

منطق و کاربرد تحلیل همبستگی کانونی در پژوهش‌های اجتماعی

علی یوسفی (دانشیار جامعه‌شناسی دانشگاه فردوسی مشهد، نویسندهٔ مسؤول)

yousofi@um.ac.ir

عاطفه باقریان (کارشناس ارشد پژوهش علوم اجتماعی دانشگاه فردوسی مشهد)

f.bagheriyan85@gmail.com

مینا شجاعی (کارشناس ارشد پژوهش علوم اجتماعی دانشگاه فردوسی مشهد)

shojaemina@hotmail.com

چکیده

از آنجایی که در تحقیقات اجتماعی غالباً تمایل داریم اثر هم‌زمان چندمتغیر مستقل را بر چند متغیر وابسته بسنجیم و یا این‌که مفاهیم مورد بررسی، به طور معمول سازه‌هایی هستند که خود دارای چندین بعد هستند و از طرفی اعتبار استنباط آماری تکنیک‌های رایج چندمتغیره در چنین مواردی محل تردید است، استفاده از تکنیک تحلیل همبستگی کانونی، اجتناب ناپذیرمی‌شود. هدف نوشتار حاضر نیز اثبات ضرورت استفاده از این تکنیک در وضعیت‌هایی است که محقق درصدد بررسی روابط میان دو مجموعه از متغیرهاست و دیگر تکنیک‌های چندمتغیره پاسخگو نیستند؛ لذا ضمن معرفی منطق و کاربرد تکنیک تحلیل همبستگی کانونی با ارایهٔ یک مثال روشن کاربردی با استفاده از نرم‌افزار SPSS و تفسیر خروجی‌های مربوط، سعی در نشان‌دادن نقاط قوت این تکنیک داشته است و هم‌زمان، محدودیت‌ها و الزامات استفاده از این تکنیک در تحقیقات کمی اجتماعی را گوشزد می‌کند. درعین حال همانند دیگر روش‌های آماری چند-متغیره، نتایج تحلیل همبستگی کانونی نیز باید در معرض روش‌های اعتبارسنجی برای کسب اطمینان از نتایج، گذاشته شود که در این نوشتار به آن‌ها اشاره شده است.

کلیدواژه‌ها: تحلیل همبستگی کانونی، ضریب همبستگی کانونی، متغیر کانونی، تابع کانونی، ریشهٔ کانونی.

۱. مقدمه

مفاهیم مورد بررسی در علوم انسانی و به‌خصوص در علوم اجتماعی اغلب سازه‌های پیچیده و چندبعدی بوده‌اند و در هنگام تحلیل، هر یک از این ابعاد، خود متغیری مجزا محسوب می‌شوند. از

سوی دیگر، همان‌گونه که در این علوم، به‌ندرت نظریات تبیینی تک‌علتی مطرح شده است، کمتر با نظریاتی مواجه می‌شویم که علت‌های متعدد، موجد تنها یک اثر واحد شده باشند و یا این‌که متغیر وابسته مورد مطالعه، تنها یک بعد داشته باشد؛ بنابراین محققان اغلب در موقعیت‌هایی قرار می‌گیرند که در واقع با چندین متغیر مستقل از سویی و چندین متغیر وابسته از سوی دیگر مواجه‌اند. تکنیک‌های چندمتغیره رایج، در تحلیل داده‌هایی با ساختار مذکور، دچار نارسایی‌های عمده‌ای هستند (تک، ۱۹۹۶؛ گارسن، ۲۰۰۸؛ اندرسون، ۲۰۰۳). تحلیل رگرسیون چندگانه، قادر به پیش‌بینی مقدار یک متغیر وابسته واحد، توسط یک تابع خطی است که از مجموعه‌ای از متغیرهای مستقل حاصل شده است، در حالی که در بسیاری از موقعیت‌ها محقق، خواستار پیش‌بینی هم‌زمان چندین متغیر وابسته، توسط مجموعه‌ای از متغیرهای مستقل است. تحلیل همبستگی کانونی^۱ که از مدل‌های آماری چندمتغیره است، تحقق این هدف را ممکن می‌کند (گرین، ۱۹۷۸). همبستگی کانونی، برخلاف سایر روش‌ها که به بررسی متغیرها به صورت جدا از هم می‌پردازند، می‌تواند روابط بین چند مجموعه از متغیرها را نشان داده و بررسی کند؛ علاوه‌براین، اگر بین مجموع متغیرها دو یا سه رابطه وجود داشته باشد، تکنیک همبستگی کانونی قادر به تشخیص و نشان‌دادن آن‌ها است؛ بنابراین تحلیل همبستگی کانونی این قابلیت را دارد که از لحاظ فنی، داده‌هایی متشکل از چندین مجموعه متغیر را تحلیل کند، این امر از لحاظ نظری نیز اثبات می‌شود (شری، ۲۰۰۵: ۳۷). به طور کلی وقتی تکنیک‌های سطوح پیشین، کارایی خود را از دست می‌دهند، همبستگی کانونی به‌عنوان آخرین تلاش محققان برای تحلیل داده‌ها و به‌عنوان مناسب‌ترین و قدرتمندترین تکنیک گسترش‌یافته تحلیل‌های چندگانه، مورد استفاده قرار می‌گیرد. نکته قابل توجه این است که در این تکنیک امکان ورود داده‌های غیر کمی نیز وجود دارد (هایر و دیگران، ۱۹۹۸: ۱).

برای اثبات ضرورت استفاده از تکنیک همبستگی کانونی، لازم است چند نوع مختلف از داده‌هایی را که در یک پژوهش ممکن است به‌دست آید و نیز تکنیک آماری چندگانه متناسب برای تحلیل هر یک از این انواع، مورد توجه قرارگیرد. احتمالاً رایج‌ترین شرایط وقتی است که در آن پژوهشگر تعداد کمی متغیر برای تحلیل در دست داشته باشد. در این موقعیت، پژوهشگر برای هر

1. Canonical Correlation analysis

بخش تحلیل دارای یک متغیر وابسته (که قصد دارد آن را تبیین کند) و تعداد محدودی متغیرهای مستقل است. شاید بهترین روش در این جا رگرسیون چندمتغیری باشد؛ اما وقتی تعداد متغیرها در مجموعه داده‌هایی که باید در معادله (یا معادلات) رگرسیون منظور شود افزایش یابد، این روش یک ابزار پژوهشی پردردسر می‌شود. در نهایت، به نقطه‌ای می‌رسیم که تفسیر بامعنا یافته‌ها در رگرسیون ناممکن می‌شود، مگر آن‌که از برخی فنون کاهش داده‌ها^۱ استفاده کنیم. تحلیل عاملی احتمالاً عمومی‌ترین وسیله حصول این هدف است، هرچند از طریق آن با کاربرد عامل‌ها به جای متغیرها، می‌توان به صرفه‌جویی قابل توجهی دست یافت؛ اما روش تحلیل عاملی نیز از لحاظ نظری دارای یک نقص جدی است. تحلیل عاملی برخلاف رگرسیون، بین متغیرهای وابسته و مستقل تمایزی قایل نمی‌شود؛ یعنی بین متغیرهای تبیین‌کننده و تبیین‌شونده تفاوتی نمی‌گذارد. درحقیقت در تحلیل عاملی اگر همه متغیرها وارد شوند، این احتمال خیلی زیاد است که متغیرهای مستقل و وابسته مخلوط شوند؛ زیرا متغیرهایی که با یکدیگر همبستگی نسبی قوی دارند، در یک عامل دارای بارهای عاملی قوی خواهند بود. در این گونه موارد، پژوهشگر می‌تواند برای حل این تناقض، چنانچه زمینه‌های نظری کافی در دست داشته باشد، نخست مجموعه داده‌ها را به متغیرهای مستقل و وابسته تقسیم و سپس هر بخش را جداگانه تحلیل عاملی کند؛ اما این استراتژی پژوهشی علاوه بر ملاحظات نظری، دارای دو جنبه مهم است که باید به آن توجه شود. نخست اگر بیش از یک عامل وجود داشته باشد که باید وابسته در نظر گرفته شود، کاربرد رگرسیون مسأله‌ساز می‌شود. دومین ملاحظه در این روش، به ماهیت خود تحلیل عاملی مربوط می‌شود. تحلیل عاملی، عامل‌ها را بر اساس همبستگی‌های درونی بین متغیرهایی که در دسترس آن است، برمی‌گزیند و تلاش می‌کند که واریانس تبیین‌شده در آن‌ها را بیشینه سازد؛ بنابراین، عامل‌هایی که استخراج می‌شوند به احتمال زیاد متغیرهای مرکبی نیست که بهترین تبیین را برای مجموعه متغیرهای وابسته فراهم کند. این مطلب ممکن است قابل قبول باشد، مشروط بر آن‌که پژوهشگر بخواهد عامل‌های متغیر مستقل و وابسته را مستقل از یکدیگر برگزیند؛ اما اگر هدف کلی پژوهش، تبیین هرچه بیشتر مجموعه متغیرهای وابسته از روی متغیرهای مستقل تا حد ممکن باشد، استراتژی پژوهش برای چنین مقصودی طرح‌ریزی نشده است. این عمل بسیار مؤثر تحلیل

1. Data reduction

همبستگی کانونی است. تحلیل همبستگی کانونی به عنوان درون داد اساسی خود با دو مجموعه متغیر که به هر یک از آنها می توان یک معنای نظری داد، سروکار دارد. استراتژی اساسی آن عبارت است از استنتاج یک ترکیب خطی از مجموعه هایی از متغیرها به گونه ای که همبستگی بین دو ترکیب خطی بیشینه شود (هومن، ۱۳۸۰: ۲۴۷).

آنچه اهمیت استفاده از تکنیک همبستگی کانونی در تحقیقات علوم اجتماعی را نشان می دهد، این واقعیت است که محققان در این علوم با مفاهیمی سروکار دارند که بسیط نبوده و برعکس از ابعاد متعددی ساخته شده اند؛ به عبارت دیگر، محققان در هر دسته از متغیرهای مستقل و وابسته با چندین سازه مواجه هستند، به گونه ای که حتی اگر تنها یک متغیر وابسته در مدل تحلیلی وجود داشته باشد، با لحاظ کردن هر یک از ابعاد این متغیر به عنوان یک متغیر مجزا، در واقع چند متغیر وابسته در مدل تحلیلی وجود خواهد داشت؛ بنابراین استفاده از تکنیک همبستگی کانونی در تحقیقات اجتماعی علاوه بر کاهش خطای نوع اول^۱ که ناشی از استفاده از چند رگرسیون ساده خواهد بود، می تواند برآورد دقیقی از مجموعه روابط درونی دو دسته متغیرهای مستقل و وابسته ارائه دهد.

براین اساس هدف نوشتار حاضر معرفی منطق و کاربرد این تکنیک در تحقیق اجتماعی در قالب مثال های روشن با استفاده از نرم افزار SPSS و تفسیر خروجی های مربوط و در نهایت معرفی محدودیت ها و الزامات استفاده از این تکنیک در تحقیقات کمی اجتماعی است.

۲. مبانی نظری

۲.۱. تعریف تحلیل همبستگی کانونی

تحلیل همبستگی کانونی یکی از اعضای خانواده تکنیک های آماری خطی چندمتغیره است که در حدود ۷۰ سال پیش توسط اچ. هتلینگ (۱۹۳۶-۱۹۳۵) توسعه یافته است (شری، ۲۰۰۵: ۳۸). تحلیل همبستگی کانونی متضمن افراز کردن^۲ مجموعه متغیرها، به دو مجموعه است. مجموعه X و

۱. این خطا هنگامی رخ می دهد که محقق به اشتباه فرض صفر را رد کند. شایان ذکر است که هرچه تعداد آزمون های

مورد استفاده بیشتر باشد، احتمال رخداد این نوع خطا بیشتر خواهد شد.

۲. افراز یک مجموعه؛ یعنی تبدیل یک مجموعه به زیر مجموعه هایش به طوری که اشتراک زیر مجموعه ها با هم، مجموعه

تهی باشد و اجتماع آنها برابر مجموعه افراز شده باشد.

مجموعه Y ؛ بنابراین هدف، به دست آوردن ترکیب‌های خطی $U = ax$ و $V = by$ است به قسمی که U و V دارای بزرگترین همبستگی ممکن باشند. چنین ترکیب‌های خطی می‌توانند اطلاعاتی درباره بستگی‌های بین دو مجموعه از متغیرها به دست دهند (جانسون و ویچرن، ۱۳۷۹: ۳۲۰).

تحلیل همبستگی کانونی دارای بعضی خاصیت بیشینه‌گی^۱ مشابه با خاصیت‌های تحلیل مؤلفه اصلی است؛ اما درحالی‌که تحلیل مؤلفه اصلی بستگی‌های درونی بین مجموعه‌ای از متغیرها را در نظر می‌گیرد، تمرکز همبستگی کانونی بر همبستگی بین دو گروه از متغیرهاست (جانسون و ویچرن، ۱۳۷۹: ۳۲۰). تحلیل همبستگی کانونی در واقع یک مدل آماری چندمتغیره است که مطالعه ارتباطات داخلی بین دو مجموعه از متغیرها را ممکن می‌کند. یک مجموعه از متغیرها را تحت عنوان متغیرهای مستقل و بقیه را تحت عنوان متغیرهای وابسته در نظر می‌گیرند. برای هر یک از دو مجموعه، متغیر کانونی^۲ مربوط ساخته می‌شود. تفاوت همبستگی کانونی با رگرسیون چندمرحله‌ای، این است که در همبستگی کانونی مجموعه‌ای از متغیرهای وابسته وجود دارند، درحالی‌که در رگرسیون چندمرحله‌ای تنها یک متغیر وابسته وجود دارد (لیچ، بارت و مورگان، ۲۰۰۵).

تحلیل همبستگی کانونی مستلزم انجام عملیات کانونی^۳ مختلف است و موجب می‌شود تا ضریب همبستگی کانونی بین دو متغیر به حداکثر میزان خود برسد (گارسون، ۲۰۰۸). این تکنیک را می‌توان از لحاظ مفهومی به عنوان روش گام به گام در نظر گرفت که در گام نخست دو بعد (یک بعد از هر مجموعه) که رابطه قوی‌تری در مقایسه با سایر زوج‌ها داشته باشد، انتخاب می‌شود. در گام بعدی زوج دیگری از ابعاد، که دارای قوی‌ترین همبستگی با هم باشد، انتخاب می‌شوند و همین‌طور تا آخر. بیشترین تعداد زوج ابعاد که توسط این روش مشخص می‌شود، با تعداد متغیرهای مجموعه کوچک‌تر برابر است.

رگرسیون چندگانه فقط زمانی می‌تواند مورد استفاده قرار گیرد که یک متغیر وابسته برای تبیین وجود داشته باشد. وقتی تعداد متغیرهای وابسته بیشتر از یکی باشد، تحلیل واریانس چندمتغیره

۱. خاصیت بیشینه‌گی (ماکسیمال) به این معناست که در این تکنیک توابع کانونی با بیشترین مقدار واریانس مشترک از دو

دسته متغیرهای مستقل و وابسته استخراج می‌شوند.

می تواند راه گشا باشد؛ ولی فقط زمانی که متغیرهای مستقل غیرکمی باشند، این در حالی است که تنها روشی که این محدودیت ها را ندارد، تکنیک همبستگی کانونی است (هایر و دیگران، ۱۹۹۸: ۱).

همچنین این روش به تکنیک تحلیل تمایز^۱ نیز شباهت دارد و همانند تحلیل تمایز قادر به استخراج توابع تفکیکی برای ابعاد هر دسته از متغیرهاست، با این تفاوت که در این جا، ابعاد طوری استخراج می شوند که حداکثر همبستگی را با متغیرهای دسته دیگر داشته باشند (هایر و دیگران، ۱۹۹۸: ۴). بنابراین همبستگی کانونی، توابع خطی یا ابعاد مطلوب از هر مجموعه متغیر را طوری شناسایی می کند که حداکثر ارتباط بین دو تابع خطی (مستقل و وابسته) وجود داشته باشد. یکی از خصوصیات متمایز همبستگی کانونی این است که باعث پیدایش و شکل گیری عملیات کانونی چند-مرحله ای می شود. هر عمل کانونی مستقل از عملیات کانونی دیگر انجام می گیرد که این خود گویای وجود ارتباطات مختلف بین مجموعه های متغیرهای مستقل و وابسته است و در کل جنبه های مختلف همبستگی بین دو مجموعه متغیر را آشکار می کند^۲ (اندرسون، ۲۰۰۳).

۲. مفاهیم اصلی تحلیل همبستگی کانونی

۲.۲.۱. متغیرهای کانونی

ترکیب هایی خطی هستند که نماینده جمع بهینه شده دو یا چند متغیر هستند. متغیرهای کانونی یک نوع از متغیرهای پنهان هستند.

۲.۲.۲. ضریب همبستگی کانونی^۳

میزان همبستگی کلی بین دو ترکیب خطی از متغیرهای مستقل و وابسته را نشان می دهد؛ به عبارتی ضریب همبستگی بین دو متغیر کانونی را نشان می دهد (هایر و دیگران، ۱۹۹۸). مجموع مجذورات کلیه ضرایب همبستگی کانونی، همبستگی کانونی ادغام شده^۴ نامیده می شود و نشان می دهد

1. Discriminant analysis

۲. برای مطالعه بیشتر رجوع شود به:

Levine, M.S. 1977. Canonical analysis and factor comparisons. Sage University papers on Quantitative Applications in the Serial Sciences 07-006. Sage Publications, Beverly Hills.

3. Canonical Correlation Coefficient

4. Pooled canonical correlation

که یک مجموعه از متغیرها چه میزان از تغییرات مجموعه متغیرهای دیگر را تبیین یا پیشگویی می‌کند (واریانس مشترک) (شری، ۲۰۰۵: ۴۰).

۲.۲.۳. تابع کانونی^۱

همبستگی یا ارتباط بین دو ترکیب خطی است، که یکی از این ترکیب‌های خطی، از دسته متغیرهای مستقل و دیگری از دسته متغیرهای وابسته حاصل شده است. در هر تابع همبستگی کانونی دو متغیر کانونی وجود دارد. یک متغیر کانونی حاصل ترکیب خطی بهینه از متغیرهای وابسته و دیگری ترکیب خطی بهینه از متغیرهای مستقل است که متغیر کانونی کوواریت^۲ نیز نامیده می‌شود (تک، ۱۹۹۶: ۳۳۶، اندرسون، ۲۰۰۳، گارسون، ۲۰۰۸).

۲.۲.۴. وزن‌های کانونی^۳

این مقدار تعیین‌کننده اهمیت نسبی سهم هر یک از متغیرهای اصلی در ضرایب همبستگی کانونی است. این وزن‌ها مشابه وزن بتاها در تحلیل رگرسیون تفسیر می‌شوند. نسبت وزن هر یک از متغیرها با کنترل سهم متغیرهای دیگر محاسبه می‌شود. یک وزن کانونی برای هر یک از متغیرهای اصلی در هر مجموعه‌ای محاسبه می‌شود، بنابراین اگر مجموعه وابسته شامل ۵ متغیر اصلی باشد و ۳ تابع کانونی (همبستگی کانونی) استخراج شده باشد، در کل ۱۵ وزن کانونی محاسبه خواهد شد (تک، ۱۹۹۶: ۳۳۵).

۲.۲.۵. بارهای کانونی^۴

همبستگی ساده خطی میان متغیرهای اصلی و متغیرهای کانونی متناظر آن‌ها است که همبستگی‌های ساختاری^۵ نیز نامیده می‌شوند. این‌ها را می‌توان به عنوان فاکتورهای بارگذاری نام برد. در واقع، متغیرهای کانونی استخراج شده از هر دسته؛ همانند عوامل یا متغیرهای پنهان در تحلیل عاملی هستند. هر یک از متغیرهای اصلی در هر دسته بر روی متغیر کانونی استخراج شده از همان دسته، دارای بار کانونی است. جدول مربوط به بارهای کانونی را گاهی ساختار عاملی^۶ نیز می‌نامند. مجذور

1. Canonical Function
2. covariate canonical variable
3. Weight Canonical
4. Canonical Loading
5. Canonical Structure Correlation
6. Factor structure

بارهای کانونی، قدرت متغیرهای کانونی برای تبیین متغیرهای اصلی را نشان می‌دهد. بنابراین در مجموع برای بارهای کانونی ۳ نوع کاربرد ذکر شده است:

۱. با توجه به مقدار بارهای متغیرهای اصلی بر روی هر یک از متغیرهای کانونی می‌توان متغیرهای کانونی استخراج شده را نام‌گذاری و تفسیر کرد (دقیقاً مانند تحلیل عاملی، در این‌جا نیز تغییری که بار بیشتری بر متغیر کانونی دارد، سهم بیشتری در نام‌گذاری خواهد داشت) و در نهایت، می‌توان متوجه شد کدام یک از ابعاد مجموعه متغیر مستقل ارتباط بیشتری با ابعاد متغیرهای وابسته دارند. یک قاعده کلی در این زمینه وجود دارد که هر تغییری که بار آن بیش از ۰/۳ است در تفسیر و نام‌گذاری متغیر کانونی دخیل خواهد بود و اگر کمتر باشد، خیر (گارسون، ۲۰۰۸).

۲. اگر مقدار بار کانونی را به توان ۲ برسانیم (مجذور بار کانونی)، میزان واریانس متغیر اصلی که توسط متغیر کانونی توضیح داده شده است، به دست می‌آید.

۳. مجموع مجذورات بارهای کانونی یک دسته از متغیرها، مقدار واریانس کل یک مجموعه از متغیرها را که توسط متغیر کانونی استخراج شده از همان دسته توضیح داده می‌شود، را بیان می‌کند (گارسون، ۲۰۰۸).

۲.۲.۶. بارهای عاملی متقاطع^۱

بارعاملی متقاطع، نشان‌دهنده همبستگی بین متغیر مستقل یا وابسته با متغیر کانونی مخالف آن است به عنوان مثال ممکن است متغیرهای مستقل با متغیرهای کانونی وابسته دارای همبستگی باشد (گارسون، ۲۰۰۸). بارهای متقاطع کانونی جایگزین مناسبی هستند برای بارهای کانونی و وزن‌های کانونی (دایلون و گلدستین، ۱۹۸۴) که ارتباط بین هر متغیر اصلی در هر مجموعه را، با متغیر کانونی استخراج شده از دسته دیگر، نشان می‌دهند.

۲.۲.۷. ریشه‌های کانونی^۲

مجذور ضریب همبستگی کانونی است که در واقع بیانگر تمام ابعاد متعامد موجود در هر دسته از متغیرهاست. از ریشه‌های کانونی جهت ارزیابی این‌که تا چه حد می‌توان یکی از مجموعه متغیرها را

1. Canonical Cross-Loadings
2. Canonical roots

پیش‌بینی کرد، استفاده می‌شود. ریشه‌های کانونی در واقع، تخمینی از مقدار واریانس مشترک بین دو متغیر کانونی هستند. مجذور ضریب همبستگی کانونی، مقدار ویژه نیز نامیده می‌شود^۱ (هایر و دیگران، ۱۹۹۶؛ گارسون، ۲۰۰۸).

۲.۲.۸. شاخص افزونگی^۲

این شاخص مقداری از کل واریانس متغیرهای اصلی در هر مجموعه است که توسط یکی از متغیرهای کانونی استخراج شده از دسته دیگر، تبیین می‌شود. در هر تابع کانونی می‌توان این شاخص را هم برای متغیرهای مستقل و هم برای متغیرهای وابسته تعریف کرد؛ برای مثال شاخص افزونگی در متغیر وابسته، مقداری از واریانس متغیر وابسته را نشان می‌دهد که توسط متغیر کانونی مستقل توضیح داده شده است (شری، ۲۰۰۵: ۴۰). نکته قابل توجه درباره شاخص افزونگی این است که هر چند ریشه‌های کانونی یا همان مجذور همبستگی‌های کانونی، میزانی از واریانس تابع کانونی استخراج شده از مجموعه متغیرهای متناظر را نشان می‌دهد، باین حال نمی‌توان این شاخص را به راحتی به عنوان تبیین‌گر واریانس موجود در متغیرهای وابسته قلمداد کرد؛ چرا که ریشه‌های کانونی نشان‌دهنده میزان واریانس تبیین شده، تابع کانونی استخراج شده از مجموعه متغیرهای مستقل یا وابسته هستند نه میزان واریانس متغیرهای اصلی در هر مجموعه (مارک و پترسون، ۱۹۷۲). این مسأله وقتی آشکارتر می‌شود که بدانیم توابع کانونی استخراج شده، نماینده کاملی برای همه واریانس موجود در متغیرهای اصلی در هر مجموعه نیستند. شاخص افزونگی، آماره‌ای است که این مشکل را مرتفع می‌سازد؛ چرا که این شاخص نه از مجذور همبستگی بین توابع کانونی؛ بلکه از مجموع مجذور همبستگی‌های بین متغیرهای اصلی یک مجموعه با متغیر کانونی مجموعه متناظر حاصل شده است (داگلاس و لائو، ۱۹۶۸).

۲.۲.۹. مجموع ضرایب افزونگی^۳

این شاخص حاصل جمع مقداری از کل واریانس موجود در یک مجموعه از متغیرهاست که توسط همه متغیرهای کانونی استخراج شده از همان مجموعه از متغیرها تبیین می‌شود (داگلاس و لائو، ۱۹۶۸: ۳۴).

1. Eigenvalue
2. Redundancy index
3. Pooled redundancy coefficient

۲.۱۰. آزمون‌های معناداری

برای بررسی معنادار بودن روابط کانونی استخراج شده (روابط میان زوج متغیرهای کانونی)، چهار نوع آزمون معناداری محاسبه می‌شود که هر یک در ادامه توضیح داده می‌شود:

۱- اثر پیلایی: اثر پیلایی از مجموع مجذورات همبستگی‌های کانونی حاصل می‌شود:

$$V = \sum_{i=1}^s r_i^2$$

۲- اثر هتلینگ: بسیار شبیه اثر پیلایی است و روش محاسبه آن چنین است:

$$U = \sum_{i=1}^s \frac{r_i^2}{1 - r_i^2}$$

۳- لامبدای ویلکز که توسط بارتلت پیشنهاد شده است، معناداری اولین همبستگی کانونی را آزمون می‌کند، اگر کمتر از ۰/۵ باشد به این معناست که ارتباط معناداری بین دو دسته از متغیرها وجود دارد. درجه آزادی در این آزمون برابر است با $P, P*Q$ برابر است با تعداد متغیرها در دسته اول و Q برابر است با تعداد متغیرها در دسته دوم. این آماره معناداری اولین همبستگی کانونی را آزمون می‌کند نه لزوماً همبستگی‌های بعدی را^۱. به همین دلیل به آن آزمون تشخیصی بزرگترین ریشه نیز می‌گویند (گارسون، ۲۰۰۸). روش محاسبه مقدار آن نیز به شکل زیر است:

$$\lambda = \prod_{i=1}^s (1 - r_i^2)$$

۴- آماره رویز: این آماره بر اساس بزرگترین ریشه محاسبه می‌شود:

بزرگترین ریشه-۱ / بزرگترین مقدار ویژه^۲

۱. برای تشخیص معنادار بودن سایر همبستگی‌های کانونی باید به جدول تحلیل کاهش بعد (Dimension reduction analysis) در خروجی نرم‌افزار Spss مراجعه شود. در این جدول معناداری سایر ریشه‌های کانونی با توجه به آماره لامبدا و توزیع فیشر مشخص می‌شود. شایان ذکر است در صورتی که اولین همبستگی کانونی معنادار نباشد هیچ همبستگی کانونی دیگر معنادار نخواهد بود.

۲. مقادیر ویژه، نسبتی از واریانس موجود در متغیر کانونی هستند که توسط همبستگی کانونی مربوطه تبیین می‌شوند.

درباره این که کدام یک از چهار آزمون متداول بهترین هستند، توافق وجود ندارد و به نظر می‌رسد، هیچ یک از آن‌ها در مقایسه با دیگری توانمندترین یا مقاوم‌ترین نیست. مقایسه توان‌های این چهار آزمون نشان می‌دهد که اگر واریانس مجموع متغیرهای وابسته در یک ارزش ویژه واحد (در ترکیب خطی واحد) متمرکز شود، مقادیر لامبدای ویلکز، اثر پیلای و اثر هتلینگ تقریباً توانمندی یکسانی دارند. ملاک اساسی‌تر برای انتخاب مشخصه آزمون، میزان مقاوم بودن آزمون تقریبی F در برابر تخطی از مفروضه‌های اساسی کاربرد آن است. برای آن که آزمون F معتبر تلقی شود، فرض می‌شود که: ۱- گروه نمونه به گونه‌ای تصادفی از جامعه مورد مطالعه انتخاب شده است. ۲- مشاهده‌ها مستقل از یکدیگرند. ۳- مشاهده‌ها از توزیع چند متغیری نرمال پیروی می‌کنند. ۴- گروه‌های هر عامل دارای ماتریس‌های واریانس-کوواریانس درون گروهی مشترک هستند (تک، ۱۹۹۶: ۳۳۸؛ هایر و دیگران، ۱۹۹۸: ۸).

۳. منطق بنیادی همبستگی کانونی

برای روشن شدن روند این تکنیک فرض کنید مجموعه متغیرهای پیش‌بینی‌کننده، شامل P متغیر و مجموعه متغیرهای وابسته، شامل q متغیر باشد، متغیرهای X در بین خود دارای همبستگی متقابل هستند و متغیرهای Y نیز در داخل خود همبستگی دارند و همچنین میان متغیرهای X و Y نیز به صورت دوجه دو همبستگی وجود دارد و این همان نقطه تفاوت میان این روش و روش‌های اقتصادسنجی و رگرسیون است؛ زیرا در اقتصادسنجی، متغیرهای پیش‌بینی‌کننده باید از هم مستقل باشند در صورتی که می‌دانیم در دنیای واقع، در عمل چنین چیزی کمتر وجود دارد (کلاتری، ۱۳۸۲).

در این جا برای بیان مفهوم اصلی همبستگی کانونی، از پژوهشی که توسط هتلینگ (۱۹۳۶) انجام شده و روش همبستگی کانونی در آن به کار گرفته شده، استفاده می‌شود. این پژوهش دربرگیرنده نتایج آزمون‌های سرعت خواندن (x_1)، قدرت خواندن (x_2)، سرعت محاسبه (y_1) و قدرت محاسبه (y_2) است که بر روی ۱۴۰ دانش‌آموز کلاس هفتم صورت گرفته است. سؤال مورد نظر این بود که آیا توانایی خواندن (که توسط x_1 و x_2) اندازه‌گیری می‌شود) با توانایی محاسبه (که توسط y_1 و y_2)

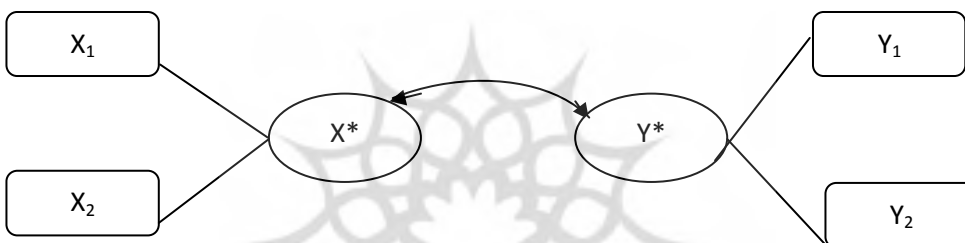
اندازه‌گیری می‌شود) همبستگی دارد یا نه؟ روش پاسخ به این پرسش توسط تحلیل همبستگی کانونی، عبارت است از پیدا کردن ترکیب خطی X_1 و X_2 یعنی:

$$X^* = U = a_1 X_1 + a_2 X_2$$

و نیز ترکیب خطی Y_1 و Y_2 یعنی:

$$Y^* = V = b_1 Y_1 + b_2 Y_2$$

این ترکیبات خطی طوری انتخاب می‌شوند که همبستگی بین U و V به حداکثر مقدار ممکن برسد. همان‌طور که گفته شد؛ این ایده تا حدودی به تجزیه مولفه‌های اصلی مشابه است، با این تفاوت که در این جا همبستگی به جای واریانس، به حداکثر مقدار خود رسانده می‌شود.



شکل ۱- ساختار منطقی تکنیک تحلیل همبستگی کانونی

بعد از استاندارد کردن متغیرهای X_1, X_2, Y_1, Y_2 به منظور دستیابی به واریانس یک، هتلینگ بهترین حالت برای U و V را به شرح زیر به دست آورد:

$$U = -2/78 X_1 + 2/27 X_2, \quad V = -2/44 Y_1 + 1/00 Y_2$$

در این حالت ضریب همبستگی بین U و V ، $0/62$ است. می‌توان مشاهده کرد که U تفاوت بین قدرت خواندن و سرعت خواندن و V تفاوت بین قدرت محاسبه و سرعت محاسبه را برآورد می‌کند، بنابراین به نظری می‌رسد که دانش آموزانی که بین X_1 و X_2 (سرعت و دقت خواندن) تفاوت زیادی وجود دارد، دارای تفاوت زیادی بین Y_1 و Y_2 نیز هستند، در نتیجه این جنبه از خواندن و محاسبه بیشترین همبستگی را نشان می‌دهد.

یادآور می‌شود که در تجزیه (آنالیز) رگرسیون چندگانه، یک متغیر، Y ، با دو یا چند متغیر

X_1, X_2, \dots, X_p در ارتباط است؛ زیرا هدف از این تجزیه، بررسی نحوه ارتباطات Y با متغیرهای X

است. از این نقطه نظر همبستگی کانونی حالت تعمیم یافته رگرسیون چندگانه است که در آن چندین متغیر Y به طور همزمان با چندین متغیر X ارتباط داده می شوند.

در عمل، از مجموعه ای از داده ها، بیشتر از یک جفت متغیر کانونی قابل محاسبه است. اگر تعداد p متغیر X_1, X_2, \dots, X_p و تعداد q متغیر Y_1, Y_2, \dots, Y_q وجود داشته باشد، در این صورت می توان حداقل، p یا q جفت متغیر کانونی داشت؛ یعنی می توان روابط زیر را برقرار کرد:

$$U_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

$$U_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$$

$$U_r = a_{r1}x_1 + a_{r2}x_2 + \dots + a_{rp}x_p$$

$$V_1 = b_{11}y_1 + b_{12}y_2 + \dots + b_{1q}y_q$$

$$V_r = b_{r1}y_1 + b_{r2}y_2 + \dots + b_{rq}y_q$$

$$V_r = b_{r1}y_1 + b_{r2}y_2 + \dots + b_{rq}y_q$$

در این معادلات r حداقل مقدار p و q است. این معادلات طوری انتخاب می شوند که از طرفی همبستگی بین U_1 و V_1 حداکثر باشد و از طرف دیگر U_2 و V_2 با U_1 و V_1 همبستگی نداشته باشند و تا آخر. پس هر یک از این جفت متغیرهای کانونی (U_1, V_1) ، (U_2, V_2) و (U_r, V_r) بعد مستقلی از رابطه بین دو مجموعه از متغیرهای (x_1, x_2, \dots, x_p) و (y_1, y_2, \dots, y_q) را نشان می دهند. اولین جفت (U_1, V_1) بالاترین همبستگی ممکن را دارا بوده و به عنوان مهم ترین معادلات به-شمار می آیند. دومین جفت (U_2, V_2) بیشترین همبستگی را بعد از جفت اول (U_1, V_1) نشان می دهند. در نتیجه در درجه دوم اهمیت قرار دارند و الی آخر (منلی، ۱۳۹۰). شایان ذکر است همبستگی کانونی ادغام شده^۱ مجموع مجذورات کلیه ضرایب همبستگی کانونی است که نشان می دهد، یک مجموعه از متغیرها چه میزان از تغییرات مجموعه متغیرهای دیگر را تبیین یا پیشگویی می کند (کلاتری، ۱۳۸۲). نکته دیگر این که در تحلیل همبستگی کانونی رابطه متغیرها درون هر یک از مجموعه ها از طریق محاسبه وزن های کانونی کنترل می شود؛ یعنی در مجموعه های X و X رابطه هر متغیر با متغیر دیگر کنترل می شود. این کار مانع بروز مسأله هم خطی چندگانه می شود (فرنر و لارکر، ۱۹۸۱: ۳۲).

۴. ملاحظات داده‌ای در تحلیل همبستگی کانونی

تحلیل همبستگی کانونی، عضوی از خانواده تحلیل‌های چندمتغیره است و ضروری است ملاحظات داده‌ای لازم برای تحلیل‌های چندمتغیره در این‌جا نیز رعایت شود. برخی ملاحظات ضروری برای کاربرد تحلیل همبستگی کانونی عبارت‌اند از:

- سطح اندازه‌گیری داده‌ها باید کمی (فاصله‌ای یا نسبی) باشد. البته فومل و لارکر (۱۹۸۱) معتقدند که تکنیک همبستگی کانونی این قابلیت را دارد که متغیرهای رتبه‌ای را نیز وارد تحلیل کند. این قابلیت زمانی کاربرد دارد که داده‌ها آمیزه‌ای از مشخصه‌های کمی و کیفی باشند. در این صورت هر مشخصه کمی به صورت یک متغیر و هر مشخصه کیفی، بر فرض با g صفت کیفی، $g-1$ متغیر دو مقوله‌ای، با مقادیر ظاهری صفر و یک، شناخته می‌شوند؛ علاوه بر این هایر (۱۹۹۶) می‌گوید که شرط نرمال بودن متغیرهای مستقل و وابسته، خیلی ضروری نیست و می‌توان متغیرهای غیرنرمال را نیز وارد تحلیل کرد. البته وی معتقد است نرمال بودن متغیرها اجازه می‌دهد تا همبستگی بین آن‌ها به حداکثر ممکن برسد. با این حال این انعطاف باعث می‌شود که این تکنیک این قابلیت را داشته باشد که داده‌هایی که کمی نیستند را به متغیرهای تصنعی دو مقوله‌ای (صفر و یک) تبدیل و وارد مدل تحلیلی کند.

- برای انجام آزمون‌های معناداری نیاز به داده‌هایی با توزیع نرمال چندمتغیره است و از آن‌جا که دستیابی به توزیع نرمال چندمتغیره مشکل است، بنابراین راحت‌ترین راه این است که داده‌ها تا جایی که امکان دارد به فرض نرمالی تک‌متغیری نزدیک شود (هایر، ۱۹۹۶: ۶).

- ارتباط بین متغیرها باید خطی باشد. شرط خطی بودن رابطه، هم برای ارتباط بین متغیرهای اصلی با متغیرهای کانونی و هم برای ارتباط بین متغیرهای کانونی باید رعایت شود (شری، ۲۰۰۵: ۴۰).

۱. توزیع نرمال چندمتغیره توزیعی است که در آن متغیرهای x_1, \dots, x_n با هم دارای تابع توزیع احتمال نرمال هستند. در حالت دومتغیره تابع توزیع احتمال به شکل زیر است:

$$g(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} e^{-(\mathbf{y}-\boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu})/2}$$

شرط لازم برای این‌که چند متغیر دارای توزیع احتمال نرمال چندمتغیره باشند این است که همه آن‌ها دارای توزیع نرمال باشند؛ ولی این شرط کافی نیست. برای اطلاعات بیشتر رجوع شود به کتاب:

Rencher.A.Alvin,(2002), Methods of Multivariate Analysis, Wiley Interscience Publication

- مجموعه متغیرهای مستقل و مجموعه متغیرهای وابسته دارای توزیع نرمال چندمتغیری باشند (شری، ۲۰۰۵: ۴۰).

- متغیرها همگن باشند؛ یعنی واریانس موجود در هر دو دسته از متغیرها، تقریباً یکسان باشند.

- حجم نمونه طوری تعیین شود که نسبت مناسبی بین تعداد متغیرها و تعداد مشاهدات برقرار شود. توضیح این که به خاطر قابلیت تحلیلی بالای تکنیک همبستگی کانونی، ممکن است محققان وسوسه شوند تعداد زیادی متغیر را اندازه‌گیری کنند، این در حالی است که افزایش تعداد متغیرها اغلب باعث طولانی شدن ابزار اندازه‌گیری و در نتیجه کاهش حجم نمونه خواهد شد و این مسأله خود باعث می‌شود که ضریب همبستگی به درستی برآورد نشود، همچنین اگر حجم نمونه بیش از اندازه لازم باشد، ممکن است آزمون ضریب همبستگی کانونی به اشتباه معنی‌دار شود. توصیه می‌شود به ازای هر متغیر حداقل ۱۰ مشاهده صورت گیرد تا این توازن برقرار شود (مارک و پترسون، ۱۹۷۲: ۲۴).

علاوه بر ملاحظات داده‌ای مذکور لازم است متغیرهای مورد بررسی به لحاظ نظری به دو دسته مستقل و وابسته قابل تقسیم باشند یا مفهومی که به عنوان متغیر وابسته مورد بررسی قرار می‌گیرد، متشکل از چندین بعد باشد، به گونه‌ای که هر یک از ابعاد آن به عنوان یک متغیر وابسته محسوب شوند. هدف از انجام تحلیل نیز یافتن روابط همبستگی باشد نه روابط علی^۱.

۵. اهداف کاربردی همبستگی کانونی

تحلیل همبستگی کانونی سه هدف عمده را دنبال می‌کند:

- ۱) تعیین این که آیا دو مجموعه از متغیرهای مستقل و وابسته اندازه‌گیری شده، با هم رابطه دارند یا خیر. همچنین تعیین شدت و جهت این رابطه.
- ۲) استخراج مجموعه‌ای از وزن‌ها برای هر مجموعه از متغیرهای مستقل و وابسته، به طوری که از این وزن‌ها تعدادی ترکیب خطی حاصل شود که درعین حال حداکثر ارتباط را با ترکیب خطی

۱. رابطه متغیر X و Y زمانی علی نامیده می‌شود که X علت حضور Y باشد. در رابطه همبستگی تغییرات X و Y به هم وابسته است؛ ولی X الزاماً علت Y نیست.

حاصل شده از متغیرهای دسته دوم داشته باشد، همچنین توابع خطی دیگری پس از اولین تابع خطی استخراج می شود تا واریانس توضیح داده نشده باقی مانده را مستقل از ترکیب خطی اول، تبیین کند. (۳) آزمون این فرض که آیا روابط مشاهده شده بین متغیرهای مستقل و وابسته قابلیت تعمیم به جمعیت آماری را دارد یا خیر. این هدف توسط چهار آماره معناداری (اثر هتلینگ، لامبدای ویلکس، ضریب روی و اثر پیلائی) محقق می شود (شری، ۲۰۰۵؛ اندرسون، ۲۰۰۳؛ کلارک، ۱۹۷۵).

۶. معرفی چند کاربرد از تحلیل همبستگی کانونی

اینک به منظور توضیح موارد کاربرد تکنیک همبستگی کانونی به طور خلاصه تعدادی از پژوهش های شناخته شده که از تکنیک مزبور برای تحلیل داده های خود استفاده کرده اند، معرفی می شوند:

مثال ۱: در مطالعه ای که توسط تعدادی از محققان روسی انجام شد، تلاش شد تا با مطالعه ویژگی های ۴۷ کشور، صحت نظریه توکویل، مبنی بر وجود رابطه بین نابرابری اقتصادی و بی ثباتی سیاسی، آزمون شود. پنج شاخصی که برای اندازه گیری نابرابری اقتصادی مورد استفاده قرار گرفته بودند، عبارتند از: (۱) نحوه تقسیم زمین های کشاورزی (۲) ضریب جینی درآمد (GINI) (۳) درصد کشاورزانی که به صورت مزدبگیر روی زمین های کشاورزی کار می کنند (۴) شاخص تولید ناخالص ملی (GNP) و (۵) درصد کشاورزان، همچنین بی ثباتی سیاسی در کشورهای مورد بررسی نیز توسط (۴) شاخص زیر اندازه گیری شد (۱) میزان وقوع جنگ های درونی (۲) بی ثباتی در دموکراسی (۳) بی ثباتی در رهبری (۴) درصد گروه های مخالف درون کشور.

محققان برای تحلیل داده های گردآوری شده از ۴۷ کشور نمونه، از تکنیک همبستگی کانونی استفاده کردند. اولین ماتریس داده ها (T_{111}) که مربوط به همبستگی های درون دسته متغیرهای مستقل (X ها) بودند و همچنین دومین ماتریس همبستگی (T_{222}) که مربوط به همبستگی های درون دسته متغیرهای وابسته (Y ها) بودند، رابطه متقابل متغیرها (چیزی شبیه مسأله هم خطی چندگانه^۲ در

1. Within-correlations
2. Multi collinearity

رگرسیون چندگانه) را بررسی می‌کنند. پیش فرض تکنیک همبستگی کانونی این است که این روابط از نوع خطی هستند.

سومین ماتریس همبستگی Γ_{xy} مربوط به همبستگی‌های بین متغیرهای مستقل و متغیرهای وابسته است که به آن همبستگی‌های بین متغیری^۱ می‌گویند و مانند همبستگی پیرسون تفسیر می‌شوند.

جدول ۱- ماتریس همبستگی بین متغیرهای دسته اول و دوم

مأخذ: (تک، ۱۹۹۶)

<u>Correlation between set-1 and set-2</u>				
	y_1	y_2	y_3	y_4
x_1	Γ_{11}	Γ_{12}	Γ_{13}	Γ_{14}
x_2	Γ_{21}	Γ_{22}	Γ_{23}	Γ_{24}
x_3	Γ_{31}	Γ_{32}	Γ_{33}	Γ_{34}
x_4	Γ_{41}	Γ_{42}	Γ_{43}	Γ_{44}
x_5	Γ_{51}	Γ_{52}	Γ_{53}	Γ_{54}

در این بررسی از هر دسته دو متغیر کانونی که دو زوج کانونی را تشکیل می‌دهند، استخراج شدند که عبارت‌اند از: (x_1^*, y_1^*) و (x_2^*, y_2^*)

هر یک از این متغیرهای کانونی، حکم عوامل پنهانی را دارند که در تحلیل عاملی، از متغیرهای مشاهده‌شده استخراج می‌شوند. همبستگی‌های بین متغیرهای اصلی در هر مجموعه و متغیرهای کانونی استخراج‌شده نیز همانند بارهای عاملی، در تحلیل عاملی تفسیر می‌شوند و بارهای کانونی نامیده می‌شوند. ماتریس شماره ۲ به ضرایب همبستگی بین متغیرهای اصلی در هر دو مجموعه و هر یک از متغیرهای کانونی استخراج‌شده، مربوط است. این ضرایب را همبستگی‌های ساختاری نیز می‌نامند.

1. Between-correlation

جدول ۲- ماتریس همبستگی بین متغیرهای اصلی و متغیرهای کانونی

مأخذ: (تک، ۱۹۹۶)

	Factor structure, for set-2		Factor structure for set-2	
	x_1^*	x_2^*	y_1^*	y_2^*
x_1	$R_{x_1x_1^*}$	$R_{x_1x_2^*}$	y_1	$R_{y_1y_1^*}$ $R_{y_1y_2^*}$
x_2	$R_{x_2x_1^*}$	$R_{x_2x_2^*}$	y_2	$R_{y_2y_1^*}$ $R_{y_2y_2^*}$
x_3	$R_{x_3x_1^*}$	$R_{x_3x_2^*}$	y_3	$R_{y_3y_1^*}$ $R_{y_3y_2^*}$
x_4	$R_{x_4x_1^*}$	$R_{x_4x_2^*}$	y_4	$R_{y_4y_1^*}$ $R_{y_4y_2^*}$
x_5	$R_{x_5x_1^*}$	$R_{x_5x_2^*}$		

در نهایت در این تحقیق محققان به دو ترکیب خطی از متغیرهای کانونی استخراج شده به شرح زیر دست یافته‌اند: (تک، ۱۹۹۶: ۳۳۶)

$$x^*_1 = 0.547x_1 + 0.836x_2$$

$$x^*_2 = 0.837x_1 - 0.549x_2$$

$$y^*_1 = -0.359y_1 + 1.116y_2$$

$$y^*_2 = 1.072y_1 - 0.182y_2$$

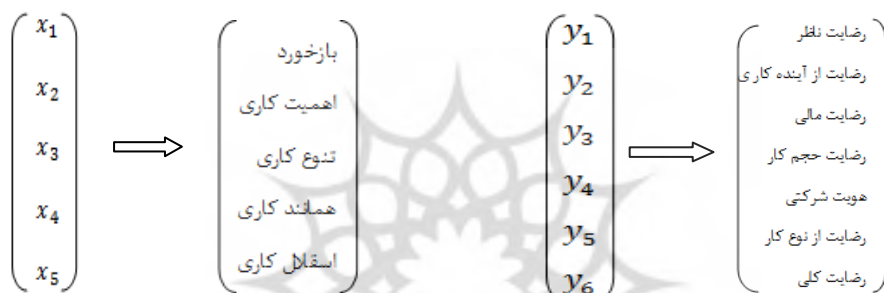
این چهار متغیر کانونی استخراج شده، دو زوج کانونی را تشکیل می‌دهند، بنابراین محققان در نهایت به دو ریشه کانونی رسیدند که هر یک مقداری از واریانس متغیر کانونی استخراج شده از مجموعه متغیرهای وابسته (متغیرهای بی‌ثباتی سیاسی) را توضیح می‌دهد. این ریشه‌ها را مقادیر ویژه نیز می‌نامند که در جدول (۳) نشان داده شده است.

جدول ۳- مقادیر ویژه و همبستگی‌های کانونی

مأخذ: (تک، ۱۹۹۶)

همبستگی کانونی	مجدور همبستگی کانونی	درصد تراکمی	درصد	مقدار ویژه	ریشه کانونی
۱	۱/۵۸۷	۸۴/۲۱۵	۸۴/۲۱۵	۰/۸۷۳	۰/۶۱۳
۲	۰/۲۹۸	۱۵/۷۸۵	۱۰۰	۰/۴۷۹	۰/۲۲۹

مثال ۲: دون هام^۱ (۱۹۷۷) به بررسی اثرات ساختار سازمان یافته و رضایت شغلی پرداخته است. این پژوهش، تلاش می کند حدى را بررسی کند که در آن میزان اندازه های رضایت شغلی و ویژگی های شغلی با هم ارتباط معنادار داشته باشند. با استفاده از ابزار نمونه گیری دون هام، اندازه هایی از $p=5$ متغیر ویژگی شغلی و $q=7$ متغیر رضایت شغلی را برای $n=784$ شعبه از شعبات یک شرکت کالای خرده فروشی بزرگ به دست آورده است. سؤال اصلی پژوهش این است که آیا اندازه های رضایت شغلی با ویژگی های شغلی ارتباط دارد؟ متغیرهای ویژگی های شغلی با X و متغیرهای رضایتمندی شغلی با Y ، به صورت زیر تعریف می شوند:



تعداد متغیرهای دسته کوچک تر ۵ است:

$$\text{Min}(p,q)=\text{min}(5,7)=5$$

بنابراین، ۵ همبستگی کانونی، به عبارتی ۵ ترکیب خطی خواهیم داشت؛ برای مثال اولین زوج متغیر کانونی، عبارت است از:

$$U_1 = 0.42y_1 + 0.21y_2 + 0.17y_3 + 0.2y_4 + 0.44y_5$$

$$V_1 = 0.42x_1 + 0.22x_2 - 0.3x_3 + 0.1x_4 + 0.29x_5 + 0.52x_6 - 0.12x_7$$

و اولین و تنها همبستگی کانونی معنادار استخراج شده برابر است با: $P_1 = 0.55$

نکته قابل ذکر این است که اولین همبستگی کانونی که از اولین زوج کانونی استخراج شده به دست می آید، بزرگترین ضریب همبستگی بین دو دسته متغیرهاست و توان دوم آن که ریشه کانونی نامیده می شود، بیشترین مقدار واریانس متغیرهای مستقل را توضیح می دهد. مجموع این ریشه ها که ریشه-

های کانونی ادغام شده نامیده می شوند، مجموع واریانس‌ها که توسط زوج‌های کانونی استخراج شده توضیح داده می شوند را نشان می دهد.

شایان ذکر است عملیات کانونی پس از رسیدن به اولین ضریب همبستگی حاصل از اولین زوج کانونی، متوقف نمی شود؛ بلکه زوج‌های کانونی بعدی، مستقل از زوج کانونی اول و البته با ضرایب کانونی کمتر، استخراج شده و واریانس باقی مانده را توضیح می دهند. این زوج‌ها به تعداد متغیرهای موجود در دسته کوچکتر استخراج می شوند؛ ولی لزوماً همه آن‌ها معنادار نیستند، بنابراین با توجه به آماره‌های معناداری که توضیح داده خواهد شد، تنها همبستگی‌های کانونی‌ای تفسیر می شوند که معنادار باشند.

۷. تحلیل همبستگی کانونی با spss: یک مثال کاربردی

اینک بر اساس داده‌های یک تحقیق پیمایشی در مشهد^۱، مراحل استفاده از تکنیک تحلیل همبستگی کانونی با استفاده از نرم‌افزار SPSS معرفی و خروجی‌های مربوط تفسیر می شود. در این مثال، هدف یافتن ارتباط بین دو مجموعه متغیر است: متغیر تحریک‌جویی به‌عنوان متغیر مستقل با دو بعد تنوع‌طلبی (X_1) و هیجان‌طلبی (X_2) و متغیر رفتار پرخطر رانندگی به‌عنوان متغیر وابسته با سه بعد خطا (Y_1)، لغزش (Y_2) و تخلف (Y_3). فرضیه مورد بررسی این است که «با تشدید تحریک‌جویی رانندگان، رفتارهای پرخطر آن‌ها افزایش پیدا می کند».

در برنامه SPSS، منوی مجزایی برای اجرای دستور تحلیل همبستگی کانونی وجود ندارد. برای اجرای این تحلیل دو روش وجود دارد. یک روش، افزودن منوی تحلیل همبستگی کانونی به‌عنوان یک Add-on به نرم‌افزار SPSS نصب شده است. با این کار، دستور تحلیل همبستگی کانونی به زیرمنوی correlate در منوی Analyze اضافه می شود. برای این کار باید از منوی Add-ons و سپس Programmability Extension وارد سایت IBM شده و بسته مخصوص تحلیل همبستگی کانونی را

۱. پیمایش مذکور در سال ۱۳۹۰ با نمونه‌ای تصادفی متشکل از ۱۶۰ راننده خودرو در مشهد، توسط نگارنده اول این نوشتار (یوسفی) انجام گرفته است.

دانلود کرد، سپس نرم‌افزار Python را نصب کرده و پس از انجام تنظیمات مربوط، داده‌ها را تحلیل کنیم. به دلیل پیچیده بودن این روش به خصوص برای کاربران غیر حرفه‌ای، ترجیح داده می‌شود که از روش ساده‌تر؛ یعنی نوشتن دستورها در Syntax در SPSS استفاده شود. دستور لازم برای اجرای تحلیل همبستگی کانونی کوتاه و مختصر است. پس از وارد کردن داده‌ها در SPSS، باید یک فایل Syntax را باز کرده و دستور زیر را در آن اجرا کرد:

```
Manova x1 x2 with y1 y2 y3
/ discrim all alpha(1)
/ Print=sign(eigen dim).
```

نتایج تحلیل همبستگی کانونی بین دو مجموعه متغیر تحریک‌جویی و رفتار پرخطر رانندگی متناسب با مقاصد تحلیل در جداول ۴ تا ۷ درج شده است که در زیر به مهم‌ترین آن‌ها اشاره می‌شود.^۱ یادآوری این نکته لازم است که در هر تابع کانونی همواره دو متغیر کانونی وجود دارد که یکی به عنوان متغیر وابسته و دیگری به عنوان متغیر مستقل منظور می‌شود؛ علاوه بر این، تکیه تحلیل همبستگی کانونی بر توابعی است که بین زوج متغیرهای کانونی آن‌ها، همبستگی معناداری وجود داشته باشد.^۲

مطابق جدول ۴ همبستگی میان متغیر کانونی تحریک‌جویی و رفتار پرخطر رانندگی حدود ۰/۵۳ است. مجذور این ضریب (۰/۲۸)، حاکی از واریانس مشترک بین این دو مجموعه متغیر (ترکیب خطی) است. بر این اساس با آگاهی از متغیر کانونی تحریک‌جویی، ۲۸ درصد از تغییرات متغیر کانونی رفتار پرخطر رانندگی قابل پیش‌بینی خواهد بود.

۱. مثال تحلیلی حاضر با اندک تغییراتی از مقاله ذیل گرفته شده است:

یوسفی، علی؛ قاسمی ادکان، فاطمه؛ تقوی گرجی، حسین. (۱۳۹۱). تحلیل چندگانه رفتارهای پرخطر رانندگی در شهر مشهد. *مجله جغرافیا و مخاطرات محیطی* (۴) ۶۵-۷۸.

۲. با توجه به این که تعداد توابع کانونی همواره برابر با تعداد متغیرهای مجموعه کوچک‌تر است و چون در این‌جا مجموعه کمتر، متغیر تحریک‌جویی است که خود دارای دو جزو است - لذا تنها دو تابع کانونی محتمل است که تنها یک تابع آن معنادار است. بر این اساس چون دومین زوج از متغیرهای کانونی، همبستگی معناداری با هم ندارند، در ادامه از تشریح تابع دوم کانونی، صرف نظر و جداول مربوط نیز حذف می‌شود.

جدول ۴- نتایج محاسبه همبستگی کانونی تحریک جویی و رفتار پرخطر رانندگی (تعداد آزمودنی = ۱۶۰ نفر)

شماره توابع کانونی	مقادیر ویژه	همبستگی کانونی	مجذور همبستگی	سطح معنی داری
۱	۰/۴۰۷	۰/۵۳	۰/۲۸	۰/۰۰۰
۲	۰/۰۱۰	۰/۱۰	۰/۰۱	۰/۴۴۴

وزن‌های کانونی نشان‌دهنده سهم نسبی هر یک از متغیرهای اصلی در میزان همبستگی کانونی و بارهای کانونی نیز نشان‌دهنده سهم هر متغیر اصلی در تشکیل متغیر کانونی مربوط است. نتایج محاسبه وزن‌ها و بارهای کانونی برای دو مجموعه متغیر رفتار پرخطر رانندگی و تحریک‌جویی مطابق جداول (۵) و (۶) محاسبه شده که در ذیل به نتایج عمده آن اشاره می‌شود:

- وزن‌های کانونی: در تابع رفتار پرخطر رانندگی، بیشترین وزن (۰/۷۱-) یا اهمیت نسبی مربوط به متغیر خطا و در تابع تحریک‌جویی مربوط به متغیر شدت هیجان در رانندگی (۰/۶۳-) است. در حالت نخست با یک واحد افزایش در خطای رانندگی، میزان همبستگی دو تابع مذکور به اندازه ۷۱ صدم واحد کاهش و در حالت دوم با یک واحد افزایش در شدت هیجان رانندگی، میزان همبستگی دو تابع مذکور ۶۳ صدم واحد کاهش می‌یابد. این در شرایطی است که تأثیر سایر متغیرها در مجموعه مربوط، ثابت نگه داشته شود.

- بارهای کانونی: در تشکیل متغیر کانونی رفتار پرخطر رانندگی بیشترین سهم مربوط به تخلف (۰/۸۹-) و در متغیر کانونی تحریک‌جویی مربوط به هیجان‌طلبی رانندگان (۰/۹۲-) است. چنان‌چه با تأکید بر مقدار بیشتر بار کانونی در هر مجموعه، متغیر کانونی مربوط را نام‌گذاری کنیم در این صورت می‌توان متغیر کانونی رفتار پرخطر رانندگی را «رفتار پرخطر متخلفانه رانندگی» و متغیر کانونی تحریک‌جویی را «تحریک‌جویی هیجان‌طلبانه رانندگی» نام‌گذاری کرد.

جدول ۵- وزن و بار کانونی اجزا در تابع رفتار پرخطر رانندگی (تعداد آزمودنی = ۱۶۰ نفر)

اجزای تابع	وزن کانونی (b)	بار کانونی
خطا	-۰/۷۱	-۰/۸۰
تخلف	-۰/۶۶	-۰/۸۹
لغزش	-۰/۴۴	-۰/۳۵

جدول ۶- وزن و بار کانونی اجزا در تابع تحریک جویی (تعداد آزمودنی = ۱۶۰ نفر)

اجزای تابع	وزن کانونی (a)	بار کانونی
تنوع طلبی	-۰/۴۸	-۰/۸۶
شدت هیجان	-۰/۶۳	-۰/۹۲

باتوجه به نتایج حاصل از محاسبه ضریب همبستگی (ρ) میان دو متغیر کانونی تحریک جویی (X^*) و رفتار پرخطر رانندگی (Y^*) و همچنین وزن های کانونی (a, b) مربوط، نمودار علی تجربی این دو تابع را می توان مطابق شکل (۲) ترسیم کرد.



شکل ۲- نمودار علی تجربی رفتارهای پرخطر رانندگی و تحریک جویی

۸ نتیجه گیری

وارد کردن همزمان چند متغیر وابسته در مدل تحلیلی، و همچنین در نظر گرفتن تمامی ابعاد متغیرها و مفاهیم مورد بررسی، در تحلیل داده های تحت پردازش، نقطه قوت اصلی تکنیک همبستگی کانونی است. این تکنیک با فراهم کردن امکان تبیین همزمان چند متغیر وابسته، علاوه بر کاهش میزان خطای نوع اول که از کاربرد چندین رگرسیون به صورت مجزا ناشی می شود، با ارائه توابع خطی از هر دسته از متغیرها، تحت عنوان زوج های کانونی، بیشترین واریانس مشترک بین متغیرها را استخراج می کند. از آنجایی که در تحقیقات علوم اجتماعی به طور غالب تمایل داریم اثر همزمان چند متغیر مستقل را بر چند متغیر وابسته بسنجیم و یا این که مفاهیم مورد بررسی معمولاً سازه هایی دارای چندین بعد هستند، تکنیک همبستگی کانونی، ابزار مفید و قدرتمندی برای بررسی این نوع روابط است. تکنیک همبستگی کانونی در اصل یک تکنیک توصیفی است که به این سؤال پاسخ می دهد که چه نوع ارتباطی بین دو مجموعه از متغیرها وجود دارد و قدرت و ماهیت این ارتباطات چگونه است؟ این

روش به خصوص زمانی که پژوهشگر اطلاعات ابتدایی اندکی از متغیرها و روابط آنها دارد، بسیار مفید است. این روش از جهاتی به رگرسیون چندگانه، تحلیل تمایزی و تحلیل عاملی شباهت دارد. با این حال، همانند دیگر روش‌های آماری چندمتغیره، نتایج تحلیل همبستگی کانونی نیز باید در معرض روش‌های اعتبارسنجی، برای کسب اطمینان از نتایج، گذاشته شود. مستقیم‌ترین روش این است که دو زیرمجموعه از داده‌ها (اگر حجم نمونه اجازه می‌دهد) تشکیل شود و هر یک از مجموعه‌ها به‌طور جداگانه مورد تجزیه و تحلیل همبستگی کانونی قرار بگیرد، سپس نتایج را می‌توان در توابع کانونی، بارگذاری کانونی و مانند آن مورد مقایسه قرار داد. اگر تفاوت‌های عمده‌ای مشاهده شد، محقق باید بررسی بیشتری برای اطمینان از نتایج خود انجام دهد (گیتینز، ۱۹۸۵). روش دیگر، ارزیابی حساسیت نتایج نسبت به حذف یک متغیر وابسته یا مستقل است؛ چون همبستگی کانونی ضریب همبستگی را ماکزیمم می‌کند و ممکن است وزن‌های کانونی با حذف یک متغیر تغییر کنند، برای اطمینان از ثبات نتایج باید ضریب همبستگی کانونی چندگانه محاسبه شود و هر بار یک متغیر وابسته یا مستقل حذف شود (هایر و دیگران، ۱۹۹۸: ۱۶).

با این حال، چون روش‌های تشخیصی زیادی برای این امر وجود ندارد، باید به محدودیت‌های تکنیک توجه شود که بیشترین تأثیر را بر نتایج و تفسیر می‌گذارند. محدودیت‌های تحلیل همبستگی کانونی عبارت‌اند از:

- ۱- همبستگی کانونی نشان‌دهنده واریانس مشترک مجموعه‌ای از متغیرهایی است که توسط ترکیب‌های خطی استخراج شده‌اند نه همه واریانس موجود در متغیرهای اصلی.
- ۲- وزن‌های کانونی که در محاسبه توابع کانونی مشتق شده‌اند، مقدار زیادی بی‌ثباتی دارند.
- ۳- تفسیر متغیرهای کانونی ممکن است سخت باشد؛ چرا که آنها برای به‌حداکثر رساندن رابطه محاسبه شده‌اند نه برای تفسیر ابعاد پنهان دسته‌ای از متغیرها؛ مانند تحلیل عاملی. ضمن این که چرخش در تحلیل همبستگی کانونی همانند تحلیل عاملی توسط SPSS پشتیبانی می‌شود؛ ولی تامپسون (۱۹۸۴) معتقد است که اگرچه این کار به ضرایب ساختاری و ضرایب کانونی ساده‌تری منجر می‌شود؛ ولی به‌شدت با منطق بنیادین همبستگی کانونی مبنی بر تفکیک متغیرهای مستقل و

وابسته از یکدیگر، تناقض دارد؛ زیرا در تحلیل عاملی این تفکیک رعایت نمی‌شود و بر همین اساس عمل چرخش در آنجا مجاز است (گارسون، ۲۰۰۸).

در پایان یادآور می‌شود که اگرچه تحلیل همبستگی کانونی نسبت به مفروضات اولیه آماری نظیر خطی بودن، نرمال بودن، هم‌وابستگی و هم‌خطی چندگانه، چندان حساسیتی ندارد؛ اما رعایت آن‌ها، تفسیر نتایج را آسان‌تر و رضایت‌بخش‌تر می‌کند.

کتاب‌نامه

۱. بی اف. جی، منلی. (۱۳۹۰). *آشنایی با روش‌های آماری چندمتغیره*. ترجمه پرویز نصیری و پریسا حاج قربانی. تهران: نشر پیام‌رسان.
۲. بی اف. جی، منلی. (۱۳۷۳). *آشنایی با روش‌های آماری چندمتغیره*. ترجمه محمد مقدم، سید ابوالقاسم محمدی شوطی و مصطفی آقایی سربزه. تبریز: انتشارات پیشتاز علم.
۳. جانسون، ریچارد آ؛ ویچرن، دین دبلیو. (۱۳۷۹). *تحلیل آماری چندمتغیره کاربردی*. ترجمه حسین نیرومند. مشهد: دانشگاه فردوسی مشهد.
۴. خوش‌سیما، غلامرضا. (۱۳۸۴). «تجزیه و تحلیل همبستگی متعارف چابکی و عملکرد استراتژیک سازمان‌های تولیدی». *مجله مدیریت صنعتی*. (۱۱). صص ۱۱۵-۱۴۳.
۵. رضائی، عبدالحمید؛ سلطانی، افشین. (۱۳۸۲). *مقدمه‌ای بر رگرسیون کاربردی*. اصفهان: دانشگاه صنعتی اصفهان. مرکز نشر.
۶. ساعی، علی. (۱۳۸۶). *روش تحقیق در علوم اجتماعی (با رهیافت عقلانیت انتقادی)*. تهران: نشر سمت.
۷. سرفه، کریم؛ امینی، فاطمه. (۱۳۸۸). *اصول و روش‌های تجزیه و تحلیل چندمتغیره*. تهران: انتشارات دانش-پرور.
۸. سریواستوا، کارتر. (۱۳۷۰). *آمار چندمتغیره کاربردی*. ترجمه ناصر اقامی. ابوالقاسم بزرگ‌نیا. مشهد: آستان قدس رضوی. بنیاد فرهنگی رضوی.
۹. فراهانی، حجت‌الله؛ عریضی، حمیدرضا. (۱۳۸۴). *روش‌های پیشرفته پژوهش در علوم انسانی*. اصفهان: جهاد دانشگاهی. واحد اصفهان.
۱۰. فرشادفر، عزت‌الله. (۱۳۸۴). *اصول روش‌های آماری چندمتغیره*. کرمانشاه: انتشارات طاق بستان دانشگاه رازی.

۱۱. کانتی ماردیا جان کنت، جان بی.بی. (۱۳۷۶). تحلیل چندمتغیره. ترجمه محمد مهدی طباطبایی. ویراستار: علی عمیدی. تهران: مرکز نشر دانشگاهی.
۱۲. کرلینجر، فردان؛ الازار جی، پدهازور. (۱۳۶۶). رگرسیون چندمتغیره در پژوهش رفتاری. مترجم: حسن سرایی. جلد دوم. تهران: انتشارات مرکز نشر دانشگاهی.
۱۳. کلاتری، خلیل. (۱۳۸۲). پردازش و تحلیل داده‌ها در تحقیقات اجتماعی اقتصادی. تهران: نشر شریف.
۱۴. هومن، حیدرعلی. (۱۳۸۰). تحلیل داده‌های چندمتغیری در پژوهش رفتاری. تهران: نشر بیک فرهنگ.
۱۵. یوسفی، علی؛ قاسمی ادکان، فاطمه؛ تقوی گرگی، حسین. (۱۳۹۱). «تحلیل چندگانه رفتارهای پر خطر رانندگی در شهر مشهد». *مجله جغرافیا و مخاطرات محیطی*. (۴). صص ۶۵-۷۸.
16. Alpert, Mark I. and Peterson, Robert A. (1972). On the Interpretation of Canonical Analysis. *Journal of Marketing Research* 9 (May). p 187
17. Anderson. T. W. (2003). *An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics)*. New York: Wiley InterScience
18. Bagozzi, R. P; Fornell, C. & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*. 16. Pp 437-454.
19. Bartlett, M.S. (1947). The general canonical correlation distribution. *Ann Mathematical statistics*. 18. pp 1-17.
20. Clarke, D. (1975). *Understanding canonical correlation analysis Concepts and techniques in modern geography 3*. Geo.Abstacts. Norwich: UK.
21. Constantine, A. G; James, A.T. (1958). On the general canonical correlation distribution. Division of Mathematical Statistics. G.S.I.R.O.
22. Douglas. S; Love. W. (1968). A General Canonical Correlation Index. *Psychological Bulletin*. 70. Pp 160-163.
23. Dillon, W. R., and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. New York: Wiley.
24. Garson, G. D. (2008). Canonical correlation. Retrieved July 4, 2009 from <http://faculty.chass.ncsu.edu/garson/PA765/canonic.htm>.
25. Gittins, R. (1985). *Canonical Analysis: a Review with Applications in Ecology*. Berlin: Springer-verlage.
26. Green, P. E., and Douglas Carroll, J. (1978). *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.

27. Hotelling, H. (1963). Relations between two sets of variables. *Biometrika*. 28 pp. 321-377.
28. Hardoon, R. D; Szedmak, S; Shawe-Taylor, J. (2003). *Canonical correlation analysis. An overview with application to learning methods*. London: Department of computer of London.
29. Levine, M.S. (1977). Canonical analysis and factor comparisons. *Sage University papers on Quantitative Applications in the Serial Sciences* 07-006. Sage Publications. Beverly Hills.
30. Nancy L. Leech; Karen C. Barrett; and George A. Morgan. (2005). SPSS for Intermediate Statistics: Use and Interpretation. *Lawrence Erlbaum Psychological Bulletin*. 70. Pp 160-63
31. Rencher, A. Alvin. (2002). *Methods of Multivariate Analysis*. Wiley Interscience Publication
32. Tacq, J. J. A. (1996). *Multivariate analysis techniques in social science research. from problem to analysis* / Jacques Tacq. Thousand Oaks, CA: SAGE Publications.
33. Thompson, B. (1985). *Canonical Correlation Analysis: Uses and Interpretations*. Sage Publications. Thousand Oaks. CA.