روش‌های مناسب برای شناسایی داده‌های پرت و نحوه‌ی برخورد با آن در آمارگیری‌های از هزینه و درآمد خانوار

مجری طرح: مجتبی گنجعلی

سمانه افتخاری

شهاب جولانی

احسان بهرامی سامانی

گروه پژوهشی طرح‌های فنی و روش‌های آماری

پژوهشکده‌ی آمار

۱۳۸۷

پیشگفتار

تحلیل رگرسیونی یک ابزار مهم آماری است که به طور معمول در اکثر علوم مورد استفاده قرار می‌گیرد. اغلب، فرض‌های مورد علاقه‌ای بر روی ساختار رگرسیونی به منظور تسهیل ساختن و محاسبه مدل در نظر گرفته می‌شود. برخی از فرض‌های معمولی که بر رگرسیون نهاده می‌شود شامل نرمال بودن متغیر پاسخ و همگونی واریانس پاسخ است. به همین دلیل بیشتر اوقات پیش از مدل بندی داده‌ها، نیازمند استفاده از انواع تبدیلات بر روی پاسخ مورد علاقه به منظور رسیدن به فرض نرمال بودن هستیم (باکس و کاکس، ۱۹۶۴). همچنین همگونی واریانس نیز بایستی پیش از ورود به مرحلهٔ تحلیل داده‌ها مورد بررسی قرار گیرد. یک روش مفید می‌تواند بررسی تغییرات پاسخ در سطوح مختلف متغیرهای کمکی مدل باشد (کوک و ویسبرگ، ۱۹۸۳).

از میان روش‌های مختلف برازش مدل رگرسیونی، روش کمترین توان‌های دوم (LS) به دلیل قدمت و راحتی محاسبات عموما مورد استفاده قرار می‌گیرد. اما، امروزه آگاهی بیشتری نسبت به خطرات رخداد نقاط دور افتاده در برازش مدل حاصل شده است (راسو و لروی، ۱۹۸۷). نقاط دور افتاده در داده‌های واقعی بسیار رخ می‌دهند، و اغلب از کنار این نقاط با بی توجهی گذر می‌شود چرا که امروزه داده‌ها توسط رایانه‌ها بدون توجه دقیق پردازش

می‌گردد. نه تنها متغیر پاسخ می‌تواند دور افتاده باشد بلکه متغیرهای تبیینی مدل نیز می‌توانند عامل دور افتادگی قلمداد شوند که به آنها نقاط نافذ گفته می‌شود. هر دوی این گونه نقاط دور افتاده و نافذ می‌توانند موجب عدم کارایی یک تحلیل کمترین توان‌های دوم معمولی گردند. اغلب، چنین نقاط پر نفوذی از چشم کاربر آمار پنهان می‌مانند به دلیل اینکه این نقاط از مانده‌های روش LS قابل تشخیص نیستند.

دو روش مفید به منظور مقابله با نقاط دورافتاده و نافذ وجود دارد، که یکی ابزارهای تشخیصی در رگرسیون (کوک، ۱۹۷۷ و ۱۹۷۹؛ بلسلی و همکاران، ۱۹۸۰) و دیگری رگرسیون استوار می‌باشد (هابر، ۱۹۷۳ و ۱۹۸۱؛ راسو، ۱۹۸۴؛ راسو و یوهای، ۱۹۸۴). این دو روش دارای هدفی یکسان اما عملکردی در خلاف جهت یکدیگر هستند. در ساختار رگرسیون تشخیصی ابتدا یک رگرسیون برازش داده شده و سپس تحقیق برای تشخیص نقاط دورافتادهٔ بالقوه انجام می‌گیرد تا مدل دوباره با استفاده از داده‌های خوب برازش داده شود در حالی که در رگرسیون استوار، مدلی که مناسب اکثریت داده‌ها باشد برازش داده می‌شود و سپس مشاهداتی که دارای مانده‌های بزرگ از این مدل استوار باشند به عنوان نقاط دورافتاده شناسایی می‌گردند.

همچنین می‌توان تحلیل تاثیر موضعی را به عنوان روشی دیگر به منظور بررسی حساسیت برازش مدل نسبت به فرض‌های مختلف استفاده نمود. با استفاده از این روش نیز می‌توان از طریق حذف یا کم وزن نمودن نقاط مشکوک به شناسایی نقاط دورافتاده یا نافذ با رسم نمودار تاثیر پرداخت. در پژوهش حاضر روشی جدید برای استفاده از خمیدگی نرمال جابه‌جایی درستنمایی به عنوان ابزار انتخاب و مقایسه میان رگرسیون‌های استوار مختلف معرفی می‌گردد.

در فصل ۱ به مرور سوابق و ضرورت بحث و بررسی نقاط دور افتاده در داده‌های طرح هزینه و درآمد خانوار می‌پردازیم. فصل ۲ مروری بر مدل‌های رگرسیونی، تبدیلات مناسب و مانده‌های متداول به عنوان ابزارهای رگرسیون تشخیصی خواهد داشت. همچنین در این فصل آماره‌های تشخیص نفوذ و تحلیل تأثیر موضعی مورد بررسی قرار می‌گیرد. انواع رگرسیون استوار نیز در فصل ۳ بررسی می‌شود. در فصل ٤، مدل‌های متغیر پاسخ محدود معرفی می‌شوند. در فصل ٥، داده‌های طرح هزینه و درآمد خانوارهای شهری مورد استفاده در این پژوهش معرفی و مدل رگرسیون معمولی برازش داده شده و انواع روش‌های تشخیصی برای این مدل مورد بررسی قرار می‌گیرند. همچنین در این فصل انواع رگرسیون‌های استوار برای این داده‌ها برازش داده می‌شود و با استفاده از تحلیل تأثیر موضعی با یکدیگر مقایسه می‌شوند. در گروه مطالعاتی مذکور، جناب آقای دکتر گنجعلی عضو هیأت علمی دانشگاه شهید بهشتی به عنوان مجری طرح و خانم سمانه افتخاری و آقایان شهاب جولانی و احسان بهرامی سامانی در گروه عضویت داشته‌اند که بدین‌وسیله از زحمات بی‌دریغ یکایک این افراد تشکر و قدردانی می‌شود.

نشریه‌ی حاضر، حاصل تلاش گروه مطالعاتی فوق‌الذکر است که امید است مورد توجه و استفاده‌ی مسئولین مسئولین و کارشناسان ذیربط قرار گیرد.

گروه پژوهشی طرح‌های فنی و روش‌های آماری

فهرست مندرجات

[۱] Difference of Fits

[۲] Difference of Betas

[۳] Covariance Ratios

[۴] Least-Absolute Deviations

[۵] Least Trimmed of Squares

فصل ۱

مقدمات

۱.۱ مقدمه‌ای بر طرح آمارگیری هزینه و درآمد و مرور سوابق

طرح آمارگیری هزینه و درآمد خانوارهای شهری و روستایی یکی از طرح‌های آماری گسترده در کشور می‌باشد که با قدمتی ۳۸ ساله یکی از قدیمی‌ترین و مهم‌ترین طرح‌های آمارگیری اجرا شده توسط مرکز آمار ایران است. مجریان این طرح هر ساله با مطالعات و بررسی‌های لازم، تطبیق آن را با استانداردها و توصیه‌های بین المللی، بررسی و تغییرات مورد نیاز در راستای تکامل آن را تهیه و اجرای آن را به کارشناسان مربوط واگذار می‌کنند. هدف کلی این طرح، برآورد میانگین هزینه‌های خوراک، غیرخوراک و هزینه کل و نیز برآورد میانگین درآمد برای یک خانوار در سطح نقاط شهری و روستایی هریک از استان‌ها و کل کشور می‌باشد. به علت وسعت پوشش مکانی و نیز نوع اقلام مورد پرسش، نتایج حاصل از این طرح بسیاری از اطلاعات مورد نیاز اقتصادی همچون شناسایی الگوی مصرف، توزیع درآمد و نیز میزان و تحولات برخورداری خانوارها از امکانات و تسهیلات زندگی و اجتماعی را در مقیاس داخلی و

بین المللی فراهم می آورد که همه اینها در سیاست‌گذاری‌های کلان اقتصادی دولت و بخش خصوصی مفید واقع می‌شود.

در این طرح مشخصات اساسی خانوار از جمله خصوصیات اجتماعی اعضای خانوار که شامل سن، وضعیت سواد (باسواد یا بی‌سواد) و تحصیل (شاغل به تحصیل یا خیر)، وضع فعالیت (شاغل، بیکار جویای کار، دارای درآمد بدون کار، محصل، خانه‌دار و سایر) و وضع زناشویی (دارای همسر، بی‌همسر بر اثر فوت، بی‌همسر بر اثر طلاق و هرگز ازدواج نکرده) است، ثبت می‌گردد و همچنین مشخصات محل سکونت و تسهیلات و لوازم عمده زندگی مانند نحوه تصرف محل سکونت (ملکی عرصه و اعیان، ملکی اعیان، اجاری، رهن، در برابر خدمت، مجانی و سایر) و نوع سوخت عمده مصرفی خانوار (نفت سفید، گازوئیل، گاز، برق و سایر) مورد بررسی قرار گرفته و در سطحی گسترده به بررسی و ثبت هزینه‌های خانوار در قالب دو بخش خوراکی و غیر خوراکی پرداخته شده است. از جمله اقلام مورد توجه در قسمت هزینه‌های خوراکی می‌توان به هزینه‌های مربوط به نان، گوشت، لبنیات و خشکبار مصرفی خانوار اشاره نمود. در بخش هزینه‌های غیر خوراکی به هزینه‌های پوشاک، مسکن و تسهیلات آن، هزینه‌های بهداشتی و درمانی خانوار، هزینه‌های حمل ونقل خانوار، هزینه‌های ارتباطات، خدمات فرهنگی و تفریحی، آموزش و تحصیل و سرمایه گذاری خانوار توجه گردیده است. همچنین درآمدهای خانوار که شامل مواردی همچون درآمدهای حاصل از مشاغل حقوق بگیری و آزاد و درآمدهای متفرقه می‌باشد نیز ثبت گشته است.

با توجه به اهمیت اطلاعات درآمد خانوار در مباحث مربوط به توزیع درآمد، آمارهای فقر و شاخص‌های رفاه اقتصادی، ارزیابی‌ها و سیاست گذاری‌های اقتصادی با اتکا به این مهم، ضروری و حائز اهمیت فراوان می‌باشد. همچنین آگاهی از هزینه و درآمد خانوارهای

کشور می‌تواند منعکس کنندهٔ میزان و ترکیب هزینه‌ها و درآمدها و چگونگی استفاده خانوارها از امکانات و تسهیلات زندگی در سطح ملی و منطقه‌ای باشد. باید توجه داشت که استفاده از مدل‌های مناسب آماری برای پیش‌بینی و بررسی ساختاری اطلاعات بدون در نظر گرفتن خصوصیات توزیعی متغیرهای به کاررفته، اغلب برآوردهای گمراه کننده‌ای را نتیجه می‌دهد. در این میان شناسایی داده‌های پرت به دلیل تاثیر قابل ملاحظه‌ای که بر برازش مدل القا می‌نماید از اهمیت ویژه‌ای برخوردار است.

۱.۱.۱ مرور سوابق

طرح داده کاوی و کاربرد آن در کیفیت داده‌ها

بر اساس گزارش طرح مطالعاتی داده کاوی و کاربرد آن در کیفیت داده‌ها (این طرح از طریق مرکز آمار ایران قابل دسترسی می‌باشد)، اغلب اطلاعات و داده‌های موجود در طرح هزینه و درآمد خانوار توزیع‌های ناشناخته یا پیچیده‌ای دارند که به راحتی نمی‌توان آن توزیع ها را شناسایی نمود و مورد استفاده قرار داد. بنابراین برای تحلیل داده‌ها و اطلاعات فوق استفاده از روش‌هایی که نیاز به دانستن توزیع متغیرها ندارد از اهمیت خاصی برخوردار است. خوشه بندی را می‌توان هم به روش پارامتری و هم ناپارامتری انجام داد که در صورت عدم امکان دسترسی به اطلاعات مورد نیاز برای استفاده از روش پارامتری، می‌توان از خوشه بندی ناپارامتری که با توزیع داده‌های موجود سروکار نداشته و اغلب با استفاده از معیارهای تشابه و عدم تشابه انجام می‌گیرد، استفاده نمود. اغلب در پایگاه داده‌های با حجم بالا تعداد متغیرها، حجم داده‌ها یا هر دو خیلی زیاد است که طرح هزینه و درآمد خانوار نمونه بارزی از آن است، برای خوشه بندی داده‌ها در این پایگاه‌ها یک روش بسیار مفید و ارزشمند تحلیل خوشه‌ای و

تحلیل مناسب در خصوص گروه‌بندی متغیرها تحلیل عاملی است. از آن‌جا که تعداد متغیرها و همچنین تعداد خانوار نمونه در طرح هزینه و درآمد خانوار زیاد می‌باشد، استفاده از روش‌های مرسوم خوشه بندی همچون خوشه بندی سلسله مراتبی یا دو مرحله‌ای، پیشنهاد نمی‌شود. لذا برای انجام تحلیل خوشه‌ای از الگوریتم کلارا که یکی از روش‌های خوشه بندی پم (بخش بندی اطراف مدوئید) است و از جمله روش‌های خوشه بندی در داده کاوی است استفاده شده است. این روش‌ها در مقایسه با روش‌های سلسله مراتبی نیاز به صرف وقت خیلی کمتر و امکانات سخت افزاری معمولی دارند. از آنجا که در روش خوشه بندی تعداد خوشه‌ها را نمی‌توان با استفاده از استدلالی علمی از قبل تعیین کرد، لذا از حداقل انتخاب تعداد خوشه شروع به خوشه بندی نموده و این کار را تا آن جا که نمودار سایه نمای مناسب به دست آید (ماکسسیمم مقدار میانگین سایه نما) ادامه می‌دهند. (برای راهنمایی بیشتر به گزارش طرح مطالعاتی داده کاوی و کاربرد آن در کیفیت داده‌هادر مرکز آمار مراجعه کنید).

در این طرح برای شناسایی و اصلاح مشاهدات دور افتاده از خوشه بندی استفاده گردیده که ابتدا سرجمع‌های هر بخش پرسشنامه را محاسبه کرده و سپس به خوشه بندی هزینه‌های مربوط به هر بخش پرسشنامه پرداخته شده است، بدین منظور خوشه بندی با استفاده از متغیرهایی چون بعد خانوار، فصل آمارگیری و هزینه هر بخش (شامل بخش هزینه‌های خوراک، پوشاک و ...) و نیز هر ۱۳ بخش به طور کلی درسطح مناطق روستایی کل کشور و استان تهران برای تمامی خانوارهای نمونه فصل ۱ سال ۱۳۸۴ انجام شده است. برای این کار با استفاده از الگوریتم خوشه بندی کلارا، خانوارها در بخش‌های هزینه خوشه بندی شده‌اند که بر اساس ضریب سایه نما که تعلق داده‌ها به خوشه‌ها را نمایش می‌دهد، بزرگترین میانگین ضریب سایه نما که بیانگر بهترین تعداد خوشه‌هاست انتخاب شده است.

سپس داخل خوشه‌ها برای هر بخش داده‌های دورافتاده شناسایی شده‌اند. داده‌هایی که در خارج از فاصله ۱/۵ برابر دامنه میان چارکی (متغیر مورد نظر در آن بخش) قرار دارند را به عنوان مشاهدات دور افتاده بالقوه معرفی شده‌اند. سپس با استفاده از سایر فراداده‌های موجود و مرتبط با خانوار در فایل داده‌های مربوط به خانوار دور افتاده بودن یا نبودن اطلاعات خانوار تعیین شده است.

یکی از مشکلات روش گفته شده در بالا نیازمندی به فراداده‌هاست که اغلب جمع آوری نشده‌اند یا به طور کامل دردسترس نمی‌باشند. همچنین روش ذکر شده تنها به مسئلهٔ داده‌های دور افتاده در داده‌های خام توجه می‌نماید که روش تشخیصی آن نیز بر مبنای درنظر گرفتن توزیعی متقارن (استفاده از نمودار جعبه‌ای) برای متغیرهای مورد نظر است که ممکن است در عمل رخ ندهد. از طرف دیگر مسئلهٔ داده‌های دور افتاده زمانی اهمیت بیشتری خواهد داشت که مدل‌بندی و استخراج اطلاعات از داده‌های موجود مورد نظر باشد. این امر خود مستلزم استفاده از تعاریفی دقیق تر از مفهوم داده‌های دور افتاده می‌باشد که گاه به آنها داده‌های با نفوذ بالا اطلاق می‌گردد و تمامی مفاهیم گفته شده تنها در قالب مدل‌بندی معنی و مفهوم واقعی خود را می‌یابد. لذا در پژوهش حاضر به مسئلهٔ داده‌های پرت زمانی که به مرحلهٔ استخراج اطلاعات به منظور پیش بینی یا استنباط وضع موجود توجه گردد، اهمیت داده می‌شود و مسئلهٔ داده‌های پرت و شناسایی آنها در راستای دستیابی به مدلی بهتر مورد توجه قرار می‌گیرد.

۲.۱ نقاط‌دورافتاده

مسئلهٔ نقاط دور افتاده همواره در برازش یک مدل مورد توجه قرار می‌گیرند و معرفی آنها منوط به داشتن معیاری برای میزان دوری از اکثریت داده‌هاست که بر طبق مانده‌های مرتبط با مدل مورد استفاده تعریف می‌گردند. یکی از آشناترین این مدل‌ها، مدل رگرسیونی ساده می‌باشد که شامل برازش مدل‌هایی به یک پاسخ پیوسته است که امید داریم به مقادیر چندین متغیر کمکی وابسته باشد. همچنین در رگرسیون، فرض می‌کنیم که متغیر پاسخ به همراه خطا مشاهده گردیده در حالی که متغیرهای کمکی مقادیر معلومی بدون خطا هستند (که البته در عمل فرض واقع بینانه‌ای نمی‌باشد) و رابطه‌ی میان این دو مجموعه از متغیرها به مجموعه‌ای از پارامترهای نامعلوم که توسط روش کمترین توان‌های دوم برآورد می شوند، وابسته است.

در رگرسیون خطی، متغیر پاسخ به صورت خطی به بردار پارامترهای نامعلوم وابسته است. دو کلاس کلی‌تر از مدل‌ها شامل مدل‌هایی است که غیر خطی در پارامتر هستند و مدل‌های تعمیم‌یافته که درآن‌ها، احتیاجی به پیوسته بودن پاسخ نبوده و وابستگی میان میانگین متغیر پاسخ و متغیرهای مستقل توسط تابع ربط برقرار می‌شود. توجه ما در اینجا بیشتر به مدل‌هایی است که به صورت رگرسیون چندگانه باشند یعنی پاسخ به چندین متغیر کمکی وابسته باشد.

برونلی[۱] (۱۹٦۰)، رگرسیون چندگانه را توصیف کرده و به برآورد پارامترها و آزمون فرض پارامترهای مدل پرداخته است، اما از روش‌های نموداری همچون رسم داده‌ها، یا مانده‌های مدل برازش شده برای تشخیص نقاط دور افتاده استفاده نکرده است. از زمان انتشار کتاب برونلی، تفاوت‌های تحسین برانگیزی در تحلیل‌های آماری رخ داده که با وجود نرم افزارهای قوی گرافیکی آماری به راحتی می‌توان خصوصیات مدل‌های مختلف را با رسم نمودارهای

[۱] Brownlee

گوناکون بررسی کرد.

دراین بخش، به روش‌های نموداری برای ساختن و بررسی مدل‌های رگرسیونی توجه خواهیم نمود. یکی از کاربردهای این نمودارها که به نمودارهای تشخیصی[۲] معروف هستند، شامل تشخیص پرت بودن یک یا چند مشاهده می‌باشد. داده‌هایی که دور افتاده هستند، به حالت‌های مختلفی می‌توانند رخ دهند. در زیر چندین حالت آورده شده است.

۱ – خطاهایی که ممکن است در متغیر پاسخ یا متغیرهای کمکی و یا هر دو وجود داشته باشند. که ممکن است اینگونه خطاها ناشی از اشتباه در اندازه‌گیری‌ها یا خطاهای ایجاد شده در ورود داده‌ها باشند.

۲ – خطا ناشی از در نظر گرفتن شکل صحیح رابطه ریاضی مدل می‌باشد.

۳ – ممکن است به تبدیل مقیاس مانند لگاریتم پاسخ یا دیگر تبدیل‌ها، نیاز باشد (باکس و کاکس[۳] ، ۱۹۶۴).

۴ – قسمت سیستماتیک مدل و مقیاس آن ممکن است درست بوده اما توزیع خطاهای پاسخ متفاوت از نرمال باشند، برای مثال دم‌هایی کلفت تر از توزیع نرمال داشته باشد، که در نتیجه استفاده از روش ماکسیمم درستنمایی برازش مدل را محدود می‌سازد (دراپر و اسمیت، ۱۹۶۶).

یک منبع بالقوه‌ی خطاهای بزرگ (در مطالعاتی که کمیت داده‌ها کوچک باشد)، حذف ارقام اعشاری داده‌هاست که باید با دقت زیادی انجام گیرد و حتی‌الامکان از نرم‌افزارهایی استفاده شود که پیش زمینه‌ی خود کار حذف اعشار را نداشته باشند، تفاوت ویژهٔ میان داده‌ها و مدل، تعیین کننده‌ی آسان یا سخت بودن تشخیص ناهنجاری در مدل است.

---

[۲] Diagnostic Plots

[۳] Box and Cox

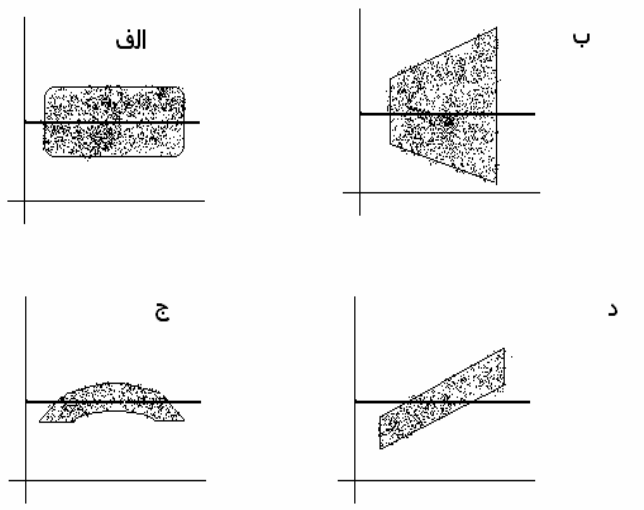
در این قسمت به منظور مطالعهٔ مانده‌ها و نقاط پر نفوذ به بررسی مشاهدات عجیبی می‌پردازیم که بر اساس یکی از سه مورد اول ذکر شده در بالا ناشی شده باشند. بدین منظور در زیر مطالب ساده ولی بسیار مهمی را در ردیابی این قبیل داده‌ها به روش گرافیکی مرور می‌کنیم. مبانی نظری مربوط به این روش‌ها در فصل ۲ گزارش آورده شده است. حالت ۴ نیز در روش‌های رگرسیون استوار در فصل ۳ بررسی خواهد شد.

۱.۲.۱ نمودارهای مانده‌ها

نمودارهای بسیاری برای بررسی تفاوت‌های ساختاری مدل با داده‌ها استفاده می‌شود. این نمودارها بر اساس مانده‌ها رسم می‌گردند. در یک مدل به طور کلی، مانده به تفاضل مقدار برازش شده تحت مدل از مقدار مشاهده شده گفته می‌شود. نمودارهایی که در کنترل فرض‌های مدل و بررسی انحرافات از مدل، مناسب می‌باشند به قرار زیرند.

۱ — نمودار دنبالهٔ زمانی مانده‌ها

برای بررسی وجود تاثیر گذشت زمان (در این حالت فرض می‌کنیم که زمان به عنوان یک متغیر مستقل در مدل استفاده شود) در مقادیر مانده‌های مدل از نمودار مانده‌ها در مقابل زمان استفاده می‌گردد. اگر از دور به نمودار زمانی نگاه کنیم و تجسمی از یک مرز افقی برای مانده‌ها بدست آوریم که می‌توان آن را به صورت قسمت الف از شکل ۱.۱ نشان داد، این نشان می‌دهد که اثر زمانی در دراز مدت روی داده‌ها موثر نیست. اما اگر شکل مانده‌ها شبیه یکی از اشکال ب یا ج یا د در نمودار ۱.۱ باشد، نتیجه می‌گیریم که اثر زمانی در این حالات، در مدل به حساب نیامده است. اگر رفتاری مانند نمودار ب در شکل ۱.۱ رخ دهد می‌توان نتیجه گرفت که واریانس ثابت نبوده بلکه با زمان افزایش می‌یابد.



شکل ۱.۱: حالات ممکن نمودارهای مانده‌ها

نمودار د در شکل ۱.۱ نشان دهنده‌ی نیاز به یک جمله‌ی خطی برحسب زمان در مدل است و نمودار ج در شکل ۱.۱ گویای این مطلب است که باید جملات خطی و درجه‌ی دوم برحسب زمان به مدل اضافه شوند. البته ترکیب‌ها و تغییراتی از این اثرها نیز می‌توانند رخ‌دهند (مثلاً درمورد نمودار د شکل ۱.۱ شیب در جهت مخالف باشد و نظایر آن).

۲ — نمودار مانده‌ها نسبت به مقادیر برازش یافته:

با رسم مقادیر مانده‌ها در مقابل مقادیر برازش یافته، باز مانند نمودار دنباله‌ی زمانی می‌توان انحرافات مدل را بررسی نمود.

۱ — اگر نمودار همانند نمودار الف شکل ۱.۱ یک مرز افقی را نشان دهد، درست بودن تحلیل‌های کمترین توان‌های دوم (و نرمال بودن در مدل رگرسیون خطی) تایید می‌گردد. درمقابل اگر اشکال ب یا ج یا د از شکل ۱.۱ رخ دهند نتایج به صورت زیر خواهند بود: