آزمون t

(R)

تدوین: مرکز تحلیل آماری خوارزمی

❋❋❋

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

یکی از آزمون های پرکاربرد در آمار آزمون t است. آزمون t در زمانی کاربرد دارد که مقدار انحراف معیار مشخص نباشد. در این حالت نمی توان از آزمون z استفاده کرد و از آزمون t استفاده می کنیم. انجام این آزمون دارای پیش فرض هایی است که باید قبل از انجام آزمون در نظر گرفته شود.

۱. توزیع داده ها باید نرمال باشد. (با استفاده از آزمون کولموگرف-اسمیرنوف)
۲. داده ها دارای مقیاس فاصله ای یا نسبی باشند. (همانند سن، نمره آزمون های روانی مانند افسردگی و...)
۳. دو گروهی که قصد مقایسه آن ها را داریم مقیاس اسمی داشته باشند. مانند مقایسه دو گروه از مادران، یا مقایسه دو برند از ماشین.

آزمون t در نرم افزارR برای متغیرهای تک نمونه، متغیرهای جفتی و متغیرهای دو نمونه ای قابل انجام است. در ادامه چگونگی انجام آزمون در نرم افزار بیان شده است.

مطالبی که در این متن بیان شده عبارتنداز:

آزمون t تک نمونه

آزمون t دو نمونه ای

آزمون t دو نمونه ای در حالت واریانس های برابر

آزمون t دو نمونه ای در حالت واریانس های ناآزمون t تک نمونه

آزمون t دو نمونه ای

آزمون t دو نمونه ای در حالت واریانس های برابر

آزمون t دو نمونه ای در حالت واریانس های برابر

آزمون t برای نمونه های جفتیبرابر

آزمون t برای نمونه های جفتی

آزمون t تک نمونه

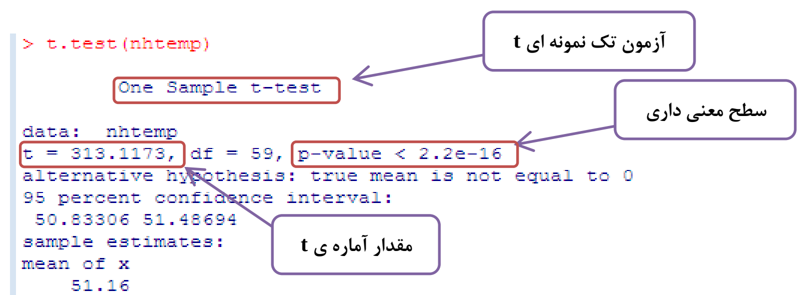
زمانی که انحراف معیار در مسئله مجهول باشد از آزمون t استفاده می کنیم. در این حالت آماره $t=\frac{x-\mu}{\frac{s}{\sqrt{n}}}$ است، در اینجا s انحراف معیار نمونه، و جایگزین σ یعنی انحراف معیار جمعیت می شود. از آماره ی t و آزمون t زمانی استفاده می کنیم که:

الف) توزیع داده ها نرمال و n کوچک باشد. در نتیجه توزیع t دارای n-1 درجه آزادی است.

ب) اگر n بزرگ باشد در نتیجه می توان گفت توزیع نرمال است.

همانطور که گفته شده یکی از شروط اولیه برای انجام آزمون t نرمال بودن داده ها است. جهت مطالعه بیشتر برای آزمون کلموگرف-اسمیرنوف و دیگر آزمون ها برای تشخیص نرمال بودن توزیع داده ها به فایل "آزمون نرمال بودن داده ها" در پایگاه اینترنتی مرکز تحلیل آماری خوارزمی مراجعه نمایید.

جهت انجام آزمون t تک نمونه ای داده های nhtemp را از سری داده های پیش فرض نرم افزار انتخاب می کنیم. با توجه به آزمون داده های مورد نظر نرمال هستند در نتیجه پیش فرض مورد نظر را دارد. جهت انجام آزمون t تنها لازم است عبارت ***t.test (نام داده)*** را تایپ نمایید. در صورت انجام صحیح مراحل خروجی شبیه به تصویر پایین نمایش داده می شود.



همانطور که در خروجی مشاهده می کنید با تایپ یک متغیر، نرم افزار آزمون تک نمونه ای t را انجام می شود. در گزارش ارائه شده توسط نرم افزار مقدار آماره ی t=۳۱۳.۱۱۷۳ و مقدار p کمتر از سطح معنی داری بدست آمده است. در نتیجه فرض t در این آزمون رد می شود. فرض صفر و مقابل در آزمون بالا به صورت زیر است.

$$\begin{cases} H_0: \mu = 0 \\ H_1: \mu \neq 0 \end{cases}$$

چنانچه عبارت دیگری غیر از عبارت بالا در نرم افزار نوشته نشود نرم افزار مقدار فرض مقابل را به طور پیش فرض صفر و مقدار سطح آزمون را ۹۵٪ در نظر گرفته می گیرد. عبارت کامل برای فرمان آزمون t به صورت زیر است. خروجی نرم افزار برای هر دو عبارت پایین یکسان و مانند تصویر بالا است.

```
>t.test(nhtemp)

> t.test(nhtemp, alternative=c("two.sided"),mu=0,conf.level=0.95)
```

سطح معنی داری (پیش فرض ۹۵٪) است. و برآورد فاصله میانگین بر حسب فرض جایگزین

عدد مورد نظر برای آزمون میانگین. پیش فرض نرم افزار صفر است.

فرض جایگزینی توسط رشته ای از کارکترها بیان می گردد. پیش فرض آن **two.sided** است. عبارت های **greater** و**less** نیز برای حالت بزرگتر از و کوچکتر از به کار می رود. شما می توانید از حروف اول آن نیز برای به کارگیری استفاده کنید. همچنین می توانید از عبارت اختصاری **alt** به جای **alternative** استفاده نمایید.

آزمون t دو نمونه ای

آزمون t دو نمونه ای براساس آمار $t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ و فرض نرمال بودن داده ها و یا تقریبا نرمال بودن داده ها می باشد. مشاهده می کنید که مخرج متفاوت تر از آزمون یک نمونه ای است. اساسا اگر ما بتوانیم فرض بیشتری داشته باشیم مبنی بر اینکه دو نمونه ای ها انحراف معیار یکسان (ناشناخته) دارند، این ساده سازی می شود.

آزمون t دو نمونه ای در حالت واریانس های برابر

دو سری داده های زیر را در نظر بگیرید.

a=(۷۹.۹۸,۸۰.۰۴,۸۰.۰۲,۸۰.۰۴,۸۰.۰۳,۸۰.۰۳ ,۸۰.۰۴ ,۷۹.۹۷ ,۸۰.۰۵ ,۸۰.۰۳ ,۸۰.۰۲ ,۸۰.۰۰ ,۸۰.۰۲)

b=(۸۰.۰۲, ۷۹.۹۴ ,۷۹.۹۷ ,۷۹.۹۸,۷۹.۹۷ ,۸۰.۰۳ ,۷۹.۹۵)

در این مسئله فرض بر آن است که دو نمونه دارای توزیع نرمال و واریانس های برابر هستند. در این حالت با رسم نمودار می توان دید که میانگین اعداد در سری A بیشتر از B هستند. در این مسئله می خواهیم آزمون برابری میانگین های دو سری را با آزمون t بررسی نماییم. عبارت زیر را در نرم افزار بنویسید.

```
>t.test(a,b,var.equal=TRUE)
```

با نوشتن عبارت مختصر بالا تمامی پیش فرض های مدنظر در نرم افزار اعمال می شود در این حالت نیز برای اعمال فرض های مدنظر خود می توان فرامینی شبیه به حالت آزمون t تک نمونه ای را نوشت. دستور کلی فرمان برای آزمون دو نمونه ای به صورت زیر است.

```
t.test(x, y, mu, alternative, conf.level, paired,var.equal)
```

x,y: بردارمشاهدات

mu: فرض صفر آزمون

alternative: فرض مقابل در آزمون

conf.level: سطح آزمون

paired: آزمون زوجی است یا مستقل

var.equal: برابری واریانس ها

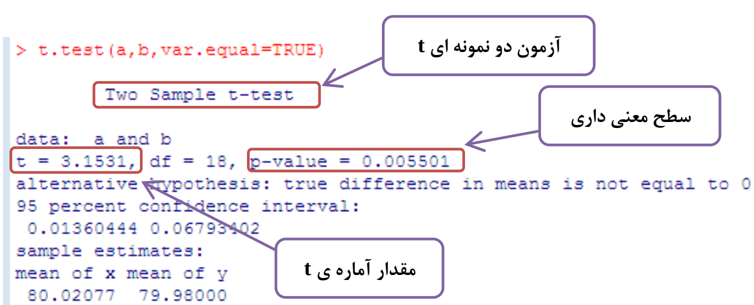
فرض صفر در این حالت به صورت زیر است.

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

گزارش نمایش داده شده در هر دو عبارت در نرم افزار یکسان است.

```
> t.test(A,B,var.equal=TRUE)

>t.test(a,b,alt="two.sided",var.equal=TRUE)
```



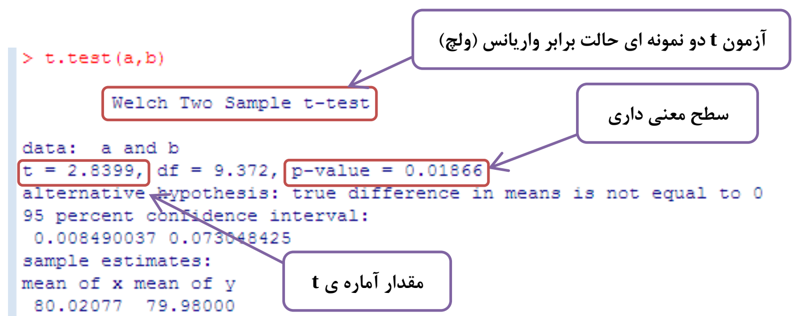
همانطور که در خروجی نرم افزار مشاهده می کنید چنانچه داده ها دارای برابری واریانس باشند از آماره t دو نمونه ای برای محاسبه استفاده می شود. در اینجا آماره t و مقدار p برآورد شده است که با توجه به مقدار p کمتر بودن آن نسبت به مقدار سطح معنی داری ۹۵٪ در نتیجه فرض صفر رد می شود و دلیلی بر پذیرش برابری میانگین های دو گروه در حالت برابری واریانس ها وجود ندارد.

آزمون t دو نمونه ای در حالت واریانس های نابرابر

چنانچه واریانس ها نابرابر باشند مخرج در آماره ی t برای محاسبه از نظر ریاضی سخت تر است اما با R اینگونه نیست. تنها تفاوت در این حالت آن است که عبارت var.equal=TRUE را در نرم افزار نمی نویسیم.

```
>t.test(A,B)
```

برای این حالت نیز از داده های حالت قبل استفاده می کنیم. تنها با این تفاوت که در اینجا برابری واریانس را قرار نمی دهیم. گزارش نرم افزار به صورت زیر خواهد بود.



هنگامی که آزمون t برای دو نمونه با احتمال عدم برابری واریانس ها استفاده می شود آماره ای ترکیبی به نام welch از آماره ی t ساخته می شود که آزمون مبتنی بر آن خواهد بود. با توجه به کمتر بود مقدار p (مقدار معنی داری) از سطح ۹۵٪ در نتیجه می توان گفت این آزمون فرض صفر برابری میانگین ها را رد کرده است.

لازم به ذکر است که مقدار آماره ی t و مقدار p در این حالت (عدم برابری واریانس ها) با حالت برابری واریانس ها کمی با یکدیگر متفاوت است. زمانی که ما واریانس ها را برابر فرض می کنیم، مدل سازی توزیع آزمون آماری یک توزیع t درجه کمتری از آزادی دارد. علاوه بر این حوزه کمتری در دنباله وجود دارد و بنابراین مقدار p کمتر است. چنانچه مقادیر درجه ی آزادی را در این دو حالت مشاهده نمایید در حالت اول درجه آزادی بیشتر از حالت دوم است.

آزمون t برای نمونه های جفتی

آزمون های t جفت شده یا همتا از مدل آماری مختلف استفاده می کنند. به جای اینکه فرض کنیم که دو نمونه، نمونه های مستقل نرمال هستند اگر چه شاید با میانگین و انحراف مهیار متفاوت باشند، آزمایش نمونه های همتا فرض می کند که دو نمونه در دنباله های مشترک سهیم هستند.

مدل اساسی این است $Y_i = x_i + \varepsilon_i$ که ε_i تصادفی است.

ما می خواهیم آزمایش کنیم که آیا ε_i میانگین صفر دارند و یا نه؟ به منظور انجام دادن آن هر کس می تواند x_i را از Y_i تفریق کند و سپس یک آزمون t تک نمونه منظم انجام دهد.

در واقع R همه این کار را انجام می دهد شما تنها به مشخص کردن تابع parid= TRUE نیاز دارید که تابع t.test را می طلبد. برای انجام آزمون مربوطه به مثال زیر توجه نمایید.

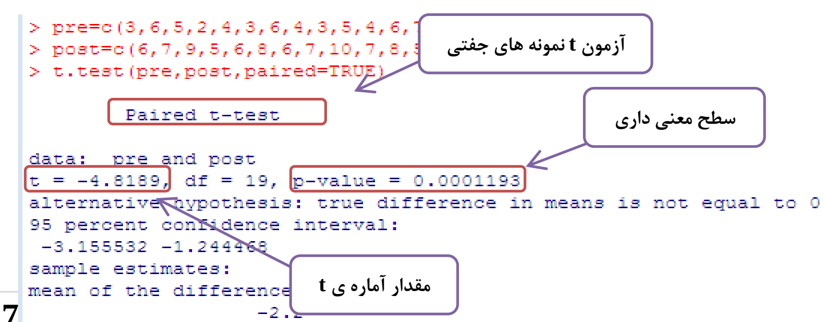
مثال: بیست بیمار در نظر گرفته شده است. فاکتور سلامتی افراد قبل و بعد از تست دارویی خاص، اندازه گیری شده است. می خواهیم آزمون کنیم آیا دارو اثر قابل توجهی در فاکتور سلامتی افراد داشته است و یا خیر؟ فاکتورهای سلامتی از ۰ تا ۱۰ عددگذاری شده اند. فرض میشود داده ها دارای توزیع نرمال هستند.

قبل از مصرف دارو:۹,۶,۸,۶,۳,۴,۵,۷,۶,۴,۵,۳,۴,۶,۳,۴,۲,۵,۶,۳

بعد از مصرف دارو:۹,۸,۹,۷,۴,۷,۹,۶,۵,۸,۷,۱۰,۷,۶,۸,۶,۵,۹,۷,۶

عبارت های زیر را در نرم افزار بنویسید.

```
>pre=c(3,6,5,2,4,3,6,4,3,5,4,6,7,5,4,3,6,8,6,9)
>post=c(6,7,9,5,6,8,6,7,10,7,8,5,6,9,7,4,7,9,8,9)
>t.test(pre,post,paired=TRUE)
```



همان طور که مشاهده می شود آزمون گزارش شده، آزمون t برای داده های جفتی است. مقدار آماره t مقدار p محاسبه شده است. با توجه به مقدار p آزمون مورد نظر رد می شود. رد فرض صفر در این آزمون به این معنا است که تفاوت معناداری بین میانگین شاخص سلامت قبل و بعد از مصرف دارو وجود دارد. و مصرف دارو در تغییر شاخص سلامت موثر بوده است. فرض صفر در اینجا برابری میانگین در حالت قبل و بعد از مصرف دارد است.

$$\begin{cases} H_0: \mu_{pre} = \mu_{post} \\ H_1: \mu_{pre} \neq \mu_{post} \end{cases}$$

لازم به یادآوری است که این نوع نمونه داده ها از هم مستقل نیستند. چنانچه در این نوع نمونه ها دقت لازم در تشخیص نوع نمونه نباشد ممکن است نتایج بدست آمده متفاوت و در نتیجه تحلیل اشتباهی بدست آید. در نتیجه باید در تشخیص جامعه ای با دو نمونه ی مستقل و نمونه های جفتی دقت لازم را داشت.

منبع:

- مبانی برنامه نویسی آماری با R، تالیف: جان براون، دانکن مرداک، ترجمه: دکتر منوچهر بابانژاد،مبین ملکشاه
- آشنایی، تحلیل و برنامه نویسی با نرم افزار R، موسی گلعلی زاده، حمیدرضا مسافری،دانشگاه تربیت مدرس، شهریور ۹۴
- موسسه فرهنگی دیجیتال بهکامان www.behkaman.ir
- http://www.irantahgig.ir

- R: A Language and Environment for Statistical Computing, Reference Index, The R Development Core Team, Version 2.4.0 (2006-10-03)