



# دانشگاه شهید چمران اهواز

دانشکده علوم ریاضی و کامپیوتر

گروه آمار

پایان نامه کارشناسی ارشد

عنوان:

آزمون کلی فرض‌های مدل رگرسیونی خطی

نگارنده:

محمد رضا مهاجرنیا

استادان راهنما:

دکتر عبدالرحمن راسخ

دکتر قاسم تارمست

استاد مشاور:

دکتر صادق رضایی

اسفند ۱۳۸۹

عَنْ مُحَمَّدٍ  
عَلَيْهِ السَّلَامُ  
لَا إِلَهَ إِلَّا اللَّهُ

به سپاس و ستایش تو، من سگتہ زبان راجہ امکان زبان کشایی، و این آشفته حال راجہ یارای سخن آرای،  
اما با همین زبان سگتہ می گویم کہ سپاس من ہمہ این است کہ در تنگنای ظلمات خاک، هنوز شعلہ های افلاکی  
اشتیاق و انتظار در من نموده است و گاہ گاہ چون عطری ناگمان در سرتاسر جانم می پیچد و نام تو را آواز  
می کند، ای مہربان دوست داشتنی تو بالطف بی پایان خود سپاس مرا بپذیر.

نخست در برابر کرامی پدر و مادرم کہ بالیدنم تجلی عشق و محبتشان و موفقیت هایم بہ پشتوانہ می دعای خیرشان  
بودہ زانوی ادب می زنم. ہمسرم کہ بہترین دوست و راہنایم و وجودش امید بخش زندگی من است،  
خواہر و برادر خوبم کہ بہترین مشوقان من در زندگی بودند را سپاس گویم.

از اساتید راهنا جناب آقای دکتر راسخ که وجود ایشان، همچون تکیه گاهی استوار، توانایی رستن به من بخشیده و گفتار نغزشان قدم هایم در طی این راه پرفراز و نشیب بوده است. پشتکار و خستگی ناپذیری ایشان در تحصیل علم را سرلوحه خویش قرار می دهم و آرزوی کامیابی و شادکامی در تمام مراحل زندگی برای ایشان از خداوند متعال خواهانم، باشد که همواره بنده را مورد لطف و عنایت خویش قرار دهد، و جناب آقای دکتر تارمست که با سعی صدر یاریم کردند و از استاد گرامی جناب آقای دکتر رضایی استاد مشاور محترم که انسانیت و علم را به من آموختند، سپاس فراوان دارم.

از داوران گرامی جناب آقای دکتر چینی پرداز و جناب آقای دکتر علوی و نیز استاد ناظر محترم، جناب آقای دکتر آخوند و اساتید بزرگوار که در طول مدت تحصیل در خدمتشان بوده و از راهنمایی های ارزنده آن ها بهره مند شدم صمیمانه سپاسگزاری می کنم.

از تک تک دوستان مهربانم که حق دوستی را به خوبی ادا کردند سپاسگزارم.

تقدیم به پدر و مادر عزیزتر از جانم و، همسر مهربانم  
آنمان که گامهایشان، طنین سنگین را جاری ساخت و دست‌هایشان چه روشن،  
تاریکی ام را دور کرد.

پس باد آنچه را بخشد

باغ سبز دلتان پر گل باد

که هر آن چیز که با من همراه است

ریشه از دست شامی گیرد

در برابر وجود کرامتشان زانومی ادب بر زمین میزنم و بادی ملو از

عشق، محبت و خضوع

بر دستشان بوسه می زنم

### چکیده پایان نامه

نام خانوادگی: مهاجرنیا	نام: محمدرضا
عنوان پایان نامه: آزمون کلی فرض های مدل رگرسیونی خطی	
استادان راهنما: دکتر عبدالرحمن راسخ، دکتر قاسم تارمست	
درجه تحصیلی: کارشناسی ارشد	رشته: آمار
محل تحصیل: دانشگاه شهید چمران اهواز	دانشکده: علوم ریاضی و کامپیوتر
تاریخ فارغ التحصیلی: ۸۹/۱۲/۲۱	تعداد صفحات: ۱۰۴
واژه های کلیدی: آماره حذف، آزمون هموارسازی نیمن، تعیین نقاط پرت، اعتبارسنجی مدل، آزمون امتیاز	
چکیده	
<p>یکی از مدل های بسیار مهم در آمار مدل خطی است که در آن نخستین فرض خطی بودن می باشد و بیانگر این است که متغیر پاسخ با متغیرهای مستقل رابطه ای خطی دارد. فرض دوم همگن بودن واریانس خطاهاست که نشان دهنده ی ثابت بودن واریانس متغیرهای تصادفی پاسخ نیز می باشد. سومین فرض مورد نظر ناهمبستگی خطاها (پاسخ ها) می باشد؛ و آخرین فرض این است که توزیع متغیرهای تصادفی پاسخ شرطی شده روی ماتریس متغیرهای مستقل نرمال می باشد. در این رساله، به دست آوردن یک فرآیند کلی و کامل برای آزمودن چهار فرض مدل های خطی مورد توجه قرار گرفته است. آزمون بر پایه ی بردار باقی مانده های استاندارد شده و آزمون هموارسازی نیمن بنا نهاده شده است. اگر فرآیند کلی نقض حداقل یکی از فرض ها را نشان دهد، مولفه های آماره آزمون کلی برای تشخیص فرض یا فرض های نقض شده می توانند به کار گرفته شوند، همچنین از فرآیند می توان برای ساختن آماره حذف مربوطه و تشخیص نقاط غیرعادی استفاده کرد. در انتها مثال کاربردی درآمدهای صادرات غیرنفتی ایران طی سال های ۱۳۵۲ تا ۱۳۸۴ را مورد تجزیه و تحلیل قرار داده و علاوه بر آن صحت برقراری این چهار فرض را در این داده ها به وسیله ی آزمون های سابق و آزمون کلی معرفی شده بررسی می کنیم.</p>	

# فهرست

صفحه	عنوان
۱	۱ کلیات
۲	۱-۱ مقدمه
۳	۲-۱ تاریخچه
۴	۳-۱ دلایل طرح مساله
۵	۴-۱ ساختار پایان نامه
۷	۲ تعاریف و مفاهیم اساسی
۸	۱-۲ مقدمه
۸	۲-۲ مدل رگرسیونی خطی و فرضیه‌های آن
۹	۱-۲-۲ برآورد آماری
۹	۱-۱-۲-۲ خصوصیات برآوردکننده‌ها
۱۰	۲-۱-۲-۲ برآورد کننده‌های حداکثر درست‌نمایی
۱۱	۳-۱-۲-۲ برآورد کننده‌های حداقل مربعات
۱۲	۲-۲-۲ برآورد پارامترهای مدل رگرسیونی خطی
۱۵	۳-۲-۲ باقی مانده‌ها
۱۵	۱-۳-۲-۲ مقدمه
۱۶	۲-۳-۲-۲ باقی مانده‌های بازگشتی
۱۸	۳-۳-۲-۲ باقی مانده‌های استاندارد شده
۱۸	۴-۳-۲-۲ باقی مانده‌های استیودنت شده
۲۰	۵-۳-۲-۲ باقی مانده‌های نرمال شده
۲۱	۳-۲ فرض صفر و فرض مقابل
۲۳	۴-۲ چولگی
۲۴	۵-۲ کشیدگی
۲۵	۶-۲ آزمون هموارسازی
۲۵	۱-۶-۲ تعریف مساله
۲۵	۲-۶-۲ آزمون $X^2$ پیرسن
۲۶	۳-۶-۲ آزمون هموارسازی نیمین

۲۸	۳ معرفی آماره آزمون کلی فرض‌های مدل رگرسیون خطی و آماره‌های جزئی مربوط به آن
۲۹	۱-۳ مقدمه
۲۹	۲-۳ آماره‌های آزمون در یک نگاه
۳۱	۳-۳ معرفی آماره‌های $\hat{K}_1^2$ و $\hat{K}_p^2$ و توزیع آنها
۳۳	۴-۳ معرفی آماره‌ی $\hat{K}_p^2$ و توزیع آن
۳۳	۵-۳ معرفی آماره‌ی $\hat{K}_i^2$ و توزیع آن
۳۵	۶-۳ نتیجه
۳۵	۷-۳ آزمون فرض‌های مدل رگرسیونی خطی
۳۶	۸-۳ نقاط غیرعادی
۳۸	۱-۸-۳ روش مبتنی بر آماره حذف بر اساس آماره کلی $(\Delta \hat{G}_i^2)$
۴۰	۴ بسط نظری روش آزمون کلی فرض‌های مدل رگرسیون خطی
۴۱	۱-۴ مقدمه
۴۱	۲-۴ خاصیت بهینگی نیرومند
۴۲	۳-۴ تأثیر مجانبی جایگذاری برآوردگرهای پارامترها در برخی از انواع آماره‌ها
۴۲	۱-۳-۴ شرایط منظم
۴۲	۲-۳-۴ یافتن توزیع حدی $\hat{T}_n$
۴۳	۳-۳-۴ نتیجه
۴۴	۴-۴ ساختن آماره آزمون کلی بر اساس آزمون هموارسازی نیمین
۴۶	۱-۴-۴ به دست آوردن آماره آزمون امتیاز و توزیع آن تحت $\beta$ و $\sigma^2$ معلوم
۴۸	۱-۱-۴-۴ نتایج برای $n$ های بزرگ (وقتی $n \rightarrow \infty$ )
۴۹	۲-۱-۴-۴ آماره‌ی آزمون و توزیع آن وقتی $\sigma^2$ و $\beta$ معلومند
۵۰	۲-۴-۴ به دست آوردن آماره آزمون امتیاز و توزیع آن تحت $\beta$ و $\sigma^2$ نامعلوم
۵۸	۱-۲-۴-۴ آماره‌ی آزمون و توزیع آن وقتی $\sigma^2$ و $\beta$ نامعلومند
۵۹	۵ شبیه‌سازی
۶۰	۱-۵ مقدمه
۶۰	۲-۵ شبیه‌سازی از مدل واقعی



۶۲	۱-۲-۵ شبیه سازی در حالت نقض فرض نرمال بودن
۶۴	۲-۲-۵ شبیه سازی در حالت نقض فرض هم‌وابستگی
۶۵	۳-۲-۵ شبیه سازی در حالت نقض فرض خطی بودن تابع ربط
۶۶	۴-۲-۵ شبیه سازی در حالت نقض فرض استقلال خطاها
۶۸	۶ بررسی اعتبار فرض‌ها در داده‌های صادرات غیر نفتی ایران
۶۹	۱-۶ مقدمه
۶۹	۲-۶ کالاهای صنعتی
۷۰	۳-۶ توزیع جغرافیایی صادرات غیرنفتی ایران
۷۳	۴-۶ آزمون بررسی فرض‌های مدل رگرسیونی خطی روی داده‌های درآمد حاصل از صادرات نفتی و غیر نفتی طی سال‌های ۱۳۵۲ تا ۱۳۸۴ بر اساس آماره‌ی جدید
۸۱	۵-۶ آزمون‌های تشخیص نقاط غیرعادی به‌روش مبتنی بر آماره حذف بر اساس آماره کلی آزمون $(\Delta \hat{G}_i^* [i])$
۸۵	۶-۶ بررسی فرض‌های مدل رگرسیونی خطی در داده‌های مربوط به درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۳
۹۲	۷-۶ نتیجه‌گیری و پیشنهادها
۹۴	واژه‌نامه
۹۹	منابع
۱۰۴	چکیده‌ی انگلیسی

## فهرست نمودار ها

صفحه	عنوان
۷۷	نمودار ۱-۶ نمودار باقی مانده های استاندارد شده در مقابل مقادیر برازش داده شده
۷۸	نمودار ۲-۶ نمودار باقی مانده های استاندارد شده در مقابل زمان
۷۹	نمودار ۳-۶ نمودار مقادیر متغیر وابسته در مقابل زمان
۸۰	نمودار ۴-۶ نمودار متغیر وابسته در مقابل متغیرهای توضیحی
۸۱	نمودار ۵-۶ نمودار هیستوگرام برای باقی مانده های استاندارد شده ی مدل
۸۳	نمودار ۶-۶ نمودار $p$ - مقدار و میزان درصد تغییرات آماره های حذف بر اساس آماره ی آزمون کلی و مولفه های آن در مقابل شماره زمان وقوع مشاهده
۸۴	نمودار ۷-۶ نمودار میزان درصد تغییرات آماره های حذف بر اساس آماره ی آزمون کلی و مولفه های آن در مقابل $p$ - مقدار
۸۷	نمودار ۸-۶ نمودار هیستوگرام برای باقی مانده های استاندارد شده ی مدل
۸۸	نمودار ۹-۶ نمودار باقی مانده های استاندارد شده در مقابل مقادیر برازش داده شده و در مقابل زمان
۸۸	نمودار ۱۰-۶ نمودار احتمال نرمال
۸۹	نمودار ۱۱-۶ نمودار متغیر وابسته در مقابل متغیرهای توضیحی
۹۰	نمودار ۱۲-۶ نمودار $p$ - مقدار آماره های حذف بر اساس آماره ی آزمون کلی و مولفه های آن در مقابل شماره زمان وقوع مشاهده
۹۱	نمودار ۱۳-۶ نمودار میزان درصد تغییرات آماره های حذف بر اساس آماره ی آزمون کلی و مولفه های آن در مقابل شماره زمان وقوع مشاهده

## فهرست جدول ها

عنوان	صفحه
جدول ۱-۵ احتمال رد فرض صفر در سطح $\alpha = 0.05$ با ۲۰۰۰ تکرار	۶۱
جدول ۲-۵ توان محاسبه شده‌ی آزمون در حالت‌های مختلف نقض فرض نرمال بودن	۶۳
جدول ۳-۵ توان محاسبه شده‌ی آزمون وقتی مدل واقعی $Y_i = x_i + x_i^\gamma \varepsilon_i$ با $\gamma = 1/5$ باشد و مدل $Y_i = \beta_0 + \beta_1 x_i + \sigma \varepsilon_i$ برازش داده شده باشد	۶۵
جدول ۴-۵ توان محاسبه شده‌ی آزمون وقتی مدل واقعی $Y_i = x_i + \sigma_i x_i$ و $\sigma_i = \sigma_1$ وقتی $i \leq n/2$ و $\sigma_i = \sigma_2$ وقتی $i > n/2$ با $(\sigma_1, \sigma_2) = (1, 1/5)$ باشد و مدل $Y_i = \beta_0 + \beta_1 x_i + \sigma \varepsilon_i$ برازش داده شده باشد	۶۵
جدول ۵-۵ توان محاسبه شده‌ی آزمون وقتی مدل واقعی $Y_i = x_i + \beta_1 x_i^\gamma + \varepsilon_i$ باشد و مدل $Y_i = \beta_0 + \beta_1 x_i + \sigma \varepsilon_i$ برازش داده شده باشد	۶۶
جدول ۶-۵ توان محاسبه شده‌ی آزمون وقتی مدل واقعی $Y_i = x_i + \varepsilon_i$ باشد و مقادیر $\varepsilon_i = \frac{1}{\sqrt{i}} \sum_{j=1}^i \varepsilon_j^*$ باشند	۶۷
جدول ۱-۶ داده‌های درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۴	۷۴
جدول ۲-۶ آماره‌های توصیفی باقی‌مانده‌های مدل رگرسیونی برازش داده شده روی داده‌های درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۴	۷۵
جدول ۳-۶ مقادیر برآورد حداقل مربعات پارامترهای مدل رگرسیونی برازش داده شده روی داده‌های درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۴ و آزمون $t$ مربوطه	۷۵
جدول ۴-۶ آنالیز واریانس مدل	۷۶
جدول ۵-۶ مقادیر آماره‌ی آزمون کلی و آماره‌های جزئی به همراه $p$ -مقدار آنها و تصمیم مناسب برای آزمون	۷۶

- جدول ۶-۶ ویژگی‌های توصیفی درصد تغییرات آماره‌های حذف بر اساس آماره‌ی  
آزمون کلی و مولفه‌های آن ۸۲
- جدول ۶-۷  $p$ -مقدار و اندازه‌ی درصد تغییرات آماره حذف مشاهداتی که بر اساس  
آماره‌ی حذف کلی و مولفه‌هایش، غیرعادی تشخیص داده شده‌اند ۸۲
- جدول ۶-۸ آماره‌های توصیفی باقی‌مانده‌های مدل رگرسیونی برازش داده شده روی  
داده‌های درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی سال‌های  
۱۳۵۲ تا ۱۳۸۳ ۸۵
- جدول ۶-۹ مقادیر برآورد حداقل مربعات پارامترهای مدل رگرسیونی برازش داده  
شده روی داده‌های درآمد حاصل از صادرات نفتی و غیرنفتی ایران طی  
سال‌های ۱۳۵۲ تا ۱۳۸۳ و آزمون  $t$  مربوطه ۸۵
- جدول ۶-۱۰ مقادیر آماره‌ی آزمون کلی و آماره‌های جزئی به همراه  $p$ -مقدار آنها و  
تصمیم مناسب برای آزمون برای داده‌های درآمد حاصل از صادرات  
نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۳ ۸۶
- جدول ۶-۱۱ ویژگی‌های توصیفی درصد تغییرات آماره‌های حذف بر اساس آماره‌ی  
آزمون کلی و مولفه‌های آن برای داده‌های درآمد حاصل از صادرات  
نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۳ ۸۶
- جدول ۶-۱۲  $p$ -مقدار و اندازه‌ی درصد تغییرات آماره حذف مشاهداتی که بر اساس  
آماره‌ی حذف کلی و مولفه‌هایش برای داده‌های درآمد حاصل از صادرات  
نفتی و غیرنفتی ایران طی سال‌های ۱۳۵۲ تا ۱۳۸۳ غیرعادی تشخیص  
داده شده‌اند ۸۷

فصل اول

کلیات

تحلیل رگرسیونی یکی از روش‌های متداول کاربردی آمار می‌باشد که در بسیاری از رشته‌های علمی مانند دارو سازی، زیست شناسی، کشاورزی، اقتصاد، مهندسی، جامعه‌شناسی، زمین‌شناسی و ... به کار برده می‌شود. اهداف تحلیل رگرسیونی عبارتند از:

(۱) پیدا کردن یک رابطه بین متغیرهای مستقل و متغیر وابسته.

(۲) پیش‌بینی متغیر وابسته توسط متغیرهای مستقل.

(۳) مشخص کردن متغیر یا متغیرهای مستقلی که بیشترین تاثیر را روی متغیر وابسته دارد.

(ژین‌یان و ژیاوگانگ‌سو<sup>۱</sup>، ۲۰۰۹، صفحه ۴)

به این منظور پس از جمع‌آوری اطلاعات مورد نیاز و قبل از برآزش مدل رگرسیونی، آن‌ها را آزمون کرده تا برقراری شرایط لازم برای تحلیل رگرسیونی بررسی شود، شرایط لازم همان فرض‌های مدل رگرسیونی خطی است که در فصل دوم راجع به آن بحث خواهد شد. در این پایان‌نامه، به دست آوردن یک فرآیند کلی و کامل برای آزمودن همزمان چهار فرض مدل‌های رگرسیونی خطی مورد توجه قرار گرفته است. آزمون بر پایه‌ی بردار باقی مانده‌های استاندارد و آزمون هموار سازی نیمین بنا شده است. اگر فرآیند کلی نقض حداقل یکی از فرض‌ها را نشان دهد، مولفه‌های آماره آزمون کلی برای تشخیص فرض یا فرض‌های نقض شده می‌توانند به کار گرفته شوند، همچنین از فرآیند می‌توان برای ساختن آماره حذف مربوطه و تشخیص نقاط غیرعادی استفاده کرد. آماره‌هایی که تابعی از بردار باقی مانده‌ها هستند،

<sup>۱</sup> Xin Yan & Xiao Gang Su

معمولاً ویژگی‌های توزیعی پیچیده‌ای دارند، لذا جنبه‌های توزیعی آزمون کلی پیشنهاد شده مجانبی است.

## ۲-۱ تاریخچه

روش‌های موجود برای بررسی اعتبار فرض‌های مدل به دو نوع دسته‌بندی می‌شود، یکی روش‌های ترسیمی و دیگری روش‌های آزمون معنی داری متداول می‌باشند. روش‌های ترسیمی را به خاطر سهولت و راحتی در دوره‌های مقدماتی و پایه آموزش می‌دهند. روش‌های ترسیمی را می‌توان در کارهای کوک و ویسبرگ (۱۹۸۲)، آتکینسون (۱۹۸۵) و کوک (۱۹۹۸) مشاهده نمود. آنها به کمک روش ترسیمی به شناسایی نقاط پرت و تاثیرگذار نیز پرداخته‌اند. مقالات زیادی در زمینه‌ی آزمون فرض‌های مدل نوشته شده و آزمون‌های متعددی به وجود آمده است. برای مثال آزمون‌های کلموگروف-اسمیرنوف (۱۹۳۳) و شاپیرو-ویلک (۱۹۶۵) و اندرسون-دارلینگ (۱۹۵۲) برای بررسی فرض نرمال بودن و آزمون‌های بارتلت (۱۹۳۷) و لون (۱۹۶۰) و برون-فریست (۱۹۷۴) برای بررسی فرض ثابت بودن واریانس خطاها و آزمون‌های دوربین-واتسون (۱۹۵۰ و ۱۹۵۱) و بروش-گودفری (۱۹۷۸ و ۱۹۷۹) برای بررسی فرض ثبات واریانس متغیرهای وابسته شرطی شده روی متغیرهای پیشگو و... به طور خاص چند نمونه از مقالاتی که برپایه‌ی بردار خطای استاندارد شده به رشته‌ی تحریر در آمده‌اند عبارتند از: توکی (۱۹۴۹)، روشی برای آزمون خطی بودن مدل ارائه کرده‌است؛ دوربین و واتسون (۱۹۵۰ و ۱۹۵۱) با تعریف ماتریس واریانس کوواریانس خطاها به بررسی برقراری فرض

ثابت بودن واریانس خطاها و استقلال آنها می‌پردازد؛ و بیکل<sup>۱</sup> (۱۹۷۸) به بررسی فرض ثابت بودن واریانس خطاها می‌پردازد؛ انسکمب (۱۹۶۱) و انسکمب و توکی (۱۹۶۳) نرمال بودن خطاها را آزمون می‌کنند. کوک و ویسبرگ (۱۹۸۲) همچنین آتکینسون (۱۹۸۵) علاوه بر بررسی فرضها به روش ترسیمی، در مورد اعتبار فرضهای مدل‌های رگرسیونی خطی به طور خلاصه بحث نموده‌اند. خصوصیات توزیعی باقی‌مانده‌ها را می‌توان در کارهای دوربین و واتسون (۱۹۵۰)، انسکمب و توکی (۱۹۶۳)، تیل<sup>۲</sup> (۱۹۶۵) و آتکینسون (۱۹۸۵) ملاحظه نمود.

ممکن است در اینجا این سوالات مطرح شوند که: " آیا ممکن است این آزمون‌ها را به طور همزمان انجام داد؟ به طور کلی چرا باید آزمونی مطرح شود که این فرضها را همزمان آزمون کند؟! از چه روش جدیدی برای ادغام این آزمون‌ها استفاده می‌شود؟ چه تفاوتی بین این روش با سایر روش‌های قبلی مثل بونفرونی یا... وجود دارد؟ توان این آزمون در مقایسه با آزمون‌های سابق چگونه است؟

### ۳-۱ دلایل طرح مساله

بررسی اعتبار روش‌های استنباطی، تا حد زیادی به اعتماد کردن به صحت اعتبار فرض‌های مدل بستگی دارد. در زمینه‌ی بررسی برقراری فرض‌های مدل توجه زیادی شده و روش‌ها و آزمون‌های متعددی پیشنهاد شده است.

<sup>۱</sup> Bickel

<sup>۲</sup> Theil



اشکال این آزمون‌های مجزا این است که هر یک از آنها تحت برقرار بودن شروط دیگر انجام می‌شود، لذا در عمل و در بسیاری از موارد رد فرضیه ممکن است به دلیل برقرار نبودن سایر فرض‌های مدل باشد. از سوی دیگر انجام چهار آزمون مجزا باعث افزایش خطای نوع اول خواهد شد. بنابراین به یک آزمون کلی برای فرض‌های مدل رگرسیونی خطی نیازمندیم، که میزان خطای نوع اول را کنترل کند و نیز بتواند در صورتی که استفاده‌کننده علاقه‌ای به اینکه کدام فرض نقض می‌شود نداشته باشد، مورد استفاده قرار گیرد. اگر چنین آزمونی مشخص کند که حداقل یکی از فرض‌ها برقرار نیست، آنگاه آزمون راهنما ممکن است برای تعیین فرض‌هایی که نقض شده‌اند مورد استفاده قرار گیرد.

در این رساله چنین آزمون کلی معرفی شده است، آزمون بر اساس بردار باقی‌مانده‌های استاندارد شده و آزمون هموارسازی نیمین می‌باشد، به علاوه مؤلفه‌های این آزمون جامع می‌توانند به عنوان آزمون‌های راهنما برای تعیین فرض‌هایی که نقض شده‌اند مورد استفاده قرار گیرد. آماره‌هایی که تابعی از بردار باقی‌مانده‌های استاندارد هستند، معمولاً ویژگی‌های توزیعی پیچیده‌ای دارند، لذا جنبه‌های توزیعی آزمون کلی پیشنهاد شده مجانبی است.

#### ۴-۱ ساختار پایان‌نامه

این رساله به طور نظری و کاربردی تمام آنچه را که بیان شده مطرح می‌کند. در فصل دوم به تعریف مفاهیم اساسی و کلیدی این پایان‌نامه پرداخته می‌شود. در فصل سوم آماره‌ی آزمون، آماره‌های جزئی و توزیع آنها بر اساس بردار باقی‌مانده‌های استاندارد شده معرفی خواهند شد و در انتها آماره‌ی حذف ساخته شده بر اساس آماره‌ی کلی مطرح می‌شود. در فصل چهارم قضایای

---

مورد نیاز مطرح خواهد شد و با توجه به ایده‌ی نیم‌ن آماره‌ی آزمون کلی و توزیع آن محاسبه می‌شود، در فصل پنجم به کمک شبیه‌سازی‌های کامپیوتری توان این آماره با آماره‌های به دست آمده به کمک روش‌های قدیمی مقایسه می‌شود و در فصل آخر با به کار بردن مجموعه داده‌های صادرات غیر نفتی ایران طی سال‌های ۵۲ تا ۸۴ به بررسی فرض‌های مدل رگرسیون خطی در این داده‌ها پرداخته می‌شود.

## فصل دوم

### تعاریف و مفاهیم اساسی

## ۱-۲ مقدمه

در این فصل به تعریف مفاهیم اساسی و مورد نیاز در این پایان نامه پرداخته شود، باشد که از لحاظ تئوری در درک هر چه بهتر این رساله به خواننده یاری رساند.

در ابتدا مدل رگرسیونی خطی معرفی و فرض های آن ارائه خواهد شد. با تعریف برآوردهای حداکثر درستنمایی و حداقل مربعات و بیان ویژگی های برآورد کننده ها، برآوردهای پارامترهای مدل رگرسیونی خطی را مطرح نموده و پنج نوع از باقی مانده ها معرفی می شوند. در ادامه فرض صفر و فرض مقابل آزمون بررسی فرض های مدل خطی را تعریف کرده و پس از آن چولگی و کشیدگی، که پایه و اساس به وجود آمدن دو آماره جزء ابتدایی می باشند بیان می شود. در بخش بعدی چند نمونه از آزمون های سابق برای بررسی فرض های مدل رگرسیونی خطی که در فصل یک نیز معرفی شدند، توضیح داده خواهند شد. در پایان نظریه نیمین در مورد هموارسازی که پایه ی بحث فصل چهارم می باشد، مطرح می شود.

## ۲-۲ مدل رگرسیونی خطی و فرضیه های آن

یکی از مدل های بسیار مهم در آمار مدل رگرسیونی خطی است، که مبین رابطه بین یک بردار  $n \times 1$  قابل مشاهده از پاسخ ها ( $Y$ ) و یک ماتریس  $n \times p$  قابل مشاهده ی از مقادیر متغیرهای پیشگو ( $X$ ) می باشد. در مدل رگرسیونی خطی رابطه ی بین  $X$  و  $Y$  به صورت زیر است:

$$Y = X\beta + \sigma\varepsilon \quad (1-2)$$

که  $\beta$  یک بردار  $p \times 1$  از ضرایب نامعلوم و  $\sigma$  پارامتر مقیاس نامعلوم و  $\varepsilon$  بردار  $n \times 1$  از متغیرهای خطای غیرقابل مشاهده است. علاوه بر این فرض می شود که  $\varepsilon$  های شرطی شده