مجله دیابت و لیپید ایران. دوماهنامه مرداد- شهریور۱۳۹۲؛ دوره ۱۲(شماره ۶) ۵۰۶-۴۷۹

سری آمار: همبستگی و رگرسیون

محمد اصغری جعفرآبادی[۱*]، اکبر سلطانی [۲]، سیده مومنه محمدی [۳]

چکیده

در بسیاری از مطالعات علوم پزشکی و از جمله مطالعات مشاهده‌ای، بررسی رابطه میان متغیرها به عنوان اهداف پژوهشی تعریف می‌شوند. هدف از این مقاله، معرفی مفاهیم و روش‌های ساده و کاربردی آماری بررسی رابطه شامل انواع همبستگی و رگرسیون در این مطالعات می‌باشد.

مبانی و نحوه انجام محاسبات شاخص‌ها و آزمون فرض‌های بررسی رابطه، نحوه محاسبه فاصله اطمینان و نحوه گزارش کردن یافته‌ها برای همبستگی پیرسون، همبستگی رتبه‌ای اسپیرمن، همبستگی درون کلاسی، رگرسیون خطی ساده و رگرسیون خطی چندگانه، ارائه گردید.

برای بررسی رابطه بین دو متغیر کمی نرمال تحلیل همبستگی پیرسون، دو متغیر کمی غیر نرمال و یا رتبه‌ای تحلیل همبستگی اسپیرمن، توافق اندازه‌گیری‌های تکراری درون یک فرد یا بین مجموعه‌ای از افراد همبستگی درون کلاسی، و همچنین برای توصیف رابطه میان متغیر(های) مستقل و وابسته، پیش‌بینی متغیر وابسته بر اساس متغیر(های) مستقل و کنترل متغیرهای مخدوشگر، در قالب یک معادله، تحلیل رگرسیونی خطی ساده و چندگانه به کار می‌روند. در هر یک از وضعیت‌های فوق، علاوه بر P-Value برای تصمیم‌گیری، اندازه اثر مناسب رابطه و فاصله اطمینان آن باید گزارش شود. بر اساس مثال‌های مطالعاتی، محاسبات آزمون فرض‌ها و فاصله اطمینان‌های مزبور انجام و نتایج آنها ارائه گردید.

برای بررسی رابطه، یک یا مجموعه‌ای از عوامل خطر با یک متغیر کمی، تحلیل‌های ارائه شده با توجه به موقعیت و هدف مطالعه توصیه می‌شود.

واژگان کلیدی: رابطه، توافق، همبستگی، رگرسیون، پیرسون، اسپیرمن، درون کلاسی

۱- مرکز تحقیقات آموزش علوم پزشکی، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران

۲- مرکز تحقیقات غددو متابولیسم، پژوهشکده علوم بالینی غدد و متابولیسم، دانشگاه علوم پزشکی تهران، تهران، ایران

۳- گروه علوم تشریحی، دانشکده پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

***نشانی:** تبریز، خیابان گلگشت، خیابان عطار نیشابوری، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، گروه آمار و اپیدمیولوژی، کدپستی: ۵۱۶۶۶۱۴۷۱۱، تلفن: ۲- ۰۴۱۱۳۳۵۷۵۸۰، نمابر: ۰۴۱۱۳۳۴۰۶۳۴، پست الکترونیک: asgharimo@tbzmed.ac.ir

تاریخ دریافت: ۱۳۹۲/۳/۲۹ **تاریخ درخواست اصلاح:** ۱۳۹۲/۰۴/۲۷ **تاریخ پذیرش:** ۱۳۹۲/۰۵/۵

[Downloaded from journals.tums.ac.ir on 2025-03-10]

مقدمه

در بسیاری از مطالعات علوم پزشکی از جمله مطالعات مشاهده‌ای، یافتن روند تغییرات دو متغیر نسبت به یکدیگر، از اهداف مطالعه است. مثلاً آیا با افزایش سن افراد، فشار خون آنها هم افزایش می‌یابد، یا آیا با افزایش سن افراد، وزن آنها کاهش می‌یابد.

در مطالعه‌ای با هدف بررسی (RDW) Red Cell Distribution Width، با شدت بیماری (مبتنی بر Disease Activity Index (DAI))، با افزایش RDW، شدت بیماری هم افزایش یافت[۱] و یا در مطالعه Godarzi و همکاران (۲۰۱۲)، با افزایش نگرش افراد در مورد بیماری دیابت، مقادیر فشار خون آنها کاهش یافت[۲].

اهداف اختصاصی مطروحه معمولاً بررسی هم‌راستایی یا ناهم‌راستایی متغیرهای مورد بررسی، و ارزیابی رابطه این متغیرها با عوامل خطر می‌باشد. بنابراین از نظر آماری، بررسی رابطه مورد نظر است که در این مورد نیاز به انجام آزمون فرض آماری می‌باشد[۳].

در حالت کلی در بررسی روند تغییرات دو متغیر اهداف یافتن، ۱) رابطه بین دو متغیر، ۲) پیش‌بینی یک متغیر از روی متغیر(های) دیگر یا ۳) میزان توافق بین دو اندازه‌گیری است که به ترتیب با ۱) تحلیل همبستگی، ۲) تحلیل رگرسیون و ۳) ضریب توافق بررسی می‌شوند. در تحلیل همبستگی بر یافتن قوت رابطه بین دو متغیر، تحلیل رگرسیون بر یافتن شکل رابطه بین دو متغیر و تحلیل توافق بر میزان انطباق دو اندازه‌گیری تاکید می‌کند.

همبستگی

با توجه به طرح مطالعه و ماهیت متغیرهای مطالعه روش‌های مختلفی برای بررسی رابطه بین متغیرها وجود دارد. برای بررسی رابطه بین متغیرهای کمی همبستگی پیرسون، متغیرهای کمی غیر نرمال و رتبه‌ای همبستگی اسپیرمن (یا τ کندال) استفاده می‌شود. برای بررسی رابطه بین متغیرهای کیفی انواع تحلیل کای‌دو [٤] به کار می‌رود که در سری‌های آتی بدان پرداخته خواهد شد.

برای روشن شدن مفهوم همبستگی، به مثال زیر توجه نمایید.

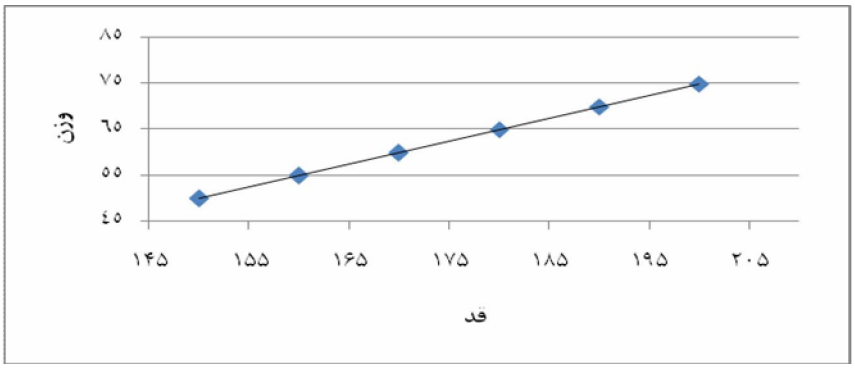
مثال ۱

معمولاً وزن افراد با بزرگتر شدن قد آنها افزایش می‌یابد. فرض کنید در مطالعه‌ای از بین افراد حاضر در یک جامعه آماری، با شرط ورود قد بالای ۱۵۰ سانتی‌متر افرادی انتخاب شوند که به ازای هر ۱۰ سانتی‌متر افزایش در قد آنها، ۵ کیلوگرم به وزن آنها افزوده می‌شود، در این صورت روند افزایش وزن به ازای هر واحد افزایش در قد کاملاً مشخص است. به عبارت دیگر، مثلاً اگر فردی با قد ۱۵۰ سانتی‌متر، ۵۰ کیلوگرم وزن داشته باشد دقیقاً می‌توان گفت که یک فرد با قد ۱٦۰ سانتی‌متر، وزنی برابر ۵۵ کیلوگرم خواهد داشت. اندازه‌گیری‌های قد و وزن برای ٦ نفر به صورت زیر خواهد بود:

قد	وزن
۱۵۰	۵۰
۱٦۰	۵۵
۱۷۰	٦۰
۱۸۰	٦۵
۱۹۰	۷۰
۲۰۰	۷۵

حال اگر این نقاط روی یک محور مختصات رسم گردد مشاهده می‌شود که همه‌ی نقاط روی یک خط راست قرار می‌گیرند (نمودار ۱):

[Downloaded from journals.tums.ac.ir on 2025-03-10]



نمودار ۱- همبستگی کامل و مستقیم بین قد و وزن

دراین حالت گفته می‌شود که بین قد و وزن همبستگی کامل وجود دارد و چون با افزایش قد، وزن نیز افزایش می‌یابد، یعنی تغییرات قد و وزن هم‌راستا هستند، بنابراین یک همبستگی کامل و مستقیم بین این دو متغیر وجود دارد.

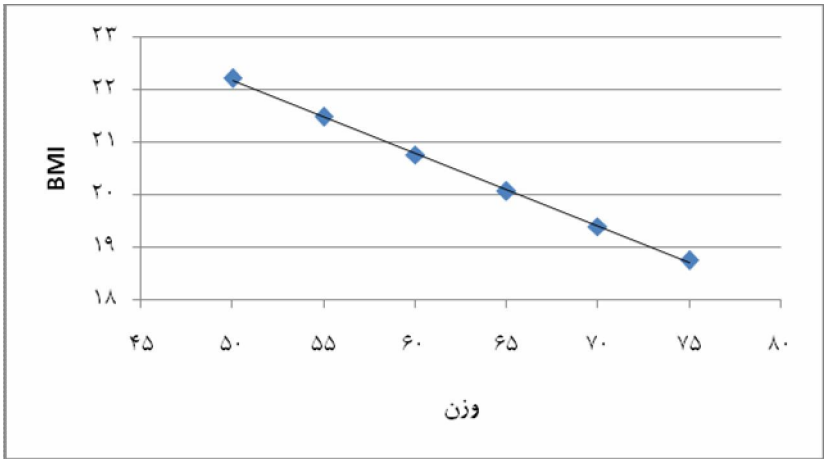
نکته: رسم نمودارها با استفاده از نرم‌افزار MSExcel 2007 و استفاده از منوی Insert و سپس انتخاب و رسم گزینه Scatter از داخل منوی نمودارها حاصل شده است.

مثال ۲

حال فرض کنید برای همین مشاهدات نمایه توده بدنی (Body Mass Index (BMI)) محاسبه گردد، نتایج به صورت زیر خواهد بود:

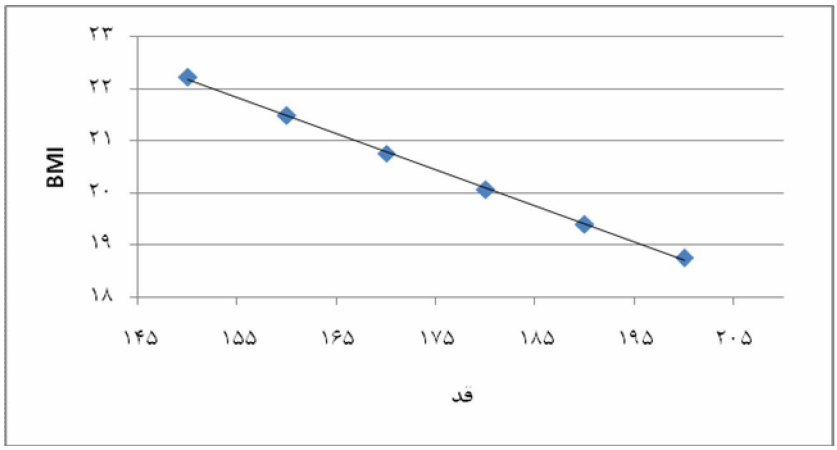
قد	وزن	BMI
۱۵۰	۵۰	۲۲/۲
۱۶۰	۵۵	۲۱/۵
۱۷۰	۶۰	۲۰/۸
۱۸۰	۶۵	۲۰/۱
۱۹۰	۷۰	۱۹/۴
۲۰۰	۷۵	۱۸/۸

برای محاسبه BMI از فرمول ^{۲} (قد (متر))/ (وزن (کیلوگرم)) =BMI استفاده می‌شود. همان‌طور که مشاهده می‌شود، در بررسی روند تغییرات وزن (یا قد) با روند تغییرات BMI، با افزایش ۵ کیلوگرم در وزن (و در نتیجه ۱۰ سانتی‌متر در قد)، BMI به اندازه ۰/۷ واحد کاهش می‌یابد. حال اگرنقاط فوق روی یک نمودار رسم گردد، ملاحظه می‌شود که این نقاط روی یک خط راست به صورت زیر قرار می‌گیرند (نمودارهای ۲ و ۳):



نمودار ۲- همبستگی کامل و معکوس بین BMI و وزن

[Downloaded from journals.tums.ac.ir on 2025-03-10]



نمودار ۳- همبستگی کامل و معکوس بین BMI و قد

در این حالت گفته می‌شود که بین وزن (یا قد) و BMI همبستگی کامل وجود دارد و چون با افزایش وزن (یا قد)، BMI کاهش می‌یابد، یعنی تغییرات قد و وزن در جهت مخالف با BMI هستند، بنابراین یک همبستگی کامل و معکوس بین وزن (قد) با BMI وجود دارد.

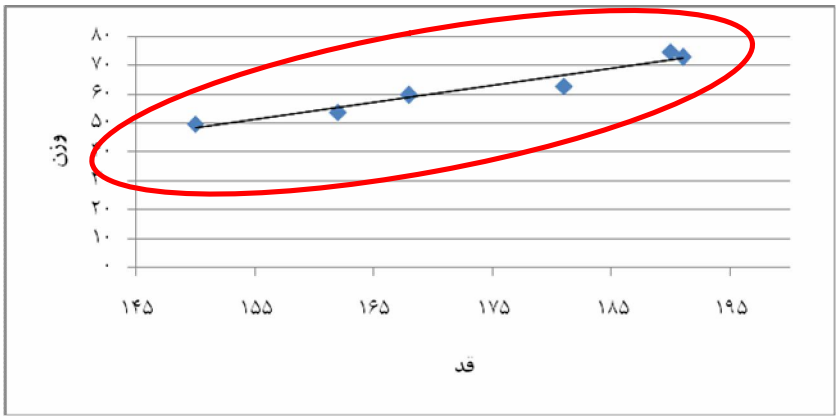
مثال فوق یک حالت فرضی برای روشن شدن مفهوم همبستگی بود. ممکن است در عمل معمولاً اعدادی مشاهده نمی‌شوند که همگی آنها دقیقاً روی یک خط قرار گیرند (و یک قاعده ریاضی رابطه بین آنها را مشخص نماید). برای روشن شدن موضوع دو مثال زیر را ببینید.

مثال ۳

ممکن است برای دو متغیر وزن و قد اعداد زیر مشاهده شوند:

قد	وزن
۱۵۰	۵۰
۱۶۲	۵۴
۱۶۸	۶۰
۱۸۱	۶۳
۱۹۱	۷۳
۱۹۰	۷۵

دراین صورت یک قاعده ریاضی دقیق نمی‌توان یافت که مثلاً به ازای یک سانتی‌متر افزایش در قد افراد وزن آنها چقدر تغییر می‌کند، یعنی اگر نقاط مربوط به این دو متغیر روی یک محور مختصات رسم گردد دقیقاً روی یک خط راست قرار نمی‌گیرند، اما آنها را می‌توان در داخل یک بیضی با شیب مثبت قرارداد (نمودار ٤):



نمودار ۴- همبستگی مستقیم بین وزن و قد

[Downloaded from journals.tums.ac.ir on 2025-03-10]

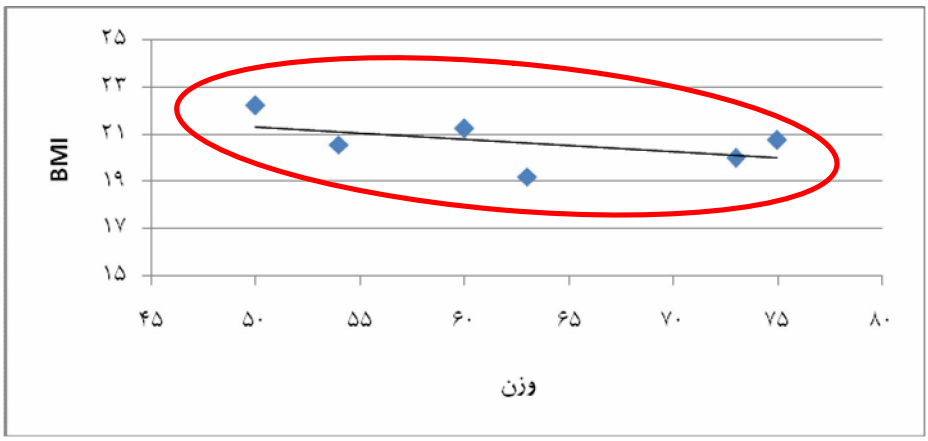
در این صورت یک رابطه مستقیم (و نه کامل) بین وزن و قد افراد وجود دارد.

مثال ٤

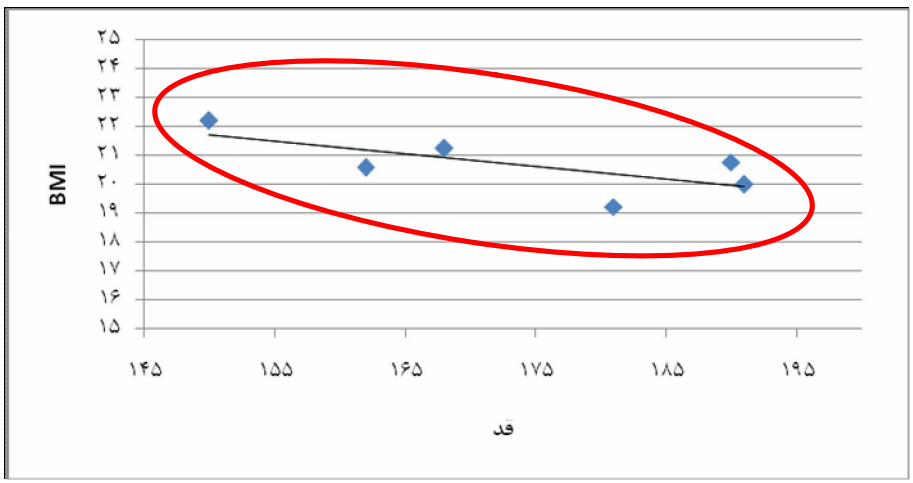
اگر مقادیر BMI برای این افراد محاسبه شود:

قد	وزن	BMI
۱۵۰	۵۰	۲۲/۲۲
۱۶۲	۵٤	۲۰/۵۸
۱۶۸	٦۰	۲۱/۲٦
۱۸۱	٦۳	۱۹/۲۳
۱۹۱	۷۳	۲۰/۰۱
۱۹۰	۷۵	۲۰/۷۸

ملاحظه می‌شود که یک قاعده ریاضی دقیق نمی‌توان یافت که مثلاً به ازای افزایش ۵ کیلوگرم در وزن افراد (یا ۱۰ سانتی متر در قد افراد) مقدار شاخص BMI چقدر کاهش می‌یابد، یعنی اگر نقاط مربوط به وزن (یا قد) در مقابل BMI روی یک محور مختصات رسم گردد، دقیقاً روی یک خط راست قرار نمی‌گیرند، اما آنها را می‌توان در داخل یک بیضی با شیب منفی قرار داد (نمودارهای ۵ و ٦):



نمودار ۵- همبستگی معکوس بین BMI و وزن



نمودار ۶- همبستگی معکوس بین BMI و قد

[Downloaded from journals.tums.ac.ir on 2025-03-10]

در این صورت یک رابطه معکوس بین وزن (یا قد) افراد و BMI وجود دارد.

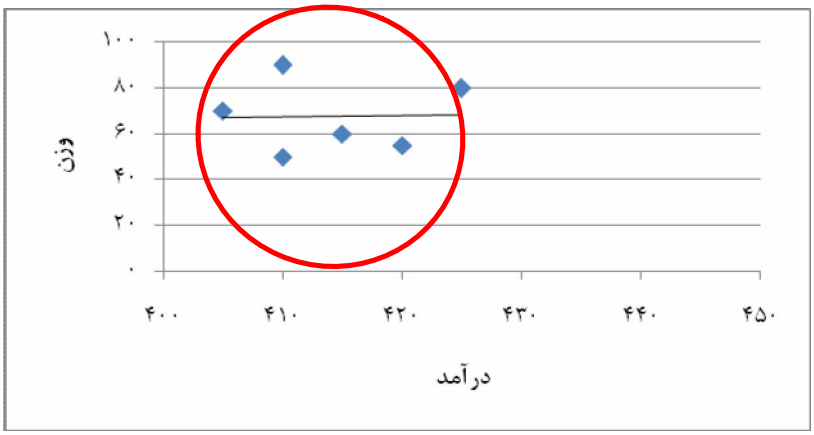
نکته: در مورد مثال‌های فوق، به طور ضمنی ملاحظه می‌شود که رابطه مورد نظر در اطراف یک خط راست بررسی می‌شود، یا به عبارت دیگر رابطه مورد نظر در قالب یک رابطه خطی ارزیابی می‌شود. علاوه بر وجود رابطه مستقیم و معکوس (چه به شکل خط راست و چه در داخل یک بیضی) ممکن است دو متغیر با یکدیگر رابطه‌ای نداشته باشند. برای روشن شدن موضوع دو مثال زیر را ببینید.

مثال ۵

برای بررسی رابطه وزن افراد با میزان درآمد آنها، اعداد زیر را در نظر بگیرید:

درآمد	وزن
٤١٠	٥٠
٤٤٠	٥٤
٤٢٠	٦٠
٤٠٥	٦٣
٤٣٠	٧٣
٤١٠	٧٥

پس از رسم نقاط، نمودار زیر حاصل می‌شود (نمودار ۷):



نمودار ۷- عدم همبستگی بین وزن و درآمد

ملاحظه می‌شود که پراکنش نقاط در داخل نمودار الگوی خاصی ندارد و نقاط در داخل یک دایره قرار می‌گیرند. در این حالت بین دو متغیر رابطه‌ای وجود ندارد.

ضریب همبستگی پیرسون

هدف بررسی رابطه خطی بین مشاهدات دو متغیر است. ملاک بررسی این رابطه خطی، شاخص ضریب همبستگی (مثال۲). عدد صفر، عدم همبستگی(مثال ۵) را نشان می‌دهد.

(گشتاوری) پیرسون نام دارد. این شاخص باید به گونه‌ای باشد که بتواند رابطه مستقیم یا معکوس را نشان دهد و همچنین بتواند دامنه ممکن برای شدت رابطه را بیان نماید. این شاخص مقادیر بین ۱- تا ۱ را اختیار می‌کند که عدد ۰ در وسط این دامنه قرار می‌گیرد؛ مقدار عددی ۱ حداکثر همبستگی (کامل) مستقیم را نشان می‌دهد (مثال۱) مقدار ۱- حداکثر همبستگی (کامل) معکوس را نشان می‌دهد در عمل ممکن است این مقادیر مشاهده نشود به عبارت

[Downloaded from journals.tums.ac.ir on 2025-03-10]

دیگر همبستگی مستقیم بین ۰ تا ۱ دیده شود (مثال۳) و یا این که همبستگی معکوس بین ۱- تا ۰ مشاهده شود (مثال٤). برای محاسبه همبستگی از فرمول آن استفاده می-شود (ضمیمه ۱). بر اساس این فرمول، به عنوان مثال، مقدار همبستگی در مثال‌های ۱ تا ۵، مقادیر همبستگی پیرسون بر اساس فرمول فوق به ترتیب برابر ۱، ۱-، ۰/۹۷، ۰/۵۱ (برای وزن) و صفر به دست می‌آید.

نکته: همان طورکه اشاره شد، همبستگی در راستای خط بررسی می‌شود، به عبارت دیگر ضریب همبستگی پیرسون تنها رابطه خطی بین دو متغیر را بررسی می‌کند و قادر نیست رابطه از نوع غیرخطی را بررسی نماید و یا حداقل ضعیف عمل می‌نماید (مثال ٦ را ببینید).

مثال ٦

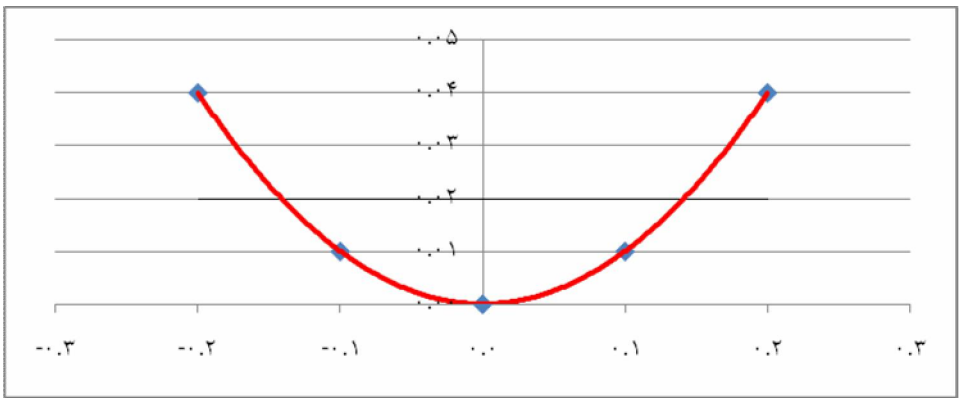
فرض کنید بررسی رابطه بین قد و (قد)^{۲} (یعنی قد به توان ۲) مد نظر باشد، به عبارت دیگر اگر قد با X و (قد)^{۲} با Y (مخرج کسر فرمول BMI) نشان داده شود، در این صورت بین قد و (قد)^{۲} رابطه $Y=X^2$ وجود دارد که این رابطه یک رابطه دقیق ریاضی ولی از نوع غیرخطی و به عبارت دقیق‌تر از نوع منحنی درجه دوم است:

قد	قد به توان ۲
۱/۵	۲/۲۵
۱/٦	۲/۵٦
۱/۷	۲/۸۹
۱/۸	۳/۲٤
۱/۹	۳/٦۱

برای سادگی محاسبات، مشاهدات قد افراد منهای میانگین شده که در این صورت برای قد و (قد)^{۲} اعداد زیر حاصل خواهد شد:

قد	قد به توان ۲
۰/۲ -	۰/۰٤
۰/۱ -	۰/۰۱
۰	۰
۰/۱	۰/۰۱
۰/۲	۰/۰٤

بر اساس اعداد فوق، نمودار زیر نتیجه خواهد شد (نمودار ۸):



نمودار ۸ - رابطه معکوس بین قد و قد به توان ۲

اگر در فرمول ارائه شده (ضمیمه ۱)، اعداد جدول فوق جایگزین شوند و همبستگی محاسبه گردد برای آن عدد صفر حاصل می‌گردد، اما همان طور که در شکل فوق ملاحظه می‌شود یک رابطه دقیق بین دو متغیر روی منحنی درجه دوم مشاهده می‌شود. در نتیجه، همبستگی پیرسون فقط رابطه از نوع خطی را می‌تواند بررسی کند ولی قادر نیست رابطه از نوع غیرخطی را نشان دهد؛ در این مثال، یک رابطه دقیق ریاضی از نوع منحنی (درجه دوم) وجود داشت ولی مقدار همبستگی برای آن برابر صفر به دست آمد. لازم به ذکر است که در مثال ۵ نیز همبستگی برای نقاط برابر

[Downloaded from journals.tums.ac.ir on 2025-03-10]

صفر به دست می‌آید، ولی درآن مثال همبستگی صفر به معنی عدم وجود رابطه (از نوع خطی یا غیرخطی) است.

نکته

الف) برای محاسبه همبستگی اگر چه نمادهای X و Y تعریف شده است که معمولاً X نمادی برای متغیر مستقل و Y نمادی برای متغیر وابسته است، ولی با این وجود نقش متغیرها در این شاخص مهم نیست، به عبارت دیگر، نیازی نیست که مشخص گردد متغیر مستقل و وابسته کدام است یعنی اگر جای X و Y عوض شود باز هم مقدار یکسانی برای همبستگی به دست می‌آید.

ب) همبستگی، به تغییرات نظیر جمع (تفریق) و ضرب (تقسیم) در متغیرها بستگی ندارد، به عبارت دیگر اگر چنین تغییراتی در متغیرها صورت گیرد، مقدار همبستگی تغییر نمی‌کند. به عنوان مثال، اگر رابطه بین قد و وزن بر اساس کیلوگرم و متر برای این دو متغیر ۰/۷ به دست آید، بر اساس گرم و متر، گرم و سانتیمتر، و کیلوگرم و متر، نیز همین مقدار به دست می‌آید. به عبارت دیگر، ضریب همبستگی به واحد اندازه‌گیری متغیرها بستگی ندارد و یک کمیت بدون بعد است.

در **تفسیر همبستگی** نکات زیر باید مورد نظر قرار گیرد:

۱) معنی‌داری ضریب همبستگی: در این مورد پس از محاسبه مقدار ضریب همبستگی یک آزمون برای بررسی معنی‌دار بودن این رابطه انجام می‌شود که شاخص آزمون آن t (یا z) و معنی‌داری آن براساس P-value مشخص می‌شود (ضمیمه ٤).

۲) جهت رابطه: مقادیر مثبت ضریب همبستگی رابطه مستقیم یعنی هم راستا بودن تغییرات دو متغیر و مقادیر منفی ضریب همبستگی رابطه معکوس یعنی مخالف بودن تغییرات دو متغیر را نشان می‌دهد.

۳) شدت همبستگی: هرچه مقدار همبستگی به ۱ و ۱- (یا قدرمطلق آن به یک) نزدیک‌تر باشد در این صورت قوت رابطه بیشتر خواهد بود و هرچه مقدار آن به صفر نزدیک شود در این صورت شدت رابطه کمترخواهد بود. بنابراین بر اساس قدر مطلق آن می‌توان قوت رابطه را مشخص نمود. Cohen (۱۹۸۹)، نقاط ۰/۱، ۰/۳ و ۰/۵ را معرفی می‌کند، مقادیر مطلق همبستگی کمتر از ۰/۱ همبستگی ناچیز، بین ۰/۱ و ۰/۳ همبستگی ضعیف، بین ۰/۳ و ۰/۵ همبستگی متوسط و بزرگتر از ۰/۵ همبستگی قوی را نشان می‌دهد [۵].

٤) علی و معلولی بودن همبستگی: که در این مورد به طور مفصل بحث خواهد شد.

پیش‌فرض‌های همبستگی

شایان ذکر است که آزمون همبستگی بر اساس این پیش‌فرض‌ها انجام می‌شود که داده‌ها به طور تصادفی جمع‌آوری شده‌اند و دو متغیر به طور همزمان دارای توزیع نرمال هستند. اگر این پیش فرض برقرار نباشد، محاسبات همبستگی جای اصل داده‌ها روی رتبه‌های متناظر با آنها صورت می‌گیرد که ضریب همبستگی دیگری با عنوان ضریب همبستگی اسپیرمن را نتیجه می‌دهد که در سری‌های آتی معرفی خواهد شد.

نکته

اگر از دیدگاه صرف ریاضی، همبستگی معیاری برای بررسی تغییرات همزمان (کواریانس استاندارد شده) بین متغیرهاست و از روی مقادیر آن نمی‌توان یک رابطه علی و معمولی را نتیجه گرفت. برای این منظور به مثال‌های زیر توجه نمایید:

مثال ۷

با گرم شدن هوا در فصل تابستان، میزان مصرف بستنی و یا نوشیدنی‌های خنک به طور روزانه افزایش می‌یابد و به عبارت دیگر روزهای گرم‌تر نوشیدنی‌ها یا بستنی‌های بیشتری را می‌طلبد. در همین حین در شهرهای ساحلی، به دلیل افزایش تعداد موارد شنا در دریا، تعداد غرق شدن‌ها در

[Downloaded from journals.tums.ac.ir on 2025-03-10]

دریا نیز افزایش می‌یابد، بنابراین یک همبستگی قوی بین میزان مصرف بستنی و غرق شدن در دریا وجود دارد.

مثال ۸

در برهه‌ای از تاریخ به دلیل نشست مواد نفتی در آمریکا، آلودگی‌های زیست محیطی در دریا ایجاد شده بود، در همین زمان روزانه تعداد کودکان مبتلا به نوعی بیماری در چین افزایش یافت. در این مورد نیز بین آلودگی‌های نفتی آمریکا و تعداد بیماری کودکان چین همبستگی وجود داشت.

در دو مثال فوق وجود همبستگی وجود یک رابطه علی و معولی را تضمین نمی‌کند، بنابراین در تفسیر همبستگی به عنوان ابزاری برای بیان رابطه بین دو متغیر باید دقت نمود و بدون در دست داشتن شواهد نظری مناسب، یک همبستگی به عنوان یک رابطه علی و معمولی تلقی نگردد. در مثال ۷، دلیل وجود همبستگی یک عامل مشترک (گرما) بوده است حال آن که در مثال ۸، عامل مشترکی نیز برای این همبستگی وجود نداشت و تنها عامل شانس دلیل این همبستگی است.

همبستگی یک ابزار است و استفاده نامناسب از ابزار نیز ممکن است اشتباه و گمراه کننده باشد و حتی ممکن است در پی آن تصمیم آنی در مورد برخی از اعمال بالینی گرفته شود که مضر و خطرناک باشد. برای برقراری یک رابطه علی و معلولی باید شرط‌های Bradford Hill (۱۹٦٥) شامل قوت رابطه (Strength)، سازگاری (Consistency)، اختصاصی بودن (Specificity)، تقدم و تاخر زمانی (Temporality)، رابطه دوز-پاسخ (Biological gradient)، معقول بودن (Plausibility)، پیوستگی (Coherence)، شواهد آزمایشی (Experiment) و تشابه (Analogy) برقرار باشد [٦].

اگر دو متغیر به صورت منظم و در قالب یک فرایند علی و معمولی با یکدیگر رابطه داشته باشند ایجاد تغییرات در یک متغیر سبب ایجاد تغییرات در متغیر دیگر شده و در نتیجه دو متغیر با یکدیگر همبستگی خواهند داشت. به عنوان مثال، همبستگی نسبتاً قوی بین قد و وزن افراد وجود دارد که این همبستگی در نتیجه یک رابطه علی و معمولی قابل مشاهده است؛ به عبارت دیگر با افزایش قد افراد، توده بدنی افراد نیز زیاد می‌شود و این موضوع سبب افزایش وزن آنها می‌گردد. همبستگی بین قد و وزن نمایشی از این رابطه علی و معمولی بین دو متغیر است. یعنی رابطه علی و معلولی بین دو متغیر است که این همبستگی را ایجاد نموده است. از طرف دیگر، وجود همبستگی بین دو متغیر یک رابطه علی و معلولی را ایجاب نمی‌نماید و تنها می‌توان گفت که دو متغیر تغییرات همزمان دارند.

ضریب همبستگی اسپیرمن و موارد کاربرد:

ضریب همبستگی رتبه‌ای اسپیرمن، توسط اسپیرمن (۱۹۰٤) معرفی شد و برای بررسی رابطه بین دو متغیر کمی غیرنرمال و یا رتبه‌ای به کار می‌رود. در محاسبه این همبستگی برای متغیرهای کمی غیرنرمال، ابتدا برای هر یک از داده‌ها، رتبه‌های متناظر با آنها اختصاص داده می‌شود و این رتبه‌ها در فرمول محاسبه ضریب همبستگی پیرسون (ضمیمه ۱) قرار می‌گیرند، عدد حاصل همان ضریب همبستگی رتبه‌ای اسپیرمن است. به عنوان مثال برای داده‌ها مثال ۳، مقدار ضریب همبستگی رتبه‌ای اسپیرمن برابر ۰/۹٤ به دست می‌آید (ضمیمه ۲).

نکته: این شاخص همبستگی یکنوای بین دو متغیر را اندازه‌گیری می‌کند و به دلیل تبدیل داده‌ها به رتبه‌های متناظر با آنها ماهیت غیر خطی داده‌های اصلی در محاسبه آن دخالتی نداشته و در نتیجه این شاخص در موارد ارزیابی روابط غیر خطی نیز به کار می‌رود. همبستگی رتبه‌ای اسپیرمن، تمایل برای بزرگ‌تر (کوچک‌تر) شدن مقادیر رتبه‌های یک متغیر را در ازای بزرگ‌تر (کوچک‌تر) شدن مقادیر رتبه‌های متغیر دیگر ارزیابی می‌کند. همچنین این شاخص به دلیل داشتن این خصوصیت، تحت تاثیر داده‌های پرت و یا پراکندگی بیش از حد موجود در داده‌ها قرار نمی‌گیرد.

[Downloaded from journals.tums.ac.ir on 2025-03-10]

آزمون معنی‌داری همبستگی

در تحلیل یک رابطه یا اختلاف، از نظر آماری زمانی رابطه یا اختلاف مورد نظر وجود دارد که معنی‌دار شود [۳]. بنابراین پس از محاسبه همبستگی باید بررسی شود که آیا همبستگی محاسبه شده از نظر آماری معنی‌دار است یا خیر؟ برای بررسی معنی‌داری همبستگی می‌توان از جدول معنی‌داری همبستگی (جدول ضمیمه ۳) استفاده کرد. مثلاً با مراجعه به این جدول، به ازای حجم نمونه ٦ و سطح معنی‌داری ۰/۰۵، همبستگی با حداقل مقدار ۰/۸۱ معنی‌دار خواهد بود. در نتیجه هر دو مورد همبستگی محاسبه شده فوق، در سطح معنی‌داری ۰/۰۵ معنی‌دار بودند.

برای آزمون، علاوه بر جدول ضمیمه ۳، آزمون معنی‌داری همبستگی بر اساس فرمولی به دست می‌آید (ضمیمه ٤) که این فرمول به صورت معادل برای بررسی معنی‌داری ضرایب همبستگی پیرسون و اسپیرمن به کار می‌رود. با توجه به فرمول، همبستگی پیرسون و اسپیرمن محاسبه شده برای داده‌های مثال ۳، معنی‌دار بودند(ضمیمه ٤).

مثال ۹

در مطالعه‌ای با هدف بررسی رابطه بین (RDW) Red Cell Distribution Width، و شدت بیماری (مبتنی بر Disease Activity Index (DAI))، با افزایش RDW، شدت بیماری هم افزایش یافت[۱]. در این مطالعه، همبستگی پیرسون بین این دومتغیر برابر ۰/٦۲ به دست آمد که با توجه به حجم نمونه ٤۷ نفر، و مراجعه به جدول ضمیمه ۳ (که در آن مقادیر بالاتر از ۰/۲۹ معنی‌دار می‌شوند)، آزمون همبستگی معنی‌دار بودن این همبستگی را تایید نمود. بنابراین بین دو متغیر مطالعه، همبستگی معنی‌دار، مستقیم و در حد قوی وجود داشت (r= ۰/٦۲ , P< ۰/۰۵).

مثال ۱۰

در مطالعه Godarzi و همکاران (۱۳۹۰)، با افزایش نگرش افراد، مقادیر فشار خون آنها کاهش یافت [۲]. در این مطالعه، همبستگی پیرسون بین این دومتغیر برابر ۰/۱٤ – به دست آمد که آزمون همبستگی، معنی‌دار بودن این همبستگی را تایید نمود (با محاسبه t و مراجعه به جدول t [۷]. بنابراین بین دو متغیر مطالعه، همبستگی معنی‌دار، مستقیم و در حد ضعیف وجود داشت (r= – ۰/۱٤ , P< ۰/۰۵). در این مورد به دلیل بزرگ بودن حجم نمونه (n = ۲۰۰)، و در نتیجه حساسیت بالا، نتیجه آزمون معنی‌دار شد ولی همان‌طور که ملاحظه می‌شود این همبستگی ضعیف است و باید در تفسیر آن با احتیاط عمل نمود. در این مورد، اندازه اثر نشان می‌دهد که این رابطه فاقد اهمیت بالینی است [۳].

محدودیت‌های همبستگی

در موقعیت‌های زیر ممکن است کاربرد همبستگی با اشتباه همراه باشد:

۱) استفاده از همبستگی پیرسون برای بررسی روابط غیر خطی
۲) استفاده صرف از همبستگی برای تبیین یک رابطه علی و معلولی
۳) استفاده از همبستگی زمانی که داده‌ها از زیرگروه‌های مختلفی تشکیل شده باشند.
٤) استفاده از همبستگی پیرسون زمانی که داده پرت وجود داشته باشد.
۵) استفاده از همبستگی زمانی که مقدار یک متغیر از پیش‌تعیین شده باشد مثلاً یک شرط ورود مطالعه باشد و روی آن تحدید صورت گرفته باشد.
٦) استفاده از همبستگی برای زمانی که اندازه‌گیری‌ها پایایی (Reliability) لازم را ندارند (که سبب کم‌برآورد شدن همبستگی (Correlation Attenuation) می‌گردد).
۷) استفاده از همبستگی به جای توافق که در این صورت همبستگی بالای بین دو اندازه‌گیری صرفاً همراستایی آنها را نشان می‌دهد و انطباق دو اندازه‌گیری را نشان نمی‌دهد.

[Downloaded from journals.tums.ac.ir on 2025-03-10]

بررسی توافق بین اندازه‌گیری‌ها (ضریب همبستگی درون کلاسی) ((Intra Class Correltion Coefficient (ICC))

در برخی از ارزیابی‌ها در علوم پزشکی، نیاز است ۱) یک اندازه‌گیری توسط یک نفر، چند بار تکرار شود، ۲) توسط چند نفر یک اندازه‌گیری یا ارزیابی صورت گیرد یا ۳) یک اندازه‌گیری در طول زمان دو یا چند بار تکرار شود. در هر یک از این موارد باید میان یا بین اندازه‌گیری‌های انجام شده توافق لازم وجود داشته باشد که در اصطلاح آن را توافق درونی ارزیاب (Intra Rater Reliabilty)، توافق بین ارزیاب‌ها (Inter Rater Reliability)، یا تکرارپذیری (Reproduceability) می‌نامند. شاخص حاصل اندازه‌ای از همگنی یا توافق اندازه‌گیری‌ها را نشان می‌دهد و زمانی به مقادیر حاصل از اندازه‌گیری اعتماد کرد و به عبارتی پایایی لازم را دارد که شاخص حاصل مقدار خاصی را کسب کند.

شاخص ICC زمانی به کار می‌رود که اندازه‌گیری‌های تکراری از یک آزمودنی انجام شده باشد. در واقع این شاخص اصلاحی از شاخص‌های همبستگی به گونه‌ای فراهم می کند که علاوه بر همراستایی بتواند تفاوت بین اندازه‌گیری‌ها را نیز در ارزیابی لحاظ کند و بدین ترتیب اندازه‌ای از توافق ونه صرفاً همراستایی فراهم کند.

نکته: برای ارزیابی توافق، شاخص های مختلفی ارائه شده است که متناسب با ماهیت داده‌ها در موقعیت‌های خاص خود استفاده می‌شوند؛ برای بررسی توافق بین دو اندازه‌گیری کیفی اسمی شاخص کاپای کوهن [۵]، توافق بین بیش از دو اندازه‌گیری کیفی اسمی شاخص کاپای فلایز [۸]، توافق بین دو اندازه‌گیری کیفی رتبه‌ای شاخص کاپای وزنی [۹]، توافق بین دو یا بیش از دو اندازه‌گیری کمی شاخص ضریب همبستگی درون کلاسی (ICC) [۱۰]، استفاده می‌شوند.

مثال ۱۱

در بررسی پایایی آزمون- باز آزمون (Test Retest) به منظور ارزیابی ثبات (Stability) یک ابزار اندازه‌گیری نظیر پرسشنامه، در یک فاصله زمانی مشخص (مثلاً دو هفته) و بررسی تکرارپذیری (Repeatability) آن ابزار، شاخص ICC استفاده گسترده‌ای دارد. در مطالعه‌ای با هدف بررسی روایی و پایایی پرسشنامه کیفیت زندگی ویژه افراد دیابتی (Iranian Diabetes Quality of Life: IDQOL)، پایایی ثبات این ابزار با روش آزمون-بازآزمون، توسط شاخص ICC بررسی گردید و نتیجه حاصل، حاکی از تایید پایایی ثبات پرسشنامه بود (مقدار شاخص برابر ۰/۸۱ که بزرگتر از ۰/۷ به دست آمد) [۱۱].

نکته: این شاخص شکل‌های مختلفی دارد که بر حسب موقعیت مورد استفاده، مشخص می‌شود [۱۰]. انجام محاسبات این شاخص در شکل‌های مختلف، با توجه به پیچیدگی آن توسط نرم‌افزارهای آماری انجام می‌شود.

تفاوت همبستگی و توافق

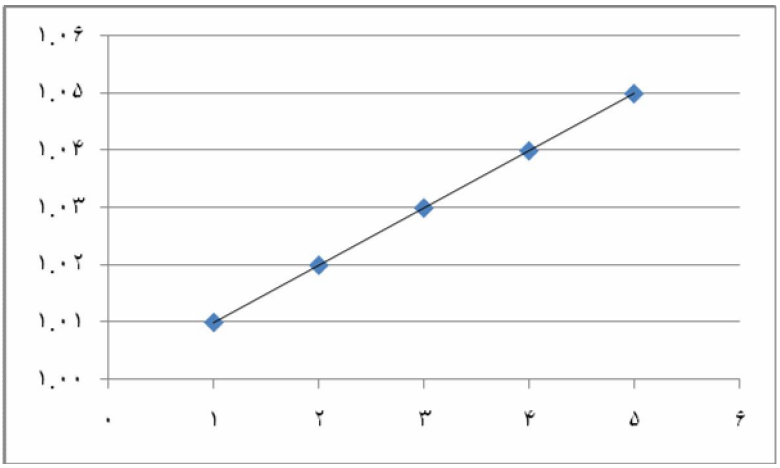
همان‌طور که اشاره شد همبستگی توافق را نتیجه نمی‌دهد و صرفاً از آن همراستایی را می‌توان استنباط کرد و توافق علاوه بر همراستایی برابری مقادیر را نیز نیاز دارد. به عبارت دیگر، ممکن است بین دو اندازه‌گیری همبستگی خیلی قوی وجود داشته باشد ولی این دو اندازه‌گیری توافق نداشته باشند. برای روشن شدن موضوع مثال زیر را ببینید.

مثال ۱۲

X_1	$= 1+.01\times X_1)(X_2$
۱	۱/۰۱
۲	۱/۰۲
۳	۱/۰۳
٤	۱/۰٤
۵	۱/۰۵

[Downloaded from journals.tums.ac.ir on 2025-03-10]

در این مورد با توجه به این که X_2 بر اساس یک رابطه خطی مشخص از X_1 ساخته شده است (۰/۰۱ برابر X_1 با عدد ۱ جمع شده است)، در نتیجه همبستگی بین این دو متغیر برابر ۱ (همبستگی کامل مستقیم) به دست می‌آید. نمودار ۹ این رابطه را نشان می‌دهد.



نمودار ۹- رابطه کامل و مستقیم بین دو متغیر X_1 و X_2

اما مقدار ICC برای این دواندازه‌گیری برابر ۰/۰۴ به دست می‌آید (توافق ناچیز). بنابراین اگر چه همبستگی کامل و در نتیجه همراستایی کامل در این دو اندازه‌گیری مشاهده می‌شود، ولی این همراستایی به معنای توافق نیست.

رگرسیون

همان‌طور که قبلاً اشاره شد همبستگی و رگرسیون از بسیار جهات شبیه هم هستند. به عبارت دیگر، برای بررسی رابطه بین متغیرها می‌توان همبستگی و رگرسیون را به کار برد با این تفاوت که در رگرسیون، تصویری از تغییر همزمان دو متغیر با یکدیگر نیز به صورت گرافیکی (منظور خط رگرسیون) آورده می‌شود. در بررسی همبستگی بین دو متغیر، یک خط راست در نظر گرفته شده و بررسی می‌گردد که چقدر پراکنش نقاط حاصل از دو متغیر به این خط نزدیک است. به عبارت دیگر هدف، یافتن خطی است که به بهترین وجه از میان نقاط در محور مختصات می‌گذرد. این خط طوری از میان نقاط بگذرد که کمترین فاصله بین نقاط و این خط ایجاد گردد. این خط همان خط رگرسیونی است که بر اساس آن نه تنها رابطه بین دو متغیر بررسی می‌شود، به علاوه از طریق این خط می‌توان مقادیر متغیر وابسته Y را بر اساس مقادیر متغیر مستقل X پیش‌بینی کرد (در مثال ۱ برای فردی که دارای قد ۱۶۰ است، وزنی برابر ۵۵ کیلوگرم پیش‌بینی می‌شود).

همان طورکه اشاره شد هدف یافتن خطی است که بهترین برازش را روی نقاط داشته باشد. این خط طوری محاسبه می‌شود که مجموع فواصل عمودی نقاط بالا و پایین آن کمترین مقدار ممکن را داشته باشد. چون فاصله برحسب توان دوم اختلاف بین نقطه و خط تعریف می‌شود به این ایده در یافتن خط رگرسیون، ایده کمترین مربعات (توان دوم) معمولی یا (Ordinary: Least Squres OLS) گفته می‌شود. براساس این ایده یک خط رگرسیون حاصل می‌گردد و یا به عبارت بهتر معادله خط رگرسیون (یک خط راست با رابطه Y=a+bX) یافت می‌شود. جهت شیب خط در واقع جهت رابطه و نزدیکی نقاط به این خط، قوت رابطه دو متغیر را (از دیدگاه همبستگی) نشان می‌دهد. خط رگرسیون در واقع رابطه متوسط بین X و Y است که در

[Downloaded from journals.tums.ac.ir on 2025-03-10]

نمونه خاصی از مشاهدات به دست می‌آید. با یافتن مقادیر a و b، خط رگرسیون تعیین می‌شود.

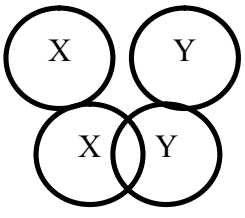
مثال ۱۳

برای داده‌های مثال ۳، با انجام رگرسیون وزن روی قد، شیب رگرسیون و عرض از مبدا و فاصله اطمینان ۹۵ درصدی آنها به ترتیب برابر (۰/۷۸ و ۰/۴۰) ۰/۵۹ و (۷/۵- و ۷۳/۲ - ۴۰/۳-) به دست آمد. همچنین آزمون معنی‌داری برای هر یک از کمیت‌های فوق نیز معنی‌دار بودن آنها را تایید نمود (محاسبات در ضمیمه ۵). بنابراین معادله خط رگرسیون برای این داده‌ها به صورت $Y = -۴۰/۳ + ۰/۵۹X$ به دست آمد.

مقدار ۰/۵۹ برای شیب خط رگرسیون نشان می‌دهد که با افزایش یک سانتی‌متری در قد افراد، **به طور متوسط** وزن آنها به اندازه ۰/۵۹ کیلوگرم (۵۹۰ گرم) افزایش می‌یابد (با توجه به علامت مثبت شیب). به عبارت دیگر، **شیب رگرسیون به طور متوسط میزان تغییرات در متغیر وابسته را به ازای یک واحد افزایش در متغیر مستقل نشان می‌دهد.** با توجه به این که فقط شیب رگرسیون در بررسی رابطه بین دو متغیر استفاده می‌شود (و عرض از مبدا بیشتر از جهت محاسبه معادله خط و پیش‌بینی مورد استفاده است)، بنابراین تفسیر شیب رگرسیون حائز اهمیت است. همچنین، با توجه به این که عدد صفر (یعنی عدم رابطه) در داخل فاصله اطمینان شیب رگرسیون حضور ندارد، بنابراین بین دو متغیر رابطه معنی‌داری وجود داشت. در عمل، برای توصیف ۱- و۱ به ۱ تبدیل می‌شوند).

مزیت ضریب تعیین در این است که معیاری برای قوت همبستگی بین دو متغیر فراهم می‌کند. دیاگرام زیر این موضوع را روشن‌تر می‌کند:

(۱)



(۲)

دایره X ۱۰۰٪ تغییرات (واریانس) X و دایره Y ۱۰۰٪ تغییرات Y را نشان می‌دهد.

۱) زمانی که X و Y تغییرات همزمان (همبستگی) نداشته باشند (مثال۵)، دو دایره X و Y یعنی واریانس‌های دو متغیر همپوشانی نخواهند داشت و به عبارتی این دو متغیر از هم مسستقل هستند[۱۲].

رابطه بین دو متغیر شیب رگرسیون نقش دارد. همچنین آزمون معنی‌داری شیب رگرسیون نشان داد که شیب به طور معنی‌داری با صفر اختلاف داشته و بین دو متغیر رابطه معنی‌داری وجود داشت.

به علاوه، با توجه به این که عدد صفر در داخل فاصله اطمینان عرض از مبدا حضور ندارد، بنابراین عرض از مبدا به طور معنی‌داری با صفر اختلاف داشت. در این مورد نیاز است به منظور **پیش‌بینی** دقیق عرض از مبدا در معادله رگرسیونی حضور داشته باشد. در صورتی که عرض از مبدا معنی‌دار نباشد می‌توان آنرا از معادله رگرسیونی حذف کرد و به معادله ساده Y=bX یعنی خط گذرنده از مبدا دست یافت. در این مثال، آزمون معنی‌داری عرض از مبدا نیز نشان داد که عرض از مبدا به طور معنی‌داری با صفر اختلاف داشت.

ضریب تعیین (تبیین)

شاخصی که از لحاظ مفهوم با ضریب همبستگی خیلی نزدیک است، توان دوم ضریب همبستگی (r) است که به آن ضریب تعیین گفته می‌شود. به توان دوم رساندن ضریب همبستگی، سبب می‌شود که مقادیر شاخص جدید بین ۰ و ۱ قرار بگیرد که صفر در آن عدم رابطه و یک حداکثر رابطه را نشان می‌دهد، همچنین (به عنوان یک عیب این شاخص) توان دوم سبب می‌شود که جهت رابطه توسط این شاخص قابل شناسایی نباشد (چون با اعمال این تابع

[Downloaded from journals.tums.ac.ir on 2025-03-10]

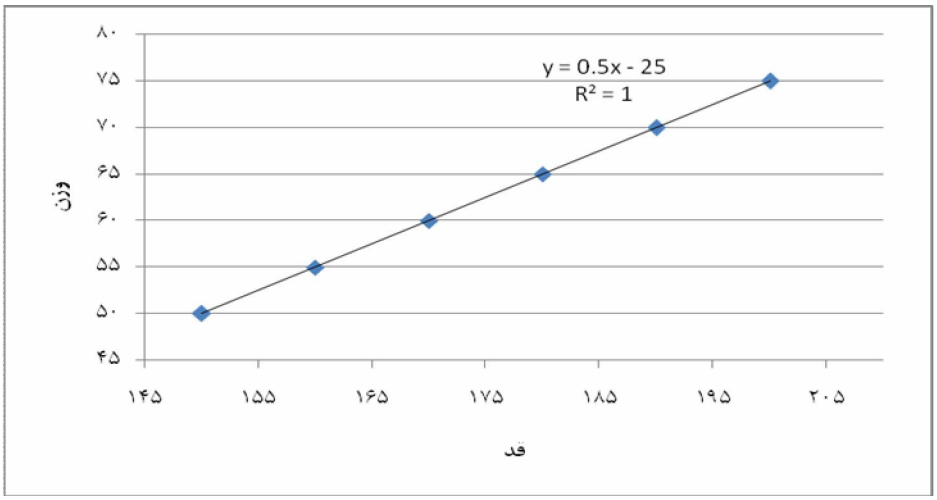
۲) زمانی که بین X و Y تغییرات همزمان (همبستگی) وجود داشته باشد، (مثال۱ تا ٤)، دراین صورت دو دایره X و Y همپوشانی دارند و ضریب تعیین، این همپوشانی را ارائه می‌دهد. به عبارت دیگر، **ضریب تعیین درصدی از تغییرات متغیر Y است که توسط X به اشتراک گذاشته می‌شود یا بیان می‌شود.**

ضریب تعیین در واقع تغییرات تبیین شده توسط رگرسیون را نسبت به تغییرات کل (تغییرات تبیین شده و تبیین نشده یا خطا) نشان می‌دهد و هر چه مقدار آن بزرگ‌تر باشد نشان دهنده این است که رگرسیون بهتر توانسته است رابطه بین متغیر(های) مستقل و متغیر وابسته را تبیین کند. در ضمن توان دوم ضریب همبستگی را می‌توان به عنوان R^2 در نظر گرفت ولی محاسبه این شاخص بر اساس ضریب همبستگی اسپیرمن اشتباه است.

مثال ١٤

محاسبه معادله خط رگرسیونی برای مثال ۱ ، ۲، ۳ و ٤:

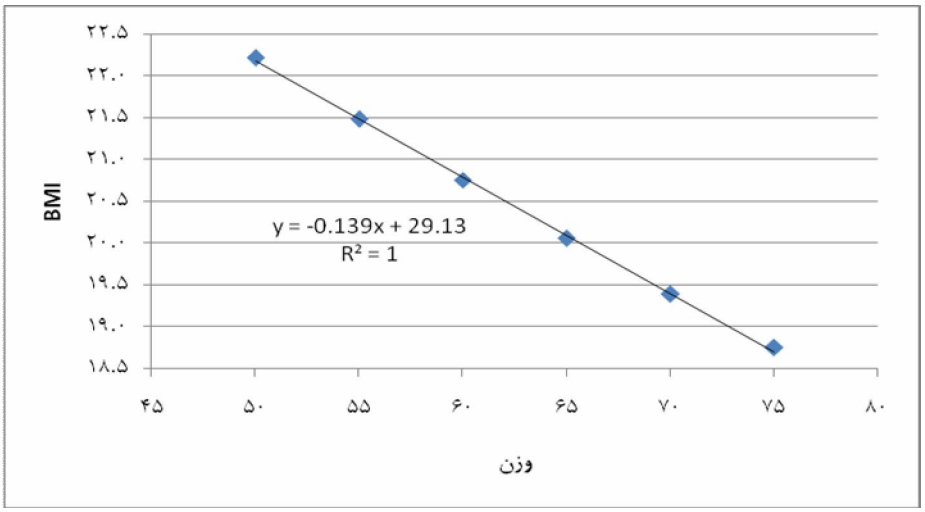
محاسبات مربوط با استفاده از نرم‌افزار MSExcel 2007 و با استفاده از دستور Insert > Scatter و پس از رسم نمودار با استفاده از دستور Add Trend Line و اضافه نمودن گزینه‌های مرتبط با R^2 و معادله خطر رگرسیون، حاصل شده است.



نمودار ۱۰- خط رگرسیون برای مثال ۱

در این مثال، معادله خط رگرسیون به صورت $Y = 0.5X - 25$ به دست آمد، به عبارت دیگر مقدار b و a در معادله به ترتیب برابر ۰/۵ و ۲۵- حاصل شد. با افزایش یک سانتی‌متری در قد افراد، وزن افراد به طور متوسط به اندازه ۰/۵ کیلوگرم افزایش می‌یابد (افزایش به دلیل علامت مثبت b). مقدار $R^2 = 1$ نیز بیانگر این است که در این مشاهدات متغیر وزن را می‌توان بر اساس متغیر قد به طور کامل (۱۰۰٪) پیش‌بینی کرد. به عبارت دیگر، مقدار واقعی مشاهده شده برای متغیر وزن به ازای قد ۱۸۰ سانتی متری برابر ٦٥ کیلوگرم بود که بر اساس این خط نیز دقیقاً همین مقدار به دست می‌آید ($Y = 0.5X - 25 = 0.5 \times 180 - 25 = 65$) و در نتیجه مقدار خطای این پیش‌بینی برابر صفر است (نمودار ۱۰).

[Downloaded from journals.tums.ac.ir on 2025-03-10]

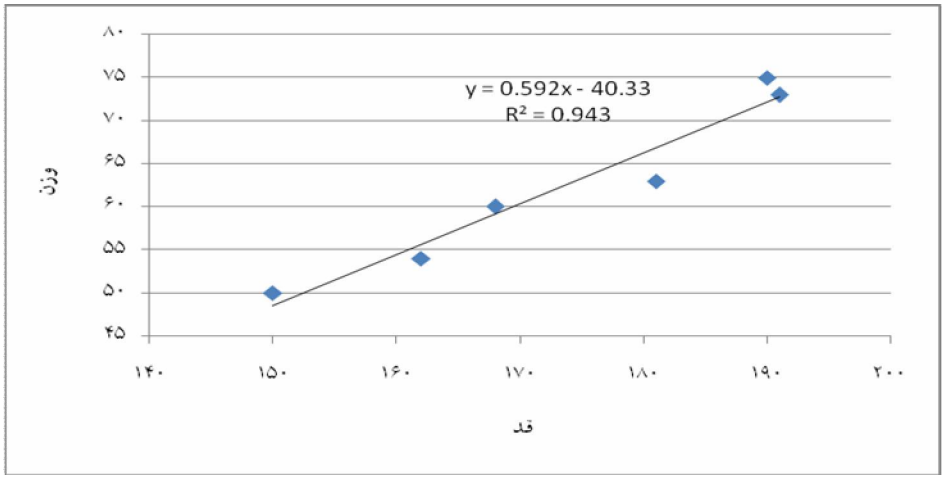


نمودار ۱۱- خط رگرسیون برای مثال ۲

در این مثال، معادله خط رگرسیون به صورت:

$$Y = -0.139X + 29.13$$

به دست آمد، به عبارت دیگر مقدار b و a در معادله به ترتیب برابر ۰/۱۳۹- و ۲۹/۱۳ حاصل شد. با افزایش یک کیلوگرم در وزن افراد، BMI افراد به طور متوسط به اندازه ۰/۱۳۹ کیلوگرم کاهش می‌یابد (کاهش به دلیل علامت منفی b). مقدار $R^2 = 1$ نیز بیانگر این است که در این مشاهدات متغیر BMI را می‌توان بر اساس متغیر وزن به طور کامل (۱۰۰٪) پیش‌بینی کرد. به عبارت دیگر، مقدار واقعی مشاهده شده برای متغیر BMI به ازای وزن ٦٠ کیلوگرم برابر ۲۰/۸ بود که بر اساس این خط نیز دقیقاً همین مقدار به دست می‌آید ($Y = -0.139X + 29.13 = -0.139 \times 60 + 29.13 = 20.8$) و در نتیجه مقدار خطای این پیش‌بینی برابر صفر است (نمودار ۱۱).



نمودار ۱۲- خط رگرسیون برای مثال ۳

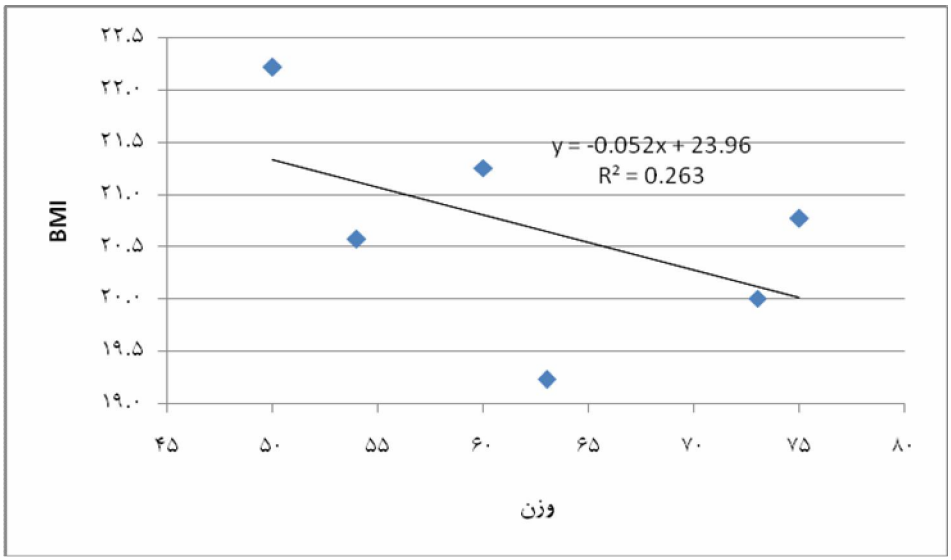
در این مثال، معادله خط رگرسیون به صورت:

$$Y = 0.592X - 40.33$$

به دست آمد. به عبارت دیگر، مقدار b و a در معادله به ترتیب برابر ۰/۵۹۲ و ٤٠/۳۳ - حاصل شد. با افزایش یک سانتی‌متری در قد افراد، وزن افراد به طور متوسط به اندازه ۰/۵۹۲ کیلوگرم افزایش می‌یابد (افزایش به دلیل علامت مثبت b). مقدار $R^2 = 0.943$ نیز بیانگر این است که در این مشاهدات ۹٤/۳ از تغییرات متغیر وزن را می‌توان بر اساس

[Downloaded from journals.tums.ac.ir on 2025-03-10]

متغیر قد پیش‌بینی کرد. به عبارت دیگر، مقدار واقعی مشاهده شده برای متغیر وزن به ازای قد ۱۸۱ سانتی متری برابر ٦٣ کیلوگرم بود که بر اساس این خط نیز مقدار آن برابر ٦٦/٨٢ به دست می‌آید (Y = 0.592X - 40.33 = 66.82= 40.33- 0.592×181) و در نتیجه مقدار خطای این پیش‌بینی برابر ٣/٨٢ می‌باشد (نمودار ١٢).



نمودار ١٣- خط رگرسیون برای مثال ۴

در این مثال، معادله خط رگرسیون به صورت:

Y = -0.052X + 23.96

به دست آمد. به عبارت دیگر، مقدار b و a در معادله به ترتیب برابر ۰/۰۵۲- و ۲۳/۹۷ حاصل شد. با افزایش یک کیلوگرم در وزن افراد، BMI افراد به طور متوسط به اندازه ۰/۰۵۲ کیلوگرم کاهش می‌یابد (کاهش به دلیل علامت منفی b). مقدار $R^2 = 0.263$ نیز بیانگر این است که در این مشاهدات ٢٦/٣ درصد از تغییرات متغیر وابسته BMI را می‌توان بر اساس متغیر وزن پیش‌بینی کرد. به عبارت دیگر، مقدار واقعی مشاهده شده برای متغیر BMI به ازای وزن ٦٠ کیلوگرم برابر ٢١/٣ بود که بر اساس این خط مقدار به دست می‌آید (Y = -0.052X + 23.96 = -0.052 ×60 + 23.96 =20.84) و در نتیجه مقدار خطای این پیش‌بینی برابر ٠/٠٤ است (نمودار ١٣).

همان‌طور که ملاحظه می‌شود هر چه مقدار همبستگی بیشتر باشد (مثال ١ در مقایسه با مثال ٣ یا مثال ٢ در مقایسه با مثال ٤) مقادیر پیش‌بینی شده متغیر وابسته به ازای مقدار معلومی از متغیر مستقل، به مقدار واقعی متغیر وابسته نزدیک‌تر می‌شود. در نتیجه هر چه قوت رابطه همبستگی بیشتر باشد، پیش‌بینی خط رگرسیون دقیق‌تر خواهد بود. تفاوت بین مقدار واقعی متغیر وابسته و مقدار پیش‌بینی شده توسط مدل خطای خط رگرسیونی در پیش‌بینی یا همان خطای پیش‌بینی یا مانده (Residual) است.

نکته: مقدار عددی که بر اساس خط و معادله رگرسیونی به ازای مقدار مشخصی از متغیر مستقل به دست می‌آید، در واقع برآورد میانگین متغیر وابسته به ازای آن مقدار مشخص از متغیر مستقل است. بنابراین همان‌طور که ملاحظه شد، پس از یافتن معادله رگرسیونی، می‌توان مقادیر متغیر(های) مستقل را در معادله قرار داد و برآوردی از مقدار متغیر وابسته را نتیجه می‌دهد که علاوه بر برآورد عددی نیاز است برآورد فاصله‌ای یا فاصله اطمینان آن نیز گزارش شود [٣].

[Downloaded from journals.tums.ac.ir on 2025-03-10]

برای مشاهده نحوه محاسبه فاصله اطمینان پیش‌بینی به ضمیمه ۷ مراجعه شود.

نکته: ضریب b متناسب با همبستگی است (ضمیمه ۵) و این موضوع سبب می‌شود که رابطه بین دو متغیر توسط شیب خط رگرسیون ارزیابی شود. بنابراین همان طور که در تفسیر r چهار نکته رعایت گردید، درتفسیر شیب خط رگرسیونی نیز ٤ نکته زیر باید رعایت شود:

۱)معنی‌داری: در این مورد نیز پس از محاسبه شیب خط رگرسیونی شاخص آزمون با فرمول $b\div SE(b)$ محاسبه می‌شود (ضمیمه ۵) و بر اساس آن p-value حساب می‌شود که مقادیر کمتر از ۰/۰۵ آن معنی‌داری رابطه را نشان می‌دهد.

۲) جهت رابطه: مشابه همبستگی، علامت مثبت جهت مستقیم رابطه و علامت منفی جهت معکوس رابطه را نشان می‌دهد.

۳) شدت رابطه: در این مورد از روی شیب معمولی نمی‌توان شدت را مشخص کرد چون از لحاظ نظری این ضریب می‌تواند مقادیر در دامنه (∞-و∞+) اختیار نماید و در نتیجه در این دامنه گسترده امکان تعیین شدت رابطه نیست ولی اگر متغیرهای مستقل و وابسته استاندارد شوند و شیب برای متغیرهای استاندارد شده رگرسیون محاسبه می‌شود (که آن را شیب استاندارد شده رگرسیونی می‌نامند)، مشابه r انتظار می‌رود مقادیر آن بین ۱ و ۱- قرار گیرد و مشابه با (دسته بندی) r می‌توان شدت رابطه را بر اساس آن تعیین کرد.

٤) **علی و معمولی بودن رابطه:** مشابه همبستگی، وجود رابطه رگرسیونی به معنای وجود رابطه علی و معمولی نیست.

نکته

در تحلیل رگرسیون مشابه با تحلیل واریانس، جدول تحلیل واریانس یا ANOVA تهیه می‌شود که شامل تغییرات کل (SSY)، تغییرات حاصل از رگرسیون (SSReg) (درصدی از واریانس کل است که توسط رگرسیون تبیین می‌شود) و تغییرات ناشی از خطا (SSres) (درصدی از واریانس کل است که توسط رگرسیون تبیین نمی‌شود) می‌باشد (ضمیمه ٦).

رگرسیون چندگانه

زمانی که یک متغیر وابسته روی یک متغیر مستقل رگرسیون می‌شود، این مدل، رگرسیون خطی ساده نامیده می‌شود. زمانی که تعداد متغیرهای مستقل بیش از یک مورد گردد، این مدل، رگرسیون خطی چندگانه (multiple) نامیده می‌شود. در مدل چندگانه، به دلیل امکان ورود متغیرهای مستقل، کنترل کردن اثر آنها نیز وجود دارد که در اصطلاح اثر آنها تعدیل (adjust) می‌شود. بنابراین در این مدل، علاوه بر اهداف **توصیف رابطه و پیش‌بینی**، که در مدل رگرسیون خطی ساده نیز قابل دستیابی بود، هدف **کنترل متغیرهای مخدوشگر** نیز قابل دستیابی است. در این مورد حتی می‌توان اهدافی نظیر مقایسه میانگین یک متغیر کمی در گروه‌های تعریف شده متغیرهای کیفی را با استفاده از وارد نمودن متغیرهای کیفی در مدل بررسی نمود (مثال ۱۵).

البته با توجه به پیچیدگی محاسبات این مدل، برای انجام محاسبات باید از نرم‌افزارهای آماری استفاده شود. مدل آماری این تحلیل به صورت:

$$Y= a+b_1X_1+b_2X_2+ \ldots+b_kX_k$$

بیان می‌شود که در a عرض از مبدا و b_1، b_2 و ... و b_k به ترتیب ضرایب رگرسیونی برای متغیرهای مستقل X_1، X_2 و ...و X_k هستند.

مثال ۱۵

در مطالعه‌ای با هدف بررسی رابطه بین فعالیت بدنی با ابعاد کیفیت زندگی (Quality of Life) و سابقه بیماری‌ها، یافته‌های مرتبط به صورت زیر گزارش شده است [۱۳]:

[Downloaded from journals.tums.ac.ir on 2025-03-10]

جدول ۱- رابطه بین فعالیت بدنی با ابعاد کیفیت زندگی (Quality of Life) و سابقه بیماری (n=۲۷۳)

متغیر مستقل	B	فاصله اطمینان ۹۵ درصدی	P-Value
بعد روانشناختی	۰/۵۲٤	(۰/۲۹٦ ۰/۷۵۳)	۰/۰۰۱
بعد اجتماعی	۰/۵۷۱	(۰/۲۳٤ ۰/۹۰۹)	۰/۰۰۱
بعد محیطی	۰/۳۳۳	(۰/۰٦۸ ۰/۵۹۷)	۰/۰۱٤
بعد بدنی	۰/۳۱٤	(-۰/۱٦۲ ۰/۷۹۱)	۰/۱۹۵
سابقه دیابت	۲/٦۷۰	(۰/۰۳۳ ۵/۳۰۷)	۰/۰٤۷
سابقه بیماری‌های قلبی	۳/۵٤۸	(۰/۱٦٦ ٦/۹۳۰)	۰/۰٤۰

متغیر وابسته: میزان فعالیت بدنی

منبع : [۱۳]

در این مطالعه، بین میزان فعالیت بدنی با ابعاد روانشناختی، اجتماعی، محیطی و بدنی کیفیت زندگی، و همچنین با سابقه بیماری‌های دیابت و قلبی رابطه معنی‌داری نشان داد. همان‌طور که ملاحظه می‌شود در مدل فوق، ترکیبی از متغیرهای کمی و کیفی مستقل حضور دارند. لازم به ذکر است برای ورود متغیرهای کیفی به مدل آنها را باید به متغیرهای نشانگر (با کدهای صفر و یک به ترتیب برای عدم حضور و حضور یک ویژگی) تبدیل نمود، مثلاً متغیر سابقه دیابت در این مدل با کد صفر برای افراد بدون سابقه دیابت و کد ۱ برای افراد با سابقه دیابت تبدیل شده بود.

همچنین زمانی که یکی از روابط توصیف و یا تفسیر می‌شود با تعدیل روی سایر متغیرهای حاضر در مدل است. به عنوان مثال، با افزایش ۱ واحدی نمره بعد روانشناختی کیفیت زندگی، با تعدیل روی سایر متغیرهای حاضر در مدل، میزان فعالیت بدنی، به طور متوسط، به اندازه ۰/۵۲٤ واحد افزایش یافته است.

تفسیر متغیرهای کیفی مستقل مشابه مقایسه میانگین‌ها صورت می‌گیرد. به عنوان مثال، با تعدیل روی سایر متغیرهای حاضر در مدل، میانگین فعالیت بدنی در افراد با سابقه بیماری دیابت، به طور متوسط به اندازه ۲/٦۷۰ واحد در مقایسه با افراد بدون سابقه بیماری دیابت بیشتر بوده است. از این دیدگاه، تحلیل رگرسیون چندگانه مشابه تحلیل کواریانس عمل می‌کند.

توانایی پیش‌بینی افزون

در تحلیل رگرسیون چندگانه می‌توان سهم یک یا دسته‌ای از متغیرهای مستقل را در میزان افزایش در پیش‌بینی متغیر وابسته بررسی نمود که به آن، توانایی پیش‌بینی افزون (Added Predictive Ability) گفته می‌شود. به عنوان مثال، در داده‌های مثال ۱۵، اگر دو مجموعه متغیر یعنی دسته اول شامل ابعاد کیفیت زندگی (ابعاد روانشناختی، اجتماعی، محیطی و بدنی) و مجموعه دوم شامل سابقه بیماری‌های دیابت و قلبی، تعریف شود و دو مدل رگرسیونی که مدل اول بر اساس متغیرهای مجموعه اول و مدل دوم بر اساس مجموع متغیرهای مجموعه اول و دوم برازش شود (یعنی متغیرهای سابقه بیماری‌های دیابت و قلبی در مدل دوم به مجموعه متغیرهای ابعاد کیفیت زندگی در مدل اول افزوده شده است)، مقادیر R^2 برای این دو مدل به ترتیب برابر ۰/۰۸ و ۰/۱۳ به دست می‌آید که اختلاف بین این دو مقدار برابر ۰/۰۵ است. به عبارت دیگر، با افزودن متغیرهای سابقه بیماری‌های دیابت و قلبی در مدل دوم به مجموعه متغیرهای ابعاد کیفیت زندگی در مدل اول، مقدار پیش‌بینی مدل به

[Downloaded from journals.tums.ac.ir on 2025-03-10]

اندازه ۵ درصد افزایش یافته است و در نتیجه توانایی پیش‌بینی افزون متغیرهای سابقه بیماری‌های دیابت و قلبی برابر ۵ درصد می‌باشد.

در بسیاری مطالعات، علاوه بر متغیرهای اصلی مطالعه، اغلب متغیرهای دیگری حضور دارند که اندازه‌گیری آنها از لحاظ زمانی، اقتصادی و نیروی انسانی مورد نیاز برای اندازه‌گیری، ارزان است. استفاده از این رویکرد سبب می‌شود تا با افزودن متغیرهای مزبور به عنوان مجموعه‌هایی به مدل، توان پیش‌گویی مدل را افزایش داد. اصطلاحاً این مدل‌ها، مدل‌های سلسله مراتبی نامیده می‌شوند و بسته به تعداد مجموعه متغیرهایی که در این مدل ها وارد می‌شوند نام‌گذاری می‌شوند. مثلاً اگر تعداد مجموعه‌هایی که وارد مدل می‌شوند دو مجموعه (مشابه مدل فوق) باشند در این صورت، مدل سلسله مراتبی دو مرحله‌ای (Tw step hierarchical modeling) ایجاد خواهد شد. از این رویکرد با افزودن هر بار یک متغیر به متغیرهای قبلی نیز می‌توان استفاده کرد که در این صورت پیش‌بینی افزون هر متغیر حاصل خواهد شد.

نحوه گزارش یافته‌ها در تحلیل رگرسیونی

نحوه ارائه یافته‌های حاصل از تحلیل‌های آماری حائز اهمیت است [۱۴]. برای یافته‌های حاصل از مثال ۳ نحوه ارائه آنها در تحلیل رگرسیونی به صورت جدول ۲ ارائه خواهد شد:

جدول ۲- رابطه بین متغیرهای وزن و قد افراد بر اساس تحلیل رگرسیون خطی (n = ۶)

متغیر مستقل	B (SE)	Beta	P-Value
قد	۰/۵۹ (۰/۰۷)	۰/۹۷	۰/۰۰۱

متغیر وابسته: وزن

R^2 = ۰/۹۴

که در آن Beta ضریب استاندارد شده رگرسیونی است و در رگرسیون خطی ساده (رگرسیون با یک متغیر مستقل)، این شاخص برابر ضریب همبستگی پیرسون است. به همین دلیل از این ضریب برای بیان شدت رابطه بین متغیرهای مستقل و وابسته (مشابه با آنچه در بحث همبستگی مطرح شد)، استفاده می‌شود. به علاوه در رگرسیون چندگانه، از این ضریب به عنوان یک **شاخص نسبی** برای مقایسه شدت ارتباط متغیرهای مستقل مختلف و اولویت‌بندی رابطه آنها با متغیر وابسته استفاده نمود. در مقایسه ضریب B (شیب رگرسیون)، به عنوان یک **شاخص مطلق** رابطه است (یعنی تغییرات مطلق متغیر وابسته را به ازای تغییر در متغیر مستقل نشان می‌دهد) و از آن نمی‌توان برای بیان شدت رابطه و همچنین مقایسه این شدت برای متغیرهای مستقل مختلف استفاده نمود. به علاوه، انتظار می‌رود ضریب استاندارد رگرسیونی مشابه همبستگی در دامنه (۱ و ۱-) قرار می‌گیرد ولی B با توجه به ابعاد اندازه‌گیری آن ممکن است هر مقداری را اختیار نماید.البته به جای SE در ستون دوم، فاصله اطمینان ۹۵ درصدی شیب رگرسیون را می‌توان با عنوان (95% CI) آورد که در این مثال به صورت ((۰/۷۸ , ۰/۴۰) ۰/۵۹) نمایش داده می‌شود.

نکته: معمولاً تحلیل رگرسیونی یک بار به صورت انفرادی برای متغیرهای مستقل انجام می‌شود و یک بار هم با حضور همه متغیرهای مستقل که به ترتیب نتایج تعدیل نشده (Un-adjusted) و تعدیل شده (Adjusted) را نتیجه می‌دهد که در این صورت جدول آن در دو بخش فوق ارائه می‌گردد.

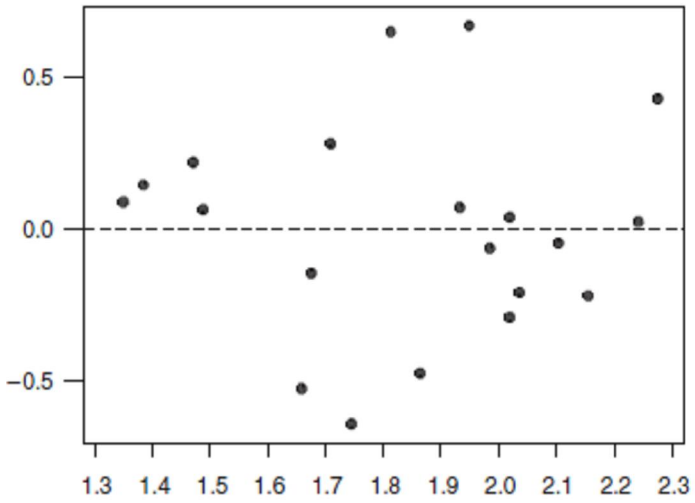
پیش‌فرض‌های رگرسیون

اعتبار یافته‌های حاصل از تحلیل‌های رگرسیونی بر مبنای پیش‌فرض‌هایی استوار است و زمانی نتایج معتبر و قابل

[Downloaded from journals.tums.ac.ir on 2025-03-10]

استفاده هستند که این پیش‌فرض‌ها برقرار باشند. یکی از پیش‌فرض‌های این تحلیل تصادفی بودن متغیر وابسته می‌باشد (در تحلیل همبستگی هر دو متغیر باید تصادفی باشند ولی در رگرسیون لازم نیست متغیرهای مستقل تصادفی باشند)، که این موضوع با انتخاب تصادفی نمونه‌ها در مرحله طرح مطالعه قابل انجام است. از پیش‌فرض‌های دیگر رگرسیون که برای انجام استنباط در مورد ضرایب رگرسیونی نیاز است، نرمال بودن توزیع مانده‌ها، همگنی واریانس مانده‌ها، عدم همبستگی بین مانده‌های رگرسیونی و مشابه همبستگی، خطی بودن رابطه بین متغیرها می‌باشد.

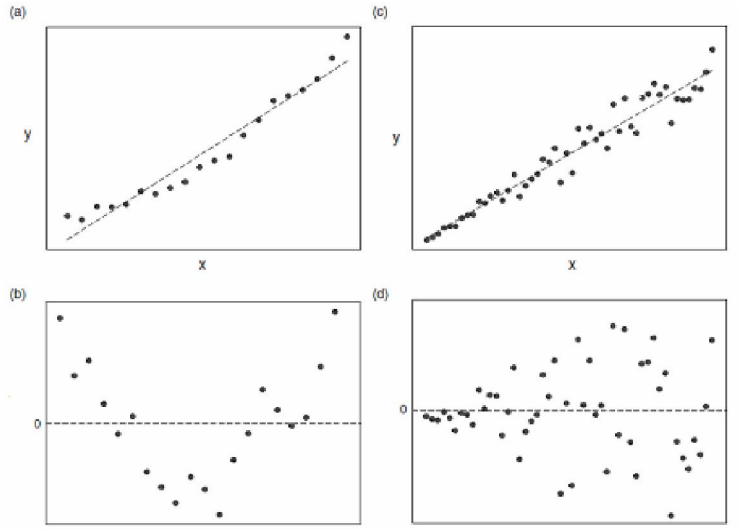
برای بررسی پیش‌فرض‌های فوق، می توان از شاخص‌ها و روش‌های نموداری استفاده کرد. برای بررسی عدم همبستگی میان مانده‌ها از شاخص دوربین- واتسون استفاده می‌شود که این شاخص در دامنه ۰ تا ٤ تغییر می کند که به ترتیب همبستگی کامل مثبت و منفی میان مانده‌ها را نشان می‌دهند. مقادیر در دامنه ۱/۵ تا ۲/۵ در این شاخص برقراری پیش‌فرض عدم همبستگی میان مانده‌ها را تایید می‌کند. در روش‌های نموداری، نمودار پراکنش متغیرهای مستقل در مقابل متغیر وابسته معیاری از خطی بودن رابطه و یا همگنی پراکندگی موجود را نشان می‌دهد. در صورتی که رابطه مورد بررسی خطی باشد و پراکندگی موجود همگن باشد در این صورت در نمودار مانده‌ها در مقابل مقادیر پیش‌بینی شده، مانده‌ها باید در اطراف خط صفر و به ازای مقادیر پیش‌بینی شده، پراکنش یکنواخت و بدون شکلی داشته باشند (نمودار ١٤).



نمودار۱۴- مانده ها (محورعمودی) در مقابل مقادیر پیش‌بینی شده (محورافقی) در صورت برقراری پیش‌فرض همگنی واریانس‌ها و خطی بودن رابطه

مواردی از نقض این پیش‌فرض‌ها در نمودار ۱۵ نشان داده شده است:

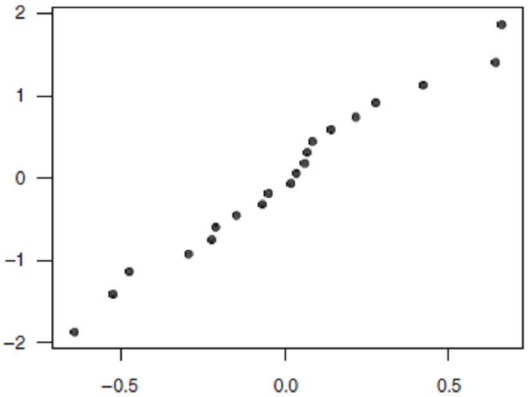
[Downloaded from journals.tums.ac.ir on 2025-03-10]



نمودار ۱۵- بررسی پیش‌فرض‌های خطی بودن و همگنی واریانس تحلیل رگرسیونی

برای شکل‌های (b) و (d) مانده ها (محورعمودی) در مقابل مقادیر پیش‌بینی شده (محورافقی) رسم شده‌اند

در نمودار ۱۵، در قسمت (a)، پراکنش متغیر وابسته Y در مقابل متغیر مستقل X یک رابطه غیر خطی را نشان می‌دهد که در قسمت (b)، برای همین متغیرها و با رسم نمودار مانده‌ها (محور عمودی) در مقابل مقادیر پیش‌بینی شده (محور افقی)، غیر خطی بودن و منحنی‌وار بودن شکل رابطه به وضوح مشخص می‌باشد. در قسمت (c)، پراکنش متغیر وابسته Y در مقابل متغیر مستقل X، افزایش پراکندگی متغیر Y را بـه ازای افزایـش مقادیر X نـشان می‌دهد که در قسمت (d)، برای همین متغیرها و با رسم نمودار مانده‌ها (محور عمودی) در مقابل مقادیر پیش‌بینی شده (محور افقی)، افزایش پراکندگی مانده‌ها با افزایش مقادیر پیش‌بینی شده کاملاً واضح می‌باشد. برای بررسی نرمالیتی نمودار احتمال نرمال (Normal Probabilty Plot) استفاده می‌شود که در این نمودار، مانده‌ها در مقابل امتیازهای نرمال استاندارد (و معمولاً توسط نرم‌افزارهای آماری) رسم می‌شوند. زمانی توزیع مانده‌ها نرمال خواهد بود که نقاط حاصل روی خط راست قرار بگیرند (نمودار ١٦).



نمودار ۱۶- نرمال بودن مانده‌ها در تحلیل رگرسیونی در صورت برقراری پیش‌فرض

در این نمودار، امتیازهای نرمال مانده‌ها (محور عمودی) در مقابل مقادیر مانده‌ها (محور افقی) رسم شده‌اند

[Downloaded from journals.tums.ac.ir on 2025-03-10]

محدودیت‌های رگرسیون

۱- در استفاده از تحلیل رگرسیونی برای پیش‌بینی، لازم به ذکر است وجود خطا در پیش‌بینی صرفاً از نوع خطای تصادفی نیست و ممکن است بخشی از آن به دلیل عدم کفایت مدل و مناسب نبودن آن باشد.

۲- پیش‌بینی در خارج از دامنه متغیرهای مستقل مجاز نیست.

۳- اندازه‌گیری‌های تکراری از افراد در داده‌ها وجود نداشته باشد که در این صورت پدیده برگشت به میانگین رخ خواهد داد و افراد با مقادیر بالاتر از میانگین در اولین اندازه‌گیری مقادیرکوچکتری در اندازه‌گیری‌های بعدی نشان خواهند داد.

٤- همان‌طور که در تحلیل همبستگی هم بدان اشاره شد، وجود رابطه رگرسیونی، یک رابطه علی و معلولی را تضمین نمی‌کند.

انواع دیگر رگرسیون

علاوه بر مدل‌های رگرسیونی خطی فوق که برای متغیرهای وابسته کمی با توزیع نرمال استفاده می‌شوند، انواع دیگر رگرسیون‌ها شامل رگرسیون لجستیک دو حالتی، چند حالتی، رگرسیون رتبه‌ای و رگرسیون پواسن وجود دارند که به ترتیب برای متغیرهای وابسته اسمی دوحالتی، اسمی چندحالتی، رتبه‌ای و شمارشی (کمی گسسته) به کار می‌روند که در سری‌های آتی معرفی خواهند شد.

نتیجه‌گیری نهایی

همبستگی و رگرسیون را برای بررسی رابطه میان متغیرها می‌توان به کار برد و تحلیل رگرسیونی علاوه بر این هدف، اهداف پیش‌بینی و کنترل متغیرهای مخدوشگر را نیز دنبال می‌کند. برای استفاده از این تحلیل‌ها نیاز است پیش‌فرض‌های خاصی در نظر گرفته شود و همچنین استفاده از این تحلیل‌ها با محدودیت‌های خاصی نیز مواجه می‌باشد که حین استفاده از این تحلیل‌های نیاز است در نظر گرفته شوند.

ضمیمه‌ها

ضمیمه ۱

فرمول ضریب همبستگی

برای محاسبه همبستگی از فرمول **زیر استفاده می‌شود.**

$$r=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$$

$$=\frac{\sum X_iY_i-\frac{(\sum X_i\sum Y_i)}{n}}{\sqrt{[\sum X_i^2-\frac{(\sum X_i)^2}{n}][\sum Y_i^2-\frac{(\sum Y_i)^2}{n}]}}$$

مفهوم این فرمول برای اولین با توسط سر فرانسیس گالتون (۱۸۷۷) [۱۵] معرفی و فرمول رسماً توسط کارل پیرسون (۱۸۹٦) [۱٦] ایجاد گردید.

برای روشن شدن موضوع مثال زیر را ببینید.

مثال ۱

محاسبه همبستگی پیرسون برای داده‌های مثال ۳ مقاله به صورت زیر می‌باشد:

گام‌های محاسبات:

۱) محاسبه حاصل‌ضرب وزن در وزن (۱)

۲) محاسبه حاصل‌ضرب قد در قد (۲)

۳) محاسبه حاصل‌ضرب وزن در قد (۳)

٤) محاسبه مجموع برای همه ستون‌ها (٤)

[Downloaded from journals.tums.ac.ir on 2025-03-10]

وزن	قد	وزن × وزن (۱)	قد× قد (۲)	وزن × قد (۳)	
۵۰	۱۵۰	۲۵۰۰	۲۲۵۰۰	۷۵۰۰	
۵٤	۱٦۲	۲۹۱٦	۲٦۲٤٤	۸۷٤۸	
٦۰	۱٦۸	۳٦۰۰	۲۸۲۲٤	۱۰۰۸۰	
٦۳	۱۸۱	۳۹٦۹	۳۲۷٦۱	۱۱٤۰۳	
۷۳	۱۹۱	۵۳۲۹	۳٦٤۸۱	۱۳۹٤۳	
۷۵	۱۹۰	۵٦۲۵	۳٦۱۰۰	۱٤۲۵۰	
۳۷۵	۱۰٤۲	۲۳۹۳۹	۱۸۲۳۱۰	٦۵۹۲٤	**مجموع (٤)**

۵) قرار دادن مجموع‌های حاصل در قسمت‌های ۱ تا ٤ فوق در فرمول محاسباتی همبستگی پیرسون به صورت زیر:

$$= \frac{۶۵۹۲۴ - \frac{۳۷۵ \times ۱۰۴۲}{۶}}{\sqrt{[۲۳۹۳۹ - \frac{۳۷۵^۲}{۶}][۱۸۲۳۱۰ - \frac{۱۰۴۲^۲}{۶}]}}$$

$$= \frac{۶۵۹۲۴ - ۶۵۱۲۵}{\sqrt{[۲۳۹۳۹ - ۲۳۴۳۷.۵][۱۸۲۳۱۰ - ۱۸۰۹۶۰.۷]}}$$

$$= \frac{۷۹۹}{\sqrt{[۵۰۱.۵][۱۳۴۹.۳]}} = \frac{۷۹۹}{۸۲۲.۶۱} = ۰.۹۷۱$$

محاسبه همبستگی اسپیرمن برای داده‌های مثال ۳ مقاله به صورت زیر انجام می‌شود.

گام‌های محاسبات:

۱) محاسبه رتبه‌های وزن (اختصاص رتبه به وزن‌ها به ترتیب از کوچکترین رتبه ۱، دومین رتبه ۲ و ...) (۱)

۲) محاسبه رتبه‌های قد (اختصاص رتبه به قدها به ترتیب از کوچکترین رتبه ۱، دومین رتبه ۲ و ...) (۲)

۳) محاسبه حاصل‌ضرب رتبه وزن در رتبه وزن (۳)

٤) محاسبه حاصل‌ضرب رتبه قد در رتبه قد (٤)

۵) محاسبه حاصل‌ضرب رتبه وزن در رتبه قد (۵)

٦) محاسبه مجموع برای همه ستون‌ها (٦)

ضمیمه ۲

مثال ۲

وزن	قد	رتبه وزن (۱)	رتبه قد (۲)	رتبه وزن × رتبه وزن (۳)	رتبه قد × رتبه قد (٤)	رتبه قد × رتبه وزن (۵)
۵۰	۱۵۰	۱	۱	۱	۱	۱
۵٤	۱٦۲	۲	۲	٤	٤	٤
٦۰	۱٦۸	۳	۳	۹	۹	۹
٦۳	۱۸۱	٤	٤	۱٦	۱٦	۱٦
۷۳	۱۹۱	۵	٦	۲۵	۳٦	۳۰
۷۵	۱۹۰	٦	۵	۳٦	۲۵	۳۰
(٦) مجموع		۲۱	۲۱	۹۱	۹۱	۹۰

۱) قرار دادن مجموع‌های حاصل در قسمت‌های ۱ تا ٦ فوق در فرمول محاسباتی همبستگی پیرسون به صورت زیر:

[Downloaded from journals.tums.ac.ir on 2025-03-10]

$$= \frac{۹۰ - \frac{۲۱ \times ۲۱}{۶}}{\sqrt{[۹۱ - \frac{۲۱^۲}{۶}][۹۱ - \frac{۲۱^۲}{۶}]}} = \frac{۹۰ - ۷۳.۵}{\sqrt{[۹۱ - ۷۳.۵][۹۱ - ۷۳.۵]}}$$

$$= \frac{۱۶.۵}{\sqrt{[۱۷.۵][۱۷.۵]}} = \frac{۱۶.۵}{۱۷.۵} = ۰.۹۴$$

ضمیمه ۳

جدول معنی‌داری ضرایب همبستگی در سطوح ۰/۰۵ و ۰/۰۱

مقادیر r برای احتمالات دودنباله‌ای			مقادیر r برای احتمالات دودنباله‌ای		
حجم نمونه	۰/۰۵	۰/۰۱	حجم نمونه	۰/۰۵	۰/۰۱
۳	۱/۰۰	۱/۰۰	۲۳	۰/۴۱	۰/۵۳
٤	۰/۹۵	۰/۹۹	۲٤	۰/٤۰	۰/۵۲
۵	۰/۸۸	۰/۹٦	۲۵	۰/٤۰	۰/۵۱
٦	۰/۸۱	۰/۹۲	۲٦	۰/۳۹	۰/۵۰
۷	۰/۷۵	۰/۸۷	۲۷	۰/۳۸	۰/٤۹
۸	۰/۷۱	۰/۸۳	۲۸	۰/۳۷	۰/٤۸
۹	۰/٦۷	۰/۸۰	۲۹	۰/۳۷	۰/٤۷
۱۰	۰/٦۳	۰/۷٦	۳۰	۰/۳٦	۰/٤٦
۱۱	۰/٦۰	۰/۷۳	٤۰	۰/۳۱	۰/٤۰
۱۲	۰/۵۸	۰/۷۱	۵۰	۰/۲۸	۰/۳٦
۱۳	۰/۵۵	۰/٦۸	٦۰	۰/۲۵	۰/۳۳
۱٤	۰/۵۳	۰/٦٦	۷۰	۰/۲٤	۰/۳۱
۱۵	۰/۵۱	۰/٦٤	۸۰	۰/۲۲	۰/۲۹
۱٦	۰/۵۰	۰/٦۲	۹۰	۰/۲۱	۰/۲۷
۱۷	۰/٤۸	۰/٦۱	۱۰۰	۰/۲۰	۰/۲٦
۱۸	۰/٤۷	۰/۵۹	۱۱۰	۰/۱۹	۰/۲٤
۱۹	۰/٤٦	۰/۵۸	۱۲۰	۰/۱۸	۰/۲۳
۲۰	۰/٤٤	۰/۵٦	۱۳۰	۰/۱۷	۰/۲۳
۲۱	۰/٤۳	۰/۵۵	۱٤۰	۰/۱۷	۰/۲۲
۲۲	۰/٤۲	۰/۵٤	۱۵۰	۰/۱٦	۰/۲۱

ضمیمه ٤

آزمون معنی‌داری همبستگی

برای بررسی معنی‌داری همبستگی از فرمول زیر استفاده می‌شود:

$$t = \frac{r}{\frac{\sqrt{۱ - r^۲}}{\sqrt{n - ۲}}} = \frac{r \times \sqrt{n - ۲}}{\sqrt{۱ - r^۲}}$$

[Downloaded from journals.tums.ac.ir on 2025-03-10]

$$\frac{\sqrt{1-r^2}}{\sqrt{n-2}}$$

که در آن ، خطای معیار همبستگی است. این فرمول به صورت معادل برای بررسی معنی‌داری ضرایب همبستگی پیرسون و اسپیرمن به کار می‌رود.

مثال ٤

بررسی معنی‌داری ضرایب محاسبه شده در مثال‌های ۱ و ۲

مقدار شاخص آزمون برای مثال ۱:

$$t = \frac{0.97}{\frac{\sqrt{1-0.97^2}}{\sqrt{6-2}}} = \frac{0.97 \times \sqrt{6-2}}{\sqrt{1-0.97^2}} = 8.2$$

مقدار شاخص آزمون برای مثال ۲:

$$t = \frac{0.94}{\frac{\sqrt{1-0.94^2}}{\sqrt{6-2}}} = \frac{0.94 \times \sqrt{6-2}}{\sqrt{1-0.94^2}} = 5.51$$

برای بررسی معنی‌دار بودن این دوشاخص باید به جدول t [۷] با ٤ (=۲ - ٦) درجه آزادی مراجعه نمود (دلیل منهای ۲ برای محاسبه درجه آزادی آن است که دو متغیر در فرمول همبستگی حضور دارند که به ازای هر یک از آنها یک محدودیت اعمال می‌شود، پس درجه آزادی برابر تعداد نمونه منهای ۲خواهد بود). مقدار P-Value برای شاخص‌های فوق به ترتیب برابر ۰/۰۰۱ و ۰/۰۰۵ به دست می‌آید.

ضمیمه ۵

نحوه محاسبه شیب رگرسیون و عرض از مبدا، فاصله اطمینان و آزمون معنی‌داری آنها

$$b = r\frac{S_Y}{S_X}$$

$$a = \overline{Y} - b \times \overline{X}$$

که در آن SX و SY به ترتیب

$$S_X = \sqrt{\sum (X_i - \overline{X})^2}$$

$$= \sqrt{[\sum X_i^2 - \frac{(\sum X_i)^2}{n}]}$$

$$S_Y = \sqrt{\sum (Y_i - \overline{Y})^2}$$

$$= \sqrt{[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}]}$$

که محاسبات آنها مشابه مخرج کسر همبستگی خواهد بود. پس از محاسبه b، عرض از مبدا a محاسبه می‌شود:

$$a = \overline{Y} - b\overline{X}$$

خطای معیار عرض از مبدا

$$SE_a = S \times \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}}$$

خطای معیار شیب

$$SE_b = S \div \sqrt{\sum (X_i - \overline{X})^2}$$

که در آن

$$S = \sqrt{\frac{\sum (Y_i - \overline{Y})^2 - b^2 \times \sum (X_i - \overline{X})^2}{n-2}}$$

برای محاسبه فاصله اطمینان ۹۵ درصدی از فرمول $b \pm t \times SE_b$ استفاده می‌شود که در آن t بر اساس جدول t [۷] با n-۲ درجه آزادی به دست می‌آید. همچنین فرمول $t = b \div SE_b$ شاخص آزمون را نتیجه دهد که برای رگرسیون با حضور یک متغیر مستقل در مدل، مقدار آن با شاخص آزمون t آزمون همبستگی یکسان است.

مثال ۵

محاسبه ضرایب رگرسیون، فاصله اطمینان آنها و آزمون معنی‌داری برای داده‌های مثال ۳ مقاله

گام‌های محاسبات:

۱) محاسبه SY به صورت

$$= \sqrt{[۲۳۹۳۹ - \frac{۳۷۵^2}{۶}]}$$

$$= \sqrt{[۲۳۹۳۹ - ۲۳۴۳۷.۵]}$$

$$= \sqrt{۵۰۱.۵} = ۲۲.۳۹$$

[Downloaded from journals.tums.ac.ir on 2025-03-10]

۲) محاسبه SX به صورت

$$= \sqrt{[۱۸۲۳۱۰ - \frac{۱۰۴۲^۲}{۶}]}$$
$$= \sqrt{[۱۸۲۳۱۰ - ۱۸۰۹۶۰.۷]}$$
$$= \sqrt{۱۳۴۹.۳} = ۳۶.۷۳$$

۳) محاسبه b به صورت

$$= .۹۷۱ \times \frac{۲۲.۳۹}{۳۶.۷۳} = .۵۹$$

که در آن ۰/۹۷۱ مقدار همبستگی پیرسون است که در مثال ٦ محاسبه شده بود. با توجه به این محاسبات، همبستگی پیرسون و شیب خط رگرسیونی هر دو با هدف بررسی رابطه بین متغیرها به کار می‌روند و همچنین علامت آنها یکی است. اما به دلیل وابستگی b به SY و SX، بعد اندازه‌گیری آن برخلاف r، به بعد اندازه‌گیری Y و X بستگی دارد.

٤) محاسبه a به صورت

$$a = \frac{۳۷۵}{۶} - ۰.۵۹ \times \frac{۱۰۴۲}{۶} = ۶۲.۵ - ۱.۵۹ \times ۱۷۳۷ = -۴۰.۳$$

۵) محاسبه S به صورت

$$S = \sqrt{\frac{۵۰۱.۵ - (.۵۹)^۲ \times ۱۳۴۹.۳}{۶-۲}} = \sqrt{۷.۰۹} = ۲.۷$$

٦) محاسبه خطای معیار عرض از مبدا به صورت

$$SE_a = ۲.۷ \times \sqrt{\frac{۱}{۶} + \frac{۱۷۳.۷^۲}{۱۳۴۹.۳}} = ۱۲.۶$$

۷) محاسبه خطای معیار شیب به صورت

$$SE_b = ۲.۷ \div \sqrt{۱۳۴۹.۳} = .۰۷$$

۸) محاسبه فاصله اطمینان عرض از مبدا به صورت زیر خواهد بود:

$$-۴۰.۳ \pm ۲.۶ \times ۱۲.۶$$
$$= -۴۰.۳ \pm ۳۲.۸$$
$$= (-۷۳.۲ \quad -۷.۵)$$

در محاسبات فوق مقدار t بر اساس جدول t [۷] وبه ازای درجه آزادی ٤ و خطای ۰/۰۵ در آزمون دودنباله‌ای (اطمینان ۹۵ درصد) به دست آمد. با توجه به این که عدد صفر در داخل فاصله اطمینان حضور ندارد، بنابراین عرض از مبدا به طور معنی‌داری با صفر اختلاف داشت. در این مورد نیاز است به منظور **پیش‌بینی** دقیق عرض از مبدا در معادله رگرسیونی حضور داشته باشد. در صورتی که عرض از مبدا معنی‌دار نباشد می‌توان آن‌را از معادله رگرسیونی حذف کرد و به معادله ساده Y=bX یعنی خط گذرنده از مبدا دست یافت.

۹) محاسبه فاصله اطمینان شیب به صورت زیر خواهد بود:

$$.۵۹ \pm ۲.۶ \times .۰۷$$
$$= .۵۹ \pm .۱۹$$
$$= (.۴۰ \quad .۷۸)$$

با توجه به این که عدد صفر (یعنی عدم رابطه) در داخل فاصله اطمینان حضور ندارد، بنابراین بین دو متغیر رابطه معنی‌داری وجود داشت. بنابراین، در عمل، برای توصیف رابطه بین دو متغیر شیب رگرسیون نقش دارد.

۱۰) محاسبه شاخص آزمون برای عرض از مبدا به صورت زیر خواهد بود:

$$t = -۴۰.۳ \div ۱۲.۶$$
$$= -۳.۲$$

که قدر مطلق آن از مقدار t بر اساس جدول t [۷] وبه ازای درجه آزادی ٤ و خطای ۰/۰۵ در آزمون دودنباله‌ای (یعنی ۲.٦) بزرگتر است و در نتیجه عرض از مبدا به طور معنی‌داری با صفر اختلاف داشت.

۱۱) محاسبه شاخص آزمون برای شیب به صورت

$$t = ۰.۵۹ \div .۰۷$$
$$= ۸.۲$$

که از مقدار t بر اساس جدول t [۷] وبه ازای درجه آزادی ٤ و خطای ۰/۰۵ در آزمون دودنباله‌ای (یعنی ۲/٦) بزرگتر است و در نتیجه شیب به طور معنی‌داری با صفر اختلاف داشته و بین دو متغیر رابطه معنی‌داری وجود داشت. همان طور که ملاحظه می‌شود، مقدار شاخص آزمون برای برررسی معنی‌داری شیب رگرسیون دقیقاً برابر مقدار شاخص آزمون برای بررسی معنی‌داری همبستگی در ضمیمه ٤ است. بنابراین در بررسی رابطه بین دو متغیر، رگرسیون خطی ساده و همبستگی پیرسون

[Downloaded from journals.tums.ac.ir on 2025-03-10]

دقیقاً یکسان عمل می‌کنند.

ضمیمه ٦

نحوه انجام محاسبات جدول ANOVA و تشکیل آن

تغییرات کل (SSY)، تغییرات حاصل از رگرسیون (SSReg) و تغییرات ناشی از خطا (SSres) به صورت زیر محاسبه می‌شوند:

الف- محاسبه SSY:

$SSY=(SY)^2$

ب- محاسبه SSReg:

$SSReg=SSY(R^2)$

درصدی از واریانس کل است که توسط رگرسیون تبیین می‌شود.

ج- محاسبه SSres:

$SSres=SSY(1-R^2)$

درصدی از واریانس کل است که توسط رگرسیون تبیین نمی‌شود.

مثال ٦

محاسبات جدول ANOVA برای داده های مثال ۳ مقاله به صورت زیر می‌باشد:

الف- محاسبه SSY:

$$= [۲۳۹۳۹ - \frac{۳۷۵^۲}{۶}]$$

$$= [۲۳۹۳۹ - ۲۳۴۳۷.۵]$$

$$= ۵۰۱.۵$$

ب- محاسبه SSReg:

$$SSReg = ۵۰۱.۵ \times .۹۴۳ = ۴۷۳.۱۲$$

ج- محاسبه SSres:

$$SSres = ۵۰۱.۵ \times (۱ - .۹۴۳) = ۲۸.۳۸$$

بر اساس اطلاعات محاسبه شده فوق، جدول تحلیل واریانس (ANOVA) به صورت زیر تهیه می‌شود:

تحلیل واریانس (ANOVA)

منبع خطا	SS	درجات آزادی	واریانس	F	P-Value
رگرسیون	٤٧٣/١٢	١	٤٧٣/١٢	٦٦/٦٩	٠/٠٠١
خطا	٢٨/٣٨	٤	٧/٠٩		
کل	٥٠١/٥	٥			

درجه آزادی کل برابر تعداد کل نمونه منهای یک، درجه آزادی رگرسیون برابر ۱ و درجه آزادی خطا برابر تعداد کل نمونه منهای ۲ می‌شود. واریانس برابر SS تقسیم به درجه آزادی می‌شود.

ضمیمه ۷

نحوه انجام محاسبات فاصله اطمینان ۹۵ درصدی پیش‌بینی

برای محاسبه فاصله اطمینان ابتدا نیاز است خطای معیار پیش‌بینی به دست آید:

$$SE = S\sqrt{\frac{۱}{n} + \frac{(x_p - \bar{x})}{\sum (x_i - \bar{x})^۲}}$$

که در آن S در فرمول محاسبه فاصله اطمینان و آزمون a و b به دست آمد و فرمول فاصله اطمینان به صورت زیر است:

$$y \pm (t_{n-2} \times SE)$$

لازم به ذکر است که پیش‌بینی بر اساس مدل رگرسیونی باید در دامنه متغیرهای مستقل استفاده شده در مدل باشد، پیش‌بینی در خارج از این دامنه اعتبار ندارد.

مثال ۳

محاسبه مقدار پیش‌بینی و فاصله اطمینان X = ۱۸۰ بر اساس داده‌های مثال ۳ مقاله در این صورت مقدار پیش‌بینی بر اساس معادله y = 0.5921x - 40.336، مقدار پیش‌بینی وزن به ازای قد ۱۸۰ سانتی‌متر برابر ٦٦/٢٤ کیلوگرم خواهد بود. برای محاسبه فاصله اطمینان:

۱) گام اول: محاسبه خطای معیار پیش‌بینی به صورت:

$$SE = ۵.۶۴ \times \sqrt{\frac{۱}{۶} + \frac{(۱۸۰ - ۱۷۳.۷)}{۱۳۴۹.۳}}$$

$$= ۲.۳۳$$

[Downloaded from journals.tums.ac.ir on 2025-03-10]

که در آن S = ۵/٦٤ قبلاً محاسبه شده بود.

۲) گام دوم: محاسبه فاصله اطمینان به صورت:

$$= ۶۶.۲۴ \pm ۲.۲۶ \times ۲.۳۳$$

$$= ۶۶.۲۴ \pm ۵.۲۷$$

$$= (۶۰.۹۷ \quad ۷۱.۵۱)$$

که در آن شاخص t بر اساس ٤ درجه آزادی و ۹۵٪ اطمینان از جدول t [۷] به دست آمد.

سپاسگزاری

از همکاری صمیمانه دانشکده بهداشت دانشگاه علوم پزشکی تبریز سپاسگزاریم. این مقاله منبع تأمین مالی نداشته است.

مأخذ

1- Abedimanesh N, Alipour B, Soumi MH, Ostadrahimi A, Abedimanesh S, Asghari Jafarabadi M*, et al*. The relationship between Red Cell Distribution with Levels and the Disease Activity in Ulcerative Colitis. *Govaresh*. 2013 17(3):183-8.

2- Godarzi M, Ebrahimzadeh E, Rabi A, SaeediPour B, Asghari Jafarabadi M. Evaluation the relationship Between Knowledge, Attitude and Participation with Self Efficacy in Type 2 Diabetes Patients in Karaj. *Journal of Diabetes and Lipid Disorders*. 2012 11(3):269-81.

3- Asghari Jafarabadi M, Mohammadi SM. Statistical Series: An Introduction to Inferential Statistics (Point Estimation, Confidence Interval and HypothesisTesting). *Journal of Diabetes and Metabolic Disorders*. 2013 Under Press [In Persian].

4- Agresti A. *Categorical Data Analysis*. 2nd Ed. New York: John Wiley and Sons; 2003.

5- Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.

6- Bradford Hill A. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. 1965 58 295-300.

7- Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Tests for Comparing the Means. *Journal of Diabetes and Metabolic Disorders*. 2013 Under Review [In Persian.[

8- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971 76(5):378-82.

9- Cohen J. Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*. 1968 70 (4):213-20.

10- Shrout P, Fleiss JL. Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin*. 1979 86(2):420-8.

11- Mirfeizi M, Asghari Jafarabadi M, Mehdizadeh Toorzani, Mohammadi SM, Dehghan Azad M, Vizheh Mohammadi A*, et al*. Feasibility, Reliability and Validity of the Iranian Version of the Diabetes Quality of Life Brief Clinical Inventory (IDQOL-BCI). *Diabetes Research and Clinical Practice*. 2012 96(2):237-47.

12- Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Probability and Distributions. *Journal of Diabetes and Lipid Disorders*. 2013 12(2):101-17 [in Persian.[

13- Javadivala Z , Ahmad K , Allahverdipour H, Asghari Jafarabadi M , Tallebian H. Modeling the Relationship between Physical Activity and Quality of Life in Menopausal-aged Women: A Cross-Sectional Study. *J Res Health Sci*. 2013 13(2): Under Press.

14- Asghari Jafarabadi M, Mohammadi SM. Statistical Series: Summarizing and Displaying Data. *Journal of Diabetes and Lipid Disorders*. 2013 12(2):83-100 [in Persian].

15- Galton F. Typical laws of heredity. *Proc R Inst Great Britain*. 1877 8:282-301.

16- Pearson K. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Phil Trans R Soc Lond Series A*. 1896 187:253-318.

[Downloaded from journals.tums.ac.ir on 2025-03-10]