آزمون t استودنت

(S-PLUS)

تدوین: مرکز تحلیل آماری خوارزمی

❋❋❋

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

آزمون t استودنت برای ارزیابی میزان هم‌قوارگی یا یکسان بودن و نبودن میانگین نمونه‌ای با میانگین جامعه در حالتی به کار می‌رود که انحراف معیار جامعه مجهول باشد چون توزیع t در مورد نمونه‌های کوچک با استفاده از درجات آزادی تعدیل می‌شود، می‌توان از این آزمون برای نمونه‌های بسیار کوچک استفاده نمود. در این ادامه چگونگی انجام این تست در نرم افزار را بیان می کنیم.

مطالب بیان شده در این متن:

آزمون تک نمونه ای tاستودنت

بررسی نرمال بودن توزیع

آزمون دو نمونه ای tاستودنت

آزمون هم واریانسی دو جامعه

آزمون تک نمونه ای t استودنت

تعداد زیادی از پژوهش ها مشتمل بر دو نمونه یا بیشتر از داده ها است. اما این امر همیشه اتفاق نمی افتد و گاهی محقق از یک نمونه واحد از مشاهدات برای مطالعه در مورد فقط یک جامعه استفاده می کند. حالاتی که در آن فرد از آزمون تک نمونه ای استفاده می کند، دو نوع عمده است:

۱- ممکن است فرد بخواهد توزیع یک نمونه را با یک توزیع فرضی مانند توزیع نرمال مقایسه کند. این یک سوال در مورد تطابق توزیع با یک توزیع نظری است.
۲- ممکن است فرد بخواهد از طریق آماره های یک نمونه در مورد پارامترهای یک جامعه واحد استنباط کند. این کار ممکن است برای تعیین اینکه آیا نمونه مورد نظر از یک جامعه شناخته شده است یا برای برآورد کردن پارامترهای یک جامعه ناشناخته به کار می رود.

زمانی از این نمونه استفاده می شود که واریانس جامعه نامعلوم و حجم نمونه کم(کمتر از ۲۵ یا ۳۰) باشد.

t-test تک نمونه ای آزمون می کند که آیا میانگین یک جامعه برابر مقدار مفروضی است یا نه؟ به عبارت دیگر آزمون های زیر بر قرار است:

$$(1) \quad \begin{cases} H_0: \mu \le \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

$$(2) \quad \begin{cases} H_0: \mu \ge \mu_0 \\ H_1: \mu \ne \mu_0 \end{cases}$$

$$(3) \quad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \ne \mu_0 \end{cases}$$

فرض هایی که برای این آزمون وجود دارد:

۱. مشاهدات دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشند.

۲. مشاهدات مستقل از هم باشند.

در نتیجه داریم: $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$, $\bar{x} = \frac{\sum x_i}{n}$

آنگاه با فرض درست بودن H_0 آماره آزمون را از رابطه زیر محاسبه می کنیم.

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

الف) اگر $T \ge t(n-1, \alpha)$ فرض H_0 رد می شود.

ب) اگر $T \le -t(n-1, \alpha)$ فرض H_0 رد می شود.

ج) اگر $|T| \ge t(n-1, \alpha)$ فرض H_0 رد می شود.

عبارت $t(n-1, \alpha)$ مقدار بالای توزیع t استودنت با n-1 درجه آزادی است و α خطای نوع اول است.

انجام آزمون تک نمونه ای t استودنت در نرم افزار S-PLUS

برای انجام این آزمون از داده های موجود در نرم افزار استفاده می کنیم. آیکون object Explorer در نوار استاندارد را زده، تا پنجره مربوطه باز شود. مراحل زیر را طی نمایید.

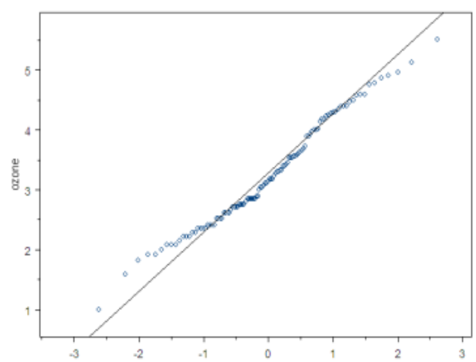
Search path > data > air

داده های air برای انجام آزمون مورد نظر انتخاب شده است. در این سری داده ۴ متغیر و ۱۱۱ مشاهده برای هر متغیر مشاهده می شود. برای داده های موجود می خواهیم فرض $\begin{cases} H_0: \mu = 3.2 \\ H_1: \mu \neq 3.2 \end{cases}$ را آزمون کنیم. قبل از انجام آزمونt، ضروری است که داده ها را از نظر وجود مقادیر پرت و شکل توزیع بررسی کنیم، برای این منظور نمودارهای QQ-Normal, Box Plot و Density مربوط به داده ها را رسم می کنیم. (ما در اینجا محاسبات را بر روی داده ی ozone انجام می دهیم.)

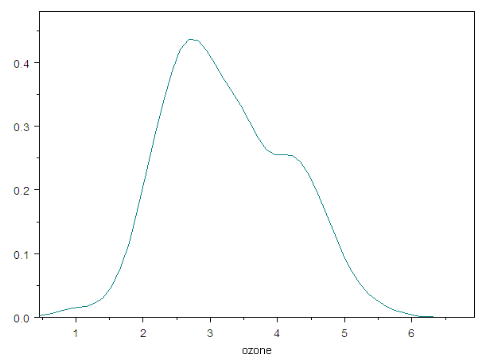
برای آموزش چگونگی رسم نمودار در نرم افزار s-plus به صفحه ی آموزش نرم افزار در سایت مرکز تحلیل آماری خوارزمی مراجعه نمایید. به آدرس زیر:

http://www.kharazmi-statistics.ir/fa/narmafzarsplus.aspx

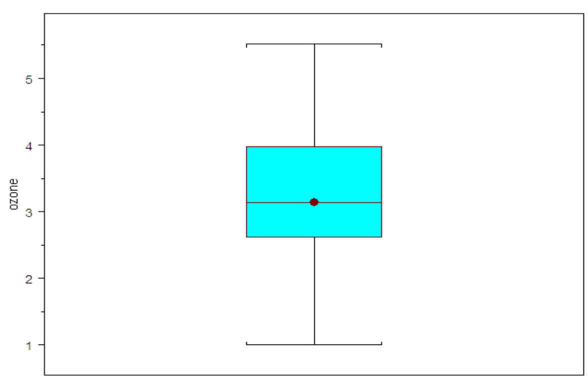
تصاویر نمودارهای رسم شده برای داده ی ozone به صورت زیر است.



نمودارQQ-Normal



نمودار Density

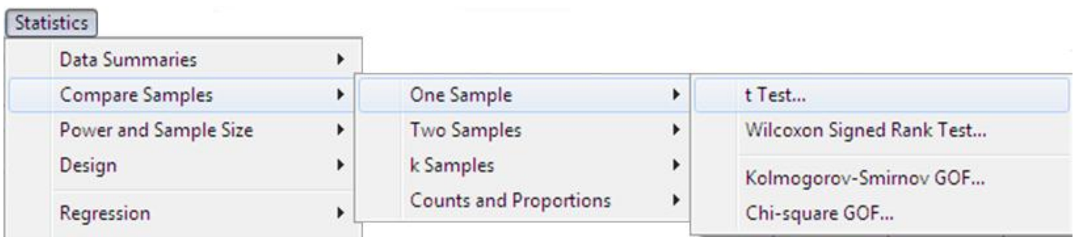


نمودارBox Plot

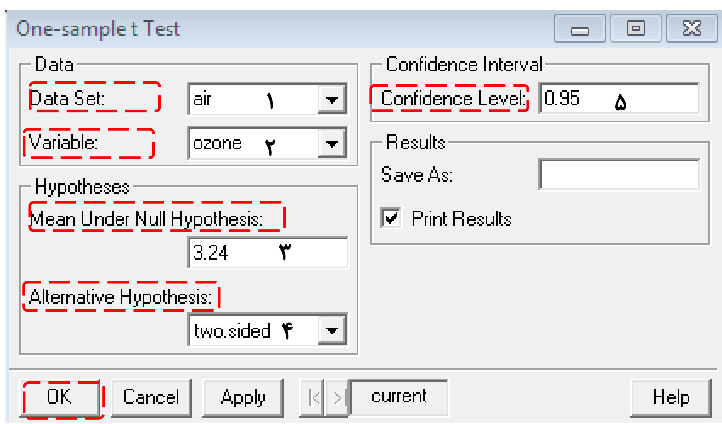
نمودار ها نشانگر آن است که چولگی توزیع مشاهدات به راست است. (یعنی داده ها به سمت مقادیر کوچکتر متمایلند.) با وجود این توزیع را نرمال در نظر می گیریم زیرا چولگی بسیار شدید نیست. باید در نظر داشت که

نرمال بودن توزیع یکی از پیش فرض های آزمون t است. و در صورتی که توزیع نرمال نبود باید از معادلات ناپارامتری استفاده کنیم. برای انجام آزمون t استودنت مراحل زیر را طی نمایید.

Statisctics > compare samples > one sample> t Test…



پس از آن پنجره ای مشابه پنجره پایین باز می شود. توضیحات هر کادر متناسب با شماره قرار گرفته سده در پایین داده شده است.



۱. مجموعه داده را انتخاب می کنیم.
۲. متغیری که قصد داریم آزمون را بر روی آن انجام دهیم انتخاب می کنیم.
۳. مقدار میانگینی که در آزمون مد نظر است را می نویسیم.
۴. فرض مقابل فرض دو طرفه است. در نتیجه گزینه ی two.sided را انتخاب می کنیم. بقیه گزینه ها greater و less است و به معنای این است که در فرض مقابل میانگین بزرگتر از و یا کوچکتر از مقداری باشد که مدنظر داریم.
۵. سطح اطمینان ۹۵٪ را درنظر می گیریم.

پس از پر کردن فیلدهای مورد نظر دکمه ی ok را بزنید. نتایج نمایش داده می شود.

تحلیل آزمون:

```
     One-sample t-Test

data:  ozone in air
t = 0.0921, df = 110, p-value = 0.9268
alternative hypothesis:  mean is not equal to 3.24
95 percent confidence interval:
 3.080299 3.415268
sample estimates:
 mean of x
  3.247784
```

همانطور که مشاهده می شود مقدار آماره ی t ، درجه آزادی، و مقدار p-value محاسبه شده است. با توجه به مقدار p-value که از مقدار α=۰.۰۵ بزرگتر است می توان گفت با اطمینان ٪۹۵ دلیلی بر رد فرض وجود ندارد.

آزمون دو نمونه ی مستقل t استودنت

زمانی که می خواهیم پارامترهای دو جامعه را مقایسه کنیم از آزمون های دو نمونه ای استفاده می کنیم در این آزمون ها فرض صفر به صورت $H_0: \theta_1 = \theta_2$ است. که در آن θ_1 و θ_2 پارامترهای دو جامعه هستند.

برای مقایسه میانگین های دو جامعه به حجم های n_1 و n_2 به شرطی که حجم نمونه ها کوچکتر از ۲۵ یا ۳۰ باشد و هر دو نمونه دارای توزیع نرمال و واریانس های دو جامعه نیز با برابر باشند، (یا لااقل تفاوت فاحشی با هم نداشته باشند تا بتوان آنها را برابر فرض کرد) از آزمون دو نمونه ای t استفاده می کنیم. به عبارت دیگر:

$$n_1 \sim N(\mu_1, \sigma) \quad , \quad n_2 \sim N(\mu_2, \sigma)$$

در این آزمون یکی از حالت ها زیر را مطرح است:

$$(1) \qquad \begin{cases} H_0: \mu_1 - \mu_2 \leq a \\ H_1: \mu_1 - \mu_2 > a \end{cases}$$

$$(2) \qquad \begin{cases} H_0: \mu_1 - \mu_2 \geq a \\ H_1: \mu_1 - \mu_2 < a \end{cases}$$

$$(3) \qquad \begin{cases} H_0: \mu_1 - \mu_2 = a \\ H_1: \mu_1 - \mu_2 \neq a \end{cases}$$

برای محاسبه آماره آزمون با فرض درست بودن H_0 دو حالت خواهیم داشت:

I. واریانس دو جامعه مجهول ولی برابر که از واریانس ادغامی استفاده می کنیم (S_p^2)

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_1-2}$$

سپس آماره آزمون از رابطه :

$$T = \frac{\bar{x}_1 - \bar{x}_2 - a}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

محاسبه می شود که دارای توزیع t استودنت با $d.f = n_1 + n_2 - 2$ د رجه آزادی است.

II. در رابطه حالتی که واریانس دو جامعه مجهول و نابرابر باشند آنگاه آماره آزمون از رابطه:

$$T = \frac{\bar{x}_1 - \bar{x}_2 - a}{S_p\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

محاسبه می شود که دارای توزیع t استودنت با درجه آزادی زیر می باشد:

$$d.f = \frac{(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}}$$

$\bar{x}_1$ میانگین نمونه اول S_1^2 واریانس نمونه ی اول،n_1 حجم نمونه اول و $\bar{x}_2$ میانگین نمونه دوم S_2^2 واریانس نمونه ی دوم،n_2 حجم نمونه دوم.

سپس این مقدار T با مقادیر جدول توزیع t استودنت مقایسه ی می کنیم:

الف) اگر $T \geq t_{(d.f,\alpha)}$ فرض H_0 رد می شود.

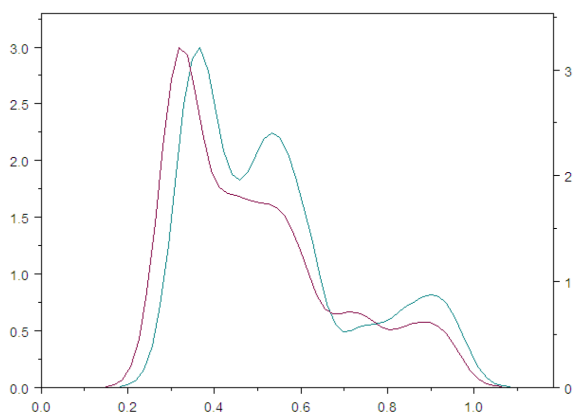
ب) اگر $T \leq -t_{(d.f,\alpha)}$ فرض H_0 رد می شود.

ج) اگر $|T| \geq t_{(d.f,\alpha)}$ فرض H_0 رد می شود.

$t_{(d.f,\alpha)}$ مقدار بالائی توزیع t استودنت با d.f درجه آزادی است. α خطای نوع اول است.

انجام آزمون دو نمونه ی مستقل t استودنت در نرم افزارS-PLUS

برای انجام آزمون ما از داده های Sensors از سری داده هایی که از قبل در نرم افزار به صورت پیش فرض قرار داده شده است استفاده می کنیم. در این سری داده می خواهیم دو متغیر V_7 و V_5 را با هم آزمون کنیم. در ابتدا باید از نرمال بودن توزیع دو جامعه مورد بررسی اطمینان حاصل کنیم. همانند حالت قبل سری نمودارهای جعبه ای، density و QQ-Normal را رسم می کنیم.



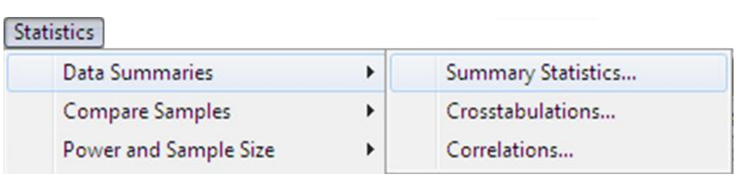
نرمال بودن توزیع دو جامعه: با توجه به نمودار Density کاملا واضح است که این سری داده ها چولگی به سمت راست دارند. در نتیجه توزیع داده ها نرمال نیست و باید از معادلات ناپارمتری استفاده کرد. در اینجا از این فرض صرف نظر کرده و مسیر بدست آوردن را بدست می آوریم اما در داده های حقیقی تحلیل که از این سری داده بدون دارا بودن توزیع نرمال بدست می آید صحیح نیست.

از دیگر شروط برای انجام این آزمون برابر بودن واریانس دو جامعه است. اگر واریانس دو جامعه با هم برابر باشند، آماره آزمون t دارای توزیع دقیقt استودنت خواهد بود ولی در حالت عدم برابری واریانس های جوامع، این آماره دارای توزیع تقریبی t استودنت است.

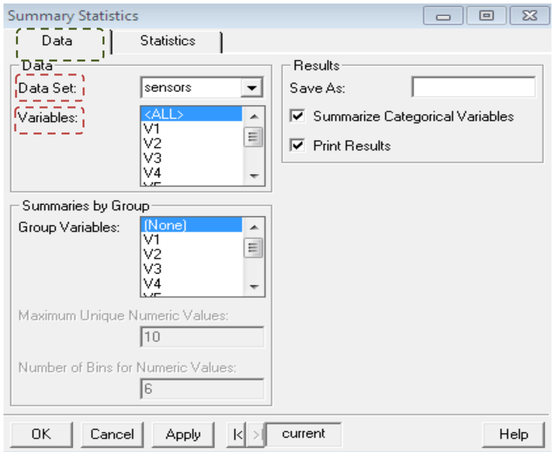
آزمون هم واریانس دو جامعه

برای انجام آزمون برابری واریانس ها لازم است مسیر زیر طی شود.

Statistics > Data Summaries > Summary Statistics…



با انجام این مراحل پنجره روبرو باز می شود.



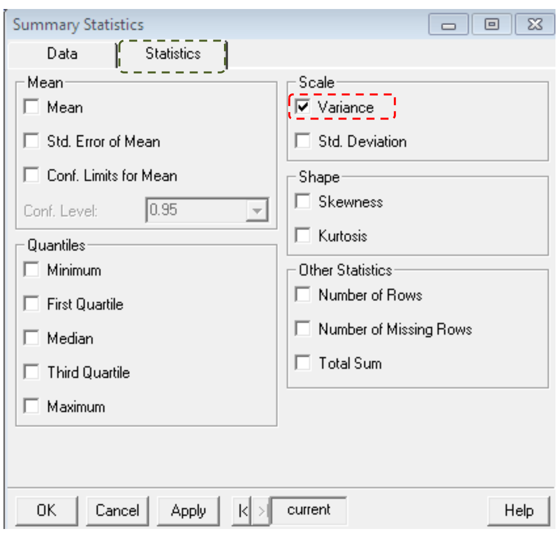
زبانه ی Data

✓ در کادر data set نام فایل مورد نظر را انتخاب می کنید.

✓ در کادر variable دو متغیری که قصد آزمون هم واریانسی آنها را دارید انتخاب کنید.

باقی کادرها را در حالت پیش فرض افزار قرار داده و تغییری در آنها ندهید.

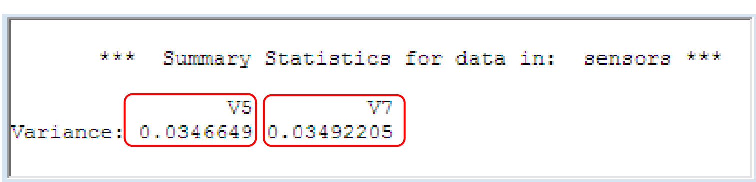
زبانه ی statistics



در این زبانه انواع متغیرهای مرکزی و پراکندگی وجود دارد ولی ما در اینجا تنها به محاسبه ی واریانس نیاز داریم. در نتیجه تنها مربع مربوط به واریانس را تیک دار کرده و مابقی را غیر فعال می کنیم.

با زدن دکمه ok محاسبات انجام می شود و نتیجه به صورت زیر نمایش داده می شود.

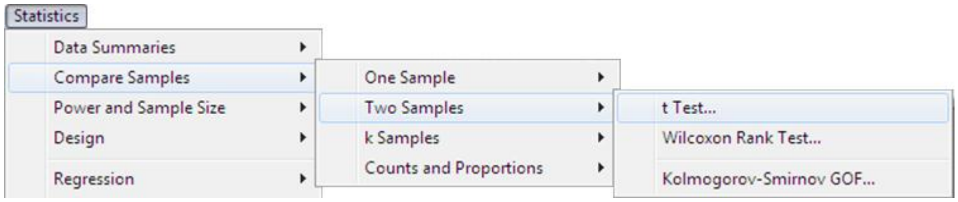
مقدار $VAR_{V5} = 0.0346649$ و $VAR_{V7} = 0.034922045$ است،که تقریبا با هم برابر است. در نتیجه متغیرهای مورد نظر هم واریانس هستند.



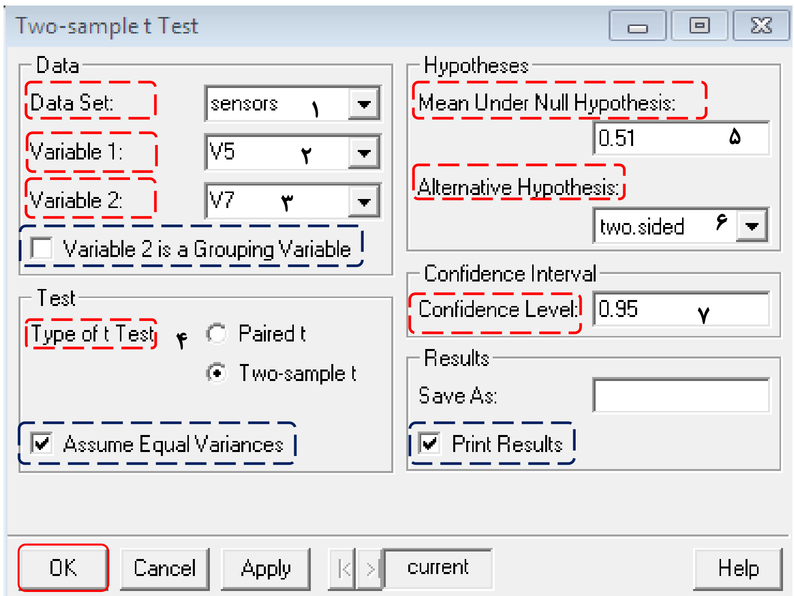
```
*** Summary Statistics for data in:  sensors ***

                V5         V7
Variance: 0.0346649 0.03492205
```

برای انجام آزمون T استودنت مراحل زیر را طی نمایید:

Statistics > Compare Samples > Two Sample > t Test…



کادر زیر ظاهر می شود.که توضیحات هر یک از کادرها با شماره های قرار گرفته شده در کنارشان بیان شده است.



۱. نام سری داده های مورد بررسی
۲. متغیر اول که قصد بررسی آن را دارید انتخاب کنید. (یعنی V5)
۳. متغیر دوم که قصد بررسی آن را دارید انتخاب کنید. (یعنی V7)

اگر متغیر دوم، گروه بندی شده باشد گزینه Variable 2 is A Grouping Variable را تیک دار می کنیم. در این مثال متغیر دوم گروه بندی شده نیست در نتیجه تیک دار نمی کنیم.

۴. در این قسمت Paired t به معنای داده های جفتی است. یعنی داده هایی که از یک فرد خاص در دو شرایط گرفته شده است مثلا وضعیت سلامت فرد قبل و بعد از خوردن یک دارو. و قسمت Two-sample t مربوط به مشاهدات دو نمونه ای است. در این مثالم شاهدات از نوع دو نمونه ای است.

گزینه ی assume equal variances مربوط به فرض برابری واریانس است که با توجه به آزمون برابری واریانس و برابر بودن واریانس ها عبارت را تیک دار می کنیم.

۵. در این قسمت مقدار میانگینی که می خواهیم فرض صفر را تعریف کنیم قرار می دهیم. فرض آزمون به صورت روبرو است.
$$\begin{cases} H_0: \mu_1 - \mu_2 = 0.51 \\ H_1: \mu_1 - \mu_2 \neq 0.51 \end{cases}$$

۶. بنا به آزمون ارائه شده در خط قبل آزمون دو طرفه را انتخاب می کنیم.

۷. سطح اطمینان ۹۵٪ را در نظر بگیرید.

گزینه ی Print Results برای چاپ نتایج است. و سپس گزینه ی ok را بزنید. نتایج به صورت زیر می باشد.

تحلیل آزمون:

```
      Pooled-Variance Two-Sample t-Test

data:  x: V5 in sensors , and y: V7 in sensors
t = -15.9237, df = 158, p-value = 0
alternative hypothesis:  difference in means is not equal to 0.51
95 percent confidence interval:
 -0.01788888  0.09861388
sample estimates:
 mean of x mean of y
  0.534525 0.4941625
```

همان طور که مشاهده می کنید عدد مربوط به p-value = 0 کمتر از سطح معنی داری ۰.۰۵ است. در نتیجه با توجه به مطالب بالا چون عدد محاسبه شده از ۰.۰۵ کمتر است در نتیجه فرض صفر رد می شود. یعنی اختلاف بین میانگین دو نمونه برابر ۰.۵۱ نیست.

اگر به جای عدد ۰.۵۱ عدد 0 را قرار دهیم به این معناست که میانگین دو جامعه با هم برابر است. به شکل زیر:

$$H_0: \mu_1 - \mu_2 = 0 \qquad \text{یا} \qquad H_0: \mu_1 = \mu_2$$

که با توجه به آماره ی بدست آمده رد و یا تایید می شود. در آن صورت در ارائه در صورت عدم رد فرض صفر در گزارش نوشته می شود دلیلی بر رو فرض صفر وجود ندارد در نتیجه تفاوت معنی داری در میانگین های دو جامعه وجود ندارد. و در صورت رد فرض صفر نوشته می شود فرض صفر پذیرفته نشده و رد می شود. بدین معنا که تفاوت معنی داری بین میانگین های دو جامعه وجود دارد.

منبع:

- آموزش کاربردی و آموزش نرم افزار s-plus، گردآوردنگان دکتر مسعود نیکوکار، حبیب تربتی قره باغ، مرضیه سهیلی راد، راضیه اولاد دیلمقانیان، انتشارات گسترش علوم پایه،۱۳۸۴
- کاربرد نرم افزار s-plus در تحلیل آماری، تالیف: آیتین سعادت، مهدی مختارپور، زینب نوروزی، انتشارات جهاد دانشگاهی واحد اصفهان، ۱۳۸۹