

پیش‌بینی روند‌ها با استفاده از تجزیه و تحلیل رگرسیونی راهنمایی برای مدیران کتابخانه‌ها

نوشته حاضر راهنمایی‌های دقیقی را برای کاربرد «تجزیه و تحلیل رگرسیون خطی ساده»^۱ در پیش‌بینی روندهای خدمات کتابداری و اطلاع‌رسانی ارائه می‌کند. پیش‌بینی، عنصری بااهمیت در برنامه‌ریزی و تصمیم‌گیری خدمات کتابداری و اطلاع‌رسانی است و با حرفه کتابداری و اطلاع‌رسانی رابطه نزدیک دارد. حوزه‌های احتمالی کاربرد پیش‌بینی به قرار زیر است:

الف. رشد مواد کتابخانه

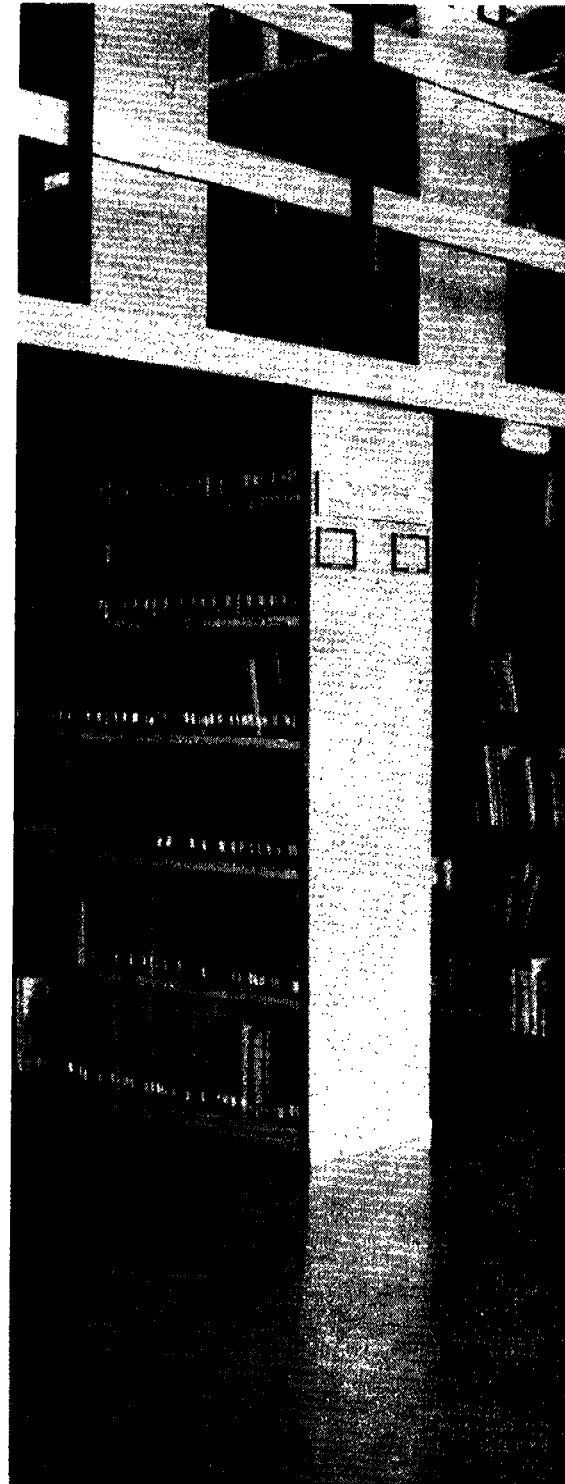
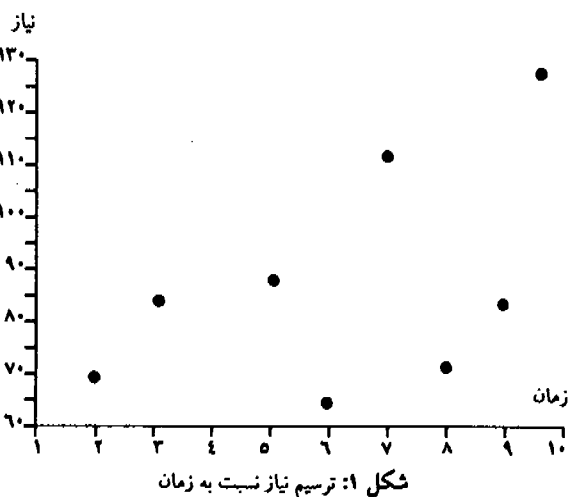
ب. هزینه مواد کتابخانه

ج. نیاز برای خدمات کتابداری و اطلاع‌رسانی

اطلاع از دانش مربوط به مفاهیم پایه در آمار ریاضی برای استفاده از تجزیه و تحلیل رگرسیونی لازم است. به همین منظور فهرستی از متون مقدماتی مناسب در انتهای نوشته آمده است. در این نوشته برای سهولت امر، کلیه ملزومات متفاوت برای پیش‌بینی، برحسب نوع نیاز پیش‌بینی توضیح داده خواهد شد.

یکی از فنون مشهور پیش‌بینی «تجزیه و تحلیل سری زمانی»^۲ است که تجزیه و تحلیل آماری نیازهای گذشته را برای تولید پیش‌بینی بکار می‌برد. پیش‌فرض اصلی این روش آن است که روندهای مفروض در آینده نیز ادامه خواهند یافت.

روش پیش‌بینی مورد بحث در این نوشته تحت عنوان «رگرسیون خطی ساده» شناخته شده است. اگر نیازهای گذشته را نسبت به زمان ترسیم کنیم (شکل ۱)، در پی یافتن خط مستقیمی هستیم که نیازها را به فواصل زمانی مربوط می‌سازد. مثلاً در عوض تلاش برای رسم تقریبی یک خط مستقیم، تجزیه و تحلیل رگرسیونی برای یافتن خط مستقیمی استفاده می‌شود که بیشترین تناسب را با نقطه‌های رسم شده داشته باشد. فرض می‌شود که نیازهای آینده بر روی این خط مستقیم قرار گیرند.



پیش فرض ها

مهمترین پیش فرض این است که روند بدست آمده از تجزیه و تحلیل نیازهای گذشته در زمان آینده نیز ادامه خواهد یافت. چنانچه شرایط اقتصادی به طور بارز تغییر یابد، بدیهی است که پیش بینی های مربوط به نیروی انسانی در یک حرفه نباید صرفاً متکی بر روندهای گذشته باشد. اگر عقیده بر این نیست که روندهای گذشته در آینده ادامه یابد، لازم است از پیش بینی بسیار پیچیده ای استفاده شود.

پیش فرض دیگر این است که رابطه میان زمان و نیاز را می توان به جای سیم منحنی کامل تر (شکل ۳)، با رسم خط مستقیم (شکل ۲) نشان داد.

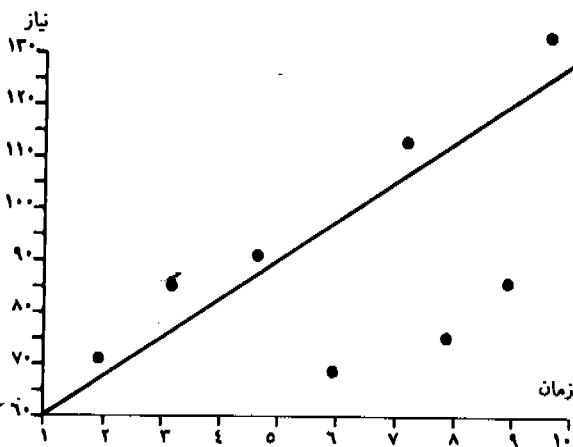
به علاوه، پیش فرضهای فنی دیگری نیز وجود دارند که بایستی مورد توجه قرار گیرند (صفحات بعد).

مثالها

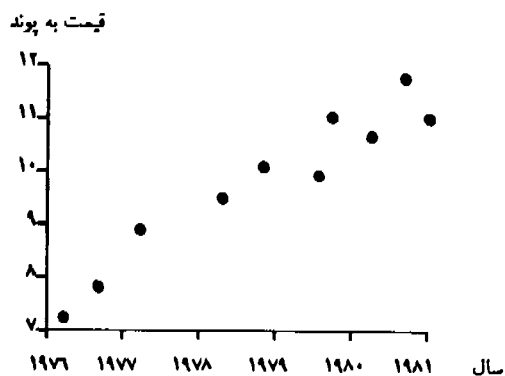
فنون تجزیه و تحلیل رگرسیونی به کمک دو مجموعه از داده های نمونه توضیح داده خواهند شد. جدول شماره ۱ میانگین کتابهای دانشگاهی بریتانیا را در هر شش ماه در طول سالهای ۱۹۷۶ الی ۱۹۸۰ نشان می دهد. نمودار آمار قیمت ها در شکل شماره ۴ ترسیم شده است. واضح است که جستجو برای ترسیم یک خط مستقیم مناسب کار عاقلانه ای است. لازم است که میانگین قیمت ها را برای شش ماه اول سال ۱۹۸۱ برآورد کنیم. جدول شماره ۲ تعداد بازدید سالانه از یک کتابخانه را در سالهای ۱۹۶۰ الی ۱۹۷۷ به دست می دهد. نمودار این اعداد در شکل شماره ۵ رسم گردیده است. در اینجا نیز بی مناسبت نیست که برای ترسیم خط مستقیم مناسبی تلاش کنیم. لازم است که تعداد بازدیدها را برای سالهای ۱۹۷۸-۱۹۸۷ برآورد نماییم. مجموعه سوم داده های نمونه برای انجام تمرین داده می شود.

جدول شماره ۱. میانگین قیمت کتابها در هر شش ماه (در ۱۹۷۶-۱۹۸۰) به پوند

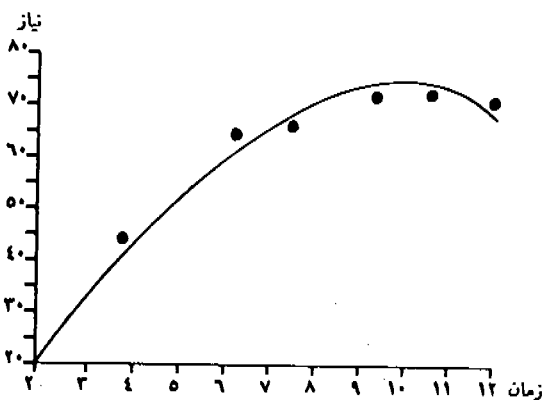
فاصله زمانی	قیمت
۷۶	۷/۲۱
۷۶/۵	۷/۷۱
۷۷	۸/۹۹
۷۷/۵	۹/۴۳
۷۸	۹/۹۳
۷۸/۵	۹/۹۷
۷۹	۱۱/۰۴
۷۹/۵	۱۰/۵۵
۸۰	۱۱/۷۵
۸۰/۵	۱۱/۳۵



شکل ۲: برازنده خط مستقیم برای نمودار نیازها نسبت به زمان

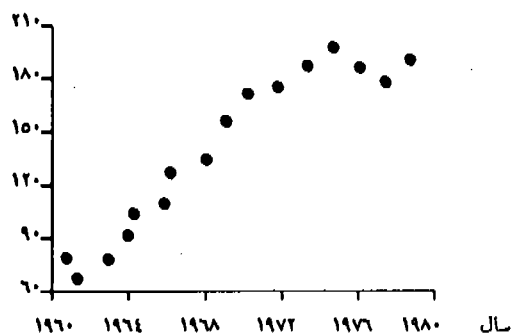


شکل شماره ۴. نمودار قیمت کتابهای دانشگاهی بریتانیا نسبت به زمان



شکل ۳: برازنده منحنی برای نمودار نیازها نسبت به زمان

بازدید (۱۰۰۰ مورد)



شکل شماره ۵. نمودار تعداد بازدید سالانه نسبت به زمان

جدول شماره ۳. میانگین قیمت کلیه کتابهای بریتانیا (در ۱۹۸۷-۱۹۸۱) به پوند

سال	فاصله زمانی	قیمت
۱۹۸۷	ژانویه - ژوئن	۷/۳۲
	جولای - دسامبر	۷/۳۱
۱۹۷۹	ژانویه - ژوئن	۷/۶۵
	جولای - دسامبر	۸/۰۶
۱۹۸۰	ژانویه - ژوئن	۸/۸۳
	جولای - دسامبر	۹/۳۵
۱۹۸۱	ژانویه - ژوئن	۹/۶۷
	جولای - دسامبر	۱۰/۴۱

محاسبات آماری

در انجام محاسبات آماری برای پیش‌بینی، اشاره به آخرین دوره به عنوان دوره صفر که اطلاعات مورد نیاز برای آن وجود دارد، مناسب است. دوره قبل از آن، دوره ۱-، سپس دوره ۲-، و مثلاً تا دوره ۱۷- ادامه می‌یابد. بدین ترتیب پیش‌بینی برای دوره بعدی، در حقیقت پیش‌بینی برای دوره ۱ خواهد بود. جدولی با مشخصات زیر (رجوع شود به جدولهای شماره ۴ و ۵) باید به ترتیب محاسبات آماری برای داده‌ها تکمیل گردد. در این جدول هر ردیف مربوط به دوره‌ای از زمان است و ستونها عبارتند از:

- | | |
|-----|--------------------|
| (۱) | سال |
| (۲) | دوره |
| (۳) | نیاز |
| (۴) | مربع دوره |
| (۵) | نیاز برای دوره x |
| (۶) | مربع نیاز |

تعداد دوره‌ها در جدول و مجموع ستونها نیز به شرح زیر لازم است:

- | | | |
|-----|--------------------------|---------------|
| (۱) | تعداد دوره‌ها | n |
| (۲) | مجموع دوره | مجموع (x) |
| (۳) | مجموع نیاز | مجموع (y) |
| (۴) | مجموع مربع دوره | مجموع (x^2) |
| (۵) | مجموع نیاز برای دوره x | مجموع (xy) |
| (۶) | مجموع مربع نیاز | مجموع (y^2) |

جدول شماره ۳ میانگین قیمت کلیه کتابهای بریتانیا را در هر شش ماه در طول سالهای ۱۹۷۸ الی ۱۹۸۱ نشان می‌دهد. لازم است که مطالعه کننده رساله در حین مطالعه هر فصل، از این مجموعه داده‌ها برای انجام تمرینات استفاده نماید.

جدول شماره ۴. تعداد ۱۰۰۰ مورد بازدید سالانه از کتابخانه (در ۱۹۷۷-۱۹۶۰)

فاصله زمانی	بازدیدها
۱۹۶۰	۷۱
۱۹۶۱	۶۳
۱۹۶۲	۸۰
۱۹۶۳	۸۹/۵
۱۹۶۴	۹۵
۱۹۶۵	۱۰۱
۱۹۶۶	۱۱۳/۵
۱۹۶۷	۱۱۷
۱۹۶۸	۱۳۵
۱۹۶۹	۱۴۹
۱۹۷۰	۱۵۶
۱۹۷۱	۱۶۸
۱۹۷۲	۱۶۹
۱۹۷۳	۱۷۱/۵
۱۹۷۴	۱۸۵
۱۹۷۵	۱۸۲/۵
۱۹۷۶	۱۷۶/۵
۱۹۷۷	۱۸۷/۵

جدول شماره ۴. محاسبات آماری برای داده‌های قیمت کتاب

سال	x	y	مربع x	xy	مربع y
(۱)	(۲)	(۳)	(۴)	(۵)	(۶)
۷۶	-۹	۷/۳۱	۸۱	-۶۵/۷۹	۵۳/۴۴
۷۶/۵	-۸	۷/۷۱	۶۴	-۶۱/۶۸	۵۹/۴۴
۷۷	-۷	۸/۹۹	۴۹	-۶۲/۹۳	۸۰/۸۲
۷۷/۵	-۶	۹/۴۳	۳۶	-۵۶/۵۸	۸۸/۹۳
۷۸	-۵	۹/۹۳	۲۵	-۴۹/۶۵	۹۸/۶۱
۷۸/۵	-۴	۹/۹۷	۱۶	-۳۹/۸۸	۹۹/۴۰
۷۹	-۳	۱۱/۰۴	۹	-۳۳/۱۲	۱۲۱/۸۸
۷۹/۵	-۲	۱۰/۵۵	۴	-۲۱/۱۰	۱۱۱/۳۰
۸۰	-۱	۱۱/۷۵	۱	-۱۱/۷۵	۱۳۸/۰۶
۸۰/۵	۰	۱۱/۳۵	۰	۰	۱۲۸/۸۲
جمع	-۴۵	۹۸/۰۳	۲۸۵	-۴۰۲/۴۸	۹۸۰/۷۰

در واقع، تعداد دو ستون از مجموع کل ستونها به راحتی به قرار زیر قابل محاسبه است:
(۲) مجموع دوره

$$\text{مجموع } (x) = \frac{n(n-1)}{2}$$

(۴) مجموع مربع دوره

$$\text{مربع } (x) \text{ مجموع} = \frac{n(n-1)(2n-1)}{6}$$

سه نوع آمار افزوده نیز به قرار زیر باید محاسبه گردد:

(۱) «مجموع صحیح مربع برای دوره‌ها»

$$S_{xx} = \frac{n(n-1)(n+1)}{12}$$

(۲) «مجموع صحیح حاصلضربها»

$$S_{xy} = \text{مجموع } (xy) \frac{(n-1) \times \text{مجموع } (y)}{2}$$

(۳) «جمع مجموع مربعات»

$$S_{yy} = \text{مجموع } (y) \times \text{مجموع } (y) \frac{\text{مجموع } (y)}{n} - \text{مربع } (y) \text{ مجموع}$$

جدول شماره ۵. محاسبات آماری برای داده‌های استفاده از کتابخانه.

سال	x	y	مربع x	xy	مربع y
(۱)	(۲)	(۳)	(۴)	(۵)	(۶)
۱۹۶۰	-۱۷	۷۱	۲۸۹	-۱۲۰۷	۵۰۴۱
۱۹۶۱	-۱۶	۶۳	۲۵۶	-۱۰۰۸	۳۹۶۹
۱۹۶۲	-۱۵	۸۰	۲۲۵	-۱۲۰۰	۶۴۰۰
۱۹۶۳	-۱۴	۸۹/۵	۱۹۶	-۱۲۵۳	۸۰۱۰/۲۵
۱۹۶۴	-۱۳	۹۵	۱۶۹	۱۲۳۵	۹۰۲۵
۱۹۶۵	-۱۲	۱۰۱	۱۴۴	-۱۲۱۲	۱۰۲۰۱
۱۹۶۶	-۱۱	۱۱۳/۵	۱۲۱	-۱۲۴۸/۵	۱۲۸۸۲/۲۵
۱۹۶۷	-۱۰	۱۱۷	۱۰۰	۱۱۷۰	۱۳۶۸۹
۱۹۶۸	-۹	۱۳۵	۸۱	-۱۲۱۵	۱۸۲۲۵
۱۹۶۹	-۸	۱۴۹	۶۴	-۱۱۹۲	۲۲۲۰۱
۱۹۷۰	-۷	۱۵۶	۴۹	-۱۰۹۲	۲۴۳۳۶
۱۹۷۱	-۶	۱۶۸	۳۶	-۱۰۰۸	۲۸۲۲۴
۱۹۷۲	-۵	۱۶۹	۲۵	-۸۴۵	۲۸۵۶۱
۱۹۷۳	-۴	۱۷۱/۵	۱۶	-۶۸۶	۲۹۴۱۲/۲۵
۱۹۷۴	-۳	۱۸۵	۹	-۵۵۵	۳۴۲۲۵
۱۹۷۵	-۲	۱۸۲/۵	۴	-۳۶۵	۳۳۳۰۶/۲۵
۱۹۷۶	-۱	۱۷۶/۵	۱	-۱۷۶/۵	۳۱۱۵۲/۲۵
۱۹۷۷	۰	۱۸۷/۵	۰	۰	۳۵۱۵۶/۲۵
جمع	-۱۵۳	۲۴۱۰	۱۷۸۵	-۱۶۶۶۸	۳۵۴۰۱۶

جدولهای شماره ۶ و ۷ محاسبات تصویری را نشان می‌دهند.

جدول شماره ۶. محاسبه مجموع مربع داده‌های قیمت کتاب

$$S_{xx} = \frac{n(n-1)(n+1)}{12} = \frac{10 \times 9 \times 11}{12} = 82/5 \quad (۱)$$

$$S_{xy} = \text{مجموع } (xy) + \frac{(n-1) \times \text{مجموع } (y)}{2} = \quad (۲)$$

$$-۴۰۲/۴۸ + \frac{۹۸/۰۳ \times ۹}{2} = ۳۸/۶۶$$

$$S_{yy} = \text{مجموع } (y) \times \text{مجموع } (y) \frac{\text{مجموع } (y)}{n} - \text{مربع } (y) \text{ مجموع} \quad (۳)$$

$$۹۸۰/۷۰ - \frac{۹۸/۰۳ \times ۹۸/۰۳}{10} = ۱۹/۷۱$$

جدول شماره ۷. محاسبه مجموع مربعات برای داده‌های استفاده از کتابخانه

$$S_{xx} = \frac{n(n-1)(n+1)}{12} = \quad (1)$$

$$\frac{18 \times 17 \times 19}{12} = 484/5$$

$$S_{xy} = \text{مجموع}(xy) + \frac{(n-1) \times \text{مجموع}(y)}{2} = \quad (2)$$

$$-16668 + \frac{17 \times 2410}{2} = 2817$$

$$S_{yy} = \text{مجموع}(y^2) - \frac{(\text{مجموع}(y))^2}{n} = \quad (3)$$

$$254016 - \frac{2410 \times 2410}{18} = 31243/78$$

معادله خط مستقیم

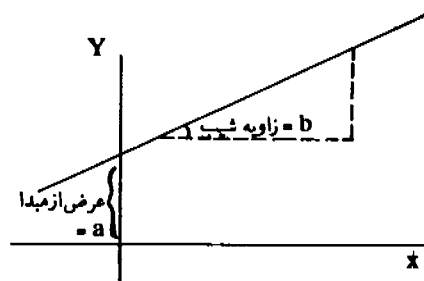
شکل معادله خطوط مستقیم عبارت است از $y = a + bx$ در معادله فوق، حروف a و b پارامترهایی هستند که موقعیت خط مزبور را تثبیت می‌کنند. پارامتر a به ازای $x = 0$ تنظیم‌کننده y و پارامتر b زاویه شیب خط است (شکل ۶).

آماري که برای داده‌ها محاسبه گردیده‌اند به منظور محاسبه مقادیر تخمینی پارامترهای a و b به کار می‌روند. آمار مذکور از طریق فرمولهای زیر به دست می‌آیند:

$$b = \frac{S_{xy}}{S_{xx}}$$

$$a = \frac{\text{مجموع}(y)}{n} + \frac{(n-1)b}{2}$$

محاسبات تصویری در جدولهای شماره ۸ و ۹ نشان داده شده‌اند. اکنون برازنده‌ترین خط مستقیم برای داده‌ها را می‌توان رسم کرد (شکلهای ۷ و ۸).



شکل شماره ۶. معادله خط مستقیم

جدول شماره ۸. محاسبه پارامترهای خط مستقیم برازنده برای داده‌های قیمت کتاب

$$b = \frac{S_{xy}}{S_{xx}} = \frac{38/66}{42/5} = 0/469$$

$$a = \frac{\text{مجموع}(y)}{n} + \frac{(n-1)b}{2} = \frac{98/03}{10} + \frac{9 \times 0/469}{2} = 11/912$$

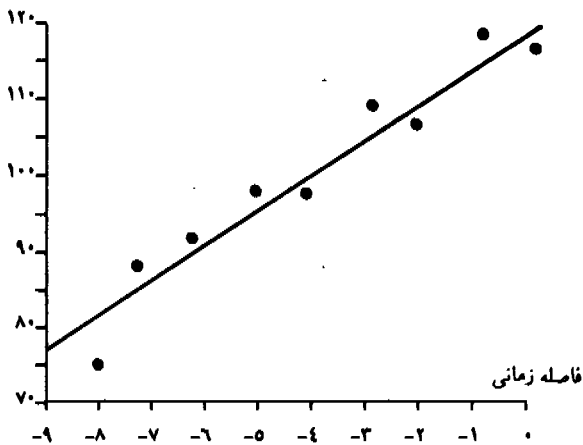
جدول شماره ۹. محاسبه پارامترهای خط مستقیم برازنده برای داده‌های استفاده از کتابخانه

$$b = \frac{S_{xy}}{S_{xx}} = \frac{2817}{484/5} = 7/88$$

$$a = \frac{\text{مجموع}(y)}{n} + \frac{(n-1)b}{2} = \frac{2410}{18} + \frac{17 \times 7/88}{2} =$$

$$133/889 + 66/960 = 200/9$$

۱۰ × میانگین قیمت



شکل شماره ۷. رسم خط مستقیم برازنده برای داده‌های قیمت کتاب

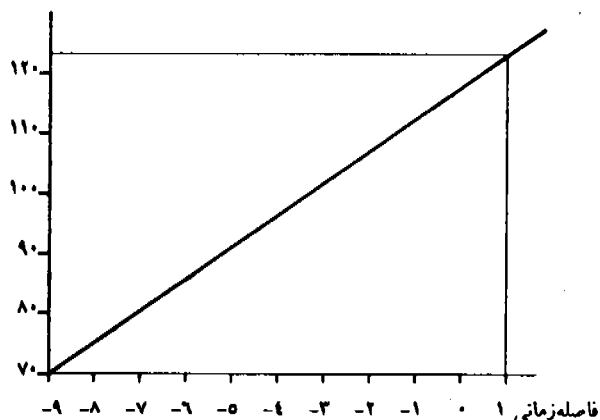
پیش‌بینی

نیازهای آینده بوسیله خط مستقیم برازنده با روند نیازها پیش‌بینی می‌شود. این کار ممکن است به وسیله نمودار صورت پذیرد (شکلهای شماره ۹ و ۱۰).

با وجود این، محاسبه نیاز پیش‌بینی شده بوسیله معادله خط مستقیم مورد اطمینان‌تر است. نیاز پیش‌بینی شده برای فاصله‌های متوالی x به وسیله معادله زیر محاسبه می‌گردد:

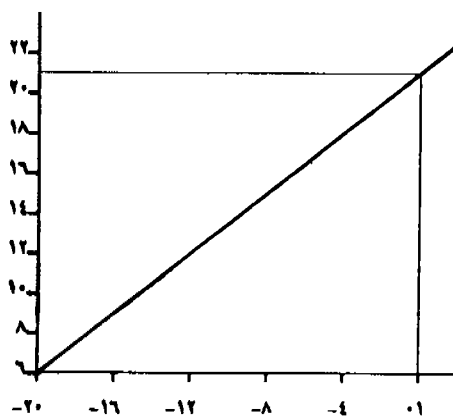
$$y = a + bx$$

۱۰ × میانگین قیمت



شکل شماره ۹. پیش‌بینی قیمت کتاب از طریق ترسیم خط مستقیم
برازنده با داده‌ها

بازدیدها
(هر ۱۰۰۰ مورد)



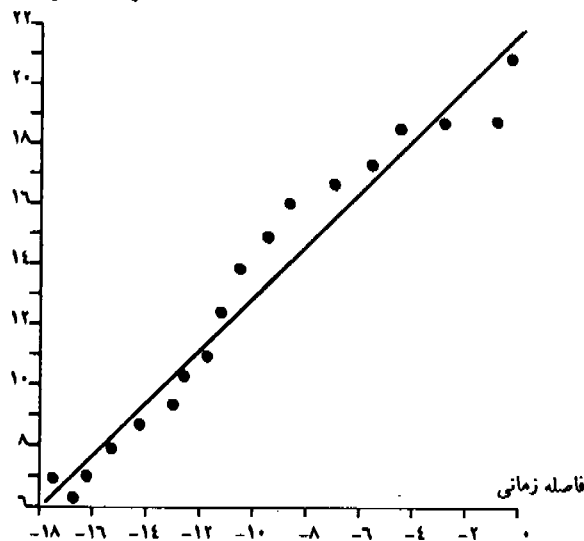
شکل شماره ۱۰. پیش‌بینی میزان استفاده از کتابخانه از طریق ترسیم خط
مستقیم برازنده با داده‌ها

باقیمانده‌ها

باقیمانده عبارت از اختلاف بین ارزش بدست آمده بوسیله خط مستقیم برازنده و ارزش نیاز واقعی در یک فاصله زمانی خاص است. باقیمانده‌های مربوط به داده‌های نمونه در جدولهای شماره ۱۲ و ۱۳ محاسبه شده‌اند.

باقیمانده‌ها امکان آزمون صحت برخی از پیش‌فرض‌ها را که با استفاده از تجزیه و تحلیل رگرسیونی ساده بوجود می‌آیند، فراهم می‌آورند. این کار عموماً با ترسیم نمودار باقیمانده‌ها نسبت به

تعداد بازدیدها
(هر ۱۰۰۰۰ مورد)



شکل شماره ۸. رسم خط مستقیم برازنده برای داده‌های قیمت کتاب

پیش‌بینی‌های مربوط به مثالها در جداول شماره ۱۰ و ۱۱ نشان داده شده‌اند.

جدول شماره ۱۰. پیش‌بینی قیمت کتاب از طریق معادله خط مستقیم

به ازای $x = 1$ خواهیم داشت:

$$y = a + bx = 11/912 + 0/469x =$$

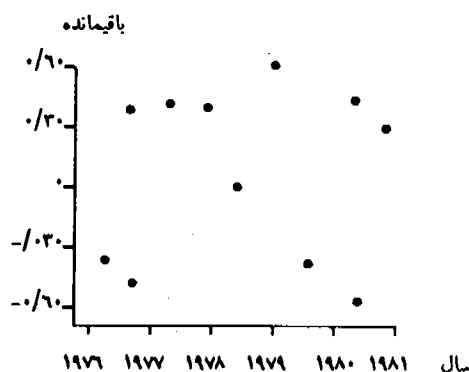
$$11/912 + 0/469 \times 1 = 11/912 + 0/469 = 12/38$$

جدول شماره ۱۱. پیش‌بینی میزان استفاده از کتابخانه از طریق معادله خط مستقیم

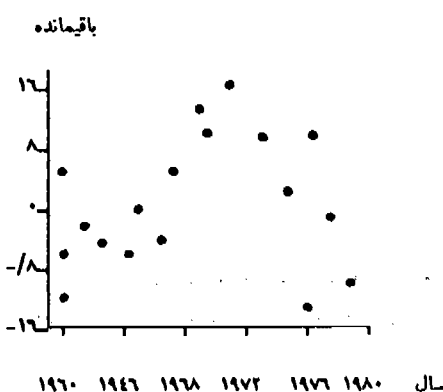
سال	پیش‌بینی
x	$y = a + bx = 200/9 + 7/88$
۱	۲۰۸/۸
۲	۲۱۶/۷
۳	۲۲۴/۵
۴	۲۳۲/۴
۵	۲۴۰/۳
۶	۲۴۸/۲
۷	۲۵۶/۱
۸	۲۶۳/۹
۹	۲۷۱/۸
۱۰	۲۷۹/۷

جدول شماره ۱۲. باقیمانده‌های مربوط به داده‌های قیمت کتاب

سال	x	y	پیش‌بینی	باقیمانده
۷۶	-۹	۷/۳۱	۷/۶۹	-۰/۳۸
۷۶/۵	-۸	۷/۷۱	۸/۱۶	-۰/۴۵
۷۷	-۷	۸/۹۹	۸/۶۳	-۰/۳۶
۷۷/۵	-۶	۹/۴۳	۹/۱۰	-۰/۳۳
۷۸	-۵	۹/۹۳	۹/۵۷	-۰/۳۶
۷۸/۵	-۴	۹/۹۷	۱۰/۰۴	-۰/۰۷
۷۹	-۳	۱۱/۰۴	۱۰/۵۱	-۰/۵۳
۷۹/۵	-۲	۱۰/۵۵	۱۰/۹۷	-۰/۴۲
۸۰	-۱	۱۱/۷۵	۱۱/۴۴	-۰/۳۱
۸۰/۵	۰	۱۱/۳۵	۱۱/۹۱	-۰/۵۶



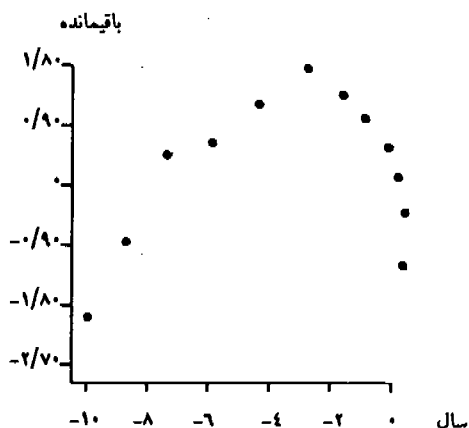
شکل شماره ۱۱. نمودار باقیمانده‌ها نسبت به زمان برای داده‌های قیمت کتاب



شکل شماره ۱۲. نمودار باقیمانده‌ها نسبت به زمان برای داده‌های استفاده از کتابخانه

جدول شماره ۱۳. باقیمانده‌های مربوط به داده‌های استفاده از کتابخانه

سال	x	y	پیش‌بینی	باقیمانده
۱۹۶۰	-۱۷	۷۱	۶۶/۹۴	۴/۰۶
۱۹۶۱	-۱۶	۶۳	۷۴/۸۲	-۱۱/۸۲
۱۹۶۲	-۱۵	۸۰	۸۲/۷۰	-۲/۷۰
۱۹۶۳	-۱۴	۸۹/۵	۹۰/۵۸	-۱/۰۸
۱۹۶۴	-۱۳	۹۵	۹۸/۴۶	-۳/۴۶
۱۹۶۵	-۱۲	۱۰۱	۱۰۶/۳۴	-۵/۳۴
۱۹۶۶	-۱۱	۱۱۳/۵	۱۱۴/۲۲	-۰/۷۲
۱۹۶۷	-۱۰	۱۱۷	۱۲۲/۱۰	-۵/۱۰
۱۹۶۸	-۹	۱۳۵	۱۲۹/۹۸	۵/۰۲
۱۹۶۹	-۸	۱۴۹	۱۳۷/۸۶	۱۱/۱۴
۱۹۷۰	-۷	۱۵۶	۱۴۵/۷۴	۱۰/۲۶
۱۹۷۱	-۶	۱۶۸	۱۵۳/۶۲	۱۴/۳۸
۱۹۷۲	-۵	۱۶۹	۱۶۱/۵۰	۷/۵۰
۱۹۷۳	-۴	۱۷۱/۵	۱۶۹/۳۸	۲/۱۲
۱۹۷۴	-۳	۱۸۵	۱۷۷/۲۶	۷/۷۴
۱۹۷۵	-۲	۱۸۲/۵	۱۸۵/۱۴	-۲/۶۴
۱۹۷۶	-۱	۱۷۶/۵	۱۹۳/۰۲	-۱۶/۵۲
۱۹۷۷	۰	۱۸۷/۵	۲۰۰/۹۰	-۱۳/۴۰



شکل شماره ۱۳. نمودار باقیمانده‌ها برای منحنی تغییرات خط منحنی در نیاز

فاصله زمانی انجام می‌گیرد (شکلهای شماره ۱۱ و ۱۲).

چنین فرض شده است که تغییرات نیاز به جای منحنی بودن، خطی باشند. رابطه‌ای خط منحنی بوسیله ترسیم مشخص می‌شود که در شکل شماره ۱۳ نشان داده شده است.

پیش‌فرض مهم دیگر این است که خطاها دارای واریانس مشابهی هستند. این مطلب با رسم یک منحنی که در شکل شماره

پیش‌بینی سودمند است. یکی از روشهای مشخص کردن درجه صحت پیش‌بینی محاسبه «خطای استاندارد»^۵ آن است. ابتدا ضروری است که «خطای استاندارد رگرسیون»^۶ محاسبه گردد، مقدار خطا از فرمول زیر به دست می‌آید:

$$Ser = \left(\frac{S_{yy} - \frac{S_{xy} \times S_{xy}}{S_{xx}}}{n - 2} \right)^{\frac{1}{2}} =$$

(به جدولهای شماره ۱۴ و ۱۵ رجوع شود).

سپس برای هر کدام از فواصل زمانی x که درجه صحت آن مورد نیاز است، خطای استاندارد پیش‌بینی با استفاده از فرمول زیر محاسبه می‌گردد:

$$Sep = Ser \times \left(1 + \frac{1}{n} + \frac{\left(x + \frac{n-1}{2}\right) \left(x + \frac{n-1}{2}\right)}{S_{xx}} \right)^{\frac{1}{2}} =$$

(به جدولهای شماره ۱۶ و ۱۷ رجوع شود).

بزرگی اعداد خطای استاندارد پیش‌بینی نشانگر پایین بودن درجه صحت پیش‌بینی برای نیاز آینده است.

جدول شماره ۱۴. محاسبه خطای استاندارد رگرسیون برای داده‌های قیمت کتاب

$$Ser = \left(\frac{S_{yy} - \frac{S_{xy} \times S_{xy}}{S_{xx}}}{n - 2} \right)^{\frac{1}{2}} =$$

$$\left(\frac{19.71 - \frac{29.99 \times 29.99}{17.0}}{15} \right)^{\frac{1}{2}} =$$

$$\left(\frac{19.71 - 52.99}{15} \right)^{\frac{1}{2}} =$$

$$\left(\frac{-33.28}{15} \right)^{\frac{1}{2}} =$$

$$(-2.219)^{\frac{1}{2}} =$$

جدول شماره ۱۵. محاسبه خطای استاندارد رگرسیون برای داده‌های استفاده از کتابخانه

$$Ser = \left(\frac{S_{yy} - \frac{S_{xy} \times S_{xy}}{S_{xx}}}{n - 2} \right)^{\frac{1}{2}} =$$

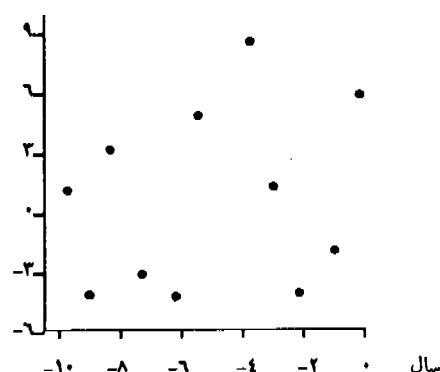
$$\left(\frac{3.1332 \times 10^{-4} - \frac{2.14 \times 2.14}{16}}{14} \right)^{\frac{1}{2}} =$$

$$\left(\frac{3.1332 \times 10^{-4} - 0.00029125}{14} \right)^{\frac{1}{2}} =$$

$$\left(\frac{1.2206 \times 10^{-4}}{14} \right)^{\frac{1}{2}} =$$

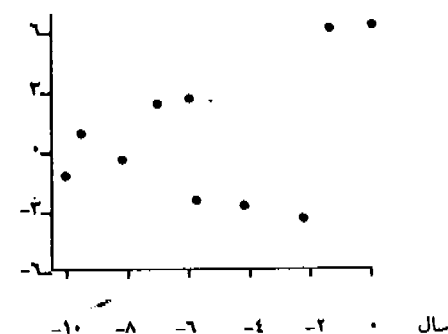
$$(8.72 \times 10^{-3})^{\frac{1}{2}} =$$

باقیمانده



شکل شماره ۱۴. نمودار باقیمانده‌ها برای خطاهای واریانس مساوی

باقیمانده



شکل شماره ۱۵. نمودار باقیمانده‌ها برای خطاهای واریانس متغیر

۱۴ نشان داده شده است، معین خواهد شد. اگر خطاها دارای واریانس واقعی باشند، منجر به رسم شکلی خواهند شد که در شکل شماره ۱۵ نشان داده شده است.

آزمون سایر پیش‌فرض‌های لازم برای بکارگیری تجزیه و تحلیل رگرسیونی ساده (که در آن خطاها ناهمبسته بوده و توزیع عادی داشته باشند) استفاده از تجزیه و تحلیل بسیار پیچیده‌ای نظیر (محاسبه آماری دوربین-واتسون^۳ و نمودار رانکیت^۴ که در بخش ارجاعات مورد بحث قرار گرفته است) را ضروری می‌سازد.

اگر پیش‌فرضهای لازم را برای کاربرد تجزیه و تحلیل رگرسیونی خطی ساده کامل و قانع‌کننده نباشند، می‌توان روش کار را به شکلی تغییر داد که از عهده تخطی از پیش‌فرضهای معینی برآید (— ارجاعات).

خطاهای استاندارد

پس از پیش‌بینی نیاز، آگاهی از درجه صحت

جدول شماره ۱۶. محاسبه خطای استاندارد پیش‌بینی برای داده‌های قیمت کتاب

$$Sep = Ser \times \left[1 + \frac{1}{n} + \frac{\left(x + \frac{n-1}{2} \right) \left(x + \frac{n-1}{2} \right)}{S_{xx}} \right]^{\frac{1}{2}} =$$

$$= 0.226 \times \left[1 + \frac{1}{10} + \frac{(x + 2/5)(x + 2/5)}{82/5} \right]^{\frac{1}{2}}$$

$$= 0.226 \left[1/1 + \frac{(x + 2/5)(x + 2/5)}{82/5} \right]^{\frac{1}{2}}$$

$$Sep = 0.226 \times \left[1/1 + \frac{0/5 \times 0/5}{82/5} \right]^{\frac{1}{2}} =$$

$$= 0.226 \times (1/1 + 0.267)^{\frac{1}{2}} =$$

$$= 0.226 \times (1/267)^{\frac{1}{2}} =$$

$$= 0.226 \times 1/21 =$$

$$= 0.050.$$

جدول شماره ۱۷. محاسبه خطای استاندارد پیش‌بینی برای داده‌های استفاده از کتابخانه

$$Sep = Ser \times \left[1 + \frac{1}{n} + \frac{\left(x + \frac{n-1}{2} \right) \left(x + \frac{n-1}{2} \right)}{S_{xx}} \right]^{\frac{1}{2}} =$$

$$= 0.92 \times \left[1 + \frac{1}{18} + \frac{(x + 1/5)(x + 1/5)}{484/5} \right]^{\frac{1}{2}} =$$

$$= 0.92 \times \left[1/0.56 + \frac{(x + 1/5)(x + 1/5)}{484/5} \right]^{\frac{1}{2}} =$$

فواصل اطمینان^۷

یکی از روشهای بسیار مفید برای تعیین درجه صحت پیش‌بینی، محاسبه فاصله اطمینان است، یعنی حدود ارزشهایی که نیاز اصلی محتمل است در آن حدود قرار گیرد.

جدول شماره ۱۸. ارزشهای تصویری^۸ t

n	درجات آزادی (n - 2)	t
5	3	3/182
10	8	2/306
15	13	2/160
20	18	2/101
25	23	2/069
30	28	2/048

دو نوع فاصله اطمینان وجود دارد که می‌توان برای پیش‌بینیها محاسبه نمود. نوع اول هنگامی به وجود می‌آید که مایل به پیش‌بینی فاصله زمانی واحدی^۸ باشیم (نظیر پیش‌بینی واحد برای قیمت کتابها در فاصله زمانی آینده). فاصله اطمینان برای پیش‌بینی از فرمول زیر به دست می‌آید:

(خطای استاندارد پیش‌بینی $t \times x \pm$ نیاز پیش‌بینی شده عامل t از جدولهای «توزیع t برای دانشجویان»^۳ استخراج می‌شود. ارزش صحیح « t » برای استفاده، به مقدار n بستگی دارد؛ مهم‌تر، جدولها به صورت قراردادی نه برحسب اندازه نمونه بلکه بسته به درجات آزادی^{۱۰} توصیف شده‌اند که در آن درجات آزادی عبارت است از:

$$n - 2 = \text{درجات آزادی}$$

ارزشهای تصویری « t » در جدول شماره ۱۸ ارائه شده است. محاسبه فاصله اطمینان هنگامی مناسب است که

جدول شماره ۱۹. محاسبه فاصله اطمینان پیش‌بینی واحد برای داده‌های قیمت کتاب

$$(t \times Sep) \pm \text{نیاز پیش‌بینی شده}$$

به ازای $x = 1$ داریم:

$$(جدول ۱۰) 12/38 = \text{نیاز پیش‌بینی شده}$$

$$t = 2/306 \text{ (جدول ۱۸)}$$

$$Sep = 0/050 \text{ (جدول ۱۶)}$$

فاصله اطمینان عبارت است از:

$$12/38 \pm (2/306 \times 0/050) =$$

$$12/38 \pm 1/250 =$$

$$11/13 \text{ الی } 13/63$$

چنانچه شرایط اقتصادی به طور بارز تغییر یابد، بدیهی است که پیش‌بینی‌های مربوط به نیروی انسانی در یک حرفه نباید صرفاً متکی بر روندهای گذشته باشد.

منابع

- دانستن مفاهیم اصلی آمار ریاضی در کاربرد «تجزیه و تحلیل رگرسیونی» لازم است. متن مقدماتی زیر در این زمینه برای کتابداران مفید است:
Cooper, A. *An introduction to Statistics for Librarians*. Loughborough: Centre for Library and Information Management 1982.
- کاربرد تجزیه و تحلیل رگرسیونی در منبع زیر مورد بحث قرار گرفته است:
Weisberg, S. *Applied Linear regression*. New York; Chichester Wiley, 1980.
- سایر فنون پیش‌بینی در منابع زیر مورد بحث قرار گرفته‌اند:
Makridakis, S. and Wheelwright, S.C. *Forecasting: methods and applications*. Santa Barbara, California; Chichester: Wiley, 1978.
- Thomopoulos, N. T. *Applied forecasting methods*. Englewood Cliffs, N. J. Prentice-Hall, 1980.
- جدولهای توزیع T و F برای مثال در منبع زیر قابل دسترسی است:
Lindley, D.V. and Miller, J.C.P. *Cambridge elementary Statistical tables*. Cambridge: Cambridge University Press, 1961.
- تعدادی از بسته‌های نرم‌افزاری کامپیوتر در زمینه آمار، تهبلائی را برای تجزیه و تحلیل رگرسیونی فراهم می‌آورند. مشخصات بسته‌های SPSS و Minitab در زیر آمده است:
Nie, N.H. (and others). *SPSS: Statistical Package for the Social Sciences*. 2nd ed. New York; London: McGraw-Hill. 1975. *SPSS Update 7-9*. New York; London: McGraw-Hill. 1981. RYAN, T.J. *Minitab: a Student handbook*. North Scituate, Mass.: Duxbury Press. 1976.

توضیحات:

1. Simple Linear regression analysis
2. Time-Series analysis
3. Durbin-Watson Statistic
4. Rankit Plotting
5. Standard error
6. Standard error of regression
7. Confidence intervals
8. Single time period
9. Students t-distribution
10. Degrees of freedom
11. F-distribution

جدول شماره ۲۰. ارزشهای تصویری F

n	درجات آزادی (n - 2)	F
5	3 و 2	9/00
10	8 و 2	4/46
15	13 و 2	3/80
18	16 و 2	3/63
20	18 و 2	3/50
25	23 و 2	3/42
30	28 و 2	3/34

گروهی از پیش‌بینی‌های همزمان برای فاصله‌هایی از زمان ضروری باشد (مثلاً آمار استفاده در مورد ۱ تا ۱۰ سال گذشته). در این صورت فاصله اطمینان از فرمول زیر به دست می‌آید:
(خطای استاندارد پیش‌بینی $\pm \frac{1}{2} (2F)$) نیاز پیش‌بینی شده مقدار حرف F از جدولهای «توزیع F» ۱۱ با درجات آزادی ۲ و (n - 2) استخراج می‌شود. برخی از ارزشهای تصویری F در جدول شماره ۲۰ نشان داده شده‌اند.
در ضمن، محاسبه فواصل پیش‌بینی همزمان در جدول شماره ۲۱ نشان داده شده است.

جدول شماره ۲۱. محاسبه فواصل پیش‌بینی همزمان برای داده‌های استفاده از کتابخانه

$$(\text{Sep}) \times \frac{1}{2} (2F) \pm \text{نیاز پیش‌بینی شده}$$

(با استفاده از جدول ۲۰ یعنی به ازای $F = 3/63$ فواصل پیش‌بینی برای مقادیر X به شرح زیر است):

X	پیش‌بینی (جدول ۱۱)	Sep (جدول ۱۷)	فواصل پیش‌بینی
۱	۲۰۸/۸	۹/۹۴	۱۸۲ الی ۲۳۵/۶
۲	۲۱۶/۷	۱۰/۱۱	۱۸۹/۴ الی ۲۴۳/۹
۳	۲۲۴/۵	۱۰/۲۸	۱۹۶/۸ الی ۲۵۲/۲
۴	۲۳۲/۴	۱۰/۴۷	۲۰۴/۲ الی ۲۶۰/۶
۵	۲۴۰/۳	۱۰/۶۷	۲۱۱/۵ الی ۲۶۹/۱
۶	۲۴۸/۲	۱۰/۸۹	۲۱۸/۸ الی ۲۷۷/۵
۷	۲۵۶/۱	۱۱/۱۱	۲۲۶/۱ الی ۲۸۶
۸	۲۶۳/۹	۱۱/۳۵	۲۳۳/۴ الی ۲۹۴/۵
۹	۲۷۱/۸	۱۱/۵۹	۲۴۰/۶ الی ۳۰۳
۱۰	۲۷۹/۷	۱۱/۸۴	۲۴۷/۸ الی ۳۱۱/۶