

## آموزش نرم افزار SAS

### تهیه و تنظیم: فاطمه آقابابایی

نرم افزار SAS از جمله تخصصی ترین نرم افزارهای آماری است. سرعت پردازش اطلاعات به نحو قابل توجهی در این نرم افزار نسبت به سایر نرم افزارهای آماری بالاست. بنابراین در کارهای تخصصی آماری که شامل اطلاعات پیچیده می باشند، بهتر است از این نرم افزار استفاده شود. نگارش های متعددی از این نرم افزار در دسترس می باشد. اولین نگارش موجود این نرم افزار نگارش تحت سیستم عامل DOS می باشد بعد از آن با ایجاد سیستم عامل های ویندوز و ... این نرم افزار قابلیت نصب تحت سیستم های عامل را در نگارش های بعدی یافت. اولین نگارش تحت ویندوز این نرم افزار نگارش 6/12 می باشد.

نگارش های بعدی تحت سیستم عامل ویندوز عبارتند از نگارش 7.0، 8.0، 8.2 و 9.1 این نرم افزار از نگارش اول به صورت، دارای سه پنجره مهم طراحی شده است که عبارتند از:

**الف:** پنجره Program Editor برای برنامه نویسی

**ب:** پنجره output مختص خروجی برنامه

**ج:** پنجره LOG برای نمایش پیغام های ویژه اجرای دستورات محاسبات آماری

**- برنامه نویسی با استفاده از نرم افزار SAS:**

برنامه نویسی با استفاده از نرم افزار SAS با یک سری عبارات اختیاری و اجباری امکان پذیر است. شیوه برنامه نویسی در محیط این نرم افزار، تقریباً شبیه برنامه نویسی با زبان C می باشد به این معنی که تمام 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 شود و سپس دستورات لازم برای تجزیه و تحلیل اطلاعات لیست شود.

**Data step:** به مرحله ورود اطلاعات و معرفی آنها به نرم افزار را Data step گویند.

مرحله دوم برای تجزیه و تحلیل اطلاعات توسط قسمت proc step امکان پذیر است.

**- شیوه معرفی داده ها با استفاده از صفحه کلید:**

شکلی کلی یک Data step برای معرفی داده ها با استفاده از صفحه کلید به صورت زیر است:

Data statment;

input statment;

Cards;

داده‌ها

;

1) بنابراین یک Data step با عبارت اجباری Data شروع شده و با عبارت Run خاتمه می‌یابد.

2) عبارت اجباری input برای مشخص نمودن لیست نام متغیرهای لازم در مورد داده‌هاست مثلاً

input x y ;

نام متغیرها را باید با حداقل یک فاصله از هم جدا کنیم و هر متغیر تنها می‌تواند، حداکثر 8 کاراکتر بگیرد.

نرم‌افزار SAS قادر است اطلاعات را به صورت متغیرهای عددی (Numeric) یا متغیرهای رشته‌ای (string) دریافت کند.

چنانچه متغیر یا متغیرهایی عددی باشند نام آن‌ها را حداکثر با 8 کاراکتر در جلوی عبارت input

یادداشت می‌کنیم اما اگر برخی از متغیرها از نوع رشته‌ای باشند، بایستی حتماً بعد از نام آن‌ها

علامت \$ را حداقل با یک فاصله استفاده نمود.

3) Cards: مرحله ورود داده‌ها فرا رسیده و بعد از Cards; هر چه بیاید به عنوان داده محسوب

می‌شود به جای Cards می‌توان Datalines نوشت.

4) Run; پایان داده‌هاست.

**:Proc step**

تجزیه و تحلیل اطلاعات در نرم افزار SAS با رویه‌های تعریف شده برای آن امکان پذیر است. هر رویه با توجه به عملکرد آن با یک نام، مشخص شده است. ممکن است بین رویه‌های مختلف نرم افزار، اشتراك وجود داشته باشد، ولي هر يك از رویه‌ها، برای يك تجزیه و تحلیل خاص در نظر گرفته شده‌اند. در يك رویه چند عبارت یا زیردستور کار تجزیه و تحلیل را توسعه می‌دهند برخی از این عبارات اختیاری و برخی دیگر اجباری هستند.

ترتیب نوشتن زیر دستورها، در برخی رویه‌ها مهم می‌باشد و بایستی به آن توجه داشت. برخی عبارات در يك رویه می‌توانند تکرار شوند و نیز برخی زیر دستورها نیز طوری ساخته شده‌اند که ممکن است استفاده همزمان از آنها در proc step جایز نباشد.

مرحله استفاده از رویه‌ها: چند زیر دستور عملکرد مشابهی در اکثر رویه‌های نرم افزار SAS دارند که قبل از معرفی انواع رویه‌های لازم و کاربرد آنها به معرفی آنها می‌پردازیم. **(1) عبارت اختیاری Var:** از این عبارت برای لیست کردن متغیرهای لازم که عملکرد رویه، روی آنها صورت خواهد گرفت استفاده می‌شود.

;

نام متغیر یا متغیرها Var

حداقل با يك فاصله باید نام متغیرها را از هم جدا کرد.

**(2) عبارت By:** برای منحصركردن عملکرد يك رویه به زیر گروه‌های يك یا چند متغیر از عبارت By استفاده می‌شود. در هنگام استفاده از این زیردستور بایستی توجه داشت که مجموعه داده‌ها بایستی قبلاً بر حسب مقادیر متغیر یا متغیرهای لیست شده در عبارت By مرتب شده باشند.

**- آمار توصیفی:****بدست آوردن خلاصه آماری:**

یکی از کارهایی که برای توصیف داده‌ها استفاده شود پیدا کردن خلاصه‌های آماری برای متغیرهاست. خلاصه‌های آماری شامل: میانگین، واریانس، انحراف معیار، ضریب تغییرات، ضریب چولگی، کشیدگی و ... می‌باشند.

در sas از رویه‌های capability, univariate, means استفاده می‌شود.

**رویه means:**

شکل کلی این دستور به صورت زیر است:

Options; نام مجموعه داده ها = data proc means

; نام متغیر یا متغیرها var

انتخاب‌هایی (options) که با این دستور به کار می‌روند عبارتند از:

**1-N:** تعداد مشاهداتی که محاسبات بر مبنای آن‌ها انجام شده است را چاپ می‌کند.

**2-mean:** میانگین متغیرهای تعریف شده در جلوی Var را چاپ می‌کند.

**3-Var:** واریانس را می‌دهد.

**4-Std:** انحراف معیار را چاپ می‌کند.

**5-Stderr:** خطای استاندارد (std/ $\sqrt{n}$ )

**6-Min:** مینیمم را چاپ می‌کند.

**7-Max:** ماکسیمم را چاپ می‌کند.

**8-CV:** ضریب تغییرات را به ما می‌دهد.

**9-Skewness:** ضریب چولگی را به ما می‌دهد.

**10-Kurtosis:** ضریب کشیدگی را به ما می‌دهد.

**11-T:** آماره t برای آزمون فرضی صفر بودن میانگین را چاپ می‌کند.

**12-PRT:** مقدار احتمال برای آزمون (p-value)

**13-maxdec=n:** تعداد ارقام اعشاری که می‌خواهیم در محاسبات مبذول شود.

**14-Range:** دامنه (Max-Min) را چاپ می‌کند.

$$\sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{مقدار: CSS-15}$$

$$\sum_{i=1}^n X_i^2 \quad \text{مقدار: USS-16}$$

### - رویه univariate

یکی دیگر از دستورالعمل‌هایی که برای محاسبه خلاصه آماری متغیرها به کار می‌رود.

این دستور مشابه دستور العمل means عمل می‌کند.

- شکل کلی این رویه به صورت زیر می‌باشد:

Proc univariate data= نام مجموعه داده‌ها options;

Var نام متغیر یا متغیرها

histogram نام توزیع / نام متغیر یا متغیرها

probplot نام توزیع / نام متغیر یا متغیرها

qqplot نام توزیع / نام متغیر یا متغیرها

این دستور مشکلات دستور means را ندارد و فقط با نوشتن نام متغیر یا متغیرها در جلوی زیر

دستور var کلیه خلاصه‌های آماری را چاپ می‌کنند.

انتخاب‌های مورد استفاده در دستور Univariate:

**1-plot:** نمودار جعبه‌ای و نمودار احتمال نرمال و نمودار ساقه و برگ را به ما می‌دهد.

**2-normal:** تست نرمال بودن داده‌ها را انجام می‌دهد.

**3-MU0=C:** آزمون تساوی میانگین با عدد C را انجام دهد.

- زیر دستور **probplot:** نمودار احتمال نرمال رسم می‌کند.

Qqplot: برای رسم چندک‌های نمونه در برابر چندک‌های توزیع

### - رویه Capability:

شکل کلی این رویه:

Proc capability options;

Var نام متغیر یا متغیرها

histogram نام توزیع / نام متغیر یا متغیرها



برای رسم چندک ها:

qqplot نام متغیر یا متغیرها

ppplot نام متغیر یا متغیرها

probplot نام متغیر یا متغیرها

برای رسم تابع توزیعی تجمعی:

cdfplot نام متغیر یا متغیرها

- تجزیه و تحلیل همبستگی:

در بسیاری از تجزیه و تحلیل‌های آماری یکی از مسائل مهم بررسی همبستگی متغیرها می‌باشد. در sas برای تجزیه و تحلیل همبستگی از رویه Corr استفاده می‌شود.

- شکل کلی proc corr:

Proc corr options;

Var نام متغیرها;

انتخاب‌ها (options):

1- pearson: آنالیز همبستگی پیرسون را انجام می‌دهد.

2- Spearman: آنالیز همبستگی اسپیرمن را انجام می‌دهد.

3- Kendall: آنالیز همبستگی کندال را انجام می‌دهد.

4- Cov: ماتریس کوواریانس را چاپ می‌کند.

$$COV = \frac{CSSCP}{n-1} = \frac{1}{n-1} \begin{bmatrix} \sum(X - \bar{X})^2 & \sum(X - \bar{X})(Y - \bar{Y}) \\ \sum(X - \bar{X})(Y - \bar{Y}) & \sum(Y - \bar{Y})^2 \end{bmatrix}$$

5- Nosimple: از چاپ خلاصه‌های آماری ممانعت می‌کند.

6- Noprob: از چاپ مقادیر احتمال p-value جلوگیری می‌کند.

7- SSCP: محاسبه ی مجموع مربعات تصحیح نشده و مجموع حاصلضرب ها

که sscp به شرح زیر است:

$$SSCP = \begin{bmatrix} n & \sum X & \sum Y \\ \sum X & \sum X^2 & \sum XY \\ \sum Y & \sum XY & \sum Y^2 \end{bmatrix}$$

**8-CSSCP:** محاسبه مجموع مربعات تصحیح شده و مجموع حاصلضربها

که CSSCP به شرح زیر است:

$$\text{CSSCP} = \begin{bmatrix} \sum (X - \bar{X})^2 & \sum (X - \bar{X})(Y - \bar{Y}) \\ \sum (X - \bar{X})(Y - \bar{Y}) & \sum (Y - \bar{Y})^2 \end{bmatrix}$$

**9-nomissing:** برای حذف مشاهدات گم شده از تجزیه و تحلیلها (در صورتی این انتخاب را به کار نبریم تعداد مشاهدات گم شده را نیز در نظر می گیرند).

**-ضریب هم بستگی اسپیرمن:**

ضریب همبستگی اسپیرمن یک معیار ناپارامتری برای سنجش ارتباط بین دو متغیری باشد که بر پایه رتبه ی مقادیر استوار می باشد و فرمول آن به شرح زیر می باشد:

$$\theta = \frac{\sum_i ((R_i - \bar{R})(S_i - \bar{S}))}{\sqrt{\sum_i (R_i - \bar{R})^2 \sum_i (S_i - \bar{S})^2}}$$

که در آن  $R_i$  رتبه ی متغیر  $X_i$  و  $S_i$  رتبه ی متغیر  $Y_i$ ،  $\bar{R} = \frac{\sum R_i}{n}$  و  $\bar{S} = \frac{\sum S_i}{n}$

می باشد.

**-ضریب همبستگی تاو کندال:**

$$\tau = \frac{\sum_{i < j} (\text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j))}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$$

که در آن:

$$\text{sgn}(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \\ -1 & \text{if } z < 0 \end{cases}$$

که در آن:

$$T_1 = \sum_k t_k(t_k - 1)/2 \quad T_0 = n(n-1)/2, T_2 = \sum_l u_l(u_l - 1)/2$$



**رویه ttest و آزمون استودنت و مقایسه میانگین ها:**

آزمون استودنت در برخی موارد برای مقایسه میانگین يك جامعه با عدد ثابت، مقایسه میانگین دو جامعه وابسته با یکدیگر و مقایسه میانگین يك متغیر در دو جامعه مستقل مورد استفاده قرار می‌گیرد.

**- آزمون مقایسه میانگین با عدد ثابت:**

در این آزمون 2 پذیره زیر بایستی برقرار باشد.

(1) متغیر مورد بررسی کمی و از نوع مقیاس نسبی یا فاصله‌ای باشد.

(2) يك نمونه تصادفی از جامعه استخراج شده باشد.

در عمل با توجه به اینکه توزیع متغیر مورد بررسی مشخص نیست و واریانس نیز معلوم نمی‌باشد.

حالت‌های زیر اتفاق می‌افتد:

**الف)** اگر حجم نمونه استخراج شده از جامعه بزرگ باشد ( $n \geq 30$ ) علی‌رغم مشخص نبودن توزیع متغیر پاسخ و واریانس آن می‌توان از آزمون استودنت با توجه به قضیه حد مرکزی استفاده کرد.

**ب)** اگر حجم نمونه کم باشد ( $n < 30$ ) آن‌گاه 2 حالت زیر اتفاق می‌افتد:

(1) اگر توزیع جامعه نرمال باشد یا بتوانیم نشان دهیم که مشاهدات از توزیع نرمال استخراج شده‌اند می‌توان از آزمون استودنت استفاده نمود.

(2) اگر توزیع مشاهدات غیرنرمال باشد، نمی‌توان با استفاده از آزمون استودنت فرضیه‌های فوق را آزمون نمود بلکه در این حالت بایستی از روش‌های ناپارامتری مثل آزمون علامت یا رتبه علامت‌دار ویلکاکسون حالت تک نمونه‌ای استفاده کرد.

**(نکته):** برای متغیرهای کیفی از آزمون‌های ناپارامتری استفاده می‌شود.

تذکر: آزمون استودنت نسبت به تقارن توزیع حساس است برای توزیع‌های نامتقارن مثل کی‌دو درست نیست.

**- مقایسه میانگین دو جامعه وابسته:**

در استفاده از آزمون استودنت برای مقایسه میانگین دو جامعه وابسته به یکدیگر، اگر شرایطی برقرار باشد می‌توان از این آزمون برای آزمون فرضیه تساوی میانگین دو جامعه در مقابل فرض عدم تساوی، بزرگتری و کوچکتری استفاده کرد.

(دو پذیره ذکر شده در حالت تک نمونه‌ای نیز باید در این‌جا برقرار باشد)

در عمل شرایط زیر اتفاق می‌افتد:

**حالت 1)** اگر حجم نمونه استخراج شده ( $n \geq 30$ ) در این صورت علی‌رغم مشخص نبودن توزیع متغیر مورد بررسی در دو جامعه می‌توان از آزمون استودنت برای آزمون فرضیه‌ها استفاده نمود.

**حالت 2)** وقتی حجم نمونه استخراج شده کوچک باشد ( $n < 30$ ) در این صورت دو حالت زیر ایجاد می‌شود:

**حالت 2- الف:** اگر مشاهدات از توزیع نرمال پیروی کنند می‌توان آزمون استودنت را استفاده کرد.

**حالت 2- ب:** چنانچه روش‌های توصیفی یا استنباطی مورد استفاده برای برآزش توزیع نرمال به مشاهدات نشان دهند که توزیع مشاهدات نرمال نیستند نمی‌توان آزمون استودنت را مورد استفاده قرار داد در این صورت از آزمون‌های ناپارامتری معادل، همانند آزمون علامت یا علامتدار ویلکاکسون حالت زوج نمونه‌ای استفاده می‌شود.

**- مقایسه میانگین در دو جامعه مستقل: چنانچه بخواهیم میانگین را در دو جامعه مستقل مقایسه**

نماییم. همانند حالت قبل در عمل شرایط زیر اتفاق می‌افتد:

دو پذیره در مورد انجام این آزمون فرضیه وجود دارد که عبارتند از:

**1)** دو نمونه تصادفی و مستقل از دو جامعه مستقل استخراج شده باشند.

**2)** متغیری که مقایسه بر اساس آن صورت می‌گیرد بایستی کمی باشد.

در عمل حالات زیر در نظر گرفته می‌شود:

حالت (1): چنانچه حجم نمونه‌های استخراج شده از دو جامعه بزرگ باشد ( $n_1 \geq 30$  و  $n_2 \geq 30$ ) هر دو در این صورت آزمون استودنت برای مقایسه میانگین دو جامعه مستقل کاربرد دارد. برای مشخص کردن مقدار آماده آزمون در این حالت بایستی بدانیم که آیا فرضیه برابری واریانس‌های دو جامعه رد می‌شود یا نه؟ بر اساس نتیجه این آزمون مقدار آماده آزمون استودنت آن تفاوت دارد. نرم‌افزارهای آماری از جمله SPSS, SAS طوری طراحی شده‌اند که قبل از انجام آزمون استودنت آزمون برابری واریانس‌ها را بر اساس توزیع فیشر انجام داده و P- مقدار آن را گزارش می‌کنند. محقق می‌تواند با توجه به نتیجه آزمون برابری واریانس‌ها یکی از دو آزمون استودنت انجام شده برای مقایسه میانگین‌ها را بر اساس پذیرفتن برابری واریانس‌ها و عدم آن انجام شده است انتخاب نماید.

حالت (2): حجم نمونه استخراج شده از دو جامعه بزرگ نباشد ( $n_1 < 30$  یا  $n_2 < 30$ )

**حالت 2- الف:** چنانچه بتوان نشان داد که توزیع متغیر مورد بررسی در دو جامعه نرمال است استفاده از آزمون استودنت مجاز است.

**حالت 2- ب:** اگر توزیع متغیر مورد بررسی در دو جامعه نرمال نباشد استفاده از آزمون استودنت درست نیست بلکه آزمون‌های ناپارامتری معادل، همانند آزمون میانه یا آزمون من ویتنی ویلکاکسون باید مورد استفاده قرار گیرد. با ذکر مثال، شکل کلی رویه ttest برای حالات مختلف شرح داده می‌شود.

- آزمون تك نمونه‌اي:

Proc ttest h0=عدد ثابت;

Var; نام متغیر مورد بررسی;

گزینه عدد ثابت = h0 برای آزمون میانگین جامعه با عدد ثابت می‌باشد.

- آزمون دو نمونه‌اي وابسته برای آزمون  $H_0 : \mu_a = \mu_b$  در مقابل  $H_1 = \mu_a \neq \mu_b$

Proc ttest ;

Paired a\*b ;

**- آزمون دو نمونه‌ای مستقل**

Proc ttest options ;

Class متغیر گروه‌بندی;

Var متغیر مورد بررسی;

**برخی option‌های این رویه عبارتند از:**

**1)  $\alpha=0 < p < 1$ :** از این انتخاب برای تغییر ضریب اطمینان فواصل اطمینان استفاده می‌شود که به طور پیش‌فرض توسط نرم‌افزار،  $\alpha = 0/05$  در نظر گرفته می‌شود.

**2)  $CI=equal$  یا  $umpu$ :** از این انتخاب برای ارائه فاصله اطمینان برای واریانس استفاده می‌شود.

$equal$  حالت فاصله اطمینان با در دنباله مساوی را در نظر می‌گیرد و  $umpu$  فاصله اطمینان را بر اساس پرتوان‌ترین آزمون ناریب به طور یکنواخت محاسبه می‌کند.

**3)  $H_0=m$ :** برای انجام آزمون فرضیه  $H_0$ ‌ای که عدد ثابت تحت آن مخالف صفر باشد.

مثال: مطالعه‌ی آماری در گذشته نشان داده است که در عرض سال به طور متوسط کارمندان 15 روز در اثر بیماری غیبت دارند. پژوهش‌گری برای 25 کارمند در سال گذشته شماره روزهای غیبت را به شرح زیر ثبت کرده است:

5	25	10	0	3	50	12	14	40	12
32	8	4	47	20	14	18	16	10	1
22	58	5	23	9					

با میزان 5 درصد بیازمائیید که در سال گذشته روزهای غیبت بیش از 15 روز است.

```

data;
input absent @@;
cards;
5 25 10 0 3 50 12 14 40 12 32 8 4 47 20 14 18 16 10 1 22 58 5 23 9
;
proc univariate mu0=15 normal plot;
var absent;
probplot absent;
proc ttest h0=15;
var absent;
run;

```

The UNIVARIATE Procedure  
Variable: absent

Moments

N	25	Sum Weights	25
Mean	18.32	Sum Observations	458
Std Deviation	15.844873	Variance	251.06
Skewness	1.19229579	Kurtosis	0.67511892
Uncorrected SS	14416	Corrected SS	6025.44
Coeff Variation	86.4894814	Std Error Mean	3.1689746

Basic Statistical Measures

Location		Variability	
Mean	18.32000	Std Deviation	15.84487
Median	14.00000	Variance	251.06000
Mode	5.00000	Range	58.00000
		Interquartile Range	15.00000

NOTE: The mode displayed is the smallest of 4 modes with a count of 2.

Tests for Location: Mu0=15

Test	-Statistic-	-----p Value-----
Student's t	t 1.047657	Pr >  t  0.3052
Sign	M -1.5	Pr >=  M  0.6900
Signed Rank	S 7	Pr >=  S  0.8549

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.875086	Pr < W 0.0055
Kolmogorov-Smirnov	D 0.167436	Pr > D 0.0704
Cramer-von Mises	W-Sq 0.186812	Pr > W-Sq 0.0073
Anderson-Darling	A-Sq 1.120087	Pr > A-Sq 0.0051

Quantiles (Definition 5)

Quantile	Estimate
100% Max	58
99%	58
95%	50
90%	47
75% Q3	23
50% Median	14

Variable: absent  
Quantiles (Definition 5)

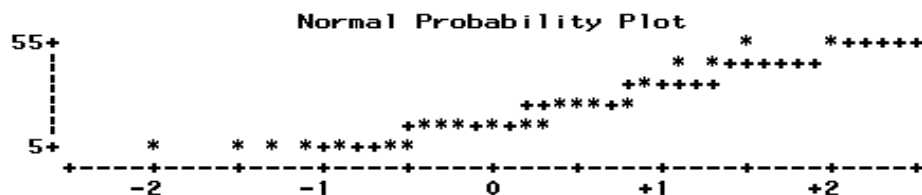
Quantile	Estimate
25% Q1	8
10%	3
5%	1
1%	0
0% Min	0

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0	4	32	11
1	20	40	9
3	5	47	14
4	13	50	6
5	23	58	22

Stem	Leaf	#	Boxplot
5	08	2	0
4	07	2	0
3	2	1	
2	0235	4	+-----+
1	00224468	8	*---+---*
0	01345589	8	+-----+

-----+-----+-----+-----+  
Multiply Stem.Leaf by 10\*\*\*+1

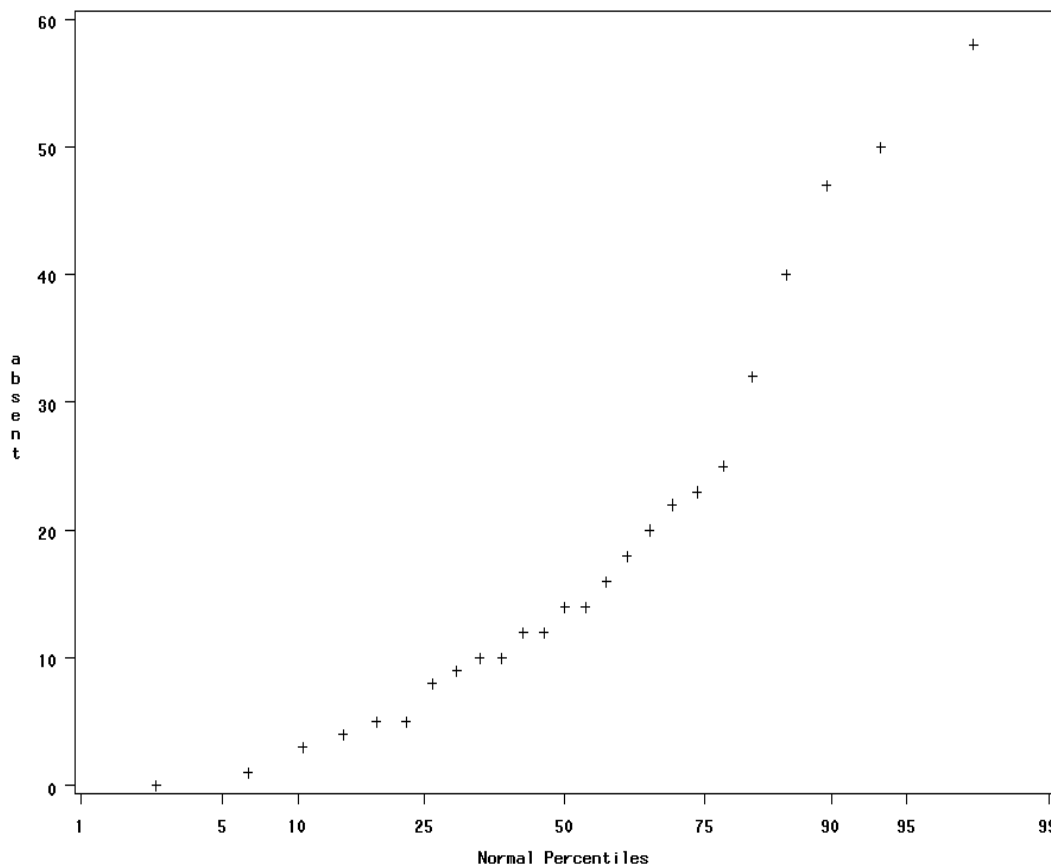


Statistics

Variable	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
absent	25	11.78	18.32	24.86	12.372	15.845	22.043	3.169	0	58

T-Tests

Variable	DF	t Value	Pr >  t
absent	24	1.05	0.3052



مثال: در يك آزمون حساب سال پنجم ابتدائي، نمرات 8 دانش آموز پسر عبارتند از:

10 و 17 و 14 و 12 و 11 و 16 و 18 و 19

و نمرات 6 دانش آموز دختر عبارتند از: 11 و 14 و 13 و 17 و 18 و 16

با فرض نرمال بودن نمرات، با میزان ده درصد بيازماييد که به طور متوسط دانش آموزان دختر و دانش آموزان پسر در درس حساب هم قوه مي باشند.

```
data;
input sex$ score @@;
cards;
m 10 m 17 m 14 m 12 m 11 m 16 m 18 m 19 f 16 f 18 f 17 f 13 f 14 f 11
;
proc ttest;
class sex ;
var score;
proc sort;
by sex;
proc univariate normal plot;
```

by sex;  
var score;  
qqplot score;  
run;

The TTEST Procedure

Statistics

Variable	sex	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
score	f	6	12.063	14.833	17.603	1.6476	2.6394	6.4735	1.0775
score	m	8	11.801	14.625	17.449	2.2334	3.378	6.8751	1.1943
score	Diff (1-2)		-3.43	0.2083	3.8464	2.2171	3.0918	5.1037	1.6697

T-Tests

Variable	Method	Variances	DF	t Value	Pr >  t
score	Pooled	Equal	12	0.12	0.9028
score	Satterthwaite	Unequal	11.9	0.13	0.8991

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
score	Folded F	7	5	1.64	0.6070

----- sex=f -----

The UNIVARIATE Procedure  
Variable: score

Moments

N	6	Sum Weights	6
Mean	14.8333333	Sum Observations	89
Std Deviation	2.63944439	Variance	6.96666667
Skewness	-0.3190464	Kurtosis	-1.1709897
Uncorrected SS	1355	Corrected SS	34.8333333
Coeff Variation	17.7940071	Std Error Mean	1.07754866

Basic Statistical Measures

Location		Variability	
Mean	14.83333	Std Deviation	2.63944
Median	15.00000	Variance	6.96667
Mode	.	Range	7.00000
		Interquartile Range	4.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 13.76581	Pr >  t	<.0001
Sign	M 3	Pr >=  M	0.0313
Signed Rank	S 10.5	Pr >=  S	0.0313

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.965793	Pr < W	0.8631
Kolmogorov-Smirnov	D 0.17076	Pr > D	>0.1500
Cramer-von Mises	W-Sq 0.026279	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq 0.177569	Pr > A-Sq	>0.2500



Quantiles (Definition 5)

Quantile	Estimate
100% Max	18
99%	18
95%	18
90%	18
75% Q3	17
50% Median	15

sex=f

The UNIVARIATE Procedure  
Variable: score

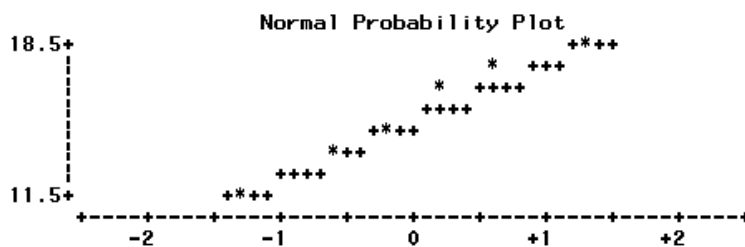
Quantiles (Definition 5)

Quantile	Estimate
25% Q1	13
10%	11
5%	11
1%	11
0% Min	11

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
11	6	13	4
13	4	14	5
14	5	16	1
16	1	17	3
17	3	18	2

Stem Leaf	#	Boxplot
18 0	1	
17 0	1	-----+-----
16 0	1	-----+-----
15		*-----*
14 0	1	-----+-----
13 0	1	-----+-----
12		
11 0	1	



sex=m

The UNIVARIATE Procedure  
Variable: score

Moments

N	8	Sum Weights	8
Mean	14.625	Sum Observations	117
Std Deviation	3.37797488	Variance	11.4107143
Skewness	-0.1320345	Kurtosis	-1.7154621
Uncorrected SS	1791	Corrected SS	79.875
Coeff Variation	23.0972641	Std Error Mean	1.19429447

Basic Statistical Measures

Location		Variability	
Mean	14.62500	Std Deviation	3.37797
Median	15.00000	Variance	11.41071
Mode	.	Range	9.00000
		Interquartile Range	6.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 12.24572	Pr >  t  <.0001
Sign	M 4	Pr =  M  0.0078
Signed Rank	S 18	Pr >=  S  0.0078

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.936352	Pr < W 0.5756
Kolmogorov-Smirnov	D 0.158014	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.038285	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.249442	Pr > A-Sq >0.2500

Quantiles (Definition 5)

Quantile	Estimate
100% Max	19.0
99%	19.0
95%	19.0
90%	19.0
75% Q3	17.5
50% Median	15.0

sex=m

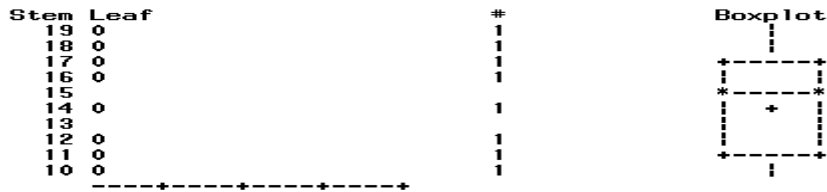
The UNIVARIATE Procedure  
Variable: score

Quantiles (Definition 5)

Quantile	Estimate
25% Q1	11.5
10%	10.0
5%	10.0
1%	10.0
0% Min	10.0

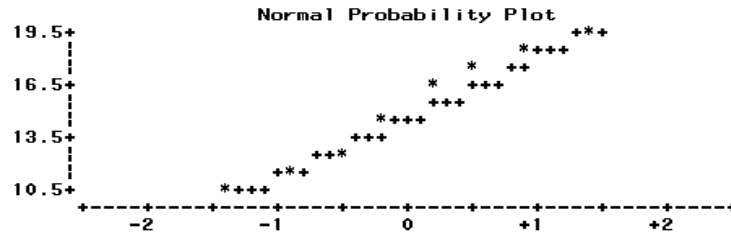
Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
10	1	14	3
11	5	16	6
12	4	17	2
14	3	18	7
16	6	19	8

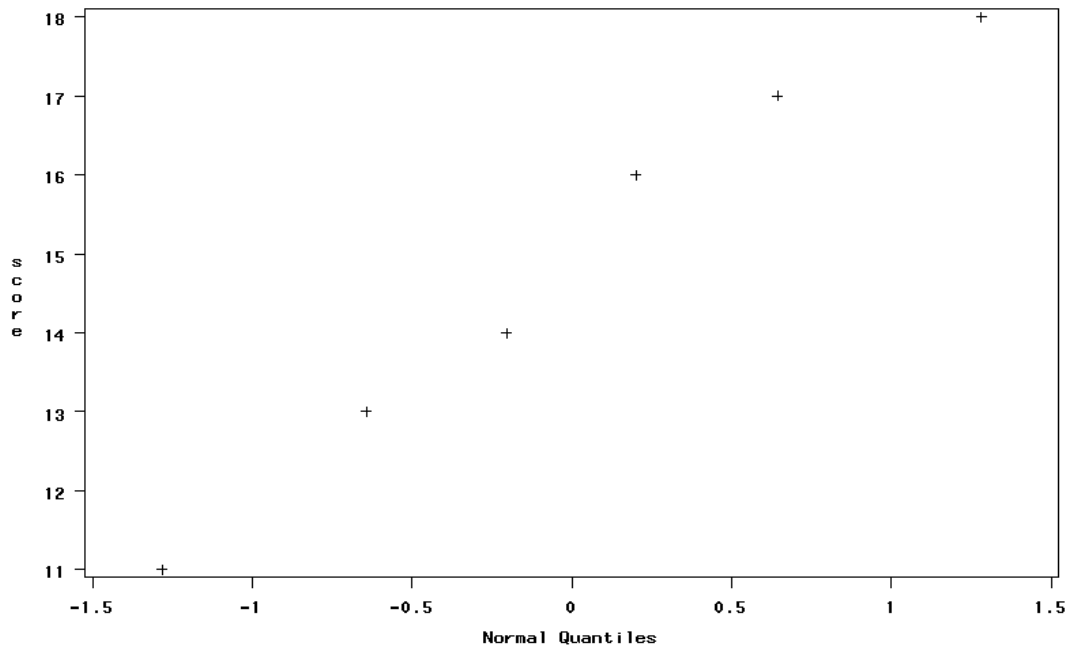
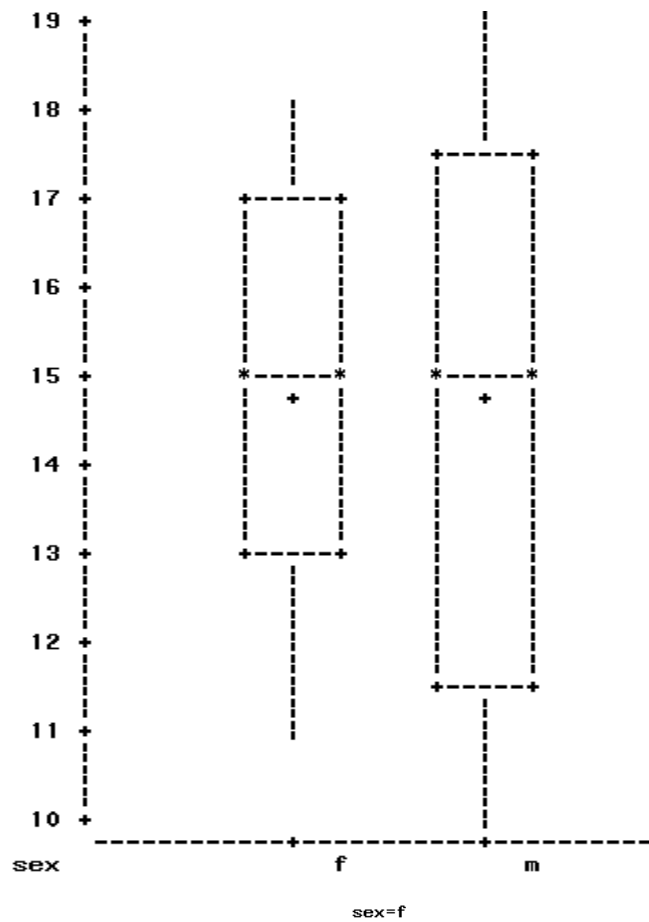


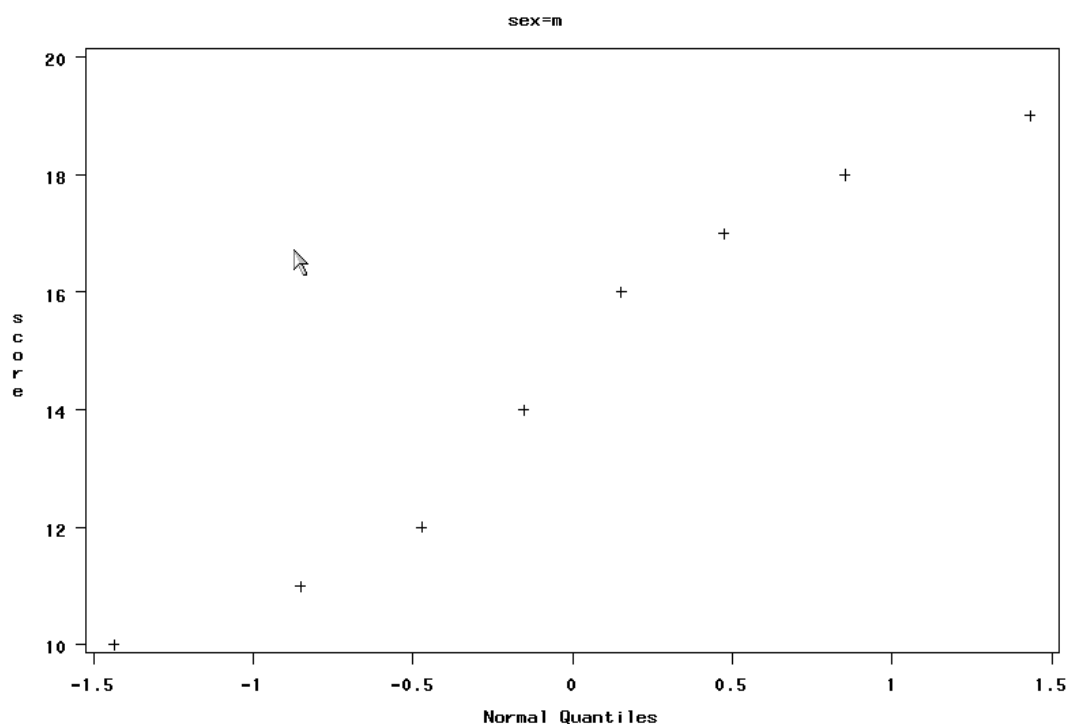
sex=m

The UNIVARIATE Procedure  
Variable: score



The UNIVARIATE Procedure  
Variable: score  
Schematic Plots





مثال: مطالعه‌ی جدیدی در توان جسمی هفت نفر بر حسب پوند قبل و بعد از یک دوره کارورزی انجام شده است. مطلوبست مقایسه توان جسمی قبل و بعد از دوره، داده‌ها به شرح زیرند:

فرد آزمون شده	1	2	3	4	5	6	7
قبل از دوره	100	110	90	110	125	130	105
بعد از دوره	115	125	105	130	140	140	125

```

data;
input before after @@;
diff=after-before;
cards;
100 115 110 125 90 105 110 130 125 140 130 140 105 125
;
proc means t prt;
var diff;
proc univariate normal plot;
var diff;
proc ttest;
paired before*after;
run;
    
```

The MEANS Procedure

Analysis Variable : diff

t Value	Pr >  t
12.05	<.0001

The UNIVARIATE Procedure

Variable: diff

Moments

	7	Sum Weights	7
N	7	Sum Observations	110
Mean	15.7142857	Variance	11.9047619
Std Deviation	3.4503278	Kurtosis	0.336
Skewness	-0.1738965	Corrected SS	71.4285714
Uncorrected SS	1800	Std Error Mean	1.30410133
Coeff Variation	21.9566314		

Basic Statistical Measures

Location		Variability	
Mean	15.71429	Std Deviation	3.45033
Median	15.00000	Variance	11.90476
Mode	15.00000	Range	10.00000
		Interquartile Range	5.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 12.0499	Pr >  t  <.0001
Sign	M 3.5	Pr >=  M  0.0156
Signed Rank	S 14	Pr >=  S  0.0156

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.840044	Pr < W 0.0995
Kolmogorov-Smirnov	D 0.296288	Pr > D 0.0624
Cramer-von Mises	W-Sq 0.127665	Pr > W-Sq 0.0377
Anderson-Darling	A-Sq 0.654365	Pr > A-Sq 0.0495

Quantiles (Definition 5)

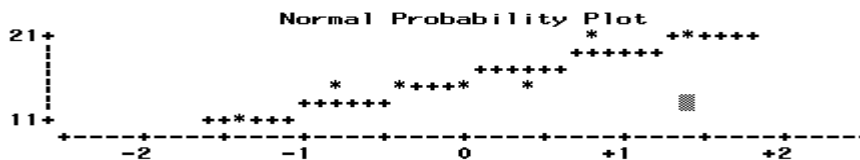
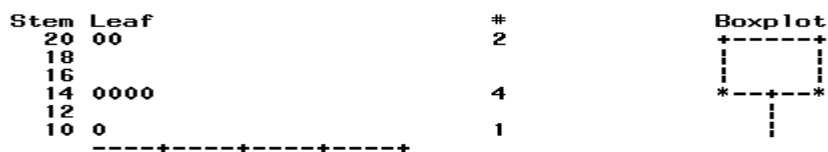
Quantile	Estimate
100% Max	20
99%	20
95%	20
90%	20
75% Q3	20
50% Median	15
25% Q1	15
10%	10

Quantiles (Definition 5)

Quantile	Estimate
5%	10
1%	10
0% Min	10

Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
10	6	15	2
15	5	15	3
15	3	15	5
15	2	20	4
15	1	20	7



The TTEST Procedure

Statistics

Difference	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
before - after	7	-18.91	-15.71	-12.52	2.2234	3.4503	7.5978	1.3041

T-Tests

Difference	DF	t Value	Pr >  t
before - after	6	-12.05	<.0001

**طرح و تجزیه و آزمایشات:** در این بخش روش استفاده از نرم افزار sas در طرح و تجزیه آزمایشات را توضیح می دهیم.

### رویه آنالیز واریانس proc anova:

شکل کلی این رویه به صورت زیر است:

Proc anova ;

Class عوامل ;

Model اثرات = متغیر وابسته ;

Means /options ; نام اثر یا اثرات

Test h= نام اثر یا اثرات e= نام اثر خطا ;

برای مقایسه جفت میانگین های تیمار می توان از option های Snk, Bon, dunnett, duncan, Scheffe, Lsd و سایر آزمون های دیگر، در زیر دستور means استفاده نمود و گزینه  $\alpha=0 < p < 1$  سطح معنی داری مورد نظر برای آزمون های مقایسه جفتی را مشخص می کند. **زیر دستور test:** هرگاه بخواهیم اثر یا اثرات مشخصی را در برابر یک اثر خطای معین آزمون کنیم از زیر دستور test می کنیم.

**گزینه hovtest:** در صورتی که آنالیز واریانس یک طرفه داشته باشیم فرض همگن بودن

واریانس ها (برابری واریانس ها:  $H_0$ ) بررسی می کند. ضمناً حجم تیمارها نباید از 3 کمتر باشد نرم افزار sas به طور پیش فرض آزمون Leven برای همگن بودن واریانس ها را انجام می دهد و

نیز می توان از آزمون های همگن بودن دیگر از قبیل HARTLEY, BROWN-

FORSYTHE-OBRIEN, COCHRAN, BARTLETT استفاده نمود که این آزمون ها را

می توان به صورت گزینه

نام آزمون **hovtest=** در زیر دستور means به کار برد.