دانشکده علوم ریاضی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد

آمار ریاضی

بررسی رفتار برآوردگرهای رگرسیونی ریج در مدل‌های خطی منفرد

نگارش

سمانه نجاریان

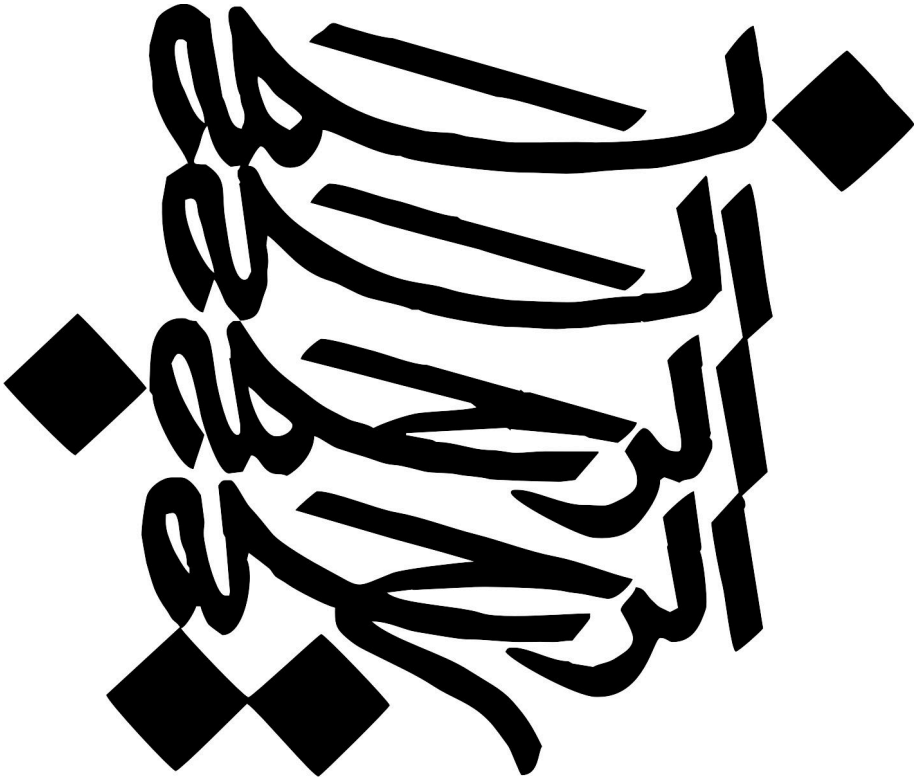
استاد راهنما

دکتر محمد آرشی

استاد مشاور

دکتر جعفر فتحعلی

شهریور ۱۳۹۰



تقدیم به:

همه آنانی که

به من اندیشیدن آموختند

تا امروز

کورسوی ظلمت را از روشنایی معرفت

بازشناسم.

فهرست مقالات مستخرج از متن پایان‌نامه

[۱] نجاریان. س. و آرشی. م. (۱۳۸۹)، شبیه‌سازی در مدل رگرسیون خطی ریج منفرد، ششمین همایش ملی آمار- دانشگاه پیام‌نور اهواز.

[2] Najarian. S. and Arashi. M. (2010), Ridge Estimation In Singular Linear Model: New Approch, The 10th Iranian Statistical Conference.

[3] Najarian. S., Arashi. M. and Golam Kibria. B. M. (2011), A Simulation Study On Some Ridge Regression Estimators, Communication in statistics- simulation Restricted and Computation.

پیشگفتار

در دهه‌های اخیر، تلاش‌های زیادی در راستای یافتن برآوردگرهای بهبودیافته انجام گرفته است. یکی از مسائلی که زمینه‌ساز این تلاش بوده، همخطی چندگانه می‌باشد. وقتی‌که ستون‌های ماتریس طرح وابسته خطی باشند، برآوردگر کمترین توان دوم بدست آمده به روش گوس- مارکف به دلیل نزدیک به صفر بودن دترمینان ماتریس $X'X$، کارایی لازم را نخواهند داشت. این مطلب منجر به ارائه برآوردگرهای بهبودیافته از جمله برآوردگر نوع ریج شده است. هدف از این پایان‌نامه، ارائه برآوردگر ریج در مدل‌های خطی محدودشده و محدودنشده و تحلیل و تفسیر ویژگی‌های برآوردگر ارائه‌شده، است. این مجموعه شامل ۵ فصل و ۴ ضمیمه می‌باشد. مطالب هر فصل بطور مختصر عبارتند از:

در فصل ۱، مقدمه‌ای بر تحلیل رگرسیونی، ارائه و چگونگی برآورد ضرایب رگرسیونی و همچنین چگونگی بروز همخطی در مدل‌های رگرسیون چندگانه خطی و روش‌های تشخیص همخطی بررسی شده است.

در فصل ۲، برآوردگر رگرسیونی ریج در مدل خطی محدودشده مورد بررسی قرارگرفته و خطای برآوردگر، محاسبه شده‌است و به کمک شبیه سازی و مثال عددی کارایی برآوردگر و برتریش بر برآوردگر کمترین توان‌های دوم محدودشده نشان داده شده است.

در فصل ۳، برآوردگر رگرسیونی ریج محدودشده و محدودنشده در مدل خطی منفرد ارائه‌شده و ویژگی‌های برآوردگر بررسی شده است.

در فصل ۴، برآوردگر رگرسیونی ریج در مدل خطی منفرد با در نظر گرفتن محدودیتی که دارای جمله خطای

تصادفی است، ارائه و ویژگی‌های برآوردگر بررسی و به کمک شبیه‌سازی کارایی برآوردگر تایید شده است.

در فصل ۵، برآوردگر رگرسیونی ریج منفرد در مدل به ظاهر نامرتبط (SUR) ارائه و ویژگی‌های برآوردگر بطور خلاصه بیان شده است.

در ضمیمه A، بحث و نتیجه‌گیری و پیشنهادها برای ادامه کار در آینده، مطرح شده‌اند.

در ضمیمه B، عملیات جبری ماتریس‌ها بمنظور سادگی درک اثبات قضایا و لم‌ها در متن پایان‌نامه بیان شده است.

در ضمیمه C، توضیحاتی در خصوص عامل تورم واریانس به عنوان عاملی برای شناخت همخطی در مدل رگرسیونی خطی داده شده است.

در ضمیمه D، توضیحات کاملتری درباره عدد شرطی که عاملی دیگر برای تشخیص مسئله همخطی است، ارائه شده است.

* در این مجموعه، قضایایی که برهان آن از نویسنده این پایان‌نامه می‌باشد با علامت ستاره مشخص شده است.

مدلهای ارائه شده در این مجموعه

فصل دوم	$$Y = X\beta + E$$ $$R\beta = r$$ $$E(E) = 0 \qquad Cov(E) = \sigma^2 I$$ $$Y \sim n \times 1 \quad X \sim n \times p \quad \beta \sim p \times 1$$ $$E \sim n \times 1 \quad R \sim q \times p \quad r \sim q \times 1$$
فصل سوم	$$Y = X\beta + E$$ $$R\beta = r$$ $$E(E) = 0 \qquad Cov(E) = \sigma^2 \Sigma$$ $$Y \sim n \times 1 \quad X \sim n \times p \quad \beta \sim p \times 1$$ $$E \sim n \times 1 \quad R \sim q \times p \quad r \sim q \times 1$$
فصل چهارم	$$Y = X\beta + E$$ $$R\beta + \nu = r$$ $$E(E) = 0 \qquad Cov(E) = \sigma^2 I$$ $$Y \sim n \times 1 \quad X \sim n \times p \quad \beta \sim p \times 1$$ $$E \sim n \times 1 \qquad \nu \sim q \times 1$$ $$R \sim q \times p \qquad r \sim q \times 1$$
فصل پنجم	$$Y = X\beta + E$$ $$R\beta = r$$ $$E(E) = 0 \qquad Cov(E) = \Sigma \otimes I_q$$ $$Y \sim n \times 1 \quad X \sim n \times k \quad \beta \sim k \times 1$$ $$E \sim n \times 1 \quad R \sim q \times p \quad r \sim q \times 1$$ $$k = \sum_{i=1}^{p} k_i \qquad n = pq$$

قدردانی

منت خدای را عزوجل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت

در پایان لازم می‌دانم از استاد گرامی، جناب آقای دکتر آرشی که راهنمایی و هدایت این پایان‌نامه را عهده‌دار بودند و تلاش زیادی در راستای جهت‌دهی به تفکرات و تلاش‌های اینجانب داشته‌اند، قدردانی کنم. همچنین از جناب آقای دکتر فتحعلی جهت مشاوره و راهنمایی‌های بی‌دریغشان کمال تشکر را دارم.

از اساتید داور محترم، آقایان دکتر داوود شاهسونی (دانشگاه صنعتی شاهرود) و دکتر مهدی عمادی (دانشگاه فردوسی مشهد) که با حضور گرمشان تصحیح و داوری پایان‌نامه را بر عهده گرفتند، سپاسگزارم.

از جناب آقای پروفسور کیبریا (دانشگاه فلوریدای آمریکا) بخاطر راهنمایی‌هایشان در استخراج مقاله‌ای از فصل یک این پایان‌نامه، تشکر می‌کنم.

از جناب آقای دکتر فکور و آقای دکتر روزبه بخاطر همکاریشان با بنده کمال تشکر را دارم.

همچنین از مسئولین کتابخانه دکتر فاطمی دانشگاه فردوسی مشهد بخصوص آقای مسئله‌گو یخاطر همکاری صمیمانه‌شان متشکرم.

سمانه نجاریان

تابستان ۱۳۹۰

علائم اختصاری

MSE : میانگین توان دوم خطا . . . . . *Mean Square Error*

$MSEM$: ماتریس میانگین توان دوم خطا . . . . . *Mean Square Error Matrix*

LSE : برآوردگر کمترین توان‌های دوم . . . . . *Least Squares Estimator*

$RLSE$: برآوردگر کمترین توان‌های دوم محدودشده . . . . . *Restricted Least Squares Estimator*

$GLSE$: برآوردگر کمترین توان‌های دوم محدودشده . . . . *Generalized Least Squares Estimator*

RE : برآوردگر ریج . . . . . *Ridge Estimator*

RRE : برآوردگر ریج محدودشده . . . . . *Restricted Ridge Estimstor*

RLM : مدل خطی محدودشده . . . . . *Restricted Linear Model*

SE : برآوردگر استاین . . . . . *Stein Estimator*

LE : برآوردگر لویی . . . . . *Lui Estimator*

PCE : برآوردگر مولفه‌های اصلی . . . . . *Principal Components Estimator*

$PLSE$: برآوردگر کمترین توان‌های دوم جزئی . . . . . *Partial Least Squares Estimator*

AUE : برآوردگر تقریبا نااریب . . . . . *Almost Unbaised Estimator*

$BLUE$: بهترین برآوردگر نااریب خطی . . . . . *Best Linear Unbaised Estimator*

SUR : مدل به ظاهر نامرتبط . . . . . *Seemingly Unrelated Regression*

AN : نرمال مجانبی ..... $Asymptotic\ Normal$

VIF : عامل تورم واریانس..... $Variance\ Infeaation\ Factor$

CN : عدد شرطی ..... $Condition\ Number$

فهرست مطالب

فصل ۱

مقدمه و دورنما

مقدمه

تحلیل رگرسیونی یکی از روش‌های آماری برای تحلیل داده‌های چندعاملی است که حوزه کاربرد آن بیشترین وسعت را دارد، نتایج پرجاذبه آن از نظر مفهومی، فرآیند ساده بکارگیری یک معادله است که ارتباط بین دسته‌ای از متغیرها را بیان می‌کند. تحلیل رگرسیونی همچنین از جهت نظری به لحاظ ظرافت و زیبایی ریاضیات دارای جذابیت است. موفقیت در تحلیل، نیاز به درک و تیزبینی در دو مقوله تئوری و مسائل عملی دارد و زمانی بروز می‌کند و خود را نشان می‌دهد که تکنیک در بکارگریری داده‌ها و اطلاعات جهان واقع قرار می‌گیرد.

تحلیل رگرسیونی فن و تکنیکی آماری برای بررسی و مدل‌بندی ارتباط بین متغیرهاست. کاربردهای رگرسیون متعدد است و تقریبا در هر زمینه‌ای از جمله مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی و بیولوژی و علوم اجتماعی صورت می‌پذیرد.

کلمه "رگرسیون" در لغت به معنای "برگشت" است. دلیل این نامگذاری به مطالعات "گالتون" در مورد رابطه قد پسر و قد پدر مربوط می‌شود. گالتون با رسم نمودار قد پسر در مقابل قد پدر (نمودار پراکنش) براساس حجم بزرگی از داده‌ها به این نتیجه رسید که اغلب افراد بلندقد، پسرانی کوتاه‌قد و پدارن کوتاه‌قد، پسرانی بلندقدتر از خود دارند و این پدیده را "برگشت به میانگین" نامید و از آن پس نام رگرسیون بر روش‌های بررسی روابط بین متغیرهای آماری باقی ماند.

به عنوان مثالی از یک مسئله که در آن تحلیل رگرسیونی می‌تواند مفید واقع شود، فرض می‌کنیم یک مهندس صنایع، توسط یک سازنده نوشابه استخدام شده که محصول تحویلی و عملیات سرویس و خدمت رسانی برای ماشین‌های فروش را تجزیه و تحلیل کند. او گمان می‌کند که زمان لازم برای اینکه یک فرد تحویل دهنده، این ماشین را سرویس دهد بستگی به تعداد موارد محصول تحویل شده دارد. مهندس ۲۵ فروشنده جزئی را که دارای ماشین فروش می‌باشند، به تصادف بازدید کرده است و زمان تحویل (برحسب دقیقه) و حجم محصول تحویل داده شده (بر حسب مورد) برای هر یک مشاهده می‌شود. بوسیله نمودار پراکنش، ارتباط بین

زمان تحویل و حجم تحویل مشخص می‌شود. نقاطی که نشان‌دهنده داده‌ها هستند در امتداد یک خط مستقیم قرار می‌گیرند، نه دقیقا روی خط. اگر $\boldsymbol{Y}$ نشان دهنده زمان تحویل و $\boldsymbol{X}$ را برای حجم تحویل شده قرار دهیم، در این صورت معادله یک خط مستقیم که این دو متغیر را به هم مربوط می‌سازد، چنین است

$$\boldsymbol{Y} = \boldsymbol{\beta}_۰ + \boldsymbol{\beta}_۱ \boldsymbol{X} \tag{۱.۱}$$

که $\boldsymbol{\beta}_۰$ عرض از مبدا و $\boldsymbol{\beta}_۱$ شیب خط است. از آنجایی‌که نقاط داده‌ها دقیقا روی یک خط مستقیم قرار نمی‌گیرند بنابراین (۱.۱) بایستی به عنوان برآوردکننده آن اصلاح و تعدیل شود. اگر اختلاف بین مقدار مشاهده‌شده $\boldsymbol{Y}$ و خط مستقیم $(\boldsymbol{\beta}_۰ + \boldsymbol{\beta}_۱ \boldsymbol{X})$ را خطای $\boldsymbol{E}$ قرار می‌دهیم، مناسب است که خطای $\boldsymbol{E}$ به عنوان یک خطای آماری تلقی شود. بدین معنی که متغیری تصادفی است که عدم برازش را بیان و اندازه ناتوانی مدل از برازش دقیق داده‌ها را برآورد می‌کند. این خطا ممکن است به لحاظ اثرات دیگر متغیرها، خطاهای اندازه‌گیری و غیره روی زمان تحویل صورت پذیرد. بنابراین مدلی با قابلیت پذیرش بیشتر برای داده‌های زمان تحویل عبارتست از

$$\boldsymbol{Y} = \boldsymbol{\beta}_۰ + \boldsymbol{\beta}_۱ \boldsymbol{X} + \boldsymbol{E} \tag{۲.۱}$$

معادله (۲.۱) یک مدل رگرسیون خطی نامیده می‌شود. در این مدل $\boldsymbol{X}$ متغیر مستقل و $\boldsymbol{Y}$ متغیر وابسته نامیده می‌شود. در هر حال این نامگذاری باعث می‌شود که این مفهوم با مفهوم استقلال آماری اشتباه شود. چون در (۲.۱) فقط یک متغیر رگرسیونی وجود دارد، این مدل را مدل رگرسیونی ساده نامیده‌اند.

۱.۱ مدل رگرسیونی چندگانه

در حالت کلی ممکن است متغیر پاسخ به k متغیر رگرسیونی $\boldsymbol{X}_۱, \boldsymbol{X}_۲, ..., \boldsymbol{X}_k$ مربوط باشد، بنابراین مدل رگرسیونی به صورت زیر نوشته می‌شود

$$\boldsymbol{Y} = \boldsymbol{\beta}_۰ + \boldsymbol{\beta}_۱ \boldsymbol{X}_۱ + \boldsymbol{\beta}_۲ \boldsymbol{X}_۲ + ... + \boldsymbol{\beta}_k \boldsymbol{X}_k + \boldsymbol{E} \tag{۳.۱}$$

چون در این مدل رگرسیونی بیش از یک متغیر درگیر است، رگرسیونی چندمتغیره نام دارد. پارامترهای $\boldsymbol{\beta}_j \quad (j = ۰, ۱, ..., k)$ ضرایب رگرسیون نامیده می‌شود. بکارگیری صفت خطی بودن در مدل فوق اشاره بر این دارد که مدل، برحسب پارامترهای $\boldsymbol{\beta}_۰, \boldsymbol{\beta}_۱, ..., \boldsymbol{\beta}_k$ خطی است و نه به دلیل اینکه $\boldsymbol{Y}$ تابعی خطی از $\boldsymbol{X}$ هاست. این مدل یک ابرصفحه در فضای kبعدی از متغیرهای رگرسیونی $\boldsymbol{X}_j$ است. پارامتر $\boldsymbol{\beta}_j$ نشان‌دهنده تغییرات مورد انتظار متغیر پاسخ به ازای یک واحد تغییر در $\boldsymbol{X}_j$ است وقتیکه همه متغیرهای رگرسیونی باقیمانده دیگر $\boldsymbol{X}_i(i \neq j)$ ثابت نگهداشته شوند. به همین جهت پارامترهای $\boldsymbol{\beta}_j(j = ۱, ۲, ..., k)$ ضرایب جزئی رگرسیون نامیده می‌شوند. مدل‌های رگرسیون چندگانه، اغلب به عنوان تقریب توابع بکار می‌روند. بدین معنی که ارتباط تابعی واقعی بین $\boldsymbol{X}_۱, \boldsymbol{X}_۲, ..., \boldsymbol{X}_k$ و $\boldsymbol{Y}$ شناخته شده نیست، اما مدل رگرسیون خطی روی دامنه تغییرات متغیرهای رگرسیونی، تقریب مناسبی از ارتباط تابعی مذکور می‌باشد.

۲.۱ برآورد پارامترهای مدل

برای برآورد ضرایب رگرسیون، روش کمترین توان‌های دوم بکار می‌رود. فرض می‌کنیم $n > k$ مشاهده در دسترس است و $\boldsymbol{Y}_i$ نمایش دهنده iامین پاسخ مشاهده شده و $\boldsymbol{X}_{ij}$ نمایش‌دهنده iامین مشاهده در jامین سطح متغیر رگرسیونی باشد. فرض می‌کنیم جمله خطای $\boldsymbol{E}$ در مدل دارای $E(\boldsymbol{E}) = ۰$ و $Var(\boldsymbol{E}) = \sigma^۲$ و خطاها ناهمبسته هستند. می‌توان مدل متناظر با (۳.۱) را بصورت زیر نوشت

$$\begin{cases} \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E} \\ E(\boldsymbol{E}) = ۰ \quad , \quad Cov(\boldsymbol{E}) = \sigma^۲ \boldsymbol{I} \end{cases} \tag{۴.۱}$$

تابع کمترین توان‌های دوم خطا بصورت زیر نوشته می‌شود

$$\begin{aligned}\boldsymbol{S}(\beta_۰, \beta_۱, ..., \beta_k) &= \sum_{i=۱}^{n} E_i^۲ \\ &= \sum_{i=۱}^{n} \left(Y_i - \beta_۰ - \sum_{j=۱}^{n} \beta_j X_{ij} \right)^۲ \\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\end{aligned}$$

با مشتقگیری از $\boldsymbol{S}$ نسبت به $\beta_۰, \beta_۱, ..., \beta_k$ و مساوی با صفر قرار دادن مشتق حاصل، در صورت نامنفرد بودن ماتریس $\boldsymbol{X}'\boldsymbol{X}$، برآوردگرهای ضرایب مدل رگرسیونی را بصورت زیر بدست می‌آوریم

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-۱}\boldsymbol{X}'\boldsymbol{Y} \tag{۵.۱}$$

بردار مقادیر برازش‌شده $\hat{Y}_i$ با مقادیر مشاهده‌شده Y_i چنین خواهدبود

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{Y} \tag{۶.۱}$$

ماتریس $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-۱}\boldsymbol{X}'$، ماتریسی $p \times p$، متقارن و خودتوان است و بنابراین $Rank(\boldsymbol{H}) = tr(\boldsymbol{H}) = p$ که ماتریس هت یا برازش نامیده می‌شود زیرا که بردار مقادیر مشاهدات را به بردار مقادیر برازش‌شده تصویر می‌کند. ماتریس برازش و خواص آن نقشی مرکزی در تحلیل رگرسیونی بازی می‌کند.

ماتریس واریانس- کوواریانس $\hat{\boldsymbol{\beta}}$ عبارتست از

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^۲(\boldsymbol{X}'\boldsymbol{X})^{-۱} \tag{۷.۱}$$

همچنین برآوردگر نااریب $\sigma^۲$ عبارتست از میانگین توان‌های دوم باقیمانده‌ها

$$\hat{\sigma}^۲ = \frac{SSE}{n-p} \tag{۸.۱}$$

که در آن $SSE = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$، مجموع توان‌های دوم باقیمانده‌هاست. حال اگر در مدل (۴.۱)، $Cov(\boldsymbol{E}) = \sigma^2\boldsymbol{\Sigma}$ باشد، که در آن $\boldsymbol{\Sigma}$ ماتریسی معیت مثبت است. برآوردگر ضرایب رگرسیونی و ماتریس واریانس- کوواریانس برآوردگرها و همچنین برآوردگر نااریب σ^2 به ترتیب به صورت زیر خواهندبود

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \tag{۹.۱}$$

و

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \tag{۱۰.۱}$$

و

$$\hat{\sigma}^2 = \frac{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})}{n - p} \tag{۱۱.۱}$$

در این صورت مقادیر برازش‌شده به فرم زیر می‌باشند

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \tag{۱۲.۱}$$

که در آن $\boldsymbol{H} = \boldsymbol{X}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1/2}$.

۳.۱ همخطی در رگرسیون خطی چندگانه

مدل‌های رگرسیونی در حوزه کاربردی وسیعی مورداستفاده قرار می‌گیرند. یک مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، همخطی چندگانه یا وابستگی خطی نزدیک بین متغیرهای رگرسیونی است. وجود وابستگی خطی نزدیک، توانایی برآورد ضرایب مدل رگرسیون را با مشکل مواجه می‌کند.

بکارگیری و تعبیر یک مدل رگرسیون چندگانه بطور ضمنی یا بطور صریح به برآوردهای ضرایب مدل رگرسیون بستگی دارد. بعضی از استنباط‌هایی که بطور معمول در مدل رگرسیونی صورت می‌پذیرد، شامل

موارد زیر است:

۱- مشخص کردن اثرات نسبی متغیرهای رگرسیونی

۲- پیش‌بینی و یا برآورد

۳- انتخاب دسته‌ای مناسب از متغیرها برای مدل

اگر هیچ رابطه خطی بین متغیرهای رگرسیونی نباشد، گفته می‌شود که متغیرهای رگرسیونی مستقلند. اگر متغیرهای رگرسیونی مستقل باشند، استنباط‌هایی مانند مواردی که در بالا ذکرشد، نسبتا بسادگی صورت می‌پذیرد. متاسفانه در بیشتر کابردهای رگرسیون، متغیرهای رگرسیونی مستقل نیستند. هنگامی‌که ارتباط خطی نزدیکی بین متغیرهای رگرسیونی وجوددارد، گفته‌می‌شود مسئله همخطی چندگانه وجوددارد.

چهار منبع اولیه همخطی وجوددارد:

۱- شیوه جمع‌آوری داده‌ها

۲- قیدهای روی مدل یا درون جامعه

۳- نوع مدل

۴- مدلی با متغیرهای رگرسیونی بیش از حد

درک اختلاف بین این منابع همخطی چندگانه اهمیت دارد، به‌طوری‌که توصیه برای تحلیل داده‌ها و تعبیر مدل نتیجه‌شده تا اندازه‌ای به علت مسئله همخطی بستگی دارد. (برای اطلاعات بیشتر در مورد همخطی چندگانه رجوع کنید به میسون، گونست و وبستر[1] (۱۹۷۵).)

[1] Mason, Gunst and Webster