

خطوطی که از مبدأ می‌گذرند، آیا رگرسیون از طریق مبدأ جواب است؟

محمد شعبانی^۱

چکیده

برازش خطی که از نقطه $(0, 0)$ می‌گذرد، مقوله‌ایست که بارها با آن در کاربرد رگرسیون خطی مواجه می‌شویم. بسیاری از نرم‌افزارهای کامپیوتری یک انتخاب حداقل مربعات دارند که بعضی اوقات بوسیله "No INT" و یا حتی "NoINT" مشخص می‌شوند. این انتخاب، بیانگر آن است که عرض از مبدأ مدل را حذف نماییم. غالباً این روشها نامعتبر بوده و یا مطابق با مناسبترین مدل نیستند. این مقاله، مسایل و پرسشهایی در مورد خط حداقل مربعات مطرح کرده و بعضی از روشهای اصلاحی را پیشنهاد می‌کند. برای مقایسه برازندگی مدل‌های اصلاحی دو اندازه ارایه می‌گردد.

۱ مقدمه

بوده، به قسمی که تک تک مقادیر y حول میانگین شرطی در حال تغییر کردن می‌باشند. این موضوع دلالت بر آن دارد که وقتی x صفر یا نزدیک آن باشد، مقادیر y حول میانگینی که صفر (یا نزدیک آن) است در حال تغییر می‌باشند تا اینکه به ناچار، بعضی از مقادیر y منفی خواهند شد.

۲ موقعیتهایی که رگرسیون از طریق مبدأ نامناسب است

آشکارترین شکست مدل رگرسیون از طریق مبدأ وقتی رخ می‌دهد که خط از نقطه $(0, 0)$ نگذرد. چنین وضعیتی در اشکال $\lambda(a)$, $\lambda(b)$, $\lambda(c)$ ارایه شده است. البته ممکن است در

موقعیتهای بسیاری وجود دارد که رابطه بین دو متغیر خطی بوده و انتظار می‌رود که خط از نقطه $x = 0, y = 0$ بگذرد. عرض از مبدأ صفر نسبت به نقطه داده‌ای و با حتی تعداد تکرارهای زیاد، دارای نفوذ بیشتری می‌باشد. قابل ذکر است که در بعضی موارد دچار تناقض می‌شویم. به این مفهوم که اگرچه وقتی $x = 0$ می‌باشد انتظار می‌رود که $y = 0$ باشد، اما واقعاً سیستم قادر نیست که $y = 0$ را مشاهده کند. بعنوان نمونه، فرآیند تولید یک محصول را در نظر بگیرید که بر اساس مصرف یک سری از مواد اولیه تأمین می‌شود. وقتی که هیچ ماده اولیه‌ای مصرف نگردد، نمی‌توان محصولی از این فرآیند مشاهده کرد. مدل رگرسیون بدون عرض از مبدأ $E(y) = bx$ ، دلالت بر آن دارد که مقدار میانگین y بطور خطی به x وابسته

^۱ دانشجوی کارشناسی ارشد، بخش آمار، دانشگاه شیراز

همگونی واریانس خطاها را در نظر بگیریم. اگر واریانس خطاها در مدل عرض از مبدأ بخواهد پایدار بماند، باید وقتی که x به سمت صفر میل می‌کند، تک تک مقادیر y که ممکن است مثبت یا منفی باشند، حول میانگین به قسمی تغییر کنند که عرض نوار حول خط رگرسیون پایدار بماند. در حالیکه ممکن است این مسئله در بسیاری از کاربردها، وقتی که هم x و هم y لزوماً مثبت هستند، برقرار باشد، اما خواص توأم همگونی و عرض از مبدأ صفر با یکدیگر ناسازگارند. در این صورت وقتی که به عرض از مبدأ صفر نزدیک می‌شویم، باقیمانده‌ها حول خط تمایل به کاهش دارند. اشکال ۳(a)، ۳(b)، ۳(c) این نکته را تشریح می‌کنند.

۳ مدل‌های اصلاحی

مدلهایی را در نظر می‌گیریم که از مبدأ می‌گذرند و برازش بهتری نسبت به مدل رگرسیون از طریق مبدأ آرایه می‌دهند. یکی از خصوصیات مدل‌های اصلاحی آرایه شده در این مقاله، قابل تفسیر بودن آنها با توجه به کاربردهای عملی می‌باشد.

۱.۳ اولین و آشکارترین مدل اصلاحی برای رگرسیون از طریق مبدأ، مدل رگرسیون خطی ساده $y = a + bx + \varepsilon$ می‌باشد.

۲.۳ حالتی را در نظر می‌گیریم که به جای y, x به عنوان متغیر وابسته باشد. متغیر $z = \frac{y}{x}$ ممکن است میانگین پایداری داشته باشد، اما در معرض تغییرات تصادفی قرار گیرد.

مدل ممکن است که می‌توان در نظر گرفت، مدل نسبت ثابت (Constant Ratio Model) می‌باشد که بفرم $\frac{y}{x} = a + \varepsilon$ نمایش داده می‌شود. تبدیل خارج قسمت $z = \frac{y}{x}$ ، منجر به مدل $z = a + \varepsilon$ می‌گردد.

۳.۳ ممکن است که ارتباطی خطی بین x و z وجود داشته باشد، بطوریکه مدل مناسب بفرم $z = a + bx + \varepsilon$ آرایه گردد، که آنرا مدل رگرسیون خارج قسمت (Quotient Regression Model) می‌نامیم. حال اگر مقادیر خطا روی برد x پایدار باشند و با توجه به $z = \frac{y}{x}$ ، خواهیم

عمل چنین حالتی رخ ندهند، اما اشکال ۱(d)، ۱(e) با واقعیت سازگارترند. بعنوان نمونه وقتی که معادله رگرسیونی معرف یک تابع کالیبراسیون باشد، ممکن است خطاهای صفر، منفی یا مثبت باشند. به قسمی که وقتی $x = 0$ است $y \neq 0$ باشد. مثلاً دستگاه درجه‌بندی شده‌ای را در نظر بگیرید که با استفاده از آن، میزان سرعت باد را بر حسب وزش باد اندازه‌گیری می‌کنند. واضح است که اگر بادی نوزد شمارنده عدد صفر را نشان می‌دهد، اما ممکن است یکدفعه آشفته‌گی ایجاد شده و طوفانی بوجود آید که سبب شود مدل رگرسیون عرض از مبدأش از صفر به عددی جابه‌جا گردد که ناشی از خطاها بوده و سبب می‌گردد که از $x = 0$ را نتوانیم $y = 0$ را نتیجه بگیریم. در اشکال ۱(a) تا ۱(e) خطوط رگرسیون از طریق مبدأ رسم شده‌اند، اما همانطوری که انتظار می‌رود ممکن است چنین مدلی مناسب نباشد. مثلاً بخاطر تحمیل عرض از مبدأ صفر، خط رگرسیون حداقل مربعات از نقطه $(\bar{x}, \bar{y})$ که در نمودار با + مشخص شد، نمی‌گذرد. موقعیت مشهور دیگری که ممکن است نامناسب بودن مدل رگرسیون از طریق مبدأ را آشکار سازد، این است که رابطه بین x و y بقسمی است که هرچند از مبدأ می‌گذرد، اما ممکن است غیر خطی باشد. بعید بنظر می‌رسد که انحرافهای بزرگ توسط روشهای نموداری نمایان گردد. در حقیقت، از طریق آنالیز باقیمانده‌ها و آزمون عدم برازش می‌توان چنین روابط غیر خطی را جستجو کرد. در روابطی که انحناء رابطه غیر خطی کوچک باشد، ممکن است که غیر خطی بودن رابطه، کمتر آشکار گردد. اشکال ۲(a)، ۲(b) این موضوع را به خوبی توضیح می‌دهند.

بر اساس اشکال ۲(a)، ۲(b) دیده می‌شود که هم رابطه واقعی و هم مدل بدون عرض از مبدأ برازش شده و مدل رگرسیون از طریق مبدأ، به رابطه واقعی خیلی نزدیک است. در شکل ۲(a) رابطه واقعی بفرم $y = 1/25x^{0.8}$ معرفی شده و در شکل ۲(b) این رابطه عبارت است از: $y = 0.8x^{1/5}$. در یک موقعیت عملی، ممکن است محسوس‌تر بودن رابطه غیر خطی (یا مدل رگرسیون خطی ساده) بوسیله تغییرات باقیمانده‌ها پنهان گردد. لذا، مناسب است که فرضهای مدل رگرسیونی شامل

داشت:

$$\frac{y}{x} = a + bx + \varepsilon \quad (۱)$$

با ضرب کردن طرفین رابطه (۱) توسط x داریم:

$$y = ax + bx^2 + x\varepsilon$$

یعنی یک مدل دو جمله‌ای بدون عرض از مبدأ و با خطایی که متناسب با x می‌باشد، خواهیم داشت. معه‌ذا تبدیل خارج قسمت، منجر به مدل رگرسیون خطی ساده با جمله خطای همگن خواهد شد.

۴.۳ ممکن است رابطه $z - x$ خطی نباشد اما z به سمت حد بالا یا پایین، وقتی x افزایش می‌یابد، میل کند. در حالیکه می‌توان رابطه را توسط یک فرم نمایی (که خارج از بحث مقاله است) معرفی کرد، اما یک مدل ساده‌تر بصورت $z = a + \frac{b}{x} + \varepsilon$ ارایه می‌دهیم. این مدل را رگرسیون با خطای مناسب (Proportional Error Regression) می‌نامیم. دوباره با جانشین کردن $\frac{y}{x}$ به جای z داریم:

$$\frac{y}{x} = a + \frac{b}{x} + \varepsilon \quad (۲)$$

$$\Rightarrow y = ax + b + x\varepsilon$$

رابطه (۲) بیانگر یک مدل خطی بوده که خطای آن با x متناسب می‌باشد. اگر ضریب b منفی باشد، گرایش $\frac{y}{x}$ به طرف یک مقدار مجانب بالایی بوده و در صورتی که ضریب b مثبت باشد، $\frac{y}{x}$ به سمت یک مقدار مجانبی پایین گرایش خواهد داشت.

مدل (۲)، خطی با عرض از مبدأ b و شیب a خواهد بود؛ معه‌ذا وقتی که خطاها در y با x افزایش یابد، تبدیل خارج قسمت مناسب بوده و سبب پایداری واریانس می‌گردد.

۵.۳ سرانجام ممکن است که رابطه $y - x$ خطی نباشد و ضمناً هیچ یک از روابط $z - x$ ، $z \frac{1}{x}$ نیز برقرار نباشند. در

چنین حالتی ممکن است مدلی به فرم:

$$\ln y = a + b \ln x + \varepsilon \quad (۳)$$

مناسب باشد. اگر از طرفین رابطه (۳)، تبدیل نمایی بگیریم و با فرض اینکه مقادیر ε کوچک باشند ($e^\varepsilon \approx 1 + \varepsilon$)، در آن صورت مدلی به فرم $y = kx^b(1 + \varepsilon)$ بدست می‌آید که آن را رگرسیون توانی (Power Regression) می‌گوییم. همانطوری که استنباط می‌گردد خطاهای مدل با x^b متناسب می‌باشد. برای انحرافات متوسط از خطی بودن (یعنی b نزدیک ۱)، مقادیر ε به طور تقریبی هم با x و هم با y متناسب خواهد بود.

۴ اندازه‌گیری سودمندی یا برازندگی مدل:

وقتی که مدل‌های اصلاحی مورد استفاده قرار گیرند، در آنصورت مقایسه مناسب بودن آنها بر اساس این موضوع که تا چه حد اندازه مدل به داده‌ها به خوبی برازش شده است، مورد نیاز می‌باشد. در رگرسیون مرسوم، همبستگی بین x, y و یا میانگین مربعات خطا (MSE) اندازه مناسبی جهت مقایسه مدل‌ها از نقطه نظر برازش بهتر به داده‌ها ارایه نمایند. در رگرسیون چندگانه (از جمله در چندجمله‌ایها)، R^2 (ضریب تعیین)، میانگین مربعات خطاها و جذر میانگین مربعات خطاها ($S = \sqrt{\text{MSE}}$) معیارهایی جهت مقایسه مدل‌ها می‌باشند. برای بعضی از مدل‌های اصلاحی پیشنهادی در این مقاله، r ، R^2 و حتی MSE نامناسب هستند. مثلاً برای مدل رگرسیونی از طریق مبدأ، R^2 وجود ندارد و برای مدل نسبت ثابت که تنها یک متغیر $z = \frac{y}{x}$ دارد، R^2 قابل تعریف نیستند.

حال سؤال این است که چگونه باید برازش رگرسیون توانی را با رگرسیون از طریق مبدأ یا رگرسیون نسبت ثابت مقایسه کرد؟ ابتدا به این نکته توجه می‌کنیم که در رگرسیون خطی ساده داریم:

$$r(y, x) = r(y, \hat{y}) \quad (۴)$$

کیللوگرمی مایعی، از بستری که حاوی عناصری است که توسط مایع جذب می‌گردد، عبور می‌نماید. غالباً ۶ تا ۶/۵ درصد از عناصر توسط وزنشان جذب می‌گردند، اما بستر با حدود ۷/۵ درصد عناصر تهیه می‌گردد تا از خشک شدن اجتناب گردد. داده‌ها در جدول ۱ خلاصه شده‌اند. در این جدول وزن مایع و میزان جذب بر حسب کیلوگرم می‌باشند. عناصر اضافی را نمی‌توان دوباره مورد استفاده قرار داد و قسمتی از این مطالعه به کاهش ضایعات و تغییرات مربوط می‌شود. ۸ بسته پوشش‌دار که برد وسیعتری از اندازه‌های مایع می‌باشند به دقت مورد ارزیابی قرار گرفته و معلوم شد که نسبت به آنچه که قبلاً بدست آورده بودیم، رابطه آشکارتری بین اندازه مایع و درصد جذب موجود است. برای اینکه بررسی کنیم که نتیجه بدست آمده ناشی از وجود اثر عجیبی نمی‌باشد، ۱۶ بسته دیگر که با برد مشابهی از اندازه‌ها پوشش داده شده بودند، بکار گرفته شدند. آنها نه تنها اثر را تأیید کردند، بلکه بر حسب عبارتی از رگرسیون مرسوم، عرض از مبدأ منفی y را پیشنهاد می‌کنند. ضمناً شواهدی موجود است مبنی بر اینکه برای بسته‌های زیر ۵۰۰ کیلوگرم، سقوط تندی در درصد جذب مشهود بوده در حالیکه برای بسته‌های بیش از ۱۰۰۰ کیلوگرم یک گرایش مجانبی دیده می‌شود (جدول ۱).

هر دو این اثرات به عرض از مبدأ منفی مربوط می‌شوند. اما آسانتر است که بر حسب جملاتی از درصد جذب تفسیر شوند و در تأیید گفته‌هایمان به نمودارهای صفحه ۲۸، جدول ۱ و جدول ۲ توجه بیشتری معطوف می‌کنیم. در صفحه ۲۸ ابتدا نمودار y بر حسب x رسم شده که دلیلی بر رابطه خطی داشته و عرض از مبدأ آن منفی می‌باشد. این نمودار بیانگر آن است که همراه با اندازه بسته‌ها، تغییرات افزایش می‌یابد. نمودار $\frac{y}{x}$ بر حسب x پیشنهاد کننده سقوط در $\frac{y}{x}$ برای مقادیر کوچک x بوده و ضمناً گرایش مجانبی را برای مقادیر بزرگ x نشان می‌دهد. نمودار $\log y$ بر حسب $\log x$ (رگرسیون توانی) مناسب به نظر نمی‌رسد، اما اگر به نمودار $\frac{y}{x}$ در برابر $\frac{1}{x}$ نظر افکنیم، مدل رگرسیون با خطای متناسب، مناسب به نظر می‌رسد.

در جایی که: $r(y, \hat{y})$ ، دلالت بر همبستگی بین مقادیر متغیر وابسته و مقادیر متغیر وابسته و مقادیر پیش بینی شده مدل دارد. لذا $r(y, \hat{y})$ شاخصی جهت مقایسه برازش مدلها، مثلاً مدل تبدیل یافته یا منحنی - خطی وار (curvilinear) با مدل‌های دیگر و یا با مدل خطی می‌باشد. معهذاً از این شاخص نمی‌توان جهت مقایسه بین یک مدل خطی با مدل دیگر (مثلاً مدل رگرسیون از طریق مبدأ با مدل نسبت ثابت یا رگرسیون خطی ساده) استفاده کرد، زیرا همه مدل‌های خطی، منجر به رابطه (۴) می‌گردند. روش دیگر برای مقایسه برازش مدلها بر اساس میانگین مربعات خطاها (MSE) می‌باشد. به طور واضح برای مدل‌های از نوع یکسان، مقایسه مستقیم بین MSE معتبر می‌باشد و مدلی که MSE کوچکتری داشته باشد، برازش بهتری داده‌ها نسبت می‌دهد.

اما ممکن است علاقه‌مند باشیم مدل با خطای پایدار را با مدلی که خطایش متناسب با x می‌باشد مقایسه نماییم. در چنین مواردی، میانگین مربعات خطای متناسب (The Mean Square Proportional Error) یا جذر آن مناسب بوده و این شاخص را با:

$$\mu^2 = \frac{1}{f} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 \quad (5)$$

نمایش می‌دهیم و جذر آن را با γ معرفی می‌کنیم. در رابطه (۵)، f تعداد درجات آزادی مناسب می‌باشد. ممکن است همانند رگرسیون خطی ساده، f لزوماً $n - 2$ نباشد بلکه در حالت کلی داریم:

$$f = n - \text{تعداد پارامترهای برآورد شده}$$

حتی ممکن است که بخواهیم از مقدار μ^2 (مشابه با R^2) یک شبه همبستگی (Quasi - Correlation) بسازیم، اما از بیان کردن این موضوع خودداری می‌کنیم، چون احتمالاً $r(y, \hat{y})$ و R^2 برای اهداف علمی کفایت می‌کنند.

۵ مثال کاربردی

در یک فرآیند شیمیایی، بسته‌های ۵۰۰ تا ۱۰۰۰

۷ برای مدل رگرسیون خطی ساده در صفحه ۲۵ با استفاده از نرم افزار MINITAB تحت WINDOWS نشان داده شده است.

۶ نتیجه گیری

در این مقاله، چند مدل ممکن برای معرفی کردن روابط خطی تقریبی که از نقطه $(0, 0)$ می گذرد، ارائه نمودیم. مسئله پایداری یا متناسب بودن تغییرات باقیمانده ها در نظر گرفته شد مدل های چند جمله ای، نمایی و گمپرتز، همچنین مدل هایی بر اساس تبدیلات پایدار کننده واریانس بدست می آیند، در این بحث گنجانده نشده اند، هر چند که می توانند به عنوان مدل های اصلاحی در نظر گرفته شوند. همچنین ممکن است روش های دیگری برای آنالیز داده ها (نظیر آنالیز باقیمانده های مدل) بکار گرفته شوند. همچنین نشان دادیم که همبستگی مرسوم پیرسن، تنها شاخص جهت مقایسه مدل های اصلاحی نبوده و همبستگی بین متغیرهای وابسته و متغیرهای پیش بینی شده $(\hat{y})$ وسیله ای جهت مقایسه مدل های منحنی - خطی ارائه می دهد.

همچنین آماره ای که بر اساس میانگین مربعات خطای متناسب ارائه می گردد، در بعضی از موارد بطور سودمندی جایگزین MSE خواهد بود. ضمناً نکته مهمی در معرفی کردن مدل های اصلاحی در این مقاله، قابل تفسیر بودن آنها در وضعیت های علمی و کاربردی می باشد و مدلی که به شکل سیستم مربوط می گردد، ممکن است به مدل دیگری که حتی بهتر برازش می گردد ترجیح داده شود، چرا که آن مدل این پژوهش علمی را نقص می کند.

این نمودار، خطی بودن تقریبی را نشان داده و به ازای برد مشاهده شده مقادیر x ، تغییرات پایداری (در $\%$) ملاحظه می گردد.

در حالیکه رگرسیون y بر حسب x منجر به خلاصه رضایت بخشی $R^2 = 98.6\%$ می گردد، اما نمودار y در برابر $\frac{1}{x}$ ، وسیله بهتری برای ارزیابی تغییرات باقیمانده ها ارائه می دهد، زیرا باقیمانده ها برای رابطه y بر حسب x از اعتبار کلی برخوردار نمی باشد. در حقیقت نمودار باقیمانده (Residual vs. Fits) صفحه ۲۹، بادبزنی به سمت راست (fan-shaped to right) می باشد. چهار مدل ارائه شده در جدول ۲ همگی خطی بوده، لذا $r(y, \hat{y})$ در انتخاب بین مدل ها هیچ کمکی نخواهد کرد، چرا که این شاخص، طبق جدول ۲ برای همه مدل ها 0.9930 گزارش شده است. چون به نظر می رسد که خطاها متناسب با x می باشند، لذا جذر میانگین مربعات خطای متناسب (γ) ، اندازه ای برای مناسب بودن برازش مدل ها ارائه می کند. جدول ۲، شاخص مزبور را برای هر یک از چهار مدل ممکن ارائه داده است. مدل با خطای متناسب نه تنها بهترین برازش را ارائه می دهد، بلکه پذیرفتنی تر می باشد، زیرا اینکه نسبت $\frac{y}{x}$ با افزایش x کاهش می یابد قابل تفسیرتر است تا بخواهیم مدل رگرسیون خطی ساده با عرض از مبدأ منفی y را بکار ببریم. به طور کلی، چون همگی مدل ها خطی هستند، شاخص $r(y, \hat{y})$ در این مثال نامناسب بوده اما آماره خطای متناسب، به طور واضحی مدل های بدون عرض از مبدأ و نسبت ثابت را از مدل های رگرسیون خطی ساده و مدل با خطای متناسب جدا می سازد. برای درک بهتر، موضوع محاسبات مربوط به $r(y, \hat{y})$

جدول ۱: جذب عناصر توسط مایع در پروسه جذب فرآیند

وزن مایع (x)	میزان جذب (y)	درصد جذب (y/x%)	وزن مایع (x)	میزان جذب (y)	درصد جذب (y/x%)	وزن مایع (x)	میزان جذب (y)	درصد جذب (y/x%)
۳۱۰	۱۴	۴/۵۲	۱۷/۱	۳۳۰	۲۰/۴	۵/۷۶	۲۱/۳	۳۷۰
۴۹۰	۲۷/۲	۵/۵۵	۲۰/۴	۴۰۰	۲۲/۵	۶/۰۹	۲۷/۴	۴۵۰
۵۸۰	۳۱/۹	۵/۵۰	۲۹/۸	۶۵۰	۲۲/۵	۵/۴۶	۲۸/۴	۵۲۰
۶۵۰	۳۴/۱	۵/۲۵	۴۳/۸	۸۰۰	۶/۲۲	۵/۹۲	۳۸/۵	۶۵۰
۸۱۰	۵۰/۴	۶/۲۲	۶۴/۳	۱۰۲۰	۶/۹۹	۶/۶۳	۵۰/۴	۷۶۰
۱۰۲۰	۷۱/۳	۶/۹۹	۸۰/۸	۱۲۰۰	۶/۳۸	۵/۸۸	۵۳/۵	۹۱۰
۱۲۳۰	۷۸/۵	۶/۳۸	۱۰۵/۶	۱۴۶۰	۶/۶۲	۶/۸۶	۷۹/۶	۱۱۶۰
۱۴۹۰	۹۸/۶	۶/۶۲				۷/۱۷	۹۸/۹	۱۳۸۰

خطوطی که از مبدأ می‌گذرند، آیا رگرسیون از طریق مبدأ جواب است؟ _____ سال پنجم، شماره اول □ ۲۵

جدول ۲: مقایسه مدل‌های خطی مربوط به داده‌های پروسه جذب در جدول ۱

مدل خطی	شیب	عنصر از مبدأ y	عنصر از مبدأ x	$r(y, \hat{y})$	جذر میانگین مربعات خطای متناسب (γ)
مدل رگرسیون خطی ساده $y = a + bx$	۰/۰۷۵۳۰	-۹/۸۹۶	۱۳۳	۰/۹۹۳۰	۰/۰۸۰۶۰۹۷
مدل بدون عرض از مبدأ $y = bx$	۰/۰۶۵۰۱۵	۰	۰	۰/۹۹۳۰	۰/۱۳۱۷۶
مدل نسبت ثابت $\frac{y}{x} = b$	۰/۰۶۰۳۰۸	۰	۰	۰/۹۹۳۰	۰/۱۱۷۷۴
مدل با خطای متناسب $\frac{y}{x} = \frac{a}{x} + b$	۰/۰۷۲۲۹	-۷/۷۲۵	۱۰۷	۰/۹۹۳۰	۰/۰۶۹۷۸

MTB>Name c3='FITS1' c4='RESI1'

MTB>Regress C1 1 C2;

SUBC> Fits 'FITS1';

SUBC> Constant;

SUBC> Residuals 'RESI1'.

Regression Analysis

The regression equation is

$y = -9.90 + 0.0753x$

Predictor	Coef	Stdev	t-ratio	p
Constant	-9.896	1.689	-5.86	0.000
x	0.075302	0.001925	39.12	0.000

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	1	17633	17633	1530.29	0.000
Error	22	253	12		
Total	23	17886			

MTB> CORR 'Y' 'FITS1'

Correlations (Pearson)

Correlation of y and FITS1 = 0.993

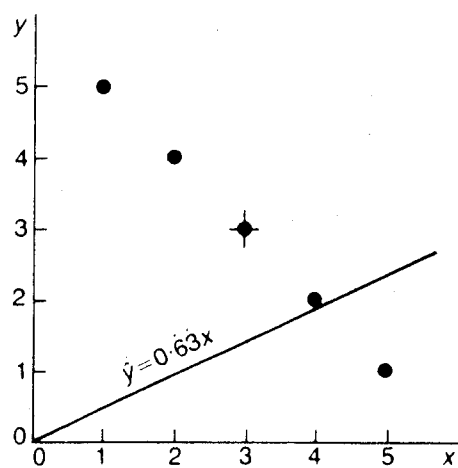
MTB>LET K1=(1/(COUNT('Y')-2))*SUM(((('Y'-'FITS1')/'FITS1'))**2)

MTB>LET K2=SQRT(K1)

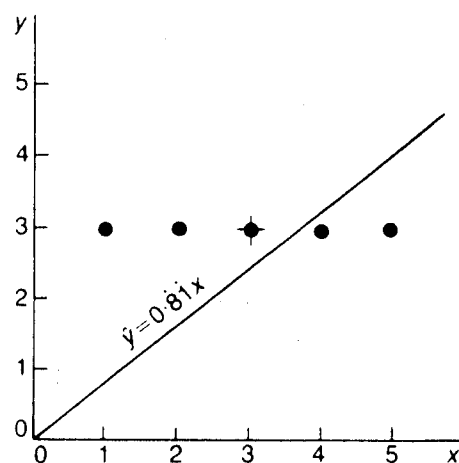
MTB>PRINT K2

Data Display

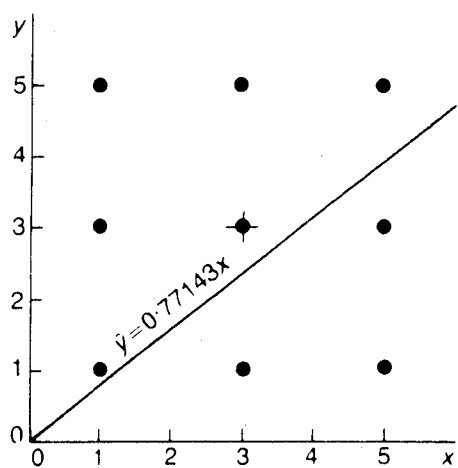
K2 0.0806097



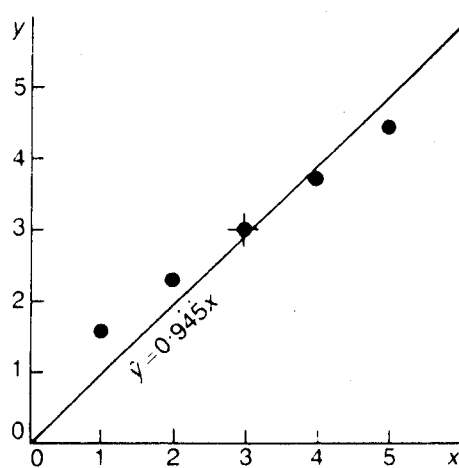
(a)



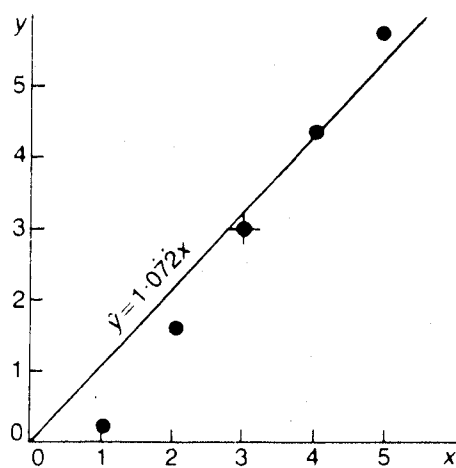
(b)



(c)



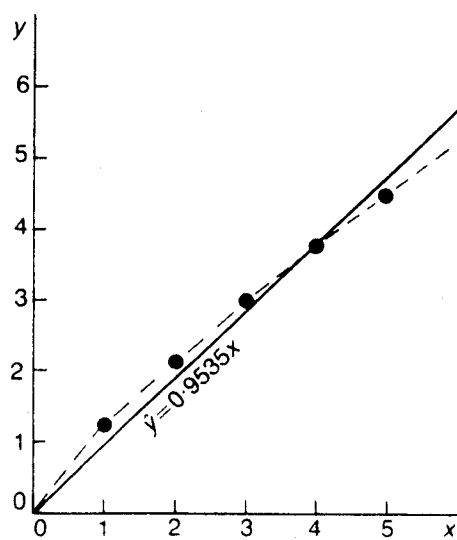
(d)



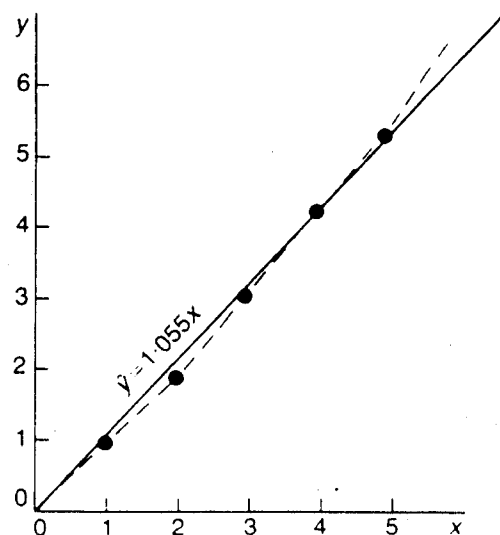
(e)

شکل ۱

خطوطی که از مبدأ می‌گذرند، آیا رگرسیون از طریق مبدأ جواب است؟ _____ سال پنجم، شماره اول □ ۲۷

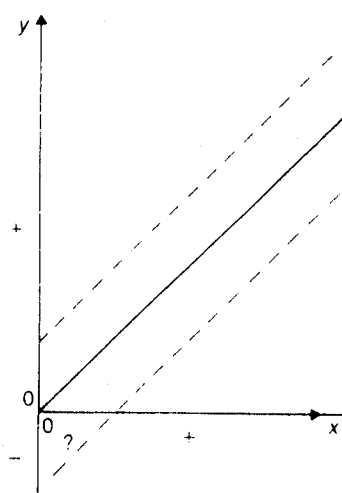


(a)

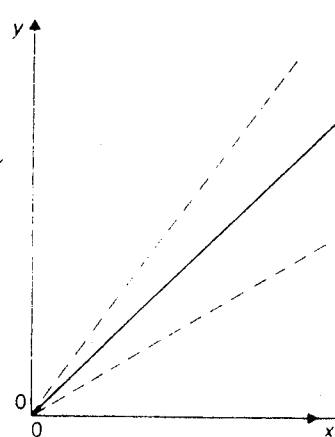


(b)

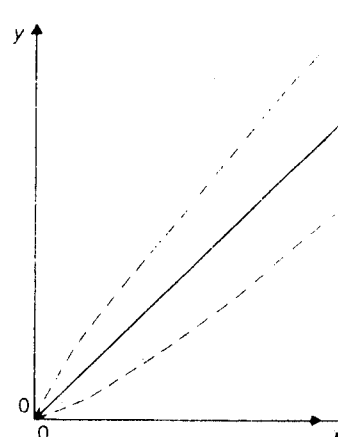
شکل ۲



(a)

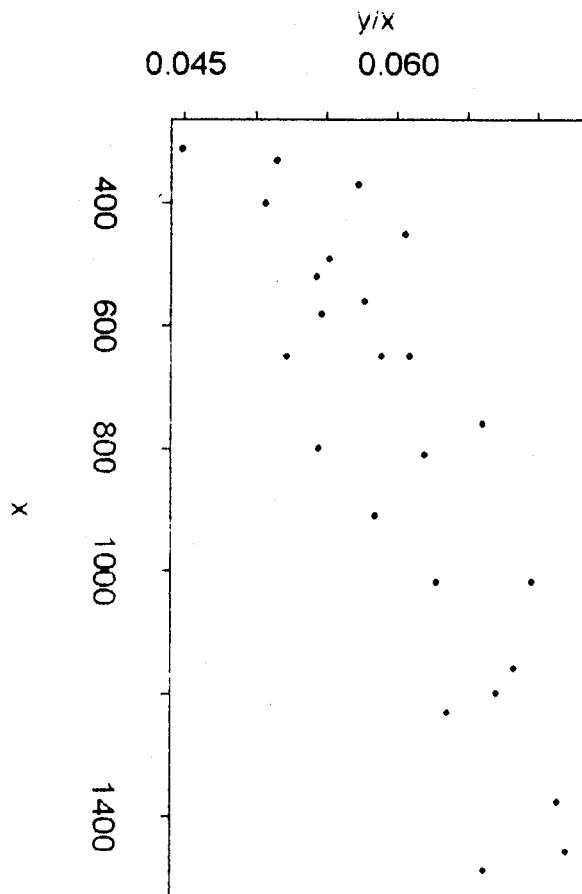
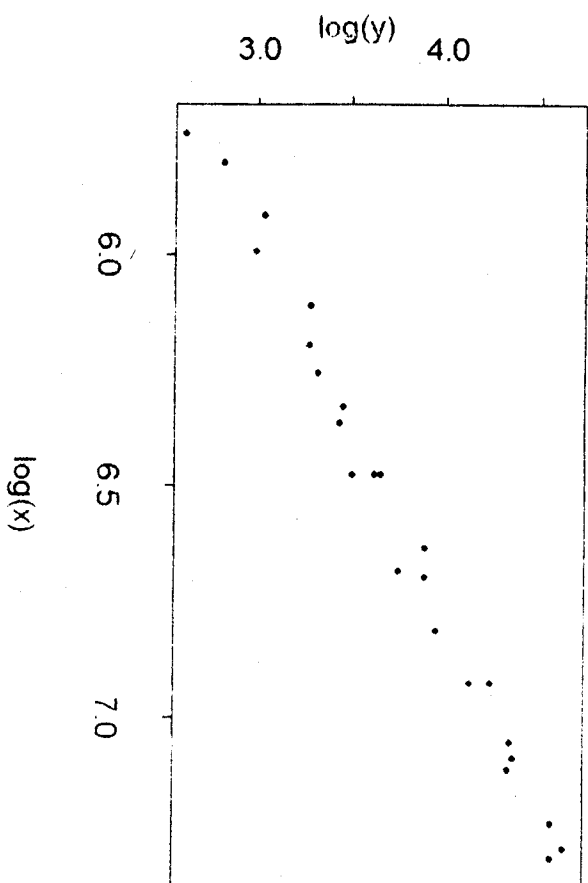
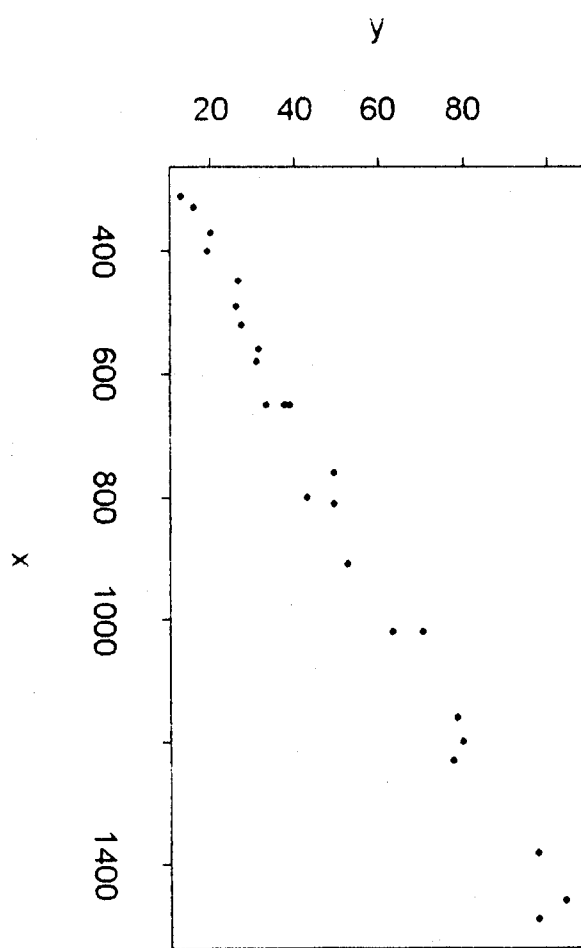
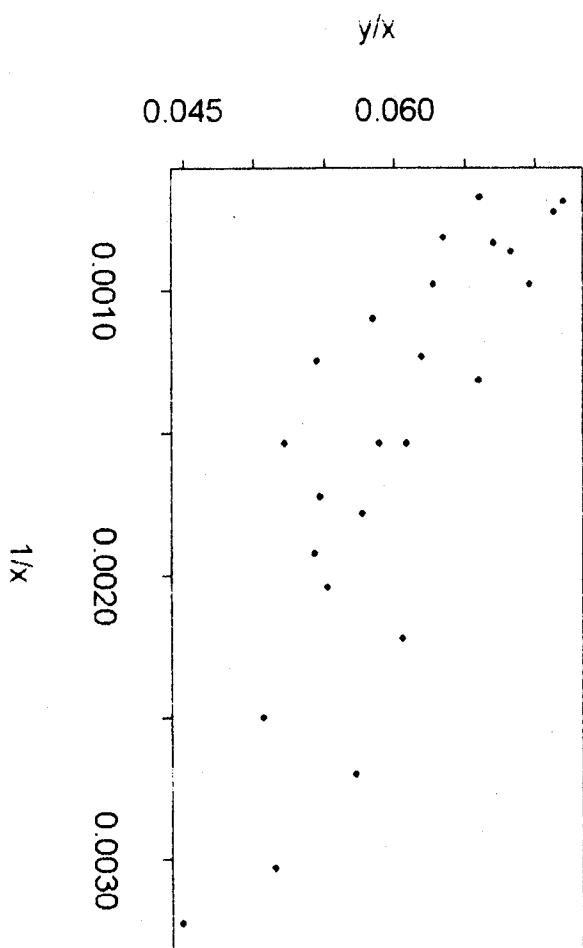


(b)

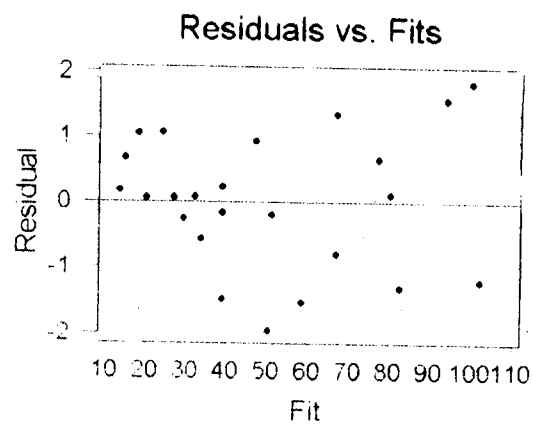
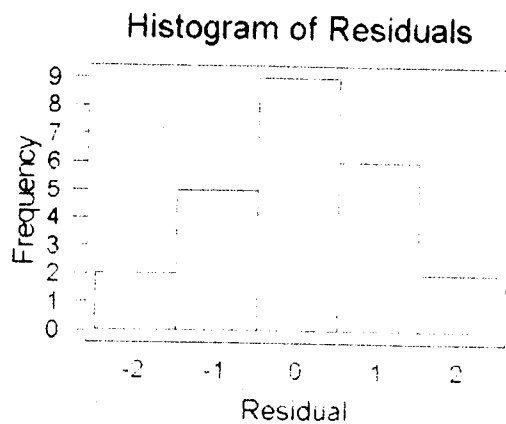
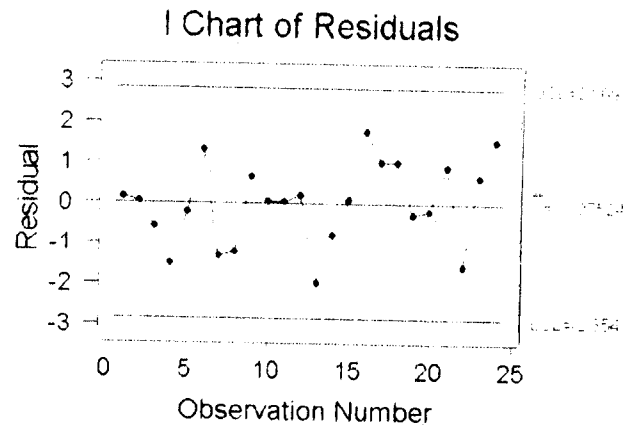
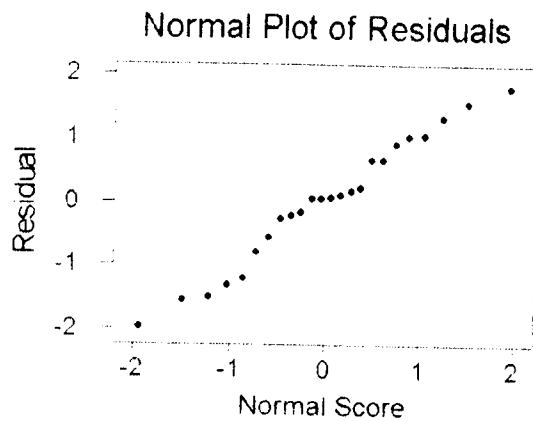


(c)

شکل ۳



ANALYSIS OF RESIDUALS



مراجع

- [1] BISSELL, A.F. (1992) lines through the origin - is NOINT the answer? , journal of Applied Statistics, vol 19, No .2, pp.193-210.
- [2] Montgomery, D.C. , Peck, E.A. (1992), Introduction to Linear Regression Analysis, 2nd ed, New York : wiley.
- [3] Weisberg, S. (1985), Applied linear Regression, 2nd ed, wiley, New York.