را فراخوانی نمائیم بطور کلی این امکان پذیر نیست ولی اگر بخواهیم اطلاعات چند متغیر را از ابتدای بانک اطلاعاتی فایل متنی در نرم افزار فراخوانی نمائیم این کار امکان پذیر است. بعنوان مثال اگر فایل متنی ورودی شامل ۵ متغیر X Y Z W V باشد و بخواهیم سه متغیر را فراخوانی نمائیم با استفاده از فرمت DATA STEP در فراخوانی اطلاعات با استفاده از فایل ورودی این کار امکان پذیر است. برای این منظور کافی است در عبارت INPUT نام سه متغیر اول را لیست نمائیم.ولی چنانچه بخواهیم سه متغیر X Z W را فراخوانی نمائیم,بایستی ابتدا تمام مشاهدات فایل متنی را فراخوانی نمود و سپس مشاهدات مربوط به متغیرهای غیر ضروری را از DATA SET حذف کرد.برای این منظور می توانیم از دستورات KEEP,DROP استفاده نمود.

شیوه ادغام دو یا چند DATA STEP

برای ادغام دو یا چند DATA STEP از دستور SET استفاده می شود.طریقه نوشتن این دستور بصورت زیر است.(ادغام دو DATA SETP به نامهای SAS1,SSA2)

```
SAS 1 SAS2
DATA NEW;
SET SAS1 SAS2;
RUN;
```

ممکن است در ادغام اطلاعات دو یا چند مجموعه داده به دنبال افزایش تعداد مشاهدات یا متغیرها باشیم در این صورت می توان با تعریف یک DATA STEP جدید توسط یک DATA STEP انجام داد.

طریقه ایجاد DATA SET توسط عملگرهای ریاضی و توابع SAS

عبارت محاسباتی در SAS که توسط عملگرهای ریاضی و توابع تعریف شده باعث ایجاد متغیرهای جدید از روی مقادیر متغیرهای تعریف شده می شوند را بایستی بعد از دستور INPUT و قبل از دستور CARDS; یا RUN; یادداشت نمود.
عملگرهای ریاضی تعریف شده به ترتیب اولویت اجرا(از چپ به راست)عبارتند از:
- (تفریق)→ + (جمع) → / (تقسیم) → * (ضرب) → **(توان)

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

بعنوان مثال اگر بخواهیم مقادیر مشاهدات متغیر X را در عدد ۲ ضرب نمائیم و مشاهدات جدید را در Y ذخیره نمائیم DATA SET می تواند بصورت زیر باشد.

```
DATA;
INPUT X;
Y=2*X;
CARDS;
2
4
3
RUN;
```

توابع ریاضی در SAS

برخی از توابع مورد نیاز در SAS عبارتند از:

1)ABS(ARGOMAN)	قدر مطلق
2)SQRT(ARGOMAN)	جذر
3)INT(ARGOMAN)	جزء صحیح
4)MIN(ARGOMAN)	کمترین
5)MAX(ARGOMAN)	بیشترین
6)LOG(ARGOMAN)	لگاریتم طبیعی
7)LOG10(ARGOMAN)	لگاریتم معمولی(اعشاری)
8)SIN(ARGOMAN)	سینوس
9)COS(ARGOMAN)	کسینوس
10)TAN(ARGOMAN)	تانژانت

سایر عبارات لازم در DATA STEP

۱-عبارت KEEP

با استفاده از این دستور میتوان متغیرهای که لازم است در DATA SET حفظ شوند را انتخاب نموده و جلوی عبارت فوق نوشت.

۲-عبارت DROP

این دستور برای حذف متغیر یا متغیرهای که ضروری نیستند از مجموعه دادهها مورد استفاده قرار می گیرد

نکته: استفاده همزمان از دو عبارت فوق در یک DATA STEP جایز نیست.

۳-دستور شرطی IF

دو صورت کلی برای استفاده از دستور شرطی IF بصورت زیر است.

1) IF(شرط) THEN (در صورت برقراری شرط) عبارت (در صورت برقراری شرط);

2) IF(شرط) THEN (در صورت برقراری شرط) عبارت ELSE (در صورت عدم برقراری شرط) ;

Example : if x=1 then y=2 else y=3;

٤-عبارت DO

از حلقه DO نیز گاهی اوقات نیز برای تعریف مقادیر مشاهدات استفاده می شود.

DO نام متغیر=مقدار اولیه TO مقدار نهائی ;
BY گام حلقه;
END;

Example no1:

```
DO X=1 TO 5;
BY 2;
END;
```

RESULT: 1,3,5

Example no2:

```
DO X=50 TO 0;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

BY -2;
END;
RESULT:50,48,…,4,2,0

Example no3:
DO X=1 TO 4;
Y=SIN(X);
END;
RESULT: 0.017, 0.034, 0.052, 0.069

Example no4:
DO X=1,2,5,7;
END;
RESULT:1,2,5,7;

Example no5:
DO X='I','II','III';
END;
RESULT: I,II,III

Example no6;
DO X='A','B','C';
END;
RESULT:A,B,C

استفاده همزمان از چند حلقه DO بصورت همزمان و بصورت حلقه های تودرتو نیز در اکثر مواقع برای تعریف بخشی از متغیرهای لازم مناسب می باشد.

Y	X1		
X2		1	2
	1	10	5
	2	4	3
	3	2	1

مثال)فرض کنید جدول فراوانی فوق را داشته باشیم برای معرفی داده ها فوق به نرم افزار می توانیم از دو برنامه زیر استفاده کنیم.

Sample 1:
DATA;
INPUT X1 X2 Y @@;
CARDS;
1 1 10 1 2 5 2 1 4
2 2 3 3 3 1 3 2 1
;
Sample 2;
DATA;
DO X1=1,2,3;
DO X2=1,2;
INPUT Y@@;
OUTPUT;
END;
END;
CARDS;

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

10 5 4 3 2 1
RUN;

۵- عبارت OUTPUT
از این عبارت برای فرستادن مقادیر مشاهدات هر یک از متغیرها به DATA STEP استفاده می شود.

تجزیه و تحلیل اطلاعات در نرم افزار (PROC STEP)
تجزیه و تحلیل اطلاعات در نرم افزار SAS با رویه های تعریف شده برای ان امکان پذیر است.هر رویه با توجه به عملکرد ان با یک نام مشخص شده است.ممکن بین رویه های مختلف نرم افزار اشتراک وجود داشته باشد ولی هر یک از رویه ها برای یک تجزیه و تحلیل خاص در نظر گرفته شده اند. در یک رویه چند عبارت یا زیر دستور کار تجزیه و تحلیل را توسعه می دهند.
برخی از عبارات اختیاری و برخی دیگر اجباری هستند.ترتیب نوشتن زیر دستورات در برخی رویه ها مهم می باشد و بایستی به ان توجه داشت.
برخی عبارات در یک رویه می توانند تکرار شوند و برخی زیر دستورات نیز طوری ساخته شده اند که ممکن است استفاده همزمان از انها در یک DATA STEP جایز نباشد.
چند زیر دستور عملکرد مشابه ای در اکثر رویه های نرم افزار SAS دارند که قبل از معرفی انواع رویه های لازم و کاربردی به معرفی انها می پردازیم.

۱-عبارت VAR
از این عبارت برای لیست کردن متغیرهای لازم که عملکرد رویه روی انها صورت خواهد گرفت استفاده می شود.

VAR نام متغیر یا متغیرها;

۲-عبارت BY
برای منحصر کردن عملکرد یک رویه به زیر گروههای یک یا چند متغیر از عبارت BY استفاده می شود.
در هنگام استفاده از این زیر دستور باید توجه داشت که فایل مجموعه داده ها بایستی قبلا بر حسب مقادیر متغیر یا متغیرهای لیست شده در عبارت BY مرتب شده باشد.

۳-عبارت ID
با استفاده از این عبارت می توان ترتیب نمایش متغیرها را بر حسب متغیرهای ذکر شده در جلوی این عبارت به نرم افزار معرفی نمود.

۴-عبارت OUTPUT
از این دستور برای ذخیره کردن مقادیر برخی از متغیرها یا شاخصها که در هر یک از رویه ها ساخته می شوند در یک مجموعه داده خروجی استفاده می شود.شکل کلی این زیر دستور به صورت زیر است.

OUTPUT OUT=DATA SET نام ;

فصل دوم: رویه های SAS

پس از تعریف مشاهدات توسط یک یا چند DATA STEP برای تجزیه و تحلیل اطلاعات از PROC STEP استفاده می شود. هر یک از رویه های نرم افزار کارکرد مشخصی دارند که با توجه به عملکرد شان توسط نامهای اختصاری نامگذاری شده اند. هر یک از رویه ها دارای زیر دستورهای اجباری و اختیاری هستند که برخی از انها بر دیگر عبارات برتری از نظر ترتیب نوشتن دارند. همچنین بعضی از دستورها را نمی توان بطور همزمان در یک دستور استفاده نمود.از طرف دیگر برخی از دستورهای رویه های نرم افزار محدودیت در تعداد موارد استفاده از ان در یک رویه را دارند. ولی برخی از عبارات را می توان چند بار بنا به ضرورت در یک رویه مورد استفاده قرار داد.
رویه های که از نظر کاربردی اهمیت زیادی دارند و در این جزو مورد بررسی قرار خواهند گرفت عبارتند از:

1)PROC PRINT
2)PROC SORT
3)PROC MEANS
4)PROC FREQUENCY
5)PROC UNIVARIATE
6)PROC TTEST
7)PROC ANOVA

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

8)PROC CORR
9)PROC GLM
10)PROC NPAR1WAY
11)PROC REGRESSION
12)PROC CATMOD
13)PROC CHART

۱)رویه PRINT

برای دیدن مشاهدات در خروجی و چک کردن طریقه مشاهدات به نرم افزار از این رویه استفاده می شود.
شکل کلی این رویه بصورت زیر است.

```
PROC PRINT OPTION;
VAR نام متغیرها ;
ID نام متغیرها;
RUN;
```

نکته : هر یک از رویه های نرم افزار SAS دارای OPTION های خاص هستند که انها را بایستی بعد از ان رویه در صورت لزوم یادداشت کرد.انتخابهای لازم در مورد عبارت اختیاری یا اجباری هر یک از رویه های نرم افزار بعد از علامت " / " تایپ می شود.
یکی از OPTION های مشترک همه رویه های نرم افزار SAS عبارت (نام DATA=(DATA SET می باشد.

Data example1:

```
INPUT X Y @@;
CARDS;
10 2 5 3 4
;
PROC PRINT DATA=Example1;
RUN;
```

نکته : دو عبارت VAR , ID عبارات اختیاری این رویه هستند.عملکرد زیر دستور VAR قبلا معرفی شد.
اگر از زیر دستور ID در این رویه استفاده نمائیم متغیرهای که در جلوی ID لیست شدهاند را به ترتیب لیست شدن انها در خروجی همراه با مقادیر انها مشاهده خواهیم کرد.

۲-رویه SORT

این دستور مقادیر مشاهدات معرفی شده به نرم افزار را بر اساس مقادیر یک یا چند متغیر مرتب می کند.
شکل کلی این رویه بصورت زیر است.

```
PROC SORT OPTION;
BY نام متغیر یا متغیرها ;
RUN;
```

یکی از OPTION های این رویه عبارت (نام DATA=(DATA SET می باشد که برای معرفی مجموعه داده ها استفاده می شود. متغیرهای را که بایستی اطلاعات بر اساس مقادیر انها مرتب شوند در جلوی عبارت BY یادداشت می کنیم.

۳-رویه MEANS

برای محاسبه شاخصهای اماری تعداد,میانگین,انحراف معیار,مینیمم و ماکزیمم در مورد متغیرهای کمی از این دستور استفاده می شود. شکل کلی این دستور بصورت زیر است.

```
PROC MEANS OPTION;
VAR نام متغیر یا متغیرها;
BY نام متغیر یا متغیرها ;
END;
```

یکی از OPTION های مهم این رویه معرفی نام DATA SET ورودی است.
از زیر دستور VAR که عبارتی اختیاری می باشد برای لیست کردن نام متغیرهای لازم استفاده می شود.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

عبارت اختیاری BY نیز برای منحصر کردن فعالیت رویه به زیر گروههای متغیرهای لیست شده در ان مورد استفاده قرار می گیرد.

۴-رویه FREQ

از این دستور برای توصیف تک بعدی متغیرهای کیفی و توصیف چند بعدی انها با استفاده از ابزار امار توصیفی استفاده می شود.
علاوه براین با استفاده از این رویه می توان برخی شاخصها و ازمونها (ازمون نیکویی برازش) را برای متغیرها (ممکن است متغیرها لازم باشد از نوع مقیاسهای فاصله ای یا نسبتی در نظر گرفته شوند) در نظر گرفته و محاسبه نمود. تجزیه و تحلیل اطلاعات جداول توافقی دو طرفه نیز با این رویه امکان پذیر است.
شکل کلی این رویه بصورت زیر است.

```
PROC FREQ OPTION;
TABLES   REQUST   /OPTION;
TEST;
EXACT OPTION;
BY نام متغیر یا متغیرها ;
OUT PUT OUT= DATA SET  نام OPTION;
WEIGHT نام متغیر یا متغیرها ;
RUN;
```

OPTION های رویه FREQ

الف) **DATA=DATA SET** : برای معرفی مجموعه داده های ورودی
ب) **NOPRINT** : برای حذف تمامی اطلاعات خروجی
ج) **ORDER=DATA** : گزارش جداول فراوانی بر اساس ترتیب معرفی مشاهدات
ORDER=FORMATED : مرتب کردن جداول فراوانی براساس مقادیر اختصاص داده شده به هر یک از سطوح مشاهدات
ORDER=FREQ : برای گزارش جداول فراوانی ساده بر اساس ترتیب نزولی فراوانی مطلق هر یک از متغیرها

جنسیت	f
مرد ۱	۵۰
۲ زن	۶۰
۱	۴۰

ORDER=DAT

جنسیت	f
۱	۵۰
۲	۶۰
۱	۴۰

ORDER=FORMTED

جنسیت	f
۱	۵۰
۱	۴۰
۲	۶۰

ORDER=FREQ

جنسیت	f
۲	۶۰
۱	۵۰
۱	۴۰

ت)**PAGE** : برای گزارش کردن هر یک از جداول فراوانی در یک صفحه جداگانه

عبارت EXACT

از این زیر دستور که یک عبارت اختیاری می باشد برای محاسبه دقیق P-مقدار و مقادیر اماره های ازمونها در مورد جداول فراوانی استفاده می شود. برخی از ازمونهای دقیقی که نرم افزار با استفاده از این عبارت محاسبه می کند عبارت است از:
۱)**AGREE** ازمونهای توافقی : براساس توزیع دقیق اماره ازمون مانند ازمون توافق کاپا , ضریب توافق کاپای وزنی , ازمون مک نمار در جدول ۲*۲
۲)**BINOMIAL** ازمون دو جمله ای
۳) **CHISQ** : برای انجام ازمونهای کی دوی پیرسون , نسبت درست نمائی
۴) **FISHER** : ازمون دقیق فیشیر
۵)**JT** : ازمون نیکوئی برازش JANCK HEARE-TERPTRA

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۶) **KAPPA** : برای انجام ازمون دقیق ضریب توافق کاپا
۷) **LRCHI** : ازمون دقیق اماره نسبت درستنمائی
۸) **MCNEM** : ازمون دقیق مک نمار
۹) **MEASURES** : ازمون معنی داری ضرایب همبستگی خطی پیرسون و اسپیرمن و غیره
۱۰) **OR** : فاصله اطمینان دقیق برای نسبت شانسها ODD RATIO
۱۱) **PCHI** : ازمون دقیق اماره کی دوی پیرسون
۱۲) **PCORR** : ازمون دقیق معنی داری ضریب همبستگی خطی پیرسون
۱۳) **SCORR** : ازمون معنی داری ضریب همبستگی خطی اسپیرمن بر اساس توزیع دقیق
۱۴) **WTKAP** : برای محاسبه ازمون دقیق ضریب کاپای وزنی
برخی از optionهای این زیر دستور که بعد از / در نظر گرفته می شوند عبارتند از:
۱) ALPHA= (0<P<1)
برای مشخص کردن ضریب اطمینان فاصله اطمینان برای نسبت شانسها
بطور پیش فرض $\alpha=0.01$ در نظر گرفته می شود.
۲) زمان بر حسب ثانیه = MAX TIME
۳) MC برای انجام ازمون دقیق مونت کارلو

عبارت TEST

از این عبارت برای انجام ازمونهای مجانبی در مورد جداول فراوانی استفاده می شود.
ازمونهای لازم برای لیست کردن در جلوی این عبارت همان اماره های لیست شده در زیر دستور EXACT می باشد. معمولا این زیر دستور به همراه عبارت TABLES در رویه استفاده می شود.

عبارت TABLES

OPTION های زیر دستور TABLES
۱) **ALPHA** : برای مشخص کردن ضریب اطمینان برای فواصل اطمینان
۲) **MEASURES** : برای محاسبه کلیه اندازه ها شامل ضرایب همبستگی و شاخص ریسک نسبی (relative risk) شاخص نسبت شانسها (odd ratio)
۳) **NOCUM** : برای حذف ستون فراوانی تجمعی از جداول فراوانی
۴) **NOPRECENT** : برای عدم گزارش هر گونه درصد فراوانی نسبی در جداول فراوانی
۵) **NOROW** : برای عدم نمایش درصد فراوانی نسبی متغیر سطر
۶) **NOCOL** : برای عدم نمایش درصد فراوانی نسبی متغیر ستون
۷) **EXPECTED** : برای نمایش فراوانی مورد انتظار در جداول توافقی
۸) **ALL** : برای محاسبه تمامی ازمونها و اندازه در مورد اطلاعات جدول
۹) **BINOMIAL** : ازمون دو جمله ای را برای جداول فراوانی یکطرفه انجام می دهد.
۱۰) **FISHER** : برای انجام ازمون دقیق فیشیر
۱۱) **MISSING** : برای در نظر گرفتن تعداد مشاهدات گمشده در محاسبه درصد های فراوانی نسبی

ازمون جداول فراوانی دوطرفه

جداول توافقی به دو دسته کلی زیر تقسیم می شود.
۱- جداول توافقی برای بررسی ارتباط یا عدم ارتباط دو متغیر کیفی
۲- جداول توافقی همبسته برای بررسی توافق یا میزان سازگاری بین دو ازمون

۱- جداول توافقی برای بررسی ارتباط یا عدم ارتباط دو متغیر کیفی

برای ازمون این نوع جداول توافقی از ازمون کی دو استفاده می شود.ازمون کی دو شامل سه اماره زیر است.
الف) ازمون دقیق فیشیر (با تصحیح پیوستگی یا بدون تصحیح پیوستگی)
ب) اماره کی دوی پیرسون PEARSON CHI-SQUARE
ج) اماره نسبت درستنمائی LIKELIHOOD RATIO
در جداول توافقی ۲*۲ **تنها** از **ازمون دقیق فیشر** برای بررسی استقلال جدول استفاده می شود.
در جداول توافقی دوطرفه (بیش از ۲*۲) اگر تعداد مشاهدات زیاد باشد به نحوی که فراوانی های نظری تحت فرضیه H_0 بیش از ۵ باشد می توان برای بررسی استقلال جدول توافقی از اماره کی دوی پیرسون استفاده

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

کرد. در صورتیکه فراوانی های از پنج کمتر باشد در مرحله اول در صورتیکه امکان داشته باشد بایستی برخی از ستونها یا سطرهای جدول توافقی را ادغام نمود. بعد از ادغام چنانچه شرایط استفاده از اماره پیرسون برقرار باشد می توان از این اماره برای بررسی استقلال استفاده نمود. در غیر اینصورت به شرط انکه فراوانی مشاهده شده صفر در جدول وجود نداشته باشد می توان از اماره نسبت درستنمائی استفاده نمود.

$$Qdf=\sum_{i=1}^{r}\sum_{j=1}^{m}\frac{(Eij-Oij)^2}{EIJ} \qquad \text{DF=(R-1)(M-1)} \qquad LR=-2\sum_{i=1}^{r}\sum_{j=1}^{m}\frac{Eij}{Oij}$$

اگر شرایط قبل برقرار نباشد بایستی از ازمون دقیق فیشیر یا ازمون دقیق برای بررسی استقلال جدول توافقی استفاده نمود.

۲- جداول توافقی همبسته برای بررسی توافق یا میزان سازگاری بین دو ازمون

معمولا این جداول که برای همبستگی بین دو ازمون که روی یکسری ازمونها استفاده می شود. این جداول بایستی مربع باشند. اگر چنین جداولی ۲*۲ باشند. برای بررسی سازگاری بین دو پاسخ که از نوع باینری می باشند از ازمون ناپارامتری مک نمار استفاده می شود.

ازمون مک نمار حالت خاصی از ازمون علامت (SIGN) حالت دو نمونه زوج شده می باشد که در ان متغیرهای پاسخ بجای اینکه رتبه ای باشند, از نوع دو سطحی (0,1) می باشد.

تعمیم ازمون مک نمار به حالتی که پاسخها چند سطحی غیر رتبه ای باشند , **ازمون همگنی غیر رتبه ای** (marginal homogeneity) است.

مثال۱ : فرض کنید نظر افراد را در مورد این سئوال جویا شده ایم (تعداد افراد ۱۰۰ نفر)

ایا موافق وضع مقررات سختری برای رانندگانی که تحت مصرف مشروبات الکلی رانندگی می کنند هستید؟

اگر پاسخها را بصورت بلی یا خیر ثبت کنیم و پس از جمع اوری نظرات این صد نفر یک فیلم آموزشی که نشان دهنده حوادث اتفاق افتاده ناشی از مصرف مشروبات الکلی است , پخش شود و دوباره سئوال را مطرح کنیم برای انکه مشخص کنیم ایا فیلم تاثیری روی نظر افراد در پاسخ به سئوال دارد بایستی از ازمون مک نمار استفاده کنیم.

نظر۱ / نظر۲	۰ خیر	۱ بلی	جمع
۰خیر	۱۰	۲۵	۳۵
۱بلی	۳۵	۳۰	۶۵
جمع	۴۵	۵۵	۱۰۰

ازمون نسبت در دو جامعه وابسته به یکدیگر

$H_0 : P_1=P_2$

$H_1 : P_1\neq P_2$

P_1 : نسبت کسانی که قبل از مشاهده فیلم موافق هستند.

P_2 : نسبت کسانی که بعد از مشاهده فیلم موافق هستند.

در صورتی که دو سری افراد (قبل از نمایش فیلم و بعد از نمایش فیلم) یکی نباشند از جدول زیر استفاده می شود.

نظر / وضعیت	۰ موافق	۱ مخالف	جمع
۰قبل از مشاهده فیلم	N1	N2	۱۰۰
۱بعد از مشاهده فیلم	N3	N4	۱۰۰
جمع	N1+N3	N2+N4	۲۰۰

برای مقایسه نسبت در دو جامعه مستقل از **ازمون دقیق فیشر** استفاده می شود.

اگر در مثال قبل پاسخها بصورت مخالف , ممتنع , موافق ثبت شود. بایستی از **ازمون همگنی حاشیه ای** برای بررسی سازگاری بین پاسخ افراد , قبل و بعد از مشاهده فیلم استفاده نمائیم.

مثال۲ : در یک نظر سنجی از ۱۰۰ نفر دو سئوال زیر مطرح شده است.

۱- ایا موافق کاندیدا توری فرد A هستید؟

۲- ایا موافق کاندیدا توری فرد B هستید؟

انگاه جنسیت افراد ثبت می شود. پاسخها به شرح زیر است.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

جنسیت ۲

سئوال۱ / سئوال۲	۰	۱	جمع
۰	۱۳	۲	۱۵
۱	۸	۲۴	۳۲
جمع	۲۱	۲۶	۴۷

جنسیت۱

سئوال۱ / سئوال۲	۰	۱	جمع
۰	۱۰	۱۳	۲۳
۱	۱۸	۱۲	۳۰
جمع	۲۸	۲۵	۵۳

مطلوبست :
الف) ایا نظر افراد در مورد کاندیدا توری فرد A,B یکسان است؟
ب) ایا افراد جنسیت های مختلف نظر متفاوتی در مورد کاندیدا توری A دارد؟
ج) ایا جنسیت تاثیری روی میزان موافقت با کاندیدا توری B دارد؟

```
DATA;
DO SEX=1,2;
DO R1=0,1;
DO R2=0,1;
INPUT C @@;
OUTPUT;
END; END; END;
CARDS;
10 13 18 12 13 2 8 24
;
PROC PRINT;  (اختیاری)
PROC FREQ;
TABLES SEX R1 R2 R1*SEX R2*SEX R1*R2 /ALL;
WEIGHT C;
RUN;
```

مثال۳:در یک بررسی گذشته نگر در مورد سرطان رحم اطلاعات زیر در مورد ۲۹۹ زن که ده سال پیش تحت عمل جراحی سرطان قرار گرفته اند کسب شده است. برای این زمان چهار متغیر زیر مشاهده شده است.
الف) دریافت یا عدم دریافت پرتو درمانی
ب) وضعیت حیات در مدت ده سال
ج) نوع عمل جراحی شامل اساسی یا محدود
د) مرحله بیماری مشتمل بر دو سطح اولیه و پیشرفته
می خواهیم به سئوالات زیر پاسخ دهیم.

۱- ایا بقای بیماری در عرض ده سال ربطی به نوع عمل جراحی استفاده شده دارد؟
۲- ایا مرحله بیماری روی زنده یا مرده بودن بیماری تاثیری دارد؟
۳- ایا نسبت بیماران زنده در بین کسانی که پرتو درمانی را دریافت کرده اند یا دریافت نکرده اند تفاوت معنی داری دارد؟
۴- در بین بیماران با جراحی محدود ایا زنده بودن بیماران وابسته به مرحله بیماری است؟

D	C	B	A	
مرحله بیماری	نوع عمل جراحی	وضعیت حیات	عدم دریافت	دریافت
اولیه	اساسی	مرده	۱۰	۱۷
		زنده	۴۱	۶۴
	محدود	مرده	۱	۳
		زنده	۱۳	۹
پیشرفته	اساسی	مرده	۳۸	۶۴
		زنده	۶	۱۱
	محدود	مرده	۳	۱۳
		زنده	۱	۱۵

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
DATA;
DO D=1,2; DO C=1,2; DO B=1,2; DO A=1,2;
INPUT COUNT@@;
OUTPUT;
END; END; END; END;
CARDS;
10 17 41 64 1 3 13 9 38 64 6 11 3 13 1 5
;
PROC FREQ;
TABLES B*C D*B B*A C*B*D;
EXACT FISHER;
WEIGHT COUNT;
RUN;
```

خروجی برنامه فوق در زیر آمده است.

```
                          The SAS System            10:51 Friday,
April 13, 2001   1

                          The FREQ Procedure

                           Table of B by C

B          C

Frequency
Percent
Row Pct
Col Pct       |         1|         2|   Total
--------------+----------+----------+
            1 |       129|        20|     149
              |     43.14|      6.69|   49.83
              |     86.58|     13.42|
              |     51.39|     41.67|
--------------+----------+----------+
            2 |       122|        28|     150
              |     40.80|      9.36|   50.17
              |     81.33|     18.67|
              |     48.61|     58.33|
--------------+----------+----------+
Total         |       251|        48|     299
              |     83.95|     16.05|   100.00


                Statistics for Table of B by C

Statistic                     DF       Value      Prob
------------------------------------------------------
Chi-Square                     1      1.5252    0.2168
Likelihood Ratio Chi-Square    1      1.5315    0.2159
Continuity Adj. Chi-Square     1      1.1609    0.2813
Mantel-Haenszel Chi-Square     1      1.5201    0.2176
Phi Coefficient                       0.0714
Contingency Coefficient               0.0712
Cramer's V                            0.0714


          Fisher's Exact Test
----------------------------------
Cell (1,1) Frequency (F)       129
Left-sided Pr <= F          0.9184
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

Right-sided Pr >= F	0.1406
Table Probability (P)	0.0590
Two-sided Pr <= P	0.2702

Sample Size = 299

با توجه به اینکه P-VALUE=0.2702 از 0.05 بیشتر است بنابراین فرض H0را می پذیریم یعنی بقای بیماری در عرض ده سال ربطی به نوع عمل جراحی استفاده شده دارد .

The SAS System 10:51 Friday, April 13, 2001 2

The FREQ Procedure

Table of D by B

D / B Frequency Percent Row Pct Col Pct	1	2	Total
1	31 10.37 19.62 20.81	127 42.47 80.38 84.67	158 52.84
2	118 39.46 83.69 79.19	23 7.69 16.31 15.33	141 47.16
Total	149 49.83	150 50.17	299 100.00

Statistics for Table of D by B

Statistic	DF	Value	Prob
Chi-Square	1	122.3342	<.0001
Likelihood Ratio Chi-Square	1	132.6134	<.0001
Continuity Adj. Chi-Square	1	119.7849	<.0001
Mantel-Haenszel Chi-Square	1	121.9251	<.0001
Phi Coefficient		-0.6396	
Contingency Coefficient		0.5388	
Cramer's V		-0.6396	

Fisher's Exact Test

Cell (1,1) Frequency (F)	31
Left-sided Pr <= F	2.623E-30
Right-sided Pr >= F	1.0000
Table Probability (P)	2.501E-30
Two-sided Pr <= P	3.129E-30

Sample Size = 299

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

با توجه به اینکه P-VALUE=3.129^30 از 0.05 کمتر است بنابراین فرض H0را رد می کنیم یعنی مرحله بیماری روی زنده یا مرده بودن بیماری تاثیری دارد.

The SAS System 10:51 Friday, April 13, 2001 3

The FREQ Procedure

Table of B by A

B A

Frequency Percent Row Pct Col Pct	1	2	Total
1	52 17.39 34.90 46.02	97 32.44 65.10 52.15	149 49.83
2	61 20.40 40.67 53.98	89 29.77 59.33 47.85	150 50.17
Total	113 37.79	186 62.21	299 100.00

Statistics for Table of B by A

Statistic	DF	Value	Prob
Chi-Square	1	1.0576	0.3038
Likelihood Ratio Chi-Square	1	1.0584	0.3036
Continuity Adj. Chi-Square	1	0.8265	0.3633
Mantel-Haenszel Chi-Square	1	1.0540	0.3046
Phi Coefficient		-0.0595	
Contingency Coefficient		0.0594	
Cramer's V		-0.0595	

Fisher's Exact Test

Cell (1,1) Frequency (F)	52
Left-sided Pr <= F	0.1817
Right-sided Pr >= F	0.8745
Table Probability (P)	0.0561
Two-sided Pr <= P	0.3405

Sample Size = 299

با توجه به اینکه P-VALUE=0.3405 از 0.05 بیشتر است بنابراین فرض H0را می پذیریم یعنی نسبت بیماران زنده در بین کسانی که پرتو درمانی را دریافت کرده اند یا دریافت نکردهاند تفاوت معنی داری دارد

The SAS System 10:51 Friday, April 13, 2001 4

The FREQ Procedure

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

Table 1 of B by D
Controlling for C=1

B D

Frequency Percent Row Pct Col Pct	1	2	Total
1	27 10.76 20.93 20.45	102 40.64 79.07 85.71	129 51.39
2	105 41.83 86.07 79.55	17 6.77 13.93 14.29	122 48.61
Total	132 52.59	119 47.41	251 100.00

Statistics for Table 1 of B by D
Controlling for C=1

Statistic	DF	Value	Prob
Chi-Square	1	106.6930	<.0001
Likelihood Ratio Chi-Square	1	116.4041	<.0001
Continuity Adj. Chi-Square	1	104.0965	<.0001
Mantel-Haenszel Chi-Square	1	106.2679	<.0001
Phi Coefficient		-0.6520	
Contingency Coefficient		0.5462	
Cramer's V		-0.6520	

Fisher's Exact Test

Cell (1,1) Frequency (F)	27
Left-sided Pr <= F	9.781E-27
Right-sided Pr >= F	1.0000
Table Probability (P)	9.372E-27
Two-sided Pr <= P	1.443E-26

Sample Size = 251

The SAS System 10:51 Friday, April 13, 2001 5

The FREQ Procedure

Table 2 of B by D
Controlling for C=2

B D

Frequency Percent

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

Row Pct Col Pct	1	2	Total
1	4 8.33 20.00 15.38	16 33.33 80.00 72.73	20 41.67
2	22 45.83 78.57 84.62	6 12.50 21.43 27.27	28 58.33
Total	26 54.17	22 45.83	48 100.00

Statistics for Table 2 of B by D
Controlling for C=2

Statistic	DF	Value	Prob
Chi-Square	1	16.1215	<.0001
Likelihood Ratio Chi-Square	1	17.0958	<.0001
Continuity Adj. Chi-Square	1	13.8486	0.0002
Mantel-Haenszel Chi-Square	1	15.7856	<.0001
Phi Coefficient		-0.5795	
Contingency Coefficient		0.5014	
Cramer's V		-0.5795	

Fisher's Exact Test

Cell (1,1) Frequency (F)	4
Left-sided Pr <= F	7.089E-05
Right-sided Pr >= F	1.0000
Table Probability (P)	6.665E-05
Two-sided Pr <= P	9.333E-05

Sample Size = 48

با توجه به اینکه P-VALUE=9.333^05 از 0.05 کمتر است بنابراین فرض H0را رد می کنیم یعنی در بین بیماران با جراحی محدود ایا زنده بودن بیماران وابسته به مرحله بیماری است

مثال٤: در موسسه مطالعات رفاهی دانمارک از کلیه افراد نمونه سئوال می شود, که ایا کارشان از نظر بدنی سخت و از نظر روانی شدید است ؟
اطلاعات دو سئوال فوق به تفکیک جنسیت به شرح زیر است.
مطلوبست پاسخ به سئوالات زیر:

۱) ایا نظر افراد در مورد سختی کار از نظر بدنی و از نظر روانی یکسان است؟
۲) ایا سختی بدنی مردان و زنان از نظر کارشان یکسان است؟
۳) ایا مشکل بودن کار از نظر روانی در مورد زنان و مردان از دیدگاه خودشان تفاوت دارد؟
۴) جواب سئوال یک به تفکیک جنسیت چگونه است؟

C	B	A		
جنسیت	وضعیت سختی کار از نظر بدنی	وضعیت سختی کار از نظر روانی		
		بله	تا حدودی	خیر

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

مردان	بلی	۱۱۳	۱۶۳	۲۷۰
	تا حدودی	۴۵	۱۰۶	۲۸۰
	خیر	۲۲۹	۳۴۳	۵۶۸
زنان	بلی	۱۰۰	۱۰۹	۲۰۲
	تا حدودی	۳۳	۸۹	۱۷۹
	خیر	۱۰۰	۱۷۹	۵۲۴

```
DATA;
DO C=1,2; DO B=1,2,3; DO A=1,2,3;
INPUT COUNT@@;
OUTPUT;
END; END; END;
CARDS;
113 163 370 45 106 280 229 343 568 100 109 202 33 89 179 100 179 524
;
PROC FREQ;
TABLES B*A C*B A*C B*A*C;
EXACT FISHER;
WEIGHT COUNT;
RUN;
```

۵-رویه TTEST

ازمون استیودنت در برخی موارد برای مقایسه میانگین یک جامعه با عدد ثابت , مقایسه میانگین دو جامعه وابسته به یکدیگر , مقایسه میانگین یک متغیر در دو جامعه مستقل مورد استفاده قرار می گیرد.

ازمون مقایسه میانگین با عدد ثابت

در این آزمون دو پذیره زیر بایستی برقرار باشد.

۱- متغیر مورد بررسی کمی از نوع مقیاس نسبتی یا فاصله ای باشد.

۲- یک نمونه تصادفی از جامعه استخراج شده باشد.

در عمل با توجه به اینکه توزیع متغیر مورد بررسی مشخص نیست و واریانس نیز معلوم نمی باشد حالتهای زیر اتفاق می افتد.

الف) اگر حجم نمونه استخراج شده از جمعیت بزرگ باشد.(n>=30) علی رغم مشخص نبودن توزیع متغیر پاسخ و واریانس ان می توان از آزمون استیودنت با توجه به قضیه حد مرکزی استفاده نمود.

ب) اگر حجم نمونه کم باشد (n<30) انگاه دو حالت زیر اتفاق می افتد.

ب۱) اگر توزیع جامعه نرمال باشد یا بتوانیم نشان دهیم که مشاهدات از توزیع نرمال استخراج شده اند می توانیم از توزیع استیودنت استفاده نمائیم.

ب۲) اگر توزیع مشاهدات غیر نرمال باشد نمی توانیم با استفاده از ازمون استیودنت فرضیه های فوق را ازمون نمود بلکه در این حالت بایستی از روشهای ناپارامتری مثل ازمونهای علامت یا رتبه علامتدار ویلکاکسون حالت تک نمونه ای استفاده کرد.

ازمون مقایسه میانگین دو جامعه وابسته

در استفاده از ازمون استیودنت برای مقایسه میانگین دو جامعه وابسته به یکدیگر , اگر شرایط برقرار باشد می توان از این ازمون برای انجام فرضیه های

$$\begin{cases} H0 : \mu 1=\mu \\ H1 : \sim H0 \end{cases}$$

استفاده کرد.

در عمل حالتهای زیر اتفاق می افتد.

الف) اگر حجم نمونه استخراج شده از جمعیت بزرگ باشد.(n>=30) علی رغم مشخص نبودن توزیع متغیر پاسخ و واریانس ان می توان از آزمون استیودنت با توجه به قضیه حد مرکزی استفاده نمود.

ب) اگر حجم نمونه کم باشد (n<30) انگاه دو حالت زیر اتفاق می افتد.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

ب۱) اگر توزیع جامعه نرمال باشد یا بتوانیم نشان دهیم که مشاهدات از توزیع نرمال استخراج شده اند می توانیم از توزیع استیودنت استفاده نمائیم.
ب۲) اگر توزیع مشاهدات غیر نرمال باشد نمی توانیم با استفاده از ازمون استیودنت فرضیه های فوق را ازمون نمود بلکه در این حالت بایستی از روشهای ناپارامتری مثل ازمونهای علامت یا رتبه علامتدار ویلکاکسون حالت زوج نمونه ای استفاده کرد.

ازمون مقایسه میانگین در دو جامعه مستقل

چنانچه بخواهیم میانگین را در جامعه مستقل مقایسه نمائیم , همانند حالت قبل در عمل حالتهای زیر اتفاق می افتد.

$$\begin{cases} H_0 : \mu_1=\mu_2 \\ H_1 : \sim H_0 \end{cases}$$

دو پذیره در انجام این ازمون فرضیه وجود دارد که عبارتند از:
۱- دو نمونه تصادفی و مستقل از دو جامعه مستقل استخراج شده باشد.
۲- متغیری که مقایسه براساس ان صورت می گیرد بایستی کمی باشد.
الف) اگر حجم نمونه های استخراج شده از جمعیت بزرگ باشد.$(n_1,n_2>=30)$
در اینصورت ازمون استیودنت برای مقایسه میانگین دو جامعه مستقل کاربرد دارد. برای مشخص کردن مقدار اماره ازمون در این حالت بایستی بدانیم که ایا فرضیه برابری واریانس های دو جامعه رد می شود یا خیر؟
براساس نتیجه این ازمون , مقدار اماره ازمون استیودنت و درجه آزادی ان را تعیین می کنیم.
نرم افزارهای اماری از جمله SAS و SPSS طوری طراحی شده اند که قبل از انجام ازمون استیودنت ازمون برابری واریانسها را براساس توزیع فیشیر انجام داده و P-VALUE ان را گزارش می کنند.
محقق می تواند با توجه به نتیجه ازمون برابری واریانس یکی از دو ازمون استیودنت انجام شده برای مقایسه میانگین ها را که براساس پذیرفتن برابری واریانسها و عدم ان انجام شده است انتخاب نماید.
ب) اگر حجم نمونه ها کم باشد $(n_1,n_2<30)$ انگاه دو حالت زیر اتفاق می افتد.
ب۱) اگر توزیع جامعه نرمال باشد یا بتوانیم نشان دهیم که مشاهدات از توزیع نرمال استخراج شده اند می توانیم از توزیع استیودنت استفاده نمائیم.
ب۲) اگر توزیع مشاهدات غیر نرمال باشد نمی توانیم با استفاده از ازمون استیودنت فرضیه های فوق را ازمون نمود بلکه در این حالت بایستی از روشهای ناپارامتری مثل ازمون میانه یا ازمون من –ویتنی ویلکاکسون استفاده کرد.

شکل کلی رویه TTEST

```
PROC TTEST OPTIONS;
CLASS VARIABLE;
PAIRED REQUEST;
VAR VARIABLE(S);
BY VARIABLE(S);
FREQ VARIABLE;
WEIGHT VARIABLE;
RUN;
```

عبارت CLASS و VAR

عبارت CLASS و VAR معمولا با یکدیگر در یک رویه استفاده می شوند و برای ازمون استیودنت حالت دو نمونه مستقل این دو عبارت کاربرد دارند.

عبارت PAIRED

عبارت PAIRED برای انجام ازمون استیودنت حالت دو نمونه وابسته به یکدیگر استفاده می شود.
استفاده همزمان از دو عبارت CLASS و PAIRED در یک رویه TTEST امکان پذیر نیست.
برخی از OPTION های این رویه عبارتند از:

۱- ALPHA= 0<P<1

از این انتخاب برای تغییر ضریب اطمینان فواصل اطمینان استفاده می شود که بطور پیش فرض ضریب اطمینان ۰,۹۵ می باشد.

۲-NONE یا UMPU یا CI=EQUL

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

از این انتخاب برای ارائه فاصله اطمینان برای واریانس استفاده می شود. EQUL حالت فاصله اطمینان با دو دنباله مساوی را در نظر می گیرد و UMPU فاصله اطمینان را براساس پرتوانترین ازمون ناااریب بطور یکنواخت محاسبه می کند.

۳- نامDATA=DATASET

۴- H0=IN

برای انجام ازمون فرضیه H0 که عدد ثابت تحت ان مخالف صفر می باشد.

CLASS برای تایپ نام متغیر گروهبندی شده که دو جمعیت مستقل را مشخص می کند , کاربرد دارد.

از زیر دستور PAIRED برای مشخص کردن زوج نمونه ها بصورت زیر استفاده می شود.

مانند PAIRED A*B C*D

عبارت اختیاری VAR نیز در ازمون استیودنت برای مقایسه میانگین دو جامعه مستقل برای لیست کردن نام متغیر با متغیرهای وابسته کاربرد دارد.

مثال۱: نمرات ۸ نفر از دانشجویان درس امار در ترم گذشته به قرار زیر است.

16 , 11 , 13 , 19 , 8 , 7 , 12 , 9

با فرض اینکه توزیع نمرات دانشجویان کلاس نرمال باشد. ایا می توان گفت میانگین نمرات دانشجویان درس امار برابر ۱۱ است؟

```
DATA;
INPUT Y@@;
CARDS;
16 11 13 19 8 7 12 9
;
PROC TTEST H0=11;
VAR Y;
RUN;
```

مثال ۲: سازمان حمایت از مصرف کنندگان می خواهد قیمت کالای بخصوصی را در دو شهر با هم مقایسه کند.برای این منظور نمونه تصادفی از ۶ فروشگاه در یک شهر و ۸ فروشگاه در شهر دیگر گرفته شده است. نتایج به شرح زیر است.

شهر ۱	۹	۱۱	۸	۱۲	۱۱	۹		
شهر ۲	۱۳	۱۴	۱۰	۱۱	۱۳	۱۲	۱۱	۱۲

الف) ایا دلیلی وجود دارد که پراکندگی قیمت این کالا در دو شهر تفاوت با یکدیگر داشته باشد؟

ب) ایا اختلاف معنی داری بین متوسط قیمت این کالا در دو شهر وجود دارد. برای انجام ازمون استیودنت هر فرضی که لازم است بپذیریم, عنوان کنید؟

```
DATA;
INPUT X Y @@;
CARDS;
1 9 1 11 1 8 1 12 1 11 1 9 2 13 2 14 2 10 2 11 2 13 2 12 2 11 2 12
;
PROC TTEST;
CLASS X;
RUN;
```

مثال ۳ : پزشک محققی می خواهد که تحقیق کند که قرص معینی اثر جانبی نامطلوب تقلیل فشار خون را بر روی مصرف کننده دارد یا نه؟

این مطالعه شامل ثبت فشار خون اولیه ۱۵ خانم است. بعد از اینکه افراد برای مدت ۶ ماه مرتبا دارو را مصرف می کنند فشار خون انها دوباره ثبت می شود. اطلاعات به شرح زیر است.

ازمودنی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
قبل از مصرف	۷۰	۸۰	۷۲	۷۶	۷۶	۷۶	۷۲	۷۸	۸۲	۶۴	۷۴	۹۲	۷۴	۶۸	۸۴
بعد از مصرف	۶۸	۷۲	۶۲	۷۰	۵۸	۶۶	۶۸	۵۲	۶۴	۷۲	۷۴	۶۰	۷۴	۷۲	۷۴

ایا مصرف قرص تاثیری روی فشار خون دارد؟

```
DATA;
INPUT  Y1 Y2@@;
CARDS;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

70 68 80 72 72 62 76 70 76 58 76 66 72 68 78 52 82 64 64 72 74 7 4 92 60 74 74 68
72 84 74
;
PROC TTEST;
PAIRED Y1*Y2;
RUN;

روشهای مورد استفاده برای برازش توزیع به مشاهدات

برای بررسی توزیع مشاهدات دو روش توصیفی و استنباطی کاربرد دارد. دلایلی وجود دارد که در عمل برای برازش توزیع نرمال که ممکن است که یکی از پذیره های ازمونهای پارامتری می باشد. معمولا توصیه می شود که از روشهای توصیفی استفاده شود.
دو دلیل عمده برای اینکه در عمل بیشتر از روشهای توصیفی برای چک کردن پذیره زیر بنایی نرمال بودن توزیع مشاهدات (در صورت لزوم) در ازمونهای پارامتری استفاده می شود این است که اولا در چنین ازمونهای تقریبا نرمال بودن مشاهدات کفایت می کند ثانیا چنانچه یک نمونه به حجم n بزرگ از توزیع نرمال تولید شده و سپس نمودار p-pیا q-q انها رسم شود ممکن است نقاط روی نمودارها از خط راست دقیقا پیروی نکند به این دلیل و اینکه ازمونهای اماری برای برازش توزیع نرمال مشاهدات انعطاف پذیر نیستند بهتر است در این حالت از روشهای توصیفی استفاده شود.

روشهای توصیفی

بهترین روش توصیفی جهت تشخیص توزیع استفاده از نمودارهای احتمال و احتمال تجمعی است
[Q-Q PLOT OR P-P PLOT]

نمودار توزیع فراوانی مشاهدات)(HISTOGRAM) تا حدودی مشخص می کند که مشاهدات حول پارامتر مکانی متقارن اند یا خیر.
تشخیص توزیع مشاهدات با استفاده از نمودار فوق امکان پذیر نیست.

نمودار P-P PLOT

با رسم نمودار پراکنش مقادیر احتمالات تجمعی مشاهدات در مقابل احتمالات تجمعی توزیع مورد نظر این امکان را فراهم می سازد تا بتوانیم تشخیص دهیم که ایا مشاهدات از توزیع مورد نظر تبعیت (یا تقریبا تبعیت) میکند یا خیر؟

نمودار Q-Q PLOT

نمودار Q-Q PLOT چندکهای توزیع مورد برازش را در مقابل چندکهای توزیع نمونه ای مشاهدات رسم کرده و این امکان را فراهم می کند تا بتوان در مورد برازش توزیع به مشاهدات حتی بطور تقریبی اظهار نظر کرد.
اگر نقاط روی نمودار P-P یا Q-Q از یک خط راست تبعیت کند گوئیم مشاهدات از توزیع مورد نظر پیروی می کند.اگر تقریبا یک خط راست توسط نقاط روی نمودار P-P یا Q-Q ظاهر شود نشان دهنده تقریبا پیروی کردن مشاهدات از ان توزیع است.

روشهای استنباطی

از ازمونهای نیکویی برازش برای بررسی استنباطی برازش توزیع خاصی به مشاهدات استفاده می شود.
معمولا ازمون نیکویی برازش پیرسون , ازمون نیکویی برازش کلموگروف-اسپیرنوف که هر دو ناپارامتری می باشد. برای این کار استفاده می شود.

توصیف متغیرهای کمی

جهت تبدیل ازمونهای دوطرفه به یکطرفه داریم:
از رویه FREQ برای توصیف تک بعدی و دو بعدی متغیرهای کیفی و همچنین استنباط دو بعدی انها استفاده می شود. در مورد توصیف متغیرهای کمی با استفاده از ابزار امار توصیفی وتا حدودی استنباط در مورد مقادیر و توزیع انها از رویه UNVARIATE استفاده می کنیم.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۶-رویه UNVARIATE

شکل کلی این رویه به شرح زیر است.

```
PROC UNIVARIATE OPTION;
BY VARIABLES;
FREQ VARIABLES;
HISTOGRAM VARIABLES / OPTION;
INSET KEYWORDS;
OUTPUT OUT=DATA نام ;
KEYWORDS=NAME(S);
PROBPLOT VARIABLE(S) / OPTION;
QQPLOT VARIABLE(S) / OPTION;
VAR VARIABLES;
WEIGHT VARIABLE;
RUN;
```

برخی از option های این رویه عبارتند از:

۱-ALL

برای محاسبه تمامی شاخصها , اماره ها و ازمونها مورد استفاده قرار می گیرد.

۲- ALPHA= 0<P<1

از این انتخاب برای تغییر ضریب اطمینان فواصل اطمینان استفاده می شود که بطور پیش فرض ضریب اطمینان ۰,۹۵ می باشد.

۳- CIBASIC TYPE= KEYWORDS ALPHA= 0<P<1

این OPTION فاصله اطمینان برای میانگین انحراف معیار و واریانس را تحت پذیره نرمال بودن مشاهدات ارائه می کند. نوع این ناحیه اطمینان می تواند بصورت UPPER, LOWER یا TWO SIDED در نظر گرفته شود.

مثال:

```
PROC UNIVARIATE CIBASIC
TYPEE=UPPER ALPHA=0.01;
```

در این انتخاب , پیش فرض نوع ناحیه اطمینان TWO SIDED و ALPHA=0.05 می باشد.

٤- CIPCTLDF TYPE=KEYWORDS ALPHA 0<P<1

برای مشخص کردن ناحیه اطمینان برای مقادیر صدکها با عدم پذیرش فرض نرمال بودن مشاهدات.
که در این OPTION نوع و احتمال خطای نوع اول عینا مشابه CIBASIC , OPTION می باشد.

۵- CI PCTL NORMAL

برای محاسبه ناحیه اطمینان برای صدکها با فرض نرمال بودن مشاهدات که TYPE و ALPHA را می توان مانند دو OPTION قبلی برای این نوع فاصله اطمینان در نظر گرفت. پیش فرضهای این انتخاب مانند دو OPTION قبلی می باشد.

٦- نام DATA=DATA SET

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۷- FREQ

برای ارائه جداول فراوانی تمامی متغیرهای موجود در DATA SET و معرفی شده در عبارتVAR

۸- MU0=M

برای مشخص کردن مقدار پارامتر مکانی تحت فرضیه H0 , که در این OPTION می توان بیش از یک مقدار برای مقایسه پارامتر مکانی با ان توسط ازمونهای پارامتری و ناپارامتری مشخص کرد.

مثال:

```
PROC UNIVARIATE
MU0= 10 15 20;
VAR Y;
RUN;
```

بنابراین این انتخاب برای انجام ازمونهای ناپارامتری تک نمونه ای همانند ازمون علامت و رتبه علامتدار ویلکاکسون برای مقایسه میانه جامعه با عدد ثابت تحت فرضیه H0 , بکار می رود.

۹- NORMAL

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

برای گزارش تمامی ازمونها نیکویی برازش برای بررسی نرمال بودن مشاهدات.(اماره شاپیرو-ویلک و سایر ازمونهای نیکویی برازش)

PLOTS -۱۰

برای رسم تمامی نمودارهای جعبه ای و احتمال نرمال برای متغیرهای معرفی شده به رویه

TRIMMED VALUE(S) TYPE=KEYWORD ALPHA= 0<P<1 -۱۱

برای محاسبه میانگین اصلاح شده مشاهدات و ارائه یک ناحیه اطمینان با ضریب اطمینان مشخص شده برای ان .

میانگین اصلاح شده حالت خاصی از میانگین پیراسته می باشد که در ان برخی از مشاهدات دنباله ابتدائی و انتهائی داده ها حذف شده و بین مقادیر باقیمانده میانگین حسابی را در نظر می گیریم.

بعنوان مثال رویه زیر میانگین اصلاح شده متغیر Y را با حذف ۵ مشاهده از دنباله ابتدائی و انتهائی مشاهدات همراه با فاصله اطمینان ۰,۹۵ برای ان گزارش می کند.

```
PROC UNIVARIATE
TRIMMED=5;
VAR Y;
RUN;
```

TYPE برای مشخص کردن نوع ناحیه اطمینان بصورت LOWER , UPPER یا TWO SIDED در نظر گرفته می شود.

از =ALPHA نیز برای تعیین ضریب اطمینان این ناحیه اطمینان استفاده می شود.

WINSORIZED=VALUE(S) TYPE=KEYWORD ALPHA= 0<P<1 -۱۲

این OPTION برای محاسبه میانگین ونیزوری مشاهدات در نظر گرفته می شود. علاوه بر این همانند میانگین اصلاح شده یک ناحیه اطمینان با ضریب اطمینان مشخص می توان برای میانگین ونیزوری مشاهدات در نظر گرفت.

مثال: اگر داده ها بصورت مقابل باشند TRIMMED=1 , WINSORIZED=1 داده ها چقدر است؟

-500 3 5 7 12 13 1000

3 5 7 12 13 TRIMMED=8

3 3 5 7 12 13 13 WINSORIZED=8

زیر دستورهای رویه UNIVARIATE

BY VARIABLE(S); -۱

برای توصیف یک یا چند متغیر کمی براساس مقادیر یک یا چند متغیر کیفی که در جلوی عبارت BY لیست می شود در نظر گرفته می شود.

FREQ VARIABLE; -۲

برای مشخص کردن یک و تنها یک متغیر شمارشی که شامل فراوانی مشاهدات می باشد مورد استفاده قرار می گیرد.

بعنوان مثال می توان داده ها را به دو صورت زیر به نرم افزار معرفی کرد.

Y	Y1	C
15	15	3
20	20	2
25	25	2

15
20
25
15

HISTOGRAM VARIABLE(S) / OPTION; -۳

از این عبارت برای رسم نمودار توزیع فراوانی مشاهداتی که متغیرهای متناظر با انها در جلوی عبارت لیست می شوند, استفاده می شود. از OPTION های این عبارت برای رسم نمودار تابع چگالی برخی از توزیعهای متعارف اماری روی نمودار هیستوگرام استفاده می شود.

برخی از OPTION های این عبارت بصورت زیراند.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

1-NORMAL 2-BETA 3-EPONENTIAL 4-GAMMA
5-LOGNORMAL 6-KERNAL 7-WEIBULL

۴- INSET KEYWORD(S);

از این عبارت برای محاسبه برخی شاخصها که توسط KEYWORD مشخص می شود , استفاده می کنیم.
برخی از این KEYWORD های این عبارت به شرح زیراند.

۱-مجمع مربعات تصحیح شده	1-CSS
۲-ماکزیمم مشاهدات	2-MAX
۳-مینیمم مشاهدات	3-MIN
۴-ضریب چولگی پیرسون و انحراف معیار ان	4-SKEWNESS
۵-مجموع مشاهدات	5-SUM
۶-ضریب تغییرات	6-CV
۷-میانگین حسابی مشاهدات	7-MEAN
۸-نمای مشاهدات	8-MODE
۹-انحراف معیار	9-STD
۱۰-ضریب کشیدگی و انحراف معیار ان	10-KURTOSIS
۱۱-تعداد مشاهدات	11-N
۱۲-دامنه مشاهدات	12-RANG
۱۳-خطای استاندارد میانگین	13-STD MEAN
۱۴-واریانس	14-VAR
۱۵-میانه	15-MEDIAN
۱۶-صدک مرتبه اول	16-P1
۱۷-صدک مرتبه پنجم	17-P5
۱۸-چارک مرتبه اول	18-Q1

۵- نام OUTPUT OUT= DATA SET

KEYWORD(S)=NAME(S);

برای مشخص کردن مقادیر برخی شاخصها تحت عنوان متغیرهای و فرستادن انها به مجموعه داده های خروجی
KEYWORD های این عبارت مشابه KEYWORDهای عبارت INSET است.

۶- PROB PLOT VARIABLE(S) / OPTION;

برای رسم نمودار احتمال مشاهدات مورد استفاده قرار می گیرد. متغیر یا متغیرهای مورد نظر حداقل با یک فاصله در این عبارت لیست می شوند.
از OPTION نیز برای مشخص کردن نوع توزیع مورد بررسی استفاده می شود.
OPTION های این عبارت همانند OPTIONهای عبارت HISTOGRAM میباشد.

۷- QQ PLOT VARIABLE(S) / OPTION;

برای رسم نمودار احتمال تجمعی مشاهدات استفاده می شود.
OPTION های ان همانند عبارت قبلی می باشد.

۸- VAR VARIABLE(S);

برای مشخص کردن نام متغیرهایی که بایستی رویه روی انها اعمال شود.

۹- WEIGHT VARIABLE;

برای مشخص کردن یک و فقط یک متغیر که شامل وزن مشاهدات می شود.
مثال۱ : یک نمونه تصادفی ۱۶تائی از لاستیک های تولیدی یک شرکت در دو شیفت صبح و عصر برای مقایسه طول عمر انها در دو شیفت گرفته شده است.

صبح	22.42 18.36 21.64 19.20 23.40 27.38 23.46
عصر	23.51 20.62 26.47 19.75 20.30 17.84 26.34 28.27 24.37

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

الف) مشاهدات را توصیف کنید؟
ب) ایا طول عمر لاستیکها از نظر پراکندگی یکسان است؟
ج) ایا می توان گفت بطور متوسط طول عمر لاستیکهای تولید شده در شیفت عصر بیشتر از شیفت صبح است؟

```
DATA;
INPUT X Y@@;
CARDS;
1  22.42  1  18.36  1  21.64  1  19.20  1  23.40  1  27.38  1  23.46 2  23.51 2  20.62
 2  26.47  2  19.75  2  20.30  2  17.84  2  26.34  2  28.27  2  24.37
;
PROC UNIVAIATE;
VAR Y;
HISTOGRAM Y / NORMAL;
QQPLOT Y / NORMAL;
BY X;
PROC TTEST;
CLASS X;
VAR Y;
RUN;
```

Statistics

Variable	X	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
Y	1	7	19.493	22.266	25.039	1.9321	2.9983	6.6024	1.1332
Y	2	9	20.291	23.052	25.813	2.4263	3.5921	6.8816	1.1974
Y	Diff (1-2)		-4.408	-0.787	2.8349	2.453	3.3505	5.2841	1.6885

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
Y	Pooled	Equal	14	-0.47	0.6485
Y	Satterthwaite	Unequal	13.9	-0.48	0.6407

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
Y	Folded F	8	6	1.44	0.6785

با توجه به اینکه P-VALUE=06785 از 0.05 بیشتر است بنابراین فرض H0را قبول می کنیم یعنی بطور متوسط طول عمر لاستیکهای تولید شده در شیفت عصر بیشتر از شیفت صبح است.

مثال۲: جهت بررسی اختلاف بین طول عمر چهار نوع باطری مارک الف , ب , ج , د , چهار باطری از هر نوع از هر باطری بطور تصادفی انتخاب شده است.نتایج به شرح زیر است.
الف) طول عمر باطریها را در کل و به تفکیک نوع توصیف نمائید؟
ب) ایا بطور متوسط طول عمر باطریهای انواع الف , ب , ج , د تفاوت معنی داری دارد؟

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

مارک	طول عمر (ساعت)
الف	10 18 15 12
ب	19 12 17 14
ج	23 20 19 21
د	20 25 21 14

```
DATA;
DO X=1 TO 4; DO I=1 TO 4;
INPUT Y@@;
OUTPUT;
END;END;
CARDS;
10 18 15 12 19 12 17 14 23 20 19 21 20 25 21 14
;
PROC UNIVARIATE;
VAR Y;
HISTOGRAM Y / NORMAL;
QQPLOT Y / NORMAL;
PROC UNIVARIATE;
VAR Y;
HISTOGRAM Y / NORMAL;
BY X;
RUN;
```

برای انجام ازمون مقایسه میانگین K جامعه مستقل(K>2) بایستی از انالیز واریانس یکطرفه استفاده کنیم.

انالیز واریانس یکطرفه

برای ازمون فرضیه های

$$\begin{cases} H_0 : \mu_1=\mu_2=\ldots=\mu_K \\ H_1 : \sim H_0 \end{cases}$$

از انالیز واریانس یکطرفه استفاده می کنیم. در واقع انالیز واریانس یکطرفه برای مقایسه میانگین K جامعه مستقل کاربرد دارد.
برای بررسی تاثیر یک متغیر مستقل با K سطح روی یک متغیر پاسخ کمی نیز از انلیز واریانس یکطرفه استفاده می شود. در انالیز واریانس چهار پذیره زیر بایستی برقرار باشد.
الف) خطاهای معادله خطی بایستی مستقل باشند.
ب) میانگین خطاها بایستی صفر باشد.
ج) واریانس خطاها یا مشاهدات بایستی برابر مقدار ثابت سیگمادو باشد.
د) توزیع مشاهدات(خطاها) بایستی نرمال باشد. تقریبا نرمال بودن توزیع مشاهدات نیز کفایت می کند.
در این ازمون مقدارحجم نمونه انتخاب شده مهم نیست. یعنی حجم نمونه بزرگ نمی تواند پذیره زیر بنائی چهارم را مرتفع سازد.
برای بررسی نرمال بودن مشاهدات در انالیز واریانس به جهت انعطاف در تصمیم گیری بهتر است در عمل از روشهای توصیفی همانند نمودار P-P یا Q-Q استفاده نمائیم.
برای انجام ازمون واریانس یکطرفه , دو طرفه و چندطرفه , وقتی سطوح اثرات تثبیت شده می باشد از رویه ANOVA استفاده می کنیم.

۷-رویه ANOVA

شکل کلی این رویه بصورت زیر است.

```
PRO ANOVA OPTIONS;
CLASS VARIABLE(S)/OPTION;
MODEL متغیر وابسته = اثر یا اثرات /OPTION;
BY VARIABLE(S);
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

FRQ VARIABLE;
MEANS اثر یا اثرات /OPTION;
TEST H=اثر یا اثرات E=اثر /OPTION;
RUN;

برخی از OPTIONهای این رویه عبارتند از:

۱- نام DATA=DATA SET

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۲- NAMELEN=N 20<N<200

برای مشخص کردن تعداد کاراکتر های نام مقادیر اثرات

۳- خروجی OUTSTAT=DATA SET

برای فرستادن نتایج جدول انالیز واریانس به مجموعه داده خروجی

زیر دستورهای رویه ANOVA

۱- زیر دستور CLASS

این زیر دستور برای مشخص کردن نام اثر یا اثرات در مدل انالیز واریانس در نظر گرفته می شود. در واقع متغیر های مستقل که بصورت متغیرهای گروهبندی شده می باشند در جلوی این عبارت حداقل با یک فاصله لیست می شوند.

چون در رویه ANOVA تنها انالیز واریانس با اثرات تثبیت شده امکانپذیر است. بنابراین در این عبارت بایستی عوامل همه با سطوح تثبیت شده باشند.

تنها OPTION این عبارت TRUNCATE می باشد که چنانچه بخواهیم به نرم افزار اطلاع دهیم که برای عوامل دارای سطوح بیش از ۱۶ سطح انالیز واریانس را با حداکثر ۱۶ سطح اول این عوامل انجام دهد. از این OPTION استفاده می کنیم.

۲- زیر دستور MODEL

عبارت MODEL یکی از زیر دستورات اجباری این رویه می باشد. در این عبارت مدل مورد نظر بصورت زیر ارائه می شود.

نام اثر یا اثرات = متغیر وابسته

مثال۱) اگر متغیر پاسخ را با Y نشان داده و بخواهیم اثر سه عامل A,B,C که تثبیت شده هستند را روی ان در بگیریم با توجه به مدلهای مختلف طرح آزمایشها می توان عبارت مدل را بصورت زیر در نظر گرفت.

MODEL Y=A|B|C;

این مدل کامل است و معادل است با:

MODEL Y=A B A*B C A*C B*C A*B*C;

نکته: تفاوت دو مدل فوق در این است که در مدل ۲ می توان اثرات را جابجا کرد. مدل ۲ در طرحهای کرت, کرت یکبار خرد شده , طرح کرت دوبار خرد شده کاربرد دارد.

مثال۲) طرح بلوکی که عامل A عامل اصلی و B عامل بلوک است.

MODEL Y=A B;

مثال۳) طرح لانه ای که bداخل A بصورت تودرتو می باشد.

MODEL Y= A B(A);

مثال٤) طرح عامل تودرتو , عامل B در داخل A بصورت تودرتو است.

MODEL Y= A B(A) C A*C C*B(A);

برخی از OPTIONهای این زیر دستور عبارتند از:

۱) INT یا INTERCEPT

این رویه بطور پیش فرض عرض از مبدا یا ضریب ثابت را در مدل در نظر گرفته ولی ازمون وجود یا عدم وجود ان را در مدل به خروجی نمی فرستد برای اینکه نتیجه ازمون معنی داری میانگین کل یا ضریب ثابت در خروجی ارائه شود از این OPTION استفاده می کنیم.

۲) NOUNI

با توجه به اینکه با این رویه می توان انالیز واریانس چند متغیره نیز انجام داد چنانچه در انالیز واریانس چند متغیره نیازی به انجام انالیز واریانسهای یک متغیره نداشته باشیم. می توانیم جهت حذف نتایج ان از خروجی از این OPTION استفاده نمائیم.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۳- زیر دستور TEST

از این زیر دستور برای انجام ازمونهای معنی داری اثر یا اثرات براساس خطای مورد نظر استفاده می کنیم.
اگر بخواهیم اثر عامل A را در مقایسه با خطائی همانند اثر متقابل A*B ازمون کنیم, زیر دستور TEST را بصورت زیر می نویسیم.

```
TEST H=A E=A*B;
```

نکته: استفاده همزمان از دو یا چند زیر دستور TEST در یک رویه ANOVA جایز است.

۴- زیر دستور MEANS

این زیر دستور برای مقایسه ازمونهای چندگانه (MULTI COMPARISON) مورد استفاده قرار می گیرد استفاده همزمان از این زیر دستور در یک رویه برای چند عامل مجاز می باشد.
با یک زیر دستور MEANS می توان ازمونهای مقایسه چندگانه متعددی را برای یک یا چند عامل انجام داد.
OPTIONهای این زیر دستور برای مشخص کردن نوع ازمونهای مقایسه چندگانه مورد استفاده قرار می گیرد. در صورتیکه اماره فیشیر در جدول انالیز واریانس معنی دار بودن اثر را نشان دهد می توان از تمامی ازمونهای مقایسه چندگانه استفاده کرد. ولی وقتی اثر عامل در جدول انالیز واریانس معنی دار نباشد , از ازمون کمترین تفاوت معنی داری (LSD) نمی توان استفاده کرد بلکه بهتر است برای کشف تفاوتهای جزئی بین میانگین های حاشیه ای متغیر پاسخ از سایر ازمونهای مقایسه چند گانه استفاده کرد.
ازمون چندگانه دانکن (DUNCAN) در بین تمامی ازمونها محافظه کار است به این مفهوم که می تواند تفاوتهای جزئی بین میانگینهای حاشیه ای را تشخیص دهد.
ازمون دانت را نیز هنگامیکه یکی از سطوح عوامل گروه شاهد (کنترل) می باشد معمولا استفاده می کنیم.
برخی از OPTIONهای این زیر دستور عبارتند از:

۱- ALPHA= 0<P<1

برای مشخص کردن احتمال خطای نوع اول ازمونهای مقایسه های چندگانه مورد استفاده قرار می گیرد.

۲- T یا LSD

ازمون کمترین تفاوت معنی داری

۳- DUNCAN

ازمون چند دامنه ای دانکن

٤- (' گروه کنترل') DUNNETT

ازمون دانت دو طرفه

٥- (' گروه کنترل') DUNNETTU

ازمون دانت یکطرفه بزرگتری

٦- (' گروه کنترل') DUNNETTL

ازمون دانت یکطرفه کوچکتری

۷- BON

ازمون مقایسه چندگانه بونفرونی

۸- SHEFFE

ازمون شفه

۹- SNK

ازمون نیومن-کولز

۱۰- TUKEY

ازمون توکی

۱۱- WALLER

ازمون والر- دانکن

۱۲- اثر خطا = E

۸-رویه *CORR*

دستور CORR برای بررسی ارتباط بین متغیرهای کمی مورد استفاده قرار می گیرد.با این دستور می توان ضرایب همبستگی خطی پارامتری و ناپارامتری بین ویژگیهای کمی یا کیفی از نوع رتبه ای را محاسبه کرد.
علاوه بر این ازمون معنی داری این ضرایب همبستگی نیز با این دستور امکان پذیر است.
اگر دو متغیر کمی مورد نظر باشد و هدف از انجام تحقیق بررسی ارتباط انها باشد , می توان از ضریب همبستگی خطی پیرسون استفاده کرد. و ارتباط انها را از نظر این شاخص تعیین نمود.ولی چنانچه حداقل یکی

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

از متغیرهای مورد بررسی کیفی از نوع مقیاس رتبه ای باشد بایستی برای بررسی ارتباط انها از ضریب همبستگی خطی ناپارامتری استفاده نمود.
ازمون معنی داری ضریب همبستگی خطی پیرسون , اسپیرمن و کندال- تاو در واقع ازمون فرضیه $H_0: \rho = 0$ در مقابل ۳ دسته H_1 زیر است.

1) $H_1: \rho \neq 0$ 2) $H_1: \rho > 0$ 3) $H_1 : \rho < 0$

که در ان ρ ضریب همبستگی خطی بین دو ویژگی در جامعه است.
شکل کلی این رویه بصورت زیر است.

```
PROC CORR OPTION;
BY VARIABLE(S);
FREQ VARIABLE;
PARTIAL VARIABLE(S);
VAR VARIABLE(S);
WITH VARIABLE(S);
WEIGHT VARIABLE;
RUN;
```

عبارت WITH یک عبارت اجباری می باشد ولی سایر عبارات بنا به ضرورت می توانند مورد استفاده قرار نگیرند.
برخی از OPTIONهای این رویه عبارتند از:

۱) نام DATA SET ورودی= DATA

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۲) NOMISSING

برای حذف مشاهدات گمشده از تجزیه و تحلیلها

۳) KENDALL

برای محاسبه ضریب همبستگی خطی کندال و ازمون معنی داری ان

۴) PEARSON

برای محاسبه ضریب همبستگی خطی پیرسون و ازمون معنی داری ان

۵) SPEARMAN

برای محاسبه ضریب همبستگی خطی اسپیرمن و ازمون معنی داری ان

۶) ALPHA:

برای محاسبه ضریب قابلیت اطمینان αی کرانباخ

۷) NOCORR

حذف نتایج مربوط به ضریب همبستگی خطی پیرسون

۸) NOPROB

حذف تمامی مقادیر P-VALUEها

۹) COV

محاسبه کواریانس

۱۰) CSSCP

محاسبه مجموع مربعات تصحیح شده و مجموع حاصلضربها

زیر دستورهای رویه CORR

زیر دستور PARTIAL

از زیر دستور PARTIAL برای لیست کردن متغیرهای که لازم است ضرایب همبستگی جزیی به شرط حذف انها محاسبه شود , مورد نظر قرار می گیرد.

زیر دستور WITH

برای مشخص کردن نام متغیرهای است که قرار است همبستگی بین ان متغیرها و متغیرهای مشخص شده در عبارت VAR , ارزیابی شود.

مثال: عبارت

```
PROC CORR;
VAR X1 X2;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
WITH X3 X4;
RUN;
```

ضرایب همبستگی خطی بین X2,X4 X2,X3 X1,X4 X1,X3 را محاسبه می کند.

مثال: اطلاعات زیر مقدار مصرف پارچه برای هر نفر (Y) , درامد هر نفر (X2) و قیمت نسبی پارچه (X3) را طی سالهای ۱۹۲۳ تا ۱۹۳۹ نشان می دهد.

سال	Y	X2	X3
۱۹۲۳	۹۹,۲	۹۶,۷	۱۰۱
۱۹۲۴	۹۹	۹۸,۱	۱۰۰,۱
۱۹۲۵	۱۰۰	۱۰۰	۱۰۰
۱۹۲۶	۱۱۱,۶	۱۰۴,۹	۹۰,۶
۱۹۲۷	۱۲۲,۲	۱۰۴,۹	۸۶,۵
۱۹۲۸	۱۱۷,۶	۱۰۹,۵	۸۹,۷
۱۹۲۹	۱۲۱,۱	۱۱۰,۸	۹۰,۶
۱۹۳۰	۱۳۶	۱۱۲,۳	۸۲,۸
۱۹۳۱	۱۵۴,۲	۱۰۹,۴	۷۰,۱
۱۹۳۲	۱۵۳,۶	۱۰۵,۳	۶۵,۴
۱۹۳۳	۱۵۸,۵	۱۰۱,۷	۶۱,۳
۱۹۳۴	۱۴۰,۶	۹۵,۴	۶۲,۵
۱۹۳۵	۱۳۶,۲	۹۶,۴	۶۳,۶
۱۹۳۶	۱۶۸	۹۷,۶	۵۲,۶
۱۹۳۷	۱۵۴,۳	۱۰۲,۴	۵۹,۷
۱۹۳۸	۱۴۹	۱۰۱,۶	۵۹,۵
۱۹۳۹	۱۶۵,۵	۱۰۳	۶۱

در سطح ۰,۰۵درصد مشخص کنید.

الف) ایا میزان مصرف پارچه با افزایش زمان , افزایش می یابد؟

R y,year =0.895 P=VALUE=0.0001 =>RH_0

ب) ایا میزان مصرف پارچه به میزان درامد افراد وابسته است؟

R y,x2 =0.05 P=VALUE=0.847 =>AH_0

ج) ایا میزان مصرف پارچه با افزایش قیمت نسبی ان کاهش یافته است؟

R y,year = -0.947 P=VALUE=0.0001 =>RH_0

ابتدا اطلاعات فوق را بوسیله یک نرم افزار داده ای (مثلا EXCEL) ثبت می کنیم.
دقت کنید داده ها را با پسوند PRN ذخیره کرده باشید.

```
DATA;
INFILE'….PRN';
INPUT YEAR Y X2 X3;
PROC CORR;
VAR Y;
WITH X2 X3 YEAR;
RUN;
```

۹- رویه GLM

از این زیر دستور برای برازش کلیه مدلهای خطی عمومی همانند انالیز واریانس یک متغیره , یکطرفه یا چند طرفه در صورت وجود یا عدم وجود عوامل با سطوح تثبیت شده یا تصادفی شده , انالیز کواریانس , انالیز رگرسیون خطی ساده و چندگانه , انالیز واریانس چند متغیره , انالیز کواریانس چند متغیره , انالیز رگرسیون چند متغیره , طرح اندازه های مکرر و غیره استفاده می شود.
شکل کلی این رویه بصورت زیر است.

```
PROC GLM OPTIONS;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
CLASS VARIABLE(S);
MODEL متغیر یا متغیرهای مستقل = متغیر یا متغیرهای وابسته / OPTIONS;
BY VARIABLE(S);
FREQ VARIABLE;
ID VARIABLE(S);
WEIGHT VARIABLE;
CONTRAST ' ضرایب مقابله  نام اثر 'برچسب مقابله / OPTIONS;
ESTIMAT ' ضرایب  نام متغیر مستقل 'برچسب فرضیه / OPTIONS;
MANOVA TEST OPTIONS / DETAILES OPTIONS;
MEANS نام اثر یا اثرات / OPTION;
OUTPUT OUT=خروجی DATA SET نام  KEYWORDS = نام متغیرها / OPTIONS;
RANDOM نام اثرات ;
REPEATED فاکتورها / OPTIONS;
TEST H= نام اثر یا اثرات E = نام اثر / OPTIONS;
RUN;
```

برخی از OPTIONهای رویه GLM عبارتند از:

۱- ALPHA= 0<P<1

از این انتخاب برای مشخص کردن ضریب اطمینان حدود اطمینان استفاده می شود.بطور پیش فرض $\alpha=0.05$ می باشد.

۲- نام DATA SET ورودی = DATA

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۳- NAMELEN=n

برای مشخص کردن طول نام مقادیر اثرات که20<=n<=200 می باشد.

۴- NOPRINT

برای حذف تمامی نتایج مربوط به بررسی نرمال بودن مشاهدات

۵- FREQ یا FORMATED یا ORDER=DATA

برای مشخص کردن ترتیب گزارش نتایج برای متغیرها

۶- OUTSTAT= خروجی DATA SET

برای فرستادن نتایج جدول انالیز واریانس به یک مجموعه داده خروجی

زیر دستورهای رویه GLM

زیر دستورهای مهم رویه GLM عبارتند از:

۱- ;OPTION/ نام متغیرهای مستقل CLASS

این عبارت یک زیر دستور اختیاری در GLM می باشد. برای مشخص کردن عوامل یا متغیرهای مستقل کیفی از این رویه استفاده می شود.

۲- ; نام متغیر FREQ

برای مشخص کردن یک و فقط یک متغیر شمارشی شامل فراوانی مشاهدات

۳- ID VARIABLE(S)

۴- ; نام متغیر WEIGHT

برای معرفی یک و تنها یک متغیر وزنی شامل وزن مشاهدات

۵- ;OPTIONS/ ضرایب مقابله نام اثر 'برچسب مقابله ' CONTRAST

از این زیر دستور برای معرفی مقابله های متعامد براساس عوامل معرفی شده در عبارت CLASS استفاده می شود.

نکته۱) محدودیتی در تعداد موارد استفاده از این زیر دستور در یک رویه GLM وجود ندارد.

نکته۲) قسمت برچسب در این زیر دستور اختیاری است.

OPTIONهای مورد استفاده در CONTRAST عبارتند از:

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۵-۱- نام اثر = E
برای مشخص کردن خطای لازم برای ازمون مقابله
۵-۲ E
برای گزارش ضرایب مقابله در خروجی
۵-۳- ;4 یا 3 یا 2 یا TYPE=1
برای مشخص کردن نوع مجموع مربعات مورد نظر برای ازمون مقابله

۶-; نام متغیرها = KEYWORDS نام DATA SET خروجی=OUTPUT OUT
برای مشخص کردن مقادیر برخی متغیرهای محاسبه شده در رویه همانند مقادیر برآورد شده و مقادیر خطاها در یک مجموعه داده خروجی مورد استفاده قرار می گیرد.

```
OUTPUT OUT=SAS1 P=PRED R=RESI;
```

زیر دستور فوق مقادیر براورد شده را تحت عنوان متغیر PRED و مقادیر خطاها را تحت عنوان متغیر RESI در مجموعه داده خروجی با نام SAS1 ذخیره می کند.

۷- ;OPTION/ نام اثرات RANDOM
از این عبارت برای مشخص کردن اثرات با سطوح تثبیت شده استفاده می شود.
یکی از انتخابهای مهم این عبارت , TEST می باشد. که در صورت استفاده از ان نرم افزار ازمون مناسب در مورد بررسی معنی داری هر اثر را از جدول EMS تشخیص داده و گزارش می کند.

۸- ;OPTIONS / ضرایب نام متغیر مستقل 'برچسب فرضیه ' ESTIMAT
برای ازمون یک ترکیب خطی از ضرایب مدل همانند زیر دستور CONTRAST کاربرد دارد.
یکی از انتخابهای این زیر دستور E می باشد که برای گزارش بردار L در خروجی کاربرد دارد.

۹- ;OPTION / نام اثر یا اثرات MEANS
برای انجام ازمونهای مقایسه های چندگانه مورد استفاده قرار می گیرد. عملکرد ان شبیه عملکردش در رویه ANOVA است.

۱۰- ;OPTIONS / متغیر یا متغیرهای مستقل = متغیر یا متغیرهای وابسته MODEL
بسته به انکه مدل برازش شده به مشاهدات , انالیز یک متغیره یا چند متغیره باشد , تعداد متغیرهای وابسته تفاوت دارد.
در انالیز یک متغیره (انالیز واریانس , رگرسیون چندگانه یا ساده و انالیز کواریانس) متغیر وابسته یک متغیر در نظر گرفته می شود.

```
PROC GLM;
CLASS A B;
MODEL Y=A B;
RUN;
```

انالیز واریانس دو طرفه توسط دو عامل A,B

```
PROC GLM;
MODEL Y=X;
RUN;
```

رگرسیون ساده Y بر X

```
PROC GLM;
MODEL Y= X1|X2;
RUN;
```

مدل رگرسیون چندگانه Y بر X1 , X2 با اثر متقابل X1*X2

```
PROC GLM;
CLASS A;
MODEL Y=A X;
RUN;
```

انالیز کواریانس توسط هم عامل X و متغیر عامل A

```
PROC GLM;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
CLASS A B C;
MODEL Y= A|B C(A) B*C(A);
```

هم عامل : یک متغیر مستقل کمی است که به نوعی به متغیر وابسته مرتبط است که هم اثرش مهم است و هم مقدار عددی اش .

نکته) انالیز کواریانس ANCOVA

انالیز واریانس چند متغیره MANOVA

انالیز کواریانس چند متغیره MANCOVA

برخی از انتخابهای این زیر دستور عبارتند از:

۱- INTERCEPT

برای گزارش ازمون ضرورت وجود میانگین کل (ضریب ثابت) در مدل

۲- NOINT

برای در نظر گرفتن مدل بدون ضریب ثابت

۳- SS4 یا SS3 یا SS2 یا SS1

برای مشخص کردن انجام انالیز براساس یکی از چهار نوع مجموع مربعات

۴- ALPHA= 0<P<1

برای مشخص کردن ضریب اطمینان فواصل اطمینان

۵- CLI

برای مشخص کردن فاصله اطمینان برای مقادیر برآورد شده فردی

۶- CLM

برای تعیین فاصله اطمینان برای میانگین مقادیر برآورد شده

۷- CLPARM

برای مشخص کردن فاصله اطمینان برای برآورد ضرایب مدل رگرسیون

۸- P

برای مشخص کردن مقادیر مشاهده شده برآورد شده و خطاها و محاسبه اماره دربین-واتسون

اماره دربین-واتسون برای تشخیص خود همبستگی یا خود همبستگی خطاها کاربرد دارد. اگر مقدار این اماره تقریبا ۲ باشد نشان دهنده عدم خود همبستگی خطاها خواهد بود . چنانچه مقدار اماره فوق بیش از ۲ باشد , خود همبستگی منفی و چنانچه مقدار اماره کمتر از ۲ باشد , خود همبستگی مثبت را بین مقادیر خطاها نشان خواهد داد.

۹- I یا INVERSE

برای گزارش وارون ماتریس طرح تعمیم یافته

$$\begin{pmatrix} (X'X)^{-1} & (X'X)^{-1}X'Y \\ Y'X(X'X)^{-1} & Y'Y - Y'X(X'X)^{-1}X'Y \end{pmatrix}$$

۱۰- XPX

برای مشخص کردن ماتریس حاصلضربها

$$\begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}$$

مثال- فرض کنید اطلاعات زیر در مورد یک طرح مربع لاتین موجود است .

	1	2	3	4
1	C=12	D=11	A=13	B=8
2	B=16	C=12	D=11	A=13
3	A=17	B=14	C=15	D=9
4	D=13	A=16	B=13	C=9

ایا مدل $Y_{ijk}=\mu+T_i+B_j+\gamma_k+\varepsilon_{ijk}$ برای این طرح مناسب است؟

```
DATA;
DO X1=1 TO 4;
DO X2=1 TO 4;
INPUT X3 $ Y @@;
OUTPUT;
END;END;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
DATA LINES;
C 12 B 16 A 17 D 13 D 11 C 12 B 14 A 16 A 13 D 11 C 15 B 13 B 8 A 13 D 9 C 9
;
PROC GLM;
CLASS X1 X2 X3;
MODEL Y=X1 X2 X3;
MEANS X1 X2 X3 /DUNCAN;
RUN;
```

برخی از کلمات کلیدی مورد استفاده در زیر دستور OUTPUT که بیشتر در انالیز رگرسیون کاربرد دارند عبارتند از:

P= -۱

برای ذخیره کردن مقادیر براورد شده

R= -۲

برای ذخیره کردن مقادیر خطاها

COOK D= -۳

معیار تاثیر فاصله کوک

DIFFITS=نام متغیر; -۴

برای ذخیره کردن مقادیر استاندارد DIFFITS براساس خطای استیودنت شده

$$(DIFFITS)_i=T_i\sqrt{\frac{hii}{1-hii}}$$

$$T_i=\frac{ei}{si\sqrt{1-hii}}$$:خطاهای استیودنت شده

$$S_i=\frac{(n-k)^{-1}MSE-\frac{e\hat{i}^2}{1-hii}}{n-k}$$

این براساس فاصله $\hat{y}i$, $\hat{y}-i$ ساخته شده است. اگر $(DIFFITS)_i$ بزرگتر از $2\sqrt{\frac{k}{n}}$ یا $3\sqrt{\frac{k}{n}}$ باشد. انگاه نقطه i ام نقطه موثر است.

نقاط غیر عادی یا دارای LEVERAGE بالا از فرمول زیر بدست می ایند. $H_{ii}=k/n$ که در ان k تعداد پارامتر ها می باشد.اگر $h_{ii}>2k/n$ یا $h_{ii}>3k/n$ باشد انگاه نقطه i ام غیر عادی است.

H=نام متغیر; -۵

برای ذخیره کردن مقادیر هت های h_{ii}

CORRATIO=نام متغیر;-۶

برای ذخیره کردن مقادیر معیار تاثیر مشاهدات براساس کواریانس برآورد پارامتر ها

LCL=نام متغیر;-۷

برای ذخیره کردن حد پایین فاصله اطمینان برای مقادیر برآورد شده فردی

LCLM=نام متغیر;-۸

برای ذخیره کردن حد پایین فاصله اطمینان برای میانگین مقادیر برآورد شده

RSTUDENT=نام متغیر;-۹

برای ذخیره کردن مقادیر خطاهای استیودنت شده

STDI=نام متغیر;-۱۰

برای ذخیره کردن مقدار خطای استاندارد میانگین مقادیر برآورد شده

UCL=نام متغیر;-۱۱

برای ذخیره کردن حد بالا فاصله اطمینان برای مقادیر برآورد شده فردی

UCLM=نام متغیر;-۱۲

برای ذخیره کردن حد پایین فاصله اطمینان برای میانگین مقادیر برآورد شده

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۱۰- رویه NPAR1WAY

از این رویه برای انجام آزمونهای ناپارامتری برای مقایسه میانه دو یا چند جامعه مستقل یا مقایسه شکل توزیع دو جامعه مستقل استفاده می شود.
در واقع با استفاده از این رویه می توان آزمونهای ناپارامتری معادل آزمون استیودنت حالت دو نمونه مستقل همانند آزمون میانه یا من-ویتنی ویلکاکسون را انجام داد.
علاوه بر این از رویه برای انجام آنالیز واریانس ناپارامتری یک طرفه می توان استفاده کرد.
آزمون آنالیز واریانس یکطرفه پارامتری در روشهای ناپارامتری , آزمون کروسکال-والیس است که تعمیم آزمون من-ویتنی-ویلکاکسون است.
نکته ۱: آزمون من-ویتنی دو پذیره مهم دارد.
۱-دو نمونه تصادفی و مستقل از دو جامعه مستقل داشته باشیم.
۲- تحت فرضیه H0 تفاوت دو جمعیت تنها بایستی در پارامتر مکان باشد , یعنی نمی توانند در پارامتر مقیاس تفاوت داشته باشند.
نکته ۲: آزمون من-ویتنی برای میانه ارجحیت دارد. آزمون میانه برای مقایسه میانه دو جامعه مستقل کاربرد دارد.
این رویه علاوه بر آزمونهای فوق قادر است آزمون کلموگروف-اسمیرنوف (KS) را برای مقایسه شکل توزیع (واریانس) دو جامعه انجام دهد.
شکل کلی این رویه بصورت زیر است.

```
PROC NPAR1WAY OPTIONS;
BY VARIABLE(S);
CLASS VARIABLE;
EXACT STATISTIC-OPTION / COMPUTATION-OPTION;
FREQ VARIABLE;
OUTPUT OUT=DATA SET نام KEYWORDS;
VAR VARIABLE(S);
RUN;
```

زیر دستورهای رویه NPAR1WAY
بعضی از زیر دستورات این رویه عبارتند از:

۱- زیر دستور BY
برای مشخص کردن متغیر یا متغیرهای گروهبندی شده که لازم است عملکرد رویه به تفکیک مقادیر انها گزارش شود.

۲- زیر دستور CLASS
برای مشخص کردن یک و تنها یک متغیر گروهبندی شده که مقادیر ان مشخص کننده جوامع مستقل (دو یا چند جامعه) می باشد , کاربرد دارد. این دستور اجباری است.

۳- زیر دستور EXACT
برای معرفی آزمونهای دقیق برای مقایسه پارامترهای جوامع بصورت ناپارامتری می باشد.
نرم افزار معمولا بطور پیش فرض P-VALUE را براساس آزمونهای مجانبی آماره آزمون گزارش می کند, چنانچه مدنظر محقق آزمون دقیق باشد, می تواند از این عبارت استفاده کند.
; عبارات محاسباتی که بصورت محدودیت در نظر گرفته می شوند/ OPTION های لازم برای مشخص کردن آماره هاEXACT
برخی از انتخابهای مورد استفاده در عمل برای قسمت شاخصها یا آماره ها عبارتند از:

1) KS
آزمون دقیق کلموگروف-اسمیرنوف دو نمونه ای
2) MEDIAN
آزمون دقیق میانه
3) ST
آزمون دقیق سیکل توکی برای مقایسه واریانسها
4) WILCOXSON

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

آزمون دقیق ویلکاکسون برای دو نمونه مستقل (من-ویتنی) و (M-W-W) برای مقایسه میانه چند جامعه مستقل کروسکال-والیس
برخی از انتخابهای محاسباتی این عبارت , عبارتند از:

۱- ALPHA= 0<P<1

برای مشخص کردن ضرایب اطمینان برای فواصل اطمینان دقیق ارائه شده در خروجی که α بطور پیش فرض ۰,۰۱ است.

۲- MAX TIME =VALUE

برای مشخص کردن ماکزیمم زمان لازم برای تکرار الگوریتم ها , در محاسبه P-VALUE دقیق آزمونها که بصورت ثانیه در نظر گرفته می شود.

۳- N=n

برای مشخص کردن تعداد نمونه های اولیه لازم در برآورد مونت کارلوی مقادیر P-VALUE در آزمونهای دقیق که بطور پیش فرض n=10000 می باشد.

۴- زیر دستور FREQ

برای معرفی یک و تنها یک متغیر شمارشی شامل فراوانی مشاهدات

۵- زیر دستور OUTPUT

OUTPUT OUT= نام مجموعه داده KEYWORDS;

برخی از کلمات کلیدی مورد استفاده در این عبارت بصورت زیراند.

۱- ANOVA

برای ذخیره کردن شاخصها و آماره ها در جدول انالیز واریانس

۲- EDF

برای ذخیره کردن آماره آزمون نیکویی برازش کلموگروف-اسپیرنوف , کرامر-ون-میسز و آزمون کوپیر برای دو نمونه مستقل

۳- MEDIAN

برای ذخیره کردن نتایج آزمون میانه

۴- WILCOXON

برای ذخیره کردن نتایج آزمون من-ویتنی-ویلکاکسون و آزمون کروسکال-والیس

۶- زیر دستور VAR

برای مشخص کردن لیست متغیرهای وابسته که بایستی آزمون روی آنها انجام شود , در نظر گرفته می شود.

آزمون علامت

حالت خاص آزمون علامت برای دو نمونه زوج شده که در آن متغیر پاسخ دودویی است (۰و۱) آزمون مک نمار نام دارد. که آن را می توان با دستور FREQ در جدولهای توافقی نوع دوم انجام داد.
آزمونهای ناپارامتری برای مقایسه میانه جامعه با عدد ثابت یا میانه دو جامعه وابسته به یکدیگر همانند آزمون علامت و آزمون رتبه علامتدار ویلکاکسون در حالت تک نمونه ای و زوج نمونه تنها توسط دستور UNIVARIATE امکان پذیر است.
آزمونهای ناپارامتری برای مقایسه میانه دو یا چند جامعه مستقل همانند آزمون میانه (MEDIAN) , من-ویتنی-ویلکاکسون (M-W-W) و آزمون کروسکال والیس (K-W) و آزمونهای ناپارامتری اسپیرنوف حالت دو نمونه ای توسط دستور NPAR1WAY امکان پذیر است.
آزمون تک نمونه ای کلموگروف-اسپیرنوف و نیکویی برازش پیرسون و سایر آزمونهای نیکویی برازش همانند آزمون شاپیرو ویلک توسط دستور UNIVARIATE امکان پذیر است.

۱۱- رویه REG

شکل کلی این رویه بصورت زیر است.

```
PROC REG OPTIONS;
BY VARIABLE(S);
FREQ VARIABLE;
ID VARIABLE(S);
VAR VARIABLE(S);
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
WIGHT VARIABLE;
ADD VARIABLE(S);
DELETE VARIABLE(S);
MODEL متغیر های مستقل= متغیر یا متغیر های وابسته / OPTIONS;
OUTPUT OUT=DATA SET نام  KEYWORDS= نام;
PLOT VARIABLE* VARIABLE;
RESTRICT EQUATIONS;
TEST EQUATIONS;
RUN;
```

بعضی از OPTIONهای این رویه عبارتند از:

۱- نام DATA SET ورودی = DATA

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۲-; نام OUT SET = DATA SET

۳-; نام REDGE DATA SET

برای انجام رگرسیون ریدج و ذخیره کردن ان در DATA SET خروجی

۴-; نام TABLE OUT= DATA SET

برای فرستادن مقادیر خطاهای استاندارد شده , فواصل اطمینان و سایر مقادیر به مجموعه داده خروجی

۵- CORR

ماتریس ضرایب همبستگی متغیر های لیست شده در VAR, MODEL

۶- ALL

برای مشخص کردن تمامی شاخصها و آماره ها برای متغیر های موجود در MODEL , VAR

۷- ALPHA= 0<P<1

برای مشخص کردن ضریب اطمینان حدود اطمینان

زیر دستورهای رویه REG

زیر دستورات این رویه عبارتند از:

۱- زیر دستور ID

ID VARIABLE(S);

برای اینکه وقتی CLI , CLM , P , R و غیره از OPTIONهای مدل گزارش شوند.
هر کدام از این مقادیر در کنار متغیر های لیست شده در ID گزارش می شوند.

۲- زیر دستور MODEL

MODEL متغیر های مستقل= متغیر یا متغیر های وابسته / OPTIONS;

برخی از OPTIONهای این زیر دستور عبارتند از:

1-SELECTION=FORWARD (or F) OR BACKWARD (or B) OR STEPWISE OR MAXR OR MINR OR RSQUARE OR ADJRSQ OR CP OR NONE

برای انتخاب یکی از روشهای انجام رگرسیون مورد استفاده قرار می گیرد.
توجه کنید که NONE برای مدل کامل است.

2- NOINT

برای انجام مدل رگرسیون بدون عرض از مبدا

3-RIDGE=K

برای مشخص کردن انجام رگرسیون ریدج

4-ADJRSQ

برای محاسبه ضریب تعیین تصحیح شده

5-B

برای ارائه جدول برآورد پارامترهای مدل در هر STEP

6-I

برای نشان دادن معکوس ماتریس مجموع مربعات و حاصلضربها (همانند GLM)

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

7-XPX

برای گزارش ماتریس مجموع مربعات و حاصلضربها

8-COLLIN

برای انجام انالیز همخطی

9-CORRB

برای گزارش ماتریس ضرایب همبستگی برآورد پارامترها

10-COVB

برای گزارش ماتریس واریانس کواریانس برآورد پارامترها

11-STB

برای گزارش عامل بزرگی واریانس برای متغیرهای توضیحی

13-CLB

برای گزارش فاصله اطمینان برای برآورد پارامترها

14-CLI

برای گزارش فاصله اطمینان برای مقادیر برآورد شده فردی

15-CLM

برای گزارش فاصله اطمینان برای مقادیر مورد انتظار متغیر وابسته

16-DV

محاسبه مقدار آماره دربین-وتسون

17-DWPROB

برای محاسبه مقدار P-VALUE برای آماره دربین-واتسون

18-P

محاسبه مقادیر برآورد شده

19-PARTIAL

برای گزارش نمودارهای خط رگرسیون برای هر مدل

20-R

برای محاسبه انالیزهای مربوط به مقادیر خطاها

21-ALL

برای محاسبه تمامی شاخصهای لازم و ضروری

22-ALPHA= 0<P<1

برای مشخص کردن ضریب اطمینان فواصل اطمینان

۳- زیر دستور OUTPUT

OUTPUT OUT=DATA SET نام KEYWORDS= نام;

برخی از کلمات کلیدی این زیر دستور عبارتند از:

۱- نام متغیر =COOKD

برای محاسبه مقادیر تاثیر فاصله کوک

۲- نام متغیر =COVRATIO

برای محاسبه معیار تاثیر استاندارد مشاهدات براساس کواریانس بتاها

۳- نام متغیر =DIFFITS

برای محاسبه معیار تاثیر مشاهدات براساس مقادیر برآورد شده یا معیار تاثیر DIFFITS

۴- نام متغیر =H

برای محاسبه مقادیر هت ها

۵- نام متغیر =LCL

برای ذخیره باند پایین فاصله اطمینان برای مقادیر برآورد شده فردی

۶- نام متغیر =LCLM

برای ذخیره باند پایین فاصله اطمینان برای میانگین متغیر وابسته

۷- نام متغیر =P OR PREDICTED

برای ذخیره مقادیر برآورد شده

۸- نام متغیر =R OR RESIDUAL

برای ذخیره مقادیر خطاها

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۹- نام متغیر =STD I
برای ذخیره انحراف معیار مقادیر برآورد شده فردی
۱۰- نام متغیر =STD P
برای ذخیره انحراف معیار میانگین مقادیر برآورد شده
۱۱- نام متغیر =STD R
برای ذخیره انحراف معیار مقادیر خطاها
۱۲- نام متغیر =STUDENT
برای ذخیره انحراف معیار مقادیر خطاهای استیودنت شده
۱۳- نام متغیر =UCL
برای ذخیره باند بالا فاصله اطمینان برای مقادیر برآورد شده فردی
۱۴- نام متغیر =UCLM
برای ذخیره باند بالا فاصله اطمینان برای میانگین متغیر وابسته

۴- زیر دستور PLOT

PLOT VARIABLE* VARIABLE;

برای رسم نمودارهای پراکنش
مثال)

PLOT (Y1 Y2)*(X1 X2) ≡ PLOT Y1*X1 Y1*X2 Y2*X1 Y2*X2;

۵- زیر دستور RESTRICT

RESTRICT EQUATIONS,…, EQUATIONS;

برای گذاشتن برخی محدودیتها روی پارامترهای مدل رگرسیون و آزمون ان فرضیه ها
در واقع برای آزمون ترکیبات خطی روی مقادیر پارامترهای مدل
مثال)

RESTRICT X1;	$H_0:\beta_1=0$
RESTRICT X1+X2=1;	$H_0: \beta_1+\beta_2=1$
RESTRICT X1=X2=X3;	$H_0: \beta_1=\beta_2=\beta_3$
RESTRICT X1=X2 , X2=X3;	$H_0: \beta_1=\beta_2 , \beta_2=\beta_3$
RESTRICT 2*X1=X2+X3 , INTERCEPT+X2=0;	$H_0: 2\beta_1=\beta_2+\beta_3 , \beta_0+\beta_1$
RESTRICT X1=X2=X3=INTERCEPT	$H_0: \beta_0=\beta_1=\beta_2=\beta_3$

۶- زیر دستور TEST

TEST EQUATIONS/OPTION;

برای انجام یک فرضیه روی پارامترهای مدل رگرسیون (همانند RESTRICT)
در اینجا TEST دارای یک OPTION برخلاف RESTRICT می باشد.

H0:L β=C

مثال)

MODEL Y=A1 A2 B1 B2;
TEST A1+A2=1;
TEST B1=0 , B2=0; ≡ TEST B1,B2;

تنها OPTION این زیر دستور PRINT می باشد که برای گزارش محاسبات مربوط به این آزمون مورد استفاده قرار می گیرد.

۱۲- رویه CATMOD

شکل کلی این رویه بصورت زیر است.

PROC CATMOD OPTIONS;
DIRECT VARIABLE(S);
MODEL RESPONSE_EFFECT=DESIGN_EFFECTS /OPTION;

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
CONTRAST 'LABEL' REQUST /OPTION;
BY VARIABLE(S);
FACTOR factor,…,factor /OPTION;
LOGIN EFFECTS;
RESPONSE function,…,function /option;
RESTRICT parameter=value,…,parameter=value;
WIGHT VARIABLE;
RUN;
```

برخی از optionهای این رویه عبارتند از:

۱- نام DATA=DATA SET

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

۲- NAMELEN=n $24 \leq n \leq 200$

برای مشخص کردن طول نام اثر یا اثرات که بطور پیش فرض ۲٤ کاراکتر می باشد.

۳- (براساس ترتیب) REQ یا (براساس فرمت متغیرها) FORMATED یا (براساس داده ها) ORDER=DATA

زیر دستورهای رویه CATMOD

۱- زیر دستور DIRECT

```
DIRECT VARIABLE(S);
```

برای مشخص کردن متغیرهای عددی (کمی)

که با این زیر دستور نرم افزار با انها همانند متغیرهای عددی (کمی) بجای متغیرهای کیفی برخورد می کند.

و تمام سطوح انها را در ماتریس طرح در نظر می گیرد. برخلاف متغیرهای کیفی که به تعداد سطوح منهای یک واحد برابر ماتریس طرح کفایت می کند.

۲- زیر دستور MODEL

MODEL متغیرهای وابسته= متغیرهای مستقل / OPTIONS;

برای مشخص کردن مدل مورد نظر استفاده می شود.

مثال:

```
MODEL R=A B;
```

فقط اثرات اصلی

```
MODEL R=A B A*B;
```

اثرات اصلی و متقابل

```
MODEL R=A B(A);
```

اثرات لانه ای (تو در تو)

```
MODEL R=A|B;
```

مدل کامل

```
MODEL R=A B(A=1) B(A=2);
```

طرح تو در تو با مقدار اثرات

```
MODEL R*S=_RESPONSE_;
```

مدل لگ خطی

```
MODEL R*S=A _RESPONSE_(A);
```

طرح عامل با اندازه های مکرر تو در تو

```
MODEL _F_=_RESPONSE_;
```

ورودی مستقیم توابع پاسخ

برخی از انتخابهای این زیر دستور عبارتند از:

1- ML=NR OR IPF

برای مشخص کردن روش برآورد درستنمائی ماکزیمم توسط روشهای نیومن-رافن (NR) یا روش INTERATIVE PROPORTIONAL FITTING در رگرسیون لجستیک و مدلهای لگ خطی از NR استفاده می شود.

2- WLS

برای محاسبه برآوردهای کمترین مربعات وزنی

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

3-NOINT

برای حذف عرض از مبدا از مدل

4-PREDICT=FREQ OR PROB

برای محاسبه نمایش مقادیر برآورد شده و مشاهده شده همراه با مقادیر خطای استاندارد انها و مقادیر خطاها اگر FREQ انتخاب شود در هر خانه جدول فراوانی های برآورد شده گزارش می شود ولی اگر PROB انتخاب شود در خانه های جدول مقادیر احتمال برآورد شده گزارش می شوند.

5-EPSILON

برای مشخص کردن شرط همگرایی برآورد درستنمائی ماکزیمم پارامترها
عدد پیش فرض 10^{-8} می باشد.

6-MAXITER= عدد

برای مشخص کردن ماکزیمم تعداد تکرارهای روشهای برآورد
بطور پیش فرض MAXITER=20 است.

7-ALPHA= 0<P<1

برای مشخص کردن ضریب اطمینان فواصل اطمینان

8-CLPARM

برای محاسبه حدود اطمینان والد برای برآورد پارامترها

9-CORRB

برای محاسبه ماتریس ضرایب همبستگی پارامترها

10-COV

برای محاسبه ماتریس واریانس کواریانس پاسخ

11-COVB

برای محاسبه ماتریس واریانس کواریانس برآورد پارامترها

12-DSEIGN

برای گزارش ماتریس طرح یا ماتریس پاسخ (در مدلهای لگ خطی)

13-ITPRINT

برای ارائه گزارش و برآورد پارامترها در هر تکرار مدل

14-NODESIGN

برای حذف نمایش ماتریس طرح

15-NOPARM

برای حذف نمایش برآورد پارامترها

16-NORESPONSE

برای حذف نتایج مربوط به گزارش ماتریس پاسخ

۳- زیر دستور CONTRAST

CONTRAST 'LABEL' REQUST /OPTION;

برای مشخص کردن مقابله برای پارامترها
مثال)

H0:α1+α2=2α3

```
CONTRAST '1 Q 2 US 3' A 1 1 -2;
CONTRAST 'JOINT EFFECT OF A&B' A  1  0  0;
                               A  0  1  0;
                               A  0  0  1;
                               B  1  0;
                               B  0  1;
```

۴- زیر دستور FACTORS

FACTORS عامل۱ نوع ان تعداد سطوح .... عاملn نوع ان تعداد سطوح, ...;

از این زیر دستور برای مشخص کردن اثرات(عوامل) استفاده می شود که چنانچه عامل رشته ای (STRING) باشد بایستی نوع ان را بعد از ان با $ مشخص کرد. در صورت عددی بودن مقادیر نیازی به ذکر نوع نیست. تعداد سطوح مورد نظر در تجزیه و تحلیل برای هر عامل نیز حداقل با یک فاصله بعد از ان قید می شود.

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

۵- زیر دستور LOGIN

LOGIN اثرات ;

برای مشخص کردن اثرات در مدل لگ خطی بصورت a b c یا a|b|c

۶- زیر دستور RESPONSE

RESPONSE FUNCTION;

برای مشخص کردن نوع تابعی که برای متغیر پاسخ در نظر می گیریم که یکی از توابع زیر است.

1-ALOGIT(S): تابع لجیت مجاور
2-CLOGIT(S) : تابع لجیت تجمعی
3-JOINT: تابع توام (احتمال توام)
4-LOGIT(S) تابع لجیت
5-MARGINAL(S): احتمالات حاشیه ای
6-MEAN(S) میانگین پاسخ

۷- زیر دستور RESTRICT

RESTRICT B=(شماره) = عدد;

برای ایجاد محدودیت روی پارامترها که هر یک از پارامترها را با B و شماره مورد نظرشان نشان می دهند.
بیش از یک پارامتر نیز می توان معرفی کرد.
مثال)

RESTRICT B1=1 B4=B6=0;

۱۳- رویه CHART

شکل کلی این رویه بصورت زیر است.

```
PROC CHART OPTIONS;
BLOCK VARIABLE(S);
BY VARIABLE(S);
HBAR VARIABLE(S) /OPTION;
PIE VARIABLE(S) /OPTION;
STAR VARIABLE(S) /OPTION;
VBAR VARIABLE(S) /OPTION;
RUN;
```

یکی از OPTIONهای این رویه عبارت است از:

نام DATA SET ورودی = DATA

برای مشخص کردن نام DATA SET ورودی در نظر گرفته می شود.

زیر دستورهای رویه PLOT

۱- زیر دستور BLOCK

برای مشخص کردن یک نمودار برای هر یک از متغیرهای لیست شده در BLOCK

۲- زیر دستور BY

برای رسم نمودار به تفکیک سطوح متغیرهای لیست شده در این زیر دستور

۳-زیر دستور HBAR

برای مشخص کردن نمودار ستونی – افقی
تنها انتخاب این زیر دستور عبارت است از:

TYPE=FREQ یا CFREQ یا PCT یا CPRCT;

۴- زیر دستور PIE

برای رسم نمودار دایره ای
تنها انتخاب این زیر دستور عبارت است از:

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

TYPE=FREQ یا CFREQ یا PCT یا CPRCT;

۵- زیر دستور STAR

برای رسم نمودار ستاره ای
تنها انتخاب این زیر دستور عبارت است از:

TYPE=FREQ یا CFREQ یا PCT یا CPRCT;

۶- زیر دستور VBAR

برای رسم نمودار ستونی-عمودی

فصل سوم : تمرینات

تمرین۱- جدول زیر میزان رضایت شغلی فردی ۷۱۵ کارگر منتخب از کارگران صنایع را در مقابل رضایت شغلی سرپرستان انها که توسط دو پرسشنامه مشابه حاصل شده است نشان می دهد , علاوه بر این تحقیق یک ارزیابی از خارج کارخانه در مورد کیفیت مدیریت صنایع انجام شده است, که میزان کیفیت را در دو سطح خوب و بد خلاصه می کند.
موارد زیر را با استفاده از نرم افزار SAS انجام داده و تفسیر نمائید.
الف) ایا رضایت شغلی کارگران ربطی به رضایت شغلی سرپرستان انها دارد؟
ب) ایا کیفیت مدیریت تاثیری روی رضایت شغلی فردی کارگران دارد؟
ج) ایا کیفیت مدیریت خوب باعث افزایش رضایت شغلی سرپرستان می شود؟
د) در بین مدیران با کیفیت خوب ایا رضایت شغلی سرپرستان تاثیری روی رضایت شغلی کارگران دارد؟

C کیفیت مدیریت	B رضایت شغلی سرپرستان	A رضایت شغلی کارگران	
		پایین	بالا
بد	پایین	۱۰۳	۸۷
	بالا	۳۲	۴۲
خوب	پایین	۵۹	۱۰۹
	بالا	۷۸	۲۰۵

```
DATA;
DO C=1,2; B=1,2;A=1,2;
INPUT @@;
OUTPUT;
END;END;END;
CARDS;
103 87 32 42 59 109 78 205
;
PROC FREQ;
TABLES A*B C*A C*B C*B*A / CHISQ;
EXACT FISHER;
RUN;
```

تمرین۲- یک مطالعه اماری در دو منطقه تهران صورت گرفته است تا نسبت خانه های مسکونی دارای سیستم حرارت مرکزی را با یکدیگر مقایسه کنند.
یک نمونه تصادفی ۳۰۰تائی از خانه های مسکونی منطقه اول نشان داده است که ۱۸۵ مورد دارای سیستم حرارت مرکزی هستند, در حالیکه یک نمونه تصادفی ۲۰۰تائی از خانه های مسکونی منطقه دوم ۷۵ مورد را دارای سیستم حرارت مرکزی ارزیابی می کند.
ایا اختلافی بین نسبت خانه های که دارای سیستم حرارت مرکزی هستند در دو منطقه وجود دارد؟

```
DATA;
DO X1=1,2; DO X2=1,2;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
INPUT COUNT@@;
OUTPUT;
END;END;
CARDS;
185 115 75 125
;
PROC FREQ;
TABLE X1*X2;
EXACT FISHER;
WEIGHT COUNT;
RUN;
```

تمرین۳- از ۲۰۰نفر که بطور تصادفی انتخاب شده اند , دو سئوال زیر پرسیده شده است.
۱- ایا به نظر شما برای کاهش بار تورم کارمندان دولت باید حقوق انان افزایش یابد؟
۲- ایا شما کارمند هستید؟
در سطح خطای ۰,۰۱ درصد ایا نظر افراد در مورد سئوال ۱ ربطی به کارمند بودن یا نبودن انها دارد؟

	سئوال ۲	
سئوال ۱	بلی	خیر
بلی	۵۰	۶۰
خیر	۱۰	۸۰

```
DATA;
DO A=1,2; DO B=1,2;
INPUT COUNT@@;
OUTPUT;
END;END;
CARDS;
50 60 10 80
;
PROC FREQ;
TABLE A*B / ALPHA=0.01;
EXACT FISHER;
WEIGHT COUNT;
RUN;
```

تمرین۴- فروشگاهی سه تا مارک متفاوت چای را در بسته های یک کیلویی برای فروش عرضه می کند , در نمونه تصادفی در یک روز تعداد بسته های چای خریداری شده با مارکهای مختلف در جدول زیر امده است ایا می توان ادعا کرد که فروش چای در رابطه با مارک ان دارای توزیع یکنواخت است؟

مارک	الف	ب	ج
تعداد بسته های فروش رفته	۱۷۵	۲۳۵	۱۹۰

```
DATA;
DO A=1,2,3;
INPUT COUNT@@;
OUTPUT;
END;
CARDS;
190 235 175
;
PROC FREQ;
TABLE A / CHISQ;
WEIGHT COUNT;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

RUN;

تمرین۵- سه نوع کود الف , ب , و ج و چهار مزرعه مختلف مورد آزمایش قرار گرفته است.
نتایج بر حسب تن به شرح زیر است.

کود	۱	۲	۳	۴
الف	۵۵	۶۵	۶۴	۶۰
ب	۶۶	۷۰	۷۵	۶۹
ج	۶۸	۷۸	۷۴	۷۲

در سطح خطای ۰,۰۵مشخص نمائید که:
الف) ایا کودها روی میزان محصول موثرند. در ان صورت از نظر بیشترین میزان محصول کدام نوع کود بر دیگر کودها برتری دارد؟
ب) ایا حاصلخیزی ۴ مزرعه یکسان است؟ کدام مزرعه نسبت به دیگر مزرعه ها حاصلخیزتر است؟

تمرین۶- مدیر یک فروشگاه زنجیره ای که از ۳ فروشگاه مشابه در ۳ نقطه شهر تشکیل شده است مایل است اثر اعمال ۳ نوع سیاست تبلیغاتی را بر روی فروش , فروشگاههای خود آزمایش کند.
او برای هر فروشگاه سیاست تبلیغاتی ویژه ای را به مدت ۳ ماه دنبال کرده است. کل فروش برای هر ماه در جدول زیر امده است.

روش تبلیغ	مهر	ابان	اذر
فروشگاه ۱- بدون تبلیغ	۲۰	۲۴	۲۸
فروشگاه ۲- پخش اعلامیه	۲۴	۲۶	۲۹
فروشگاه ۳-پخش اعلامیه و اگهی روزنامه	۲۵	۲۹	۳۰

با توجه به اطلاعات فوق:
الف) ایا میزان فروش ۳ فروشگاه یکسان است؟
ب) ایا میزان فروش , فروشگاه زنجیره ای در ۳ ماه یکسان است؟
ج) ایا پخش اعلامیه باعث افزایش فروش فروشگاه در طی ۳ ماه می شود؟
د) ایا علاوه بر پخش اعلامیه , اگهی در روزنامه ها نیز باعث افزایش فروش فروشگاه می شود؟

تمرین۷- برای بررسی میزان افت فشار در شیرهای انبساط توربین ها سه عامل دمای گاز , سرعت توربین و فشار گاز را در یک طرح عامل بررسی می کنیم.
چنانچه همه عوامل تثبیت شده باشند. عوامل موثر را مشخص کنید؟ (خطای نوع اول=0.05)
اگر کمترین میزان افت فشار مطلوب باشد بهترین سطح دمای گاز , سرعت توربین و فشار گاز کدامست؟

		دمای گاز											
		60 F				70 F				90 F			
سرعت توربین		150	200	225	300	150	200	225	300	150	200	225	300
فشار گاز	50	-2	0	-1	4	14	6	1	-7	-8	-2	-1	-2
		-3	-9	-8	4	14	0	2	6	-8	20	-2	-1
	75	-6	-5	-8	-3	22	8	6	-5	-8	1	-9	-8
		4	-1	-2	-7	24	6	2	2	3	-7	-8	3
	85	-1	-4	0	-2	20	2	3	-5	-2	-1	-4	3
		-2	-8	-7	4	16	0	0	-1	-1	-2	-7	1

تمرین۸- در تمرین قبل فرض کنید دمای گاز تثبیت شده و سرعت توربین و فشار گاز تصادفی شده باشند. در این صورت اثرات موثر را مشخص نموده و تفاوت بین سطوح هر یک از عوامل را مشخص نمائید؟

```
DATA;
DO A=60,75,90;
DO B=150,200,225,300;
DO C=50,75,85;
DO I=1,2;
INPUT Y @@;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
OUTPUT;
END;END;END;END;
DATA LINES;
-2 -3 -6 4 -1 -2 0 -9 -5 -1 -4 -8 -1 -8 -8 -2 0 -7 4 4 -3 -7 -2 4 14 14 22 24 20 16 6 0 8 6
2 0 1 2 6 2 3 0 -7 6 -5 2 -5 -1 -8 -8 -8 3 -2 -1 -2 20 1 -7 -1 -2 -1 -2 -9 -8 -4 -7 -2 -1 -8 3
3 1
;
PROC GLM;
CLASS A,B,C;
MODEL Y=A|B|C;
CONTRAST 'H0=M1+M2=2M3 MOGHABLE1' A 1 1 -2;
CONTRAST 'H0=2M1+M2+3M3=6M4 MOGHABLE2' B 2 1 3 -6;
MEANS A B C /TEST;
RUN;
```

تمرین۹- برای مقایسه برتری یکی از دو روش حفظ کردن مطالب آزمایشی انجام شده است. برای این منظور نه زوج از دانش آموزان در آزمایش شرکت داده می شوند.دانش آموزان بر طبق ضریب هوشی و زمینه علمی زوج می شوند. و آنگاه به تصادف یکی از دو روش را برای یک عضو و روش دیگر را برای زوج دوم بکار می بریم.سپس یک آزمون حافظ به تمام دانش آموزان داده شده , نمرات دانش آموزان به شرح زیر است.

زوجها	۱	۲	۳	۴	۵	۶	۷	۸	۹
روش A	۹۰	۸۶	۷۲	۶۵	۴۴	۵۲	۴۶	۳۸	۴۳
روش B	۸۵	۸۷	۷۰	۶۲	۴۴	۵۳	۴۲	۳۵	۴۶

آیا اختلاف معنی داری در کارایی دو روش وجود دارد؟

```
DATA;
INPUT Y1 Y2@@;
CARDS;
90 85 86 87 72 70 65 62 44 44 52 53 46 42 38 35 43 46
;
PROC UNIVARIATE NORMAL;
QQPLOT Y1 Y2/NORMAL;
VAR Y;
PROC TTEST;
PAIRED Y1*Y2;
PROC TTEST;
VAR Y;
RUN;
```

تمرین۱۰- تامسون در حالی که مشغول تحقیق درباره طبیعت پایه ای اشعه کاتدی بود الکترون را کشف کرد.تامسون در تجارب آزمایشگاهی ذرات باردار شده با بار منفی را جدا کرده تا بتواند نسبت جرم به بار را تعیین نماید . به نظر می رسد که این نسبت برای انواع وسیعی از شرایط آزمایشگاهی مقداری ثابت است. تامسون نتایج زیر را با دو لامپ اشعه کاتدی مختلف با بکاربردن هوا بعنوان گاز بدست آورد.

آیا نسبت جرم به بار اندازه گیری توسط دو لامپ تفاوت معنی داری دارد؟

لامپ ۱	۰,۵۷	۰,۳۴	۰,۴۳	۰,۳۲	۰,۴۸	۰,۴۰	۰,۴۰
لامپ ۲	۰,۵۳	۰,۴۷	۰,۴۷	۰,۵۱	۰,۶۳	۰,۶۱	۰,۴۸

```
DATA;
DO X=1,2;
INPUT Y@@;
Y=Y/100;
OUTPUT;
END;
```


استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد

```
CARDS;
57 53 34 47 43 47 32 51 48 63 40 61 40 48
;
PROC UNIVARIATE NORMAL;
VAR Y;
PROC TTEST;
CLASS X;
VAR Y;
RUN;
```

http://statistics.mihanblog.com

استفاده از مطالب این جزوه بدون ذکر نام وبلاگ و نام نویسنده آن ممنوع می باشد.
با تشکر فراوان از جناب آقای دکتر هادی جباری نوقابی عضو هیئت علمی آمار دانشگاه فردوسی مشهد