بسم الله الرحمن الرحیم

دانشکده علوم ریاضی

رشته آمار، گرایش آمارریاضی

پایان‌نامه کارشناسی ارشد

تحلیل متقارن آماری

نگارنده: زین‌العابدین فرامرزی پلنگر

استاد راهنما

دکتر محمد آرشی

دی ۱۳۹۷

پایان‌نامه خود را تقدیم می‌کنم به پدر، مادر و خانواده‌ام که همیشه تکیه‌گاه من بودند.

سپاس گزاری...

من بی تو دمی قرار نتوانم کرد   احسان تو را شمار نتوانم کرد

گر بر تن من زبان شود هر مویی   یک شکر تو از هزار نتوانم کرد

شکر و سپاس خدا را که توانایی به من داد تا این پایان‌نامه را به پایان برسانم. از پدر، مادر و خانواده نهایت تشکر را دارم که وجودشان روشنی‌بخش راهم است. همچنین مراتب سپاس و تشکر قلبی خود را به محضر استاد فرزانه آقای دکتر محمد آرشی که با صبر و حوصله خود مرا در انجام این کار یاری کردند ابراز می‌نمایم. در اینجا جا دارد از اساتید گرانقدر آقایان دکتر ربیعی و دکتر شاهسونی که زحمت داوری این پایان‌نامه را کشیدند تشکر کنم. در انتها از تمام دوستان و عزیزانی که در این مدت مرا یاری و کمک کردند کمال تشکر را دارم و همچنین از تمام معلمانی که از ابتدا تا اینجا زحمات زیادی برای من کشیدند تشکر و قدردانی می‌کنم.

سه خلق را خدا داد برتری   وزین خلق، نداد ش، وین برتری بر دیگری

چو خواهی ! به دیگر برتری شوی   پدر باش و مادر یا آموزگار بر دیگری

پدر، یاور و یار خلق   به مادر، نهادست خدا، جان خلق

با پدر ره فضل، هموار کرد   به مادر، نغمه‌ها آغاز کرد

که این، ره و، این آواز تو   دگر آفرینم، آموزگار است، یار و همراز تو

چو این سه ! با تو! همره بر هم شوند   به حق خلق، یک به یک نکو نام شوند

چو این سه ! وز یادت! روند   به حق! نام تو! وز یادها برند

زین‌العابدین فرامرزی پلنگر
دی ۱۳۹۷

تعهد نامه

اینجانب **زین‌العابدین فرامرزی پلنگر** دانشجوی کارشناسی ارشد رشته **آمار** **علوم ریاضی** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **تحلیل متقارن آماری** ، تحت راهنمایی **محمد آرشی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام ‘‘ دانشگاه صنعتی شاهرود ’’ یا “ Shahrood University of Technology “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

زین‌العابدین فرامرزی پلنگر
دی ۱۳۹۷

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

در این عصر محققان در صدد هستند تا در درباره تصمیمات افراد موارد مختلف که مورد رضایت یا عدم رضایت آن‌ها است، مدلی بسازند. مدل تحلیل متقارن درصدد است که این رضایت‌مندی و انتخاب را مورد بررسی قرار دهد. در اینجا بیشترین بحث در مورد تصمیم افراد برای یک کالا می‌باشد که تحلیل متقارن با استفاده از ویژگی‌ها و سطوحی که برای یک کالا در نظر می‌گیرد آن کالا را توصیف کرده و تصمیمات افراد در مورد کالا را مدل‌سازی می‌کند. در ادامه در فصل اول ابتدا تحلیل متقارن را معرفی و مفاهیم مربوط به آن را شرح می‌دهیم، در فصل دوم روش‌های برآورد را شرح داده و در فصل سوم تحلیل متقارن با اثرات ثابت گفته می‌شود، فصل چهار هم مدل با اثرات آمیخته را بیان می‌کند.

کلمات کلیدی: تحلیل متقارن، مدل اولویت مصرف، مدل بخش‌ارزشی، مدل با اثرات ثابت، مدل با اثرات آمیخته.

پیشگفتار

یکی از اولویت‌های اقتصاد خرد در بازاریابی میزان مصرف است. این میزان به‌کمک تصمیمات مصرف‌کنندگان مورد بررسی قرار می‌گیرد. تحلیل متقارن، یک روش آماری است که تصمیات افراد را اولویت‌بندی می‌کند. تحلیل متقارن در اواسط دهه شصت توسط لوسی و توکی که ریاضی‌دان و روان‌شناس بودند، شروع و توسعه یافت. در ابتدا تحلیل متقارن در جستجوی بازاریابی بود. گرین و سیرینی‌واسان (۱۹۷۰) آن را با ارائه یک مقاله ادامه دادند. اما امروزه تحلیل متقارن یک تکنیک در میدان‌های اقتصاد، جستجوی حمل‌ونقل و ... برای پیدا کردن محصول بهینه به‌وسیله ارائه ویژگی‌ها و سطوح آن است. به‌طوری که اخیرا تحولاتی شامل مدل‌هایی مانند پایه‌های تحلیل متقارن که توسط داماراجو (۲۰۱۱) ارائه شد و سازگازی تحلیل متقارن که کانینگهام (۲۰۱۰) آن را بیان کرده صورت گرفته است. در این پایان‌نامه تحلیل متقارن با مدل‌های رگرسیونی اثرات ثابت و اثرات آمیخته بیان می‌کنیم. همچنین در این کار مدل‌ها را به‌وسیله آنالیز واریانس مورد بررسی قرار داده و با استفاده از داده‌های واقعی و نرم‌افزارهای SAS و R تحلیل داده‌ها انجام می‌شود. این مجموعه شامل چهار فصل که محتوای آن‌ها به طور مختصر به صورت زیر است.

- در فصل یک، به بیان مفاهیم مقدماتی تحلیل متقارن که شامل تعاریف و چند مثال کوتاه است می‌پردازیم.
- در فصل دو، مدل‌های رگرسیونی مربوط به تحلیل متقارن و روش‌های برآورد آن‌ها را بیان می‌کنیم.
- در فصل سه، تحلیل متقارن با اثرات ثابت بررسی می‌شود.
- در فصل چهار، تحلیل متقارن مدل اثرات آمیخته بیان می‌شود.

از این مجموعه مقاله‌ای به شرح زیر استخراج شده است.

فرامرزی پلنگر، ز. آرشی، م. (۱۳۹۷) تحلیل متقارن داده‌های چای با استفاده از رگرسیون خطی با متغیرهای توضیحی کیفی . چهاردهمین کنفرانس آمار ایران. دانشگاه صنعتی شاهرود

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

مفاهیم مقدماتی تحلیل متقارن

۱.۱ مقدمه

مردم همه روزه هزاران تصمیم می‌گیرند که اکثرا ناخودآگاه یا ساده است ولی باید آن‌ها را بررسی کرد. انتخاب از بین چند تصمیم و ارزیابی آن‌ها باعث می‌شود تا افراد آن موضوع را بهتر درک کنند و تصمیم بهتری بگیرند. زمانی که در مورد موضوعی می‌خواهیم تصمیم بگیریم، اگر به صورت گروهی انجام شود بهتر است. اما باید توجه کرد که برای افراد گرفتن تصمیم جدید به‌جای تصمیم قبلی مهم است، و در این تصمیم‌گیری علوم مختلفی از جمله روانشناسی، اقتصاد و علوم محیطی تاثیر دارد (داماراجو و همکاران، ۲۰۱۱). توجه به این نکته نیز لازم است که در انتخاب کالا توسط افراد عواملی مانند اجبار، عادت، تجربه و تبلیغات تاثیر دارد (لوویر و همکاران، ۲۰۰۰). این تاثیرات طبیعت آشفته عواقب انتخاب و بررسی انگیزه مدل‌ها را منعکس می‌کند که به‌وسیله آن رفتار انتخابی فرد را می‌توان شرح داد و به‌خوبی پاسخ رفتاری فرد برای فرصت در حال تغییر را پیش‌بینی کرد که چگونه فرد در مورد موضوعات تصمیم می‌گیرد و به آن‌ها عمل می‌کند. در این فصل به بررسی مفاهیم مقدماتی تحلیل متقارن[۱] که در تصمیم‌گیری برای انتخاب به‌کار می‌رود می‌پردازیم. از آنجایی که در آمار به بررسی این موضوع توجه چندانی نشده است لازم است که در ابتدا یک سری تعاریف پایه‌ای و مفهومی مدیریتی آورده و مدل‌های مرسوم مطرح شوند. قبل از مطالعه این مجموعه توجه به این نکته

[۱] Conjoint analysis

ضروری است که در علوم مختلف تحلیل متقارن یک مفهوم مدیریتی داشته و بیشتر در تجارت از آن استفاده می‌شود. لذا در اکثر مسائل و مثال‌های این مجموعه مسئله تحلیل متقارن را از جنبه ارزیابی انتخاب محصول بررسی می‌کنیم.

۲.۱ مروری بر تحلیل متقارن

پایه کلمه تحلیل متقارن در متصل شدن یا متحد شدن است، و باعث می‌شود که مردم عوامل کلی را در زمان یکسان بررسی کنند (گوستافسون و همکاران، ۲۰۰۷). همچنین می‌دانیم که تحلیل متقارن یک تحلیل نقل و انتقالاتی و تکنیکی است که مردم به‌طور گسترده برای اندازه‌گیری انتخاب بین سرویس یا محصولات استفاده می‌کنند. محصولات شامل عناصر مختلف، پاسخ مشتریانی است که متغیرهای مشاهده شده در مطالعه تحلیل متقارن می‌باشد. متغیری که آن را برآورد می‌کنیم رتبه‌بندی اولویت‌ها[۲] است که پاسخ‌دهنده برای هر یک از محصولات اختصاص داده است. این سطوح اولویت میزان رضایت پاسخ‌دهنده از محصول[۳] را نشان می‌دهد که آن را می‌توان با تکنیک‌های آماری تخمین زد. همچنین در این روش می‌توان برای محصولی که هنوز تولید نشده پیش‌بینی‌هایی درباره میزان مقبولیت و انتخاب انجام داد و تاثیر محصول جدید را در بازار مشاهده نمود. محققان از طریق تحلیل متقارن قادرند ویژگی‌های متفاوت محصول را به گونه‌ای ترکیب کنند که مورد علاقه مشتریان باشد.

۱.۲.۱ برخی تعاریف مرتبط با تحلیل متقارن

در ادامه تعدادی از کلمات کلیدی که در تحلیل متقارن استفاده می‌شود را ارائه می‌نماییم (هایر و همکاران، ۲۰۰۶).

تعریف ۱.۲.۱. ویژگی‌ها[۴]: عواملی هستند که به‌وسیله آن یک محصول را شرح می‌دهیم یا محصول طبق این ویژگی‌ها تولید می‌شود.

به عنوان مثال یک خمیردندان را با ویژگی‌هایی مانند اندازه، نام تجاری (برند)، عطر، طعم و قیمت می‌توان توصیف کرد. این ویژگی‌ها به صورت متغیرهای توضیحی[۵] در مدل تحلیل متقارن وارد می‌شوند.

تعریف ۲.۲.۱. سطوح ویژگی[۶]: سطوح ویژگی گزینه‌هایی هستند که هر ویژگی را توصیف می‌کند. این سطوح نقاط پیشنهادی بازار را برای ویژگی‌ها بیان می‌کند. در مثال خمیردندان

[۲] Preferences
[۳] Product
[۴] Attributes
[۵] Explanatory variables
[۶] Attribute levels

مشتری با انتخاب‌هایی مانند اندازه، قیمت، برند و طعم مواجه می‌شود که هر یک از ویژگی‌ها چندین سطح می‌تواند داشته باشد. مثلا برای طعم می‌توان دو سطح نعنایی و بدون طعم را در نظر گرفت.

تعریف ۳.۲.۱. محصول یا مشخصه[۷]: ترکیبی از سطوح مختلف ویژگی‌ها را توضیح می‌دهد که به شکل یک محصول است. به‌وسیله ترکیبات مختلف از سطوح ویژگی‌ها، محصول جدید برای ارائه به پاسخ‌دهندگان (مصرف‌کنندگان) ساخته می‌شود.

تعریف ۴.۲.۱. موارد مصرف‌شده[۸]: مواردی است که برای ارزیابی افراد در مورد محصولی که استفاده کرده‌اند در نظر گرفته می‌شود.

تعریف ۵.۲.۱. بخش‌ارزشی[۹]: مقادیر بخش‌ارزشی اولویت‌های تحلیل متقارن را به‌دست می‌آورند. در حقیقت بخش‌ارزشی به اولویت یا موارد مصرف شده با هر سطح از هر ویژگی که برای هر محصول تعریف می‌شود، اشاره دارد.

تعریف ۶.۲.۱. تقسیم‌بندی مدل[۱۰]: کلاسی از مدل‌ها است که پاسخ افراد برای مقادیر بخش‌ارزشی را برای هر محصولی که با سطوح و ویژگی‌های مختلف ساخته شده بیان می‌کند.

تعریف ۷.۲.۱. مدل افزایشی[۱۱]: اگر مشتری محصولی را انتخاب کند آن محصول در اولویت قرار گرفته و شرایط آن محصول به مدل اضافه می‌شود.

۲.۲.۱ مدل متقارن مرسوم

مدل متقارن مرسوم روشی است که اصول کلاسیک تحلیل متقارن را به‌کار می‌گیرد و یک مدل افزایشی از اولویت مشتریان را با مقایسه جفتی یا روش مشخصه کامل ارائه می‌کند. برای توضیح بیشتر مثال زیر را مطرح می‌کنیم که تحلیل متقارن اولویت پاسخ‌دهندگان از یک مجموعه کوچک خودرو را شرح می‌دهد.
جدول ۱.۱ را که مربوط به سطوح و ویژگی‌های خودروی سواری است در نظر بگیرید. برای این مثال $۸۱ = ۳ \times ۳ \times ۳ \times ۳$ مشخصه داریم که هر مشخصه معرف یک محصول می‌باشد که به عنوان مثال یکی از این محصولات دارای مشخصه دو در (اندازه)، آیودی(برند)، آبی(رنگ) و ۳۸۰۰۰۰۰۰۰ تومان (قیمت) است. پس از این‌که سطوح و ویژگی‌ها شناسایی شد مشخصه‌ها توسط تحلیل‌گران ساخته می‌شود و در اختیار پاسخ‌دهندگان برای رتبه‌بندی قرار می‌گیرد. بدیهی است بررسی پاسخ‌های مشتریان در مورد مشخصه‌ها، داده‌هایی را برای استفاده در مدل

[۷] Stimuli
[۸] Utility
[۹] Part-Worth
[۱۰] Decompositional
[۱۱] Addabtive model

جدول ۱.۱: سطوح و ویژگی‌های خودرو

سطوح	ویژگی‌ها
۱. ۲در ۲. ۴ در ۳. ۵ در	سایز
۱. تویوتا ۲. فورد ۳. آیودی	برند
۱. خاکستری ۲. قرمز ۳. آبی	رنگ
۱. ۲۲۰۰۰۰۰۰۰۰ تومان ۲. ۲۶۰۰۰۰۰۰۰۰تومان ۳. ۳۸۰۰۰۰۰۰۰۰ تومان	قیمت

آماری مهیا می‌کند، که اولویت ساخت یک خودروی مشخص را که از ترکیب سطوح ویژگی‌ها به‌دست آمده تعیین می‌نماید. اخیرا پیشرفت‌هایی در این زمینه صورت گرفته و تصمیمات فراتر از خریدار در نظر گرفته شده و به سمت مارک‌های مختلف محصول سوق داده شده است. برخی از این تصمیمات مانند رای دادن، تصمیم پزشک در مورد رفتار افراد یا رتبه‌بندی اقلام در یک پرسشنامه می‌باشد (میدو_اولیورز و بوکن‌هات، ۲۰۰۹). همه این انتخاب‌ها می‌تواند توصیفی یا پیوسته باشد. چون مقادیر ارزیابی مصرف‌کنندگان از یک شی(واقعی یا فرضی) ترکیب جدایی از بخش‌ارزشی بوسیله ویژگی‌ها درست می‌کند تحلیل متقارن در شرایط یکسان، مطالعه محصول را انجام می‌دهد. با ترکیب بخش‌های ارزشی محاسبه شده برای هر سطح ویژگی اولویت‌های کلی برای هر محصول، برای هر پاسخ‌دهنده به‌دست می‌آید. در ادامه مراحل انجام یک تحلیل متقارن را به‌طور مختصر ارائه می‌کنیم.

۳.۲.۱ مراحل انجام یک تحلیل متقارن

هشت مرحله لازم است تا یک تحلیل متقارن روی محصولی انجام گیرد و نتیجه در اختیار تولیدکننده محصول جهت بهره‌برداری قرار داده شود (هایر و همکاران، ۲۰۰۶).

- **مرحله ۱**
  تعریف موضوع و جستجو در مورد مشکلات آن موضوع یا محصول
- **مرحله ۲**
  انتخاب یک تحلیل متقارن روانشناسانه به‌وسیله میانگین‌گیری از نتایج یک آزمایش

تجربی.

- انتخاب و تعریف سطوح و ویژگی‌هایی که می‌خواهیم بر روی آن‌ها مطالعه کنیم.
- طراحی یک مجموعه از مشخصات بوسیله ترکیب یک سطح از هر ویژگی به شکل مجموعه‌هایی کامل از مشخصات که توصیف محصول را نشان می‌دهد.
- اولویت‌بندی مشخصه‌ها برای قرار گرفتن در مطالعه متقارن.

- **مرحله ۳**
مجموعه داده‌ها
در این مرحله روشی که پاسخ‌دهندگان می‌توانند برای ارزیابی مشخصات از آن استفاده کنند مشخص می‌شود.

- **مرحله ۴**
ارزیابی مدل فرضی به‌وسیله تحلیل اکتشافی (توصیفی) داده‌ها.

- **مرحله ۵**
انتخاب برآوردگرهای مدل.

- **مرحله ۶**
ارزیابی مدل برازش به‌وسیله ارزیابی قابلیت اطمینان و پیشگویی درست.

- **مرحله ۷**
شرح دادن نتایج و بررسی وابسته‌های خواص مهم.

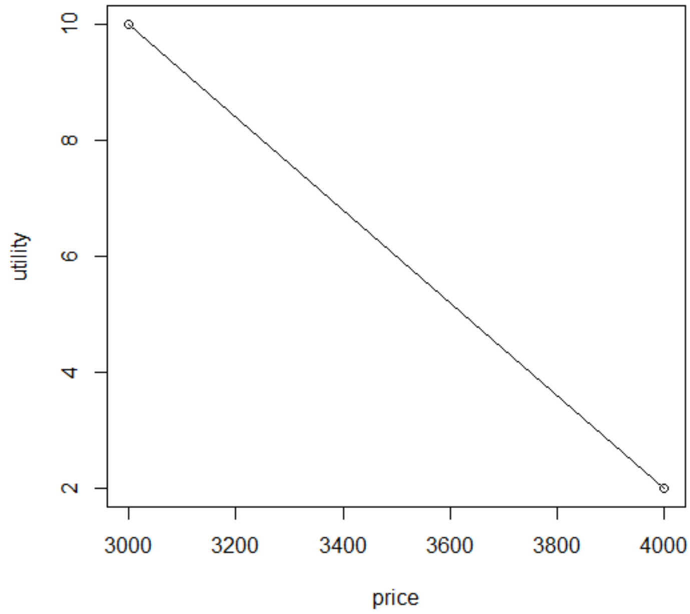
- **مرحله ۸**
اعتبار سنجی و درخواست نتایج تحلیل متقارن.

۳.۱ طراحی یک محصول

برای مطالعه روانشناسانه تحلیل متقارن نیاز به طرح مساله و جستجو در این رابطه می‌باشد. حال برای طراحی محصول ابتدا ویژگی‌های آن را انتخاب و برای هر ویژگی سطح‌هایی در نظر گرفته می‌شود، به‌گونه‌ای که مشخصات کامل محصول مورد نظر را بتوان توصیف کرد. هر محصول به‌وسیله ترکیب سطوح ویژگی‌ها ساخته می‌شود و ساخت محصول باید به گونه‌ای انجام پذیرد که قابل کنترل باشد.

۱.۳.۱ انتخاب ویژگی‌ها و سطوح

مراحل اولیه برای تحلیل متقارن در ساخت یک محصول با ویژگی‌های مختلف، رتبه‌بندی ویژگی‌ها است. به‌دست آوردن توازن بین پاسخ‌ها (باید شرایط برای پاسخ‌ها یکسان باشد) نیازمند برآورد آماری دقیق و استوار مدل است که بدین منظور باید علم پاسخ‌دهنده‌ها مورد ارزیابی قرار گیرد. اغلب محققین تعداد ویژگی‌ها را بین ۶ تا ۸ درنظر می‌گیرند که این تعداد ویژگی باعث ساخته شدن تعداد زیادی از مشخصه‌ها می‌گردد و باید از زیادی تعداد ویژگی‌ها که باعث خستگی و کسل‌کنندگی پاسخ‌دهنده می‌شود اجتناب نمود. محققین برای رفع این مشکل می‌توانند ویژگی‌های غیرضروری را با پرسیدن سوال زیر از بین ببرند.
آیا می‌توان پاسخ‌هایی را برای مطالعه در نظر گرفت که ویژگی مفید دارند، اما آن‌ها از نظر افراد رفتار یکسانی نداشته باشند؟ برای تحلیل متقارن باید از سطوح ویژگی‌هایی استفاده شود که اهمیت بیشتری دارند. یک جستجوگر با استفاده از محصول تولیدی به‌وسیله تحلیل متقارن فرضیه‌ای را در بازار ایجاد می‌کند که ببیند آیا طبق انتظار به محصول تولیدی جدید واکنش نشان داده می‌شود یا خیر. به‌عبارتی یک رقابت بین محصولات تولیدی با سطوح ویژگی‌های مختلف برای فروش ایجاد می‌شود. حال اگر یک محصول با برند خاص و قیمت بالا تولید شود مسلم است میزان استقبال از این محصول کم می‌باشد. بدین منظور وقتی ویژگی قیمت در یک نمونه ساده مانند خمیردندان مورد بررسی قرار می‌گیرد، برای افراد از اهمیت خاصی برخوردار است. برای مثال ابتدا دو سطح ۳۰۰۰ و ۴۰۰۰ تومان در نظر گرفته می‌شود و میزان مصرف هر یک از خمیردندان‌ها با این قیمت‌ها مورد بررسی قرار می‌گیرد، اگر محقق با استفاده از نمودار و نقطه درون‌یابی به این نتیجه برسد که فاصله قابل توجهی بین این دو قیمت می‌باشد پیشنهاد می‌کند که بین این دو قیمت، محصول دیگری مثلا با قیمت ۳۵۰۰ تومان تولید شودکه بتواند با دقت بیشتری مقیاس قیمت را ارزیابی کند.

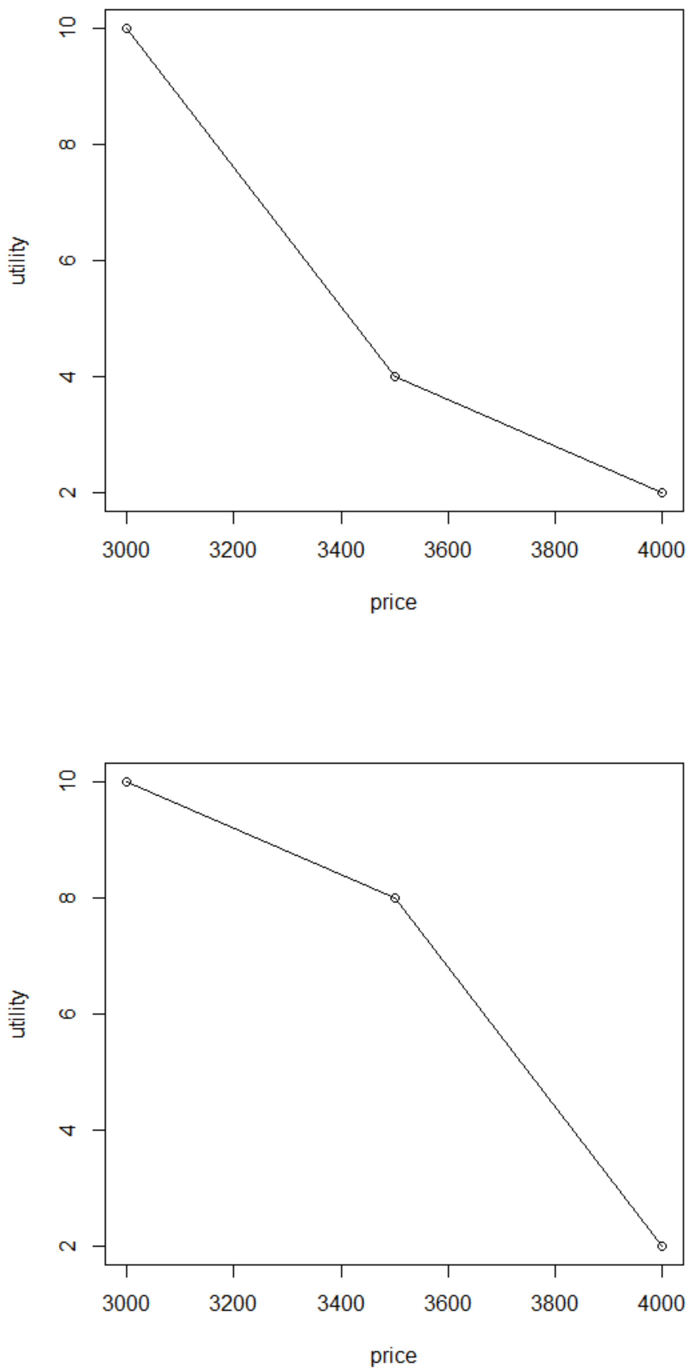


در شکل ۱.۱ قاب بالایی میزان مصرف برای قیمت ۳۰۰۰ و ۴۰۰۰ تومان را نشان می‌دهد. به‌روش درون‌یابی (میانگین‌گیری) برای قیمت ۳۵۰۰ میزان مصرف را می‌توان به‌دست آورد، در حالی که در قاب‌های وسطی و پایینی نمودار به طور دقیق مقادیر برای ۳۵۰۰ نشان داده شده است.

۲.۳.۱ ساختار محصول

وقتی که سطوح مختلف ویژگی‌ها با هم ترکیب می‌شوند مجموعه‌ای از محصولات به‌دست می‌آید. چالش اصلی که محقق در مطالعه تحلیل متقارن با آن برخورد می‌کند، تعداد زیاد مشخصه‌ها است. برای مثال در جدول ۱.۱ به تعداد $۸۱ = ۳^۴$ مشخصه داریم، حال اگر یک ویژگی دیگر با سه سطح اضافه شود تعداد مشخصه‌ها $۳۴۳ = ۳^۵$ افزایش می‌یابد. واضح است با ادامه این فرآیند، انجام آزمایش‌ها بسیار زیاد و پرهزینه می‌شود. بنابراین با تعداد زیاد مشخصه خطای برآورد زیاد می‌شود برای برآورد صحیح پارامترها و انتخاب اولویت‌ها با خطای کمتر، باید علاقمندی‌ها در پرسشنامه‌ها مورد بررسی قرار گیرد. با محدود کردن تعدادی از مشخصه‌ها می‌توان مطمئن بود که نتایج انعکاس داده شده اولویت درست افراد است.

یک روش ساده در برخورد با این مسئله این است که مشخصه‌ای که در برآورد مدل تاثیر اندکی دارد را نادیده گرفت. این مشخصه در زمان جمع‌آوروی اطلاعات وجود دارد ولی می‌توان اثر آن را در مدل به کمک اثرهای متقابل از بین برد و به منظور برآورد پارامترها از سایر مشخصه‌ها استفاده نمود.



شکل ۱.۱: نمودار میزان مصرف در مقابل قیمت برای خمیردندان

۳.۳.۱ ارائه محصول

موضوع اصلی در پاسخ‌ها برای ارائه محصولات مختلف داشتن روش واقع‌بینانه نسبت به آن محصول است. به‌کمک نمایندگی‌های این محصولات و توصیف‌هایی که می‌شود محصول مورد نظر را درک کرده و با مشخصه‌ها و روش تاثیرگذاری آن‌ها محصول ارزیابی شده و علاقه‌مندی افراد نسبت به محصول مورد رتبه‌بندی قرار می‌گیرد. محققین در انتخاب محصول از سه روش زیر استفاده می‌کنند.

۱ توصیف شفاهی: در این روش محصول به‌صورت چهره‌به‌چهره، مصاحبه تلفنی یا پرسش و پاسخ که هیچ عامل بیرونی دیگری از جمله زور یا رودربایستی دخیل نباشد معرفی می‌شود.

۲ توصیف پاراگرافی: در این روش محصول برای انتخاب مصرف‌کننده به طور کامل در یک روزنامه، مجله یا بروشور توصیف می‌شود.

۳ توصیف تصویری: برای انتخاب کالای مورد نظر توسط مصرف‌کننده، تصویری با جزئیات کامل در مقابل وی قرار می‌گیرد. این روش هرچند هزینه‌بر است اما می‌توان قابلیت‌های زیادی را به صورت تصویری برای مشتری (مصرف‌کننده) ترسیم کرد. در شکل ۲.۱، نمونه‌ای از این نوع توصیف را برای بستنی می‌توان دید.

شکل ۲.۱: نمایش تصویری از ویژگی‌ها و سطوح برای تهیه یک محصول بستنی

۴.۱ مدل اولویت مصرف

اودد (۲۰۰۸) تحلیل متقارن را مدلی برای ارزیابی اولویت‌های مصرف معرفی نمود که بیشتر برای توصیف و تحلیل اندازه‌های کمی استفاده می‌شود. مدل اولویت مصرف در تحلیل متقارن مدلی ریاضی است که اولویت‌های درنظر گرفته‌شده برای هر سطح از ویژگی‌ها را ترکیب می‌کند. لازم نیست که مدل اولویت مصرف برای همه افراد یکسان باشد، اما باید پارامترهای یکسانی در همه مدل‌ها وجود داشته باشد. فرض کنید $p = ۱,۲,\cdots,t$ ویژگی‌ها و $j = ۱,۲,\cdots,J$ محصولاتی باشند که در طرح مورد نظر برای یک محصول مطالعه می‌شود. محصول jام با ویژگی pام با x_{jp} نشان داده می‌شود. یادآوری می‌گردد که هر محصول یک سطح از ویژگی‌ها را در نظر دارد یعنی با تغییر p، j هم تغییر می‌کند و برعکس. مثال‌های ارائه شده در جدول‌ها ۲.۱ و ۳.۱ را ببینید. در جدول ۲.۱، فرض کنید ۳ ویژگی ($p = ۳$) رنگ، قیمت و اندازه مورد نظر است که در مجموع ۸ محصول ($j = ۸$) وجود دارد. برای هر ویژگی دو سطح در نظر گرفته شده است. لازم به ذکر است چون هر ویژگی ۲ سطح دارد تعداد محصول‌ها

$\binom{۲}{۱}\binom{۲}{۱}\binom{۲}{۱} = ۲ \times ۲ \times ۲ = ۸$ است. در جدول ۳.۱، مقادیر x محصولات به ازای $j = ۱,۲,\dots,۸$ و $p = ۱,۲,۳$ نشان داده شده است.

جدول ۲.۱: کدگذاری سطوح

کدها	سطح	ویژگی
۱	۱.آبی	رنگ
۱−	۲.قرمز	
۱	۱. ۲۰۰۰۰۰تومان	قیمت
۱−	۲. ۵۰۰۰۰۰ تومان	
۱	۱.کوچک	اندازه
۱−	۲.بزرگ	

جدول ۳.۱: ارائه محصول

						محصول
$x_{۱۳} = ۱$	$x_{۱۲} = ۱$	$x_{۱۱} = ۱$	کوچک	۲۰۰۰۰۰تومان	آبی	$j = ۱$
$x_{۲۳} = -۱$	$x_{۲۲} = ۱$	$x_{۲۱} = ۱$	بزرگ	۲۰۰۰۰۰تومان	آبی	$j = ۲$
$x_{۳۳} = ۱$	$x_{۳۲} = -۱$	$x_{۳۱} = ۱$	کوچک	۵۰۰۰۰۰تومان	آبی	$j = ۳$
$x_{۴۳} = -۱$	$x_{۴۲} = -۱$	$x_{۴۱} = ۱$	بزرگ	۵۰۰۰۰۰تومان	آبی	$j = ۴$
$x_{۵۳} = ۱$	$x_{۵۲} = ۱$	$x_{۵۱} = -۱$	کوچک	۲۰۰۰۰۰تومان	قرمز	$j = ۵$
$x_{۶۳} = -۱$	$x_{۶۲} = ۱$	$x_{۶۱} = -۱$	بزرگ	۲۰۰۰۰۰تومان	قرمز	$j = ۶$
$x_{۷۳} = ۱$	$x_{۷۲} = -۱$	$x_{۷۱} = -۱$	کوچک	۵۰۰۰۰۰تومان	قرمز	$j = ۷$
$x_{۸۳} = -۱$	$x_{۸۲} = -۱$	$x_{۸۱} = -۱$	بزرگ	۵۰۰۰۰۰تومان	قرمز	$j = ۸$

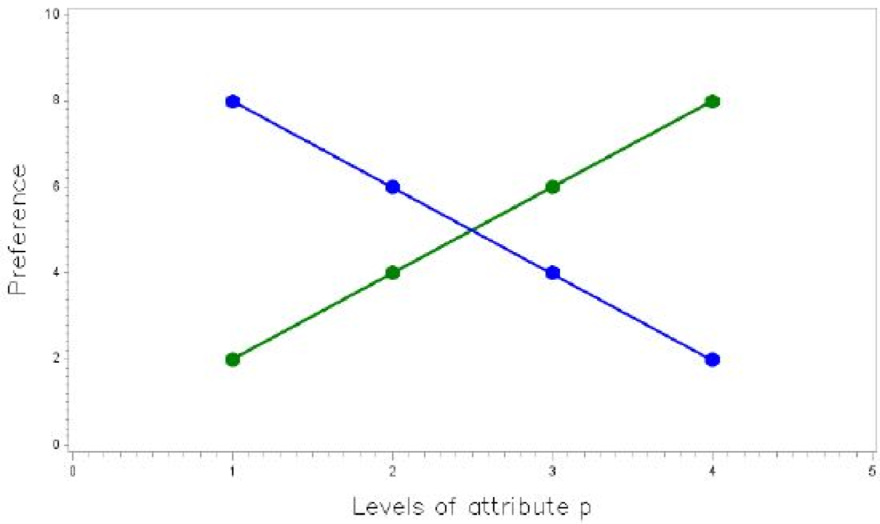
y_j امتیاز پاسخ‌دهنده برای محصول jام است، که برای برآورد موارد مصرف‌شده استفاده می‌شود. هر چه این امتیاز بیشتر باشد، محصول بیشتر مورد رضایت مشتری بوده و امکان تولید آن محصول بیشتر می‌شود. در ادامه، سه مدل مرسوم اولویت مصرف معرفی می‌شود. برای آگاهی بیشتر در مورد این مدل‌ها و خواص آن به کوآل‌تریکس (۲۰۱۱) مراجعه کنید.

۱.۴.۱ مدل برداری

فرض کنید با ترکیب سطوح مختلف چند ویژگی محصولی بدست می‌آید. در مدل برداری، اگر y_j پاسخ یا امتیاز فرد برای محصول jام باشد، آن‌گاه مدل برداری دارای رابطه زیر است

$$\begin{aligned} y_j &= \sum_{p=1}^{t} w_p x_{jp} \\ &= w_1 x_{j1} + \cdots + w_t x_{jt} \end{aligned}$$

که در آن p تعداد ویژگی‌ها و $j = 1, 2, \cdots, J$ تعداد محصول‌ها و w_p وزن پاسخ‌های فرد برای ویژگی pام می‌باشد. وزن w_p با استفاده از روابط و تکنیک‌های آماری برآورد می‌شود. در حقیقت این مدل، یک مدل رگرسیونی است. مدل برداری، برای هر فرد یک مدل ارائه می‌کند. معمولا هر وزن اختصاص داده شده به هر ویژگی برای تمام سطوح آن ویژگی یکسان است. البته ممکن است بعضی از ویژگی‌ها دارای وزن یکسان باشند. در شکل ۳.۱ یک مدل برداری نشان داده شده است.



شکل ۳.۱: مدل برداری برای ویژگی p

۲.۴.۱ مدل نقطه ایده‌آل

در این مدل سطوح ویژگی‌ها از یک مدل منحنی درجه دوم تبعیت می‌کند که ویژگی‌های سطوح پیرو یک تابع منحنی هستند و نقطه ایده‌آل سطح بهینه ویژگی‌ها را نشان می‌دهد. این مدل را با ارائه یک مثال در رابطه با شیرینی توضیح می‌دهیم. شکل ۴.۱ اندازه ویژگی‌ها و اولویت شیرینی را برای یک نفر نشان می‌دهد و همان‌طور که در شکل دیده می‌شود با افزایش سطوح ویژگی اولویت هم افزایش پیدا می‌کند تا زمانی که به نقطه ایده‌آل برسیم و در ادامه با افزایش

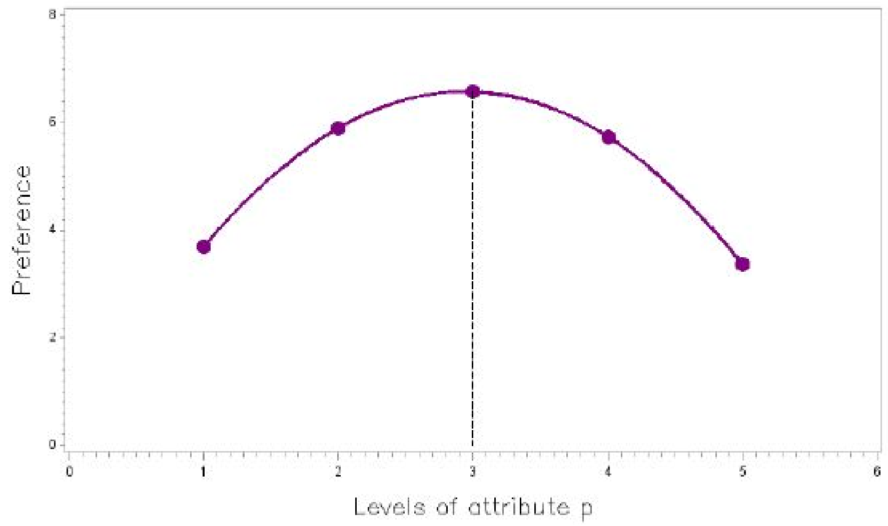
سطوح، اولویت کاهش پیدا می‌کند. (بدیهی است هرچه شیرینی بیشتر بخوریم بعداز نقطه‌ای که در این‌جا همان نقطه ایده‌آل است اشبا می‌شویم و تمایل به خوردن شیرینی بیشتر، کمتر می‌شود) رابطه ریاضی مدل نقطه ایده‌آل به صورت زیر است.

$$\begin{aligned} y_j &= -d_j^{2} \\ &= -\sum_{p=1}^{t} w_p (x_{jp} - x_p)^{2} \end{aligned}$$

که در آن

$$d_j^{2} = \sum_{p=1}^{t} w_p (x_{jp} - x_p)^{2}$$

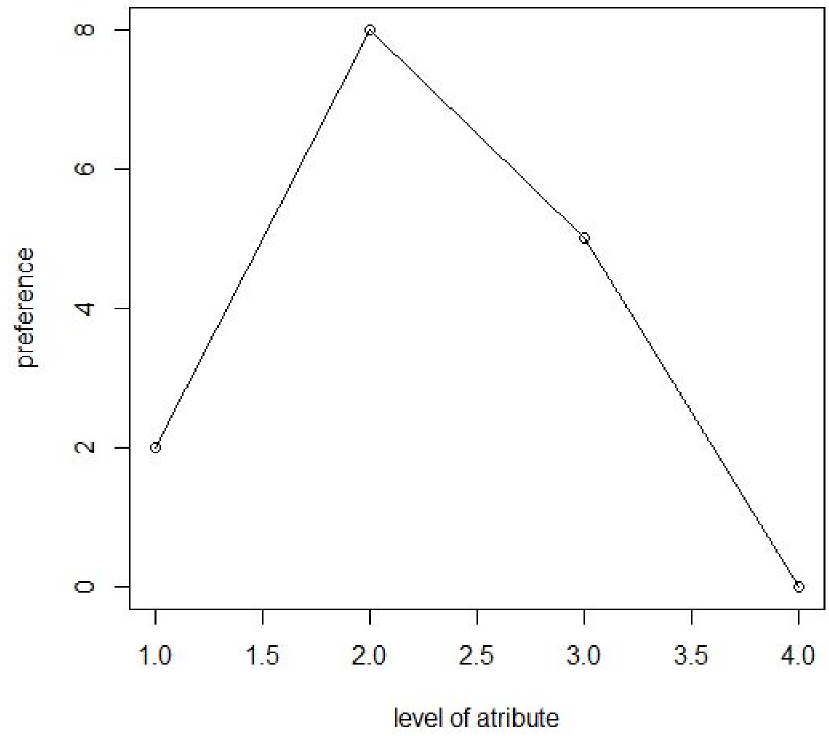
و در آن x_p نقطه ایده‌آل می‌باشد.



شکل ۴.۱: مدل نقطه ایده‌آل برای ویژگی p

۳.۴.۱ مدل تابع بخش ارزشی

این مدل نشان‌دهنده بیشترین مقدار مصرف برای مشتریان می‌باشد که شکل آن به صورت ترکیبی از خط‌های به هم متصل شده است که هر خط برآوردی از بخش ارزشی مصرف مشتریان می‌باشد.



شکل ۵.۱: شکل مدل تابع بخش–ارزشی برای ویژگی p

شکل ۵.۱ تصویر واضحی از مدل بخش‌ارزشی را نشان می‌دهد و به روشنی تغییرات سطوح از هر ویژگی دیده می‌شود. مدل ریاضی آن به صورت زیر است.

$$y_j = \sum_{p=1}^{t} f_p(x_{jp}) \quad j = 1, 2, \cdots, J$$

که در آن x_{jp} سطح دسته و f_p تابع بخش‌ارزشی سطح‌های مختلف از x_{jp} برای ویژگی pام می‌باشد، برای جزئیات بیشتر گرین و همکاران (۲۰۰۱) را ببینید. مدل بخش‌ارزشی میزان ارزش هر سطح از ویژگی‌ها در اولویت‌بندی را نشان می‌دهد. تفاوت عمده مدل بخش‌ارزشی در مقایسه با دو مدل برداری و نقطه ایده‌آل در این است که در این مدل ارزش هر سطح بر اساس مجموع توابعی از سطوح ویژگی مشخص می‌شود که نشان‌دهنده انعطاف‌پذیری بالای آن است و به مدل اجازه می‌دهد شکل‌ها برای تابع اولویت در طول هر ویژگی مختلف باشد اما مدل برداری و نقطه ایده‌آل فقط یک مقدار مصرف از ویژگی را برآورد می‌کنند (گرین و سیرینی‌واسان، ۱۹۷۸). اگر در مدل بخش‌ارزشی $f_p(x_{jp}) = w_p x_{jp}$ باشد آن‌گاه مدل برداری و اگر $f_p(x_{jp}) = -w_p(x_{jp} - x_p)^2$ باشد آن‌گاه مدل نقطه ایده‌آل حاصل می‌شود. در عین حال هرچند مدل بخش‌ارزشی جذاب است اما برای اجرای آن هزینه زیادی لازم است.

در ادامه برآورد پارامترها باید محاسبه شود که از سه مدل می‌توان استفاده کرد. در روش

مدل برداری و مدل نقطه ایده‌آل محدودیت وجود دارد مثلا در مدل برداری فقط وزن‌های w_p و در مدل نقطه ایده‌آل علاوه بر وزن‌ها، نقطه ایده‌آل x_p نیز باید برآورد شود. اما در مدل بخش‌ارزشی چنین نیست یعنی برای هر سطح از هر ویژگی برآورد به‌دست می‌آید. برای مثال، اگر بخش‌ارزشی q سطح داشته باشد آن‌گاه باید $q - ۱$ پارامتر برآورد شود. شکل ۶.۱ ارتباط بین انعطاف‌پذیری و قابلیت اطمینان را برای سه مدل نشان می‌دهد.

شکل ۶.۱: انعطاف‌پذیری و قابلیت اطمینان مدل‌های مختلف

چون در روش تحلیل متقارن برای برآورد پارامترها از پاسخ‌های ذهنی افراد استفاده می‌شود پس این روش یک روش تجربی خواهد بود. تحلیل‌گران برای پیدا کردن بخش‌ارزشی برای سطوح ویژگی‌ها، پاسخ‌های افراد در مورد مشخصه‌ها را مشاهده و مجموعه مشخصه کامل را مورد بررسی و تحلیل قرار می‌دهند. در تحلیل متقارن هدف برآورد پارامترهای تک‌تک افراد بوده و به برآورد پارامترهای توابع مصرفی افراد مورد نظر می‌پردازد. برای آگاهی بیشتر گرین و سیرینی‌واسان (۱۹۷۸) را ببینید. همچنین تحلیل متقارن مدل اولویت مشتریان را بدست می‌آورد و سطح افراد را بررسی نمی‌کند و در انتها انتخاب بهترین مدل به نوع محصول مورد مطالعه، ویژگی‌ها و انواع پاسخ‌ها بستگی دارد.

۵.۱ طراحی روش مجموعه داده‌ها

روش‌های مختلفی برای ارزیابی اولویت حاصل از ارزیابی کلی برای بهتر فهمیدن پاسخ‌ها وجود دارد. روش‌های مجموعه داده‌ها در هر مطالعه تحلیل متقارن متفاوت است. با مثالی از گرین و شافر (۱۹۹۱) در رابطه با سازه‌های مختلف آپارتمان با ۶ ویژگی و هر ویژگی دارای ۳ سطح، این روش‌ها را بررسی می‌کنیم. در حقیقت در این مثال علاقه‌مندیم به روش تحلیل متقارن و استفاده از مدل‌های مختلف موجود، میزان اقبال به انتخاب یک آپارتمان را از دیدگاه فردی

که می‌خواهد به عنوان مشتری آن را اجاره کند بررسی کنیم. در جدول ۴.۱ ویژگی‌ها و سطوح مربوط به آپارتمان را مشاهده می‌کنید.

جدول ۴.۱: مثال آپارتمان: ویژگی‌ها و سطوح آپارتمان

۱. ۱۰دقیقه ۲. ۲۰دقیقه ۳. ۳۰دقیقه	زمان پیاده‌روی تا کلاس
۱. سکوت ۲. متوسط ۳. پر سروصدا	میزان سروصدا
۱. خیلی ایمن ۲. متوسط ۳. ایمن نباشد	ایمنی آپارتمان
۱. بازسازی شده کامل ۲. آشپزخانه بازسازی شده ۳. ساختمان قدیمی	شرایط آپارتمان
۱. ۲۴ تا۳۰ متر ۲. ۱۵ تا ۲۰ متر ۳. ۹ تا ۱۲ متر	اندازه حال نشیمن
۱. ۲۲۵۰۰۰۰ تومان ۲. ۳۶۰۰۰۰۰ تومان ۳. ۵۴۰۰۰۰۰ تومان	میزان اجاره در ماه

با استفاده از ویژگی‌ها و سطوح مختلف، مشخصه‌ها یا آپارتمان‌ها ساخته می‌شود که هر کدام توصیف‌کننده مشخصه‌های یک آپارتمان می‌باشد. در مجموع به تعداد ۷۲۹ = ۳۶ مشخصه داریم. در ادامه به معرفی روش‌های مختلف مجموعه داده‌ها پرداخته و در قالب مثال ۴.۱ مشخصه‌ها را به کمک این روش‌ها می‌سازیم.

۱.۵.۱ مدل مشخصه کامل

مدل مشخصه کامل روش اساسی مصرف است که برای اندازه‌گیری ویژگی مصرف شده توسط مشتریان برای امروز (حال حاضر) به‌کار می‌رود. در این روش پاسخ‌های جمع‌آوری شده توسط

پاسخ‌دهنده‌ها ارزیابی می‌شود. هر یک از مشخصه‌های کامل[۱۲] یک سطح از ویژگی‌ها را برای بررسی و مطالعه توصیف می‌کند. برای تولید یک محصول با توجه به پاسخ‌های پاسخ‌دهنده‌ها، مشخصه‌ها اولویت‌بندی شده و مورد استفاده قرار می‌گیرد.

بر اساس مقیاس کوچکترین و بزرگترین اولویت مشخصه کامل، امتیازهایی که پاسخ‌دهنده‌ها به هر مشخصه می‌دهند ارزیابی شده و طبق این امتیازها، اولویت‌ها رتبه‌بندی می‌شود. با استفاده از مطالعه سطوح ویژگی‌ها و مشخصه‌های ساخته‌شده بوسیله سطح‌ها به این نتیجه می‌رسیم که برخی از این مشخصه‌ها ارزیابی درستی از محصول نمی‌دهند در نتیجه این مشخصه‌ها حذف می‌شوند. طبق مثال آپارتمان (جدول ۴.۱) ۷۲۹ تا مشخصه ممکن وجود دارد که این می‌تواند برای پاسخ‌دهنده‌ها خسته‌کننده باشد پس ارزیابی و توصیف هم دچار مشکل شده و ممکن است برخی از این مشخصه‌ها در زندگی واقعی افراد وجود نداشته باشد. لذا با استفاده از جفت کردن ویژگی‌ها برای طراحی، عوامل کوچک کنترل شده و به کمک آزمون فرضیه پاسخ‌ها را برآورد کرده و مجموعه‌ای در نظر می‌گیریم. در مثال آپارتمان با استفاده از یک طراحی مناسب ۱۸ مشخصه در نظر می‌گیریم به طوری که امکان برآورد سطح‌ها از هر ۶ ویژگی وجود داشته باشد. با کاهش مشخصه‌ها بدون این‌که اطلاعات مهم و ضروری از بین برود پاسخ‌ها دقیق‌تر اندازه‌گیری و اولویت‌ها بهتر مشخص می‌شود. مزیت روش مشخصه کامل برخورد با توصیف‌های واقع‌بینانه برای محصول می‌باشد که این روش می‌تواند پتانسیل واقعی محصول در زندگی واقعی را حساب کند. جدول ۵.۱ یک مشخصه تک طراحی پاسخ‌های نرخ پایه برای تحلیل متقارن را نشان می‌دهد.

جدول ۵.۱: یک مشخصه از روش مشخه کامل در مثال آپارتمان

سطح	ویژگی
۱۰ دقیقه	پیاده‌روی تا کلاس
سکوت	میزان سروصدا
متوسط	ایمنی آپارتمان
آشپزخانه بازسازی شده	شرایط آپارتمان
۲۴ تا ۳۰ متر	اندازه حال‌نشیمن
۲۲۵۰۰۰۰ تومان	اجاره در ماه
میزان مطلوبیت	□

[۱۲] Full profile model

با استفاده از روش رگرسیون کمترین توان دوم معمولی[۱۳] (OLS) مقدار بخش‌ارزشی برای ویژگی سطح‌ها برآورد می‌شود. روش دیگر برای برآورد، رگرسیون خطی را می‌توان نام برد که شامل رگرسیون پنهان[۱۴] و روش بیز سلسله مراتبی[۱۵] است که برای همه این روش‌ها پاسخ‌ها، متغیر وابسته و سطح ویژگی‌ها، متغیر پیشگو است. اگر چه در فصل ۲ به بررسی برخی روش‌های برآورد می‌پردازیم ولی روش بیز سلسله مراتبی را مورد بررسی قرار نمی‌دهیم و برای آگاهی بیشتر در خصوص آن می‌توان به کرونجی (۲۰۱۴) مراجعه کرد. مقادیر بدست آمده از مدل بخش‌ارزشی مدل تابع پاسخ‌ها را تعیین می‌کند و اختلاف این سه روش در برخورد با اولویت‌ها می‌باشد.

۲.۵.۱ مدل مبادله پایاپای

اساس این مدل در جدول مبادله پایاپای[۱۶] است. در جدول‌های مبادله پایاپای دو ویژگی از زمان، به شکل دنباله‌ای از ویژگی‌ها در نظر گرفته می شود. جدول‌های مبادله پایاپای پاسخ‌ها را ارزیابی کرده و اولویت‌ها را از مقدم‌ترین به کمترین اولویت‌بندی می‌کند. هر جدول مبادله پایاپای شامل دو ویژگی می‌باشد که داخل سلول‌های جدول قرار می‌گیرند که بوسیله آن ترکیبات مختلف بوجود می‌آید. جدول ۶.۱ جدول یک مدل مبادله پایاپای بر اساس مثال آپارتمان می‌باشد که پاسخ‌های بدست آمده از آن ارزیابی می‌شود.

جدول ۶.۱: روش مبادله پایاپای دو ویژگی در مثال آپارتمان

اندازه	۹ تا ۱۲ متر	۱۵ تا ۲۰ متر	۲۴ تا ۳۰ متر
پیاده‌روی تا کلاس			
۱۰ دقیقه			
۲۰دقیقه			
۳۰دقیقه			

توجه کنید که چون هر جدول مبادله پایاپای دو ویژگی را در بر می‌گیرد و اگر هر ویژگی دارای سه سطح باشد آن‌گاه $۹ = ۳ \times ۳$ سلول داریم، در این صورت جدول مبادله پایاپای اطلاعات اضافی را کاهش می‌دهد.

[۱۳] Ordinary least squares regression
[۱۴] Latent regression
[۱۵] Hierarichal Bayes approach
[۱۶] Trade-off

در جدول ۶.۱ فقط ویژگی اندازه آپارتمان و میزان پیاده‌روی تا کلاس مورد بررسی قرار گرفته است. با استدلال پاسخ‌ها با افزایش اندازه آپارتمان قیمت هم افزایش پیدا می‌کند. در این روش چون عدم قطعیت در نتایج وجود دارد پس اولویت‌ها درست نشان داده نمی‌شود.

۳.۵.۱ مدل متقارن انتخاب مبنا

یکی از مهمترین و به خصوص کاربردی‌ترین مدل‌های تحلیل متقارن در ایران، مدل تحلیل متقارن انتخاب مبنا[۱۷] (CBC) و انتخاب گسسته[۱۸] (DC) است. اولین بار مدل‌بندی انتخاب گسسته توسط لوسی (۱۹۵۹) و مک‌فارن (۱۹۷۳) به صورت ریاضی فرمول‌بندی شد. پس از مدل‌بندی ریاضی انتخاب گسسته با تحلیل متقارن توسط لوویر و ودورث (۱۹۸۳) صورت گرفت. برای آگاهی بیشتر گروور و وریس (۲۰۰۶) را ببینید.

امروزه مدل CBC و مدل DC به شکل گسترده‌ای در مطالعه تحلیل متقارن استفاده می‌شود. آن‌جایی که راه رضایت مشتری در یک محصول تکرار در استفاده آن است پس در دنیای واقعی افراد برای انتخاب محصول مورد نظرشان دچار چالش می‌شوند که در این صورت نظریه تصمیم‌گیری برای انتخاب کالا بوجود می‌آید. مثلا وقتی مشتری جلوی قفسه خمیردندان قرار می‌گیرد حق انتخاب یکی از گزینه‌ها را دارد که با دیگر انتخاب‌ها امتیاز یکسانی از نظر اولویت بندی ندارد. مدل CBC پایه روش مشخصه کامل می‌باشد اما در اینجا انتخاب، بهترین محصولی است که بیشترین محبوبیت را درنظر خریدار دارد. در اینجا محصول‌هایی که از روش مشخصه کامل ساخته شده‌اند به‌طور جداگانه امتیازبندی می‌شوند (هایر و همکاران، ۲۰۰۶). در حقیقت محصول‌ها به‌صورت دوبه‌دو ارزیابی و مقایسه می‌شوند و پاسخ‌ها به صورت واقع‌بینانه انتخاب می‌شود. مجددا مثال آپارتمان را در نظر بگیرید با استفاده از روش مدل CBC جدول ۷.۱ را داریم.

جدول ۷.۱: روش مدل متقارن انتخاب مبنا مثال آپارتمان

ویژگی	آپارتمان ۱	آپارتمان ۲	آپارتمان ۳
پیاده‌روی تا کلاس	۱۰ دقیقه	۳۰دقیقه	۱۰دقیقه
میزان سروصدا در آپارتمان	متوسط	خیلی آرام	پرسروصدا
ایمنی مکان آپارتمان	خیلی ایمن	متوسط	خیلی ایمن
شرایط آپارتمان	بازسازی کلی	آپارتمان قدیمی	آشپزخانه بازسازی شده
اندازه نشیمن	۹ تا ۱۲ متر	۲۴ تا ۳۰ متر	۹ تا ۱۲ متر
اجاره در ماه	۲۲۵۰۰۰۰ تومان	۵۴۰۰۰۰۰ تومان	۳۶۰۰۰۰۰ تومان

[۱۷] Choice based conjoint model
[۱۸] Discrete choice

روش برآورد در این مدل به طور مفهومی مشابه روش OLS است و می‌توان، دسته‌های دوتایی و چندتایی انتخاب کرد. برای مدل چندتایی مدل لجستیک چندگانه[۱۹] (MNL) مناسب است.

۴.۵.۱ مدل تطبیقی

مدل تطبیقی[۲۰] یک پردازش مطمئن با روش معمولی برای تحلیل متقارن است. این روش مدل‌بندی شامل مجموعه‌ای از مراحل است که با اطلاعات در دسترس می توان برآوردی برای موارد مصرف شده بخش‌ارزشی تعیین کرد و بهترین پیش‌گویی برای اولویت‌های دنیای واقعی را داشت. در این روش از یک تحلیل متقارن متکی به مجموعه اطلاعات پاسخ‌ها استفاده می‌کنیم در ادامه مراحل تحلیل متقارن تطبیقی از کانینگهام (۲۰۱۰) را ارائه می‌کنیم.

مرحله ۱: در طول این مرحله پاسخ‌ها و سطح‌های ویژگی به ترتیب اهمیت رتبه‌بندی می‌شود. جدول ۸.۱ یکی از ویژگی‌های سطوح مثال آپارتمان را که می‌توان رتبه‌بندی کرد، نشان می‌دهد.

جدول ۸.۱: مرحله اول روش تحلیل متقارن انطباقی در مثال آپارتمان

شرایط آپارتمان:
□ بازسازی کلی
□ آشپزخانه بازسازی شده
□ آپارتمان قدیمی

مرحله ۲: در این مرحله سطوح ویژگی‌ها در مرحله ۱ (پاسخ مشتری) تعیین می‌شود. جدول ۹.۱ مرحله دوم را در مثال آپارتمان نشان می‌دهد.

[۱۹] Multinomial logistic model
[۲۰] Adaptive model

جدول ۹.۱: مرحله دوم روش تحلیل متقارن انطباقی در مثال آپارتمان

بازسازی شده کلی و آشپزخانه بازسازی شده: ☐ خیلی خیلی مهم ☐ خیلی مهم ☐ مهم ☐ مهم نیست

مرحله ۳: پس از اجرای دو مرحله اول پاسخ مشخصات کلی توصیفی دریافتی به صورت مقایسات جفتی اعمال می‌شود که در جدول ۱۰.۱ آورده شده است.

جدول ۱۰.۱: مرحله ۳ روش تحلیل متقارن انطباقی در مثال آپارتمان

۱	۲
۲۰دقیقه تا کلاس	۱۰ دقیقه تا کلاس
خیلی آرام	میزان سروصدا متوسط
ایمنی متوسط	خیلی ایمن
بازسازی شده کامل	آشپزخانه بازسازی شده
نشیمن ۹ تا ۱۲ پا	نشیمن ۱۵ تا ۲۰ پا
اجاره در ماه ۳۶۰۰۰۰۰ تومان	اجاره در ماه ۵۴۰۰۰۰۰ تومان
☐	☐

حال برای انتخاب مجموعه با توجه به پاسخ‌ها، سطوحی که مقدم‌ترند اولویت‌بندی شده و هر پاسخ نیاز به ارزیابی دارد که نقطه شروع اولویت‌بندی پاسخ‌ها می‌باشد. همچنین شرایط این مرحله به گونه‌ای است که ویژگی‌ها و سطوحی که در شرایط مراحل پذیرفتنی نیستند از بین می‌روند و ویژگی‌های باقی‌مانده برای ارزیابی نشان داده می‌شوند بزرگترین مزیت روش تحلیل متقارن تطبیقی کاهش طول اطمینان بدون خطر است. به عبارت دقیق‌تر سطوحی که کمترین اهمیت را دارند و یا اهمیتی ندارند آورده نمی‌شوند و مشتری نیازی به پاسخ دادن آن‌ها ندارد.

۵.۵.۱ مدل ماکزیمم- تفاوت

محقق در برخی مطالعات به‌دنبال این است بداند که مشتریان کدام محصول را ترجیح می‌دهند. در روش‌های قبلی پرسش‌های مطرح شده از پاسخ‌دهندگان بر اساس اندازه مقیاس ویژگی‌ها بوده و اولویت سطوح اهمیت داشت. یک مشکل اساسی روش‌های مطرح شده این است که پاسخ‌دهنده بین بیشترین مقدار (بالاترین سطح) و آن چیزی که خودش می‌خواهد نمی‌تواند انتخاب کند. روش دوم محصولات را از بیشترین اهمیت به کمترین اهمیت رتبه‌بندی می‌کند. وقتی پاسخ‌دهنده رتبه چهار یا پنج مشخصه را می‌پرسد ترتیب این محصولات در مقابل ترتیب اولویت نسبتا آسان است. اگر تعداد مشخصه‌ها افزایش پیدا کند بحث در مورد این مشخصه‌های مختلف زیاد می‌شود. تشخیص بین بیشترین و کمترین اولویت مشخصه‌ها برای پاسخ‌دهندگان نسبتا آسان است، اما تشخیص مشخصه‌های میانی (مشخصه‌هایی که بین بیشترین مشخصه و کمترین مشخصه هستند) اولویت آن‌ها چگونه است مشکل می‌باشد. (کوآل‌تریکس، ۲۰۱۱) اگر چه اختلاف بین مشخصه‌های مختلف که دارای وزن یکسان هستند و مشخصه‌هایی که دارای وزن یکسان نیستند روش‌های قبلی ممکن است ارزیابی درستی از پاسخ، پاسخ‌دهنده ندهد. یک تکنیک که به این چالش چیره شود و دقت شناسایی بیشتری داشته باشد روش مقیاس اختلاف ماکزیمم[۲۱] می‌باشد. (ایساکسون و لنیسک، ۲۰۱۲) تحلیل متقارن ماکزیمم متفاوت بسته‌هایی از طبقه‌بندی محصولات مختلف را نشان می‌دهد که سناریوی بیشترین و کمترین اولویت را ارزیابی و بررسی می‌کند. هدف در هنگام استفاده از این طراحی تجربی، اطمینان از این است که هر مقدار برابر با مرتبه‌های دیگر نشان داده می‌شود، و هر جفت از مرتبه‌ها برابر با مقدار مرتبه‌های دیگر نشان و مجموعه اقلام بدرستی نشان داده می‌شود. جدول ۱۱.۱ یک مثال برای تحلیل متقارن مدل ماکزیمم متفاوت می‌باشد.

[۲۱] Max-diff

جدول ۱۱.۱: روش ارزیابی داده‌های مدل اختلاف ماکزیمم در مثال آپارتمان

کدام آپارتمان بیشترین و کدام کمترین محبوبیت را دارد؟		
بیشترین	مشخصات	کمترین
□	فاصله تا مدرسه ۱۰ دقیقه جای خیلی آرام منطقه ایمن آشپزخانه بازسازی شده نشیمن ۹ تا ۱۲ متر اجاره در ماه ۳۶۰۰۰۰۰ تومان	□
□	فاصله تا مدرسه ۱۰ دقیقه جای خیلی پرسروصدا منطقه از نظر ایمنی متوسط کامل بازسازی شده نشیمن ۲۴ تا ۳۰ متر اجاره در ماه ۵۶۰۰۰۰۰ تومان	□
□	فاصله تا مدرسه ۳۰ دقیقه از نظر سروصدا متوسط منطقه ایمن آشپزخانه بازسازی شده نشیمن ۹ تا ۱۲ متر اجاره در ماه ۵۶۰۰۰۰۰ تومان	□
□	فاصله تا مدرسه ۲۰ دقیقه جای خیلی آرام منطقه ایمن خانه قدیمی نشیمن ۹ تا ۱۲ متر اجاره در ماه ۵۶۰۰۰۰۰ تومان	□

آخرین مدل تحلیل متقارن روش خود توصیفی است که برای توضیح در خصوص این روش می‌توان به کوآل‌تریکس (۲۰۱۱) مراجعه کرد.

فصل ۲

روش‌های برآورد

۱.۲ مقدمه

در رگرسیون، تعیین روابط بین متغیرها و تحلیل روابط به‌دست آمده مورد توجه قرار می‌گیرد. به عنوان مثالی از رگرسیون، فرض کنید می‌خواهیم بدانیم که آیا مصرف سرانه سیگار با متغیرهایی نظیر درآمد سرانه، سطح تحصیلات، سهم دستمزد از درآمد کل و متوسط قیمت سیگار رابطه دارد یا خیر. یا ممکن است بخواهیم چگونگی اثر تغییرات یک متغیر بر متغیر دیگر را بررسی کنیم. رابطه بین مصرف سیگار با متغیرهای ذکر شده به‌صورت یک معادله یا الگویی است که متغیر وابسته (مصرف سیگار) را به یک یا چند متغیر پیشگو (توضیحی) مربوط می‌کند. در رگرسیون خطی متغیر پاسخ، پیوسته است ولی متغیرهای پیشگو می‌توانند پیوسته، گسسته یا رسته‌ای باشند. در مثال مصرف سیگار، متغیر وابسته برحسب متوسط تعداد بسته‌های سیگار فروخته‌شده در یک استان بر مبنای سرانه در طول یک سال محاسبه می‌شود و متغیرهای پیشگو، متغیرهای اقتصادی_اجتماعی ذکر شده در بالا می‌باشند. در یک معادله رگرسیونی، متغیر پاسخ را با Y و متغیر پیش‌گو را با $X_۱, X_۲, \cdots, X_p$ نشان می‌دهند. برای اطلاعات بیشتر به نیرومند (۱۳۸۷) مراجعه شود.

۲.۲ رگرسیون خطی چندگانه

مدل رگرسیون خطی چندگانه[1] عموما به صورت زیر است

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \quad i = 1, 2, \cdots, n \tag{۱.۲}$$

که در آن Y_i متغیر وابسته iام، X_{ij} متغیر توضیحی jام در نمونه iام، و β_j ضریب رگرسیونی jام، $j = 1, 2, \cdots, p$ و ε_i خطای تصادفی iام است که دارای خاصیت زیر می‌باشد

$$E(\varepsilon_i) = 0, \quad \mathrm{Var}(\varepsilon_i) = \sigma^2, \quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j = 1, 2, \cdots, n.$$

باید توجه داشت که برای پیدا کردن برآورد OLS ضرایب رگرسیونی، لازم نیست که فرض شود بردار خطای رگرسیونی از توزیع نرمال پیروی می‌کند.

مدل رگرسیونی خطی چندگانه (۱.۲) را می‌توان به‌صورت ماتریسی زیر نیز نشان داد.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{۲.۲}$$

که در آن

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} = [\mathbf{x}_1, \mathbf{x}_1, \cdots, \mathbf{x}_n]^\top$$

که در آن

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}., \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{x}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \mathbf{X}_{i2} \\ \vdots \\ \mathbf{X}_{ip} \end{bmatrix}$$

برآورد OLS پارامتر β از روش کمینه کردن مجموع توان‌های دوم خطا به‌صورت زیر به‌دست می‌آید (فرض کنیم $\mathbf{X}\top\mathbf{X}$ معکوس‌پذیر است)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

جزئیات برآورد در بخش ۱.۳.۳ آمده است. در اغلب مسائل کاربردی واریانس مؤلفه خطا، $\sigma^2 I_n$ نبوده بلکه بین مؤلفه‌های خطا، همبستگی وجود دارد. در این‌گونه مسائل واریانس خطا به‌صورت $\sigma^2\mathbf{V}$ بوده که در آن $\mathbf{V}$ یک ماتریس متقارن $n \times n$ است.

[1] Multiple regression

دراین حالت فرض‌های مدل رگرسیون (۱.۲) به‌صورت $E(\varepsilon) = \circ$ و $\text{Var}(\varepsilon) = \sigma^{۲}\mathbf{V}$، که در آن $\mathbf{V}$ یک ماتریس متقارن است، تغییر می‌کند. در این حالت معیار OLS به‌صورت زیر است

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-۱}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

و برآوردگر OLS پارامتر $\boldsymbol{\beta}$ به‌صورت زیر بدست می‌آید.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-۱}\mathbf{X})^{-۱}\mathbf{X}^\top \mathbf{V}^{-۱}\mathbf{Y}$$

۳.۲ الگوی خطی تعمیم‌یافته

مدل رگرسیون خطی (۱.۲) را در نظر بگیرید. این مدل حالت خاصی از الگوی کلی رگرسیون

$$Y = f(\mathbf{X}, \boldsymbol{\beta}) + \varepsilon \qquad (۳.۲)$$

است. توابع مختلفی می‌توان برای $f(\mathbf{X}, \boldsymbol{\beta})$ در رابطه (۳.۲) تعریف نمود. در رابطه (۱.۲) $f(\mathbf{X}, \boldsymbol{\beta})$ به صورت $\mathbf{X}\boldsymbol{\beta}$ است. می‌توان $f(\mathbf{X}, \boldsymbol{\beta})$ را به صورت $\boldsymbol{\beta}_۱\exp\{\boldsymbol{\beta}_۲\mathbf{X}\}$ فرض نمود که در این صورت الگوی رگرسیون غیرخطی به شکل

$$\mathbf{Y} = \boldsymbol{\beta}_۱\exp\{\mathbf{X}\boldsymbol{\beta}_۲\} + \varepsilon$$

به‌دست می‌آید. که همان‌طور که ملاحظه می‌شود این مدل نسبت به پارامترهای $\beta_۱$ و $\beta_۲$ خطی نیست. اغلب توزیع خطاها را در مدل رگرسیون، نرمال در نظر می‌گیرند. بر اساس رابطه (۳.۲)

$$E\left[\mathbf{Y} \mid \mathbf{X} = \mathbf{x}\right] = E\left[\mathbf{Y} = f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon\right] = f(\mathbf{x}, \boldsymbol{\beta})$$

که در آن $f(\mathbf{x}, \boldsymbol{\beta})$ تابع انتظار الگوی رگرسیون نامیده می‌شود. در مدل رگرسیون خطی این تابع مورد انتظار، تابعی خطی از $\boldsymbol{\beta}$ است، اما در حالت کلی $f(\mathbf{x}, \boldsymbol{\beta})$ می‌تواند غیرخطی نیز باشد. در الگوی رگرسیون غیرخطی حداقل یکی از مشتق‌های جزئی تابع مورد انتظار نسبت به پارامترها به یکی از پارامترها وابسته است.

به عنوان مثال الگوی رگرسیون (۳.۲) تابع $f(\mathbf{x}, \boldsymbol{\beta})$ را به‌صورت

$$f(\mathbf{x}, \boldsymbol{\beta}) = \beta_\circ + \sum_{j=۱}^{k} \beta_j x_j$$

در نظر بگیرید. مشتق‌های جزئی این تابع به شکل

$$\frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \beta_j} = x_j, \quad j = ۱, ۲, \cdots, p$$

بدست می‌آید. توجه کنید در این حالت مشتق‌های جزئی، تابعی از پارامترهای مجهول نیستند. به عنوان مثال دیگر الگوی رگرسیون غیرخطی زیر

$$Y = \boldsymbol{\beta}_۱ \exp\{\mathbf{X}\boldsymbol{\beta}_۲\} + \varepsilon$$

را در نظر می‌گیریم. مشتق‌های جزئی $f(\mathrm{x}, \boldsymbol{\beta})$ نسبت به $\boldsymbol{\beta}_۱$ و $\boldsymbol{\beta}_۲$ به‌صورت

$$\frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_۱} = \exp\{x\boldsymbol{\beta}_۲\}$$
$$\frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_۲} = \boldsymbol{\beta}_۱ \exp\{x\boldsymbol{\beta}_۲\}$$

به دست می‌آید. از آن‌جایی که مشتق‌های جزئی توابعی از پارامترهای مجهول $\boldsymbol{\beta}_۱$ و $\boldsymbol{\beta}_۲$ هستند، این الگو غیرخطی می‌باشد.

معمولا در مدل‌های رگرسیون خطی و غیرخطی فرض می‌شود که متغیر پاسخ Y از توزیع نرمال پیروی می‌کند. اما حالاتی در واقعیت استفاده از چنین فرضی، لزوما مورد تایید نیست. فرض کنید متغیر پاسخ، گسسته از نوع شمارشی باشد. مثلا اغلب با شمارش عیب‌ها یا پیشامدهای نادری چون آسیب‌ها، بیمارانی با امراض خاص و حتی با وقوع پدیده‌های طبیعی از قبیل زمین‌لرزه‌ها و طوفان‌های وابسته به آن مواجه می‌شویم. حالت دیگر یک متغیر پاسخ می‌تواند دودویی باشد مانند مطالعاتی که در آن‌ها متغیر پاسخ، موفقیت یا شکست (یعنی صفر یا یک) درنظر گرفته می‌شود، که تقریبا در تمام زمینه‌های علوم و مهندسی نسبتا متداول هستند. موارد زیادی نیز وجود دارند که متغیر پاسخ در آن‌ها پیوسته است، اما فرض نرمال بودن کاملا غیرواقعی است. به عنوان مثال توزیع فشارها در اجزاء مکانیکی و زمان زوال اجزاء الکترونیکی یا سیستم‌ها از توزیع نرمال پیروی نمی‌کند. این نوع پاسخ‌ها نامنفی بوده و معمولا یک رفتار چوله به راست جدی را نشان می‌دهند.

مدل‌های خطی تعمیم‌یافته[۲] (GLM) برای برازش الگوهای رگرسیون به داده‌های پاسخ تک متغیره توسعه داده شده‌اند که در آن توزیع متغیر پاسخ از خانواده نمایی است. این خانواده بسیاری از توزیع‌های مهم مانند نرمال، دوجمله‌ای، پواسون، هندسی، دوجمله‌ای منفی، نمایی و گاما را شامل می‌شود. مدل خطی تعمیم‌یافته به‌صورت

$$g[E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x}_i)] = \mathbf{x}_i^\top \boldsymbol{\beta}$$

نشان داده می‌شود. هر الگوی خطی تعمیم‌یافته سه جزء دارد:

۱ توزیع متغیر پاسخ (که گاهی اوقات ساختار خطا نامیده می‌شود)

۲ پیش‌گوی خطی که متغیرهای توضیحی را شامل می‌شود

[۲] Generalized linear model

۳ تابع پیوند $g(.)$ که پیشگوی خطی را به میانگین متغیر پاسخ مربوط می‌کند.

برای مثال الگوی رگرسیون خطی در معادله (۱.۲) را در نظر می‌گیریم. توزیع پاسخ نرمال بوده، پیش‌گوی خطی $\mathbf{X}\boldsymbol{\beta}$ و تابع پیوند $g(\boldsymbol{\alpha}) = \boldsymbol{\alpha}$ است یعنی

$$g[E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x})] = \beta_۱ x_۱ + \beta_۲ x_۲ + \cdots + \beta_k x_k$$

بنابراین، الگوی رگرسیون خطی استاندارد در معادله (۱.۲) عضوی از کلاس GLM است. براساس انتخاب تابع پیوند $g(.)$، یک GLM می‌تواند یک الگوی غیرخطی را نیز شامل شود. به عنوان مثال، اگر از تابع پیوند لگاریتمی $g(\alpha) = \ln(\alpha)$ استفاده شود آن‌گاه

$$E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x}) = \exp\{\beta_۱ x_۱ + \beta_۲ x_۲ + \cdots + \beta_k x_k\}$$

الگوی خطی تعمیم‌یافته را می‌توان به عنوان یکسان‌ساز الگوهای خطی و غیرخطی تلقی کرد که یک خانواده از توزیع‌های پاسخ نرمال و غیرنرمال را به هم ارتباط می‌دهد و برازش الگو و استنباط را در یک چارچوب مشترک انجام می‌دهد. هم‌چنین از آنجایی که، نرم‌افزارهای رایانه‌ای که از این رویکرد یکسان‌سازی حمایت می‌کنند، این روش به‌طور وسیعی در دسترس بوده و استفاده از آن رایج گردیده است. به این ترتیب، در حالی‌که کاربرد اولیه الگوهای GLM به علوم زیستی و صنایع داروسازی برمی‌گردد، در سایر زمینه‌های علوم و مهندسی نیز به سرعت گسترش یافته‌اند.

با توجه به مفهوم الگوهای خطی تعمیم‌یافته دو مورد مهم مطرح می‌گردد: ۱.توزیع پاسخ ۲.الگویی که میانگین پاسخ را به متغیرهایی از رگرسیون مربوط می‌کند. از آن‌جایی که الگوهای معین برای بعضی از توزیع‌ها نسبت به سایر توزیع‌ها مناسب‌تر هستند، این دو مورد مستقل از یکدیگر نمی‌باشند. برای مثال، اگر پاسخ دودویی باشد توزیع برنولی برای آن استفاده می‌شود که امید ریاضی این توزیع مقداری بین صفر و یک است. علاوه بر این، اگر پاسخ شمارشی باشد، از توزیع پواسون استفاده می‌شود. که پارامتر توزیع پواسون باید مقداری نامنفی باشد، پس نمی‌توان از مدلی استفاده کرد که برای این پارامتر مقادیر منفی ایجاد می‌کند.

اکنون ساختار زیر را به عنوان تعمیمی از مدل‌های دوجمله‌ای و پواسون در نظر بگیرید. فرض کنید مشاهدات پاسخ $Y_۱, Y_۲, \cdots, Y_n$ دارای میانگین‌های $\mu_۱, \mu_۲, \cdots, \mu_n$ باشند. هم‌چنین ارتباط بین پیشگوی خطی $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ و متغیر پاسخ Y از طریق تابع پیوند زیر برقرار باشد $\eta_i = g(\mu_i)$ که $i = ۱, ۲, \cdots, n$ است. از توابع پیوند معروف می‌توان به موارد زیر اشاره نمود

۱. پیوند همانی:

$$\eta_i = \mu_i$$

که برای توزیع‌های نرمال به‌کار می‌رود.

۲. پیوند لجستیک:

$$\eta_i = \ln(\frac{p}{۱ - p})$$

که از این پیوند برای توزیع دوجمله‌ای استفاده می‌شود.

۳. پیوند لگاریتمی:

$$\eta_i = \ln(\mu_i)$$

که الگوی خطی تعمیم یافته توزیع پواسون را بیان می‌کند.

۴. پیوند وارون:

$$\eta_i = \frac{1}{\mu_i}$$

را می‌توان برای توزیع‌های گاما و نمایی به‌کار برد.

۶. پیوند پروبیت:

$$\eta_i = \Phi^{-1}[E(y_i)]$$

که در آن $\Phi(.)$ تابع توزیع تجمعی نرمال استاندارد را نشان می‌دهد.

۷. پیوند لگ_لگ مکمل

$$\eta_i = \ln[\ln[1 - \mu_i]]$$

۸. پیوند خانواده توانی

$$\eta_i = \begin{cases} \mu_i^{\lambda} & \lambda \neq 0 \\ \ln[\mu_i] & \lambda = 0 \end{cases}$$

انتخاب تابع پیوند به نوعی مشابه انتخاب تبدیلی از پاسخ تلقی می‌گردد، در این حال، باید به این نکته توجه نمود که تابع پیوند، تابعی از میانگین جامعه است نه داده‌ها. برخلاف یک تبدیل، تابع پیوند از ویژگی توزیع پاسخ استفاده می‌کند. به‌طور عینی همان‌طور که استفاده نکردن از یک تبدیل درست می‌تواند مشکلاتی را در برازش الگوی خطی به بار آورد، انتخاب‌های نادرست تابع پیوند نیز می‌تواند به مشکلاتی در یک الگوی خطی تعمیم‌یافته منجر شود.

صورت‌های بی‌شماری از رگرسیون‌ها وجود دارد که می‌توان در مدل‌سازی متغیر پاسخ از آن‌ها استفاده کرد که هر کدام دارای اهمیت و شرایط خاصی است در آن شرایط مناسب‌ترین کارایی خود را دارا هستند. در ادامه چند مدل رگرسیونی نام می‌بریم.

- رگرسیون خطی ساده: به بررسی رابطه یک متغیر مستقل (پیشگو) و یک متغیر وابسته (پاسخ) پرداخته می‌شود.

- رگرسیون خطی چندگانه: اگر تعداد متغیرهای پیشگو افرایش یابد، مدل رگرسیون، خطی چندگانه نامیده می‌شود. که در رابطه (۲.۲) تعریف شده است.

- رگرسیون لجستیک: رگرسیون لجستیک برای یافتن احتمال رویداد مبنی بر موفقیت و شکست به‌کار می‌رود. تابع پیوند این مدل رگرسیونی عضو خانواده GLM به صورت

$$\operatorname{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \mathbf{X}\boldsymbol{\beta} = \sum_{j=1}^{k} x_i\beta_j, \quad i = 1, 2, \cdots, n$$

تعریف می‌شود که در آن

$$p = P(y_i = 1 \mid \mathbf{X} = \mathbf{x}) = \frac{\exp\{\beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}\}}{1 + \exp\{\beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}\}}.$$

- رگرسیون لجستیک چندگانه: به حالتی گفته می‌شود که متغیر وابسته بیش از یک سطح دارد. برای نمونه اگر بخواهیم ببینیم که چه عواملی موجب می‌شود که مخاطبان تلویزیونی، کدام یک از شبکه‌های یک یا دو یا سه را برای دیدن انتخاب کنند از رگرسیون لجستیک چندگانه استفاده می‌کنیم. در این مثال، عوامل مختلف به عنوان متغیر مستقل و انتخاب شبکه به عنوان متغیر وابسته شناخته می‌شود. این متغیر وابسته سه سطح دارد که پاسخ آن بله یا خیر است (استفاده یکی از شبکه‌ها یک، دو یا سه).

- رگرسیون چندجمله‌ای: یک مدل رگرسیون را رگرسیون چندجمله‌ای گوییم اگر توان یکی از متغیرهای مستقل بیش از یک باشد. مثلا رابطه

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

یک مدل رگرسیون چندجمله‌ای را نشان می‌دهد. در این روش رگرسیون، مناسب‌ترین الگوی برازش، خط مستقیم نیست بلکه یک منحنی است که متناسب با نقاط رسم می‌شود.

- رگرسیون پواسون: یک مدل خطی عمومی برای تجزیه و تحلیل رگرسیونی است که برای مدل داده‌های قابل شمارش و جدول‌ها احتمالات استفاده می‌شود. این مدل متعلق به خانواده GLM است، رگرسیون پواسون فرض می‌کند متغیر پاسخ Y دارای توزیع پواسون است. مدل رگرسیونی پواسون به عنوان مدل خطی لگاریتمی نیز شناخته می‌شود، به ویژه هنگامی‌که مدل‌های جدول احتمالاتی مورد استفاده قرار گیرند. به بیان دیگر، زمانی که فراوانی متغیر وابسته برای فرآیند مدل‌سازی کم باشد و به‌اصطلاح داده‌های مشاهده شده دارای فراوانی صفر باشند در این حالت مدل‌های معمولی توانایی پیش‌بینی مناسب فراوانی متغیر وابسته را ندارد. دلیل اصلی این امر فرض توزیع نرمال داده‌ها است در این شرایط مدل رگرسیونی پواسون می‌تواند مورد استفاده قرار گیرد. در رگرسیون پواسون متغیر پاسخ به صورت

$$\log(E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x})) = \mathbf{x}\boldsymbol{\beta}$$

مدل می‌شود. با داشتن بردار پارامتر رگرسیون پواسون β و متغیرهای پیشگو $\mathbf{X}$، می‌توان پیش‌بینی را به صورت زیر به‌دست آورد

$$E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x}) = \exp\{\mathbf{x}\beta\}$$

- رگرسیون رتبه‌ای: در رگرسیون رتبه‌ای تاثیر متغیر یا متغیرهای مستقل بر روی متغیر وابسته‌ای که سطوح مختلف و رتبه‌ای دارد، مدل‌بندی می‌شود.
- رگرسیون پروبیت: هنگامی‌که پاسخ (متغیر وابسته) دو حالت داشته باشد (بله یا خیر) و هدف بررسی شدت متغیر توضیحی در پیش‌بینی این دو حالت باشد، از مدل رگرسیون پروبیت استفاده می‌شود. برای اطلاعات بیشتر در خصوص رده مدل‌های GLM و کاربردهای آن مک کولاک و نلدر (۱۹۸۹) را ببینید.

۴.۲ روش برآورد

به محض اینکه با استفاده از طرح آزمایش‌های خاصی داده‌ها جمع‌آوری شوند، فرآیند برآورد آغاز می‌شود. این فرآیند اشاره به روش به‌دست آوردن ابزارهای بخش–ارزشی از داده‌های جمع‌آوری شده دارد (کوال‌تریکس، ۲۰۱۱). هنگامی که می‌خواهیم رهیافتی را برای برآورد پارامترها در نظر بگیریم، شکل متغیر پاسخ (متغیر وابسته) باید در نظر گرفته شود. گرین و سیرینی‌واسان (۱۹۷۸) رهیافت‌های مختلف برآورد را به سه دسته (۱) پاسخ‌های مقیاس‌پذیر مرتب[۳]، (۲) پاسخ‌های مقیاس‌پذیر فاصله‌ای[۴] و (۳) پاسخ‌های مقایسه‌ای جفت شده[۵] تقسیم نموده‌اند.

۱.۴.۲ پاسخ‌های مقیاس‌پذیر مرتب

برخی از روش‌های برآورد که می‌تواند برای داده‌ها با پاسخ‌های مقیاسی مرتب به کار رود به صورت زیر هستند

تحلیل واریانس یکنوا

ابتدایی‌ترین روش، رهیافت تحلیل واریانس یکنوا[۶] (MONANOVA) است که به تابع بخش–ارزشی محدود شده است. در برخی از مطالعات از این روش به روش MONANOVA کروسکال[۷]

[۳] Ordinally scaled responses
[۴] Interval scaled responses
[۵] Paired comparison responses
[۶] Monotone analysis of variance
[۷] Kruskal's monotone analysis of variance

(۱۹۶۵) نیز نام برده می‌شود. یک تبدیل یکنوا منجر به این می‌شود که داده‌ها با استفاده از تابعی که می‌تواند غیرنزولی یا غیرصعودی باشد تبدیل شوند. این روش اساسا این حقیقت را بیان می‌کند که استفاده از تبدیل‌های یکنوا روی متغیرهای پاسخ می‌تواند مجموع توان دوم باقی‌مانده‌ها[۸] (RSS) را کمینه کند.

نگاشت اولویت

رهیافت نگاشت اولویت[۹] (PREFMAP) سعی می‌کند تا تصویری از اطلاعات ارزشمند در مورد هر یک از پاسخ‌های مربوط به هر متغیر را نشان دهد. هنگامی که داده‌های نگاشت شده را با جدول‌ها مقایسه می‌کنیم، تصویرسازی کمک می‌کند که چگونه ویژگی‌های محصول، پاسخ‌های مصرف‌کنندگان را تحت تاثیر قرار می‌دهد. نمودار تصویری روشن از ارتباط بین محصولات و تفاوت هر یک از آن‌ها از دیدگاه اولویت‌بندی مصرف‌کنندگان این محصولات نشان می‌دهد. برای جزئیات بیشتر ال‌مور و همکاران (۱۹۹۹) را ببینید.

LINMAP

بهترین روش برای برآورد توابع نقطه ایده‌آل، روش برآورد برنامه‌ریزی خطی برای تجزیه و تحلیل چند بعدی اولویت‌ها[۱۰] است. زیرا این روش از یک برنامه‌ریزی خطی در جایی که سایر تکنیک‌ها از روش محاسباتی کلاسیک بهره می‌گیرند استفاده می‌کند (گرین و سیرینی‌واسان، ۱۹۷۸). مزیت دیگر این روش دستیابی به مقدار بهینه سراسری[۱۱] پارامترها است در حالی‌که در روش‌های دیگر ضمانتی برای دستیابی به این پارامترها وجود ندارد. در صورتی‌که محقق از دانش قبلی برخوردار باشد، می‌تواند محدودیت‌های خاصی را به مدل اضافه کند مثلا می‌تواند وزن‌ها را به مقادیر نامنفی محدود کند یا توابع بخش–ارزشی می‌توانند به یکنوایی بیشتری محدود شوند با افزودن محدودیت‌های دانش گذشته دقت برآوزدها افزایش پیدا می‌کند.

۲.۴.۲ پاسخ‌های مقیاس‌پذیر فاصله‌ای

زمانی‌که مقیاس پاسخ‌ها، فاصله‌ای است، مطالعات تحلیل متقارن کلاسیک، چهار روش برآورد را برای برآورد مقادیر بخش–ارزشی بیان می‌کند. روش‌های برآورد عبارت‌اند از

۱. رگرسیون OLS (رایج‌ترین روش استفاده شده در برآورد است)

۲. رگرسیون کمترین مجموع قدرمطلق خطاها[۱۲] (MSAE)

---

[۸] Residual sum of squares

[۹] Preference maping

[۱۰] Linear programming for multi dimen analysis of preference

[۱۱] Global

[۱۲] Minimum sum of absolute errors

۳. رگرسیون کلاس پنهان[۱۳]

۴. رگرسیون بیز سلسله‌مراتبی[۱۴]

روش برآورد آخری می‌تواند برای مدل‌های دیگر با پاسخ‌های متفاوت از مقیاس فاصله‌ای نیز به کار رود.

رگرسیون کمترین توان‌های دوم معمولی

رگرسیون OLS یکی از اصلی‌ترین روش‌های برآورد است که در مطالعات تحلیل متقارن به جهت کاربرد ساده‌ای که دارد به‌کار می‌رود. همچنین این روش می‌تواند با تعداد زیادی از متغیرها و سطح‌های آن‌ها مواجه شود و خطاهای استاندارد برای برآورد پارامترها ارائه دهد. رگرسیون OLS یک روش آماری است که از داده‌های نمونه استفاده می‌کند تا ارتباط درست بین متغیر وابسته (Y) و یک متغیر مستقل (X) را برآورد کند.

در یک مطالعه تحلیل متقارن، این روش مجموعه‌ای از ابزارهای بخش–ارزشی را برای هرسطح از ویژگی انتخاب شده برای هر نفر به‌دست می‌آورد. با یافتن میانگین مقادیر بخش–ارزشی روی همه پاسخ‌ها، مدل تجمیع‌شده بدست می‌آید و برآوردهای تجمیع‌شده به‌طور مساوی خوب یا به‌طور مساوی ضعیف روی همه پاسخ‌ها به‌کار برده می‌شود مدل همچنین با در نظر گرفتن اثرات متغیرهای مشاهده شده در تفسیر خود، آن‌ها را نیز به‌کار می‌گیرد از این مدل به منظور دست‌یابی به برآورد دقیق منطقی استفاده می‌شود اما زمانی که تغییرات زیادی بین پاسخ‌دهندگان وجود دارد خطای مدل نیز می‌تواند زیاد شود برای بررسی این عدم تجانس[۱۵] (ناهمگونی) بین پاسخ‌دهندگان مثال ساده‌ی انتخاب مواد روی پیتزا توسط مشتریان را در نظر بگیرید. فرض کنید نیمی از مشتریان زیتون را دوست دارند و نیمی دیگر علاقه‌ای به زیتون ندارند با به‌کار بردن روش برآورد OLS به این حقیقت دست پیدا می‌کنیم که برآوردهای تجمیع شده نشان می‌دهند که مشتریان وجود یا عدم وجود زیتون روی پیتزا را خیلی با اهمیت نمی‌دانند در حالی که قطعا چنین نتیجه‌ای درست نمی‌باشد. مشتریانی که پاسخ داده‌اند زیتون را دوست ندارند، قطعا پیتزایی که روی آن زیتون وجود داشته باشد را خریداری نمی‌کنند این مثال لزوما بررسی دقیق عدم تجانس بین مشاهدات را به‌خوبی آشکار می‌کند.

رگرسیون کلاس پنهان

فرض کنید که محققی می‌داند دو مجموعه متفاوت از مشاهدات وجود دارد و این‌که او اطلاعاتی دارد که براساس آن‌ها می‌تواند این گروه‌ها را شناسایی کند. با تجزیه و تحلیل جداگانه هر

[۱۳] Latent class regression

[۱۴] Hierarchical Bayes regression

[۱۵] heterogeneity

گروه محقق می‌تواند برآوردهای دقیقی به‌دست آورد. مدل رگرسیون کلاس پنهان می‌تواند به عنوان یک تغییر ساده از روش OLS توصیف شود. تنها تفاوتی که وجود دارد این است که برآورد OLS به صورت شرطی روی عضویت افراد به‌کار برده می‌شود (دی‌سربو، ۲۰۰۶)

رگرسیون کمینه مجموع قدرمطلق خطا

رگرسیون کمینه مجموع قدرمطلق خطا (MSAE) روشی است که اغلب زمانی استفاده می‌شود که رگرسیون OLS نامناسب یا ناکافی است. در برخی موارد، مجموع قدرمطلق خطاها به سادگی رهیافت رضایت‌بخش‌تری از تابع درجه دوم است. این روش برآورد حتی استوارتر[۱۶] از روش OLS است و به محقق اجازه می‌دهد تا محدودیت‌های پیشین روی پارامترهای برآورد شده قرار دهد (گرین و سیرینی‌واسان، ۱۹۷۸). از آنجایی که برنامه‌ریزی خطی استفاده شده است، شرط‌ها به آسانی قرار داده می‌شود (رودمن، ۱۹۷۴).

۳.۴.۲ پاسخ‌های مقایسه‌ای جفت‌شده (زوجی)

هنگامی‌که پاسخ‌ها مربوط به مقایسه‌های زوجی هستند، روش‌های برآورد فرض می‌کنند که مقایسه‌های زوجی از نظر آماری مستقل هستند این مدل‌ها مقیاس‌های زوجی را به مدل‌های احتمالاتی ارتباط می‌دهند (گرین و سیرینی‌واسان، ۱۹۷۸). روش پاسخ‌های جفتی دارای مدل‌های به صورت زیر می‌باشد که در مورد آن‌ها در قسمت‌های قبل به‌طور مختصر شرح دادیم.

۱. مدل لجیت

۲. مدل لجستیک چندجمله‌ای

۳. مدل پروبیت

[۱۶] Robust

فصل ۳

تحلیل متقارن با اثرات ثابت

۱.۳ مقدمه

روش تحلیل متقارن طبق مطالعاتی که توسط واتینک و کاتین (۱۹۸۹) و واتینک و همکاران (۱۹۹۴) انجام شد برای مصرف‌کنندگان مبادله پایاپای از بیشترین محبوبیت برخوردار است. همچنین این روش در جستجوی بازار برای تولیدکنندگان هم بیشترین استفاده را دارد. دلیل محبوبیت تحلیل متقارن بین مصرف‌کننده و تولیدکننده این است که مدیریت خاصی نسبت به سوالات پیش‌آمده از هر دو طرف دارد مثلا چرا مصرف‌کننده، یک برند خاص از یک تولیدکننده را انتخاب می‌کند. در این فصل مدل نقل و انتقالات طراحی تحلیل متقارن و استفاده آن در بازار و مدل اثرات ثابت[۱] را مورد بررسی قرار می‌دهیم. این فصل با بررسی تحلیل متقارن نقل‌وانتقالات[۲] شروع می‌شود و در ادامه روش معمولی– مدل اثرات ثابت بیان و در آخر با دو مثال کاربردی پایان می‌پذیرد. از آن‌جایی که برنامه‌های این مجموعه براساس نرم‌افزار SAS است و تنها یک بسته در نرم‌افزار R وجود دارد که به تحلیل متقارن پرداخته است، در این فصل یکی از مثال‌های کاربردی را با نرم‌افزار R تحلیل کرده تا این مجموعه دربرگیرنده هر دو نرم‌افزار پرکاربرد در این تحلیل متقارن باشد.

---

[۱] Approach - fixed effects model

[۲] Traditional

۲.۳ مدل نقل و انتقالات

امروزه مصرف‌کنندگان (افراد) برای انتخاب یک محصول با تصمیمات زیادی روبرو می‌شوند که این تصمیمات برای تولیدکننده از اهمیت زیادی برخوردار است. تولیدکننده اطلاعات بدست آمده از این تصمیمات را جمع‌آوری کرده تا بازاریابی کند. این اطلاعات برای اولویت‌بندی محصول تولیدی مفید است تا بتوان فهمید مشتریان (افراد) چه موردی را بیشتر پسند می‌کنند. اگرچه مصرف‌کننده نسبت به محصول تازه‌ای که وارد بازار می‌شود و محصول مصرفی قبلی عکس‌العمل نشان می‌دهد، ممکن است محصول قبلی محبوبیت خود را داشته باشد یا محصول جدید جای محصول قدمی را بگیرد.

این روش‌ها از مدل اولویت بخش‌ارزشی که در فصل یک بیان شد (بخش ۳.۴.۱) استفاده می‌کند. برای برآورد ویژگی‌های بخش ارزشی از برآوردگر OLS استفاده می‌شود، و پاسخ‌ها همان امتیازاتی است که مشتریان به محصول می‌دهند که برای اولویت‌بندی استفاده می‌شود. یکی از مزیت‌های برآوردگر OLS در واقع حسابی است که برای ناهمگونی پارامتر اختیاری برای سطح مصرف‌کنندگان فردی در نظر می‌گیرد. (فرویرس_اسچیتر و اتر، ۱۹۹۹) چون از ویژگی‌های ماتریس طرح این است که پاسخ‌ها کمی و متغیرهای توضیحی کیفی می‌باشد، و این یک مطالعه ساده بخش‌ارزشی است که بیشتر برای تحلیل متقارن استفاده می‌شود (فیچت و همکاران، ۲۰۱۱).

گسترش مدل‌های رگرسیونی برای بهبود قابلیت پیشگویی است که در ادامه فصل به بررسی تحلیل متقارن در مدل رگرسیونی با اثرات ثابت می‌پردازیم. در حقیقت این اثرات ثابت ضرایب رگرسیونی مدل می‌باشد. یک مدل باید به گونه‌ای ساخته شود که بتواند اختلاف بین ساختار اولویت‌های انتخابی بین مشتریان که پاسخ‌های آن متفاوت است را شرح دهد. چون پاسخ‌های مختلف براساس چندین مشخصه اجتماعی طبقه‌بندی می‌شود پس هرکدام از پاسخ‌ها تاثیرات گوناگونی بر مدل طراحی شده دارند، و بررسی این تاثیرت از اهمیت زیادی برخوردار است. برای درک بهتر موضوع در ادامه، یک مثال که توسط الیسون وچریستیکس(۱۹۹۴) مطرح شده را بررسی می‌کنیم. در این مثال پزشکان می‌خواهند زندگی افرادی که ترک اعتیاد کردن را بررسی کنند. پرسشنامه‌ای که توسط ۴۷۵ پزشک طراحی شده در جدول ۱.۳ آمده است.

جدول ۱.۳: پرسشنامه پزشکان

نظر متفاوت پزشکان در مورد خروج و حفظ درمان معتادین، ارائه بسته درمان است. در کل انواع روش درمان و حفظ زندگی که شما مایلید یا مایل نیستید طراحی کنید در این شرایط چیست؟
امتیازها از یک تا هشت میباشد به گونه‌ای که بیشترین تمایل ۱ و کمترین تمایل به درمان ۸ می‌باشد. همزمان از چند نفر برای روش درمان طراحی شده استفاده می‌شود که امتیازها و رفتار آن‌ها ثبت می‌شود.
آنتی‌بیوتیک ☐ محصولات خونی ☐ تزریق رگی مایعات ☐ تزریق رگی واسوپرسور ☐ تهویه فیزیکی ☐ دیالیز کلیه ☐ تمام تغذیه تزریقی باشد ☐ تغذیه و مایعات لوله‌ای باشد ☐
اگر یکی یا دوتا از افراد به هم وابسته بودند آنگاه شرایط پایین را چک کن
درمان پزشکی طراحی نمی‌شود طراحی‌ها به یک اندازه یا مشکل باشد

برآورد موارد مصرف‌شده برای هر ۸ نوع روش درمانی گفته شده در جدول ۱.۳ با استفاده از روش تحلیل متقارن معمولی بدست می‌آید. توجه داشته باشید در این تحلیل متقارن، تنها تفاوت روش‌های درمانی در نظر گرفته می‌شود و اختلافات میان پزشکان در تحلیل لحاظ نمی‌شود. امکان اختلاف پزشکان و نظرات مختلف آن‌ها و بررسی این اختلاف نظرها در بوکنهولت (۲۰۰۱) شرح داده شده که عوامل پایه آن در زیر گفته می‌شود:

- هزینه درمان طبق وزن و اهمیتی که پزشکان به آن می‌دهند، متفاوت است.
- پزشکان اختلاف ارزیابی خودشان از ویژگی‌های درمان که بینشان روشن نیست را ذکر کنند (یعنی آن ویژگی‌هایی که برای هر پزشک سؤال هست را بیان کنند تا با کمک همه پزشکان رفع شود) و یا شامل مطالعه دوباره قرار دهند. برای مثال درباره سوال گفته شده‌ای که در مورد اثرات اندازه درمان نیست.

- آموزش پزشکان برای درمان به صورت معمولی یا همه‌جانبه باشد، یا ممکن است بوسیله پاسخ‌های مختلفی که بین خودشان است به عنوان راهنما استفاده کنند.

این اختلاف در درمان می‌تواند عدم تجانس پنهان در وزن‌های رگرسیونی و ارزیابی در مشخصه‌ها را شرح دهد. ابتدا فهمیدن اختلاف بین مفهوم تاثیرات ثابت و تاثیرات تصادفی از اهمیت حیاتی برخوردار است. ترکیب این دو نوع اثر مدل خطی آمیخته را تشکیل می‌دهد، که در فصل بعدی به آن می‌پردازیم.

۳.۳ روش معمولی ـ مدل اثرات ثابت

مدل‌ها با اثرات ثابت شامل پارامترهایی هستند که با یک هدفی انتخاب و توسط محقق ثابت می‌شوند. در رگرسیون خطی کلاسیک ورودی‌ها خطی در نظر گرفته می‌شود و توسط محقق ثابت می‌شود. مطالعات نشان می‌دهند که پیش‌بینی کردن مدل غیر خطی غالبا راحت‌تر می‌باشد اما باعث می‌شود اریبی مدل زیاد شود. در این روش معمولی تحلیل متقارن، یک مدل رگرسیون خطی برای هر فرد که بین مشخصه‌های مختلف همبستگی ندارد در نظر گرفته می‌شود. مجموع مدل می‌تواند میانگین کلی پارامترها را برای همه پاسخ‌دهندگان تعیین کند. مطالب این بخش عمدتا برگرفته از جانسون و ویچرن (۲۰۰۷) است.

۱.۳.۳ مدل رگرسیون خطی کلاسیک

مدل رگرسیون خطی کلاسیک[۳] برای پیش‌گویی مقدار پاسخ Y به‌وسیله مجموعه‌ای از متغیرهای توضیحی X است و می‌توان اثرات متغیرهای توضیحی را بر متغیر پاسخ مشاهده کرد، که این مورد را در فصل دوم شرح دادیم. در ادامه برآورد پارامترها به روش OLS و ماکزیمم درستنمایی[۴] (ML) را شرح می‌دهیم. برای این منظور مدل رگرسیون خطی زیرا در نظر بگیرید

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \qquad (۱.۳)$$

که در آن $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)^\top$ برای مؤلفه‌های پاسخ، $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)^\top$، $\mathbf{x}_i \in R^p$، iامین بردار متغیرهای توضیحی، $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^\top$ بردار ضرایب رگرسیونی و $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^\top$ بردار مؤلفه‌های خطای تصادفی است.

[۳] Classical linear regression model

[۴] Maximum likelihood

برآورد پارامتر

در مدل رگرسیون خطی، فرض می‌کنیم $\mathbf{X}$، $\mathbf{X} \in R^{n\times(p+1)}$ غیرتصادفی با رتبه کامل ستونی بوده و ε بردار خطا با میانگین صفر و واریانس ثابت است. به عبارتی داریم

$$E(\boldsymbol{\varepsilon}) = 0, \qquad \mathrm{Var}(\boldsymbol{\varepsilon}) = E(\varepsilon^2) = \sigma^2 \mathbf{I}_n$$

در این صورت نتایج حاصل از برآورد به روش‌های OLS و ML برای پارامتر σ^2 و β یکسان خواهد بود.

- **روش OLS:**
با استفاده از مینیمم کردن مجموع توان‌های دوم باقی‌مانده‌ها (RSS)، βها را تعیین می‌کنیم که در آن:

$$\begin{aligned} RSS(\boldsymbol{\beta}) &= \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \\ &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

با مشتق‌گیری از $RSS(\boldsymbol{\beta})$ نسبت به بردار $\boldsymbol{\beta}$ و مساوی صفر قرار دادن نتیجه می‌شود

$$\frac{\partial RSS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = 0$$

لذا می‌توان نتیجه گرفت

$$(\mathbf{X}^\top \mathbf{X})\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y}$$

حال از آنجایی که ماتریس $\mathbf{X}$ رتبه کامل ستونی است پس $\mathbf{X}^\top\mathbf{X}$ معکوس‌پذیر بوده و در نهایت برآوردگر OLS بردار پارامتر $\boldsymbol{\beta}$ به صورت زیر بدست می‌آید

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}. \tag{۲.۳}$$

برآورد OLS برای β ناریب است، زیرا

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\mathbf{Y}) \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

از طرفی

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}((\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top\mathbf{Y}) \\
&= (\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top\mathrm{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱} \\
&= (\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top(\sigma^۲ I_n)\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱} \\
&= \sigma^۲(\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱} \\
&= \sigma^۲(\mathbf{X}^\top\mathbf{X})^{-۱}.
\end{aligned}$$

اگر بردار خطای مدل دارای توزیع نرمال باشد، تحت مدل رگرسیون خطی کلاسیک برآوردگر OLS دارای توزیع نرمال به صورت زیر است.

$$\hat{\boldsymbol{\beta}} \sim N_{p+۱}(\boldsymbol{\beta}, \sigma^۲(\mathbf{X}^\top\mathbf{X})^{-۱}). \tag{۳.۳}$$

توجه کنید که چون $\hat{\boldsymbol{\beta}} = (\hat{\beta}_۰, \hat{\beta}_۱, \cdots, \hat{\beta}_p)^\top$، $p+۱$ مؤلفه دارد توزیع آن نرمال $(p+۱)$ـ بعدی می‌شود. به این ترتیب برای برآورد پارامتر نامعلوم $\sigma^۲$ از باقی‌مانده‌ها استفاده می‌کنیم. اگر $\hat{\sigma}^۲$ برآوردگر OLS پارامتر $\sigma^۲$ باشد و $\hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ می‌توان نوشت

$$\begin{aligned}
\hat{\sigma}^۲ &= \frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{n-(p+۱)} \\
&= \frac{\mathbf{Y}^\top[\mathbf{I}-\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top]\mathbf{Y}}{n-(p+۱)} \\
&= \frac{\mathbf{Y}^\top[\mathbf{I}-\mathbf{H}]\mathbf{Y}}{n-(p+۱)}
\end{aligned}$$

که ماتریس $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top$ یک ماتریس خودتوان است و

$$\begin{aligned}
\hat{\varepsilon} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-۱}\mathbf{X}^\top\mathbf{Y} \\
&= (\mathbf{I}-\mathbf{H})\mathbf{Y} \sim \mathcal{N}_n\left((\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\beta}, (\mathbf{I}-\mathbf{H})^۲\sigma^۲\right)
\end{aligned}$$

از آنجایی که $\mathbf{H}$ خودتوان است، $(\mathbf{I}-\mathbf{H})$ نیز خودتوان بوده و داریم $(\mathbf{I}-\mathbf{H})^۲ = (\mathbf{I}-\mathbf{H})$ همچنین

$$\begin{aligned}
(\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\
&= ۰
\end{aligned}$$

بنابراین داریم
$$(\mathbf{I}-\mathbf{H})^{\frac{1}{2}}\frac{\hat{\varepsilon}}{\sigma} \sim \mathcal{N}_n(0,1)$$
لذا
$$\frac{\hat{\varepsilon}^\top(\mathbf{I}-\mathbf{H})\hat{\varepsilon}}{\sigma^2} \sim \chi^2$$
از آنجایی که $(\mathbf{I}-\mathbf{H})$ خودتوان است نتیجه می‌شود

$$\begin{aligned}
\mathrm{Rank}(\mathbf{I}-\mathbf{H}) &= \mathrm{tr}(\mathbf{I}-\mathbf{H}) \\
&= n - \mathrm{tr}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{tr}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] \\
&= n - \mathrm{tr}[\mathbf{I}_{p+1}] \\
&= n-(p+1)
\end{aligned}$$

در این صورت
$$\frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{\sigma^2} \sim \chi^2_{n-(p+1)}$$
و
$$E\left(\frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{\sigma^2}\right) = n-(p+1)$$
در آخر
$$E(\hat{\sigma}^2) = E\left(\frac{\hat{\varepsilon}^\top\varepsilon}{n-(p+1)}\right) = \sigma^2.$$
پس نتیجه می‌گیریم که $\hat{\sigma}^2$ یک برآوردگر نااریب برای σ^2 است.

- **روش ML:**

اگر شرایطی که در قسمت ابتدایی برآورد پارامتر گفته شد برقرار باشد برآوردگرهای OLS و ML یکی می‌شود. یعنی برآوردگر ML بردار پارامتر β همان رابطه (۲.۳) است.

برآورد واریانس و تعیین آن با استفاده از روش ML کمی متفاوت از روش OLS است. چون $\mathbf{Y}$ دارای توزیع نرمال است، تابع درستنمایی عبارت است از
$$L(\boldsymbol{\beta},\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\exp\left\{\frac{-(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right\}$$
با لگاریتم‌گیری از تابع درستنمایی داریم
$$\begin{aligned}
\log L &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(\mathbf{Y}^\top\mathbf{Y}-\mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta}-\mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{Y}+\mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}
\end{aligned}$$

با مشتق‌گیری نسبت به σ^2 نتیجه می‌شود:

$$\begin{aligned}\frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2(\sigma^2)} - \frac{(\mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{Y} + \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{X}\boldsymbol{\beta})}{2\sigma^4} \\ &= \frac{2\sigma^4}{n}\left[-\frac{n}{2(\sigma^2)} - \frac{(\mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{Y} + \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{X}\boldsymbol{\beta})}{2\sigma^4}\right] \\ &= -\sigma^2 + \frac{1}{n}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\end{aligned}$$

با توجه به این که مشتق دوم نسبت به σ^2 منفی است، با مساوی صفر قرار دادن مشتق فوق برآوردگر ML پارامتر σ^2 به صورت زیر حاصل می‌شود.

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{n}, \qquad \hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta}.$$

که در آن $\hat{\beta}$ برآوردگر ML بردار پارامتر β هست.

آزمون‌ها برای پارامترهای رگرسیونی

بخش مهمی از تحلیل‌های رگرسیونی ارزیابی متغیرهای پیشگو و پیشگویی متغیر پاسخ است و برای این ارزیابی و پیشگویی می‌توان از آزمون‌های فرضیه استفاده کرد. برای آزمون فرضیه از جدول تحلیل واریانس[۵] استفاده می‌شود که در جدول ۲.۳ آمده است.

جدول ۲.۳: جدول تحلیل واریانس

منبع تغییر	مجموع توان‌های دوم SS	درجه آزادی	میانگین SS
رگرسیونی	$\mathrm{SSR} = \hat{\boldsymbol{\beta}}X^TY - (\frac{1}{n})Y^TjY$	$p-1$	$\mathrm{MSR} = \frac{\mathrm{SSR}}{p-1}$
خطا	$\mathrm{SSE} = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = Y^TY - b^TX^TY$	$n-p$	$\mathrm{MSE} = \frac{\mathrm{SSE}}{n-p}$
کل	$\mathrm{SST} = Y^TY - (\frac{1}{n})Y^TjY$	$n-1$	

در جدول ۲.۳ $\hat{\beta}$ یک بردار $1 \times (p+1)$ بعدی از بردارهای ضرایب رگرسیونی است که از طریق روش OLS به‌دست می‌آید و j یک ماتریس $n \times n$ از یک می‌باشد. اگر تغییرات SSR خیلی بزرگ باشد، می‌توان گفت که اختلاف قابل ملاحظه‌ای بین گروه‌های وجود دارد.

آزمون کلی

دسته فرضیه‌های زیر را در نظر بگیرید:

$$\begin{cases} H_0: \ \beta_0 = \beta_1 = \cdots = \beta_p = 0 \\ H_1: \ \beta_j \neq 0 \qquad\qquad \forall \quad j = 0, 1, \cdots, p \end{cases}$$

[۵] Analysis variance

وقتی H_0 رد می‌شود می‌توان نتیجه گرفت که حداقل یکی از β_jها مخالف صفر است. از آماره آزمون $F^* = \frac{\text{MSR}}{\text{MSE}}$ می‌توان برای آزمون فرضیه صفر H_0 استفاده کرد. تحت فرضیه صفر منبع تغییرات دارای توزیع‌های زیر می‌باشد:

$$\frac{\text{SSR}}{\sigma^2} \sim \chi^2_p \qquad \frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-(p+1)}$$

چون این توزیع‌ها مستقل می‌باشند پس آماره آزمون دارای توزیع فیشر با p و $n-(p+1)$ درجه آزادی است. فرضیه صفر رد می‌شود اگر و فقط اگر

$$F^* > F_{p,n-(p+1)(\alpha)}$$

یعنی آماره آزمون بزرگتر از مقدار بحرانی باشد، که در آن $F_{p,n-(p+1)}(\alpha)$ مقدار بحرانی بالایی سطح α در توزیع F با p و $n-(p+1)$ درجه آزادی است.

آزمایش تکی ضرایب رگرسیونی برای آزمون معنی‌داری ضرایب تکی زمانی که تعداد β_jها کم باشد از آزمون t استفاده می‌کنیم که در این حالت دسته فرضیه به‌صورت زیر است

$$\begin{cases} H_0 : \ \beta_j = 0 \\ H_1 : \ \beta_j \neq 0 \quad \forall \quad j = 1, 2, \cdots, p \end{cases}$$

تحت فرضیه صفر آماره آزمون زیردارای توزیع tــاستیودنت با $n-(p+1)$ درجه آزادی است که در آن

$$t = \frac{\hat{\beta}_j}{\text{S.e.}(\hat{\beta}_j)}$$

و $Se(\hat{\beta}_j)$ خطای استاندارد $\hat{\beta}_j$ می‌باشد. فرضیه صفر رد می‌شود اگر قدرمطلق مقدار آماره آزمون از مقدار بحرانی در سطح α بزرگتر باشد. **آزمون زیر مجموعه‌ای از متغیرهای توضیحی**

گاهی اوقات ممکن است از زیر مجموعه‌ای از داده‌ها که در پیشگویی متغیر پاسخ تاثیر دارند استفاده شود که این کار را می‌توان با آزمون نسبت درستنمایی انجام داد. متغیرهایی که برای دیدن تاثیر آن‌ها بر پاسخ، آزمون می‌شوند مانند $X_{a+1}, X_{a+2}, \cdots, X_p$ برچسب‌گذاری می‌شوند.(دراپر و اسمیت، ۱۹۹۸) در این حالت دسته فرضیه‌های آزمون مورد نظر به صورت زیر است

$$\begin{cases} H_0 : \beta_{a+1} = \beta_{a+2} = \cdots = \beta_p \\ H_1 : \boldsymbol{\beta}_{(2)} \neq 0 \end{cases}$$

که در آن

$$\boldsymbol{\beta}_{(2)} = \{\beta_{a+1}, \beta_{a+2}, \cdots, \beta_p\}$$

اگر فرضیه صفر رد نشود آن‌گاه متغیرهای برچسب‌گذاری شده تاثیری بر متغیر پاسخ ندارد چون ضرایب زیر مجموعه دوم برابر صفر می‌شود.

به این دلیل که متغیرهای توضیحی به دو زیر مجموعه تقسیم شده‌اند بردار ضرایب هم به دو زیر بردار به صورت زیر افراز می‌شود

$$\mathbf{X} = \left[\mathbf{X}_{(1)_{n\times(a+1)}} \quad | \quad \mathbf{X}_{(2)_{n\times(p-q)}}\right] \quad , \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{(1)_{(a+1)\times 1}} \\ \hline \boldsymbol{\beta}_{(2)_{(p-q)\times 1}} \end{bmatrix}$$

در این حالت مدل خطی را می‌توان به صورت زیر بازنویسی کرد

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \varepsilon \\ &= \mathbf{X}_{(1)}\boldsymbol{\beta}_{(1)} + \mathbf{X}_{(2)}\boldsymbol{\beta}_{(2)} + \varepsilon\end{aligned}$$

اگر فرضیه صفر درست باشد زیر مدل درست عبارت است از

$$\mathbf{Y} = \mathbf{X}_{(1)}\boldsymbol{\beta}_{(1)} + \varepsilon$$

از آنجایی که $\boldsymbol{\beta}_{(2)} = \mathbf{O}$، SSR به $\mathrm{SSR}_{(1)}$ تبدیل شده و داریم

$$\mathrm{SSR}_{(1)} = \hat{\boldsymbol{\beta}}_{(1)}^{\top}\mathbf{X}_{(1)}^{\top}\mathbf{Y}.$$

آزمون درستنمایی تحت فرضیه صفر براساس مجموع توان‌های دوم اضافی محاسبه می‌شود که در آن

$$\text{مجموع توان‌های اضافی} = \mathrm{SSR}_{(1)} - \mathrm{SSR}$$

بنابراین فرضیه صفر در آزمون نسبت درستنمایی برای این مدل رد می‌شود اگر

$$\frac{\mathrm{SSR}_{(1)} - \mathrm{SSR}/(p-q)}{\mathrm{MSE}} > F_{(p-q),(n-p-1)}(\alpha)$$

که $F_{(p-q),n-(p+1)}(\alpha)$ مقدار بحرانی بالایی سطح α در توزیع F با $(p-q)$ و $(n-(p+1))$ درجه آزادی است.

۲.۳.۳ مثال کاربردی ماده شوینده

در این مثال در نظر داریم تا چگونگی انجام تحلیل متقارن با استفاده از یک مدل رگرسیونی با اثرات ثابت را در قالب یک مثال کاربردی شرح دهیم. یک کمپانی می‌خواهد اولویت پاسخ‌های افراد را برای ویژگی‌های مختلف ساختار یک تمیزکننده (ماده شوینده[۶]) مورد بررسی و ارزیابی قرار دهد(هایر و همکاران، ۲۰۰۶). در این مثال از نرم‌افزار SAS برای انجام تحلیل متقارن استفاده می‌شود.

این مواد شوینده دارای ۱۸ محصول هست که در هفت منطقه مورد بررسی قرار گرفته و تحلیل را برای ۸۶ پاسخ‌دهنده بررسی خواهیم کرد. برای انجام آزمایش تحلیل متقارن یک تیم، بازار داخلی را بررسی می‌کنند و تیمی دیگر براساس تحقیق آنها محصول را توسعه می‌دهند. در این مثال ۵ ویژگی با سطوح مختلف وجود دارد که در جدول ۳.۳ آمده است.

[۶] Industrial cleaner

جدول ۳.۳: ویژگی‌ها و سطوح ماده شوینده

ویژگی‌ها	سطوح
شکل محصول	۱.مخلوط ۲.کنسانتره ۳.پودر
تعداد کاربرد	۱. ۵۰ کاربرد ۲. ۱۰۰ کاربرد ۳. ۲۰۰ کاربرد
مواد ضدعفونی‌کننده	۱. دارد ۲. ندارد
تخریب‌پذیر زیستی	۱. هست ۲. نیست
قیمت	۱. ۳۵۰۰ تومان ۲. ۴۹۰۰ تومان ۳. ۷۹۰۰ تومان

در جدول ۴.۳ تعداد ۱۸ مشخصه مهم که برای ارزیابی بازار استفاده می‌شود معرفی شده است. این مشخصه‌ها با ویژگی‌ها و سطوحی که تعریف کردیم به‌دست آمده است. مثلا مشخصه اول را شرح می‌دهیم. با توجه به اعداد جدول ۴.۳ فرض کردیم در مشخصه اول، شکل محصول کنسانتره (عدد ۲ در جدول ۳.۳) و کاربرد محصول ۲۰۰ تایی (عدد ۳ در جدول ۳.۳) است. ضدعفونی‌کننده دارد (عدد ۱ در جدول ۳.۳) باعث تخریب زیست محیطی می‌شود (عدد ۱ در جدول ۳.۳) و قیمت آن ۳۵۰۰ تومان (عدد ۱ در جدول ۳.۳) است.

جدول ۴.۳: معرفی ۱۸ مشخصه در مثال ماده شوینده

قیمت	تخریب‌پذیر زیستی	ضدعفونی‌کننده	تعداد کاربرد	شکل محصول	مشخصه‌ها
۱	۱	۱	۳	۲	۱
۱	۱	۱	۳	۳	۲
۲	۲	۱	۲	۱	۳
۲	۲	۱	۳	۳	۴
۳	۱	۱	۱	۳	۵
۳	۲	۲	۳	۲	۶
۳	۱	۱	۲	۱	۷
۲	۱	۱	۳	۱	۸
۲	۱	۲	۲	۳	۹
۲	۱	۱	۱	۲	۱۰
۱	۱	۲	۲	۳	۱۱
۳	۱	۱	۲	۲	۱۲
۳	۱	۲	۳	۱	۱۳
۱	۱	۱	۱	۱	۱۴
۱	۲	۱	۲	۲	۱۵
۱	۲	۲	۱	۱	۱۶
۲	۱	۲	۱	۲	۱۷
۳	۲	۱	۱	۳	۱۸

برای این آزمایش چون متغیرهای توضیحی به‌صورت کیفی هستند آن‌ها را کدگذاری می‌کنیم. این‌گونه کد گذاری حضور یا غیبت سطح خاصی را در مشخصه نشان می‌دهد. برای هر ویژگی شامل k سطح، $k-1$ سطح مستقل برای برآورد استفاده می‌شود. برای مثال خاصیت شکل محصول سه سطح دارد که دو سطح آن برای طراحی ماتریس $\mathbf{X}$ استفاده می‌شود. روش کار به‌صورت زیر است

جدول ۵.۳: ساختن متغیرهای مصنوعی X

شکل محصول	$F_1 = \begin{cases} 1 & \text{کنسانتره} \\ 0 & \text{مخلوط} \\ -1 & \text{پودر} \end{cases}$	$F_2 = \begin{cases} 1 & \text{مخلوط} \\ 0 & \text{کنسانتره} \\ -1 & \text{پودر} \end{cases}$
تعداد کاربردها	$A_1 = \begin{cases} 1 & \text{۵۰ کاربرد} \\ 0 & \text{۱۰۰ کاربرد} \\ -1 & \text{۲۰۰ کاربرد} \end{cases}$	$A_2 = \begin{cases} 1 & \text{۱۰۰ کاربرد} \\ 0 & \text{۵۰ کاربرد} \\ -1 & \text{۲۰۰ کاربرد} \end{cases}$
ضدعفونی‌کننده و تخریب‌پذیر زیستی	$D_1 = \begin{cases} 1 & \text{نیست} \\ -1 & \text{هست} \end{cases}$	$B_1 = \begin{cases} 1 & \text{دارد} \\ -1 & \text{ندارد} \end{cases}$
قیمت	$P_1 = \begin{cases} 1 & \text{۳۵۰۰ تومان} \\ 0 & \text{۴۹۰۰ تومان} \\ -1 & \text{۷۹۰۰ تومان} \end{cases}$	$P_2 = \begin{cases} 1 & \text{۴۹۰۰ تومان} \\ 0 & \text{۳۵۰۰ تومان} \\ -1 & \text{۷۹۰۰ تومان} \end{cases}$

طراحی ماتریس X با استفاده از کدهای ساخته شده در جدول ۶.۳ آمده است.

جدول ۶.۳: ماتریس $\mathbf{X}$

مشخصه	ثابت	F_1	F_2	A_1	A_2	D_1	B_1	P_1	P_2
۱	۱	۰	۱	−۱	−۱	۱	۱	۱	۰
۲	۱	−۱	−۱	−۱	−۱	۱	۱	۱	۰
۳	۱	۱	۰	۰	۱	۱	−۱	۰	۱
۴	۱	−۱	−۱	−۱	−۱	۱	−۱	۰	۱
۵	۱	−۱	−۱	۱	۰	۱	۱	−۱	−۱
۶	۱	۰	۱	−۱	−۱	−۱	−۱	−۱	−۱
۷	۱	۱	۰	۰	۱	۱	۱	−۱	−۱
۸	۱	۱	۰	−۱	−۱	۱	۱	۰	۱
۹	۱	−۱	−۱	۰	۱	−۱	۱	۰	۱
۱۰	۱	۰	۱	۱	۰	۱	۱	۰	۱
۱۱	۱	−۱	−۱	۰	۱	−۱	۱	۱	۰
۱۲	۱	۰	۱	۰	۱	۱	۱	−۱	−۱
۱۳	۱	۱	۰	−۱	−۱	−۱	۱	−۱	−۱
۱۴	۱	۱	۰	۱	۰	۱	۱	۱	۰
۱۵	۱	۰	۱	۰	۱	۱	−۱	۱	۰
۱۶	۱	۱	۰	۱	۰	−۱	−۱	۱	۰
۱۷	۱	۰	۱	۱	۰	−۱	۱	۰	۱
۱۸	۱	−۱	−۱	۱	۰	۱	−۱	−۱	−۱

با توجه به جدول ۳.۳ مدل رگرسیون خطی کلاسیک با اثرات ثابت به‌صورت زیر است

$$y_{ij} = \beta_1 x_{j1} + \beta_{F_1} x_{j2} + \beta_{F_2} x_{j3} + \beta_{A_1} x_{j4} + \beta_{A_2} x_{j5} + \beta_{D_1} x_{j6} + \beta_{B_1} x_{j7} + \beta_{P_1} x_{j8} + \beta_{P_2} x_{j9} + \varepsilon_{ij}$$

که در آن $i = 1, 2, \cdots, 86$ تعداد پاسخ‌دهنده و $j = 1, 2, \cdots, 18$ تعداد مشخصه‌ها را نشان می‌دهد. مدل تحلیل متقارن معمولی، ویژگی‌های مختلف بخش‌ارزشی برای هر سطح از ویژگی برای هر پاسخ را برآورد می‌کند. مدل ماتریسی آن به صورت زیر است

$$y_i = \mathbf{X}_i \beta_i + \varepsilon_i \qquad i = 1, 2, \cdots, 86$$

که فرض می‌شود

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

y_i بردار پاسخ‌هایی است که پاسخ‌دهنده iام به ۱۸ مشخصه داده است یعنی امتیازاتی است که پاسخ‌دهندگان به ۱۸ مشخصه داده‌اند که این امتیازات بین ۰ تا ۱۰ می‌باشد. همچنین β_j بردار ضرایب بخش‌ارزشی است، $\mathbf{X}$ ماتریسی شامل کدهایی که از روی ویژگی‌ها و سطوح ساخته شده و ε_i بردار خطاها است.

بدون در نظر گرفتن نرمال بودن خطاها با استفاده از روش OLS برای برآورد مدل رگرسیون خطی یک مجموعه از ویژگی‌های بخش‌ارزشی را می‌توان برای هر پاسخ برآورد کرد. به عبارتی برای هر فرد پارامترها را باید برآورد کرد. برای توضیحات بیشتر و کامل کردن محاسبات به عنوان نمونه پاسخ‌دهنده اول را مورد بررسی قرار می‌دهیم.

پاسخ‌دهنده اول ($i = ۱$)

در جدول ۷.۳ برآوردهای OLS بخش‌ارزشی پاسخ‌دهنده اول آورده شده است.

جدول ۷.۳: برآورد بدست آمده بخش‌ارزشی برای نفر (پاسخ دهنده) اول در مثال ماده شوینده

$\hat\beta_۰$	$\hat\beta_{F_۱}$	$\hat\beta_{F_۲}$	$\hat\beta_{A_۱}$	$\hat\beta_{A_۲}$	$\hat\beta_{D_۱}$	$\hat\beta_{B_۱}$	$\hat\beta_{P_۱}$	$\hat\beta_{P_۲}$
۴/۱۱	$-$۰/۰۶	۰/۶۱	۰/۴۴	۰/۶۱	$-$۰/۲۱	۰/۵۴	۱/۴۴	۰/۹۴

این مقادیر نشان می‌دهد که او تمایل دارد به‌صورت کم یا زیاد از هر سطح از ۵ ویژگی استفاده کند. برای این‌که مجموعه‌ای کامل از مقادیر بخش ارزشی، که به صورت جداگانه توسط پاسخ‌دهنده اول از هر سطح از ویژگی‌ها مصرف شده، را به‌دست آوریم، محقق می‌تواند از ماتریس طراحی شده به‌وسیله کدها ($\mathbf{X}$) استفاده کند، به گونه‌ای که مجموع پارامترهای هر ویژگی باید برابر صفر شود به عبارتی باید داشته باشیم

$$\begin{cases} \beta_{F_۱} + \beta_{F_۲} + \beta_{F_۳} = ۰ \\ -۰/۰۶ + ۰/۶۱ + \beta_{F_۳} = ۰ \end{cases} \tag{۴.۳}$$

در نتیجه $\beta_{F_۳} = -۰/۵۵$.

پس به همین صورت برای پاسخ‌دهنده اول تمام ویژگی‌ها را می‌توان محاسبه نمود. جدول ۸.۳ برآوردهای دیگر ویژگی‌ها را نشان می‌دهد.

وقتی برای بدست آوردن مقدار بخش‌ارزشی، ویژگی شکل محصول را در نظر داریم که پاسخ‌دهنده اول کنسانتره را به مخلوط و پودر بودن محصول ترجیح می‌دهد. وقتی مقدار بخش‌ارزشی یک سطح از ویژگی بیشتر باشد به این معنا است که پاسخ‌دهنده آن سطح را به سطحی که مقدار بخش‌ارزشی آن کمتر است ترجیح می‌دهد. با استفاده از برآوردهای جدول ۸.۳ و مدل زیر می‌توان نمرات پاسخ ویژگی‌ها برای پاسخ‌دهنده اول را پیش‌گویی کرد.

$$\hat{\mathbf{Y}} = \mathbf{X}\hat\beta \tag{۵.۳}$$

جدول ۸.۳: مجموعه کامل مقادیر برآورد بدست آمده بخش‌ارزشی برای نفر اول در مثال ماده شوینده طبق درستی رابطه ۴.۳

$\hat\beta_0$	$\hat\beta_{F_1}$	$\hat\beta_{F_2}$	$\hat\beta_{F_3}$	$\hat\beta_{A_1}$	$\hat\beta_{A_2}$	$\hat\beta_{A_3}$
۴/۱۱	−۰/۰۶	۰/۶۱	−۰/۵۵	۰/۴۴	۰/۶۱	−۱/۰۵
$\hat\beta_{D_1}$	$\hat\beta_{D_2}$	$\hat\beta_{B_1}$	$\hat\beta_{B_2}$	$\hat\beta_{P_1}$	$\hat\beta_{P_2}$	$\hat\beta_{P_3}$
−۰/۲۱	۰/۲۱	۰/۵۴	−۰/۵۴	۱/۴۴	۰/۹۴	−۲/۳۸

که در آن $\hat{\boldsymbol\beta} = (\hat\beta_0, \hat\beta_{F_1}, \cdots, \hat\beta_{P_3})$.

پیرو نرخ پیش‌گویی که بوسیله ارزیابی پاسخ‌دهنده اول برای ۱۸ مشخصه بدست آمد بیشترین امتیاز پیش‌بینی، اولویت اول پاسخ‌دهنده نسبت به آن مشخصه را نشان می‌دهد.

جدول ۹.۳: پیش‌گویی پاسخ‌ها برای نفر اول که با استفاده از رابطه (۵.۳) در مثال ماده شوینده

مشخصه	۱	۲	۳	۴	۵	۶	۷	۸	۹
Y	۶	۳	۵	۲	۳	۱	۱	۶	۶
$\hat{Y}$	۵/۴۴	۴/۲۸	۴/۸۶	۲/۶۹	۱/۹۴	۰/۹۴	۲/۶۱	۴/۲۸	۵/۸۸
مشخصه	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸
Y	۶	۷	۴	۱	۶	۶	۶	۶	۱
$\hat{Y}$	۶/۴۴	۶/۳۶	۳/۲۸	۱/۳۶	۶/۲۸	۶/۰۳	۵/۶۱	۶/۸۶	۰/۸۶

درستی مدل را می‌توان به‌وسیله ارزیابی دوباره پیش‌گویی ویژگی‌ها با نرخ واقعی که پاسخ‌دهنده از مشخصه‌ها می‌دهد، تعیین کرد. یک راه متداول برای ارزیابی مدل، محاسبه ضریب تعیین است که با R^2 نشان داده می‌شود. هر چه مقدار آن به یک نزدیکتر باشد مدل مناسب‌تر است. ضریب تعیین به‌صورت زیر محاسبه می‌شود.

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SSE}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

وقتی که پیش‌گویی امتیازات پاسخ‌دهنده نسبتا خوب باشد مثلا $R^2 = ۰/۸۶۳$ (در مثال فوق) محقق قادر است نتیجه بگیرد مدلی که انتخاب کرده برای پیش‌گویی امتیاز پاسخ‌دهنده مناسب است. برای تعیین اندازه یک تحلیل متقارن، توسط محقق بر روی پاسخ افراد، این اهمیت دارد که محقق طرحی را مطالعه کند که پاسخ‌دهنده ندانسته ویژگی به مدل اضافه نکند. چون با اضافه کردن یک ویژگی به مدل مقادیری بدست می‌آید که تاثیرش را روی نتیجه‌گیری و اولویت‌بندی کلی نشان می‌دهد.

حال برای رتبه‌بندی ویژگی‌های، مقادیر برآورد شده ضرایب هر ویژگی را در نظر می‌گیریم ماکزیمم و مینیمم سطوح ویژگی را تعیین کرده و فاصله بیشترین و کمترین سطح برای هر ویژگی را بدست می‌آوریم بعد مجموع این فاصله را حساب می‌کنیم. برای این که نشان دهیم که هر ویژگی چه نسبتی بر اولویت‌بندی تاثیر دارد، سهم هر قسمت را نسبت به کل مقدار بدست می‌آوریم که مشاهده شود هر ویژگی چه درصدی در اولویت‌بندی سهم دارد.برای پاسخ‌دهنده اول اندازه اهمیت‌ها در جدول ۱۰.۳ آمده است.

جدول ۱۰.۳: اندازه‌های اهمیت‌ها برای پاسخ‌دهنده اول در مثال ماده شوینده

نسبت(درصد)	رتبه	ماکزیمم	مینیمم	ویژگی
۱۴/۲۴٪	۱/۱۶	۰/۶۱	−۰/۵۵	شکل محصول
۲۰/۴۰٪	۱/۶۶	۰/۶۱	−۱/۰۵	تعداد کاربرد
۵/۱۵٪	۰/۴۲	۰/۲۱	−۰/۲۱	ضدعفونی‌کننده
۱۳/۲۷٪	۱/۰۸	۰/۵۴	−۰/۵۴	زیست تخریب‌پذیر
۴۶/۹۳٪	۳/۸۲	۱/۴۴	−۲/۳۸	قیمت
۱۰۰٪	۸/۱۴			کل

با توجه به نتایج به‌دست آمده در جدول ۱۰.۳ مشاهده می‌شود بیشترین اهمیت برای اولویت‌بندی از نظر مصرف‌کننده را قیمت دارد که درصد اهمیت آن ۴۶/۹۳٪ بوده و دومین ویژگی که بیشترین تاثیر را در انتخاب محصول دارد کاربرد آن است. این نشان می‌دهد بیشترین نگرانی مردم هزینه بوده و بعد از هزینه‌ای که انجام داده‌اند کاربرد محصول اهمیت بسزایی دارد. این اولویت‌بندی را برای ۸۵ پاسخ‌دهنده دیگر نیز می‌توان انجام داد. اما از آنجایی که می‌خواهیم نتایج ۸۵ پاسخ‌دهنده را همزمان ارزیابی کنیم باید از مجموع مدل‌ها استفاده کنیم.

ترکیب مدل‌های جزئی

ساده‌ترین راه برای نشان دادن مقادیر مدل بخش‌ارزشی مطالعه تحلیل متقارن است. یکی از مهم‌ترین راه مطالعه تحلیل متقارن برآورد تکی بخش‌ارزشی است که از آن برای طبقه‌بندی گروه‌های پاسخ‌دهنده و بخش‌های مختلف بازار استفاده می‌شود. میانگین‌ها و واریانس‌ها از پارامترهای برآوردکننده در جدول ۱۱.۳ میانگین و واریانس همه پارامترها از ۸۶ پاسخ‌دهنده می‌باشد. برای هر یک از افراد پاسخ‌دهنده، یک مدل برازش شده و برآوردهای ۹گانه به‌دست آمده سرانجام برای ترکیب کردن نتایج می‌توان از ۸۶ برآورد به‌دست آمده برای هر پارامتر، میانگین و واریانس آن‌ها را محاسبه کرد.

جدول ۱۱.۳: برآورد کلی میانگین و واریانس مقادیر ثابت بخش‌ارزشی در مثال ماده شوینده

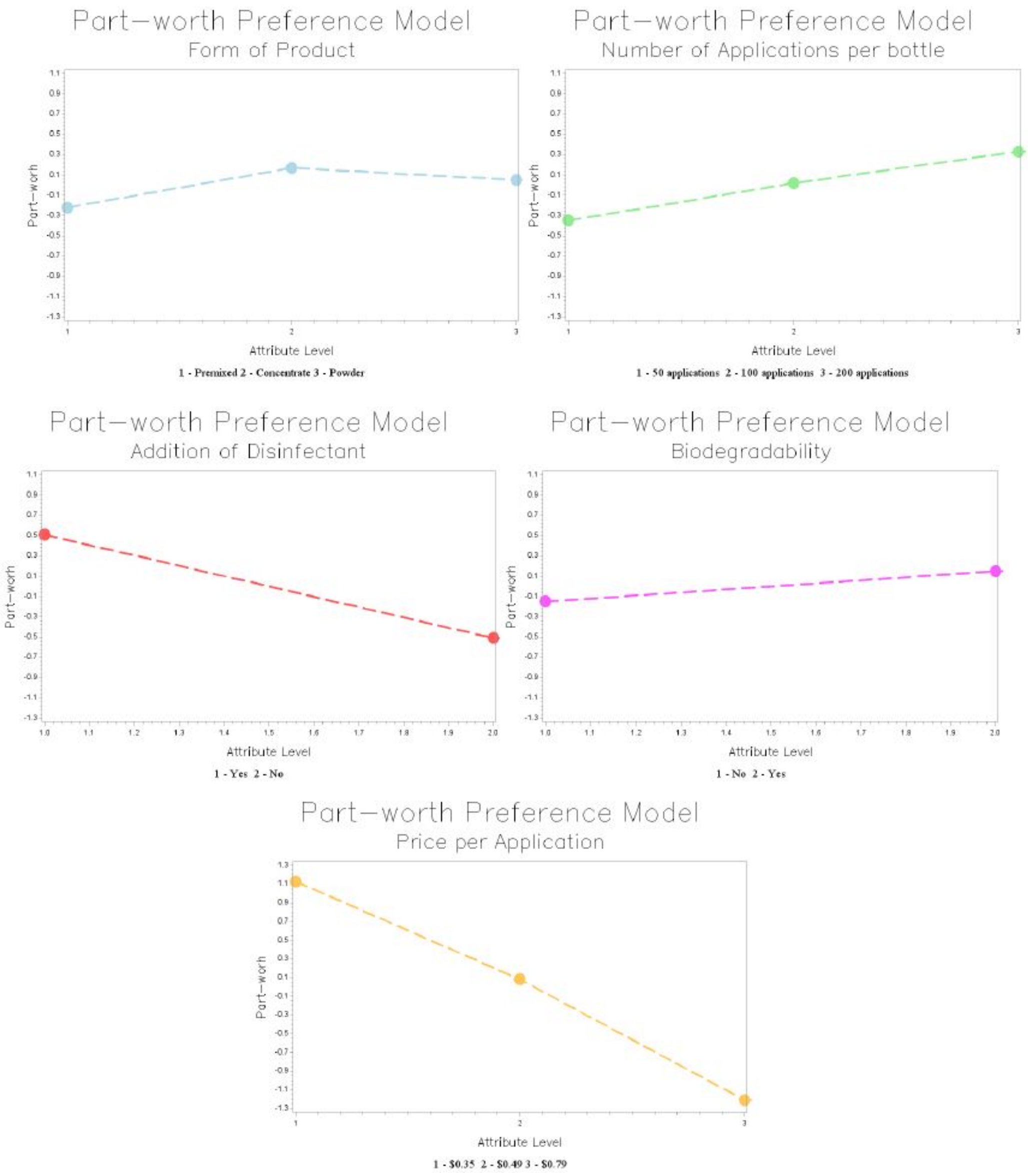
سطح	برآورد میانگین β_s	برآورد واریانس
باقی‌مانده		۲/۶۲
ضریب ثابت	۳/۷۴	۰/۶۳
مخلوط	−۰/۲۲	۰/۲۳
کنسانتره	۰/۱۷	۰/۱۵
۵۰ کاربرد	−۰/۳۵	۰/۳۲
۱۰۰ کاربرد	۰/۰۲	۰/۱۹
ضدعفونی‌کننده	۰/۵۱	۰/۳۸
زیست تخریب‌پذیر	−۰/۱۵	۰/۱۷
قیمت ۳۵۰۰ تومان	۱/۱۳	۰/۶۳
قیمت ۴۹۰۰ تومان	۰/۰۸	۰/۲۵

جدول ۱۱.۳ شامل قسمتی از برآورد سطوح بخش‌ارزشی از ماتریس ساختگی X می‌باشد. به این ترتیب برای تعیین برآورد موارد مصرفی بخش‌ارزشی تمام سطوح ویژگی، که دارای ویژگی یکسانی از نظر تحلیل رگرسیونی متغیرهای X دارند از رابطه (۷.۳) استفاده می‌شود که مقادیر برآورد شده برای تمام سطوح را در جدول ۱۲.۳ می‌توان مشاهده کرد. همچنین شکل ۱.۳ قسمت ارزشی موارد مصرف‌شده برای هر سطح از ویژگی را نشان می‌دهد.

جدول ۱۲.۳: مقادیر بخش‌ارزشی برای همه سطوح ویژگی در مثال ماده شوینده

سطح ویژگی	موارد مصرفی بخش‌ارزشی $\hat{\beta}_s$
مخلوط	−۰/۲۲
کنسانتره	۰/۱۷
پودر	۰/۰۵
۵۰ کاربرد	−۰/۳۵
۱۰۰ کاربرد	۰/۰۲
۲۰۰ کاربرد	۰/۳۳
ضدعفونی کننده (بله)	۰/۵۱
ضدعفونی کننده (نه)	−۰/۵۱
زیست‌تخریب‌پذیر(بله)	−۰/۱۵
زیست تخریب‌پدیر(نه)	۰/۱۵
۳۵۰۰ تومان	۱/۱۳
۴۹۰۰ تومان	۰/۰۸
۷۹۰۰ تومان	−۱/۲۱

شکل ۱.۳ پیش‌گویی سطوح موارد مصرف شده برای پاسخ‌دهنده‌ها را نشان می‌دهد.



شکل ۱.۳: موارد مصرف شده بخش‌ارزشی در مثال ماده شوینده

با توجه به شکل ۱.۳ برای هر سطح از هر ویژگی می‌توانیم به خوبی ببینیم که چه اتفاقی می‌افتد. خطوط آن پیوستگی ندارد اما اولویت‌ها را به خوبی نشان می‌دهد و این می‌تواند یک روش خوب برای جستجوگران باشد چون سریع می‌توانند ویژگی‌ها و سطوحی که اولویت بیشتر دارند برای مصرف‌کننده را تعیین کنند. وقتی ویژگی قیمت مطالعه می‌شود باید دقت داشت که برای برآورد مقادیر بخش‌ارزشی سطوح آن از نمودار استفاده شود چون نمودار آن به صورت خطی است. بوسیله مشاهده شکل ۱.۳ سطوح ویژگی‌ها به‌وسیله پاسخ‌های افراد برتری داده برای هر ویژگی را در جدول ۱۳.۳ می‌بینید.

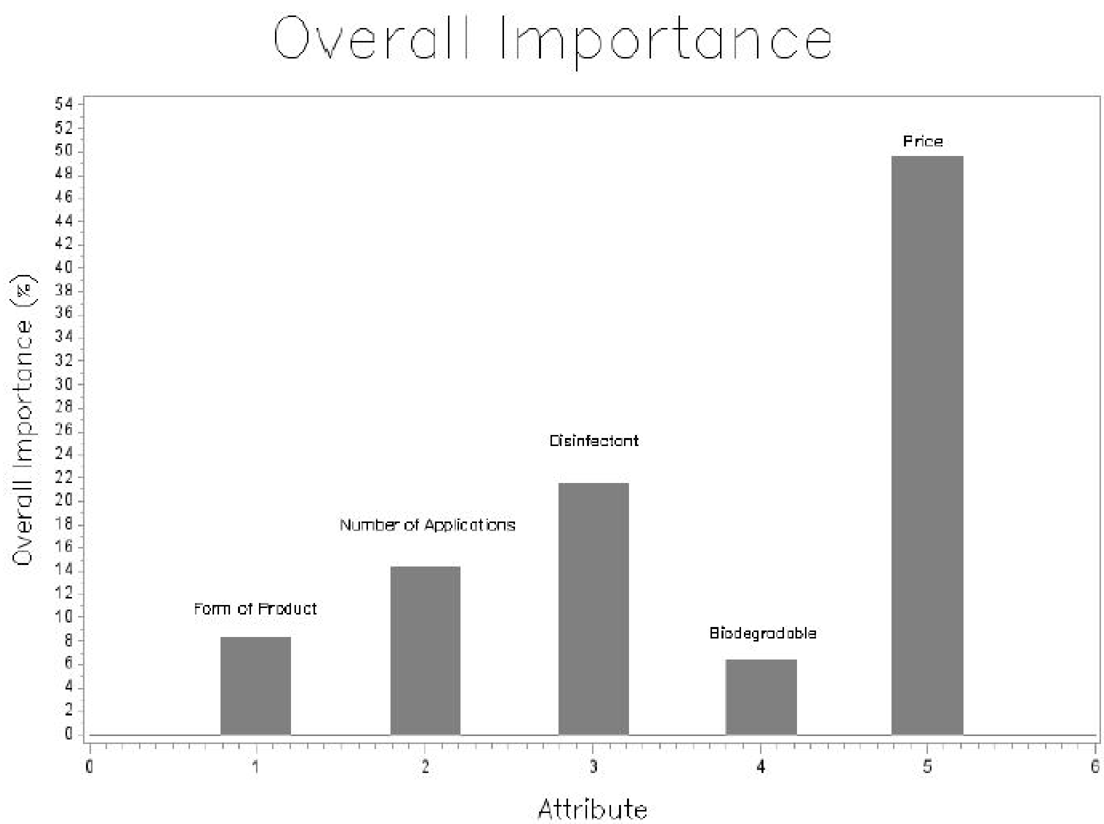
جدول ۱۳.۳: سطوح ویژگی که بیشترین اولویت را دارند در مثال ماده شوینده

کنسانتره
۲۰۰کاربرد
ضدعفونی کننده
تخریب‌پذیر زیستی
۳۵۰۰ تومان

اگرچه این موارد مصرف‌شده بخش‌ارزشی نشان‌دهنده اهمیت زیاد اولویت پاسخ‌ها است، ضروری است اهمیت کلی اولویت گروه‌ها برای ویژگی‌های مختلف ارزیابی شود. اگر موارد مصرف‌شده بخش‌ارزشی ویژگی تخریب‌پذیرزیستی را در نظر بگیریم و با دیگر ویژگی‌ها مقایسه کنیم آنگاه دیده می‌شود که این ویژگی نسبت به برخی از ویژگی‌ها از اهمیت خوبی برخوردار نیست. این بدان معنی است اگرچه زیر مجموعه بزرگی از پاسخ‌دهندگان تخریب‌پذیرزیستی را ترجیح می‌دهند، اما تجزیه زیستی محصول تاثیر زیادی برای در اولویت قرار گرفتن برای پاسخ‌دهندگان ندارد. اهمیت کلی ویژگی‌ها برای مجموع مدل مانند پاسخ‌دهنده اول مورد مطالعه و بررسی قرار می‌گیرد. میزان اهمیت هر یک از ویژگی‌ها را در جدول ۱۴.۳ و شکل ۲.۳ مشاهده می‌کنید

جدول ۱۴.۳: اندازه‌های مهم برای پاسخ‌دهنده اول در مثال ماده شوینده

ویژگی	مینیمم	ماکسیمم	رتبه	نسبت(درصد)
شکل محصول	−۰/۲۲	۰/۱۷	۰/۳۹	۸/۲۵٪
تعدا کاربرد	−۰/۳۵	۰/۳۳	۰/۶۸	۱۴/۳۸٪
ضدعفونی‌کننده	−۰/۵۱	۰/۵۱	۱/۰۲	۲۱/۵۶٪
زیست تخریب‌پذیر	−۰/۱۵	۰/۱۵	۰/۳	۶/۳۴٪
قیمت	−۱/۲۱	۱/۱۳	۲/۳۴	۴۹/۴۷٪
کل			۴/۷۳	۱۰۰٪



شکل ۲.۳: مقادیر کلی ویژگی‌های ماده شوینده (در این شکل محور افقی ویژگی و محور عمودی اهمیت کلی است)

با توجه به اهمیت کلی ویژگی‌ها، (نتایج جدول ۱۴.۳ و شکل ۲.۳)کاملا مشخص است که قیمت تاثیر بیشتری بر روی انتخاب افراد دارد. این بدان معناست که اگر قیمت یک محصول افزایش پیدا کند با توجه به شکل ۱.۳ احتمال این‌که مشتریان آن محصول را انتخاب کنند به شدت کاهش پیدا می‌کند. در بسیاری از مطالعات محققان سعی می‌کنند که ویژگی

قیمت را کنار بگذارند چون به شدت روی انتخاب تاثیر دارد و با این کار می‌خواهند برآورد بهتری برای دیگر ویژگی‌ها داشته باشند. دومین ویژگی که اهمیت بیشتری دارد، داشتن یا نداشتن ضدعفونی‌کننده است. واضح است که مصرف‌کنندگان ترجیح می‌دهند که محصول شامل ضدعفونی‌کننده باشد. این احساس وجود دارد که مصرف‌کنندگان تمیزکننده‌ای که ضدعفونی‌کننده ندارد را خریداری نمی‌کنند و اولویت محصولی که شامل ضدعفونی‌کننده نیست کاهش پیدا می‌کند. زمانی که ارزش کلی هر ویژگی مشخص شد آن‌گاه محققان می‌توانند سطوح هر ویژگی را مورد بررسی قرار دهند. در هر محصول سطوح ویژگی که از اهمیت بیشتری برخوردار است ضدعفونی‌کننده بودن و قیمت کم است که این دو سطح با هر سطح از ویژگی‌های دیگر باشد مشکلی ایجاد نمی‌کند. در محصولات تخریب‌پذیرزیستی از اهمیت زیادی برخوردار نیست هر چند تخریب‌پذیر زیستی تاثیر زیادی بر روی محیط دارد. در نهایت اکثر پاسخ‌دهندگان شکل کنسانتره بودن را ترجیح می‌دهند. مقادیر مصرف‌شده بخش‌ارزشی کلی نشان داده شده در جدول ۱۵.۳ برای پیش‌گویی موارد مصرف‌شده از ۱۸ مشخصه برای همه پاسخ‌دهندگان، استفاده می‌شود. معادله پیشگو آن همان رابطه (۵.۳) است. که معادل زیر می‌باشد

$$\hat{y}_j = \hat{\beta}_1 x_{j1} + \hat{\beta}_{F_1} x_{j2} + \hat{\beta}_{F_2} x_{j3} + \hat{\beta}_{A_1} x_{j4} + \hat{\beta}_{A_2} x_{j5} + \hat{\beta}_{D_1} x_{j6} + \hat{\beta}_{B_1} x_{j7} + \hat{\beta}_{P_1} x_{j8} + \hat{\beta}_{P_2} x_{j9}$$

که در آن $j = 0, 1, 2, \dots, 18$ مقادیر پیش‌گویی کننده برای ۱۸ مشخصه است.

جدول ۱۵.۳: پیش‌گویی موارد مصرف‌شده برای ۱۸ مشخصه در مثال ماده شوینده

مشخصه	۱	۲	۳	۴	۵	۶	۷	۸	۹
$\hat{Y}$	۶	۶	۴	۵	۳	۳	۳	۴	۳
مشخصه	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸
$\hat{Y}$	۴	۴	۳	۲	۵	۶	۴	۳	۳

سه پروفایل پیش‌بینی شده که بیشترین امتیاز را بدست آوردند (یعنی محصولاتی که می‌توانند بیشترین فروش را تضمین کنند) در جدول ۱۶.۳ آورده شده است.

جدول ۱۶.۳: مشخصه‌هایی از مثال ماده شوینده که در پیش‌گویی بیشترین اولویت را داشتند

مشخصه	شکل محصول	کاربردها	ضدعفونی‌کننده	زیست‌تخریب‌پذیر	قیمت
۱	کنسانتره	۲۰۰	بله	بله	۳۵۰۰ تومان
۲	پودر	۲۰۰	بله	بله	۳۵۰۰ تومان
۱۵	کنسانتره	۱۰۰	بله	نه	۳۵۰۰ تومان

دوباره مشاهده شد که بیشترین امتیاز مشخصه مربوط به محصولی است که دارای قیمت کم است دوتا از سه محصول دارای بیشترین کاربرد هستند. همیشه یک تولیدکننده نمی‌تواند یک محصول را با بهترین کیفیت و هزینه کم تولید کند. یک تولیدکننده باید بتواند یک تعادل کامل بین بهترین محصول و بهترین قیمت ایجاد کند. این مساله به عنوان سردرد مدیران معرفی شده است. با رقابت بین شرکت‌ها و دسترسی آسان به محصولات مختلف شرکت‌ها باید محصولات خود را در یک سطح بالاتری معرفی کنند. تفکر مطالعه تحلیل متقارن می‌تواند به شرکت‌ها و تولیدکنندگان برای تشخیص این که کدام ویژگی به طور کلی بر مصرف‌کننده تاثیر دارد کمک کند. این نشان‌دهنده این است که با انجام تحلیل متقارن می‌توان یک محصول را تغییر یا بهبود بخشید و یا این‌که یک محصول جدید تولید کرد.

روش دیگری که احتمالا می‌تواند به درک بهتر اولویت پاسخ‌دهندگان نسبت به انواع پاک‌کننده صنعتی کمک کند، این است که از یک مؤلفه تصادفی در مدل اثرات ثابت استفاده کنیم. در فصل بعد مدل اثرات تصادفی را معرفی کرده و روش تحلیل متقارن معمولی را بوسیله این مدل بهبود می‌دهیم. این مؤلفه‌های تصادفی به مدل اثرات ثابت اضافه می‌شود تا مدل اثرات آمیخته خطی ساخته شود.

۳.۳.۳ تحلیل متقارن در R

در این بخش مثال دیگری از مدل متقارن با اثرات ثابت را با استفاده از نرم‌افزار R مورد بررسی قرار می‌دهیم.

علاقه‌مندیم اولویت مصرف کنندگان چای از برند یک نوع خاص در سال ۲۰۰۴ در یک گروه از دانشجویان اقتصاد دانشگاه ورودزلاو لهستان را تعیین کنیم. برای این منظور ۱۰۰ نفر از دانشجویان انتخاب شدند و پرسشنامه‌ای در خصوص امتیاز چندین مشخصه را پر کردند. چهار ویژگی از چای در دسترس است که سه ویژگی دارای سه سطح و یک ویژگی دارای دو سطح می‌باشد که در جدول ۱۷.۳ مشاهده می‌کنید (آندرز بک و توماس بارتلوموویچ، ۲۰۱۴).

جدول ۱۷.۳: ویژگی‌ها و سطوح مربوط به داده‌های چای

ویژگی	سطوح
قیمت	۱.کم ۲. متوسط ۳. زیاد
تنوع	۱. مشکی ۲. سبز ۳. قرمز
نوع	۱.کیسه ۲. گرانوله ۳. برگ
عطر	۱. دارد ۲. ندارد

در مجموع $۵۴ = ۲ \times ۳ \times ۳ \times ۳$ مشخصه خواهیم داشت، که در این مطالعه تنها ۱۳ مشخصه مهم آن را مورد بررسی قرار دهیم.

در جدول ۱۸.۳، ۱۳ مشخصه مهم و ۵ پاسخ نمونه مربوط به ارزش‌گذاری ۵ نفر آورده شده است.

جدول ۱۸.۳: مقادیر ۴ ویژگی، ۱۳ مشخصه و ارزش‌گذاری ۵ دانشجو مربوط به داده‌های چای

منبع					ویژگی‌ها				مشخصه
S_5	S_4	S_3	S_2	S_1	عطر	نوع	تنوع	قیمت	
۵	۶	۴	۰	۸	بله	کیسه	مشکی	زیاد	۱
۱	۷	۱۰	۱۰	۱	بله	کیسه	سبز	کم	۲
۷	۴	۳	۳	۱	بله	گرانوله	سبز	متوسط	۳
۸	۹	۵	۵	۳	بله	برگ	مشکی	متوسط	۴
۶	۶	۴	۱	۹	بله	برگ	قرمز	زیاد	۵
۱۰	۳	۱	۴	۲	نه	کیسه	مشکی	متوسط	۶
۷	۷	۲	۸	۷	نه	کیسه	سبز	زیاد	۷
۱۰	۴	۰	۶	۲	نه	کیسه	قرمز	متوسط	۸
۶	۸	۰	۲	۲	نه	گرانوله	مشکی	زیاد	۹
۶	۵	۱	۹	۲	نه	گرانوله	قرمز	کم	۱۰
۶	۲	۸	۷	۲	نه	برگ	مشکی	کم	۱۱
۱۰	۱۰	۹	۵	۳	نه	برگ	سبز	متوسط	۱۲
۷	۹	۷	۲	۴	نه	برگ	سبز	زیاد	۱۳

برای تعیین مدل رگرسیون ارزشی ابتدا سطوح کیفی ۴ ویژگی مورد بررسی را با کد گذاری کمی کنیم. برای مثال چون ویژگی قیمت سه سطح دارد، دوسطح آن برای طراحی ماتریس x استفاده می‌شود جدول ۱۹.۳ مراحل کدگذاری را نشان می‌دهد.

جدول ۱۹.۳: جدول برای ساختن متغیرهای مصنوعی x در مثال چای

قیمت	$X_1 = \begin{cases} 1 & \text{کم} \\ 0 & \text{متوسط} \\ -1 & \text{زیاد} \end{cases}$	$X_2 = \begin{cases} 1 & \text{متوسط} \\ 0 & \text{کم} \\ -1 & \text{زیاد} \end{cases}$
تنوع	$X_3 = \begin{cases} 1 & \text{مشکی} \\ 0 & \text{سبز} \\ -1 & \text{قرمز} \end{cases}$	$X_4 = \begin{cases} 1 & \text{سبز} \\ 0 & \text{مشکی} \\ -1 & \text{قرمز} \end{cases}$
نوع	$X_5 = \begin{cases} 1 & \text{کیسه} \\ 0 & \text{گرانوله} \\ -1 & \text{برگ} \end{cases}$	$X_6 = \begin{cases} 1 & \text{گرانوله} \\ 0 & \text{کیسه} \\ -1 & \text{برگ} \end{cases}$
عطر	$X_7 = \begin{cases} 1 & \text{بله} \\ -1 & \text{خیر} \end{cases}$	

با توجه به مثال فوق متغیرهای کمکی کدگذاری شده در جدول ۲۰.۳ مشخص شده‌اند.

جدول ۲۰.۳: کد گذاری ۴ ویژگی داده‌های چای برای ۱۳ مشخصه

مشخصه	X_1	X_2	X_3	X_4	X_5	X_6	X_7
۱	−۱	−۱	۱	۰	۱	۰	۱
۲	۱	۰	۰	۱	۱	۰	۱
۳	۰	۱	۰	۱	۰	۱	۱
۴	۰	۱	۱	۰	−۱	−۱	۱
۵	−۱	−۱	−۱	−۱	−۱	−۱	۱
۶	۰	۱	۱	۰	۱	۰	۱
۷	−۱	−۱	۰	۱	۱	۰	−۱
۸	۰	۱	−۱	−۱	۱	۰	−۱
۹	−۱	−۱	۱	۰	۰	۱	−۱
۱۰	۱	۰	−۱	−۱	۰	۱	−۱
۱۱	۱	۰	۱	۰	−۱	−۱	−۱
۱۲	۰	۱	۰	۱	−۱	−۱	−۱
۱۳	−۱	−۱	۰	۱	−۱	−۱	−۱

برای انجام تحلیل متقارن از مدل رگرسیونی زیر استفاده می‌شود

$$U_s = b_{0s} + b_{1s}X_{1s} + b_{2s}X_{2s} + \cdots + b_{7s}X_{7s} \qquad s = 1, 2, \cdots 100$$

که در آن U_s امتیاز نفر sام و X_{is} متغیر کمکی iام برای نفر sام و b_{is} ضریب رگرسیونی متغیر X_{is} است. در این مدل عرض از مبدا با b_{0s} نشان داده شده است. برآورد ضرایب مدل رگرسیونی ساده کیفی با استفاده از بسته Conjoint در نرم‌افزار R مربوط به نفر اول به‌دست آورده شده است که در جدول ۲۱.۳ می‌بینید

جدول ۲۱.۳: برآورد ضرایب مدل داده‌های چای برای نفر اول

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
برآوردگر	۳/۳۹۳	−۱/۵۱۷۲	−۱/۱۴۱۴	۰/۴۷۴۷	−۰/۶۷۴۷	۰/۶۵۸۶	−۱/۵۱۷۲	۰/۶۲۹۳

باتوجه به جدول ۲۱.۳، مقادیر به‌دست آمده از بخش‌ـارزشی برای پاسخ‌دهنده اول نشان می‌دهد که او مایل است به‌صورت کم یا زیاد از همه سطوح ویژگی استفاده کند. در ادامه طبق (۴.۳) تمام ویژگی‌ها به‌دست می‌آید که در جدول ۲۲.۳ مشاهده می‌کنید.

جدول ۲۲.۳: مجموعه کامل مقادیر برآورد بدست آمده بخش–ارزشی برای نفر اول در مثال چای

$b_۰$	$b_۱$	$b_۲$	$b_{۲_۱}$	$b_۳$	$b_۴$
۳/۳۹	−۱/۵۱۷۲	−۱/۱۴۱۴	۲/۶۵۶۸	−۰/۴۷۴۷	۰/۶۷۴۷
$b_{۴_۱}$	$b_۵$	$b_۶$	$b_{۶_۱}$	$b_۷$	$b_{۷_۱}$
۱/۲۴۹۴	۰/۶۵۸۶	−۱/۵۱۷۲	۰/۸۵۸۶	۰/۶۲۹۳	−۰/۶۲۹۳

با استفاده از برآوردهای جدول ۲۲.۳ و مدل زیر می‌توان پاسخ پاسخ‌دهنده اول را برای ویژگی‌ها، پیش‌گویی کرد. پیش‌گویی پاسخ‌دهنده اول را در جدول ۲۳.۳ می‌بینید.

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

جدول ۲۳.۳: پیش‌گویی پاسخ‌ها برای نفر اول در مثال چای

مشخصه	۱	۲	۳	۴	۵	۶	۷
Y	۸	۱	۲	۳	۹	۲	۷
$\hat{Y}$	۶/۸۶	۲/۹۴	۰/۶۹	۳/۲۶	۸/۶۹	۱/۸	۵/۴
مشخصه	۸	۹	۱۰	۱۱	۱۲	۱۳	
Y	۲	۲	۲	۲	۳	۴	
$\hat{Y}$	۳/۴۳	۳/۴۳	۰/۸۷	۱/۶۳	۱/۸	۵/۶	

یک راه متداول برای ارزیابی مدل، استفاده از ضریب تعیین است که هرچه به یک نزدیک‌تر باشد مدل مناسب‌تر است در اینجا ضریب تعیین برابر $R^۲ = ۰/۸۱۸$ است که نشان می‌دهد مدل مناسب می‌باشد.

در ادامه برای همه افراد (۱۰۰ نفر) تحلیل مانند تحلیل فوق انجام داده و اولویت‌های افراد را مشخص می‌کنیم. همان‌طور که در جدول ۲۴.۳ می‌بینید مقادیر بخش–ارزشی برای همه سطوح ویژگی با استفاده از مدل رگرسیون چندگانه کیفی به‌دست آمده است.

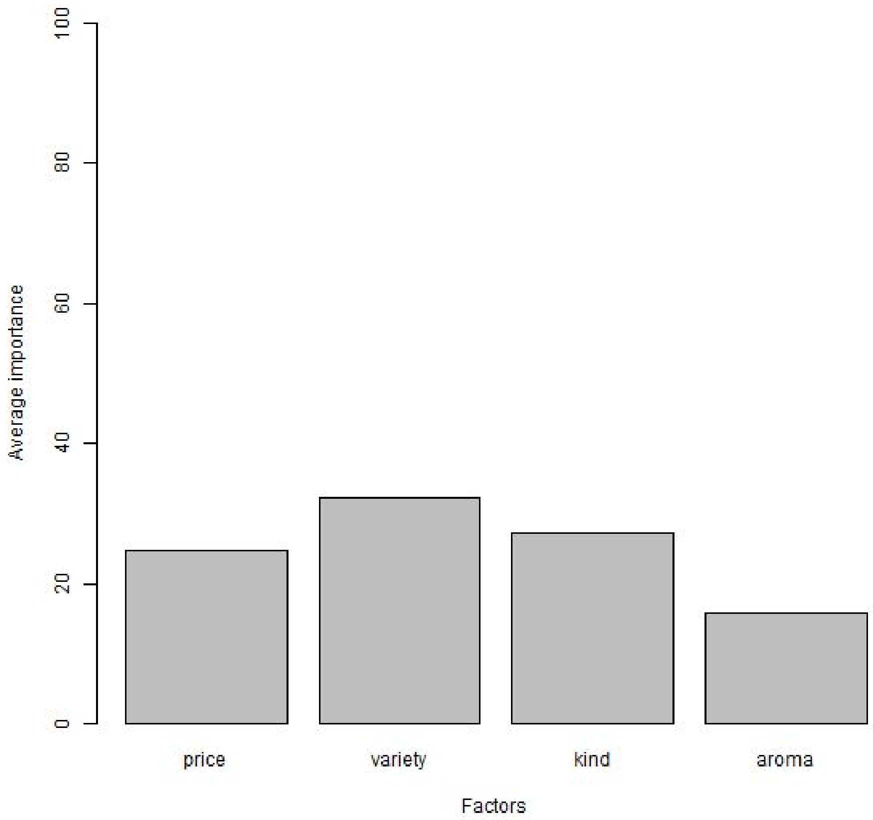
جدول ۲۴.۳: مقادیر بخش‌ـارزشی برای تمام سطوح ویژگی در مثال چای

ردیف	سطوح	برآوردها
۱	عرض‌ازمبدا	۳/۵۵۳۴
۲	کم	−۰/۲۴۰۲
۳	متوسط	۰/۱۴۳۱
۴	زیاد	−۰/۰۹۷۱
۵	مشکی	۰/۶۱۴۹
۶	سبز	۰/۳۴۹
۷	قرمز	−۰/۶۴۹۸
۸	کیسه	۰/۱۳۶۹
۹	گرانوله	۰/۸۸۹۸
۱۰	برگ	۰/۷۵۲۹
۱۱	بله	۰/۴۱۰۸
۱۲	نه	−۱۰/۴۱۰۸

حال با توجه به مقادیر به‌دست آمده بخش‌ـارزشی و مطالعه ویژگی‌ها، می‌توان گفت که در هر یک از ویژگی کدام سطح بیشترین مجبوبیت را دارد. مثلا برای قیمت (متوسط)، تنوع (مشکی)، نوع (برگ) و عطری بودن چای بیشترین محبوبیت را دارند. تا اینجا نشان دادیم که در هر ویژگی کدام سطح بیشترین محبوبیت را دارد، حال برای نشان دادن این‌که کدام ویژگی‌ها بیشترین تاثیر را بر خرید مردم دارد، مقادیر برآورد شده ضرایب هر ویژگی را در نظر می‌گیریم ماکزیمم و مینیمم سطوح ویژگی را تعیین کرده و فاصله بیشترین و کمترین سطح برای هر ویژگی را به‌دست می‌آوریم بعد مجموع این فاصله‌ها را حساب می‌کنیم و سهم هر قسمت را نسبت به کل مقدار بدست می‌آوریم که مشاهده شود هر ویژگی چه درصدی در اولویت بندی سهم دارد. در جدول ۲۵.۳ میزان سهم هر ویژگی را مشاهده می‌کنید.

جدول ۲۵.۳: اندازه‌های مهم برای پاسخ‌دهنده‌ها در مثال چای

نسبت (درصد)	رتبه	ماکزیمم	مینیمم	ویژگی
۲۴/۷۶	۱/۶۳	۰/۷۵	−۰/۸۸	نوع
۳۲/۲۲	۱/۲۵	۰/۶۱	−۰/۶۴	تنوع
۲۷/۱۵	۰/۳۸	۰/۲۴	−۰/۱۴	قیمت
۱۵/۸۸	۰/۸۸	۰/۴۱	−۰/۴۱	عطر
٪۱۰۰	۴/۱۴	کل		



شکل ۳.۳: مقادیر کلی ویژگی‌های مثال چای. در این شکل محور افقی ویژگی و محور عمودی اهمیت کلی را نشان می‌دهد

با توجه به اهمیت کلی ویژگی‌ها (شکل ۳.۳ و جدول ۲۵.۳) کاملا مشخص است که تنوع محصول بر روی انتخاب مشتریان تاثیر بیشتر نسبت به بقیه ویژگی‌ها دارد. دومین ویژگی که نسبت به بقیه تاثیر بیشتری دارد قیمت محصول است. پس با این وضعیت می‌توان نتیجه گرفت که با تغییر تنوع محصول میزان فروش آن کمتر شود.

فصل ۴

مدل اثرات آمیخته

۱.۴ مقدمه

بار دیگر مدل $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ را با اثرات ثابت در نظر بگیرید. بدیهی است در این مدل ساختار کواریانس $\mathbf{Y}$ تنها به مؤلفه خطای مدل، ε، وابسته است. اما چنانچه همبستگی در $\mathbf{Y}$ (ساختار کواریانس) ناشی از یکسری متغیرهای توضیحی خاص باشد که می‌بایستی در مدل لحاظ می‌شدند ولی در مدل‌بندی وارد نشده‌اند. دیگر نمی‌توان از این مدل استفاده کرد. در این حالت برای فهم و درک ساختار کواریانس از مدل اثرات آمیخته استفاده می‌شود. در این فصل ابتدا مدل اثرات تصادفی را معرفی کرده و در ادامه، به مدل اثرات آمیخته پرداخته می‌شود. سپس با استفاده از این تکنیک برای مثال ماده شوینده در بخش ۲.۳.۳ یک تحلیل متقارن انجام می‌گیرد.

۲.۴ مقدمه‌ای بر مؤلفه‌های تصادفی

توانایی مدل رگرسیون تصادفی بسیار زیاد است و به محقق اجازه می‌دهد تا داده‌هایی که دارای مقادیر گمشده[۱]، داده‌هایی که وابسته به زمان[۲] یا متغیر ثابت [۳]و داده‌هایی که از جنس اندازه تصادفی متفاوت هستند، را در مدل‌بندی لحاظ کند.

۱.۲.۴ مدل اثرات تصادفی یکطرفه

متغیر وابسته y_{ij} را در نظر بگیرید که در آن $i = ۱,۲,\cdots,k$ و $j = ۱,۲,\cdots,n_i$ بوده و k تعداد گروه‌ها و n_i اندازه نمونه برای گروه iام را نشان می‌دهد. مدل اثرات ثابت یکطرفه با فرض نرمال بودن به‌صورت زیر بیان می‌شود

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

که در آن

$$\varepsilon_{ij} \sim \mathcal{N}(\circ, \sigma^۲), \qquad \tau_i \sim \mathcal{N}(\circ, \sigma^۲_\tau)$$

که ε_{ij}ها دوبه‌دو مستقل، τ_iها دوبه‌دو مستقل و τ_i و ε_{ij}ها هم دوبه‌دو مستقل‌اند، $\sigma^۲$ واریانس خطا و $\sigma^۲_\tau$ واریانس مربوط به قسمت تصادفی است همچنین $\sum_{i=۱}^{k} n_i = N$ تعداد کل مشاهدات است. زمانی‌که اندازه نمونه‌ها با هم برابر باشند یعنی $n_i = n$، مدل متعادل است. پارامترهای مورد نیاز برای برآورد این مدل برآوردهای تصادفی τ_i (که برای هر گروه متفاوت هست) و پارامترهای ثابت μ، $\sigma^۲$ و $\sigma^۲_\tau$ هستند.

با توجه به فرضیات فوق نتیجه می‌شود که y_{ij} دارای توزیع نرمال با میانگین $E(y_{ij}) = \mu$ و کواریانس زیر است

$$\mathrm{Cov}(y_{ij}, y_{hl}) = \begin{cases} \sigma^۲_\tau + \sigma^۲ & (ij) = (hl) \\ \sigma^۲_\tau & i = j, j \neq l \\ \circ & i \neq j \end{cases}$$

با توجه به فرض نرمال بودن y_i، اگر J_{n_i} یک ماتریس $n_i \times n_i$ از یک باشد آن‌گاه متغیر وابسته y_i دارای توزیع نرمال به صورت زیر است

$$y_i \sim N(\mu, v_i)$$

---

[۱] Missing values
[۲] Time-varying
[۳] Fixed variables

که در آن

$$V_i = \sigma^2 I_{n_i} + \sigma_\tau^2 J_{n_i}$$
$$= \begin{bmatrix} \sigma_\tau^2+\sigma^2 & \sigma^2 & \cdots & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2+\sigma^2 & \cdots & \sigma_\tau^2 \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_\tau^2 & \sigma_\tau^2 & \cdots & \sigma_\tau^2+\sigma^2 \end{bmatrix}$$

برای برآورد پارامترهای نامعلوم از روش ANOVA استفاده می‌شود. میانگین جامعه و میانگین نمونه به صورت زیر بدست می‌آیند.

$$\bar{y_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \qquad , \bar{y} = \frac{1}{N}\sum_{i=1}^{k}\bar{y_i} \tag{1.4}$$

که مجموع توان‌های دوم آن در جدول ۱.۴ آمده است

جدول ۱.۴: جدول ANOVA برای مدل اثرات تصادفی یکطرفه

مجموع توان دوم (SS)	متغیرها
$\mathrm{SSB} = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$	بین گروه‌ها
$\mathrm{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	یک گروه
$\mathrm{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n_I}(y_{ij} - \bar{y})^2$	کل

تعداد نتایج برای زمان برآورد پارامترهای نامعلوم شامل نرخ سیگنال به نویز[۴] که ناهمگونی بین گروه‌ها در ارتباط با فاصله متغیرها را بیان می‌کند استفاده می‌شود

$$\gamma = \frac{\sigma_\tau^2}{\sigma^2}$$

در مورد متعادل شدن توزیع‌ها از مجموع توان‌های دوم استفاده می‌کنیم. در این حالت طبق فرض نرمال بودن داریم

$$\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2_{N-k}$$
$$\frac{w(\gamma)\mathrm{SSB}}{\sigma^2} \sim \chi^2_{k-1}$$

که در آن $w(\gamma) = \frac{1}{1+n_i\gamma}$ و امید میانگین توان‌های دوم به صورت زیر حاصل می‌شود

$$E(\mathrm{MSE}) = E\left(\frac{\mathrm{SSE}}{N-k}\right) = \sigma^2$$
$$E(\mathrm{MSB}) = E\left(\frac{\mathrm{SSB}}{k-1}\right) = \sigma^2 + n_0\sigma_\tau^2$$

[۴] Signal to Noise

که در آن

$$n_\circ = \frac{N - \sum_i \frac{n_i^{۲}}{N}}{k - ۱} \qquad (۲.۴)$$

در این صورت برآورد پارامترهای نامعلوم عبارت است از

$$\hat{\mu} = \bar{y}$$

$$\hat{\sigma}^{۲} = \mathrm{MSE} = \frac{\mathrm{SSE}}{N-k}$$

$$\hat{\sigma}_\tau^{۲} = \frac{\mathrm{MSB} - \hat{\sigma}^{۲}}{n_\circ} = \frac{\mathrm{SSB}/(k-۱) - \mathrm{SSE}/(N-k)}{n_\circ}$$

برای مشاهدات و جزئیات و آگاهی بیشتر به مدسن و تایرگاد (۲۰۱۱) مراجعه کنید.

- **آزمون برای مؤلفه‌های تصادفی:**

با توجه به اثرات مدل‌سازی به صورت تصادفی واریانس اثرات تصادفی باید به‌طور قابل توجهی بیشتر از صفر باشد تا بتواند به مدل کمک کند. آزمون فرضیه معنی‌داری مؤلفه تصادفی واریانس $\sigma_\tau^{۲}$ عبارت است از

$$\begin{cases} H_\circ : \sigma_\tau^{۲} = \circ \\ H_۱ : \sigma_\tau^{۲} > \circ \end{cases}$$

اگر بتوان فرضیه صفر را رد کرد آن‌گاه واریانس اثر تصادفی معنی‌دار خواهد بود یعنی اثر تصادفی واریانس بزرگتر از صفر می‌باشد و اثر تصادفی واریانس در مدل، مشارکت قابل توجهی خواهد داشت. این بدان معنا است که متغیری که در میان گروه‌ها به طور تصادفی انتخاب و اندازه‌گیری شده به طور قابل توجهی بزرگتر از متغیری است که به تنهایی توسط خطا بیان می‌شود. اگر $\sigma_\tau^{۲} = \circ$ آن‌گاه همه اثرات تصادفی صفر می‌شود.

تحت فرضیه صفر، آماره آزمون F^* دارای توزیع فیشر با $k - ۱$ و $N - k$ درجه آزادی است که در آن

$$F^* = \frac{\mathrm{MSB}}{\mathrm{MSE}} = \frac{\mathrm{SSB}/k-۱}{\mathrm{SSE}/N-k} \sim F_{(k-۱),(N-k)} \qquad (۳.۴)$$

برای طرح متعادل، توزیع آماره آزمون تحت فرضیه متقابل به صورت زیر است

$$F^* = \frac{\mathrm{MSB}}{\mathrm{MSE}} = \frac{\mathrm{SSB}/k-۱}{\mathrm{SSE}/N-k} \sim (۱ + n\gamma) F_{(k-۱),(N-k)} \qquad (۴.۴)$$

برای درک بهتر موضوع مثال زیر را از فتر(۱۹۹۶) ارائه می‌کنیم این مثال میزان علاقه برای محاسبه شوندگان مختلف برای درخواست یک شغل را مطالعه می‌کند.

مثال شهودی:

۵ مصاحبه‌کننده و ۴ مصاحبه شونده (متقاضی) داریم که به طور تصادف برای مصاحبه انتخاب می‌شوند. مدل پاسخ به‌صورت زیر است

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \alpha \sim \mathcal{N}(0, \sigma_\alpha^2), \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

در این صورت

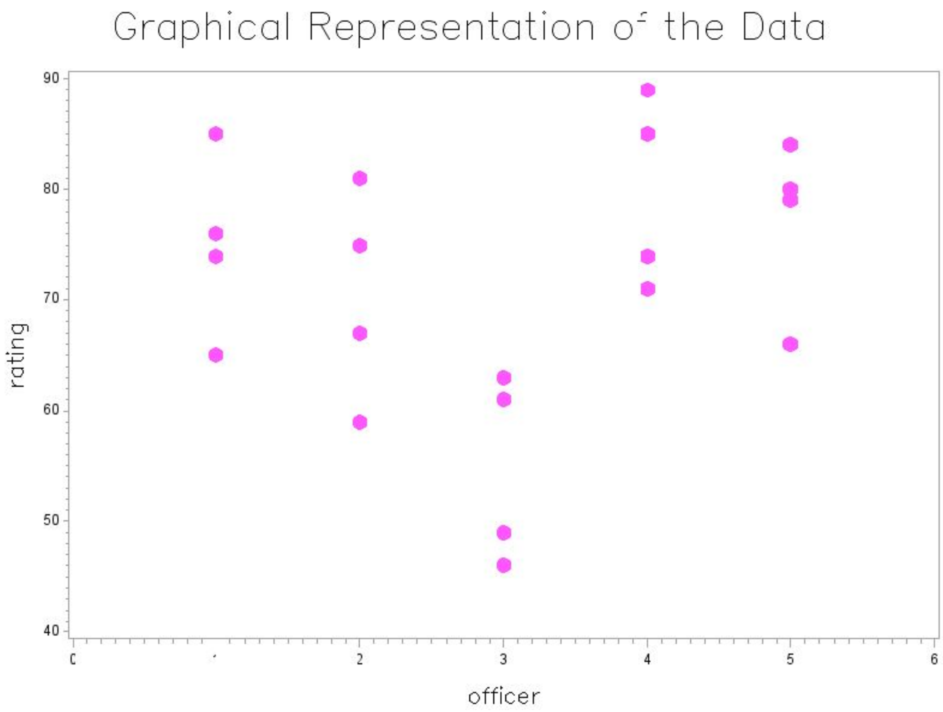
$$Y \sim \mathcal{N}(\mu, \sigma_\alpha^2 + \sigma^2).$$

در این مثال $n = 4$ بوده و ۲۰ مشاهده داریم. Y امتیازی بین ۰ تا ۱۰۰ است که مصاحبه‌کنندگان به هر متقاضی بعد از مصاحبه می‌دهند توجه به این نکته ضروری است که در امتیازات تغییر وجود دارد. داده‌های جمع‌آوری شده در جدول ۲.۴ نمایش داده شده است

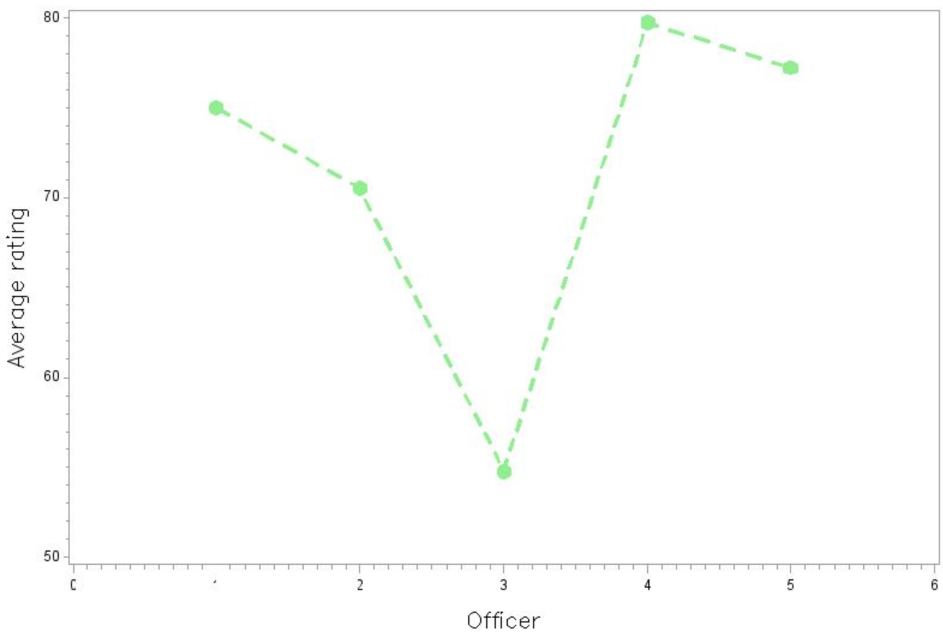
جدول ۲.۴: داده‌های درخواست شغل

مصاحبه‌کننده	امتیاز	مصاحبه‌کننده	امتیاز	مصاحبه‌کننده	امتیاز	مصاحبه‌کننده	امتیاز	مصاحبه‌کننده	امتیاز
۵	۶۶	۴	۷۴	۳	۴۹	۲	۵۹	۱	۷۶
۵	۸۴	۴	۷۱	۳	۶۳	۲	۷۵	۱	۶۵
۵	۸۰	۴	۸۵	۳	۶۱	۲	۸۱	۱	۸۵
۵	۷۹	۴	۸۹	۳	۴۶	۲	۶۷	۱	۷۴

و برای درک بهتر وضعیت نتایج بدست آمده از داده‌ها، شکل ۱.۴ را ببینید



شکل ۱.۴: نمودار امتیازات مصاحبه کنندگان در مثال درخواست شغل. محور افقی نشان‌دهنده مصاحبه‌کننده و محور عمودی امتیاز داده شده از ۰ تا ۱۰۰ است.



شکل ۲.۴: میانگین امتیازات برای هر مصاحبه کننده در مثال درخواست شغل

شکل ۱.۴ امتیازهایی که مصاحبه‌کننده‌ها به متقاضیان داده‌اند را نشان می‌دهد. شکل ۲.۴ میانگین نمراتی است که هر مصاحبه‌کننده داده است. به وضوح مشخص است که مصاحبه‌کنندگان مختلف میانگین امتیازاتی که داده‌اند متفاوت است. مثلا مصاحبه‌کننده ۳ دارای میانگین کمتری است. این نشان می‌دهد که این مصاحبه کننده سخت‌گیر است و نمرات پایین می‌دهد. سوال اینجا است که آیا اختلاف بین امتیازات مصاحبه‌کنندگان را تنها مؤلفه $\sigma^۲$ می‌تواند بیان کند یا نیاز به یک واریانس دیگر است که این اختلاف را نشان دهیم. حال می‌خواهیم آزمون زیر را انجام دهیم

$$\begin{cases} H_۰ : \sigma_\alpha^۲ = ۰ \\ H_۱ : \sigma_\alpha^۲ > ۰ \end{cases}$$

پارامترها با توجه به فرض مصاحبه شده و مقادیر آن‌ها برابر است با

$$\hat{\mu} = ۷۱/۴۵ \;, \hat{\sigma}^۲ = ۷۳/۲۸ \;, \hat{\sigma_\alpha}^۲ = ۸۰/۴۱.$$

برای آزمون داده شده مقدار آماره F را با مقدار بحرانی مقایسه می‌کنیم. با مقدار آماره‌های $F^* = ۵/۳۸۹$ و مقدار بحرانی $F_\alpha = ۰/۳۲۷۳$ فرضیه صفر در سطح معنی‌داری ۵٪ رد می‌شود و در این صورت می‌توان نتیجه گرفت که متغیری که مصاحبه‌کنندگان تعیین می‌کنند به‌طور قابل توجهی بزرگتر است از متغیری که توسط $\sigma^۲$ تعیین می‌شود. میزان تاثیر $\sigma^۲$ را می‌توان با استفاده از معیار زیر نشان محاسبه کرد

$$\frac{\sigma_\alpha^۲}{\sigma^۲ + \sigma_\alpha^۲} = \frac{۸۰/۴۱}{۷۳/۲۸ + ۸۰/۴۱} = ۰/۵۲.$$

بنابراین ۵۲٪ تغییرپذیری داده‌ها به‌وسیله اثر تصادفی بیان می‌شود و این نشان دهنده تنوع بین مصاحبه‌کنندگان است.

این مثال قدرت مدل اثر تصادفی را نشان می‌دهد. اگر مدل تحلیل متقارن با اثرات ثابت که در فصل قبل در مورد آن بحث کردیم را گسترش دهیم و بتوان اندازه اختلاف بین پاسخ‌دهندگان را شامل شود محقق می‌تواند تغییرپذیری مدل را قابل درک‌تر کند. چنانچه در مدل رگرسیون هم اثرات ثابت هم اثرات تصادفی در نظر گرفته شود به آن مدل آمیخته می‌گویند. در بخش بعد جزئیات مدل آمیخته خطی را شرح می‌دهیم.

۳.۴ مدل آمیخته خطی

صورت کلی مدل آمیخته خطی[۵] عبارت است از (جنسن و همکاران، ۲۰۰۷)

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i \qquad i = ۱, ۲, \cdots, m \tag{۵.۴}$$

[۵] Linear mixed model

توجه کنید که y_i یک بردار $n \times ۱$ از پاسخ‌ها، X_i یک ماتریس $n \times p$ برای اثرات ثابت، β یک بردار $p \times ۱$ از اثرات ثابت، Z_i ماتریسی $n \times q$ برای اثرات تصادفی، ε_i برداری $n \times ۱$ از خطاها و b_i برداری $q \times ۱$ از اثرات تصادفی نامعلوم هستند. که در آن n تعداد مشاهدات ، p تعداد اثرات ثابت، q تعداد اثرات تصادفی و m تعداد پاسخ‌های مستقل است. معمولا برای سادگی محاسبات فرض می‌شود که اثرات تصادفی و خطاها دارای توزیع نرمال با شرایط زیر است

$$\mathbf{b}_i \sim \mathcal{N}_q(۰, \mathbf{D}_i), \qquad \varepsilon_i \sim \mathcal{N}_n(۰, \mathbf{R}_i)$$

که در آن $\mathbf{D}_i$ یک ماتریس $q\times q$ قطری واریانس- کواریانس برای اثرات تصادفی (از آنجا که فرض می‌کنیم اثرات تصادفی ناهمبسته‌اند) و $\mathbf{R}_i$ یک ماتریس $n\times n$ واریانس- کواریانس برای مؤلفه خطا است. چون خطاها مستقل‌اند، ماتریس واریانس خطا $\mathbf{R}_i = \sigma^۲ I_n$ و ساختار واریانس کواریانس ساده بوده و $\mathbf{D}_i$ برای همه پاسخ‌دهندگان ثابت است از این رو $D_i = D$ می‌شود. همچنین باید توجه داشت که خطاها و اثرات تصادفی ناهمبسته‌اند.

در مدل آمیخته خطی خطاها می‌توانند مستقل یا همبسته باشند که انعطاف‌پذیری این مدل را نشان می‌دهد. در این مطالعه برای حفظ ساختار ساده مدل برای کاهش تعداد پارامترهایی که نیاز به برآورد دارند، خطاها همبسته در نظر گرفته می‌شوند. فرض می‌شود که اثرات تصادفی مستقل‌اند، و از آنجایی که ماتریس واریانس- کواریانس $\mathbf{D}$ قطری است، مدل را می‌توان تا جای ممکن توسعه داد. با توجه به مفروضات فوق، نتیجه میشود

$$\mathbf{y}_i \sim \mathcal{N}_n(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i), \qquad i = ۱, ۲, \cdots, m$$

که در آن

$$\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{R}_i$$

یک ماتریس $n \times n$ واریانس- کواریانس است. مدل‌های اثرات آمیخته برای شرح روابط بین یک متغیر پاسخ و تعدادی از متغیرها استفاده می‌شود که گروه‌بندی‌ها با توجه به یک یا بیشتر از یک عامل طبقه‌بندی انجام می‌شود. (پین‌هیرو و بیتس، ۲۰۰۴)

مدل آمیخته خطی، همان مدل خطی کلاسیک توسعه یافته است، و تنها اختلاف‌شان در این است که در مدل اثرات آمیخته یک اثر تصادفی نیز علاوه بر اثر ثابت وجود دارد.

تصمیم‌گیری برای تعیین اثرات ثابت و تصادفی از مشکلات این طرح است که وابسته به نوع چگونگی جمع‌آوری داده‌ها است.

از آنجایی که ساختار اولویت برای همه پاسخ‌دهندگان در مطالعه تحلیل متقارن یکسان نیست به مدلی نیاز داریم که تفاوت ساختار اولویت پاسخ‌دهندگان را در نظر داشته باشد.

۱.۳.۴ ساختار انباشته

راه دیگر برای نشان دادن مدل اثر آمیخته با نماد ماتریسی است. این مدل ماتریس پاسخ‌دهندگان را بالای هم قرار می‌دهد تا ماتریس بزرگی شامل اطلاعات پاسخ‌دهندگان تشکیل دهد. این

نوع نمایش برای ساده سازی محاسبات به‌کار می‌رود. در این حالت مدل $\mathbf{y}_i \sim \mathcal{N}_n(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i)$ را می‌توان برای m مشاهده به‌صورت زیر نوشت

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \varepsilon$$

که در آن $\mathbf{Y}$ یک بردار $N \times ۱$ از پاسخ‌ها برای همه مشخصه‌ها با $N = mn$، $\mathbf{X}$ ماتریس انباشته $N \times (p + ۱)$ از $X_i^\top$ها، $\boldsymbol{\beta}$ بردار $p \times ۱$ از اثرات ثابت، $\mathbf{Z}$ طرح بلوکی $N \times mp$ است، به عبارتی $\mathbf{Z} = \text{diag}(\mathbf{Z}_i)$ و $\mathbf{b}$ بردار $mp \times ۱$ شامل بردارهایی از اثرات تصادفی برای هر پاسخ فرض می‌شود که دارای توزیع نرمال $\mathbf{b} \sim \mathcal{N}(\circ, \mathbf{D})$ بوده و در آن $\mathbf{D} = \text{diag}(\mathbf{D}_i)$ است. همچنین ε یک بردار $N \times ۱$ شامل خطاها است که دارای توزیع نرمال $\varepsilon \sim \mathcal{N}_n(\circ, \mathbf{R})$ بوده و $\mathbf{R} = \text{diag}(\mathbf{R}_i)$ است. به عبارتی می‌توان نوشت

$$\begin{pmatrix} Y_۱ \\ Y_۲ \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} X_۱ \\ X_۲ \\ \vdots \\ X_m \end{pmatrix} \begin{pmatrix} \beta_۱ \\ \beta_۲ \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} z_۱ & \circ & \cdots & \circ \\ \circ & z_۲ & \cdots & \circ \\ \vdots & \cdots & \ddots & \vdots \\ \circ & \circ & \cdots & z_m \end{pmatrix} \begin{pmatrix} b_۱ \\ b_۲ \\ \vdots \\ b_m \end{pmatrix} + \begin{pmatrix} \varepsilon_۱ \\ \varepsilon_۲ \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

۴.۴ برآورد پارامترها

پارامترهای مختلف که در مدل اثرات آمیخته خطی نیاز به برآورد دارد شامل اثرات ثابت و اثرات تصادفی است. بسیاری از محققان اصطلاح برآورد را هنگام کار با اثرات تصادفی دوست ندارند، و در عوض اصطلاح پیش‌گویی را می‌پسندند. از این رو برآورد اثرات ثابت از طریق بهترین برآورد نااریب خطی[۶] (BLUE) به‌دست می‌آید و پیش‌بینی اثرات تصادفی از طریق بهترین پیش‌گویی نااریب خطی[۷] (BLUP) حاصل می‌شود. برآورد پارامتر در مدل‌های آمیخته می‌تواند دو جهت را دنبال کند. نخست وضعیتی را در نظر می‌گیریم که واریانس مؤلفه‌ها معلوم باشد. معمولا واریانس مؤلفه‌ها نامعلوم است که برآورد در این شرایط پیچیده می‌باشد، ولی زمانی که مؤلفه‌ها معلوم هستند می‌توان درک بهتری از فرآیند بدست آورد.

۱.۴.۴ برآورد با ماتریس واریانس‌ـ کواریانس معلوم

وقتی که $\mathbf{V}$ معلوم است می‌توان فرض کرد که $\mathbf{R}$ و $\mathbf{D}$ هم معلوم هستند. آن‌گاه فقط پارامترهای اثرات ثابت β و اثرات تصادفی $\mathbf{b}$ نیاز به برآورد دارند. ابتدا برآورد اثرات ثابت را بررسی می‌کنیم.
با ساده‌سازی توزیع مدل اثر آمیخته به مدل حاشیه‌ای برآوردها را با استفاده از روش ML

[۶] Best linear unbiased estimates
[۷] Best linear unbiased predictors

به سادگی می‌توان بدست آورد. مدل حاشیه‌ای به صورت زیر است:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon^*$$

که در آن $\varepsilon^* \sim \mathcal{N}(۰, V)$ است. چون Y دارای توزیع نرمال است تابع درستنمایی عبارت است از

$$L = \frac{۱}{(۲\pi)^{n/۲}|\mathbf{V}|^{۱/۲}} \exp\left\{-\frac{۱}{۲}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-۱}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}.$$

بنابراین لگاریتم تابع درستنمایی برابر است با

$$\begin{aligned}\mathrm{Log}L &= -\frac{n}{۲}\mathrm{Log}(۲\pi) - \frac{۱}{۲}\mathrm{Log}(\mathbf{V}) - \frac{۱}{۲}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-۱}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{n}{۲}\mathrm{Log}(۲\pi) - \frac{۱}{۲}\mathrm{Log}(\mathbf{V}) - \frac{۱}{۲} \\ &\{\mathbf{Y}^\top - \mathbf{V}^{-۱}\mathbf{Y} - \mathbf{Y}^\top\mathbf{V}^{-۱}\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{V}^{-۱}\mathbf{Y} + \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{V}^{-۱}\mathbf{X}\boldsymbol{\beta}\}\end{aligned}$$

برای تعیین برآورد پارامتر β از لگاریتم تابع درستنمایی نسبت به β مشتق می‌گیریم و آن را برابر صفر قرار می‌دهیم داریم

$$\begin{aligned}\frac{\partial \mathrm{Log}L}{\partial \beta} &= -\frac{۱}{۲}\left[۰ - \mathbf{Y}^\top\mathbf{V}^{-۱}\mathbf{X} - \mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{Y} + ۲\mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{X}\boldsymbol{\beta}\right] \\ &= -\frac{۱}{۲}\left[۲\mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{Y} + ۲\mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{X}\boldsymbol{\beta}\right] \\ &= ۰\end{aligned}$$

با توجه به این که $(\mathbf{X}^\top\mathbf{V}^{-۱}X)\boldsymbol{\beta} = \mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{Y}$، اگر ماتریس $\mathbf{X}^T\mathbf{V}^{-۱}\mathbf{X}$ معکوس‌پذیر باشد برآوردگر ML بردار β به صورت زیر نتیجه می‌شود

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{V}^{-۱}\mathbf{X})^{-۱}X^\top\mathbf{V}^{-۱}\mathbf{Y}$$

برای بدست آوردن پیش‌گویی اثر تصادفی مقدار امید شرطی b به شرط $\mathbf{Y}$ را تعیین می‌کنیم.

قضیه ۱.۴.۴. (جنسن، ۲۰۰۱)، طبق مفروضات این بخش اگر $\mathbf{Y}$ یک بردار $N \times ۱$ از پاسخ‌ها، $\mathbf{b}$ یک ماتریس $mp \times ۱$ از اثرات تصادفی آنگاه توزیع $\mathbf{b}$ به شرط $\mathbf{Y}$ عبارت است از

$$\mathbf{b}|\mathbf{Y} \sim \mathcal{N}_n(\mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-۱}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \mathbf{D} - \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-۱}\mathbf{Z}\mathbf{D})$$

که در آن $\mathbf{X}$ یک ماتریس انباشته $N \times (p+۱)$ از ماتریس‌های طراحی $X_i^\top$ برای اثرات ثابت، $\mathbf{Z}$ یک ماتریس انباشته قطری $N \times mp$ از ماتریس‌های طراحی $\mathbf{Z}^\top$ و $\mathbf{D}$ و $\mathbf{R}$ ماتریس‌های واریانس کواریانس هستند.

برهان. با توجه به توزیع $\mathbf{Y}$ و $\mathbf{b}$ داریم

$$\begin{pmatrix}\mathbf{Y} \\ \mathbf{b}\end{pmatrix} = \left\{\begin{pmatrix}\mathbf{X}\boldsymbol{\beta} \\ ۰\end{pmatrix}, \begin{pmatrix}\mathbf{V} & \mathbf{Z}\mathbf{D} \\ \mathbf{D}\mathbf{Z}^\top & \mathbf{D}\end{pmatrix}\right\}$$

با توجه به این که توزیع b به شرط Y دارای توزیع نرمال با میانگین

$$\boldsymbol{\mu}_{\mathbf{b}|\mathbf{Y}} = \circ + \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$$

و ساختار ماتریس واریانس کواریانس

$$\boldsymbol{\Sigma}_{\mathbf{b}|\mathbf{Y}} = \mathbf{D} - \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{D}. \tag{۶.۴}$$

است، نتیجه می‌شود پیش‌گویی اثر تصادفی عبارت است از

$$\hat{\mathbf{b}} = E(b|\mathbf{Y}) = \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}).$$

□

۲.۴.۴ برآورد با ماتریس واریانس ـ کواریانس نامعلوم

وقتی که $\mathbf{V}$ نامعلوم است ابتدا باید ساختار ماتریس‌های واریانس کواریانس $\mathbf{D}$ و $\mathbf{R}$ قبل از برآورد پارامترهای نامشخص β و b مشخص شود. این دیگر یک محاسبه مستقیم نیست چون برآوردها صورت بسته نخواهند داشت. یک راه حل بالقوه برای رفع این مشکل استفاده از یک روش تکراری برای پیدا کردن جواب بهینه است. روشی که برای برآورد در این حالت استفاده می‌شود الگوریتم (EM[۸]) است.

الگوریتم EM یک نوع الگوریتم تکراری بوده که برای محاسبه برآورد ML یا برآورد ماکزیمم درستنمایی محدود شده [۹] (REML) در حضور داده‌های گمشده یا غیرقابل مشاهده استفاده می‌شود. الگوریتم EM توسط دمپستر و همکاران (۱۹۷۷) در مقاله‌ای پایه‌ای تدوین و منتشر شد. استفاده از الگوریتم EM در این دهه‌های اخیر به شدت افزایش پیدا کرده و ارائه شد.

یک مشکل الگوریتم EM، این است که به کندی همگرا می‌شود اگر چه پیشنهادهایی برای سرعت بخشیدن به روند همگرایی در این الگوریتم وجود دارد. مینگ و وندیک (۱۹۷۷) استراتژی‌های را برای سرعت بخشیدن به این الگوریتم با حفظ سادگی و ثبات آن پیشنهاد کرده‌اند.

چار چوب نظری:

الگوریتم EM روشی است تکراری که برای پیدا کردن ماکزیمم درستنمایی از یک مدل آماری وقتی که وابستگی به متغیرهای نامعلوم قابل مشاهده نیست استفاده می‌شود. چارچوب‌های الگوریتم EM در دو مرحله اساسی دنبال می‌شود. مرحله اول مرحله برآورد است. در این مرحله تابع Q را که نشان‌دهنده امید ریاضی لگاریتم درستنمایی است تعریف می‌کنیم. این

[۸] Expectation maximization
[۹] Restricted maximum likelihood

مرحله تابع Q را در برآورد پارامتر ارزیابی می‌کند. دومین مرحله، مرحله ماکزیمم‌سازی نامیده می‌شود. این مرحله درستنمایی را به حداکثر می‌رساند.

چون اثرات تصادفی متغیرها غیرقابل مشاهده هستند فرض کنید داده گمشده بوده و پاسخ‌های داده مشاهده شده باشند. در این صورت داده‌ها را می‌توان به صورت زیر نشان داد

$$(y,b)=\{(y_i,b_i);\quad i=۱,۲,\cdots,m\} \tag{۷.۴}$$

$k(k=۱,۲,\cdots)$ تعداد تکرارها را نشان می‌دهد و تا زمانی که به همگرایی نرسیدیم فرآیند تکراری ادامه دارد، پس انتخاب اولین مجموعه از مقادیر ابتدایی الگوریتم EM به‌صورت زیر ادامه می‌یابد(وو، ۲۰۱۰)

- **مرحله E**

  در تکرار kام این مرحله لگاریتم درستنمایی شرطی داده‌های کامل و داده‌های مشاهده شده و برآورد پارامتر محاسبه می‌شود. در این حالت Q به صورت امید ریاضی شرطی تعریف می‌شود

  $$Q(\theta|\theta^{(k)})=E\left[\log L(\theta|Y,b)|Y,\theta^{(k)}\right].$$

- **مرحله M**

  این مرحله بیشترین امید ریاضی شرطی $Q(\theta|\theta^{(k)})$ را در مرحله E، با توجه به رابطه پارامترهای نامعلوم ،برای یافتن پارامترهای بروز شده پیدا می‌کند به‌طوری که داشته باشیم

  $$Q(\theta^{(k+۱)}|\theta^{(k)})\geq Q(\theta|\theta^{(k)}).$$

  این روند بارها و بارها تکرار می‌شود تا همگرایی رخ دهد این نکته مهم را هم باید توجه داشت که الگوریتم EM همگرایی در نقطه ماکزیمم دارد. نقاط شروع مهم هستند و باید نقاط مختلف را امتحان کرد تا به نقطه‌ای برسیم که الگوریتم در آن نقطه به ماکزیمم همگرایی کلی خود نزدیک شود. در ابتدا شاید به نقاط عجیبی برخورد کنیم جای تعجب ندارد چون الگوریتم EM به‌کندی همگرا می‌شود.

- **الگوریتم EM برای روش ML**

  از آنجایی که مرحله E با محاسبه امید ریاضی شرطی داده‌های لگاریتم درستنمایی شروع می‌شود، از این‌رو باید در مرحله نخست از تابع داده‌های لگاریتم درستنمایی گرفته شود. در این قسمت از مدل انباشته استفاده نمی‌شود بلکه از مدل (۵.۴) استفاده می‌شود.

  با توجه به داده‌ها

  $$(y,b)=\left\{(y_i,b_i);\quad i=۱,۲,\cdots,m\right\}.$$

تابع درستنمایی مدل آمیخته خطی داده‌های مشاهده شده با پاسخ‌های چندگانه (۵.۴) را می‌توان به‌صورت زیر نوشت

$$L(\theta|y) = \prod_{i=1}^{m} f(y_i|\theta)$$

و در آن $\theta = (\beta, \eta)$ برداری شامل پارامترهایی که باید برآورد شوند، و η نشان‌دهنده ماتریسی شامل پارامتری متمایز از ماتریس‌های واریانس کوواریانس $\mathbf{D}_i$ و $\mathbf{R}_i$ است. با توجه به مفهوم ابتدایی احتمال از مدل‌های آمیخته، تابع درستنمایی را می‌توان به شکل زیر بدست آورد

$$f(\mathbf{y}_i, \theta) = f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \mathbf{R}_i) f(\mathbf{b}_i|\mathbf{D}_i).$$

حال با توجه به $\mathbf{Y} \mid \mathbf{b} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}, \mathbf{R})$ و فرض‌های مدل آمیخته خطی $\mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}_i)$ و این فرض که $\mathbf{D}_i$ برای همه افراد یکسان است می‌توان نوشت

$$f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \mathbf{R}_i) = \frac{1}{(2\pi)^{-n/2}(|\mathbf{R}_i|)1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\right\} \quad (8.4)$$

$$= \frac{1}{(2\pi)^{-n/2}|\mathbf{R}_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{e}_i)^\top (\mathbf{R}_i)^{-1}(\mathbf{e}_i)\right\}$$

که $\mathbf{e}_i = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$ و

$$f(\mathbf{b}_i|\mathbf{D}) = \frac{1}{(2\pi)^{q/2}|\mathbf{D}|^{1/2}} exp\left\{-\frac{1}{2}\mathbf{b}_i\mathbf{D}_i^{-1}\mathbf{b}_i\right\} \quad (9.4)$$

بنابراین لگاریتم درستنمایی پاسخ‌دهنده iام برابر است با

$$\log L(\theta|\mathbf{y}_i, \mathbf{b}_i) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{R}_i| - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \quad (10.4)$$
$$\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) - \frac{q}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{D}| - \frac{2}{2}\mathbf{b}_i\mathbf{D}^{-1}\mathbf{b}_i$$

- **مرحله E** امید شرطی لگاریتم درستنمایی داده‌ها (Q) به صورت زیر حاصل می‌شود

$$Q = (\theta|\theta^{(k)}) = E\left[\log L(\theta|\mathbf{y}, \mathbf{b})|\mathbf{y}, \theta^k\right]$$
$$= E\left[\sum_{i=1}^{m}\left\{\log f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \mathbf{R}_i) + \log f(\mathbf{b}_i|\mathbf{D})|\mathbf{y}_i, \theta^{(k)}\right\}\right]$$

که $\log(\theta|\mathbf{y}_i, \mathbf{b}_i)$ را در (۱۰.۴) مشاهده کردید. مرحله E را می‌توان به‌وسیله آماره بسنده زیر که از پارامترهای واریانس کوواریانس η در ماتریس $\mathbf{R}_i$ و $\mathbf{D}$ ساده شده‌اند محاسبه کرد (وو، ۲۰۱۰)

$$\sum_{i=1}^{m} E(\mathbf{e}_i^\top \mathbf{e}_i|\mathbf{y}_i, \hat{\theta}^{(k)}) = \sum_{i=1}^{m}\left[\hat{\mathbf{e}}_i^{\top(k)}\hat{\mathbf{e}}_i^{(k)} + \mathrm{tr}\left(\mathrm{cov}(\mathbf{e}_i|\mathbf{y}_i, \theta^{(k)})\right)\right]$$

و

$$\sum_{i=1}^{m} E\left[\mathbf{b}_i^\top \mathbf{b}_i|\mathbf{y}_i,\hat{\theta}^{(k)}\right] = \sum_{i=1}^{m}\left[\hat{\mathbf{b}}_i^{\top(k)}\hat{\mathbf{b}}_i^{(k)} + \mathrm{cov}(\mathbf{b}_i|\mathbf{y}_i,\hat{\theta}^{(k)})\right]$$

که در آن

$$\begin{aligned}\hat{\mathbf{e}}_i =& \mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}\\ \hat{\mathbf{b}_i}^{(k)} =& \mathbf{D}(\hat{\eta})\mathbf{Z}_i^\top\mathbf{V}_i^{-1}(\eta^{\hat{(k)}})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}^{\hat{(k)}})\\ \mathbf{V}_i^{-1}(\hat{\eta}^{(k)}) =& \mathbf{Z}_i\mathbf{D}(\hat{\eta}^{(k)})\mathbf{Z}_i + \mathbf{R}_i(\hat{\eta}^{(k)})\end{aligned}$$

- **مرحله M**

برای این مرحله پارامترهایی که $Q(\theta|\theta^{(k)})$ را بیشترین مقدار می‌کنند برای تولید پارامتر به‌روز شده استفاده می‌شود. برآوردهای به‌روز شده پارامترها می‌توان با استفاده از رابطه زیر بدست آورد

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left[\sum_{i=1}^{m}\mathbf{X}_i^\top\hat{\mathbf{V}}_i^{-1}(\hat{\eta}^{(k)})\mathbf{X}_i\right]^{-1}\left[\sum_{i=1}^{m}\mathbf{X}_i^T\hat{\mathbf{V}}_i^{-1}(\hat{\eta}^{(k)})\mathbf{Y}_i\right] \qquad (۱۱.۴)$$

این برآورد پارامتر با زمانی که V معلوم بود یکسان است. مقادیر برآورد شده ماتریس واریانس کوواریانس برای تعیین برآورد اثرات ثابت استفاده می‌شود. تکرار نتایج باعث دقیق‌تر شدن نتیجه می‌شود. واریانس مؤلفه‌ها را می‌توان بوسیله امید شرطی $Q(\theta|\theta^{(k)})$ پیش‌گویی کرد.

قضیه ۲.۴.۴. (جنسن، ۲۰۰۷) طبق مفروضات این بخش از آنجایی که $\mathbf{Y}$ یک بردار انباشته $N \times 1$ از پاسخ‌ها و ε یک بردار انباشته از باقی‌مانده‌ها است، توزیع شرطی ε به شرط $\mathbf{Y}$ عبارت است از

$$\varepsilon|\mathbf{Y} \sim \mathcal{N}(\mathbf{R}^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}), \mathbf{R}-\mathbf{R}^\top\mathbf{V}^{-1}\mathbf{R})$$

که در آن $\mathbf{X}$ یک ماتریس $N \times p$ از ماتریس‌های طراح $\mathbf{X}_i^\top$ برای اثرات ثابت، $\mathbf{Z}$ یک ماتریس انباشته قطری $N \times mp$ از ماتریس‌های طراحی $\mathbf{Z}_i^\top$ و $\mathbf{D}$ و $\mathbf{R}$ ماتریس‌های واریانس کوواریانس هستند.

برهان. توزیع توأم $\mathbf{Y}$ و ε به شکل زیر است:

$$\begin{pmatrix}\mathbf{Y}\\ \varepsilon\end{pmatrix} = \left\{\begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ 0\end{pmatrix};\begin{pmatrix}\mathbf{V} & \mathbf{R}\\ \mathbf{R}^\top & \mathbf{R}\end{pmatrix}\right\}$$

در نتیجه ε به شرط $\mathbf{Y}$ دارای نرمال با میانگین

$$\begin{aligned}\boldsymbol{\mu}_{\varepsilon|\mathbf{Y}} &= 0 + \mathbf{R}^\top\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\\ &= \mathbf{R}^\top\mathbf{V}^{-1}(\mathbf{Y}-X\boldsymbol{\beta})\end{aligned}$$

و ساختار ماتریس واریانس ـ کواریانس

$$\mathbf{\Sigma}_{\varepsilon|\mathbf{Y}} = \mathbf{R} - \mathbf{R}^\top \mathbf{V}^{-1}\mathbf{R}$$

است. □

برای مدل آمیخته خطی برآورد پارامترها به صورت زیر محاسبه می‌شود:

$$\hat{\sigma}^{2(k+1)} = \frac{\sum_{i=1}^m E\left[e_i^{\top(k)} e_i^{(k)} | Y_i, \hat{\theta}^{(k)}\right]}{\sum_{i=1}^m n_i} = \frac{\sum_{i=1}^m E\left[e_i^{\top(k)} e_i^{(k)} | Y_i, \hat{\theta}^{(k)}\right]}{N} \qquad (12.4)$$

و در نتیجه

$$\begin{aligned}\sum_{i=1}^m E\left[\mathbf{e}_i^{\top(k)}\mathbf{e}_i^{(k)}|\mathbf{Y}_i, \hat{\theta}^{(k)}\right] &= \sum_{i=1}^m \left[\hat{\mathbf{e}}_i^{\top(k)}\hat{\mathbf{e}}_i^{(k)} + \mathrm{tr}(\mathrm{cov}(\mathbf{e}_i|\mathbf{Y}_i, \hat{\theta}^{(k)}))\right] \\ &= \sum_{i=1}^m \left[\hat{\mathbf{e}}_i^{\top(k)}\hat{\mathbf{e}}_i^{(k)} + \mathrm{tr}\left[\hat{\sigma}^{2(k)}\mathbf{I}_n + \hat{\sigma}^{4(k)}\mathbf{I}_n^\top \hat{\mathbf{V}}_i^{-1}(\hat{\eta}^{(k)})\mathbf{I}_n\right]\right] \\ &= \sum_{i=1}^m \left[\hat{\mathbf{e}}_i^{\top(k)}\hat{\mathbf{e}}_i^{(k)} + \mathrm{tr}\left[\hat{\mathbf{R}}_i^{(k)} - \hat{\mathbf{R}}_i^{(k)}\mathbf{I}_n^\top \hat{\mathbf{V}}_i^{-1}(\hat{\eta}^{(k)})\mathbf{I}_n\hat{\mathbf{R}}_i^{(k)}\right]\right]\end{aligned}$$

مؤلفه واریانس متغیرهای تصادفی و معادله (۶.۴) به صورت مشابه بدست می‌آید.

$$\begin{aligned}\hat{\mathbf{D}}^{(k+1)} &= \frac{\sum_{i=1}^m E\left[\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{Y}_i, \hat{\theta}^{(k)}\right]}{m} = \frac{\sum_{i=1}^m \left[\hat{\mathbf{b}}_i^{(k)}\hat{\mathbf{b}}_i^{\top(k)} + \mathrm{cov}(\mathbf{b}_i|\mathbf{Y}_i, \hat{\theta}^{(k)})\right]}{m} \qquad (13.4) \\ &= \frac{\sum_{i=1}^m \left[\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^\top + (\hat{\mathbf{D}}^{(k)} - \hat{\mathbf{D}}^{(k)}\mathbf{Z}_i^\top \mathbf{V}_i^{-1(k)}\mathbf{Z}_i\hat{\mathbf{D}}^{(k)})\right]}{m}\end{aligned}$$

- **مرحله نهایی:**

اگر در تکرار kام به همگرایی برسیم آن‌گاه مجموعه $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(k+1)}$، $\hat{\mathbf{D}} = \hat{\mathbf{D}}^{(k+1)}$، $\hat{\mathbf{R}} =$ $\hat{\sigma}^{2(k+1)}\mathbf{I}_n$ و $\hat{\mathbf{b}} = \hat{\mathbf{b}}^{(k+1)}$ جواب مسئله است در غیر این صورت یک واحد به k اضافه می‌شود و به مرحله E برمی‌گردد.

مثال ۱.۴.۴. این مثال توسط لسافری و لوسون (۲۰۱۲) مطرح شده که مطالعه‌ای درباره میزان مصرف کلسترول بین متغیرهای کمکی سن و جنس است هدف بررسی تاثیر سن و جنس بر روی جذب کلسترول در بدن است. مدل را به صورت زیر در نظر می‌گیریم

$$y_{ij} = \beta_0 + \beta_1 \mathrm{age}_{ij} + \beta_2 \mathrm{gender}_{ij} + b_{0i} + \varepsilon_{ij}$$

که در آن y_{ij} میزان مصرف کلسترول را در شخص iام و متغیر کمکی jام مشخص کند. به‌علاوه فرض می‌شود که $b_{0i} \sim \mathcal{N}(0, \sigma_{b_0}^2)$ و $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. با استفاده از الگوریتم EM نتیجه می‌گیریم

$$\hat{\beta} = \begin{bmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{bmatrix} = \begin{bmatrix}421/71 \\ -0/7144 \\ -55/16\end{bmatrix}$$

که در آن

$$\hat{\sigma}_b^{۲} = ۹۴/۲۸$$

پارامترهای برآورد شده نشان می‌دهد که سن افراد تاثیری بر میزان مصرف کلسترول ندارد چون تاثیر زیادی بر مدل ندارد، اما جنسیت بر عکس سن بر میزان مصرف کلسترول تاثیر دارد.

در داده‌هایی که ساختار پیچیده دارند می‌توان از روش‌های ماکزیمم درستنمایی محدود شده (REML) و نیوتون رافسون (NR) برای برآورد پارامترها استفاده کرد.

- **الگوریتم EM برای روش REML**

وقتی یک پارامتر نامعلوم با استفاده از روش REML برآورد می‌شود، از همان مراحل روش ML استفاده می‌شود. اما تفاوت روش REML با ML در این است که روش REML بخش درستنمایی را به دو قسمت تقسیم می‌کند که یکی از آن‌ها بدون اثرات ثابت است. برای بدست آوردن واریانس نامعلوم مؤلفه‌ها، بخش تابع درستنمایی که شامل اثرات ثابت نیست ماکزیمم می‌شود. (کوربیل و سیرل، ۱۹۷۹)

با استفاده از برآوردهایی که برای واریانس در مرحله قبل بدست آمد برآوردهایی برای اثرات ثابت بدست می‌آید. توصیه می‌شود که برای داده‌ها پیچیده از برآوردگر REML استفاده کنید چون واریانس مثبت است و نقاط دور افتاده تاثیری بر روی آن ندارد.

- **محاسبه برآوردها با استفاده از روش نیوتون رافسون (NR)**

یکی دیگر از روش‌های برآورد، که می‌توان پارامترهای نامعلوم را برآورد کرد روش نیوتون رافسون است. این یک رویکرد تکراری است که از روش‌های ریشه‌یابی استفاده می‌کند، در ادامه به طور مختصر در مورد آن توضیح داده می‌شود. (مک کولاک و همکاران، ۲۰۰۸)

شروع آن با تابع $f(\theta)$ می‌باشد و هدف این روش پیدا کردن ریشه با حل کردن

$$\acute{f}(\theta) = \frac{\partial f(\theta)}{\partial \theta} = ۰ \qquad (۱۴.۴)$$

است با این امید که ماکزیمم مقدار را به ما بدهد.

بسط تیلور رابطه (۱۴.۴) را حول $\theta_۰$ به‌صورت زیر در نظر بگیرید

$$\frac{\partial f(\theta)}{\partial \theta} = \acute{f}(\theta) \approx \acute{f}(\theta) + \frac{\partial^{۲} f(\theta)}{\partial \theta^{۲}}|_{\theta=\theta_۰}(\theta - \theta_۰). \qquad (۱۵.۴)$$

معادله (۱۵.۴) را برابر صفر قرار می‌دهیم در این صورت

$$\acute{f}(\theta_۰) + \frac{\partial^{۲} f(\theta)}{\partial \theta^{۲}}\ |_{\theta=\theta_۰}\ (\theta - \theta_۰) = ۰ \qquad (۱۶.۴)$$

با حل معادله (۱۶.۴) نتیجه به‌صورت زیر حاصل می‌شود

$$\theta = \theta_0 - \left[\frac{\partial^2 f(\theta)}{\partial\theta^2} \mid_{\theta=\theta_0}\right]^{-1} \acute{f}(\theta_0).$$

این کار برای ریشه‌های بعدی تکرار می‌شود تا برآوردها به‌دست آید. در این حالت مرحله $m+1$ام می‌توان نوشت

$$\theta^{m+1} = \theta^m - \left[\frac{\partial^{(m)} f(\theta)}{\partial\theta^{(m)}\partial\acute{\theta}^{(m)}} \mid_{\theta=\theta^{(m)}}\right]^{-1} \acute{f}(\theta^{(m)}).$$

لیندسترم و بیتس (۱۹۹۸) تفاوت‌های بین الگوریتم EM و روش NR را بررسی کردند. هر چند هر دو روش برای برآورد پارامتر ناشناخته سریع هستند، اما روش NR با تکرار کمتری به همگرایی می‌رسد و در همگرایی استوارتر است.

- **مثال:** مثال ۱.۴.۴ را در نظر بگیرید با استفاده از الگوریتم NR نتایج زیر به‌دست می‌آید

$$\beta = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} = \begin{bmatrix}423/98\\ -0/5719\\ -54/2712\end{bmatrix}$$

که در آن

$$\hat{\sigma}_b^2 = 110/66$$

است. این برآوردها به برآوردهای بدست آمده از الگوریتم EM در مثال قبل نزدیک است و نتایج حاصل از این مطالعه دقیق می‌باشد.

۵.۴ آزمون فرضیه

اگرچه در فصل قبل آزمون‌هایی برای اثرات ثابت ارائه شد در این قسمت علاوه بر تعمیم نتایج قبلی در حالت آزمون ترکیبات خطی، آزمون اثرات تصادفی را نیز ارائه می‌کنیم.

۱.۵.۴ آزمون برای اثرات ثابت

فقط چند آزمون برای برآورد اثرات ثابت وجود دارد. اولین مرحله آزمون اهمیت انتخاب ماتریس متقابل $\mathbf{L}$ است، که حاصل آن ترکیب خطی پارامتر برآوردپذیر $\boldsymbol{\beta}$ است و به‌صورت زیر می‌باشد

$$\begin{cases} H_0 : \mathbf{L}\boldsymbol{\beta} = 0 \\ H_1 : \mathbf{L}\boldsymbol{\beta} \neq 0 \end{cases}$$

از آزمون‌های tـ استیودنت، والد و فیشر برای آزمون دسته فرضیه‌های فوق استفاده می‌شود. آماره‌های آزمون‌های ذکر شده به ترتیب عبارت‌اند از

$$t = \frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top}}$$

$$Wald = \hat{\boldsymbol{\beta}}^\top \mathbf{L}^\top \left\{ \mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top \right\} \mathbf{L}\hat{\boldsymbol{\beta}}$$

$$F = \frac{\hat{\boldsymbol{\beta}}^\top \mathbf{L}^\top \left\{ \mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top \right\} \mathbf{L}\hat{\boldsymbol{\beta}}}{\mathrm{rank}\left(\mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top \right)}.$$

آماره آزمون والد تقریبا دارای توزیع $\chi^۲$ با درجه آزادی رتبه (Rank) ماتریس $\mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top$ است. وقتی که تعداد نقاط داده کوچک باشد آماره آزمون والد نتایج دقیق نمی‌دهد. آماره آزمون F دارای توزیع F با درجه آزادی رتبه $\mathbf{L}(\mathbf{X}^\top \hat{\mathbf{V}}^{-۱}\mathbf{X})^{-۱}\mathbf{L}^\top$ و درجه آزادی از برآورد مشتق داده‌ها است. برای مشاهده جزئیات به کاتری و نایک (۲۰۰۳) مراجعه کنید.

۲.۵.۴ آزمون برای مؤلفه‌های تصادفی

وقتی اهمیت اثرات تصادفی ارزیابی می‌شود، مؤلفه‌های واریانس این اثرات تصادفی بیشتر مورد نظر است. اگر مؤلفه واریانس صفر باشد، آن‌گاه اثر تصادفی در مدل نخواهد بود. یک محقق مدل آمیخته را به مدل ثابت ترجیح می‌دهد اگر اثرات تصادفی به‌طور قابل ملاحظه‌ای در مدل تاثیر داشته باشد (یعنی واریانس‌ها بزرگ‌تر از صفر باشد) آزمون فرضیه آن به‌صورت زیر است

$$\begin{cases} H_۰ : \sigma_b^۲ = ۰ \\ H_۱ : \sigma_b^۲ \neq ۰ \end{cases}$$

اگر فرضیه صفر رد شود، محقق متغیر وابسته را به عنوان یک اثر تصادفی در نظر می‌گیرد. یکی از راه‌های آزمون این فرضیه تعیین آماره آزمون تقریبی والد Z است. عبارت است از (کلین‌بوم و همکاران، ۲۰۰۸)

$$Z = \frac{\hat{\sigma}_b^۲}{S_{\hat{\sigma}_b^۲}}$$

هدف این آزمون مقایسه آزمون درستنمایی دو مدل (مدل با اثرات تصادفی و بدون اثرات تصادفی) با استفاده از آماره آزمون کای‌دو با درجه آزادی r است که r تعداد پارامترهای واریانس و کواریانس برآورد شده در ماتریس D است.

۶.۴ تحلیل متقارن مثال ماده شوینده

تحلیل متقارن برآوردگر OLS سطح افراد مصرف‌کننده برای پارامترهای دلخواه بین مصرف‌کنندگان ناهمگن را حساب می‌کند. در مطالعه تحلیل متقارن برآورد پارامترها و ناهمگنی اولویت

مصرف‌کنندگان را می‌توان پیگیری کرد و این پیگیری به محقق این امکان را می‌دهد که بتواند بازار را به چند قسمت تقسیم کند. مزیت تحلیل متقارن برای شرکت‌های بزرگ اثبات شده است که باعث سود آگاهانه می‌شود. متاسفانه تعداد نقاط داده‌های موجود در سطح مصرف‌کنندگان با توجه به مدت زمان محدود و مشتریان مصاحبه شده به‌طور کلی خیلی نزدیک به تعداد پارامترهای مدل است، لذا نتایج حاصل از تحلیل متقارن در مقابل با مدل آماری برای یک مصرف‌کننده تقریبا غیر ممکن است. (فریوواریس_اسچیتر و اوتر، ۱۹۹۹) یک راه حل آماری ممکن دیگر اضافه کردن مشاهدات از طریق مؤلفه تصادفی است و بنابراین استفاده از روش مدل آمیخته پیشنهاد می‌شود.

وقتی اولویت ۸۶ نفر اندازه‌گیری می‌شود لازم نیست که اولویت و افکار همه این افراد یکسان باشد. این تفاوت‌ها در اولویت پاسخ‌دهندگان، ایده اثرات تصادفی در مطالعه تحلیل متقارن را توجیه می‌کند. محقق به‌طور دقیق این تفاوت‌ها را نمی‌تواند مدل کند اما با استفاده از مؤلفه تصادفی در مدل اثرات ثابت محقق می‌تواند اثر تغییر افراد در مدل را ارزیابی کند. در این حالت

- $i = ۱, ۲, \cdots, ۸۶$ تعداد شرکت‌کنندگان در مطالعه می‌باشد.
- $j = ۱, ۲, \cdots, ۱۸$ تعداد مشخصه‌ها می‌باشد که مصرف کنندگان باید به آن‌ها امتیاز بدهند.
- p تعداداثرات ثابت و q تعداد اثرات تصادفی است.

در این مثال، چون مدل برای هر اثر ثابت شامل یک مؤلفه تصادفی می‌باشد، فرض می‌کنیم $q = p = ۹$ مدل آمیخته خطی را می‌توان به صورت زیر نوشت

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i \qquad (۱۷.۴)$$

که در آن ماتریس $n \times p$، $\mathbf{X}_i$ با استفاده از سطوح ویژگی خاص در هر مشخصه ساخته می‌شود، $\mathbf{Z}_i$ ماتریس $n \times q$ برای اثر تصادفی است، $\boldsymbol{\beta}$ یک بردار $p \times ۱$ از اثرات ثابت نامعلوم، (این اثرات برای همه پاسخ‌دهندگان ثابت است) $\mathbf{b}_i$ یک بردار $q \times ۱$ از اثرات تصادفی است (که برای همه پاسخ‌دهندگان ثابت است) و ε_i یک بردار $n \times ۱$ از خطاها می‌باشد. و فرض می‌شود که

$$\mathbf{b}_i \sim \mathcal{N}_q(\circ, \mathbf{D}_i) \;, \varepsilon_i \sim \mathcal{N}_n(\circ, \mathbf{R}_i)$$

برای مثال ماده شوینده همه پاسخ‌دهندگان دارای ماتریس سطح ویژگی یکسان هستند، چون همه آن‌ها یک ماتریس مشخصه یکسان را ارزیابی می‌کنند. این بدان معنا است که ماتریس X_i برای همه پاسخ‌دهندگان یکسان است. ماتریسی که در جدول ۴.۳ در فصل قبل طراحی شد یک مدل خاص از مدل اثر آمیخته خطی است، که از این مدل، می‌توان مدل ضرایب تصادفی را نام برد. برای این مدل $Z_i = X_i$، چون محقق اثرات تصادفی را در مدل برررسی می‌کند در

این حالت مدل آمیخته خطی به‌صورت زیر است

$$y_{ij} = \beta_1 x_{j1} + \beta_{F_1} x_{j2} + \beta_{F_2} x_{j3} + \beta_{A_1} x_{j4} + \beta_{A_2} x_{j5} + \beta_{D_1} x_{j6} + \beta_{B_1} x_{j7} + \beta_{P_1} x_{j8} + \beta_{P_1} x_{j9}$$
$$+ b_{1i} Z_{j1} + b_{F_{1i}} Z_{j2} + b_{F_{2i}} Z_{j3} + b_{A_{1i}} Z_{j4} + b_{A_{2i}} Z_{j5} + b_{D_{1i}} Z_{j6} + b_{B_{1i}} Z_{j7} + b_{P_{1i}} Z_{j8} + b_{P_{2i}} Z_{j9} + \varepsilon_{ij}$$

چون $X_i = Z_i$ در این مدل می‌توان اثرات تصادفی و اثرات ثابت را در یک گروه در نظر گرفت و بنابراین

$$\begin{aligned} y_{ij} =& (\beta_1 + b_{1i}) x_{j1} + (\beta_{F_1} + b_{F_{1i}}) x_{j2} + (\beta_{F_2} + b_{F_{2i}}) x_{j3} + (\beta_{A_1} + b_{A_{1i}}) x_{j4} + (\beta_{A_2} + b_{A_{2i}}) x_{j5} \\ &+ (\beta_{D_1} + b_{D_{1i}}) x_{j6} + (\beta_{B_1} + b_{B_{1i}}) x_{j7} + (\beta_{P_1} + b_{P_{1i}}) x_{j8} + (\beta_{P_2} + b_{P_{2i}}) x_{j9} + \varepsilon_{ij} \end{aligned}$$

که در نهایت می‌تواند به‌صورت زیر ساده‌تر شود

$$y_{ij} = \beta_{1i} x_{j1} + \beta_{F_{1i}} x_{j2} + \beta_{F_{2i}} x_{j3} + \beta_{A_{1i}} x_{j4} + \beta_{A_{2i}} x_{j5} + \beta_{D_{1i}} x_{j6} + \beta_{B_{1i}} x_{j7} + \beta_{P_{1i}} x_{j8} + \beta_{P_{2i}} x_{j9}$$

محقق با گنجاندن اثر تصادفی در مدل می‌تواند توزیع سطح ویژگی خاص برای پاسخ‌ها را به‌دست آورد. و در ادامه محقق به‌وسیله محاسبه واریانس مؤلفه‌های هر یک از سطوح ویژگی، می‌تواند پهنای ارزیابی پاسخ‌ها برای هر سطح ویژگی را مشاهده کند. با ارزیابی مؤلفه واریانس، محقق می‌تواند تغییر در میان تاثیر اولویت پاسخ‌ها بر مدل را بررسی کند. وقتی که واریانس اولویت‌ها بزرگ هست محقق برای تعیین شرایط کلی گروه پاسخ‌دهنده‌های مختلف در مورد یک ویژگی خاص دچار مشکل می‌شود. وقتی واریانس برای سطح ویژگی کوچک است به این معنا است که آیا سطح یک ویژگی خاص در انتخاب محصول کمک می‌کند یا خیر.

در این مطالعه، سه ساختار مختلف ماتریس واریانس کواریانس در مدل اثر آمیخته برای برآورد پارامترها در سطح معنی‌داری ۵٪ بررسی می‌شود. اگر چه تعدادی از پارامترها در مدل تاثیر چندانی ندارند، اما در مطالعه تحلیل متقارن حذف نمی‌شوند. یعنی این‌که در مدل تحلیل متقارن در قسمت بخش‌ارزشی پارامترهایی که تاثیر چندانی در مدل ندارند بر اولویت پاسخ‌دهندگان برای انتخاب مشخصه‌ها هم تاثیرگذار نیستند. در اینجا برای برآورد پارامترهای نامعلوم از روش ML استفاده می‌شود.

۱.۶.۴ مدل آمیخته خطی اول

اولین ساختار در نظر گرفته شده ساختار تقارن ترکیبی[10] است که دارای صورت زیر است

$$D_1 = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_1^2 \end{bmatrix}.$$

[10] Compound symmetry structure

این ساختار دارای یک واریانس برای همه اثرات تصادفی است. این بدان معناست که مدل نسبت به مدل‌های دیگر که در حال بررسی هستند دارای پارامتر کمتری است که نیاز به برآورد دارند، اما در این روش محاسبه واریانس‌های سطوح ویژگی مختلف وجود ندارد. با استفاده از این مدل برای نمونه ماده شوینده، مقادیر برآورد شده همراه با اهمیت پارامترها در جدول‌های ۳.۴ و ۴.۴ آورده شده است. در جدول ۳.۴ علامت × نشان‌دهنده معنی‌دار نبودن اثر متغیر

جدول ۳.۴: برآورد اثرات ثابت در مثال ماده شوینده با استفاده از ساختار ماتریس D_1 برای اثرات تصادفی در مدل آمیخته

اثرات		مقادیر برآورد	pــ مقدار	نتیجه
عرض از مبدا	$\hat{\beta}_1$	۳/۷۳۹۸	۰/۰۰۰۱>	
F_1	$\hat{\beta}_{F_1}$	−۰/۲۱۷۱	۰/۰۰۰۸	
F_2	$\hat{\beta}_{F_2}$	۰/۱۶۶۷	۰/۰۰۹۴	
A_1	$\hat{\beta}_{A_1}$	−۰/۳۴۵۰	۰/۰۰۰۱>	
A_2	$\hat{\beta}_{A_2}$	۰/۰۲۳۳	۰/۷۱۱۸	×
D_1	$\hat{\beta}_{D_1}$	۰/۵۱۰۲	۰/۰۰۰۱>	
B_1	$\hat{\beta}_{B_1}$	−۰/۱۵۴۱	۰/۰۰۸۴	
P_1	$\hat{\beta}_{P_1}$	۱/۱۳۱۸	۰/۰۰۰۱>	
P_2	$\hat{\beta}_{P_2}$	۰/۰۸۱۴	۰/۱۹۸۰	×

است.

جدول ۴.۴: برآورد مؤلفه‌های واریانس در مثال ماده شوینده با استفاده از ساختار ماتریس $D_۱$ برای اثرات تصادفی در مدل آمیخته

اثرات		مقادیر برآورد	pـ مقدار	نتیجه
عرض از مبدا	$\hat{\sigma}^۲$	۱/۲	۰/۰۰۰۱>	
$F_۱$	$\hat{\sigma}^۲_{F_۱}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$F_۲$	$\hat{\sigma}^۲_{F_۲}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$A_۱$	$\hat{\sigma}^۲_{A_۱}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$A_۲$	$\hat{\sigma}^۲_{A_۲}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$D_۱$	$\hat{\sigma}^۲_{D_۱}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$B_۱$	$\hat{\sigma}^۲_{B_۱}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$P_۱$	$\hat{\sigma}^۲_{P_۱}$	۰/۲۰۵۲	۰/۰۰۰۱>	
$P_۲$	$\hat{\sigma}^۲_{P_۲}$	۰/۲۰۵۲	۰/۰۰۰۱>	

۲.۶.۴ مدل آمیخته خطی دوم

دومین ساختار که مورد بررسی قرار می‌گیرد ساختار مؤلفه واریانس کامل است که صورت آن عبارت است از

$$D_۲ = \begin{bmatrix} \sigma_۱^۲ & ۰ & \cdots & ۰ \\ ۰ & \sigma^۲_{F_۱} & \cdots & ۰ \\ \vdots & \cdots & \ddots & \vdots \\ ۰ & ۰ & \cdots & \sigma^۲_{P_۲} \end{bmatrix}$$

همان‌طور که در فرضیه‌های اصلی مدل آمیخته بیان شد، در ساختار مدل اثرات آمیخته نوع دوم هر متغیر به تنهایی دارای مؤلفه واریانس مخصوص به خود است و ساختار $D_۲$ ناهمبسته بودن اثرات تصادفی را نشان می‌دهد اشاره دارد. نتایج برآوردها با استفاده از $D_۲$ در جداول ۵.۴ و ۶.۴ آورده شده است

جدول ۵.۴: برآورد اثرات ثابت در مثال ماده شوینده با استفاده از ساختار ماتریس D_2 برای اثرات تصادفی در مدل آمیخته

نتیجه	pـ مقدار	مقادیر برآورد		اثرات ثابت
	<۰/۰۰۰۱	۳/۷۳۹۸	$\hat{\beta}_1$	عرض از مبدا
	<۰/۰۰۰۱	−۰/۲۱۷۱	$\hat{\beta}_{F_1}$	F_1
	<۰/۰۰۰۱	۰/۱۶۶۷	$\hat{\beta}_{F_2}$	F_2
	<۰/۰۰۰۱	−۰/۳۴۵۰	$\hat{\beta}_{A_1}$	A_1
×	۰/۶۰۲۴	۰/۰۲۳۳	$\hat{\beta}_{A_2}$	A_2
	<۰/۰۰۰۱	۰/۵۱۰۲	$\hat{\beta}_{D_1}$	D_1
	۰/۰۰۰۹	−۰/۱۵۴۱	$\hat{\beta}_{B_1}$	B_1
	<۰/۰۰۰۱	۱/۱۳۱۸	$\hat{\beta}_{P_1}$	P_1
×	۰/۱۱۶۸	۰/۰۸۱۴	$\hat{\beta}_{P_2}$	P_2

جدول ۶.۴: برآورد مؤلفه‌های واریانس در مثال ماده شوینده با استفاده از ساختار ماتریس D_2 برای اثرات تصادفی در مدل آمیخته

نتیجه	pـ مقدار	مقادیر برآورد		اثرات ثابت
	<۰/۰۰۰۱	۱/۲۳۴۶	$\hat{\sigma}^2$	باقی‌مانده
	<۰/۰۰۰۱	۰/۵۲۰۲	$\hat{\sigma}^2_1$	عرض از مبدا
	۰/۰۰۵۹	۰/۰۶۶۱	$\hat{\sigma}^2_{F_1}$	F_1
×	.	۰	$\hat{\sigma}^2_{F_2}$	F_2
	<۰/۰۰۰۱	۰/۱۴۷۱	$\hat{\sigma}^2_{A_1}$	A_1
×	۰/۰۰۸۶	۰/۰۳۳۰	$\hat{\sigma}^2_{A_2}$	A_2
	<۰/۰۰۰۱	۰/۲۹۷۸	$\hat{\sigma}^2_{D_1}$	D_1
	۰/۰۰۰۲	۰/۰۹۴۰	$\hat{\sigma}^2_{B_1}$	B_1
	<۰/۰۰۰۱	۰/۴۳۹۴	$\hat{\sigma}^2_{P_1}$	P_1
	۰/۰۰۴۲	۰/۰۸۹۸	$\hat{\sigma}^2_{P_2}$	P_2

۳.۶.۴ مدل آمیخته خطی سوم

این مدل ماتریس واریانس‌ـ کواریانس را مطرح می‌کند که ساختاری ندارد. این ماتریس فرض اثرات تصادفی ناهمبسته را از بین می‌برد و صورت آن عبارت است از

$$D_۳ = \begin{bmatrix} \sigma_I^۲ & \sigma_{IF_۱} & \cdots & \sigma_{IP_۲} \\ \sigma_{IF_۱} & \sigma_{IF_۲}^۲ & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{IP_۱} & \cdots & \sigma_{P_۱P_۲} & \sigma_{P_۲}^۲ \end{bmatrix}$$

با توجه به $D_۳$ واریانس اثر تصادفی و واریانس عرض‌ازمبدا باید برآورد شوند همچنین رابطه بین ساختار کواریانس‌ها هم نیاز به برآورد دارد. برای مشاهده معنی‌دار بودن اثرات تصادفی و واریانس‌ها که از ماتریس واریانس کواریانس بدست آمده، جدول‌های ۷.۴ و ۸.۴ را بررسی می‌کنیم.

جدول ۷.۴: در مثال ماده شوینده با استفاده از ساختار ماتریس $D_۳$ برای اثرات تصادفی در مدل آمیخته برآورد اثرات ثابت

اثرات ثابت		مقادیر برآورد	pـ مقدار	نتیجه
عرض از مبدا	$\hat{\beta}_۱$	۳/۷۳۹۸	۰/۰۰۰۱>	
$F_۱$	$\hat{\beta}_{F_۱}$	−۰/۲۱۷۱	۰/۰۰۰۱>	
$F_۲$	$\hat{\beta}_{F_۲}$	۰/۱۶۶۷	۰/۰۰۰۱>	
$A_۱$	$\hat{\beta}_{A_۱}$	−۰/۳۴۵۰	۰/۰۰۰۱>	
$A_۲$	$\hat{\beta}_{A_۲}$	۰/۰۲۳۳	۰/۶۰۲۴	×
$D_۱$	$\hat{\beta}_{D_۱}$	۰/۵۱۰۲	۰/۰۰۰۱>	
$B_۱$	$\hat{\beta}_{B_۱}$	−۰/۱۵۴۱	۰/۰۰۰۹	
$P_۱$	$\hat{\beta}_{P_۱}$	۱/۱۳۱۸	۰/۰۰۰۱>	
$P_۲$	$\hat{\beta}_{P_۲}$	۰/۰۸۱۴	۰/۱۳۳۷	×

جدول ۸.۴: در مثال ماده شوینده با استفاده از ساختار ماتریس D_3 برای اثرات تصادفی در مدل آمیخته برآورد مؤلفه‌های واریانس

اثرات ثابت		مقادیر برآورد	pـ مقدار	نتیجه
باقی‌مانده	$\hat{\sigma}^2$	۱	۰/۰۰۰۱>	
عرض از مبدا	$\hat{\sigma}_1^2$	۰/۵۳۵۲	۰/۰۰۰۱>	
F_1	$\hat{\sigma}_{F_1}^2$	۰/۰۹۵۸	۰/۰۰۳۴	
F_2	$\hat{\sigma}_{F_2}^2$	۰/۰۱۶۳	۰/۲۴۵۸	$\times$
A_1	$\hat{\sigma}_{A_1}^2$	۰/۱۸۶۲	۰/۰۰۰۱>	
A_2	$\hat{\sigma}_{A_2}^2$	۰/۰۵۸۴	۰/۰۲۵۳	
D_1	$\hat{\sigma}_{D_1}^2$	۰/۳۰۵۵	۰/۰۰۰۱>	
B_1	$\hat{\sigma}_{B_1}^2$	۰/۰۹۷۵	۰/۰۰۰۱	
P_1	$\hat{\sigma}_{P_1}^2$	۰/۴۹۱۸	۰/۰۰۰۱>	
P_2	$\hat{\sigma}_{P_2}^2$	۰/۱۱۶۲	۰/۰۰۹۳	

وقتی که مدل سوم را مقایسه می‌کنیم مشاهده می‌شود که مقادیر برآورد برای پارامترهای اثرات ثابت در هر سه مدل یکسان است. فقط دوتا از این برآورد پارامترها معنی‌دار نیستند که نشان‌دهنده این است که پاسخ‌دهنده علاقه‌ای به این سطح از ویژگی ندارد. اگر مقادیر مصرف‌شده بخش‌ارزشی مثبت باشد نشان می‌دهد که یک ویژگی خاص برای پاسخ‌دهنده در اولویت هست (پاسخ‌دهنده یک ویژگی خاص را ترجیح می‌دهد) و اگر مقادیر مصرف‌شده بخش‌ارزشی منفی باشد نشان می‌دهد که آن ویژگی برای پاسخ‌دهنده در اولویت نیست (پاسخ‌دهنده نظر چندانی روی آن ویژگی ندارد) و اگر هم این مقدار نزدیک صفر باشد نشان‌دهنده این است که پاسخ‌دهنده نسبت به آن ویژگی بی‌تفاوت است. طبق تغییراتی که در ساختار ماتریس واریانس کواریانس اتفاق می‌افتد واریانس برآورد می‌شود. برای مدل اول مؤلفه واریانس اثرات تصادفی یکسان است و معنی‌دار هستند.

برآور واریانس‌ها در مدل دوم و سوم خیلی ساده است. تنها واریانسی که در مدل سوم معنی‌دار نیست واریانس مربوط به سطح کنسانتره ویژگی شکل محصول است. وقتی که مؤلفه واریانس یک سطح ویژگی خاص به صفر نزدیک است نشان‌دهنده این است که همه پاسخ‌دهندگان نسبت به این سطح ویژگی تمایل چندانی ندارند. در ادامه مدل‌هایی که مورد بررسی قرار گرفتند مقایسه و در مورد نتایج آن‌ها بحث می‌شود.

۷.۴ مقایسه مدل‌ها

محققان همواره با سؤال، چه مدلی بهترین برازش را دارد مواجه می‌شوند. انتخاب بهترین مدل برای داده همیشه یک کار راحت نیست. با استفاده از معیار اطلاع آکائیک[۱۱] (AIC) می‌توان بهترین مدل از بین مدل‌های مختلف که برای داده‌ها در نظر گرفته شده را انتخاب کرد. روش AIC از تعریف اطلاع کولبک_لیبلر[۱۲] استفاده می‌کند که یک معیار فاصله بین دو تابع چگالی می‌باشد. معیار AIC با استفاده از رابطه زیر حساب می‌کنیم

$$\mathrm{AIC} = ۲\log(L) + ۲d \qquad (۱۸.۴)$$

که در آن d نشان‌دهنده تعداد پارامترهایی است که برآورد شده‌اند. مدلی بهتر است که AIC آن کمتر باشد، در مثال ماده شوینده، مدل آمیخته سوم با مؤلفه‌های واریانس بدون ساختار دارای AIC کمتر بوده و به عنوان مدل بهتر انتخاب می‌شود. نتایج AIC در جدول ۹.۴ آمده است.

جدول ۹.۴: مقایسه مدل‌ها برای مثال ماده شوینده

AIC	مدل
۵۸۹۷/۴	مدل تحلیل متقارن سنتی اثرات ثابت
۵۵۹۰/۴	مدل اثرات آمیخته نوع اول
۵۴۷۳/۲	مدل اثرات آمیخته نوع دوم
۵۴۶۰/۹	مدل اثرات آمیخته نوع سوم

اضافه کردن مؤلفه‌های تصادفی به مدلی که دارای اثرات ثابت است، می‌تواند به طور مؤثری در محاسبه همبستگی بین متغیرها کمک کند. با توجه به خروجی جدول ۹.۴ مدل آمیخته خطی نوع اول بیان می‌کند که به‌وسیله اضافه کردن اثرات تصادفی به مدل برازش کلی مدل بهبود پیدا کرده است. و این تغییرات در مدل اولیه که باعث بهبود شده است به محقق این انگیزه را می‌دهد که به‌وسیله بعضی از این تغییرات بتواند ناهمگونی بین متغیرها را شرح دهد. با محدود کردن مؤلفه‌های واریانس و یکسان بودن واریانس‌ها، محققان نمی‌توانند به‌طور مؤثری ساختار واریانس کواریانس هر سطح ویژگی را مورد بررسی قرار دهند. از این رو با مدل کردن ساختار واریانس کواریانس همان‌طور که در مدل آمیخته خطی نوع دوم انجام شده، محققان می‌توانند یک مؤلفه واریانس متفاوتی برای هر سطح ویژگی را مدل کنند. این واریانس‌ها اطلاعات خوبی درباره اختلاف نمونه اولویت پاسخ‌دهندگان برای هر سطح ویژگی

[۱۱] Akaik information criteria

[۱۲] Kullback-Leibler information

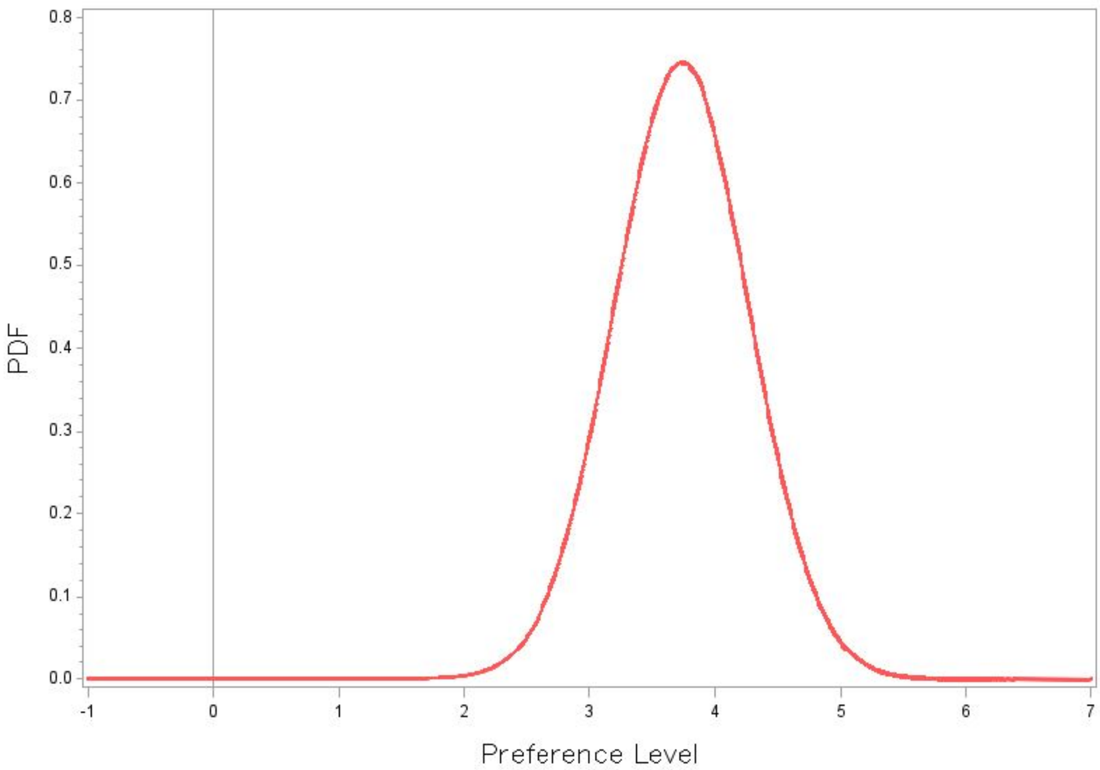
می‌دهد. مدل سوم، فرض متغیرهای تصادفی ناهمبسته را نادیده گرفته و بین واریانس‌ها ارتباط برقرار می‌کند. نتایج این مدل بهترین برآورد را می‌دهد. برآوردهای ناشی از مدل باساختار D_3 در جدول ۱۰.۴ اثر پارامترهای ثابت را همانند همه اثر واریانس‌های تصادفی در نظر می‌گیرد. به‌وسیله برآورد کردن دوباره پارامترها، پارامترهایی که برآورد نشده‌اند در مدل به‌دست می‌آیند.

جدول ۱۰.۴: مجموعه تمام برآوردهای مدل سوم برای مثال ماده شوینده

برآورد واریانس		مقدار برآورد		اثرات
۱	$\hat{\sigma}^2$			باقی‌مانده
۰/۵۳۵۲	$\hat{\sigma}^2_{d_1}$	۳/۷۳۹۸	$\hat{\beta}_1$	عرض از مبدا
۰/۰۹۵۸	$\hat{\sigma}^2_{F_1}$	−۰/۲۱۷۱	$\hat{\beta}_{F_1}$	F_1
۰/۰۱۶۳	$\hat{\sigma}^2_{F_2}$	۰/۱۶۶۷	$\hat{\beta}_{F_2}$	F_2
۰/۰۵۱۸۳	$\hat{\sigma}^2_{F_3}$	۰/۰۵۰۴	$\hat{\beta}_{F_3}$	F_3
۰/۱۸۶۲	$\hat{\sigma}^2_{A_1}$	−۰/۳۴۵۰	$\hat{\beta}_{A_1}$	A_1
۰/۰۵۸۴	$\hat{\sigma}^2_{A_2}$	۰/۰۲۳۲۶	$\hat{\beta}_{A_2}$	A_2
۰/۱۴۶۴	$\hat{\sigma}^2_{A_3}$	۰/۳۲۱۷۴	$\hat{\beta}_{A_3}$	A_3
۰/۳۰۵۵	$\hat{\sigma}^2_{D_1}$	۰/۵۱۰۲	$\hat{\beta}_{D_1}$	D_1
۰/۳۰۵۵	$\hat{\sigma}^2_{D_2}$	−۰/۵۱۰۲	$\hat{\beta}_{D_2}$	D_2
۰/۰۹۷۵	$\hat{\sigma}^2_{B_1}$	−۰/۱۵۴۱	$\hat{\beta}_{B_1}$	B_1
۰/۰۷۹۵	$\hat{\sigma}^2_{B_2}$	۰/۱۵۴۱	$\hat{\beta}_{B_2}$	B_2
۰/۴۹۱۸	$\hat{\sigma}^2_{P_1}$	۱/۱۳۱۸	$\hat{\beta}_{P_1}$	P_1
۰/۱۱۶۲	$\hat{\sigma}^2_{P_2}$	۰/۰۸۱۴	$\hat{\beta}_{P_2}$	P_2
۰/۴۱۷۲	$\hat{\sigma}^3_{P_3}$	−۱/۲۱۳۲	$\hat{\beta}_{P_3}$	P_3

محققان بوسیله برآورد میانگین سطوح خاص و واریانس این سطوح، مقادیر بخش‌ارزشی رویکردهای ضروری برای اولویت پاسخ‌دهندگان نمونه را بررسی می‌کنند. اثرات ثابت برآورد شده در مدل را می‌توان به‌طور دقیق در شکل ۱.۳ مشاهده کرد و به‌وسیله این مقادیر به‌دست آمده محققان نمونه مورد نظر را ارزیابی می‌کنند. زمانی که داده‌های ماده شوینده جمع‌آوری و مورد مطالعه قرار گرفتند، محققان برای اطمینان کار، نمونه‌ای را از بین ۸۶ پاسخ‌دهنده انتخاب کردند تا مشاهده کنند آن چیزی که برای بازاریابی در نظر گرفته‌اند درست است یا خیر. اگرچه بسیاری از کارخانه‌ها علاقه‌مند هستند تا بهترین محصول را تولید کنند اما بعضی اوقات اطلاعات به‌دست آمده باعث می‌شود تغییراتی در محصول ایجاد شود. اگرچه سطوح

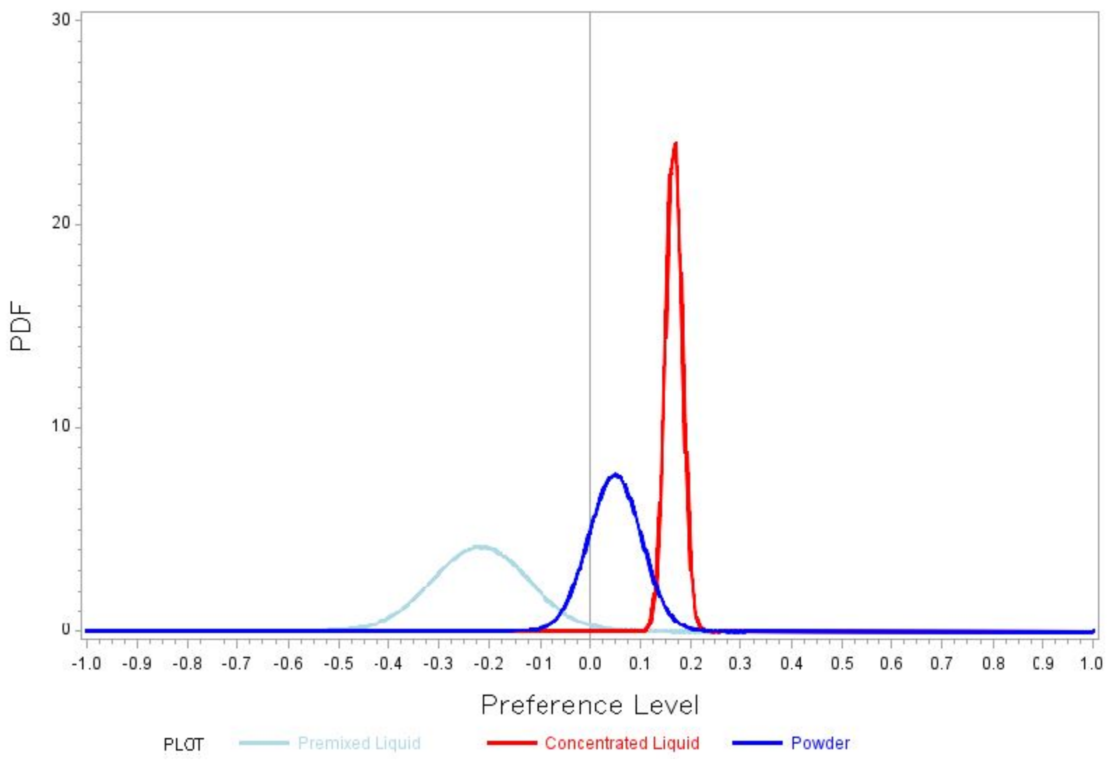
طبق بیشترین مقدار بخش‌ارزشی رده‌بندی می‌شوند اما گاهی اوقات بازار به مواردی نیاز دارد که بهترین مورد نباشد به عبارتی خصوصیات محصول ممکن است بهترین نباشد. اگر تعداد افرادی که از محصول استفاده می‌کنند زیاد شود در این صورت شرکت تولید بیشتر و در ادامه سود بیشتری خواهد داشت. موارد به‌دست آمده در برآوردها برای هر سطح اولویت افراد را برای محصول نشان می‌دهد که پاسخ‌دهندگان چقدر تمایل دارند که یک محصول یا محصولات شرکت را انتخاب کنند. شکل ۳.۴ این اولویت‌ها را نشان می‌دهد.



شکل ۳.۴: درستنمایی کلی خرید مثال ماده شوینده

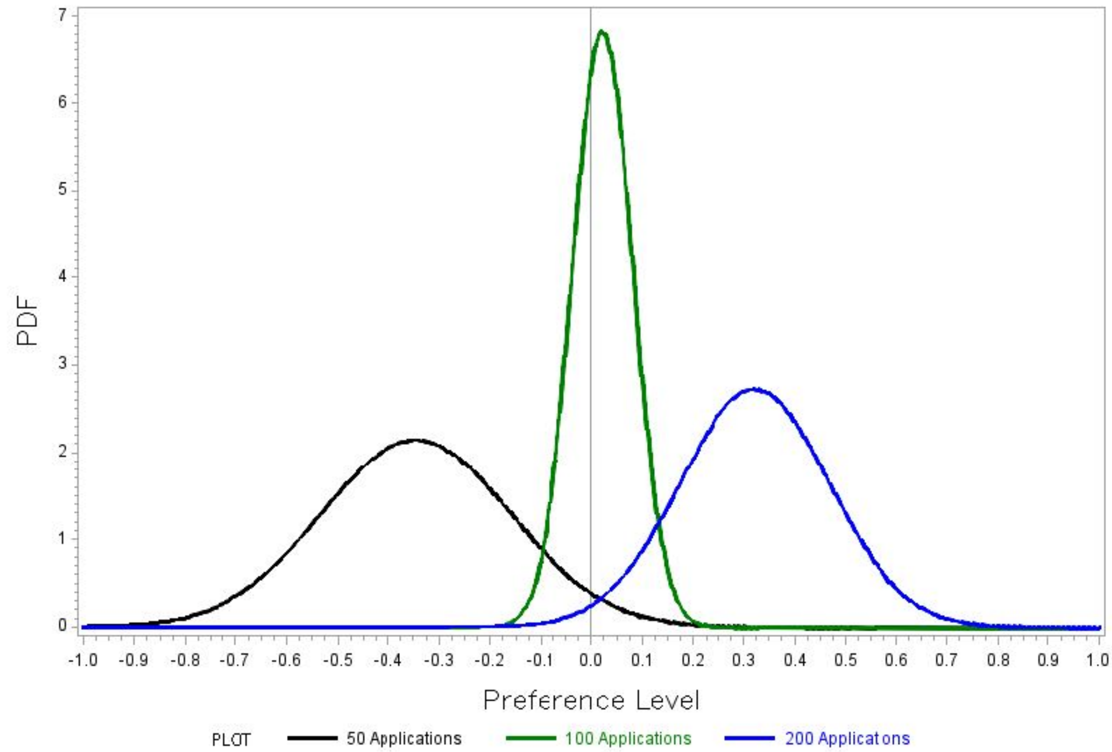
طبق آن چیزی که در شکل ۳.۴ می‌بینید مشتریان نسبتا هر ۱۸ مشخصه را انتخاب کرده‌اند. زمانی‌که متغیر تاثیر زیادی روی مدل آمیخته خطی داشته باشد، واریانس بزرگ در درستنمایی خرید دیده می‌شود. پراکندگی زیاد حول میانگین نشان می‌دهد در اولویت گروه پاسخ‌دهندگان یک تغییر بزرگ رخ می‌دهد زمانی که مشتریان مجبورند بین مشخصه‌های ممکن یکی را انتخاب کنند. این متغیرها محقق را به بررسی احتمال گروه پاسخ‌دهندگان به اجزای بازار خاص تشویق می‌کنند. اگر محققان مجبور شوند که اطلاعاتی حول پاسخ‌دهندگان اضافه کنند، محقق باید از اطلاعات طبقه‌بندی شده پاسخ‌دهندگان و عناصر درست این مدل استفاده کند. تحلیل سازگاری تحلیل متقارن (ACA) مشکل محقق در مورد اضافه کردن اطلاعات را حل می‌کند. اگرچه پژوهشگر یا محقق هیچ‌گونه اطلاعات اضافی در مورد این پاسخ‌دهندگان ندارد، اما می‌تواند با استفاده از تکنیک‌های خوشه‌بندی، گروه مشتریانی که دارای ساختار مشابه

هستند به پاسخ‌دهندگانی با مقادیر بخش‌ارزشی β_s تقسیم کند. این بخش‌های مشخص شده (خوشه‌بندی شده) برای ارزیابی بازار است که ممکن است بعضی از این بخش‌ها منحصربفرد باشد (هایر و همکاران، ۲۰۰۶). اولین ویژگی که بررسی می‌شود شکل محصول است.



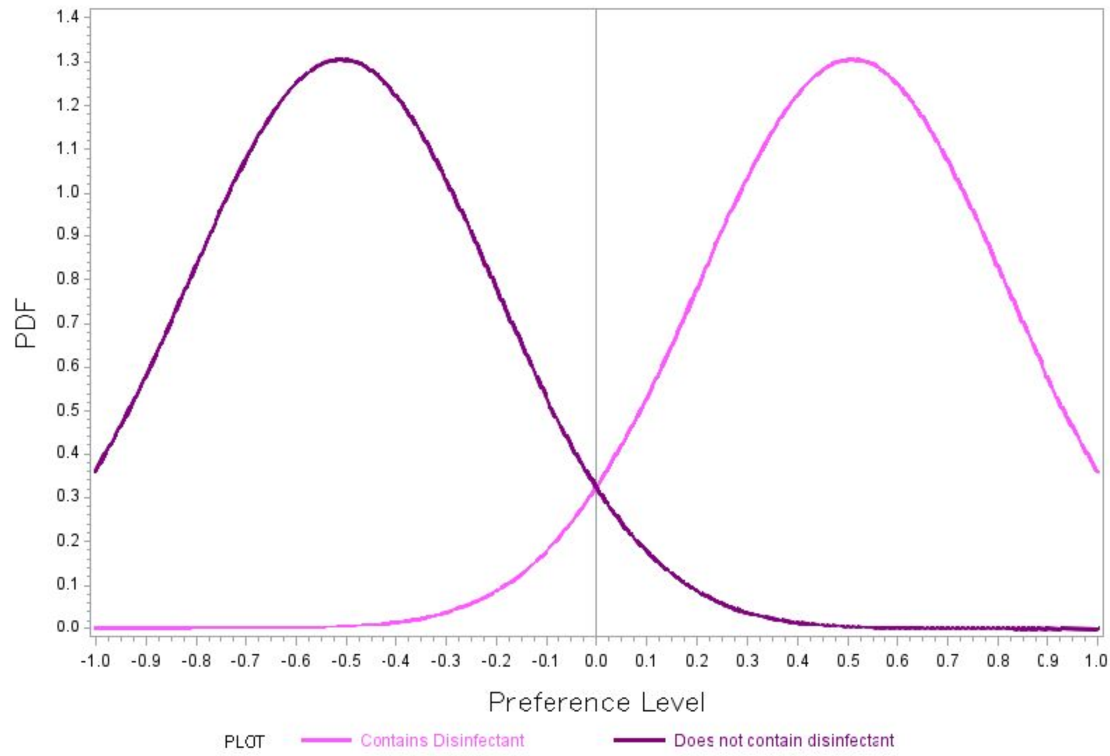
شکل ۴.۴: شکل محصول

زمانی که محقق در نظر دارد تا مقادیر بخش‌ارزشی شکل محصول را با استفاده از تحلیل متقارن تعیین کند، حالت کنسانتره بودن با توجه به انتخاب‌ها در اولویت می‌باشد. شکل ۴.۴ توزیع خط قرمز اولویت‌های مربوط به شکل کنسانتره محصول را نشان می‌دهد. واضح است که شکل کنسانتره محصول توسط پاسخ‌دهندگان ترجیح داده می‌شود. از آنجایی که واریانس این سطح ویژگی بسیار کوچک است، این نشان می‌دهد که برای پاسخ‌دهندگان این سطح از ویژگی در اولویت است. با تحقیق بیشتر در مورد شکل محصول قسمت پودر که توزیع آن با آبی تیره نشان داده شده است، می‌توان دید بخش بزرگی از پاسخ‌دهندگان این شکل محصول را ترجیح می‌دهند اگرچه به اندازه شکل کنسانتره نمی‌باشد. با این نتیجه این شرکت می‌تواند محصول خود را با این دو شکل مختلف تولید کند.



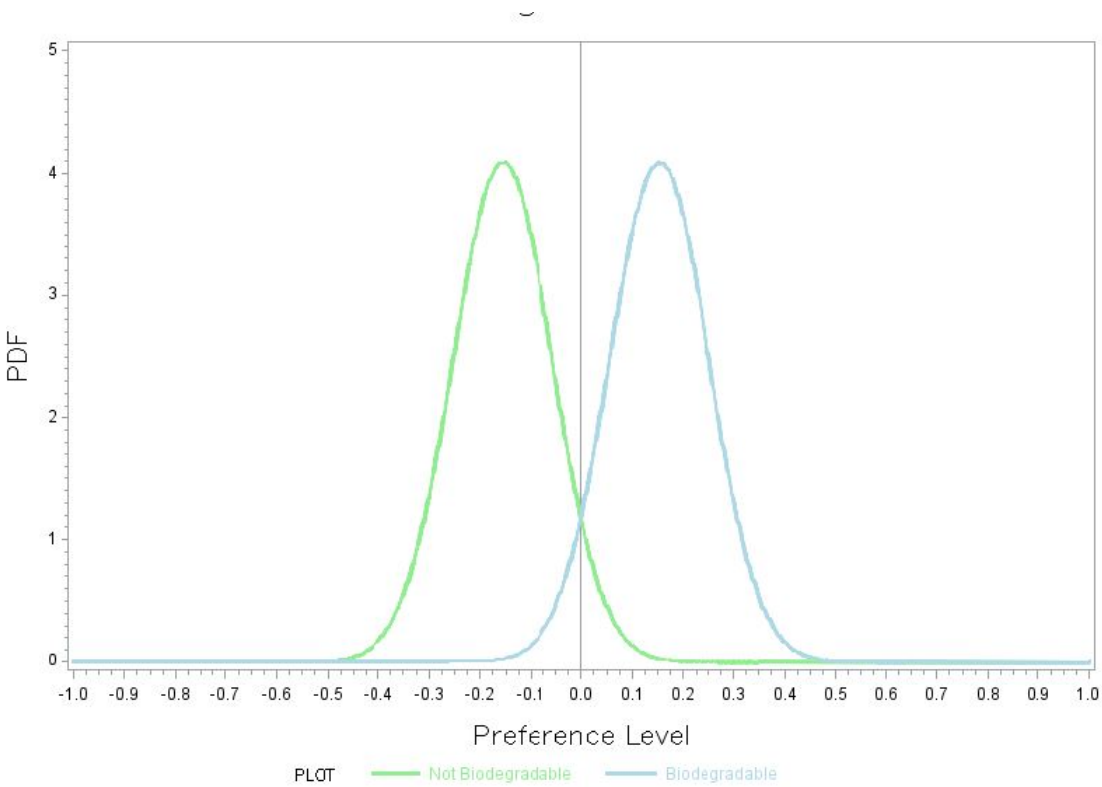
شکل ۵.۴: تعدادی از کاربردها

ویژگی دوم، تعداد کاربردها در هر ظرف می‌باشد که دارای سه سطح مختلف است. ۵.۴ انتظار می‌رفت که تقریبا تمام پاسخ‌دهندگان حداقل تعداد درخواست‌ها را نادیده بگیرند. زمانی که یک مشتری ظرفی با قیمت خاص را در نظر می‌گیرد، طبیعی است که ظرفی را انتخاب کند که دارای کاربرد بیشتر است. بخش کوچکی از جمعیت سطح کاربرد کم در هر ظرف را ترجیح می‌دهند. ظرف‌هایی که دارای کاربرد بیشتر هستند تقریبا بوسیله تمام افرادترجیح داده می‌شود. محققان باید طوری مصرف‌کننده را در نظر بگیرند و محصول را تولید کنند که درآمد مصرف‌کننده را در نظر داشته باشد به گونه‌ای که بیشترین تولید و بیشترین سود را کنند.



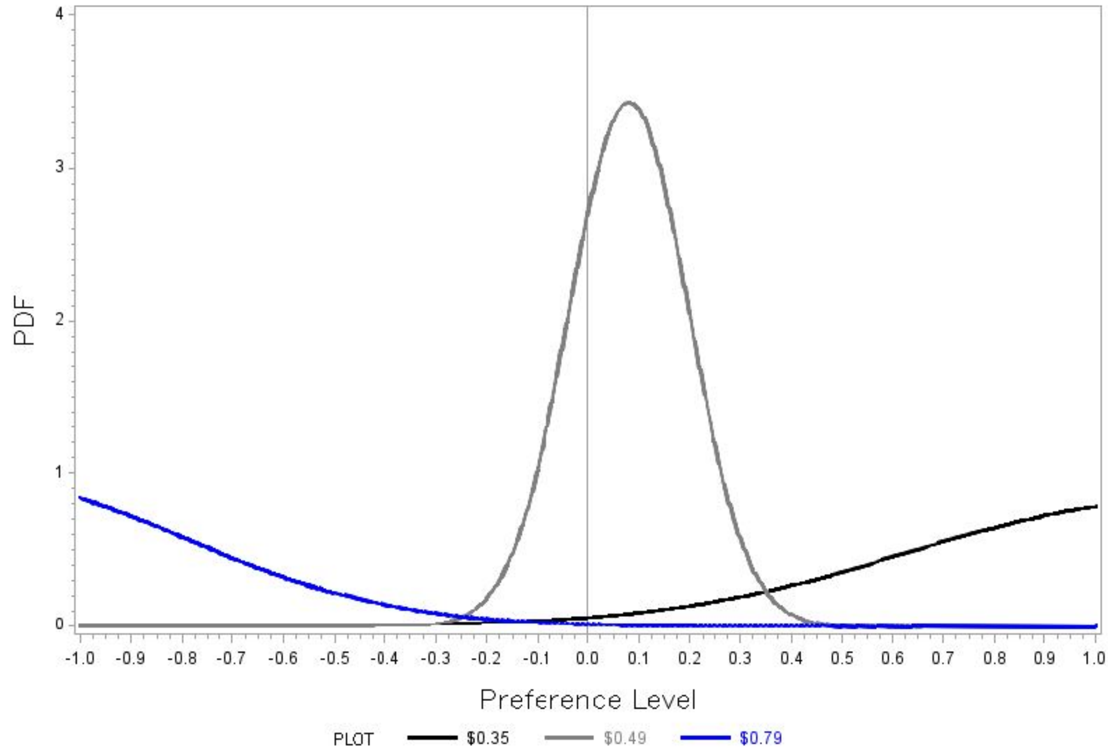
شکل ۶.۴: ضدعفونی کنندگی

وقتی در مورد میزان ضدعفونی کنندگی ماده شوینده تحقیق می‌کنیم، واضح است که این ویژگی تاثیر بزرگی برروی مدل‌های آمیخته خطی دارد. مطابق شکل ۶.۴ اگرچه نسبت بزرگی از پاسخ‌دهندگان ضدعفونی کنندگی محصول برای آن‌ها مهم است، اما شرکت باید چگونگی اهمیت این ویژگی را شرح دهد. این تردید به این دلیل است که آیا در مدل تحلیل متقارن اهمیت این ویژگی بزرگ هست؟ این سوال تاثیری بزرگ روی پیشگویی در مورد این ویژگی خواهد داشت. شرکت باید راه‌های ممکن برای کاهش این تردیدها را ارائه کند.



شکل ۷.۴: زیست‌تخریب پذیری

اگر چه مقدار بخش‌ارزشی برای ویژگی زیست‌تخریب‌پذیری خیلی بالا نیست ۷.۴ وقتی که با دیگر سطوح مقایسه می‌شود دیده می‌شود که پاسخ‌ها در کل ویژگی زیست‌تخریب پذیری ماده شوینده را ترجیح می‌دهند. پراکندگی این سطوح ویژگی بسیار کمتر از پراکندگی ویژگی ضدعفونی‌کنندگی است. این به این معنا است که تعدادی از پاسخ‌دهنده‌ها در کل در مورد مهم بودن ویژگی زیست‌تخریب‌پذیری موافق هستند. اگرچه مشتریان این ویژگی را در هنگام خرید در زمره‌ی اولین ویژگی قرار نمی‌دهند، به آن‌ها پیشنهاد می‌شود که این ویژگی را به عنوان یک فاکتور انتخاب برای مواقعی که محصولات مختلف را مقایسه می‌کنند در نظر بگیرند. چون این ویژگی یکی از مزیت‌های مهم در هنگام تولید محصول برای شرکت‌ها از نظر ایمنی محصول می‌باشد.



شکل ۸.۴: قیمت

ویژگی قیمت یکی از با نفوذترین ویژگی‌ها در تحلیل متقارن است. شکل ۸.۴ نسبت بزرگی از تغییر پاسخ افراد به قیمت را نشان می‌دهد. در مراحل ابتدایی مطالعه تحلیل متقارن ماده شوینده یک قیمت را به عنوان قیمت پایه در نظر می‌گیریم و با توجه به آن مطالعه تحلیل متقارن را شروع می‌کنیم. با دیدن تمیزکننده به طور کامل پاسخ‌دهندگان محصول با قیمت کم با هر کاربرد نسبت به محصول قیمت زیاد و هر کاربرد ترجیح می‌دهند. این انتظار می‌رود که در محیطی که کسب و کار خوب است مشتری حق دارد بهترین محصول با کمترین قیمت را انتخاب کند. عبارت بهترین محصول و کمترین قیمت برای هر فرد دارای تعریف متفاوت می‌باشد. برای سطح قیمت، متوسط و کم بودن قیمت مساله نیست باید یک سطح قیمت معقول نسبت به محصول معرفی کرد.

۸.۴ نتیجه‌گیری و پیشنهادات

با توجه به آنچه که گفته شد سعی بر این بود که تحلیل متقارن را به طور مختصر شرح دهیم. برای این که بتوانیم قابلیت و چگونگی تحلیل آن را مشاهده کنیم مثال‌های متعددی در مدل‌های رگرسیونی با اثرات ثابت، تصادفی و آمیخته مطرح شده در این میان مثال ماده شوینده را مورد بررسی قرار دادیم. با این کار به این نتیجه رسیدیم که یک کالا را می‌توان

با سطح‌های مختلف تولید کرد به گونه‌ای که مورد تایید افراد جامعه باشد در مقابل تولید یک نوع از آن کالا اگر مورد تایید نبود می‌تواند از تولید خارج شود. با نشان دادن سطوح ویژگی‌هایی که برای مردم اهمیت دارد ضمانت برای مقبولیت آن کالا تضمین می‌شود.

در ادامه به منظور پیشنهادی که می‌توان در تحقیقات آینده به بررسی آن پرداخت روش بیز سلسله مراتبی است. زمانی از آن استفاده می‌شود که برخی دشواری‌ها مانند پیچیدگی محاسبات و هزینه اجرا وجود داسته باشد. یا زمانی که جمع‌آوری داده‌ها به راحتی امکان‌پذیر نیست مثلا زمانی را در نظر بگیرید که پاسخ‌دهندگان نتوانند به طور منطقی ارزیابی‌های منطقی برای همه سطوح ویژگی تعیین کنند (کوآل‌تریکس، ۲۰۱۱). مزیت دیگر روش بیز سلسله مراتبی برای زمانی است که بین پاسخ‌دهندگان ناهمگونی وجود دارد روش تحلیل متقارن بیز سلسله مراتبی به طور معنی‌داری دقت روش‌های کلاسیک را افزایش می‌دهد. همچنین از آن‌جایی که در انتخاب سطوح توسط پاسخ‌دهندگان می‌توان عدم قطعیت را نیز لحاظ کرد می‌توان تحلیل متقارن از دیدگاه فازی نیز بررسی نمود.

و به عنوان آخرین پیشنهاد، همانطور که مشاهده کردیم در همه حالت‌هایی که از برآوردهای ML استفاده کردیم فرض نرمال بودن به داده‌ها تحمیل شد که در عمل ممکن است کاملا درست نباشد و به‌جای آن می‌توان از توزیع‌های دم پهن و چوله، مثلا tــ استیودنت چوله استفاده کرد.

مراجع

[۱] نیرومند، ح. ع. (۱۳۸۷) تحلیل رگرسیون خطی ابزاری برای تحقیق، دانشگاه فردوسی مشهد

[2] Allison, P.D. and Christakis, N.A. (1994). Logit models for sets of ranked items, *Journal of Sociological Methodology*, **24**, 199-228.

[3] Bockenholt, U. (2001). Mixed-effects analysis of rank-ordered data, *Journal of Psychometrika*, **66**(1), 45-62.

[4] Bak, A and Bart lomowicz, T. (2014) Conjoint analysis method and its implementation in conjoint R package, *Wroc law University of Economics*, Department of Econometrics and Computer Science

[5] Bak A. (2004), *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych [Decompositional preference measurement methods in marketing research]*. Wyd. AE we Wroc lawiu, Wroc law.

[6] Bak A. (2009), *Analiza Conjoint [Conjoint Analysis]*, [In:] Walesiak M., Gatnar E. (Eds.), *Statystyczna analiza danych z wykorzystaniem programu R [Statistical Data Analysis using R]*, Wydawnictwo Naukowe PWN, Warszawa.

[7] Christensen, R. (2011). *Bayesian Ideas and Data Analysis: An Intoduction for Scientists and Statisticians*, CRC Press.

[8] Cronje, T, (2014). A Mixed Model Approach to Conjoint Analysis. Msc. Thesis, University of Pretoria, South Africa

[9] Corbeil, R.R. and Searle, S.R. (1979), Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model, *Journal of Technometrics*, **18**(1), 31-38.

[10] Cunningham, C.E., Deal, K., and Chen, Y. (2010). *Adaptive Choice-Based Conjoint Analysis: A new patient-centered approach to the assessment of health service preference*, Review article, Patient 2010: 3(4), 257-273

[11] Damaraju, R., Wiley, J.B., and Chitturi, P. (2011). *Choice-Based Conjoint Analysis: Models and Designs*, CRC Press.

[12] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*. Series B (Methodological), **39**(1), 1-38.

[13] DeSarbo, W.S. (2006). *The Handbook of Market Research: Uses, Misuses and Future Advances. Chapter 19 Latent Structure Regression*. SAGE Publications, Inc.

[14] Douglas, C.J. (1969). *Categorical Conjoint Measurement* , paper presented at the annual meeting of the association of Mathematical Psychology, Ann Arbor, Michigan.

[15] Draper, N.R and Smith, H. (1998). *Applied Regression Analysis*, 3rd ed., Wiley Series in Probability and Statistics.

[16] Elmore, J.R., Heymann, H., Johnson, J., and Hewett, J.E. (1999). Preference mapping: relating acceptance of ”creaminess” to a descriptive sensory map of semi-solid, *Journal of Food Quality and Preference*. **10**, 465-475.

[17] Fichet, B., Verde, R., Piccolo, D, Vichi, M. (2011). *Classification andMultivariate Analysis of Complex Data Structures*, Studies in Classification, Data Analysis, and Knowledge Organizations, Springer Verlag, Berlin Heidelberg.

[18] Fruhwirth-Schnatter, S. and Otter, T. (1999). Conjoint Analysis Using Mixed Effect Models, Department of Statistics and Mathematics, WU Vienna Univer-sity of Economics and Business, Vienna. http://epub.wu.ac.at/806/1/document.pdf

[19] Green, P.E., Krieger, A.M., and Wind, Y. (2001). Thirty Years of Conjoint Analysis: Reflections and Prospects, *Journal of Marketing Engineering* **31**(3), S56-S73.

[20] Green, P.E. and Schaffer, C.M. (1991). *Importance Weight Effects On Self-explicated Preference Models: Some Empirical Findings* Advances in Consumer Research, Provo, UT, Association for Consumer Research, 476-482

[21] Green, P.E. and Shrinivasan, V. (1978). Conjoint analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research*, **5**(2), 103-123.

[22] Grover, R. and Vries, M. (2006). *The Handbook of Market Research: Uses, Misuses and Future Advances. Chapter 15 Conjoint Analysis: Understanding Consumer Desicion-Making*, SAGE Publications, Inc.

[23] Gustafsonn, A., Herrmann, A. and Huber, F. (2007). *Conjoint Measurement: Methods and Applications*, 4th ed. Springer Verlag.

[24] Hair, J.F., Black, B., Babin, B., Anderson, R.E., and Tatham, R.L. (2006). *Multivariate Data Analysis*, 6th edn., Pearson Prentice Hall.

[25] Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods*, Springer Verlag.

[26] Isaacson, B. and Lesnick, D. (2012). MaxDiff vs Conjoint: Which is better to measure consumer preference?, MMR Strategy Group, viewed 24 June 2013, from http://www.mmrstrategy.com/wp-content/uploads/ 2012/08/MaxDiff-vs-Conjoint-Which-is- Better-to-Measure-Consumer-Preferences.pdf

[27] Jensen, W.A., Jeffrey, B.B., and Woodall, W.H. (2007). Profile Monitoring via Linear Mixed Models, Virginia Polytechnic Institute and State University, Blacksburg. http://www.stat.vt.edu/research/$Technical_Reports$/TechReport06-2.pdf

[28] Johnson, R.A. and Wichern, D.W., (2007). *Applied Multivariate Statistical Analysis*, 6th ed., Pearson International Edition.

[29] Khattree, R., and Naik, D.N. (2003). Applied Multivariate Statistics with SAS Software, 2nd. edn., SAS Institute Inc., Cary, NC, USA.

[30] Kleinbaum, D.G., Lawrence, L.K., Nizam, A., and Muller, K.E. (2008). *Applied Regression Analysis and other Multivariate Methods*, Thomson Learning, Inc.

[31] Kruskal, J. B. (1965) Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data, *Journal of the Royal Statistical Society, Series B*, **27**, 251-63.

[32] Lasaffre, E. and Lawson, A.B. (2012). 'Bayesian Biostatistics', John Wiley and Sons, Ltd.

[33] Lindstrom, M.J. and Douglas, M.B. (1988). Newton-Rhapson and EM Algorithms for Linear Mixed-effect Models for Repeated- Measures Data, *Journal of the American Statistical Association*, **83**(404), 1014-1022.

[34] Louviere, J.J., Hensher, D.A., Swait, J.D., and Adamonice, W. (2000). *Stated Choice Methods: Analysis and Application*, Cambridge University Press.

[35] Madsen, H. and Thyregod, P. (2011). *Introduction to General and Generalized Linear Models*, CRC Press, Taylor and Francis Group.

[36] Maydeu-Olivares, A. and Böckenholt, U. (2009). *Modeling Preference Data, Chapter 12 The SAGE Handbook of Quantitative Methods in Psychology*. SAGE Publications Ltd.

[37] McCulloch, C.E., Searle, S.R. and Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed., Wiley.

[38] McCullagh P, and Nelder J (1989). *Generalized Linear Models.* second edition, Chapman and Hall/ CRC, Boca Raton.

[39] Meng, X. and Van Dyk, D. (1997) The EM Algorithm — An Old Folk-Song Sung to a Fast New Tune, *Journal of the Royal Statistical Society*. Series B (Methodological), **59**(3), 511-567.

[40] Montgomery, D.C., Peck, E.A., and Vining, G.G. (2012). *Introduction to Linear Regression Analysis*. 5e. Hoboken, NJ: Wiley.

[41] Netzer, O., Toubia, O., Bradlow, E.T., Dahan, E., Evgeniou, T., Feinberg, F.M., Feit, E.M., Hui, S.K., Johnson, J., Liechty, J.C., Orlin, J.B., Rao, V.R (2008). *Beyond conjoint analysis: Advances in preference measurements*. Springer Science + BusinessMedia, LLC 2008.

[42] Pinheiro, J.C. and Bates, D.M. (2004). *Mixed- Effects Models in S and S-PLUS*, Springer Verlag.

[43] Qualtrics (2011). Conjoint Analysis: Explaining Full Profile and Self Explicated Approaches, viewed May 2013, from http://www.qualtrics.com/docs/ ConjointOverview.pdf

R Development Core Team, R:*A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*, URL: http://cran.r-project.org/(2011).

[44] Roodman, G.M. (1974). A Procedure for Optimal Stepwise MSAE Regression Analysis, *Journal of Operational Research*, **22**(0), 393-399.

[45] Wittink, D and Cattin, P. (1989). Commercial use of conjoint analysis: An update, *Journal of Marketing*, **53**(3), 91-96.

[46] Wittink, D., Vriens, M., and Burhenne, W. (1994). Commercial use of conjoint in Europe: Results and critical reflections, International *Journal of Research in Marketing*, **11**(1), 41-52.

[47] Wu, L. (2010). *Mixed Effect Models for Complex Data*, CRC Press. Taylor and Francis Group. [Yan & Su, 2009] Yan, X. and Su, X.G. (2009). *Linear Regression Analysis: Theory and Computing*,World Science Publishing Co. Pte. Ltd

پیوست آ

برنامه‌های SAS مربوط به ماده شوینده

```
proc iml;
use rr.x;
read all into X;
use rr.y;
read all into y;
```

برآورد مقدار بخش‌ـ ارزشی

```
Beta = inv(X‘*X)*X‘*Y;
A = mean(beta‘);*global means - Aggregate model;
Estimated_fixed_effects = round(A‘,0.01);
B = var(beta‘);
Estimated_variance = round(B‘,0.01);
print "Traditional Conjoint Approach to Industrial Cleaner Example";
print Estimated_fixed_effects Estimated_variance; *Aggregate model;
print Beta;
*Determine R2 for all respondents with own beta’s;
yhat = (X*beta);
```

```
/*Stack for 86 respondents*/
Do i = 1 to ncol(Y);
YStack = YStack//Y[,i];
yhatStack = YhatStack//yhat[,i];
end;
/* End of Stack Process*/
SSR = sum((Ystack-yhatstack)##2);
SST = sum((Ystack-mean(Ystack))##2);
R2_Model_ownbetas = 1-SSR/SST;
print R2_Model_ownbetas;
YhatAgg = X*A‘;
/*Stack for 86 respondents*/
Do i = 1 to ncol(Y);
yhatAggStack = YhatAggStack//yhatAgg;
end;
SSRAgg = sum((Ystack-yhataggstack)##2);
SSTAgg = sum((Ystack-mean(Ystack))##2);
R2_Model_avebeta = 1-SSRAgg/SSTAgg;
print R2_Model_avebeta;
```

تحلیل متقارن برای پاسخ‌دهنده اول

```
yhat1 = (X*beta[,1]);
Y1 = Y[,1];
print Y1 yhat1;
SSR = sum((Y1-yhat1)##2);
SST = sum((Y1-mean(Y1))##2);
R2_resp1_Model = 1-SSR/SST;
print R2_resp1_Model;
```

ارزیابی اهمیت ویژگی‌ها برای نفر اول

```
Range_Atr_1 = 0.22+0.17;
Range_Atr_2 = 0.35+0.33;
Range_Atr_3 = 0.51+0.51;
Range_Atr_4 = 0.15+0.15;
```

```
Range_Atr_5 = 1.13+1.21;
Total_Range = Range_Atr_1+Range_Atr_2+Range_Atr_3
+Range_Atr_4+Range_Atr_5;
Range = (Range_Atr_1// Range_Atr_2// Range_Atr_3//
Range_Atr_4// Range_Atr_5)||(Total_Range*J(5,1,1));
Name2 = {'Product Form' 'Number of Applications'
'Disinfectant' 'Biodegradable' 'Price'};
Overall_Importance = round((Range[,1]/Range[,2])*100,0.01);
Title "The Overall Importance of each Attribute";
print Overall_Importance[rowname = name2];
Over_data = {1,2,3,4,5}||Overall_Importance;
name3 = {'Attribute' 'Overall_Importance'};
Create Over from Over_data[colname = name3];
append from Over_data;
```

پیشگویی برای ۱۸ مشخصه

```
Yhat = X*Estimated_fixed_effects;
rounded_yhat = round(yhat);
Print 'Estimated Profile Utilities ' Yhat rounded_yhat;
quit;
```

برنامه‌های مربوط به مثال درخواست شغل

```
Data Random_example;
input rating officer;
cards;
76 1
65 1
85 1
74 1
59 2
75 2
81 2
67 2
49 3
```

```
63 3
61 3
46 3
74 4
71 4
85 4
89 4
66 5
84 5
80 5
79 5
;
goptions reset = all;
symbol1 color = lightpurple value = dot w = 2 h = 2;
title h = 3 f = simplex 'Graphical Representation of the Data';
axis1 lable = (a = 90 f = simplex h = 1.75);
axis2 lable = ( f = simplex h = 1.75) order = (0 to 6 by 1);
proc gplot data = random_example;
plot rating*officer/ vaxis = axis1 haxis = axis2;
run;
title;
proc means data = random_example;
output out = Averages mean = ave;
var rating;
by officer;
run;
goptions reset = all i = join;
title h = 4 f = simplex 'Graphical Representation of the Means';
symbol1 color = lightgreen value = dot line = 3 h = 2 w = 2;
axis1 label = (a=90 f = simplex h = 1.75 'Average rating');
axis2 label = (f = simplex h = 1.75 'Officer') order = (0 to 6 by 1);
proc gplot data = Averages;
plot ave*officer/vaxis = axis1 haxis = axis2;
run;
```

```
title;
proc iml;
use random_example;
read all into data;
Y1 = 0;
Y2 = 0;
Y3 = 0;
Y4 = 0;
Y5 = 0;
Do i = 1 to nrow(data);
if data[i,2] = 1 then Y1 = Y1//data[i,1];
if data[i,2] = 2 then Y2 = Y2//data[i,1];
if data[i,2] = 3 then Y3 = Y3//data[i,1];
if data[i,2] = 4 then Y4 = Y4//data[i,1];
if data[i,2] = 5 then Y5 = Y5//data[i,1];
end;
Y1 = Y1[2:5,];
Y2 = Y2[2:5,];
Y3 = Y3[2:5,];
Y4 = Y4[2:5,];
Y5 = Y5[2:5,];
YY = Y1||Y2||Y3||Y4||Y5;
k = ncol(YY); *number of groups;
n = nrow(Y1);
NN = nrow(data);
GroupMean = mean(YY);
OverallMean = groupMean[,+]/k;
SE = (YY - GroupMean#J(nrow(YY), ncol(YY),1))##2;
SEE = SE[+,];
SSE = SEE[,+];
print 'The SSE is: ' SSE;
MSE = SSE/(NN-k);
print 'The MSE is: ' MSE;
SB = n#((GroupMean -
```

```
OverallMean*J(nrow(groupmean),ncol(groupmean),1))##2);
SSB = sum(SB);
print 'The SSB is: ' SSB;
MSB = SSB/(k-1);
print 'The MSB is: ' MSB;
*Estimated parameter;
mu_hat = overallmean;
print 'The estimated mean is: ' mu_hat;
SigE2_hat = MSE;
print 'The estimated error variance is: ' SigE2_hat;
SumSum = k*((n**2)/NN);
n0 = (NN - SumSum)/(k-1);
SigR2_hat = (MSB - MSE)/n0;
print 'The estimated random variables variance is: ' SigR2_hat;
*Calculate the test statistic;
F = MSB/MSE;
print 'The F test statistic is: ' F;
F_crit = Finv(0.05,NN-k,k-1);
print 'The F critical value is: ' F_crit;
if F > F_crit then print 'The Null Hypothesis will be rejected';
else print 'The Null Hypothesis will not be rejcted';
quit;
```

برنامه‌های مربوط به مثال میزان مصرف کلسترول

روش **EM**

```
Diet mixed example - Using Em with ML;
proc iml;
use rr.diet;
read all into Dat;
Age = Dat[,1];
chol = Dat[,2];
Gender = Dat[,3];
sub = Dat[,4];
n = nrow(Chol);
```

```
p = 2;
X = J(n,1,1)||Age||Gender;
Z = sub;
Y = Chol;
q = ncol(Z);
*Startin values;
Beta = J(ncol(X),1,1);
SigE = 10;
SigD = 150;
count = 0;
k = 0; *number of itterations;
Do until (count = 1);
V = Z*Z`*(SigD) +SigE*I(n);
r = Y-X*beta;
Beta_k = beta + SigE#(inv(X`*X)*X`*inv(V)*r);
SigE_k = SigE +((SigE**2)/n)*
(r`*inv(v)*I(n)*I(n)`*inv(V)*r - trace(I(n)`*inv(v)*I(n)));
SigD_k = SigD +((SigD**2)/n)*
(r`*inv(v)*Z*Z`*inv(V)*r - trace(Z`*inv(v)*Z));
*Test for convergence;
TestOld = Beta//SigE//SigD;
TestNew = Beta_k//SigE_K//SigD_K;
Test = abs(TestOld - TestNew);
Sum = Test[+];
if sum < 0.1 then count = 1;
else count = 0;
*reset values;
beta = beta_k;
SigE = SigE_k;
SE = sqrt(SigE);
SigD = SigD_k;
SD = Sqrt(SigD);
k = k+1;
if k = 300 then count = 1;
```

```
end;
print k beta SigE SE SigD SD;
random_effect = sigD*Z‘*inv(V)*(Y-X*Beta);
print random_effect;
quit;
```

روش **NR**

```
*Diet mixed example - Using NR with ML;
title ’Mixed Model Diet Example - ML method’;
proc mixed data = rr.diet method = ml;
class subsidiary;
model chol = age gender/ s;
random subsidiary/s;
run;
title;
```

مدل اثرات آمیخته

مدل ۱

```
proc mixed data = cleanerDataMix covtest method = ml;
model Y = X1 X2 X3 X4 X5 X6 X7 X8/ solution outp=ind_predict
outpm=mean_predict;
random intercept X1 X2 X3 X4 X5 X6 X7 X8/ subject=resp
type=cs solution G;
ods output SolutionF=solModel SolutionR = REffects;
run;
```

مدل ۲

```
proc mixed data = cleanerDataMix covtest method = ml;
model Y = X1 X2 X3 X4 X5 X6 X7 X8/ solution outp=ind_predict
outpm=mean_predict;
random intercept X1 X2 X3 X4 X5 X6 X7 X8/ subject=resp
type=vc solution G;
ods output SolutionF=solModel SolutionR = REffects;
run;
```

مدل ۳

```
proc mixed data = cleanerDataMix method = ml covtest ratio;
model Y = X1 X2 X3 X4 X5 X6 X7 X8/ solution outp=ind_predict
outpm=mean_predict;
random intercept X1 X2 X3 X4 X5 X6 X7 X8/ subject=resp
type=un solution G V ;
ods output SolutionF=solModel SolutionR = REffects;
run;
```

آ.۱ برنامه‌های R مربوط به داده‌های چای

```
library(conjoint)
data(tea)
caModel(y=tprefm[1,], x=tprof)
caUtilities(y=tprefm[1,], x=tprof, z=tlevn)
Conjoint(y=tpref, x=tprof, z=tlevn)
caImportance(y=tpref, x=tprof)
```

Aabstract

Conjoint analysis is a method for evaluating individual decisions and choices (specially in production), though by studying them, be able to identify the most popular choice. In this thesis, we describe the basic concepts of conjoint analysis and represent its common models with series of objective examples. The calculations related to the estimation of parameters of the models are covered by using two softwares, SAS and R.

key words: Conjoint analysis, Preference utility model, Part-Worth function model, Fixed effectes model, Mixed effectes model

Shahrood University of Technology

Faculty of Mathematical Sciences

MSc Thesis in: Mathematical Statistics

Statistical Conjoint Analysis

By: Zeinolabedin Faramarzi Palangar

Supervisor

Dr. Mohammad Arashi

Janvay 2019