



بررسی روشهای تجزیه و تحلیل رگرسیون

حسن منصوری

دانش آموخته کارشناسی ارشد حسابداری
مدرس دانشگاه علمی کاربردی استان البرز

فاطمه سادات کهکی

دانشجوی کارشناسی ارشد حسابداری

چکیده:

پژوهشهای حسابداری عموماً با استفاده از اطلاعات حسابداری شرکت‌های پذیرفته شده در بورس اوراق بهادار انجام می‌پذیرد. نتایج چنین تحقیقاتی می‌تواند به عنوان مبنا و راهکاری جهت ارزیابی و تصمیم‌گیری سرمایه‌گذاران و سایر افراد ذینفع در بورس اوراق بهادار قلمداد شود. لذا استفاده از یک روش آماری مناسب جهت تحلیل داده‌ها امری ضروری می‌باشد. تحلیل رگرسیونی به عنوان روشی کاربردی جهت تحلیل داده‌ها، در بسیاری از پژوهش‌های حسابداری مورد استفاده قرار گرفته است. شناخت دقیق انواع تحلیل‌های رگرسیونی و استفاده مناسب و بجا از آنها می‌تواند نتایج مفیدتر و اثربخش‌تری را به همراه داشته باشد. پژوهش حاضر به بررسی روش-های رگرسیونی و به کار بردن داده‌های حسابداری و نتایج حاصل از آنها در حوزه داده‌های حسابداری می‌پردازد و به دنبال ارائه جواب-های صحیح و به جا از این روش آماری در تحقیقات حسابداری می‌باشد.

کلمات کلیدی: رگرسیون، ضریب تعیین، واریانس،



۱- مقدمه:

تحلیل رگرسیونی، روشی آماری جهت بررسی رابطه بین متغیرها و به طور کلی پژوهش‌های علی است. در این روش رابطه بین متغیر یا متغیرهای مستقل با متغیر وابسته نشان داده می‌شود ساده ترین مدل رگرسیون، مدل رگرسیون خطی است که تنها شامل یک متغیر مستقل و یک متغیر وابسته با رابطه خطی می‌باشد. این مدل اولین بار توسط ودربورن و فلدر در سال ۱۹۷۲ معرفی شد و مبنای تحلیل رگرسیونی قرار گرفت دیگر مدل‌های رگرسیون شامل لجستیک، غیرخطی و ...، حالت توسعه یافته مدل رگرسیون خطی ساده می‌باشد. مدل‌های رگرسیون را می‌توان با توجه به داده‌های مورد بررسی، به انواع مختلفی تقسیم کرد. این مدل‌های رگرسیون عبارتند از: رگرسیون سری زمانی، رگرسیون مقطعی، رگرسیون میانگین و رگرسیون تجمعی.

۲- مبنای نظری تحقیق:

روشهای مقداری:

روش های مقداری به منظور برآورد ارقام بهای تمام شده و برنامه ریزی و بودجه بندی به نحو گسترده مورد استفاده قرار می گیرد نکته مهم در بکار گیری این روش ها این است که این روش ها ابزارهایی برای حل مسائل هستند و در استفاده از این روش ها باید حسابداران از قضاوت و تجربه حرفه ای نیز استفاده کنند چون این روش ها بیان کننده و در برگیرنده تمام واقعیت های موجود نمی باشند. اساس این روش ها بر پایه استفاده از اطلاعات گذشته برای بررسی رفتار هزینه ها و استفاده از این روند برای مقاصد پیش بینی و برنامه ریزی آتی است.

هدف و کاربرد روش های برآورد هزینه :

۱- بودجه بندی ۲- خرید یا ساخت از بیرون ۳- مبنای روش های مدیریتی چون هزینه یابی بر مبنای هدف

شش گام برای برآورد هزینه:

- ۱- تعریف موضوع هزینه
- ۲- تعیین عوامل ایجاد کننده هزینه
- ۳- جمع آوری داده ها
- ۴- رسم نمودار براساس داده ها
- ۵- انتخاب و به کارگیری روش برآورد هزینه
- ۶- ارزیابی صحت و دقت داده های برآوردی

روش های برآورد هزینه و تعیین گرایش ارقام هزینه:

۱- روش بالاترین و پایین ترین حد

در این روش با استفاده از حد بالا و پایین هزینه، معادله خط هزینه به شرح ذیل را بدست می آوریم.

$$y=a+bx$$

که در این رابطه

y = مقدار هزینه برآوردی

x = عامل ایجاد کننده هزینه (محرک هزینه ۹

a = مقدار ثابت برابر مقدار y وقتی x صفر است.

b = شیب خط

که در این رابطه مقدار b با استفاده از فرمول ذیل محاسبه می شود:

تفاوت بین عامل هزینه بالاترین و پایین ترین نقاط / تفاوت بین بالاترین و پایین ترین نقاط = b

و پس از محاسبه b مقدار a را می توان بدست آورد.



۲- استفاده از شاخص‌های آماری (میانگین، انحراف استاندارد و مفهوم توزیع نرمال)

قبل از شرح این روش باید با مفاهیم اولیه آن آشنا شد:

۱- شاخص‌های آماری شامل:

الف- معیارهای تمایل به مرکزیت همچون میانه، میانگین و مد

ب- معیارهای پراکندگی (در اطراف میانگین) شامل: واریانس و انحراف استاندارد

۲- مفهوم توزیع نرمال

میانگین:

حاصل جمع مقادیر مجموعه ارقام بر تعداد ارقام مجموعه که توسط فرمول ذیل بیان می‌شود:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \sum_{i=1}^n x_i$$

واریانس:

معیاری برای سنجش میزان انحراف مقادیر ارقام مجموعه از میانگین آن طبق فرمول ذیل

$$s^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1} \quad \text{واریانس}$$

در فرمول بالا صورت کسر مجموع اختلاف هریک از ارقام از میانگین است که برای حذف اثر منفی و مثبت به توان ۲ رسیده است.

انحراف معیار: در فرمول واریانس درک ماهیت مقدار پراکندگی به علت به توان ۲ رسیدن آسان نیست به همین دلیل ریشه دوم واریانس را محاسبه می‌کنند که انحراف معیار نامیده می‌شود البته از فرمول ذیل می‌توان مستقیماً انحراف معیار را محاسبه نمود

$$s = \sqrt{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}}$$

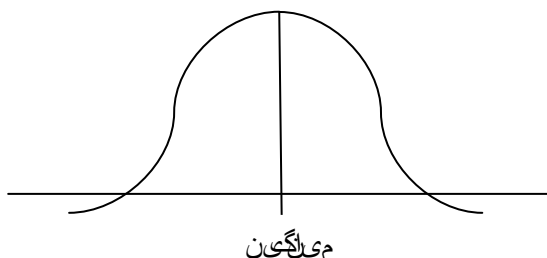
استفاده از مفهوم توزیع نرمال:

با استفاده از میانگین و انحراف استاندارد می‌توان احتمال مشاهده مقداری معین را در یک فاصله عددی برآورد کرد مثلاً احتمال اینکه ارقام در یک فاصله انحراف استاندارد در دو طرف میانگین قرار بگیرند ۶۸٪ خواهد بود این موضوع با مفهوم توزیع نرمال و جدول مقادیر Z قابل توضیح است.

در پیش‌بینی با استفاده از میانگین و انحراف استاندارد دو فرض مهم باید وجود داشته باشد:

۱- نمونه انتخاب شده معرف جامعه آماری باشد

۲- توزیع جامعه آماری نرمال باشد توضیح اینکه توزیع نرمال بدین معنی است که مقادیر با الگوی یکنواخت در دو طرف میانگین پراکنده شده باشند و احتمال قرار گرفتن یک مقدار معین در دو طرف میانگین با هم برابر بوده و پراکندگی مقادیر تشکیل منحنی به شکل زنگ بدهد.





لازم به توضیح است که Z بیانگر ضریب انحراف استاندارد در تعیین فاصله در اطراف میانگین است. مقادیر مختلف Z در جدول توزیع نرمال آورده شده است. فرض کنید $Z=1$ یعنی فاصله یک انحراف سطح زیر منحنی نرمال به اندازه یک انحراف استاندارد از میانگین که این عدد نشان دهنده احتمال قرار گرفتن ارقام در آن فاصله است: به عنوان مثال فرض کنید که $Z=1$ باشد و در جدول $Z=1$ عدد 0.2420 را نشان می‌دهد این بدان معنی است که در توزیع نرمال $0.2420 \times 100 = 24.20\%$ از مشاهدات در فاصله یک انحراف استاندارد از میانگین قرار دارد و $0.7580 = 1 - 0.2420$ از مشاهدات در فاصله یک انحراف استاندارد در دو طرف میانگین قرار دارند.

۳- تجزیه و تحلیل رگرسیون همبستگی

نوع مهم دیگری از تجزیه و تحلیل‌های آماری که درامبرآورد و پیش‌بینی کاربردگسترده‌ای دارد تجزیه و تحلیل رگرسیون و همبستگی است. این روش در طرح ریزی و کنترل ارقام بهای تمام شده و یا پیوند آن با سطح فعالیت (تولید) بکار برده می‌شود. زیرا تغییرات ارقام بهای تمام شده در سطوح مختلف فعالیت آثار مهمی بر طرح ریزی و کنترل دارد. برخی از این ارقام در یک دامنه مربوطه فعالیت نسبتاً ثابت می‌ماند، بعضی متناسب با نوسان سطح فعالیت تغییر می‌کند و بالاخره گرایش پاره‌ای به روشنی قابل توجیه نیست. برای تجزیه و تحلیل و توجیه چنین گرایش‌های متنوع روش‌های مقداری بی‌اندازه مفید است. تجزیه و تحلیل رگرسیون همبستگی، برای تعیین ماهیت وابستگی بین ارقام بهای تمام شده و سطح فعالیت (تولید) به کاربرده می‌شود. در این روش بیش از سایر روش‌های مقداری از ارقام و اطلاعات حسابداری مدیریت استفاده می‌شود زیرا به کمک آن می‌توان مقادیر یک یا چند متغیر مستقل را با یک متغیر تابع پیوند داد و مقدار متغیر اخیر را پیش‌بینی کرد. این پیش‌بینی‌ها در اغلب مسایل بودجه بندی طرح ریزی و کنترل نقش اطلاعات کلیدی را دارا است. روش رگرسیون مبتنی بر بدست آوردن یک رابطه خطی بین محرک هزینه و موضوع هزینه است برای این کار از خط رگرسیون براساس اطلاعات نمونه بدست آمده که فرض می‌شود بیانگر خط رگرسیون جامعه است استفاده می‌شود. پس از این مرحله آزمون‌هایی برای اطمینان از غیرتصادفی بودن و صحت رابطه بدست آمده صورت می‌گیرد. پس مراحل را می‌توانیم به صورت ذیل شرح دهیم [۱].

مراحل انجام تجزیه و تحلیل رگرسیون

- ۱- انتخاب متغیر وابسته (موضوع هزینه)، متغیر مستقل (محرک هزینه) و نوشتن معادله خط رگرسیون
- ۲- برآورد پارامترهای خط رگرسیون با استفاده از روش حداقل مجزورات
- ۳- بکارگیری آزمون‌ها و ابزارها برای سنجش صحت و قابلیت اتکا خط رگرسیون بدست آمده

۱- معرفی متغیرهای تابع و مستقل و تعریف معادله خط رگرسیون:

در مرحله تبیین و تشخیص مدل بر استدالات منطقی نوع ارتباطی که در جامعه بین متغیر تابع و متغیر یا متغیرهای مستقل وجود دارد تعیین می‌گردد. این ارتباط به شکل مدل ریاضی زیر بیان می‌گردد:

$$Y = \alpha + \beta X + \epsilon$$

Y معرف متغیر تابع است که باید برآورد شود

α معرف ضریب ثابتی است که در جامعه آماری وجود دارد

β معرف ضریب متغیر مستقل X در جامعه آماری است و $\sum$ معرف مقادیر مازاد تصادفی مربوط به جامعه آماری است و تفاوت بین مقادیر برآوردی Y و مقادیر مشاهده شده واقعی را بیان می‌کند.



۲- استفاده از روش حداقل مربعات برای برآورد پارامترهای مدل

برای محاسبه پارامترهای خط رگرسیون، روشهای مختلفی وجود دارد که یکی از آنها روش حداقل مجزورات است که با حل دو معادله و دو مجهول (a و b) زیر انجام می‌شود:

$$\sum y_i = na + b \sum x_i$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2$$

برای حل آن دو متغیر a و b وجود دارد و مقادیر متغیرها به شرح زیر معلوم است:

$\sum y_i$ مجموع ارقام مخارج غیر مستقیم هفتگی بخش اورژانس در ۶ هفته (نمونه انتخابی)

$\sum x_i$ مجموع تعداد بیماران بخش اورژانس در ۶ هفته انتخابی می‌باشد

$\sum x_i y_i$ مجموع حاصلضربهای مقادیر زوجهای از y_i و x_i (مثلاً $x_1 y_1$ و $x_3 y_3$)

مجموع مقادیر مجذور شده متغیر مستقل یعنی تعداد بیماران هر هفته می‌باشد. $\sum x_i^2$

بیمارستان سلامت

ارقام لازم برای تجزیه و تحلیل رگرسیون همبستگی

هفت	هفتگی y_i	y_i^2	x_i	x_i^2	$x_i y_i$
1	2250000	5,062,500,000,000	110	12,100	247,500,000
2	2580000	6,656,400,000,000	132	17,424	340,560,000
3	2800000	7,840,000,000,000	184	33,856	515,200,000
4	2340000	5,475,600,000,000	115	13,225	269,100,000
5	2150000	4,622,500,000,000	91	8,281	195,650,000
6	2400000	5,760,000,000,000	148	21,904	355,200,000
	14,520,000	35,417,000,000,000	780	106,790	1,923,210,000

حال چنانچه مقادیر متغیرهای معلوم در معادله های (۱) و (۲) جایگزین شود ارزش پارامتری a, b در معادله رگرسیون به شرح زیر به دست می‌آید:



$$14/520/000 = 6a + 780b$$

$$1/923/210/000 = 780a + 106/790b$$

برای حل دو معادله دو مجهول بالا، ابتدا معادله (۱) در عدد ۱۳۰ ضرب و سپس از معادله (۲) کم می‌شود:

$$130 * 14/520/000 = 130 * 6a + 130 * 780b$$

$$1/887/600/000 = 780a + 101/400b$$

$$1/923/210/000 = 780a + 106/790b$$

$$35/610/000 = 5390b$$

$$B = 35/610/000/5390 = 6607$$

مقدار پارامتر b معرف شیب خط رگرسیون است یعنی در مقابل هر واحد اضافی از x مقدار y معادل ۶۶۰۷ ریال افزایش می‌یابد. چون در مثال بیمارستان سلامت، متغیر x

مربوط به تعداد بیماران است، لذا می‌توان اظهار داشت که به ازای هر بیمار اضافی، مخارج غیر مستقیم هفتگی بخش اورژانس در حدود ۶۶۰۷ ریال اضافه می‌شود.

حال که مقدار پارامتر b محاسبه شده است مقدار a با جایگزین کردن مقدار b در یکی از دو معادله بالا به شرح زیر بدست می‌آید:

$$14/520/000 = 6a + 780 * 6607$$

$$a = 1/561/100$$

مقدار پارامتر a که محل تقاطع خط رگرسیون را با محور y نشان می‌دهد، معرف جزء ثابت مخارج غیر مستقیم هفتگی بخش اورژانس است. معادله رگرسیون در این مثال عبارت است از:

$$y = 1561100 + 6607x$$

بطوریکه از معادله بالا پیداست جمع مخارج غیر مستقیم هفتگی بخش اورژانس مشتمل بر ۱/۵۶۱/۱۰۰ ریال اقلام ثابت و ۶۶۰۷ ریال اقلام متغیر به ازای هر یک نفر بیمار می‌باشد [۲].

۳- سنجش ماهیت وابستگی بین متغیر وابسته و مستقل و دقت برآورد

در این مرحله باید با یکسری از ابزارها و آزمون‌ها صحت، دقت و مفروضات مدل رگرسیون را بررسی کنیم.

الف- تجزیه و تحلیل همبستگی

اگرچه معادله خط رگرسیون، وابستگی بین متغیر مستقل و متغیر تابع را بیان می‌کند اما میزان این وابستگی را دقیق مشخص نمی‌کند. روش محاسبه پارامترهای a و b نیز مناسبترین خط رگرسیون را بدست می‌دهد اما ممکن است وابستگی واقعی

و منطقی بین متغیر مستقل و متغیر تابع موجود نباشد. به منظور سنجش تغییرات وابستگی میان این دو متغیر ها و برازش R^2 با معادله 29th, Feb. 2016

ضریب ہمبستگی

این ضریب که معمولاً با حرف r نشان داده می شود، معیاری مهم برای سنجش وابستگی بین متغیر مستقل متغیر تابع است. مقدار ضریب r بین -1 و $+1$ نوسان می کند که اولی معرف همبستگی کامل منفی و دومی معرف همبستگی کامل مثبت می باشد. همبستگی مثبت در مواردی وجود دارد که با افزایش (کاهش) متغیر مستقل متغیر تابع افزایش (کاهش) یابد. همبستگی منفی عکس حالت بالا را نشان می دهد و با افزایش مقادیر متغیر مستقل مقادیر متغیر تابع کاهش می یابد

برای محاسبه مقدار r روشهای مختلفی وجود دارد که یکی از این روشها به شرح فرمول زیر است:

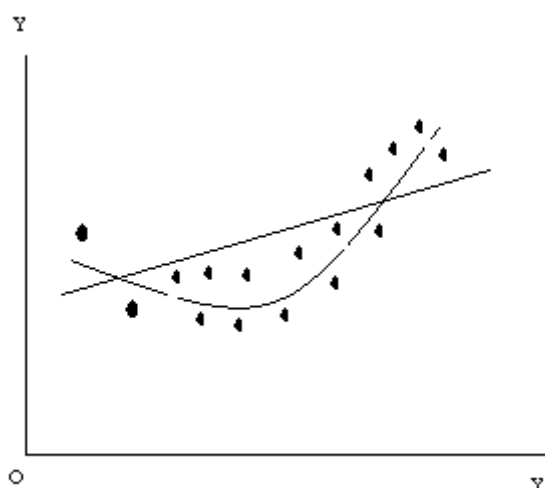
$$R = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum y_i^2 - (\sum y_i)^2} \times \sqrt{n \sum x_i^2 - (\sum x_i)^2}}$$

کلیه مقادیر متغیر فرمول بالا در جدول (۸-۵) موجود است و با جایگزینی کردن آن مقدار ۲ بدست می‌آید:

$$r^* = \frac{(6)(1923210000) - (780)(14520000)}{\sqrt{6 \cdot 106790 - (780)^2} \cdot \sqrt{6 \cdot (3541700000000) - (14520000)^2}}$$

r=0/9189398

هرچه مقدار ۲ به ۱+ و ۱- نزدیکتر باشد، وابستگی بین متغیر مستقل و متغیر تابع بیشتر است. مقدار ۲ در مثال بالا وابستگی مثبت زیادی را بین مخارج غیر مستقیم هفتگی بخش اورژانس نشان می دهد. هرچه مقدار ۲ به صفر نزدیکتر باشد بی اهمیتی وابستگی مشخص می گردد. در تفسیر مقدار ۲ باید توجه و دقت کافی به این موضوع مبذول شود که این ضریب تنها میزان وابستگی خطی بین متغیر تابع و متغیر مستقل را اندازه گیری می کند. مثلاً در موردی که وابستگی غیر خطی زیادی بین X و Y موجود می باشد مقدار ۲ محاسبه شده نسبتاً کم و معرف وابستگی خطی ضعیفی خواهد بود در شکل (۱۰-۵) زیر





* در مقایسه مقادیر r با میزان وابستگی لازم است توجه شود که این دو با نسبت یکسان افزایش یا کاهش نمی‌یابد. مثلاً $r=0/8$ معرف وابستگی دوبرابر قوی تر نسبت به مورد $r=0/4$ نیست بلکه نشان دهنده وابستگی چهار برابر قوی ترمی باشد، همینطور $r=0/75$ معرف وابستگی نه برابر قوی تر از مورد $r=0/25$ است. بطور کلی اگر مقدار r دوبرابر شود میزان وابستگی 2^2 افزایش می‌یابد و اگر r سه برابر گردد، میزان وابستگی نیز 3^2 خواهد بود.

به منظور اینکه مفهوم و درک بهتری از r بدست آید. معمولاً آن را به توان ۲ رسانیده و ضریب تعیین می‌نامند. r^2 علامت منفی احتمالی r را از بین می‌برد و مقایسه میزان وابستگی بین دو متغیر را آسان می‌کند. بنابراین ضریب تعیین نیز معیاری برای ارزیابی معادله خط رگرسیون (که از روی ارقام نمونه محاسبه شده است) محسوب می‌گردد. این معیار نیز میزان وابستگی خطی بین x و y را ارزیابی می‌کند. در مثال بخش اورژانس مقدار ضریب تعیین به شرح زیر محاسبه می‌شود:

$$r^2 = .9189398^2$$

$$0/84r^2 \approx /84$$

ضریب تعیین در واقع حاصل تقسیم دو واریانس است. یعنی ضریب تعیین در واقع حاصل تقسیم واریانس خط رگرسیون YC در اطراف میانگین ارقام؛ به واریانس کل ارقام است.

$$r^2 = \frac{\sum(y_c - \bar{y})^2 \div (n-1)}{\sum(y_i - \bar{y})^2 \div (n-1)}$$

اگر کلیه ارقام مشاهده شده در نمونه روی خط رگرسیون قرار گیرد، یعنی $r^2 = /84$ باشد، واریانس خط رگرسیون نسبت به میانگین، مساوی واریانس ارقام مشاهده شده نسبت به میانگین خواهد شد. به عبارت دیگر یک وابستگی کامل بین متغیر مستقل و وابسته وجود دارد و می‌توان کل واریانس ارقام را با متغیر مستقل (تعداد بیماران) بیان نمود. در مثال قبل که $r^2 = /84$ بود یعنی ۸۴ درصد از واریانس کل را می‌توان با تعداد بیماران بیان نمود و مابقی که از فرمول $(1 - r^2)$ بدست می‌آید مربوط به عوامل گوناگون دیگر است، که اصطلاحاً به آن ضریب عدم تعیین می‌گویند. ممکن است بخشی از این واریانس بیان نشده را با اضافه کردن متغیرهای مستقل دیگر بیان نمود، اما در نهایت درصدی مربوط به نوسان‌های تصادفی و سایر عوامل باقی خواهد ماند.

ب- آزمون برآورد پارامترها: در مرحله اعتبار دهی مدل رگرسیون برآوردی، ابتدا قابلیت اطمینان ضرایب مدل رگرسیون مورد تجزیه و تحلیل قرار می‌گیرد. در مدل رگرسیون ساده، تنها ضریب b به عنوان برآوردی از ضریب β جامعه مورد آزمون قرار می‌گیرد. متداولترین آزمون برای ارزیابی ضریب b استفاده از انحراف استاندارد این ضریب در آزمون t است.



می‌کند. به عبارت دیگر قابلیت اطمینان ضریب برآورد شده را مورد سنجش قرار می‌دهد. کوچک بودن انحراف استاندارد b به معنای نزدیکی این ضریب برآوردی به پارامتر جامعه است. قرمول محاسبه انحراف استاندارد b به شرح ذیل است:

$$s_b = \frac{SEE}{\sqrt{\sum (X - \bar{X})^2}}$$

$$SEE = \sqrt{\frac{\sum (y - y_c)^2}{n - 2}}$$

چنانچه ضریب b تصادفی بدست آمده باشد (یعنی پارامتر β جامعه مساوی صفر باشد) ارزش t محاسبه شده در محدوده رد فرض $\beta = 0$ قرار نمی‌گیرد (در توزیع t)

$$t_c = \frac{b - \beta}{s_b} \text{ یا } t_c = \frac{b - 0}{s_b} = \frac{b}{s_b}$$

چنانچه t_c بزرگتر یا کوچکتر از ارزش t مستخرج از جدول توزیع t (با توجه به ضریب اشتباه نوع اول α و درجات آزادی) باشد می‌توان فرض $\beta = 0$ را رد کرد و ضریب b بدست آمده را قابل اطمینان دانست.

انحراف استاندارد خط رگرسیون: انحراف استاندارد خط رگرسیون را می‌توان بیانگر متوسط اشتباه در تخمین y تلقی نمود که انحرافات را پیرامون خط رگرسیون نشان می‌دهد.

ج- سنجش صحت مفروضات مدل رگرسیون شامل:

- ۱- خطی بودن رابطه
 - ۲- همگن بودن واریانس
 - ۳- نرمال بودن توزیع جامعه آماری
 - ۴- همبسته نبودن واریانس بیان نشده در دوره های مختلف
 - ۵- همبسته نبودن متغیرهای مستقل با یکدیگر
- سنجش صحت این مفروضات، به کمک واریانس بیان نشده که اصطلاحاً واریانس اضافی نامیده می‌شود و استفاده از نمودارهای مختلف انجام می‌گیرد.
- روش دیگر برای آزمون وجود این پدیده آزمون *Durbin-Watson* است این آزمون با محاسبه مقدار d به کمک رابطه ذیل صورت می‌گیرد.



$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

در فرمول بالا

$$e_t = (y_t - y_c)$$

y_t معرف ارزش محاسبه شده متغیر تابع در دوره t است

y_c معرف ارزش واقعی مشاهده شده متغیر تابع در دوره t است.

چنانچه d محاسبه شده طبق فرمول صفر یا چهار باشد به ترتیب *Auto correlation* (خود همبستگی کامل) مثبت و منفی وجود دارد و گر d به دو نزدیک باشد *Auto correlation* وجود ندارد و مقادیر e تصادفی است.

۵-ج - فرض همبسته نبودن متغیرهای مستقل با یکدیگر:

در مواردی که بیش از یک متغیر مستقل وجود دارد امکان همبستگی این متغیرها وجود دارد در صورت وجود این همبستگی پدیده *Multi Collinearity* رخ می‌دهد که موجب بیش از واقع برآورد شدن انحراف استاندارد ضرایب مدل رگرسیون می‌شود اثر عملی وجود این پدیده این است که علی‌رغم بالا بودن ضریب تعیین، ضرایب مدل رگرسیون علائمی دارند که برخلاف انتظارات منطقی اقتصادی است.

یکی از روش‌های آزمون *Multi Collinearity* محاسبه ماتریس همبستگی یعنی محاسبه ضریب همبستگی بین تمامی ترکیبات ممکن متغیر تابع و متغیرهای مستقل است برای قبول فرض نباید بین هر زوج از متغیرهای مستقل همبستگی وجود داشته باشد (معمولاً همبستگی زیر ۵/۱ مورد قبول است) ضریب همبستگی بین هر زوج از متغیرهای مستقل از رابطه ذیل قابل محاسبه است:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

آنالیز واریانس (ANOVA)

توان خط رگرسیون را برای بیان مقادیر مشاهده شده متغیر وابسته ارزیابی می‌کند. یعنی صحت کلی مدل رگرسیون را بررسی می‌کند. اساس این روش بر تقسیم انحراف کل به دو قسمت است، قسمت بیان شده توسط خط رگرسیون و قسمت بیان نشده توسط خط رگرسیون. می‌توان این رابطه را به صورت ذیل نوشت: [۴]



می‌توان با رابطه بالا را به توان ۲ رساند و برای جمع مقادیر صادق خواهد بود:

$$\sum (y_i - \bar{y})^2 = \sum (y_c - \bar{y})^2 + \sum (y_i - y_c)^2$$

جدول شماره ۱ اجزای ANOVA

منشا واریانس	مجموعه مجذور انحرافات	درجات آزادی	واریانس	
بیان شده توسط خط رگرسیون	$\sum (y_c - \bar{y})^2$	$V_1 = 1$	$\frac{\sum (y_c - \bar{y})^2}{1}$	واریانس بین گروه‌ها
بیان نشده توسط خط رگرسیون (مازاد اضافی)	$\sum (y_i - y_c)^2$	$V_2 = n - 2$	$\frac{\sum (y_i - y_c)^2}{n - 2}$	واریانس درون گروه‌ها
واریانس کل	$\sum (y_i - \bar{y})^2$	$n - 2$		

توجه شود که اساس روش تقسیم انحراف کل به دو قسمت است. با استفاده از این دو قسمت اماره F برای سنجش صحت مدل استفاده می‌شود درجه‌ات آزادی نیز برای هر واریانس طبق فرمول ذیل محاسبه می‌شود.

$$v_1 = K - 1$$

$$v_2 = N - K$$

K تعداد متغیر مستقل

N تعداد کل داده‌ها

چون در رگرسیون ساده یک متغیر مستقل وجود دارد $K=1$ خواهد بود. حال به منظور آزمون اهمیت مدل رگرسیون می‌توان از آزمون F استفاده کرد. در این آزمون فرض می‌شود که

$$H_0: \beta_1 = \beta_2 = \dots \beta_k = 0$$



$$F = \frac{\sum (y_c - \bar{y})^2 / 1}{\sum (y_i - y_c)^2 / n - 2}$$

ارزش محاسبه شده F طبق جدول و فرمول بالا با ارزش جدول توزیع F در سطح اشتباه نوع اول α و درجات آزادی ۱ و $n-2$ مقایسه می‌شود و چنانچه بزرگتر از ارزش جدول باشد، فرض مساوی بودن مقادیر β با صفر رد می‌شود. به بیان دیگر، مدل رگرسیون با اهمیت تلقی می‌گردد.

استفاده از معادله خط رگرسیون:

عمده ترین مورد استفاده از معادله خط رگرسیون در فرآیند طرح ریزی و بودجه بندی است. بدین صورت که ارقام بهای تمام شده در ارتباط با معیارهای متفاوت فعالیت مورد تجزیه و تحلیل قرار می‌گیرند تا بهترین متغیر مستقل انتخاب و براساس آن معادله رگرسیون تشکیل و مبنایی برای بودجه بندی ایجاد گردد. باید در نظر داشت که روش رگرسیون در مقایسه با سایر روش‌ها برآورد بهای تمام شده همچون روش حد بالا و پایین به برآورد های دقیق تری منتج می‌شود. اما دو نکته مهم در ارزیابی نتایج رگرسیون حائز اهمیت است.

۱- محاسبات رگرسیون و پارامترها براساس ارقام تاریخی صورت می‌گیرد که ممکن است برآورد خوبی از وضعیت آتی نباشد.

۲- از ضریب همبستگی نباید رابطه علت و معلولی نتیجه گرفته شود هرچند که در هنگام بالا بودن ضریب همبستگی احتمال وجود چنین رابطه ای وجود دارد.

رگرسیون خطی مرکب:

در رگرسیون خطی مرکب دو یا چند متغیر مستقل همراه با هم برای برآورد متغیر وابسته مورد استفاده قرار می‌گیرد، چون این افزایش می‌تواند ضریب تعیین را افزایش دهد و درصد بیشتری از واریانس کل (پراکندگی ارقام مشاهده شده در نمونه نسبت به میانگین آن) را بیان کند. در رگرسیون مرکب نیز ابتدا پارامترهای مدل محاسبه و سپس میزان وابستگی با محاسبه ضریب همبستگی و ضریب تعیین ارزیابی می‌گردد و فرم کلی رگرسیون مرکب به شکل ذیل است:

$$y_c = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$



مرکب به شرح ذیل است:

$$r^2 = \frac{\text{وابستگی بیان نشده}}{\text{وابستگی کل}} = \frac{\sum (y_e - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

نظر به اینکه ضریب تعیین تحت تاثیر تعداد متغیرهای مستقل منظور شده در مدل نیز قرار می‌گیرد، لازم است که این ضریب به تناسب تعداد متغیرهای مستقل تعدیل شود. و برای تفسیر همبستگی از ضریب تعدیل شده استفاده نمود. فرمول تعدیل ضریب تعیین به شرح ذیل است:

$$\text{Adjusted } r^2 = 1 - \frac{(1 - r^2)n - 1}{n - k}$$

در رابطه بالا n معرف تعداد مشاهدات و k معرف تعداد متغیرهای مستقل بکار گرفته شده در مدل رگرسیون مرکب است.

منحنی یادگیری

در روش‌های شرح داده شده تاکنون فرض بر خطی بودن روابط بین محرک هزینه و موضوع هزینه بود اما همواره این شرط برقرار نمی‌باشد منحنی یادگیری نمونه‌ای از این استثنا می‌باشد. اساس منحنی یادگیری بر این مفهوم استوار است که میزان کار لازم برای ساخت یک واحد محصول با پیشرفت کار تا مدت معینی کاهش می‌یابد. ما در اینجا موارد ذیل را در مورد این نوع هزینه‌ها شرح می‌دهیم.

۱- مفهوم و نمودار منحنی یادگیری

۲- انواع منحنی یادگیری

۱- مفهوم و نمودار منحنی یادگیری

منحنی یادگیری از این مفهوم استنتاج می‌شود که با افزایش کارایی کارکنان ناشی از یادگیری و افزایش مهارت کارکنان میزان ساعت کار لازم برای تولید هر واحد محصول و در نتیجه هزینه دستمزد مستقیم (کار مستقیم کاهش خواهد یافت. این منحنی بیانگر رفتار غیر خطی هزینه می‌باشد. نمودار منحنی یادگیری مفهوم اثر یادگیری را به خوبی بیان می‌دارد در نمودار (۱) با افزایش تولید جمع اقلام متغیر هزینه افزایش می‌یابد ولی این افزایش با یک نسبت نزولی صورت می‌گیرد. نمودار (۲) نیز متوسط زمان انباشته را برای هر واحد نمایش می‌دهد که با افزایش تولید هزینه تولید (کار یا دستمزد مستقیم) هر واحد کاهش می‌یابد



انواع منحنی یادگیری:

الف- الگوی یادگیری متوسط زمان انباشته

ب- الگوی یادگیری زمانی واحد اضافی

الگوی یادگیری متوسط زمان انباشته :

در این الگو فرض می‌شود که هنگامی که تعداد تولید دو برابر می‌شود میانگین زمان انباشته برای هر واحد تولید به درصد ثابتی کاهش می‌یابد. مثلاً مفهوم نسبت یادگیری ۸۰٪ در این الگو به این معنی است که هنگامی که تعداد تولید دو برابر می‌شود میانگین زمان انباشته برای هر واحد ۸۰٪ میانگین زمان برای تولید اولیه خواهد بود.

معادله ریاضی این الگو به شرح ذیل است:

$$y = aX^b$$

که در آن

y = متوسط زمان انباشته هر واحد

x = تعداد چند برابر شدن تولید

a = زمان لازم برای اولین گروه تولید

b = اثر یادگیری یا نرخ یادگیری که از تقسیم لگاریتم نسبت یادگیری بر لگاریتم ۲ بدست می‌آید.

$$b = \frac{\ln(\text{نسبت یادگیری})}{\ln 2}$$

الگوی مدل یادگیری زمان واحد اضافی:

در این مورد هنگامی که تعداد تولید دو برابر می‌شود زمان واحد اضافی (زمان لازم برای تولید آخرین واحد) با نسبت ثابتی کاهش خواهد یافت یعنی هنگامی که تولید از $1x$ به $2x$ واحد افزایش یابد با نسبت یادگیری ۸۰٪ زمان لازم برای تولید آخرین واحد در سطح $2x$ ، ۸۰٪ زمان لازم برای تولید آخرین واحد در سطح $1x$ خواهد بود. در این الگو معادله ریاضی همانند الگوی متوسط خواهد بود با این تفاوت که در این الگو مقدار y نشان دهنده زمان لازم برای تولید آخرین واحد است [۳].



نتیجه‌گیری:

در آمارتحلیل رگرسیون یک نام جمعی برای فنونی است که برای الگو سازی و تحلیل داده‌های مربوط به متغیر وابسته و داده‌های مربوط به یک یا چند متغیر مستقل به کار می‌رود. بنابراین در تحلیل رگرسیون ما نه تنها می‌توانیم ارتباط دو یا چند متغیر با هم را از طریق محاسبه ضریب همبستگی محاسبه نماییم بلکه می‌توانیم از این ضریب به عنوان پایه‌ای برای پیش‌بینی مقدار یک متغیر از متغیر دیگر استفاده کنیم و در اینجا است که راه رگرسیون از ضریب همبستگی جدا می‌شود بنابراین هدف از رگرسیون برآورد تاثیر متغیرها و پیش‌بینی آنها است.

منابع:

- [1] شباهنگ، رضا، (۱۳۹۰)، حسابداری مدیریت، جلد اول، سازمان حسابرسی، شماره ۱۳۱
- [2] چن لین، بلوچر، (۱۳۹۳)، حسابداری مدیریت، ترجمه علی پارسائیان، انتشارات ترمه
- [3] عمرانی، محمد، حسابداری صنعتی ۳
- [4] مومنی، منصور، تحلیلهای آماری با استفاده از spss