استاد راهنما :
آقای دکتر خسرو ملاح

مقدمه ای بر اقتصاد سنجی

AN INTRODUCTION TO ECONOMETRICS

A. A. Walters

آ. آ. والتــرز

ترجمــه :
خسرو ملکــی

۱۰۶۵۴

مقدمه

انتخاب کتاب "مقدمه ای براقتصاد سنجی" باتوجه به سوابق مترجم درامرترجمه وبراساس راهنمائی های ارزشمند جناب آقای دکترملاح صورت گرفت. ترجمه مزبورکه ازصفحات ۸۰ تا ۱۴۴۰ کتاب است جهت ارائه بعنوان پایان نامه دوره لیسانس سیاسی دانشکده اقتصاد وعلوم سیاسی دانشگاه ملی میباشد. امیداست این ترجمه که درزمینه جدیدی ازاقتصاد است مورداستفاده پژوهشگران قراربگیرد.

دراینجالازم میداندمجددا" ازکلیه استادان که همواره ارشادکننده اینجانب بوده اندبویژه آقای دکترملاح برای کمکهائیکه درترجمه این پایان نامه نموده اندتشکر نماید.

خسروملکی

۱۰۶۵۴

فهرست مندرجات

صفحه

مقدمه رگرسیون وهمبستگی

فصل چهارم ــ روابط بین متغیرها

بخش ۱ رگرسیون وروابط نظری	۱
بخش ۲ کمترین مجذورات	۹
بخش ۳ واریانس‌مانده ها وضریب همبستگی	۱۵
بخش ٤ همبستگی های غیرخطی	۱۷
بخش ٥ برآورد مقدارآلفا	۲۰
بخش ٦ محاسبه رگرسیون	۲۰
بخش ۷ رگرسیون وروابط علت ومعلولی	۲۲
بخش ۸ خطا درمتغیرها	۲٦
سوالات برای بحث	۳۲

فصل پنجم ــ رگرسیون متعدد (مرکب)

بخش ۱ ضرائب رگرسیون	۳۹
بخش ۲ همبستگی چندتائی	٤٤
بخش ۳ ضریب همبستگی متعدد R^2	٤٦
بخش ٤ ضریب همبستگی نسبی	٤۸
بخش ٥ نتیجه گیری	٥٥
سوالات برای بحث	٦۰

رگرسیون[1] وهمبستگی[2]

درقسمت دوم ابزاراصلی رگرسیون معرفی ومورد بحث قرارداده شده است. قسمت عمده اقتصاد مربوط میشود به پیش‌بینی تأثیرمتغیرها بررروی یکدیگر. رگرسیون وهمبستگی روشهائی هستند که برای تجزیه وتحلیل وابستگی ونتیکه اثرات خطی میباشند بکارگرفته میشوند. فصل عمده این قسمت فصل چهارم است که بعنوان زیربنائی است که مابقی این کتاب برروی آن بناشده است. تبحریاحداقل آشنائی نزدیک به مدلهای ابتدائی رگرسیون که دارای دومتغیرمیباشند قدم اول ولازمی است برای آشنائی به مسائل رگرسیون. ولی هدف اصلی فصل چهارم تفسیرتصاویررگرسیون وهمبستگی است. هدف طرح عمل رگرسیون — نمایان نمودن اقلا" چند نمونه از استثنائات میباشد. فصل پنجم نمونه دومتغیری راکاملترکرده وبه نمونه های سه یاچند متغیری میپردازد. درفصل پنجم چندان افکارجدیدی وارد نشده (بامقایسه بافصل چهارم) ولی متأسفانه لزوما" ازجبرمیزان متنابهی استفاده گردیده است. باوجوداین، ازآنجائیکه کارعملی دررگرسیون شامل حداقل رگرسیون نسبی ومتعدد میگردد درک کامل مبانی اصلی لازم است. فصل ششم درمورد نمونه گیری خصوصیات مدلهای رگرسیون وهمبستگی بحث می نماید ولزومی ندارد که برای درک مابقی کتاب این فصل خوانده شود.

۱— *Regression*

۲— *Correlation*

٤– روابط بین متغیرها

رگرسیون ،روابط نظری؛ اغلب دانش‌های علمی را میتوان بصورت برقراری رابطه یا نسبتی بین دو یا چند مقدار را تنظیم وتوجیه کرد · تحقیق بمنظور دستیابی به برقراری نظم کلی بین متغیرها یکی از گرایش‌های عمده علوم فیزیکی واجتماعی را تشکیل میدهد وروش‌هائی که ازطریق آنها میتوان چنین روابطی راکشف وترسیم نمود ابزاراصلی "علم اقتصادسنجی" راتشکیل میدهند · دراین فصل کاربرد این ابزاربطورنسبتاً مبسوطی مورد بحث قرارمیگیرد ·

بعنوان نمونه مسئله قدیمی اقتصاد کلاسیک جدید (نئوکلاسیک) رادرنظربگیرید یعنی شکل وشیب منحنی تقاضا · نظریه تقاضای مصرف کننده ، باپیش‌بینی واکنش یک مصرف کننده مستقل (یایک خانواده) درقبال تغییرقیمت بازارمربوط میگردد · چنین فرض میکنیم که "سلیقه ها" ثابت است ، درآمد واقعی مصرف کننده تغییرنمیکند واینکه کالای مورد نظرچنان جزء کوچکی ازکل بودجه مصرف کننده راتشکیل میدهد که تغییرات قیمت آن مصرف کننده راازلحاظ مادی بطور چشم گیری دروضع بهتریابدتری قرارنمیدهد · وانگهی فرض براین است که مصرف کننده بطور یکنواختی (Consistent) عمل میکند که البته تعریف "یکنواختی" برحسب اینکه نظریه ماتاچه حد ممکن است پیچیده باشد ، تغییرمیکند · ولی صرفنظرازریزه کاریهای آن ، "قانون تقاضا" براساس این نظریه بطورکم وبیش روشنی چنین مطرح میگردد که: اگرقیمت کالائی کاهش یابد مقدار بیشتری ازآن کالاخریداری میشود · این قانون که بیشتربرای یک فرد مصرف کننده درنظرگرفته شده است میتوان بسهولت آن رادرشرایط بازارنیزتعمیم داد · برای رسیدن به منحنی تقاضای تنهاکاریکه بایدانجام داداین است که مقادیرکالائی که درقیمت معیّنی توسط افرادخریداری شده است باهم جمع کنیم · چون طبق پیش‌بینی قانون فوق الذکربا کاهش قیمت، افراد مقدارخرید خود راافزایش میدهند پس‌میتوان چنین نتیجه گیری کرد که بطورکلی حجم خریدها افزایش خواهد یافت · پس میتوان گفت که قانون فوق الذکرهم درمورد افراد صدق میکند وهم درمورد بازار ·

اینک چنین فرض میکنیم که کالای مورد نظرتوسط یک انحصارگرعرضه میشود · این شخص قیمت کالایش راخود تعیین میکند وسپس بازار را درمقابل این قیمت مورد مشاهده قرارمیدهد · بدیهی است که انحصارگرمایل است که راجع به ماهیت تقاضا اطلاعاتی راکسب کند چون علاقمند بانتخاب

قیمتی است که بیشترین سودبرایش حاصل کند . حال فرض میکنیم که این انحصارگر "آزمایش های" خود را بمرحله اجرا درآورد باین معنی که هرماه قیمت را درسطوح مختلفی تعیین کند وسپس مقدار کالائی که درآن قیمت بفروش رفته است مورد مشاهده قرارد هد . اگراین مشاهده تغییرقیمت و فروش کالا را روی منحنی تقاضارسم کنیم منحنی بدست آمده دارای شیب نزولی است که قانون یا نظریه فوق الذکرآن راپیش بینی کرده است .

ولی حتی دراین شرایط آزمایشی مشاهدات ما دقیقاً بانظریه فوق الذکرتطبیق نمیکند . "سلیقه ها" ممکن است ازیک زمان نسبت بزمان دیگرتغییرکند واینکه قیمت کالا های جانشین ومکمل نیزتغییرکنند ویادرآمد ها افزایش یافته باشند تمام فرضیات مربوط به نظریه فوق الذکرممکن است هنگام آزمایش نقض شوند . بنابراین درآمد مورد مشاهده لااقل دارای دوجزء اصلی است، پیش بینی نظری و "خطاهای" تجربی که شامل خطاهای اندازه گیری نیزمیشود . ولی کسی نمیتواند ادعاکند (یالااقل دراقتصاد چنین ادعاکند) که اگرخطاهای تجربی وجود نمیداشت اطلاعات داده شده میتوانست همواره بانظریه فوق الذکر دقیقاً مطابقت پیدا کند .

حتی درشرایط آزمایشی کامل نظریه فوق ممکن است تنها یک پیش بینی ناقص ازاطلاعات فرض شده رابعمل آورد . ازاین روی همیشه یک جزء سوم نیزدرنتایج حاصله ازآزمایش وجود دارد و آن عبارت است از: خطاهای ناشی ازاینکه نظریه خود به تنهائی نمایشگریک توصیف تقریبی ازواقعیتها است .

خطاهای موجود درآزمایش چنان حائزاهمیت اندکه همواره بایستی آنهارامورد توجه دقیق قرارداد . سودمندی این نظریه بمقدارزیادی این امید رادامن میزندکه اهمیت خطاها نسبت به قدرت توجیه کننده سیستماتیک نظریه ناچیزاست . درواقع روش تجربی نه تنها بایستی روشنگر نظریه مورد بحث باشد ــــ یعنی رابطه سیستماتیک بین متغییرها ــــ بلکه بایستی کیفیت خطاهارانیزتوجیه کند .

مثال فرضی آن انحصارگری رادرنظربگیرید که میخواست "منحنی تقاضای" خودراکشف کند . فرض کنیم که نظریه مورد بحث پیش بینی میکند که رابطه بین قیمت ومقدارفروش رفته یک رابطه خطی است که دارای شیب منفی است + . مامیتوانیم "نظریه" یاخط نظری رگرسیون رابطورزیرنشان دهیم:

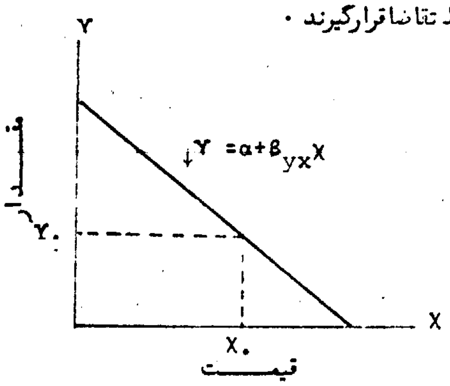
+ توجه داشته باشید که مانظریه تقاضاراخیلی دقیق ترازحالت معمولی بکارمیبریم . دراقتصاد سنجی معمولاً ضروری تریامطلوب تراست که از فرضیات محدودتری استفاده کنیم تاآنجاکه بطورضمنی توسط صرف نظریه بیان میشود .

(۴ـ۱) $Y = \alpha + \beta_{yx} X$ برای مقدار معینی از X که در آن $\alpha > 0$ و $\beta_{yx} < 0$ است.
در معادله فوق Y مقدار کالای خریداری شده، X قیمت α، β_{yx} نیز مقادیر ثابتی هستند. ترتیب زیرنویس yx β_{yx} نشانه آن است که Y متغیر تابع و X متغیر مستقل است یعنی با معلوم بودن مقدار X، میتوان مقدار Y را پیش‌بینی کرد. حال فرض میکنیم که نظریه ما تصریح میکند که یک رابطه خطی وجود دارد ولی صرفنظر از پیش‌بینی اینکه خط دارای شیب منفی است، نظریه مقدار عددی این شیب و یا محل دقیق آن را بما نشان نمیدهد. فرضیه ما صرفاً مبتنی بر این است که یک رابطه خطی منفی وجود دارد، ولی این کار آزمایشی است که باید کشف نماید که آیا ۱ـ اطلاعاتی وجود دارند که با این فرضیه تطبیق نمایند؟ و یا دقیق‌تر اینکه ۲ـ محل وشیب خط تقاضا را برآورد نمایند. ترسیم هندسی این نظریه در شکل (۴ـ۱) نشان داده شده است. رابطه تقاضا به شرح زیر تفسیر میشود:

"با معلوم بودن قیمت X_0 سپس مقدار بفروش رفته Y_0 خواهد بود"

این رابطه عبارت است از پیش‌بینی مقدار فروخته شده روی محور عمودی و بر اساس قیمت داده شده روی محور افقی. مقدار کالای فروخته شده متغیر تابع است که بر اساس اطلاع از قیمت که متغیر مستقل است، پیش‌بینی میشود. (خوانندگان توجه خواهند داشت که ما جای محورها را از روش معمولی نشان دادن منحنی های تقاضا که در کتب کشورهای آنگلوساکسون بکار برده میشود، عوض کرده ایم. هیچگونه تغییر عمده ای در اینجا داده نشده است و ترسیم چنین نموداری با موازین پذیرفته شده در علوم و آمار بهتر مطابقت پیدا میکند.)

تنها در یک آزمایش کامل و یک نظریه کامل است که تمام نتایج مشاهده شده روی خط قرار میگیرند. در یک دنیای واقعی بهترین چیزی که میتوان انتظار داشت این است که قیمت ها و مقادیر مشاهده شده در نزدیکی خط تقاضا قرار گیرند.

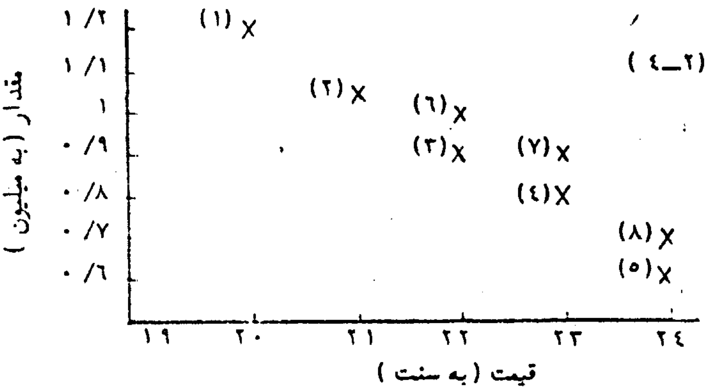


شکل (۴ـ۱)

مامیتوانیم مقادیرخریداری شده دربازارراهمانطوریکه باقیمت های مختلفی که توسط یک انحصارگرتعیین میشودرسم کنیم • فرض میکنیم که مقادیرزیرمورد مشاهده قرارگرفته است •

مقدار	قیمت (به سنت)	ماه
۱/۲۰۰/۰۰۰	۲۰	۱
۱/۰۰۰/۰۰۰	۲۱	۲
۹۰۰/۰۰۰	۲۲	۳
۸۰۰/۰۰۰	۲۳	٤
٦۰۰/۰۰۰	۲٤	٥
۱/۰۰۰/۰۰۰	۲۲	٦
۹۰۰/۰۰۰	۲۳	۷
۷۰۰/۰۰۰	۲٤	۸

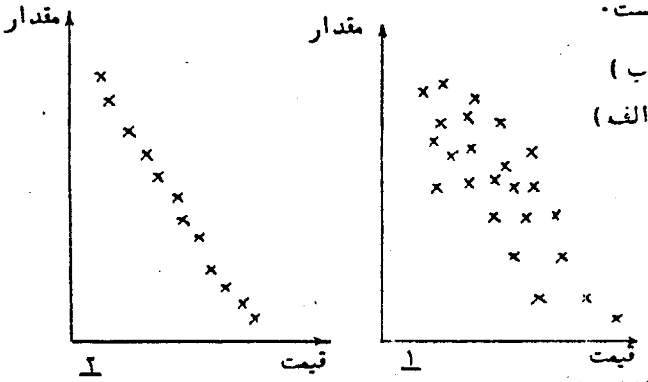
اگراطلاعات فوق راروی منحنی ایکه دارای دومتغیراست رسم کنیم، (نموداریکه اصطلاحاً نمودار پراکندگی نامیده میشود) نمودارزیربدست میآید •



شکل (۲ــ٤)

بایدتوجه داشته باشیدکه این مشاهدات دقیقاً روی خط قرارنمیگیرند • خطاهائی راه خودرابمیان نتایج بدست آمده پیداکرده اندوبایستی مسئله راباتوجه بخطاهای موجود موردبررسی قرار داد • دراینصورت میتوان رابطه خطی فرضیه ودرصورت امکان محل خط راتعیین وکشف کرد •

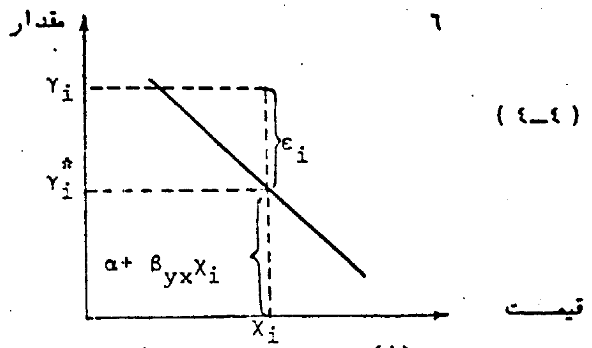
با درنظرگرفتن مثال داده شده بدیهی است که نوعی رابطه منفی بین متغیرها وجود دارد . ولی همچنین مسلم است که خطاهای مربوط به فرضیه و آزمایش زیاد هم ناچیز و قابل چشم پوشی نیست . مشاهدات درامتداد یک خط مستقیم نشان داده نشده بلکه دارای یک پراکندگی سیگار مانند است که جهت آن ازشمال غربی به جنوب شرقی امتداد دارد . هرچه این مسیرسیگار مانند نازک تر باشد ، قدرت پیش بینی نظریه بیشتر و احتمال وجود خطاها کمترخواهد شد . ولی اگر این مسیرسیگار مانند وصدای پراکندگی بیشتر و خیلی ضخیم باشد ، خطاهای موجود نسبت به ارزش پیش بینی کننده نظریه ، قابل توجه خواهد بود . چنین وضعی درنمودارهای زیرنشان داده شده است : نمودار (الف) ارقامی را نشان میدهد که دقیقاً با شرایط نظریه (تئوری) مطابقت دارد ، درحالیکه نمودار (ب) موردی را نشان میدهد که خطاها بسیار قابل ملاحظه هستند و نظریه دارای قدرت پیش بینی کننده چندانی نیست .



۱- نمودار پراکندگی (ب)

۲- نمودار پراکندگی (الف)

برای بررسی این اطلاعات بایستی نظریه خطی ساده را دوباره طرح ریزی نمود بطوریکه این دفعه دقیقاً "خطاها" را شامل شود . به عبارت دیگر این بار نظریه بایستی بشکل آماری طرح ریزی شود بطوریکه بتواند اطلاعات داده شده را دقیقاً توصیف نماید . اینک X_i و Y_i را که به ترتیب نماینده قیمت و مقدار است برای ماه آیم (ith) می نویسیم . بدیهی است که نمیتوانیم بسهولت X_i ، Y_i را درمعادله (۱-٤) جایگزین نمائیم چون Y_i تحت تأثیر خطاها قرار میگیرد . اینک (εi) را که نماینده خطاها است در Y_i می نویسیم یعنی آنکه "خطاها" درمدل مورد نظر در طی ماه آیم (ith) درمتغیرتابع نشان داده میشود . ما درمعادله (۲-٤) $Y_i = \alpha + \beta_{yx} X_i + \varepsilon i$ تصریح میکنیم که خطا اضافه شده است . اضافه کردن خطا درمعادله درنمودار زیرنشان داده شده است .



شکل (٤ـ٤)

با معلوم بودن مقدار X_i (قیمت) در ماه (i) مقدار X_i را مشاهده میکنیم . این مقدار ممکن است بدو جزء تجزیه شود . ابتدا آنکه مقدار Y از طریق نظریه پیش‌بینی میشود که این صرفاً مقدار Y است که از طریق رابطه خطی پیش‌بینی میشود . در این صورت این مقدار را Y_i^* مینامیم . بدین ترتیب خواهیم داشت:

$$Y_i^* = \alpha + \beta_{yx} X_i$$

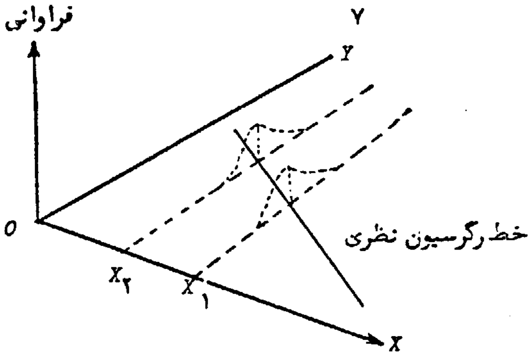
$$Y_i = Y_i^* + \varepsilon_i$$

یا بعبارت دیگر: (یا اخلال) خطا + مقدار پیش‌بینی شده = مقدار مشاهده شده

اهمیت قاطع خطا چنان است که گاهی عنوان کم اهمیت تر "اخلال" بآن خطاب میشود . (بنابراین از این ببعد کلمه "خطا" باعبارت "اخلال مشاهده شده" توصیف میشود .)

مشکل این است که یک نام مناسب برای "اخلال ها" پیدا شود . در اقتصاد سنجی از عبارت "اخلال ها" برای نشان دادن اثر عواملی که از نظریه صرف شده است، استفاده میشود . در استعمال نظریه تقاضا، میتوان بعضی از علل اختلاف ها از مدل را تشخیص داد . بعنوان مثال شرایط جوّی، جنگ و تغییرات سلیقه تعدادی از دلایل مهم برای اختلاف بین پیش‌بینی و واقعیت را تشکیل میدهد .

موضوع واقعاً مهم است که: توزیع اخلال ها از چه قواعدی متابعت میکنند؟ روش عادی، تشخیص این مطلب است که اخلال ها مانند "متغیرهای اتفاقی" عمل میکنند . برای اینکه دقیق باشیم، اخلال ممکن است بعنوان متغیری که بطور نرمال توزیع شده است و دارای میانگین صفر است، توصیف میشود . در هر مورد یک انحصارگر قیمت X_i را انتخاب میکند ما علاوه بر داشتن یک قسمت خطی سیستماتیک، دارای اخلالی (Disturbance) هستیم که از یک توزیع نرمال جمعیت بدست آمده است . بنابراین بطور نموداری میتوان چنین تصور کرد که توزیع نرمال اخلال در جهت محور Y دور از خط رگرسیون قرار میگیرد . اگر فراوانی (Frequency) را روی یک محور سوم اندازه گیری کنیم، شکل مدل ما بصورت شکل (٥ـ٤) خواهد شد .



شـــکل (٥ـ٤)

تغییررسمی مدل این است که وقتی انحصارگریک قیمت راتعیین میکند فقط یک بازده وجـود داردکه بطورسیستماتیکی توسط نظریه پیش‌بینی میشودکه آن مقدار Y^* است. ولی این مقدارواقعی بدست آمده Y برای یک قیمت معین است. ما میتوانیم این موضوع رابطورتجربی تفسیرکنیـــم. انحصارگرقیمت ۲۲ سنت راماه پس ازماه تعیین میکند ودرهرمورد مقدارفروخته شده راثبت میکنـــد. نظریه مورد بحث بیان میکند که متوسط مقدارفروخته شده طبق قانون زیربدست میآید :

$$\text{متوسط مقدارفروخته شده} = Y^* = \alpha + \beta_{yx}(22)$$

اینک مقادیرواقعی نظری α و β_{yx} رابه ترتیب ۳۰ و ۰/۵ ــ ثبت میکنیم.

$$Y^* = 20 - 0/5\ (22) = 9$$

ازاین روی باقیمت ۲۲ سنت متوسط مقدارفروخته شده ۹ است. ولی مقدارفروخته شـــده برای هرماه معین ممکن است بعلت اخلال(ε) کم وبیش ازاین رقم ۹ تجاوزکند. فروش واقعی بـــا اضافه کردن تعدادی (ε) بمقدارمتوسط که ازتوزیع عادی نرمال استخراج شده است، بدست میآید. درهرماه که ۲۲ سنت منظورمیشود، تنهاچیزی که تغییرمیکنداستخراج ازتوزیع عادی نرمال است. گوئی هرماه تعدادی به ۹ میلیون اضافه شده که بطوراتفاقی ازکلاهی بیرون کشیده ایم که درآن تکه ــ هائی گذاشته ایم که نمایشگرفراوانی های نسبی تعدادی است که ازتوزیع نرمال بدست آمده است.

دراینجابایدمکثی کنیم وبه تحقیق بپردازیم که چرا " خطاهاومعلول های حذف شده از ــ مدل " بایدازطریق یک متغیراتفاقی نشان داده شود. بنظرفوق العاده مصنوعی خواهدآمداگـــر " حذفیات وخطاها " راطوری بحساب آوریم که گوئی ازطریق یک قانون تصادفی بوجودآمـــده اند. این قوانین برای توصیف نتایج آزمایش‌های تکرارنشدنی ای بوجودآمده اندکه هرکسی میتوانـــد درصورت امکان مالی، آن آزمایش‌هاراانجام دهد. شواهدکافی نتوانسته است که روش "خط یاشیر" راردکند ــــ ولی این امرباتعیین قیمت آزمایش توسط یک انحصارگربخصوص، یکسان نیست. تنها پاسخ ــــ نه پاسخ نامطلوبی نیزهست ــــ این است که این تسمیه اصطلاح اخلال معمـــولا

بهترین تسمیه ای است که درحال حاضروجود دارد . این درواقع شامل جیزی جزاین نیست که نامی به جهالت خود داده باشیم وبرای آن یک شکل مقداری وسیع درنظربگیریم .

تسمیه اصطلاح اخلال اغلب بصورت متغیری که بطورنرمال واتفاقی توزیع شده است بیان میشود ، ولی ضروری نیست . گواینکه اغلب مناسب است که بطورنرمال توزیع شده باشد . درواقع تنها باین احتیاج است که برحسب یک قانون احتمالی خاص توزیع شود ومهم ترازهرچیزدیگر قواعد توزیع احتمالی نباید به مقدار X بستگی داشته باشد .

صرفنظرازاینکه مقدار X تاچه حد تعیین شود ، توزیع اپسیلن ها (εs) نباید تحت تاثیر قرارگیرد . دریک مفهوم احتمالی توزیع اپسیلن ها (εs) بایستی مستقل ازایکس ها (Xs) باشد . این فرضیه حسابیه تفصیل درفصل های بعدی مورد برسی قرارخواهدگرفت ولی دراین مرحله ضروری است که یک روش تعقلی راارائه دهیم . برعکس تصورکنید که توزیع اپسیلن ها (εs) درواقع به X بستگی دارد . مثلافرض شید که همانطورکه X افزایش میابد ، مقدار (ε) احتمالا بیشتر مثبت است تامنفی . پس بدیهی است که (ε) شامل یک جزئی است که بطورسیستماتیکی بمقدار X انتخاب شده بستگی دارد . این جزء از (ε) بااطلاع ازمقدار X قابل پیش بینی است بنابراین دیگر جزئی ازبی اطلاعی ماراتشکیل نمیدهد ونتیجه ظهوراین اصل میشود که توزیع اپسیلن ها (εs) نبایستی مستقل ازمقدار X باشد .

قبل ازآنکه به تجزیه وتحلیل این مطلب بپردازیم ، بهتراست نوعی آسان سازی راارائه دهیم . چون بایستی دومقدارثابت α و β_{yx} رااندازه گیریم ، بهتراست این کارا یکی یکی انجام دهیم باین معنی که اول (β_{yx}) رااندازه میگیریم وسپس به سنجید ن (α) میپردازیم . بدین ترتیب درحل مسئله ابتداباین مسئله (بطورموقتی) مواجه میشویم که چگونه (α) راازمعادله حذف کنیم . مقدار (α) محل خط راتعیین میکند ــــ بابالا وپایین بردن آن درروی محور Y . بنابراین معقول است که موقعیت خط راهمانطورکه ازمیان "میانگین های حسابی نمونه مشاهدات Y و X عبور میکند "تعریف کنیم ــــ وبعداًمشاهده خواهیم کرد که این امردرعمل چگونه انجام خواهد شد . سپس اگرمحورهای X و Y راجابجاکنیم بطوریکه ازمیان میانگین های X و Y عبورکنند ، دومحور وخط درمنطقه میانگین ها بایکدیگرتلاقی میکنند .

اگردارای مقیاس‌های زیرباشیم

$\varepsilon'i=\varepsilon i-\varepsilon,$ که درآن $\varepsilon=(1/n)\sum_i \varepsilon i$

$xi=\chi_i-\chi,$ که درآن $\chi=(1/n)\sum_i \chi i$

$yi=\gamma_i-Y,$ که درآن $Y=(1/n)\sum_i \gamma i$

که درآن حروف xi, yi, ϵi نشان دهنده "انحرافات میانگین است"، بخش جداشده ممکن است صفرتلقی گردد — چون خط ازمیانگین مبداً مختصات (جدید) (Intercept) عبورمیکند · بنابراین ابتداروی پیداکردن شیب β_{yx} درمعادله تنیه میکنیم یعنی معادله $yi=\beta_{yx}\chi_i+\epsilon i$ وپس ازحل این معادله به بررسی برآورد مقدار (α) میپردازیم ·

کمترین مجذورات: (Least Squares)

باتوصیف چنین مدلی اینک به بررسی کشف روش‌هائی برای برآورد β_{yx} (وبعدا α) ازیک نمونه مشاهدات میپردازیم · فرض‌براین است که متغیرهاازروش مدل یاالگوی مورد نظرمتابعت میکنند (ولی توجه کنید که بخاطرسهولت کار (εi) رابصورت انحراف ازمیانگین نمونه بجای ϵi مینویسیم) (۳ـ٤) $yi=\beta_{yx}\ \chi i+\ \varepsilon i$ ودراین معادله تنهامشکل مااین است که بتوانیم برآورد مقدار β راازنمونه مشاهدات xi (قیمت) و yi (مقدار) بدست آوریم · بعلاوه چنین فرض‌میکنیم که xi و yi بدون هرگونه خطائی دقیقاًمشاهده واندازه گیری میشوند · براساس‌نمونه xi و yi بایدچنین نتیجه گیری کنیم که بهترین برآورد β_{yx} چیست ·

روش‌های مختلفی برای مطابقت دادن خط رگرسیون باچنین اطلاعاتی وجود دارد · ساده‌ترین تکنیک این است که اغلب خطی رارسم کنیم که بنظربینند به بهترین وجهی نمایشگررابطه خطی اطلاعات داده شده است · گواینکه ازاین روش‌بخوبی یادنشده است ولی این روش‌مطابقت دادن خطبااعدادازطریق چشم اغلب برای مقاصدمتعددی میتواند مفیدواقع شود · تنهامشکل اصلی این روش این است که آن روشی عینی نیست — چون وجودبسیاری ازمشاهده کنندگان وبسیاری از خطوط امکان پذیراست · وحتی هنوزکاملاً روشن نشده است که کسی بتواندخواص‌این روش‌رابه بهترین طرزممکنه مورد بهره برداری قراردهد · روش‌دیگراین است که تنهابالاترین وپایین ترین نقاط xi ویاشایدبالاترین وپایین ترین نقاط yi رابهم وصل کنیم · این میتواندیک روش‌نسبتاًمعقولانه ای باشدولی ازآنجائیکه شالوده این مدل راصریحاًبکارنمی‌برد، بدلایل‌غیرتجربی نادیده گرفتن کلیه

مشاهدات درفاصله بین دومنتهاالیه، افراط کارانه بنظرخواهدآمد ۰ روشی که بیشترمورد تحسین آمارگران قرارگرفته است "روش کمترین مجذورات" است. بمعادله زیرتوجه کنید :

$$yi= \beta_{yx}xi + \varepsilon i$$

که دراین مورد ما دارای تعداد $x_1, x_2 \ldots x_n$ مشاهدات مستقل هستیم

و

$y_1, y_2 \ldots y_n$

که درآن میانگین نمونه x, y, ε صفراست. حال اگرتمام معادله رادر xi ضرب کنیم خواهیم داشت:

$$yixi= \beta_{yx}x^2i + \varepsilon ixi \quad (i=1, 2 \ldots, n) \qquad (4-4)$$

این روش تعداد (n) معادله $(i= 1, 2 \ldots, n)$ ازنوع معادله (٤ـ٤) رادراختیارماقرار خواهدداد ۰ اینک یک معادله کلی راازجمع معادله (٤ـ٤) برای تعداد (n) مشاهدات درست میکنیم، که حاصل جمع آن معادله زیرمیشود :

$$\sum_{i=1}^{n}yixi= \beta_{yx}\sum_{i=1}^{n}x^2i + \sum_{i=1}^{n}\varepsilon ixi \qquad (4-5)$$

مقدار $\Sigma yixi$ حاصل جمع x و y است. وقتیکه معادله فوق برتعداد مشاهدات (n) تقسیم شود، معادله "کواریانس" (همپراش) نامیده میشود یعنی کواریانس $(x,y)=(1/n)\ \Sigma yixi$ کواریانس شبیه واریانس است ولی بجای مجذورکردن متغیرآن رادردیگری ضرب میکنیم ۰

درمعادله (٤ـ٥) مقدارحاصل جمع x و y رادرسمت چپ معادله حساب میکنیم چون تمام مقادیرایکسرها (xs) وایگرک ها (ys) رامیدانیم ۰درسمت راست معادله بسهولت میتوانیم جمع مجذورات xi راحساب کنیم یعنی Σx^2i (البته مقدار β رانمیدانیم ولی سعی میکنیم برآوردی ازآن رابدست آوریم ۰) جمله دوم سمت راست معادله رانمیتوانیم حساب کنیم چون مقدار (εi) را نمیدانیم، ولی میتوان اندازه این جمله رابراساس سایراجزاء؟ Σx^2i استنتاج کرد ۰ بدیهی است که Σx^2i جمع مجذورات است وطبیعی است که دقیقاً میتواند مثبت باشد+ ۰ ازطرف دیگرحاصل جمع

+ یک موردغیرجالب وعجیب وجودخواهدداشت وقتیکه $\Sigma x^2i = 0$ است وآن وقتی است که $x_1 = x_2 = \ldots = xn = 0$ (اگرانحصارکرقیمتی راتغییرندهد سپس نمیتوان چیزی راراجع بقانون تقاضا استنتاج کرد ۰)

$\Sigma\varepsilon i x i$ لزوماً ممکن است مثبت نباشد . چون فرض کرده ایم که (εi) مستقل ازمقدار x توزیع پیدا میکند پس میتوانیم جمع ایکس‌ها (xs) رادرجمع کل وزن های ثابتی فرض کنیم . اپسیلن ها (εs) تنها برحسب قانون احتمالات تغییرمیبابند . برای هرمقدارمعین xi میانگین توزیع احتمالی اپسیلن ها (εs) صفراست . بنابراین جمع $\Sigma xi\varepsilon i$ بصورت جمع وزنی متغیرهای اتفاقی که دارای میانگین صفراست بحساب میآوریم . درست همانطورکه درمورد نظریه معمولی نمونه گیری، متحمل ترین مقدار این جمع صفراست، ولی درعمل میتوان انتظارداشت که اگرآزمایش بدفعات زیادی تکرارشود ، مقادیر بطوریکنواخت تری توزیع خواهند شد ـــــ این مقادیرگاهی مثبت، گاهی منفی وگاهی هم دراطـــراف میانگین صفربراکنده خواهند شد .

بنابراین محتمل است که $\Sigma\varepsilon i x i$ نسبت به $\Sigma x i^2$ کوچکترشود وکواریانس (ε) و (x) احتمالا نسبت به واریانس x همانطورکه اندازه نمونه (n) افزایش میابد ، کوچکتروکوچکترشود .بمعادلـــه (۴ـ۵) بشکل زیرتوجه کنید که تمام آن بر $\Sigma x i^2$ تقسیم شده است:

$$\frac{\Sigma yixi}{\Sigma xi^2} = \beta_{yx} + \frac{\Sigma\varepsilon ixi}{\Sigma xi^2} \qquad (۴ـ۶)$$

بنظرمیآید که آخرین جمله آن نسبتاًکوچک است . پس این امرنشان میدهد که ماآخرین جمله رانادیده میگیریم وآن رابعنوان برآوردی برای β_{yx} بکارمیبریم .

$$\frac{\Sigma yixi}{\Sigma xi} = \hat{\beta}_{yx} \qquad (۴ـ۷)$$

دراین فرمول کلاهکی که روی قرارداردنشان میدهد که این برآوردی است ازاطلاعات داده شده ولی تفاوت بین برآورد ومقدارواقعی آن بشرح زیراست:

$$\hat{\beta}_{yx} - \beta_{yx} = \frac{\Sigma\varepsilon ixi}{\Sigma xi^2} \qquad (۴ـ۸)$$

اینک درنمونه گیری میانگین متوجه میشویم همانطورکه اندازه نمونه راافزایش دادیم، احتمال یک انحراف معین بین میانگین نمونه گیری ومیانگین واقعی ، کوچکتروکوچکترمیشود ومیانگین نمونـــه نیزدارای این گرایش است که بیشتردراطراف مقدارواقعی تراکم پیداکند . بهمین ترتیب همانطورکه اندازه نمونه افزایش میابد ، مقدار $\Sigma\varepsilon ixi/\Sigma xi^2$ دارای این گرایش است که نزدیکترونزدیکتر بمقدارصفرتمرکزپیداکند . بدین ترتیب برآوردهای $\hat{\beta}_{yx}$ همانطورکه اندازه نمونه افزایش میابـــد نزدیکترونزدیکتربه β_{yx} تراکم پیدامیکند . بنابراین میتوان چنین نتیجه گیری کرد ،درصورتیکه (ε) مستقل از x توزیع پیداکند ، برآورد $\hat{\beta}_{yx}$ ارزیاب (Estimator) ثابتی از β_{yx} خواهدبود . بطور کلی همانطورکه اندازه نمونه افزایش میابد ، مقداربرآورد شده $\hat{\beta}$ دارای این گرایش خواهدبودکـــه بمقدارواقعی β نزدیکترونزدیکترشود .