

آزمون کراسکال-والیس

(S-PLUS)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

آزمون کراسکال-والیس (Kruskal-wallis) آزمونی ناپارامتریک می‌باشد که هم‌ارز آزمون پارامتریک تجزیه و تحلیل واریانس (ANOVA) است. به بیان بهتر، در این آزمون به بررسی تفاوت بین مقادیر یک متغیر در بین چند جامعه مستقل پرداخته می‌شود. برای نمونه، محققى درصد است تفاوت بین سطح تحصیلات و رضایت شغلی کارکنان را بسنجد. وی برای این که بتواند از آزمون کراسکال-والیس استفاده کند، رضایت شغلی را به گونه‌ای سنجیده است که نمرات پیوسته از نمره ۰ تا ۱۰۰ را در مورد این متغیر در اختیار دارد. آزمون کراسکال-والیس، گونه‌ای بسط یافته از آزمون من-ویتنی است و باید تعداد سطوح زیاد و متغیری پیوسته وجود داشته باشد. در ادامه چگونگی انجام آزمون کراسکال والیس در نرم افزار s-plus بیان شده است.

مطالبی که در این متن آورده شده عبارتند از:

معرفی آزمون کراسکال-والیس

آزمون کراسکال-والیس در نرم افزار S-PLUS

مثال ۱: آزمون کراسکال-والیس از روش فراخوانی داده

مثال ۲: آزمون کراسکال-والیس از روش ورود داده

مثال ۳: آزمون کراسکال-والیس از روش برنامه نویسی

معرفی آزمون کراسکال-والیس

آزمون رتبه ای کراسکال-والیس یک جایگزین ناپارامتریک در مقابل تحلیل آنالیز واریانس یک طرفه است. فرض صفر در اینجا آن است که میانگین X (پارامتر) در تمامی گروه ها یکسان است و فرض مقابل آن است که میانگین X حداقل در یکی از گروه ها متفاوت است. برخلاف تحلیل واریانس یک طرفه در این روش نیازی به نرمال بودن داده ها نیست. چرا که آزمون کراسکال-والیس از جمله آزمون های ناپارامتریک است. در صورتیکه پیش فرض های لازم برای انجام آنالیز واریانس یک طرفه برقرار نباشد، ناچاریم از معادل های ناپارامتریک برای مقایسه چند جامعه بهره بگیریم.

در این آزمون می خواهیم فرض مبنی بر برابری میانگین ها در برابر فرض مقابل مبنی بر نابرابری میانگین های $\mu_1, \mu_2, \dots, \mu_k$ آزمون کنیم.

آزمون کراسکال والیس به شرح زیر است:

یک نمونه n_i از هر i ($i=1,2,\dots,k$) در نظر میگیریم. این k نمونه را با هم ترکیب می کنیم تا یک نمونه $n = n_1, n_2, \dots, n_k$ تایی بدست آید و آن را با اعداد درست $1,2,\dots,n$ رتبه بندی می کنیم. فرض کنید R معدل تمام رتبه ها R_i برای نمونه i -ام در این رتبه بندی باشد. بر این اساس آماره زیر را می سازیم:

$$V = \sum_{i=1}^n n_i \left(\bar{R}_i - \frac{n+1}{2} \right)^2 = \sum_{i=1}^n n_i (\bar{R}_i - \bar{R})^2$$

اگر H_0 درست باشد، انتظار داریم که \bar{R}_i ها تقریباً برابر باشند، پس V را در سطح معین، مثلاً $\alpha = 0.05$ می توانیم بسنجیم. در نتیجه به توزیع احتمال v نیاز داریم. کراسکال-والیس با محاسباتی نه چندان پیچیده نشان داده اند که آماره:

$$\frac{12v}{n(n+1)}$$

تحت فرض H_0 تقریباً دارای توزیع $\chi^2_{(k-1)}$ است و پس از محاسباتی می توان آن را بر حسب $S_i = n_i R_i$ ها به صورت زیر به نام آماره ی کراسکال-والیس در آورد.

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{S_i^2}{n_i} - 3(n+1)$$

در این آماره n_i اندازه نمونه i -ام، k تعداد نمونه ها، $n = \sum_{i=1}^n n_i$ اندازه نمونه ترکیبی و S_i مجموع رتبه ها برای نمونه i -ام در نمونه ترکیبی می باشد. اگر $H > \chi^2_{(k-1, 1-\alpha)}$ فرض H_0 را با در سطح α رد می کنیم.

در این آزمون تقریبی باید:

۱. داده ها نمونه های مستقل، از جامعه هایی باشند که شکل آنها یکسان است.

۲. اندازه نمونه ها تا حد امکان با هم برابر باشند.

۳. اندازه هر نمونه حداقل ۵ باشد.

۴. داده ها به صورت تصادفی از جامعه استخراج شده باشد.

نکته: در صورتی که $k=2$ باشد، یعنی داشتن ۲ نمونه، آزمون کراسکال والیس معادل آزمون من ویتنی می باشد.

آزمون کراسکال والیس در نرم افزار S-PLUS

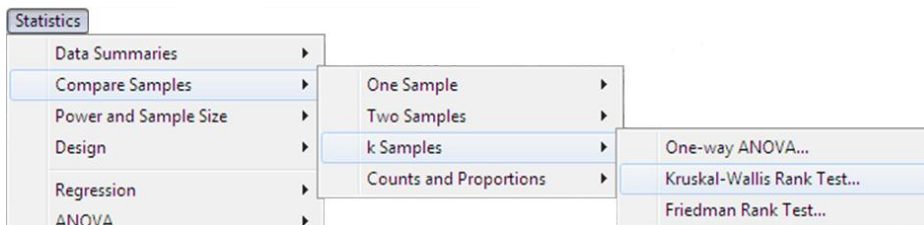
برای انجام آزمون مربوطه و درک بهتر آزمون دو مثال بیان می کنیم.

مثال ۱: داده های mpip را فراخوانی می کنیم. (برای آموزش چگونگی فراخوانی داده به فایل "[معرفی نرم افزار و ورود داده ها](#)" در صفحه ی آموزش S-PLUS پایگاه اینترنتی مرکز تحلیل آماری خوارزمی مراجعه نمایید).

	1	2	3	4	5	6
	ved	angina	education	prior.mi	nyha	rales
1	726.65	2.00	6.00	0.00	1.00	0.1
2	0.08	0.00	5.00	1.00	1.00	2.1
3	4.29	1.00	2.00	1.00	4.00	2.1
4	0.00	0.00	5.00	0.00	1.00	3.1
5	0.00	0.00	7.00	0.00	3.00	0.1
6	3.05	0.00	1.00	0.00	1.00	0.1
7	324.33	0.00	7.00	1.00	2.00	1.1
8	0.13	1.00	1.00	0.00	2.00	0.1
9	0.17	2.00	4.00	1.00	2.00	0.1
10	3.85	0.00	8.00	0.00	1.00	1.1
11	0.38	0.00	1.00	0.00	1.00	1.1

چنانچه داده های مربوطه را باز کردید برای انجام آزمون کراسکال-والیس مسیر زیر را طی نمایید.

Statistics > Compare Sample > K Samples > Kruskal-Wallis Rank Test...



پس از آن پنجره ی زیر باز می شود.

مجموعه داده ای که می خواهیم بررسی کنیم و یا صفحه ای که داده ها را در آن وارد کرده اید را انتخاب می کنید.

متغیر مورد نظر برای بررسی را در این قسمت قرار می دهیم. در اینجا متغیر **ved** انتخاب شده است.

متغیر گروه بندی شده را در این قسمت قرار می دهیم در اینجا متغیر **Angina2** به عنوان متغیر گروه بندی انتخاب شده است.

Kruskal-Wallis Rank Sum Test

Data Set: mpip

Variable: ved

Grouping Variable: angina2

Print Results:

OK Cancel Apply k > current Help

برای چاپ نتایج لازم است این گزینه را تیک دار کنید.

پس از انتخاب این مراحل دکمه ok را زده تا نتایج و گزارش نرم افزار نمایش داده شود. چنانچه مراحل به درستی انجام شود نتایجی مشابه تصویر پایین نمایش داده می شود.

```

Report1
Kruskal-Wallis rank sum test

data: ved and angina2 from data set mpip
Kruskal-Wallis chi-square = 1.0874, df = 1, p-value = 0.297
alternative hypothesis: two.sided
  
```

همانطور که مشاهده می شود در گزارش ارائه شده مقدار $p\text{-value} = 0.297$ است که در سطح معنی داری ۰.۰۵ از مقدار ۰.۰۵ بیشتر است. در نتیجه دلیلی بر رد فرض صفر وجود ندارد. در نتیجه می توان

گفت می پذیریم در سطح ۰.۰۵ متوسط متغیرهای ved به طور متوسط برای تمامی طبقات engine2 یکسان است.

مثال ۲: در این مثال نمره های آزمون ورودی دانشجویان سه دانشگاه A,B,C داده شده است. می خواهیم با $\alpha = 0.05$ بیازماییم که نمرات دانشجویان هر سه دانشگاه به طور متوسط یکسان است یا نه؟

A:	۱۹	۱۱/۷	۱۷/۸	۱۴/۸	۱۳/۹			
B:	۱۸/۲	۱۴/۸	۱۳/۱	۱۲/۶	۱۵/۲	۱۲/۸		
C:	۱۳/۴	۱۴/۱	۱۲/۳	۱۲/۳	۱۴/۷	۱۳/۹	۱۳/۸	۱۴/۳

دانشجویان هر سه دانشگاه از نظر متوسط سطح درس به طور یکسان اند: H_0
 رد فرض H_0 : H_1

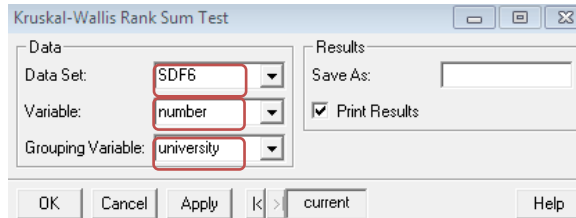
برای انجام آزمون مورد نظر داده های مربوطه را در نرم افزار S-PLUS وارد نمایید. (برای آموزش چگونگی ورود داده های جدید به فایل "[معرفی نرم افزار و ورود داده ها](#)" در صفحه ی آموزش S-PLUS پایگاه اینترنتی مرکز تحلیل آماری خوارزمی مراجعه نمایید.)

پس از وارد کردن داده های مورد نظر مسیر زیر را طی نمایید.

Statistics > Compare Sample > K Samples > Kruskal-Wallis Rank Test...

مطابق پنجره ی پایین کادرها را پر می کنیم.

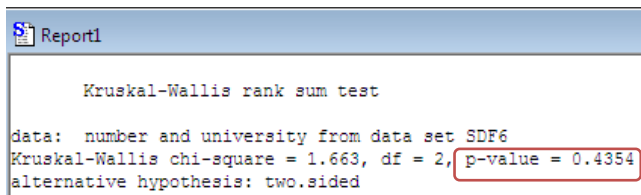
	1	2
	university	number
1	1.00	19.00
2	1.00	11.70
3	1.00	17.80
4	1.00	14.80
5	1.00	13.90
6	2.00	18.20
7	2.00	14.80
8	2.00	13.10
9	2.00	12.60
10	2.00	15.20
11	2.00	12.80
12	3.00	13.40
13	3.00	14.10
14	3.00	12.30
15	3.00	12.30
16	3.00	14.70
17	3.00	13.90
18	3.00	13.80
19	3.00	14.30



پس از زدن دکمه ی OK خروجی پایین گزارش داده می شود.

همانطور که در خروجی ها مشاهده می کنید $p\text{-value} = 0.4354$ از $\alpha = 0.05$

بزرگتر است بنابراین دلیلی بر رد فرض H_0 نداریم یعنی با احتمال ۹۵٪ می پذیریم دانشجویان ۳ دانشگاه به طور متوسط از نظر سطح درسی یکسان اند.



یکی دیگر از روش ها برای انجام آزمون کراسکال والیس در نرم افزار S-PLUS استفاده از دستورات است. برای درک بهتر به مثال زیر توجه نمایید.

مثال ۳: داده های زیر مقدار نیکوتین ۳ نوع سیگار پرمصرف در یک منطقه مربوط به ۲۰ نمونه تصادفی است. هدف مقایسه میانگین نیکوتین این ۳ نوع سیگار است.

مقدار نیکوتین								نوع سیگار
		۶۷	۶۶.۵	۶۵.۵	۶۶	۶۵.۵	۶۶	نوع اول
۶۵	۶۴	۶۴.۵	۶۴.۵	۶۵	۶۴	۶۵	۶۴.۵	نوع دوم
		۶۵.۵	۶۴.۵	۶۶	۶۴.۵	۶۵	۶۵.۵	نوع سوم

ابتدا داده های زیر را وارد S-PLUS می کنیم. داده باید در قالب ۲ بردار وارد شوند. بردار اول باید حاوی مقادیر وابسته و بردار دوم شامل مقادیر متغیر گروه بندی باشند.

```

> x<-c(66,65.5,66,65.5,66.5,67,
+ 64.5,65,64,65,64.5,64.5,64,65,
+ 65.5,65,64.5,66,64.5,65.5)
> a<-c(rep(1,6),rep(2,8),rep(3,6))
> kruskal.test(x,a)

Kruskal-Wallis rank sum test

data: x and a
Kruskal-Wallis chi-square = 11.9468, df = 2, p-value = 0.0025
alternative hypothesis: two.sided

```

نحوه ی ورود داده

نحوه ی تعریف کردن گروه بندی داده ها

فرمال کراسکال-والیس

داده ها را به صورت روبرو وارد کرده و پس از آن فرمان آزمون کراسکال والیس می نویسیم.

همانطور که در خروجی ها مشاهده می

کنید $p\text{-value} = 0.0025$ که از $\alpha = 0.05$ کوچکتر است بنابراین فرض H_0 رد می شود یعنی با احتمال ۹۵٪ میانگین نیکوتین این ۳ نوع سیگار تفاوت معنی داری با هم دارند و یکسان نیستند.

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
- محاسبات آماری با s-plus، انور محمدی، انتشارات گسترش علوم پایه، ۱۳۹۱
- <http://spss-amar.vcp.ir>

- S-PLUS 8 for Windows User's Guide, May 2007, Insightful Corporation, Seattle Washington, 1700 Westlake Avenue N, Suite 500, WA 98109-3044,USA