

فصل 4: تحلیل آماری

تهیه و تنظیم:
دکتر روح اله تولایی
(عضو هیأت علمی دانشگاه شهید بهشتی)

نسخه شماره یک، پاییز 1393

هدف از این فصل:

- ✓ این فصل به بحث در باره شیوه های عمومی تحلیل داده ها و روشهای آماری می پردازد و بر **زمینه های مورد استفاده** این روشها و مفروضه های هر یک از آنها تمرکز دارد.
- ✓ در این فصل باید بتوانید روش های آماری **مناسب** را برای تحلیل داده های تحقیق خود انتخاب کنید و **تفسیر** یافته های حاصل از تحلیل را به دقت به عمل آورید.

آماده کردن داده ها برای تحلیل آماری:

جزوه آموزشی
روش تحقیق پیشرفته

نخستین اقدامی که بایستی پس از گردآوری داده ها انجام دهید، **بررسی و پردازش** داده های گردآوری شده است تا بتوان تحلیل

آماری را درباره آنها به کار برد (Delwiche and Slaughter, 2008).

مهمترین اقدامات در این مرحله عبارتند از:

الف) مرتب کردن و تنظیم داده ها

ب) کدگذاری داده ها

ج) سازماندهی داده ها

آماده کردن داده ها برای تحلیل آماری_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

الف) مرتب کردن و تنظیم داده ها

برای این کار بایستی اقدامات زیر صورت پذیرد:

1- داده هایی که توسط خود شرکت کنندگان یا محقق ثبت شده مجدداً بررسی شود تا حد امکان اشتباه موجود را تشخیص و اصلاح نماییم. مانند سن 223 = اصلاح یا حذف داده (داده بدون جواب در تحلیل آماری)

2- همه داده ها را در قالب و فرمت یکسان در بیاورید. مانند سن = 22، بیست و دو، یک ماه دیگر 22، که همگی باید یکسان شوند

آماده کردن داده ها برای تحلیل آماری_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ب) کدگذاری داده ها

- برای تحلیل داده ها به وسیله نرم افزار آماری، بایستی آنها را کدگذاری و به صورت داده های عددی در بیاورید.
- برای کدگذاری لازم است که از ثبات آن اطمینان حاصل شود، مخصوصاً وقتی این کار توسط بیش از یک نفر انجام میشود.
- روشهای آماری چندی، مانند روش آلفای کرونباخ به منظور سنجش قابلیت اعتماد (پایایی یا ثبات) کدگذاری به کار میرود.
- مانند کدگذاری اطلاعات جمعیت شناختی یا کدگذاری طیف لیکرت.

آماده کردن داده ها برای تحلیل آماری_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ب) کدگذاری داده ها_ ادامه

مثال: کدگذاری اطلاعات جمعیت شناختی یا کدگذاری طیف لیکرت.

	Age	Gender	Highest degree	Previous experience in software A
Participant 1	34	male	College	Yes
Participant 2	28	female	Graduate	No
Participant 3	21	female	High school	No

Table 4.1 Sample demographic data in its original form.

	Age	Gender	Highest degree	Previous experience in software A
Participant 1	34	1	2	1
Participant 2	28	0	3	0
Participant 3	21	0	1	0

Table 4.2 Sample demographic data in coded form.

آماده کردن داده ها برای تحلیل آماری_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ج) سازماندهی داده ها

- برای استفاده از نرم افزارهای آماری، لازم است که داده ها قبلاً به صورت خاصی وارد کامپیوتر شوند تا تحلیل آماری معینی بر روی آنها انجام شود.
- مثلاً برای انجام آزمون t برای دو نمونه مستقل با نرم افزار SPSS (برای مقایسه میانگین های دو گروه داده ها)، لازم است داده های دو گروه در یک ستون وارد شوند. اما برای انجام آزمون t برای دو نمونه وابسته برعکس باید داده ها را در دو ستون مجزا وارد کرد. همینطور برای تحلیل واریانس (ANOVA)، همبستگی و اندازه گیری مکرر.

روش های آمار توصیفی:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ بعد از اینکه داده ها گردآوری و مورد تنظیم قرار گرفتند، لازم است تا با آزمون های آمار توصیفی، به ماهیت مجموعه داده ها پی ببرید.
- ✓ برخی شاخصهای توصیفی مرسوم برای این کار (در قالب گرایش به مرکز، پراکندگی و منحنی توزیع نرمال) عبارتند از: فراوانی، میانگین، میانه، نما و واریانس، انحراف معیار و دامنه تغییرات.

روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

الف) شاخص های گرایش به مرکز (Central Tendency)

مهمترین شاخصها برای توصیف گرایش به مرکز یک مجموعه داده

ها عبارتند از (Rosenthal and Rosnow, 2008):

1- **میانگین**: متوسط حسابی از یک مجموعه داده ها می باشد.

2- **میانه**: عددی است که در وسط داده ها قرار دارد.

برای مثال داده های زیر تعداد کلمات تایپ شده هفت کاربر در یک دقیقه است:

{15, 19, 22, 29, 33, 45, 50}

حال میانه این مجموعه داده ها، 29 می باشد.

روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

الف) شاخص های گرایش به مرکز_ ادامه

3- **نما**: مقداری است که بیشترین تکرار را در مجموعه داده ها

دارا میباشد.

برای مثال داده های زیر تعداد ساعاتی است هفت کاربر در هفته با

اینترنت کار میکنند:

{12, 15, 22, 22, 22, 34, 34}

حال نما در این مجموعه داده ها، 22 می باشد.

روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ب) شاخص های پراکندگی (Measures of Spread)

به وسیله این شاخص می خواهیم بدانیم تا چه اندازه داده ها در اطراف نقطه تمرکز پراکنده اند؛ که مهمترین آنها عبارتند از:

1- دامنه تغییرات (Range):

فاصله میان بزرگترین و کوچکترین مقدار در مجموعه داده ها را اندازه گیری می کند. هرچه دامنه طولانی تر باشد، مجموعه داده ها گسترده تر است.

برای مثال داده های زیر تعداد کلمات تایپ شده هفت کاربر در یک دقیقه است:

{15, 19, 22, 29, 33, 45, 50}

حال دامنه این مجموعه داده ها، $50 - 15 = 35$ می باشد.

روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ب) شاخص های پراکندگی_ ادامه

2- واریانس (Variance):

میانگین مجذور تفاوت (انحراف) میان هر یک از مقادیر داده ها با میانگین آنها می باشد.

3- انحراف معیار (Standard Deviation):

جذر واریانس می باشد.

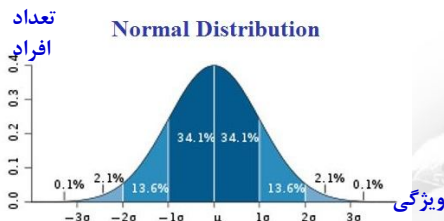
✓ هر چه واریانس و انحراف معیار مجموعه داده ها بزرگتر باشد، پراکندگی داده ها بیشتر است.

روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ج) منحنی توزیع نرمال (بهنجار)

- روشی که معمولاً برای توصیف توزیع فراوانی یک مجموعه داده ها بکار میرود، توزیع نرمال (بهنجار) است.
- منحنی توزیع نرمال با استفاده از میانگین و انحراف معیار داده ها به دست می آید.



روش های آمار توصیفی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

ج) منحنی توزیع نرمال_ ادامه

- آزمودن اینکه یک مجموعه داده ها به صورت نرمال توزیع شده یا نه، اقدام مهمی در انتخاب نوع آزمون معنی داری آماری است.
- در آزمونهای پارامتریک، فرض بر آن است که داده ها به صورت نرمال یا تقریباً نرمال توزیع شده اند.
- اگر مجموعه داده ها دارای توزیع نرمال نباشد، یا باید داده ها را تغییر دهیم تا به صورت نرمال درآیند، یا باید برای تحلیل آنها از آزمونهای ناپارامتریک (non-parametric tests) استفاده نماییم.

آزمونهای معنی داری آماری:

جزوه آموزشی
روش تحقیق پیشرفته

- به وسیله آزمونهای معنی داری، می توان **میانگین های چند گروه** را با یکدیگر مقایسه کرد.
- مهمترین آزمونهای معنی داری شامل دو مورد زیر می باشد:
 - (1) **آزمون t:** حالت ساده شده ی تحلیل واریانس است و فقط برای **دو گروه** یا دو موقعیت آزمایشی استفاده می شود.
 - (2) **تحلیل واریانس (ANOVA):** وقتی یک مطالعه با **بیش از دو گروه** یا **دو موقعیت آزمایشی** سر و کار دارد باید از این تحلیل استفاده کند.

(1) آزمون های t:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ یکی از روش های معتبر آماری است که استفاده وسیعتری برای **مقایسه دو میانگین** دارد. معمولاً 2 نوع آزمون t متناسب با طرح آزمایشی انتخاب می شود که عبارتند از:
 - 1- **آزمون t برای نمونه های مستقل:** برای **دو گروه مستقل** که مورد مقایسه قرار می گیرند استفاده می شود.
 - 2- **آزمون t برای نمونه های وابسته (همتا شده):** برای **دو میانگین** از داده های مربوط به **یک گروه** (یا دو گروه به هم وابسته) استفاده می شود.

1) آزمون های t_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

✓ در انجام آزمونهای t یک مقدار t بدست می آید، که هرچه مقدار آن بزرگتر باشد نشان دهنده آن است که **فرضیه صفر (تفاوت میان دو میانگین) رد می شود (فرضیه بدیل تایید می شود)**.

✓ به عبارت دیگر هر مقدار t بدست آمده از مقدار t مربوط به 0.95 بزرگتر باشد (با آلفای 0.05 و درجه آزادی 15)، نشان می دهد احتمال اینکه **دو میانگین متفاوت باشند بیشتر است**.

تمرین:

جزوه آموزشی
روش تحقیق پیشرفته

فرضیه های صفر و بدیل را برای سوال زیر در دو حالت:
1- آزمون t برای نمونه های مستقل و 2- آزمون t برای نمونه های وابسته (همتا شده)، بنویسید.

«آیا نرم افزار پردازش واژگان W-P (Word-Prediction)

می تواند مدت زمان تایپ کردن کاربران خود را نسبت به

نرم افزار پردازش واژگان معمولی (Standard Word-Processing)

تغییر دهد؟»

(سوال دو دامنه - اگر بگوییم کاهش دهد؟ می شود آزمون t یک دامنه)

تمرین_ ادامه:

✓ نمونه فرضیه های قابل طرح در آزمون t برای نمونه های مستقل:

H_0 : میان میانگین های گروه کاربران نرم افزار پردازش واژگان W-P و گروه کاربران نرم افزار پردازش واژگان معمولی تفاوت وجود ندارد.

H_1 : میان میانگین های گروه نرم افزار پردازش واژگان W-P و گروه نرم افزار پردازش واژگان معمولی تفاوت وجود دارد.

✓ نمونه فرضیه های قابل طرح در آزمون t برای نمونه های وابسته (همتا شده):

H_0 : میان زمان تایپ کردن با نرم افزار پردازش واژگان W-P و نرم افزار پردازش واژگان معمولی تفاوت وجود ندارد. (برای یک گروه کاربر مورد آزمایش در دو موقعیت)

H_1 : میان زمان تایپ کردن با نرم افزار پردازش واژگان W-P و نرم افزار پردازش واژگان معمولی تفاوت وجود دارد.

تمرین_ ادامه:

Group	Participants	Task completion time	Coding
No prediction	Participant 1	245	0
No prediction	Participant 2	236	0
No prediction	Participant 3	321	0
No prediction	Participant 4	212	0
No prediction	Participant 5	267	0
No prediction	Participant 6	334	0
No prediction	Participant 7	287	0
No prediction	Participant 8	259	0
With prediction	Participant 1	246	1
With prediction	Participant 2	213	1
With prediction	Participant 3	265	1
With prediction	Participant 4	189	1
With prediction	Participant 5	201	1
With prediction	Participant 6	197	1
With prediction	Participant 7	289	1
With prediction	Participant 8	224	1

Table 4.4 Sample data for independent-samples t test.

Participants	No prediction	With prediction
Participant 1	245	246
Participant 2	236	213
Participant 3	321	265
Participant 4	212	189
Participant 5	267	201
Participant 6	334	197
Participant 7	287	289
Participant 8	259	224

Table 4.5 Sample data for paired-samples t test.

2) تحلیل واریانس (ANOVA):

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ یک روش آماری است که برای مقایسه میانگین دو یا چند گروه بکار میرود.
- ✓ زمانیکه تنها میانگین دو گروه مورد مقایسه قرار میگیرد، تحلیل واریانس به آزمون t خلاصه میشود.
- ✓ در آزمونهای تحلیل واریانس، مقداری به نام F محاسبه میشود، لذا به آنها «آزمون F» نیز میگویند. (مربوط به مقدار حاصل از توزیع فراوانی F است که جدول استاندارد آن بر حسب دو درجه آزادی محاسبه میشود)
- ✓ انواع اصلی تحلیل واریانس عبارتند از: الف) تحلیل واریانس یک طرفه، ب) تحلیل واریانس عاملی.

تحلیل واریانس_ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

الف) تحلیل واریانس یک طرفه (One-way ANOVA):

- ✓ برای تحقیقات میان-گروهی با یک متغیر مستقل که دارای سه یا چند موقعیت آزمایشی است، استفاده می شود.
- ✓ مثال: متغیر مستقل: نرم افزارهای پردازش واژگان برای ورود داده ها به کامپیوتر (متغیر مستقل دارای سه حالت است)
- ✓ متغیر وابسته: زمان صرف شده برای انجام وظیفه
- ✓ موقعیت های آزمایشی: نرم افزار پردازش واژگان معمولی (گروه گواه)، نرم افزار W-P، و نرم افزار بیان صوتی.

تحلیل واریانس_ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

الف) تحلیل واریانس عاملی (Factorial ANOVA):

- ✓ برای تحقیقات میان-گروهی با دو یا چند متغیر مستقل که دارای سه یا چند موقعیت آزمایشی است، استفاده می شود.
- ✓ مثال: متغیرهای مستقل: نوع متن، و نرم افزارهای پردازش واژگان برای ورود داده ها به کامپیوتر
- ✓ متغیر وابسته: زمان صرف شده برای انجام وظیفه
- ✓ موقعیت های آزمایشی: (کپی برداری یا تالیفی) * (نرم افزار پردازش واژگان معمولی، نرم افزار w-p، یا نرم افزار بیان صوتی) = 6 موقعیت

تحلیل واریانس_ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

□ تحلیل واریانس اندازه گیری های مکرر:

- ✓ در طرح های آزمایشی خصوصاً فناوری اطلاعات، یافتن نمونه های آماری بزرگ دشوار است. (مثلاً 12 نفر * 6 موقعیت)
- ✓ برای حل این مشکل، در این روش، از طرح درون-گروهی استفاده میشود. یعنی فقط یک گروه آزمایشی انتخاب شده و از هر شرکت کننده میخواهیم که در موقعیت های آزمایشی مختلف ما قرار گیرد.
- ✓ تحلیل واریانس اندازه گیری مکرر، میتواند یک سطح (برای واریانس یک طرفه و بررسی تاثیر یک متغیر مستقل)، یا چند سطح (برای واریانس عاملی و بررسی تاثیر چند متغیر مستقل) داشته باشد.

تحلیل همبستگی (شناسایی رابطه):

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ یکی از هدفهای مشترک تحقیقات مدیریت فناوری اطلاعات، تحلیل همبستگی و شناسایی رابطه **میان دو متغیر** مستقل و وابسته (نه لزوماً علی معلولی) است.
- ✓ مهمترین روش آماری برای تحلیل همبستگی، ضریب همبستگی گشتاوری پیرسون (ضریب ۲ پیرسون) و **آزمون پیرسون** است.
- ✓ مقدار ضریب ۲ پیرسون **بین +1 تا -1 تغییر می کند**. هرچه به +1 نزدیکتر میشود نشاندهنده رابطه خطی کاملاً مثبت میان دو متغیر و بالعکس است. یعنی با افزایش متغیر مستقل، متغیر وابسته نیز افزایش می یابد و بالعکس.
- ✓ اگر ضریب ۲ پیرسون **نزدیک به صفر** باشد، نشاندهنده عدم رابطه خطی میان دو متغیر است. یعنی افزایش یا کاهش متغیر مستقل، رابطه ای با مقدار متغیر وابسته ندارد.

تحلیل همبستگی_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ **مثال:** سوال تحقیق: آیا میان تجربه کار با کامپیوتر کاربران و زمان صرف شده آنها برای وارد کردن داده ها با نرم افزار واژه پرداز معمولی ارتباطی وجود دارد؟
- ✓ **متغیر مستقل:** تجربه کار با کامپیوتر. **متغیر وابسته:** زمان صرف شده برای وارد کردن داده ها با نرم افزار واژه پرداز معمولی.
- ✓ اگر در تحلیل همبستگی میان تجربه کار با کامپیوتر و زمان صرف شده با نرم افزار معمولی، مقدار ضریب ۲ پیرسون برابر عدد **-0.723** شود (در سطح معنی داری $p < 0.05$)، به این معناست که میان این دو متغیر **همبستگی وجود دارد**، و مقدار **منفی** بدست آمده به این معناست که هرچه مقدار متغیر مستقل افزایش یابد، مقدار متغیر وابسته کاهش می یابد.

تحلیل رگرسیون (Regression):

جزوه آموزشی
روش تحقیق پیشرفته

✓ با تحلیل رگرسیون می توان رابطه میان **یک** متغیر وابسته و **چندین** متغیر مستقل را تعیین کرد.

✓ تحلیل رگرسیون با **دو هدف اصلی** انجام می شود:

1- **ساختن الگو (مدل):** برای برقراری رابطه کمی میان یک متغیر وابسته و چندین متغیر مستقل انجام می شود.

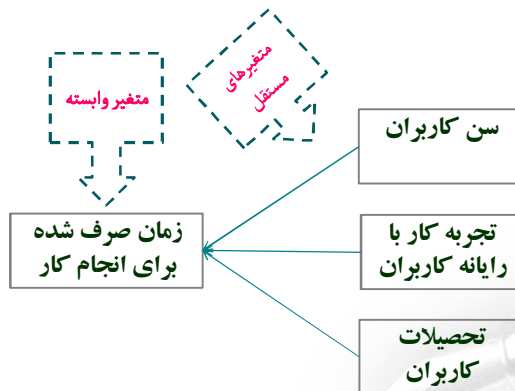
2- **پیش بینی:** در این حالت میخواهیم تعدادی از عوامل شناخته شده را که «متغیرهای پیش بین» نامیده میشوند، توسط متغیر وابسته که «متغیر ملاک» نامیده میشود، پیش بینی شود.

✓ این دو هدف کاملاً **به یکدیگر مرتبطند**. لذا باید مدل مناسبی را تدوین کرد که مقادیر متغیر ملاکی را که مد نظر شماست به خوبی پیش بینی کند.

تحلیل رگرسیون_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

✓ **مثال:** چه عواملی مرتبط با کاربران، بر میزان زمان صرف شده برای انجام کار آنها با سیستم اطلاعاتی مدیریت منابع اثر گذار است؟



آزمونهای آماری ناپارامتریک:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ تمام روشهای تحلیل آماری که تاکنون بیان شد، مستلزم **مفروضه های آزمون آماری پارامتریک** بودند که مهمترین آنها عبارتند از:
 - 1- داده ها باید از جامعه ای که دارای توزیع نرمال است گردآوری شده باشند (توزیع داده ها تقریباً به صورت نرمال باشد).
 - 2- اندازه گیری متغیر مورد مطالعه باید حداقل با استفاده از مقیاس فاصله ای انجام شده باشد (فاصله میان هر دو مقدار در جامعه برابر باشد مانند فاصله 1 و 2 با 5 و 6).
 - 3- برای آزمونهایی که مقایسه میان میانگین گروه های مختلف را نشان میدهند، واریانس داده های حاصل از هر گروه باید تقریباً مساوی باشد.
- ✓ در تحقیقاتی که یک یا چند مفروضه فوق وجود نداشته باشد، باید از **روشهای آماری ناپارامتریک** (non-parametric test) استفاده نمود.

آزمونهای آماری ناپارامتریک_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ برخی موارد استفاده از آزمون ناپارامتریک: جامعه آماری تحقیق دارای توزیع نرمال نباشد. یا متغیرهای تحقیق با استفاده از مقایسه های غیر فاصله ای مانند مقیاس لیکرت یا بلی و خیر، سنجیده شوند.
- ✓ البته آزمونهای آماری ناپارامتریک مانند آزمون مجذور کای (chi square) «آزاد از مفروضه» نیستند و برای اجرای آنها باید **اولاً** داده ها از یکدیگر مستقل باشند (تنها در یک گروه قرار گیرند) و **ثانیاً** حجم نمونه مناسب باشد (حداقل 30 نمونه).

آزمونهای آماری ناپارامتریک_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ **مثال: سوال:** آیا سن کاربران بر ترجیح آنان برای استفاده از **موس** یا **صفحه لمسی** در استفاده از رایانه تاثیری دارد؟
- ✓ **آزمودنی ها:** دو گروه کاربر شامل یک گروه 20 نفر بزرگسال جوان تر از 60 سال و گروه دیگر 20 نفر 60 سال و مسن تر.
- ✓ پس از انجام تحقیق و پرسش درخصوص ترجیح آنان جدول فراوانی زیر بدست آمده است:

ابزار مورد ترجیح		متغیر	
صفحه لمسی	موس	جوانتر از 65	سن
14	6	60 و مسن تر	
4	16		

✓ اجرای آزمون مجذور کای = 10.1
(با درجه آزادی یک و $p < 0.05$)

نتیجه میگیریم با سطح اطمینان 95 درصد، میان سن و ابزار مورد ترجیح رابطه وجود دارد.

آزمونهای آماری ناپارامتریک_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

- ✓ برای بسیاری از آزمونهای پارامتریک، «**بدیل**» آزمون های ناپارامتریک نیز وجود دارد که مهمترین آنها عبارتند از:
- 1- اگر داده های گردآوری شده از **دو نمونه مستقل** باشند (مانند طرح میان گروهی)، چنانچه مفروضه ها صادق باشد، آزمون **t** برای نمونه های مستقل استفاده می شود. در صورتیکه که مفروضه ها صادق نباشد، آزمون **U** من-ویتنی یا آزمون گردش والد-ولفowitz (WALD-WOLFOWITZ RUNS TEST) استفاده می شود.

آزمونهای آماری ناپارامتریک_ ادامه:

جزوه آموزشی
روش تحقیق پیشرفته

2- اگر دو مجموعه داده های مربوط به **یک گروه** کاربر را مقایسه می کنید (مانند طرح درون گروهی)، با صادق بودن مفروضه ها، یک آزمون t برای نمونه های وابسته استفاده می شود. اگر مفروضه ها صادق نباشد، می توان از آزمون رتبه ای علامت ویلکاکسون استفاده کرد.

3- اگر **سه یا بیشتر مجموعه داده ها** را مقایسه میکنید و مفروضه ها صادق نباشند، ولی **نمونه ها مستقل** باشند، تحلیل واریانس یک طرفه رتبه ای کروسکال-والیس (که بسط داده شده آزمون χ^2 -ممن-ویتنی است) استفاده می شود. وقتی هم نمونه ها وابسته باشند، می توان از آزمون تحلیل واریانس دو طرفه فریدمن استفاده کرد.