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

محمد شعبانی [1]

چکیده

برازش خطی که از نقطهٔ $(0, 0)$ می‌گذرد، مقوله‌ایست که بارها با آن در کاربرد رگرسیون خطی مواجه می‌شویم. بسیاری از نرم‌افزارهای کامپیوتری یک انتخاب حداقل مربعات دارند که بعضی اوقات بوسیلهٔ "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

موقعیتهای بسیاری وجود دارد که رابطهٔ بین دو متغیر خطی بوده و انتظار می‌رود که خط از نقطهٔ $x = 0, y = 0$ بگذرد. عرض از مبدأ صفر نسبت به نقطهٔ داده‌ای و یا 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 می شود. وقتیکه هیچ مادهٔ 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 بوده، به‌قسمی 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 رگرسین از طریق مبدأ وقتی رخ می‌دهد که خط از نقطهٔ $(0, 0)$ نگذرد. چنین وضعیتی در اشکال $1(a)$, $1(b)$, $1(c)$ ارایه شده است. البته ممکن است در

[1] دانشجوی کارشناسی ارشد، بخش آمار، دانشگاه شیراز

عمل چنین حالاتی رخ ندهند، اما اشکال (d)۱ ,(e)۱ 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 $x = ۰$ است $y \neq ۰$ 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 آشفتگی 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 $x = ۰$ نتوانیم $y = ۰$ را نتیجه بگیریم. در اشکال (a)۱ تا (e)۱ 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 نقطهٔ $(\bar{x}, \bar{y})$ 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 رابطهٔ 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 رابطهٔ غیر خطی کوچک باشد، ممکن است که غیر خطی بودن رابطه، کمتر آشکار گردد. اشکال (a)۲ ,(b)۲ این موضوع را به خوبی توضیح می‌دهند.

بر اساس اشکال (a)۲ ,(b)۲ دیده می‌شود که هم رابطهٔ واقعی و هم مدل بدون عرض از مبدأ برازش شده و مدل رگرسیون از طریق مبدأ، به رابطهٔ واقعی خیلی نزدیک است. در شکل (a)۲ رابطهٔ واقعی بفرم $y = ۱/۲۵x^{۰/۸}$ معرفی شده و در شکل (b)۲ این رابطه عبارت است از: $y = ۰/۸x^{\frac{۱}{۲}}$. در یک موقعیت عملی، ممکن است محسوس‌تر بودن رابطهٔ غیر خطی (یا مدل رگرسیون خطی ساده) بوسیلهٔ تغییرات باقیمانده‌ها پنهان گردد. لذا، مناسب است که فرضهای مدل رگرسیونی شامل همگونی واریانس خطاها را در نظر بگیریم. اگر واریانس خطاها در مدل عرض از مبدأ بخواهد پایدار بماند، باید وقتیکه x به سمت صفر میل می‌کند، تک تک مقادیر y که ممکن است مثبت یا منفی باشند، حول میانگین به‌قسمی تغییر کنند که عرض نوار حول خط رگرسیون پایدار بماند. در حالیکه ممکن است این مسئله در بسیاری از کاربردها، وقتیکه هم x وهم y لزوماً مثبت هستند، برقرار باشد، اما خواص توأم همگونی و عرض از مبدأ صفر با یکدیگر ناسازگارند. در این صورت وقتیکه به عرض از مبدأ صفر نزدیک می‌شویم، باقیمانده‌ها حول خط تمایل به کاهش دارند. اشکال (a)۳ ,(b)۳ ,(c)۳ 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 ارایه می‌دهند. یکی از خصوصیات مدلهای اصلاحی ارایه شده در این مقاله، قابل تفسیر بودن آنها با توجه به کاربردهای عملی می‌باشد.

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

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

مدل ممکنی 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 $z = a + bx + \varepsilon$ ارایه 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

داشت:

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

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

$$y = ax + bx^{۲} + x\varepsilon$$

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 جملهٔ خطای همگن خواهد شد.

۴.۳ ممکن است رابطهٔ $z - x$ 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 $\frac{y}{x}$ به‌جای z داریم:

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

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

رابطهٔ (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 رابطهٔ $y - x$ خطی نباشد و ضمناً هیچ‌یک از روابط $z - x$، $z\frac{1}{x}$ نیز برقرار نباشند. در چنین حالتی ممکن است مدلی به فرم:

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

مناسب باشد. اگر از طرفین رابطهٔ (۳)، تبدیل نمایی بگیریم و با فرض اینکه مقادیر ε کوچک باشند ($e^{\varepsilon} \approx ۱ + \varepsilon$)، در آن صورت مدلی به فرم $y = kx^{b}(۱ + \varepsilon)$ 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:

وقتیکه 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) اندازهٔ مناسبی جهت مقایسهٔ مدلها از نقطه نظر برازش بهتر به داده‌ها ارایه نمایند. در رگرسیون چندگانه (از جمله در چندجمله‌ایها)، $R^{۲}$(ضریب تعیین)، میانگین مربعات خطاها و جذر میانگین مربعات خطاها ($S = \sqrt{\text{MSE}}$) معیارهایی جهت مقایسهٔ مدلها می‌باشند. برای بعضی از مدلهای اصلاحی پیشنهادی در این مقاله، $R^{۲}$، r و حتی MSE نامناسب هستند. مثلاً برای مدل رگرسیونی از طریق مبدأ، $R^{۲}$ وجود ندارد و برای مدل نسبت ثابت که تنها یک متغیر $z = \frac{y}{x}$ دارد، $R^{۲}$ قابل تعریف نیستند.

حال سئوال 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:

$$r(y, x) = r(y, \hat{y}) \qquad (۴)$$

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 مقایسهٔ برازش مدلها، مثلاً مدل تبدیل یافته یا منحنی–خطی وار (curvilinear) با مدلهای دیگر و یا با مدل خطی می‌باشد. معهذا 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 همهٔ مدلهای خطی، منجر به رابطهٔ (۴) می‌گردند. روش دیگر برای مقایسهٔ برازش مدلها بر اساس میانگین مربهات خطاها (MSE) می‌باشد. به طور واضح برای مدلهای از نوع یکسان، مقایسهٔ 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 :

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

نمایش می‌دهیم و جذر آن را با γ معرفی می‌کنیم. در رابطهٔ (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 کیلوگرمی 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 بستهٔ 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، رابطهٔ آشکارتری بین اندازهٔ مایع و درصد جذب موجود است. برای اینکه بررسی کنیم که نتیجهٔ بدست آمده ناشی از وجود اثر عجیبی نمی‌باشد، ۱۶ بستهٔ 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 صفحهٔ ۲۸، جدول ۱ و جدول ۲ توجه بیشتری معطوف می‌کنیم. در صفحهٔ ۲۸ ابتدا نمودار y بر حسب x رسم شده که دلیلی بر رابطهٔ خطی داشته و عرض از مبدأ آن منفی می‌باشد. این نمودار بیانگر آن است که همراه با اندازهٔ بسته‌ها، تغییرات افزایش می‌یابد. نمودار $\frac{y}{x}$ برحسب 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 $\log y$ برحسب $\log x$ (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.

این نمودار، خطی بودن تقریبی را نشان داده و به ازای برد مشاهده شدهٔ مقادیرx، تغییرات پایداری (در$\frac{y}{x}$) ملاحظه می‌گردد.

در حالیکه رگرسیون y برحسب x منجر به خلاصهٔ رضایت بخشی $R^۲ = ۹۸.۶\%$ می‌گردد، اما نمودار $\frac{y}{x}$ در برابر $\frac{1}{x}$، وسیلهٔ بهتری برای ارزیابی تغییرات باقیمانده‌ها ارایه می‌دهد، زیرا باقیمانده‌ها برای رابطهٔ y بر حسب x از اعتبار کلی برخوردار نمی‌باشد. در حقیقت نمودار باقیمانده (Residnal vs.Fits) صفحهٔ ۲۹، بادبزنی به سمت راست (fan-shaped to right) می‌باشد. چهار مدل ارایه شده در جدول ۲ همگی خطی بوده، لذا $r(y,\hat{y})$ در انتخاب بین مدلها هیچ کمکی نخواهد کرد، چرا که این شاخص، طبق جدول ۲ برای همهٔ مدلها ۰/۹۹۳۰ گزارش شده‌است. چون به‌نظر می‌رسد که خطاها متناسب با x می‌باشند، لذا جذر میانگین مربعات خطای متناسب (γ)، اندازه‌ای برای مناسب بودن برازش مدلها ارایه می‌کند. جدول ۲، شاخص مزبور را برای هر یک از چهار مدل ممکن ارایه داده است. مدل با خطای متناسب نه تنها بهترین برازش را ارایه می‌دهد، بلکه پذیرفتنی‌تر می‌باشد، زیرا اینکه نسبت $\frac{y}{x}$ با افزایش x کاهش می‌یابد قابل تفسیرتر است تا بخواهیم مدل رگرسیون خطی ساده با عرض از مبدأ منفی y را بکار ببریم. به طور کلی، چون همگی مدلها خطی هستند، شاخص $r(y,\hat{y})$ در این مثال نامناسب بوده اما آمارهٔ خطای متناسب، به طور واضحی مدلهای بدون عرض از مبدأ و نسبت ثابت را از مدلهای رگرسیون خطی ساده و مدل با خطای متناسب جدا می‌سازد. برای درک بهتر، موضوع محاسبات مربوط به $r(y,\hat{y})$، γ برای مدل رگرسیون خطی ساده در صفحهٔ ۲۵ با استفاده از نرم‌افزار MINITAB تحت WINDOWS نشان داده شده است.

۶ نتیجه‌گیری

در این مقاله، چند مدل ممکن برای معرفی کردن روابط خطی تقریبی که از نقطهٔ (۰,۰) می‌گذرد، ارایه نمودیم. مسئلهٔ پایداری یا متناسب بودن تغییرات باقیمانده‌ها در نظر گرفته شد مدل‌های چندجمله‌ای، نمایی و گمپرتز، همچنین مدل‌هایی بر اساس تبدیلات پایدار کننده واریانس بدست می‌آیند، در این بحث گنجانده نشده‌اند، هرچند که می‌توانند به‌عنوان مدل‌های اصلاحی در نظر گرفته شوند. همچنین ممکن است روشهای دیگری برای آنالیز داده‌ها (نظیر آنالیز باقیمانده‌های مدل) بکار گرفته شوند. همچنین نشان دادیم که همبستگی مرسوم پیرسن، تنها شاخص جهت مقایسهٔ مدلهای اصلاحی نبوده و همبستگی بین متغیرهای وابسته و متغیرهای پیش‌بینی شده ($\hat{y}$) وسیله‌ای جهت مقایسهٔ مدلهای منحنی–خطی ارایه می‌دهد.

همچنین آماره‌ای که بر اساس میانگین مربعات خطای متناسب ارایه می‌گردد، در بعضی از موارد بطور سودمندی جایگزین MSE خواهد بود. ضمناً نکتهٔ مهمی در معرفی کردن مدلهای اصلاحی در این مقاله، قابل تفسیر بودن آنها در وضعیتهای علمی و کاربردی می‌باشد و مدلی که به شکل سیستم مربوط می‌گردد، ممکن است به مدل دیگری که حتی بهتر برازش می‌گردد ترجیح داده شود، چرا که آن مدل این پژوهش علمی را نقص می‌کند.

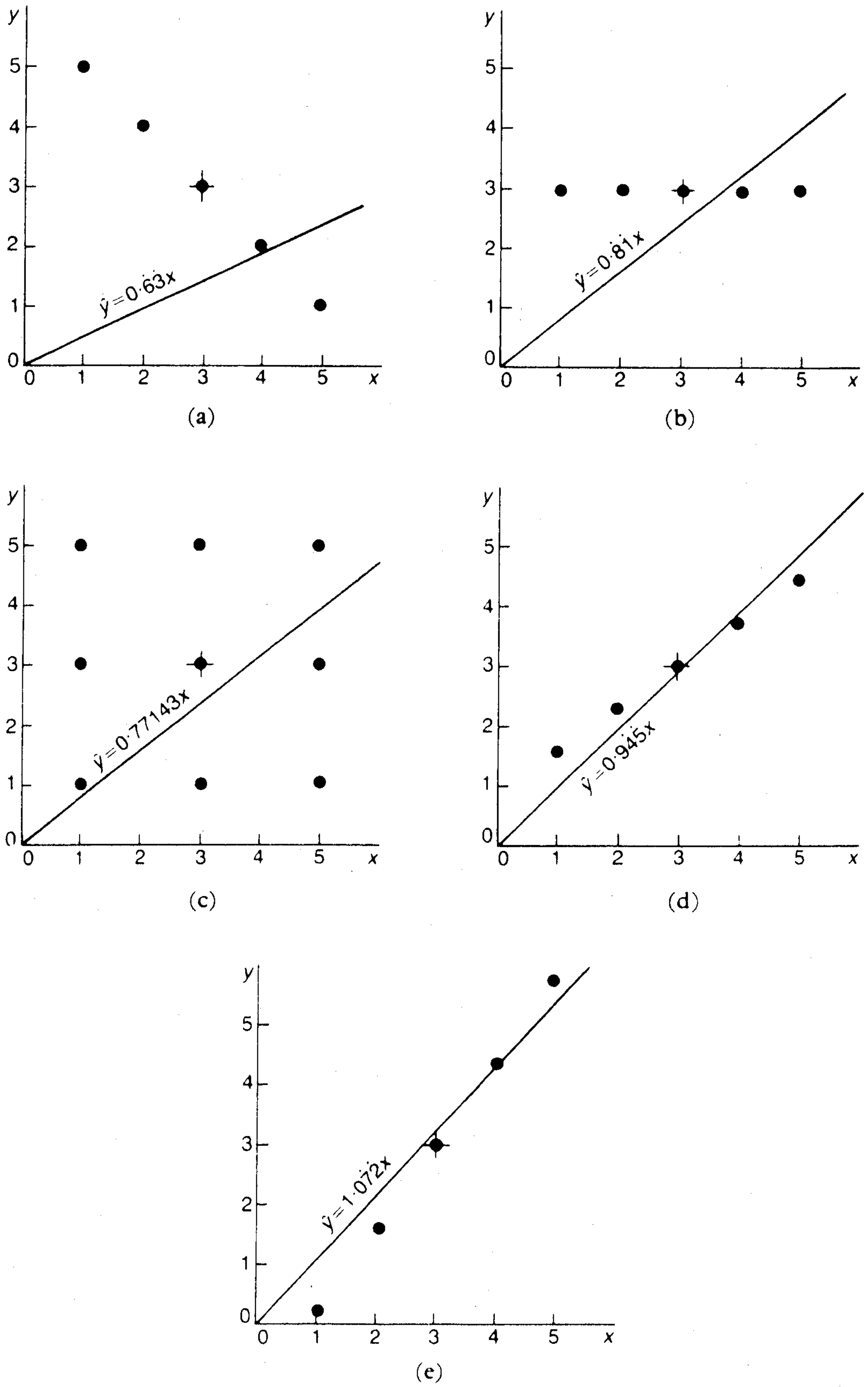
جدول ۱: جذب عناصر توسط مایع در پروسه جذب فرآیند

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	۳۳۰	۱۷/۱	۵/۱۸	۳۷۰	۲۱/۳	۵/۷۶
۴۹۰	۲۷/۲	۵/۵۵	۴۰۰	۲۰/۴	۵/۱۰	۴۵۰	۲۷/۴	۶/۰۹
۵۸۰	۳۱/۹	۵/۵۰	۵۶۰	۳۲/۵	۵/۸۰	۵۲۰	۲۸/۴	۵/۴۶
۶۵۰	۳۴/۱	۵/۲۵	۶۵۰	۳۹/۸	۶/۱۲	۶۵۰	۳۸/۵	۵/۹۲
۸۱۰	۵۰/۴	۶/۲۲	۸۰۰	۴۳/۸	۵/۴۸	۷۶۰	۵۰/۴	۶/۶۳
۱۰۲۰	۷۱/۳	۶/۹۹	۱۰۲۰	۶۴/۳	۶/۳۰	۹۱۰	۵۳/۵	۵/۸۸
۱۲۳۰	۷۸/۵	۶/۳۸	۱۲۰۰	۸۰/۸	۶/۷۳	۱۱۶۰	۷۹/۶	۶/۸۶
۱۴۹۰	۹۸/۶	۶/۶۲	۱۴۶۰	۱۰۵/۶	۷/۲۳	۱۳۸۰	۹۸/۹	۷/۱۷

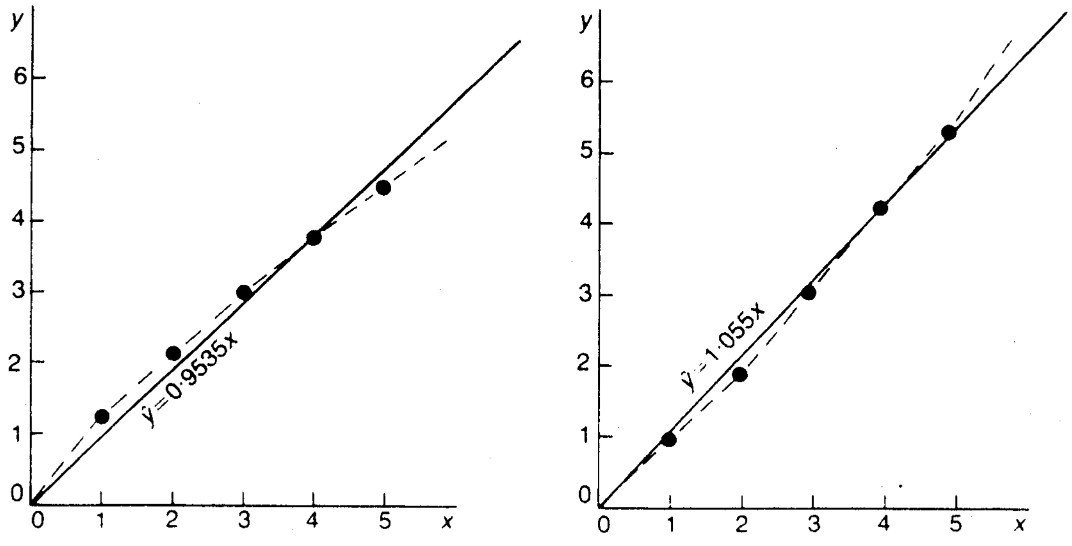
جدول ۲: مقایسه مدلهای خطی مربوط به داده‌های پروسهٔ جذب در جدول ۱

مدل خطی	شیب	عنصر از مبدأ y	عنصر از مبدأ x	$r(y,\hat{y})$	جذر میانگین مربعات خطای متناسب (γ)
مدل رگرسون 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
```

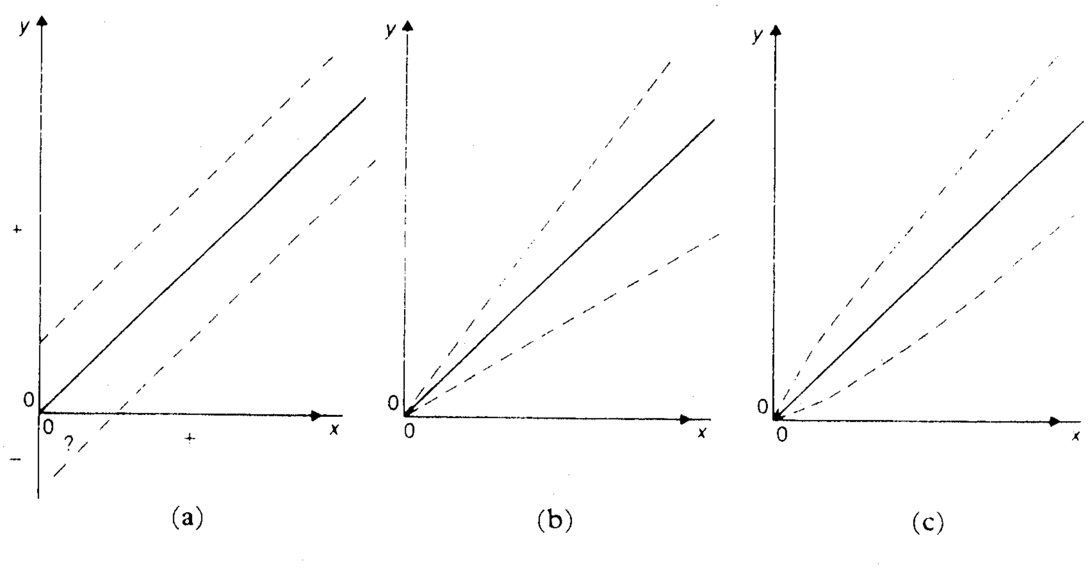


شکل ۱



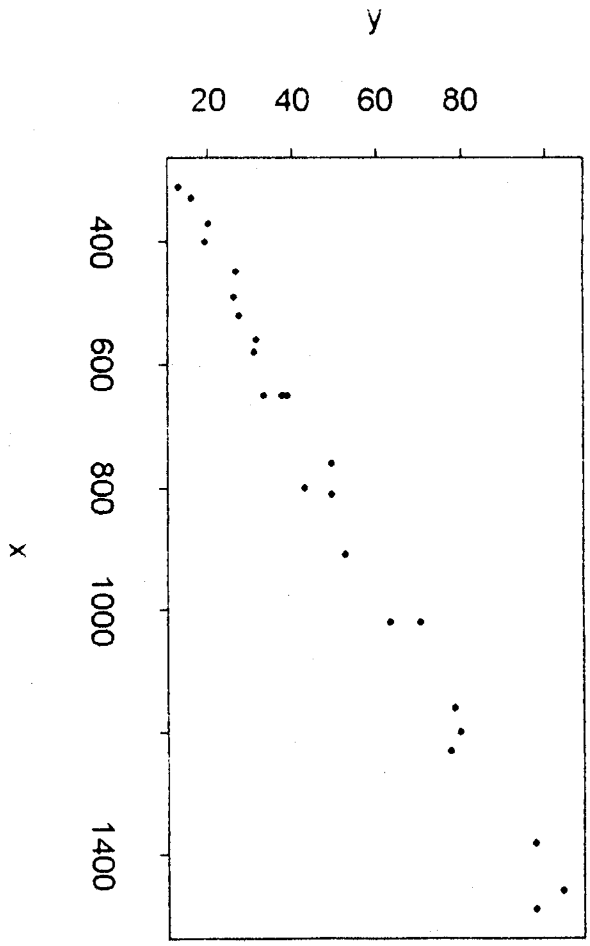
(a) (b)

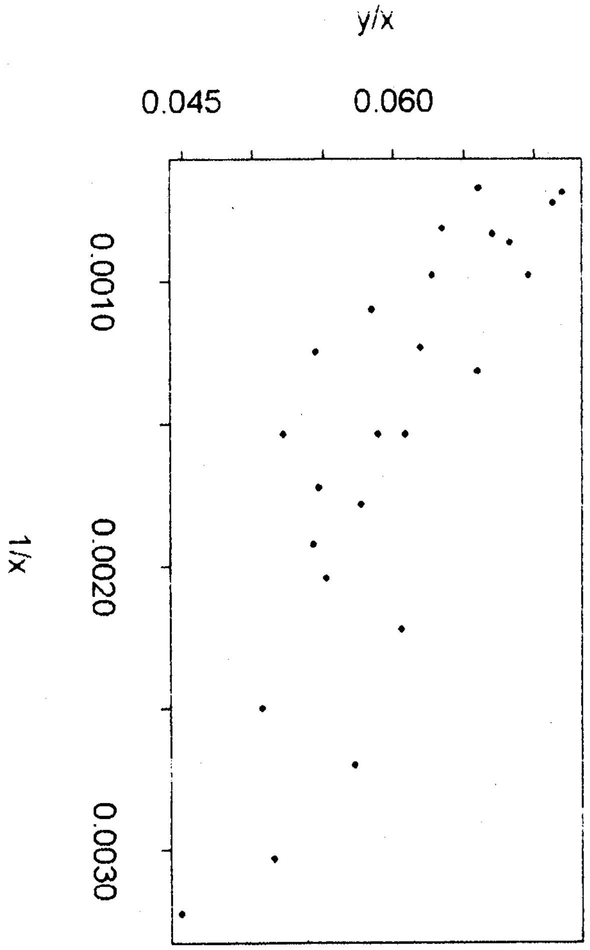
شکل ۲

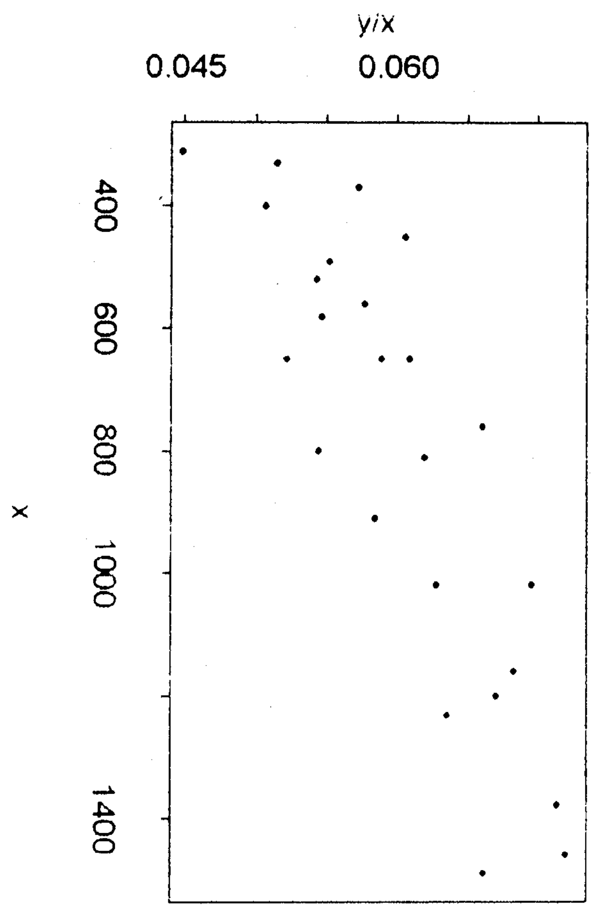


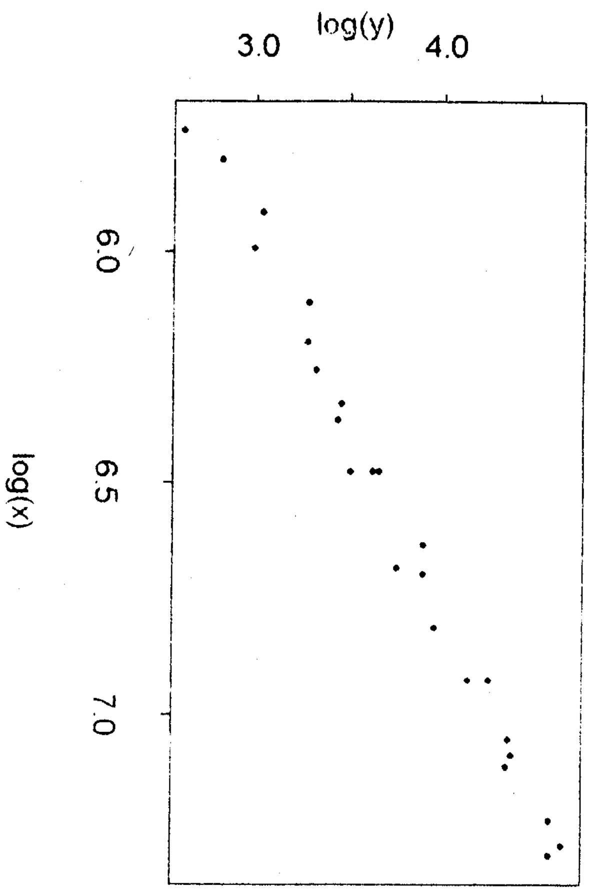
(a) (b) (c)

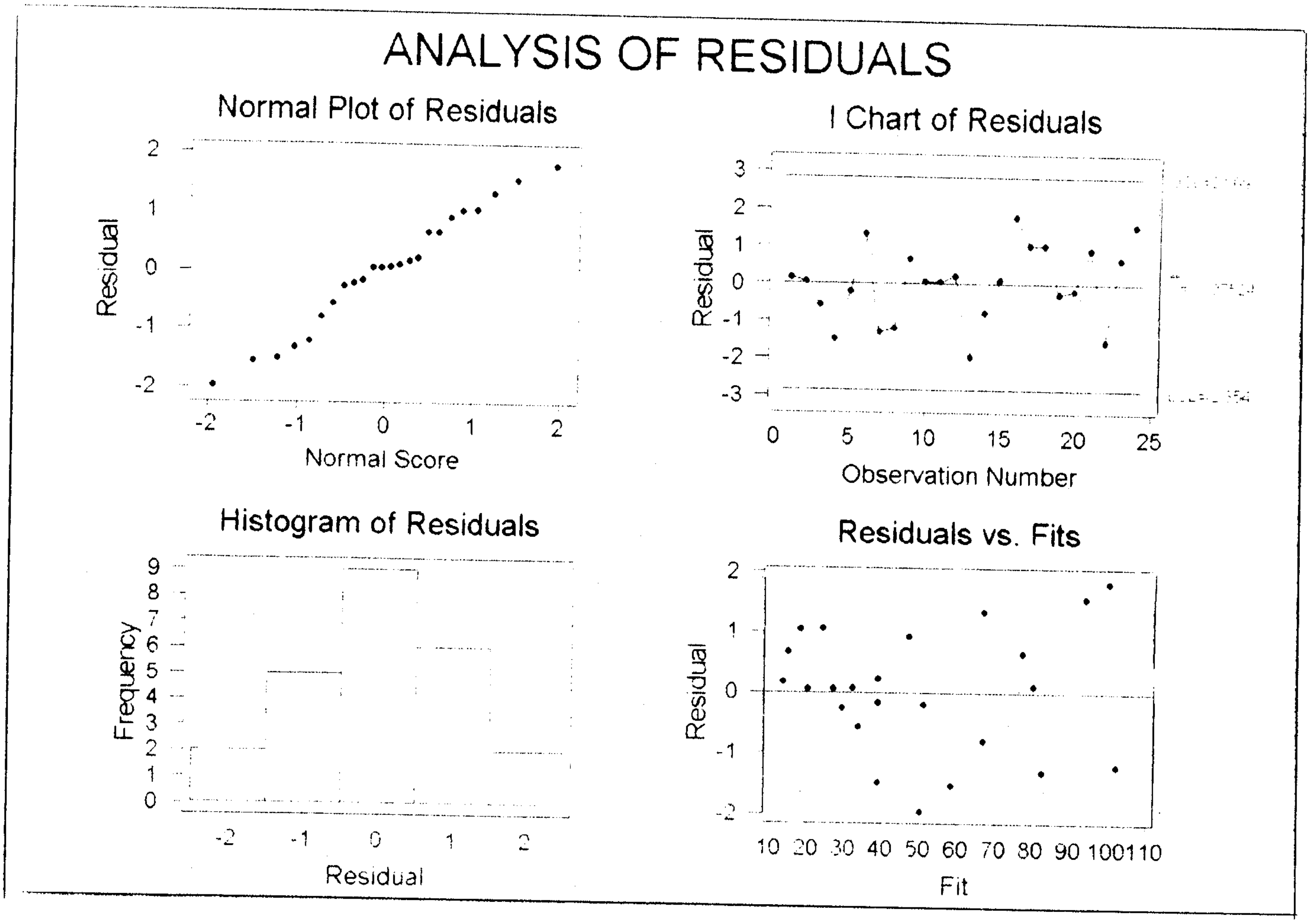
شکل ۳


y
20
40
60
80
x
400
600
800
1000
1400


y/x
0.045
0.060
1/x
0.0010
0.0020
0.0030


y/x
0.045
0.060
x
400
600
800
1000
1400


log(y)
3.0
4.0
log(x)
6.0
6.5
7.0



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.