

هم خطی چندگانه

هم خطی حالتی است که در آن بین دو متغیر پیش‌بین، همبستگی قوی وجود دارد. هم خطی چندگانه حالتی است که در آن بیش از دو متغیر پیش بین همبستگی قوی با یکدیگر دارند. هم خطی چندگانه می‌تواند تفسیر نتایج رگرسیون چندگانه را تحریف کند (میزر و دیگران، ۱۳۹۱: ۲۳۶).

هم خطی چندگانه زمانی اتفاق می‌افتد که متغیرهای مستقل چیز یکسانی را اندازه بگیرند. زمانی که دو یا چند متغیر مستقل همبستگی بالایی داشته باشند امکان وجود "هم خطی چندگانه" زیاد است. در ارتباط با گویه‌های یک متغیر نیز مساله هم خطی چندگانه وجود دارد. گویه‌هایی که مفهوم یکسانی را اندازه می‌گیرند باید با یکدیگر همبستگی بالایی داشته باشند، اما همبستگی‌های بزرگتر از ۰.۹ بین هر گویه می‌تواند موجب مشکلات آماری شود (تاباچینک و فیدل، ۲۰۰۱).

وقتی که هدف پژوهش تنها بیشینه کردن ضریب تعیین (R^2) نیست، بلکه فهم تأثیر متقابل متغیرهای پیش‌بین است، هم خطی چندگانه ممکن است موجب مشکلات متعددی در تحلیل شود. یکی از مشکلاتی که از وجود هم خطی چندگانه ناشی می‌شود آن است که مقادیر ضرایب رگرسیون متغیرهای مستقل که همبستگی بالایی دارند تحریف می‌شود. اغلب، این ضرایب تا حدودی پایین هستند و حتی ممکن است از نظر آماری معنی‌دار هم نباشند. دومین مشکل آن است که خطاهای استاندارد وزن‌های رگرسیون متغیرهای پیش‌بین که هم خطی چندگانه دارند ممکن است متورم^۱ شود که در نتیجه فواصل اطمینان آن‌ها بزرگ خواهد شد، تا جایی که بعضی وقت‌ها این فواصل اطمینان، صفر را نیز در بر می‌گیرد. اگر چنین باشد به یقین نمی‌توان گفت که افزایش اندازه متغیر پیش‌بین موجب افزایش یا کاهش متغیر ملاک می‌شود. مشکل سوم آن است که اگر هم خطی چندگانه به اندازه کافی زیاد باشد عملیات ریاضی درونی خاص (مانند معکوس کردن ماتریس) مختل شده و برنامه آماری با سوت کشیدن^۲ متوقف می‌شود (میزر و دیگران، ۱۳۹۱: ۲۳۷).

دلایل بروز هم خطی چندگانه

یک علت رایج هم خطی چندگانه این است که پژوهش‌گران هم خرده‌مقیاس‌های یک پرسشنامه و هم نمره کل پرسشنامه را به عنوان متغیرهای پیش‌بین به کار می‌برند. بسته به این که خرده‌مقیاس‌ها چگونه محاسبه شده باشند، این امکان وجود دارد که ترکیب آن‌ها با نمره کل پرسشنامه دارای همبستگی تقریباً کامل باشد. بنابراین باید فقط خرده‌مقیاس‌ها یا نمره کل را پرسشنامه را به کار ببریم نه این که همه آن‌ها را با هم وارد تحلیل کنیم. علت رایج دیگر هم خطی چندگانه این است که متغیرهایی در تحلیل وجود دارند که سازه مشابهی را می‌سنجند. ما باید همه این متغیرها به جز یک متغیر را از تحلیل کنار بگذاریم، یا احتمال ترکیب آن‌ها را با یک روش معقول و منطقی مدنظر قرار بدهیم.

¹ inflated

² screeching halt

تشخیص هم‌خطی چندگانه

زمانی که فقط دو متغیر بررسی می‌شوند، پیدا کردن رابطه هم‌خطی از طریق وجود همبستگی بالا آسان است. فقط کافی است که همبستگی‌های پیروسون بین متغیرهای موجود در تحلیل را به عنوان پیش‌درآمد (پیش‌شرط) آزمون‌هایی مانند رگرسیون چندگانه بررسی کنیم. وقتی رابطه بین دو متغیر بسیار قوی است پرچم قرمز برافراشته می‌شود. به عنوان یک قاعده کلی سرانگشتی پیشنهاد می‌کنیم زمانی که همبستگی بین دو متغیر 0.70 یا بالاتر است احتمالاً نباید آن‌ها را در رگرسیون یا هر تحلیل چندمتغیری دیگری با هم به کار برد. اگر همبستگی بالاتر از 0.80 باشد، تقریباً بدون شک مشکل خواهیم داشت، اما با مقادیر کمتر از این مقدار هم ممکن است دشواری‌هایی وجود داشته باشد (میزر و دیگران، ۱۳۹۱: ۲۳۸).

زمانی که بیشتر از دو متغیر بررسی می‌شوند از پارامتر تحمل استفاده می‌کنیم. این پارامتر موجب می‌شود با بیرون‌راندن متغیرهای پیش‌بینی که با سایر متغیرهای مستقل همبستگی بسیار بالایی دارند روش را در برابر خطر هم‌خطی چندگانه حفظ کند. از نظر مفهومی، پارامتر تحمل، مقدار واریانس متغیر پیش‌بین است که به وسیله سایر متغیرهای پیش‌بین تبیین نشده است (R^2 بین متغیرهای پیش‌بین - ۱). دامنه مقادیر پارامتر تحمل از 0 تا 1 است و اندازه‌های پایین‌تر پارامتر تحمل نشان می‌دهد که بین متغیرهای پیش‌بین روابط قوی‌تری وجود دارد (میزر و دیگران، ۱۳۹۱: ۲۳۹). اگر اندازه پارامتر تحمل در دامنه 0.40 باشد جای نگرانی دارد، اما اگر در دامنه 0.10 باشد مشکل آفرین است (استیونس، ۲۰۰۲). هر چه مقدار پارامتر تحمل به عدد 1 نزدیک‌تر باشد نشان می‌دهد احتمال وجود هم‌خطی چندگانه کمتر است. لازم به ذکر است که پارامتر تحمل برای تمام متغیرهای پیش‌بین ارائه می‌شود و پارامتر تحمل هر متغیر را باید جداگانه ارزیابی کرد.

مثال

از مثال رگرسیون چندمتغیره برای یافتن هم‌خطی چندگانه استفاده می‌کنیم. متغیرهای مستقل عبارتند از: تعداد ساعات مطالعه، علاقه به درس آمار و نمره درس SPSS در مقطع کارشناسی، متغیر وابسته هم نمره درس SPSS در مقطع ارشد است. در نتیجه ما سه متغیر پیش‌بین (مستقل) و یک متغیر ملاک (وابسته) خواهیم داشت. می‌خواهیم بدانیم که آیا بین متغیرهای مستقل هم‌خطی چندگانه وجود دارد یا خیر.

جدول ۳-۳- متغیرهای ملاک و پیش‌بین

متغیرهای پیش‌بین	متغیر ملاک
ساعات مطالعه	نمره درس SPSS (ارشد)
علاقه به آمار	
نمره درس SPSS (کارشناسی)	

اگر تنها دو متغیر مستقل وجود داشت از روش همبستگی پیروسون استفاده می‌کردیم و چنانچه ضریب همبستگی بین دو متغیر مستقل بیشتر از 0.70 بود نشان از وجود هم‌خطی بین متغیرهای مستقل داشت و باید تنها یکی از متغیرهای

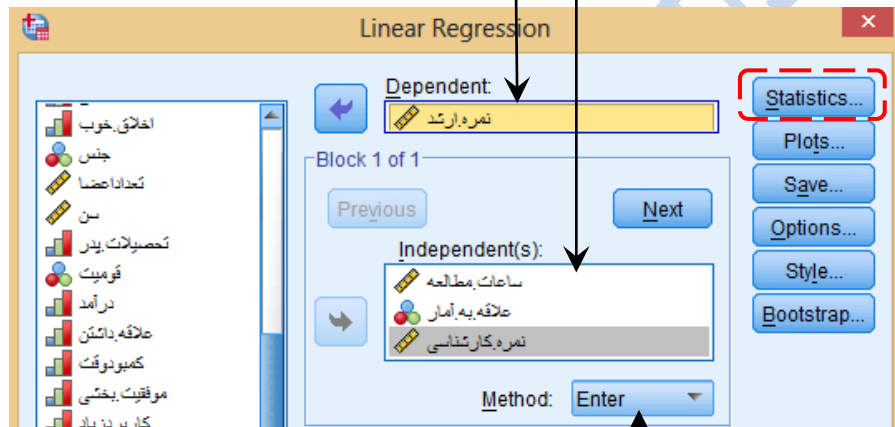
مستقل را در تحلیل رگرسیون استفاده می‌کردیم. اما در مثال کتاب سه متغیر مستقل داریم و بهتر است از دستور بررسی هم‌خطی چندگانه که در دستور رگرسیون خطی وجود دارد استفاده کنیم. می‌خواهیم بررسی کنیم که آیا مسأله هم‌خطی چندگانه بین سه متغیر مستقل یا پیش بین تعداد ساعات مطالعه، علاقه به درس آمار و نمره درس SPSS (کارشناسی) وجود دارد یا خیر.

اجرا:

مسیر زیر را دنبال می‌کنیم:

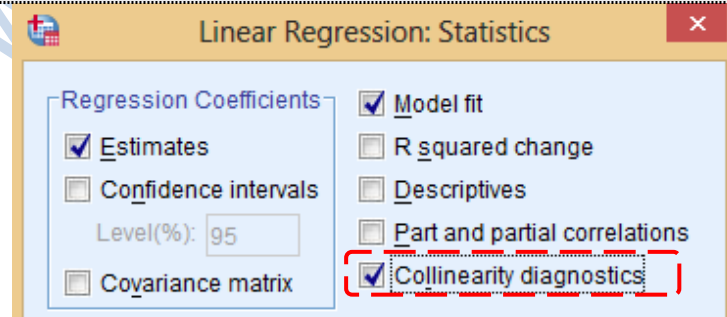
Analyze ---> Regression ---> Linear

در کادر Dependent متغیر ملاک یا وابسته را وارد می‌کنیم، یعنی متغیر نمره ارشد.
در کادر Independent متغیرهای مستقل را وارد می‌کنیم: متغیرهای ساعات مطالعه، نمره کارشناسی و علاقه به آمار. سپس گزینه Statistics را انتخاب می‌کنیم.



روش‌های متعددی برای انجام رگرسیون وجود دارد. روش همزمان یا Enter از رایج‌ترین آن‌هاست.

در کادر Statistics گزینه تشخیص هم‌خطی یا Collinearity diagnostics را انتخاب می‌کنیم.



نتایج:

از میان خروجی‌ها و جداول ارائه شده، جدول ضرایب را جهت بررسی هم‌خطی متغیرهای مستقل بررسی می‌کنیم. در این جدول نتایج ستون آماره تحمل یا تولرانس نشان دهنده میزان هم‌خطی متغیرهای مستقل است.

همان‌طور که ذکر شد از نظر مفهومی، پارامتر تحمل، مقدار واریانس متغیر پیش‌بین است که به وسیله سایر متغیرهای پیش‌بین تبیین نشده است. اگر اندازه پارامتر تحمل در دامنه ۰/۴۰ باشد جای نگرانی دارد، اما اگر در دامنه ۰/۱۰ باشد مشکل‌آفرین است (استیونس، ۲۰۰۲). مطابق نتایج به‌دست‌آمده میزان آماره تحمل در بین سه متغیر مستقل بیشتر از مقدار ۰/۴۰ است و حداقل مقدار آماره تحمل برابر با ۰/۶۳ است. نتایج به‌دست‌آمده نشان می‌دهد که میزان هم‌خطی بین متغیرهای مستقل نگران‌کننده نیست.

آماره تحمل متغیر نمره کارشناسی ۰/۶۳۲ است که بدین معناست که حدود ۶۳ درصد از واریانس متغیر نمره کارشناسی ارشد توسط دیگر متغیرهای مستقل (ساعات مطالعه و علاقه آمار) تبیین نشده است. چنانچه میزان آماره تحمل متغیری حدود ۰/۴۰ یا کمتر باشد باید اقدام به بازنگری این متغیر کرده و یا آن را از تحلیل حذف کرده و یا اقدام به ترکیب این متغیر با متغیرهایی که دارای ارتباط با آن‌هاست، نمود.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics
	B ضریب غیراستاندارد	Std. Error خطای استاندارد	Beta ضریب استاندارد شده			Tolerance آماره تحمل
1 (Constant)	13.166	1.369		9.620	.000	
ساعات مطالعه	.352	.049	.633	7.115	.000	.638
علاقه به آمار	1.512	.318	.343	4.750	.000	.969
نمره کارشناسی	-.029	.092	-.028	-.312	.756	.632