سیاوش امینی

آزمون همبستگی‌های چند عاملی در تحقیقات جامعه شناسی

گاهی اوقات هنگام تجزیه و تحلیل ارقام ، محقق جامعه شناس مجبور است که بیش از سه متغیر را همزمان مورد بررسی قرار دهد . اجبار محقق از آن جهت است که‌دربرخی از موارد در بررسی همبستگی متغیرها ، درصد قابل توجهی اختلاف غیر قابل توجیه‌ظاهر میگردد . تا جائیکه این واریانس غیر قابل توجیه‌مربوط به‌اشتباهات سنجش نباشد ،می‌توان نتیجه گرفت[1]: متغیرهای غیر وابسته ای که‌تابحال مورد توجه‌قرارگرفته‌اند کاملاازمتغیرهای وابسته تفکیک نشده اند و یا بعبارت ساده تر متغیرهای دیگری وجود دارند که بطورمنظم بر متغیرهای وابسته تأثیر میگذارند . بنابراین‌تأثیر متغیرهائی که‌محقق ازآنها غافل مانده موجب یک اختلاف توجیه ناشدنی میگردند . بدیهی است تجسم رابطه واقعی بین بیش از سه متغیر با استفاده از روش همبستگی و یا همبستگی شرطی کار بی اندازه‌مشکلی است. تجزیه و تحلیل همبستگی های سه عاملی با روش های قیاسی‌امکان پذیراست .اما زمانیکه تعداد متغیرها از سه تجاوز کند باید روش دیگری را بکار برد که بلا لوک[2] Blalock, آنرا تکمیل کرده است . این روش را می‌توان در واقع روش گسترده تجزیه‌وتحلیل جداول نامید . اختلاف این روش با روش تجزیه و تحلیل جداول آنست که الگوی همبستگی بطور

---

۱)ر.ک. به
Mayntz,R., Holm, K., Huebener, P., Einfuehrung in die Methoden der Empirischen Soziologie, 3. Aufl. Westdeutscher Verlag/ opladen 1972

2-Blalock, H.M.: Casual Interferences in Non-experimental Research, Chapel Hill, 1961.

قیاسی از ارقام گرفته نشده ، بلکه الگوهای همبستگی بین متغیرها بر اساس نظریه‌هائیست که بعداً" با توجه به محاسبات همبستگی از مواد جمع آوری شده با همبستگی های جزئی آزمایش میشوند . سپس از میان الگوها یکی از آنها که با مواد جمع آوری شده‌از همه‌بیشتر متناسب است انتخاب میشود . برای استفاده از این روش پیچیده باید قبلاً" یک سلسله فرضیات راجع به صفات الگوهای نظری ارائه دهیم :

۱ــ اولین اصل آنست که با این روش فقط الگوهای‌علی "قابل‌آزمون هستند . بعبارت‌دیگر رابطهٔ بین متغیرها نبایدرجعتی (رورسیبل) باشد .

مثلاً" چنانچه $A \rightarrow B$ تأثیر بگذارد در اینصورت تأثیر $B \rightarrow A$ مجازنیست.همبستگی‌های کارکردی(فونکسیونل)که در آنها همچنان که متغیر A برمتغییر B اثر میگذارد ، متغیر B نیز بر متغیر A تأثیر میگذارد در روش بلالوک قابلیت آزمون ندارند .

۲ــ همبستگی بین متغیرها باید خطی بوده و ارزش متغیرها قابل جمع باشد (همگی بر روی یک بعد) در غیراین‌صورت نمی‌توان از ضریب همبستگی استفاده کرد.بنابراین‌هر جا رابطهٔ بین متغیرهاغیرخطی(مثلاً" درجه۲) باشد ، در اینصورت باید با استفاده‌از لگاریتم آنها را بصورت خطی در بیاوریم .

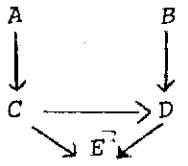
۳ــ کلیهٔ متغیرها بایددقیقاً" اندازه گیری و در الگو مطرح شوند .

۴ــ اثر کلیهٔ متغیرهای خارج از حوزهٔ الگو نباید منظم بلکه باید تصادفی باشد، بعبارت دیگر همبستگی های علی الگو نباید بوسیلهٔ این نوع متغیرها متأثر شود ، در غیر این‌صورت باید متغیر خارج از الگو را بداخل الگو منتقل نمود .

۵ــ متغیرهای خارج از الگو نباید دارای همبستگی باشند .

۶ــ اختلاف متأثر از اشتباهات سنجش در کار تحقیق باید تا حدامکان‌اندک‌باشد .

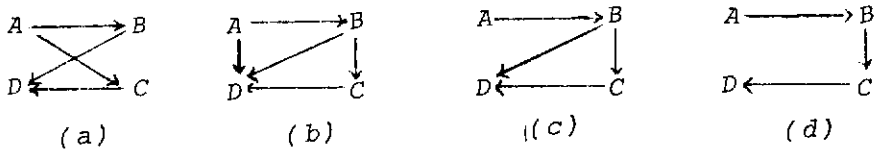
برای آزمایش الگوی نظری با مواد جمع آوری شده‌ی تجربی ، باید نه تنها تعداد متغیرها معین باشد ، بلکه کلیهٔ این متغیرها با توجه به جامعهٔ آماری محاسبه شده و فرضیه هائی نیز در زمینهٔ همبستگی های علی بین متغیرها در دست باشد . یک همبستگی علی بین۵ متغیر را می‌توان از طریق نمودار زیر نشان داد :



تصویر : ۱

با استفاده از این الگو می‌توان فرضیه هائی در باره ضریب همبستگی متغیرها

ارائه داد تا بتوان با توجه به ضریب همبستگی جزئی آنها را بطور تجربی آزمایش کرد. برای آنکه اساس این روش را بهتر توجیه کنیم از یک الگوی چهار عاملی استفاده‌می‌کنیم.



تصویر ۲:

ابتدا تصویر a را بررسی میکنیم. در این تصویر میتوان همبستگی بین متغیرهای A و B و C و D را که البته شدت آنها تا معلوم است بسادگی دید. ولی نمی‌توان راجع به همبستگی های علی بطور دقیق نظر داد و نمی‌توان گفت که همبستگی جزئی خاصی بین هر یک از متغیرها موجود است. ولی با حذف همبستگی بین متغیر A و C (عدم توجه به همبستگی درتصویر b می‌توان راجع به‌الگواظهارنظردقیق‌تری‌کرد. درتصویر b چنانچه‌اثر B ازمحاسبه‌حذف شود، همبستگی بین A و C نیز حذف میشود، زیرا A فقط از طریق B بر روی C اثر میگذارد. یعنی $r_{AC.B=0}$ بمثابهٔ یک نوع آزمایش تجربی‌است‌که در آن متغیر B ثابت فرض شده و بشرط وجود همبستگی بین متغیرها تغییرات متغیر A هیچگونه تغییری در میانگین متغیر وابسته C ایجاد نمی‌کند. چنانچه علاوه بر این، همبستگی بین متغیر A و D نیز حذف شود (تصویر C) و بدین ترتیب متغیر A فقط یک همبستگی غیر مستقیم با C (از طریق متغیر B به D) داشته‌باشد، می‌توان‌گفت که:

$$r_{AC.BC=0}$$

بعبارت ساده تر با ثابت نگهداشتن متغیرB و C همبستگی بین A و D ازبین‌میرود. چنانچه این الگو معادل با ارقام جمع آوری شده باشد باید همبستگی های جزئی (Partial Correlation) $r_{AC.B}$ و $r_{AD.BC}$ هر دو مساوی صفر باشند. حال چنانچه همبستگی بین متغیرهای B و D نیز حذف شود (تصویر d)، در اینصورت یک زنجیر علی‌ساده بدست میآید که‌درآن‌فرمول های زیر معتبر میشوند:

$$r_{AC.B=0}$$

$$r_{AD.BC=0}$$

$$r_{BD.C=0}$$

با توجه باین موضوع همزمان همبستگی های جزئی اساسی $r_{AD.B}$ و $r_{AD.C}$ معادل صفر میگردند. در بارهٔ این زنجیر علی ساده میتوان گفت: وقتی همبستگی یک متغیر تأثیر گذار محاسبه میشود، همبستگی بین دیگر متغیرها (حق بطور غیر مستقیم) از بین میرود. ازطرف دیگر این اصل معتبر است که همبستگی بین متغیر اول ومتغیر آخرمعادل حاصل ضرب همبستگی بین متغیرهای تأثیر گذارندهٔ دیگر است یعنی:[۱]

$$r_{AD}=r_{AB}.r_{BC}.r_{CD}$$

درعین حال این همبستگی همیشه کوچکتر از همبستگی مستقیم بین هرجفت متغیرهای مربوط است. بعبارت ساده ترطرف چپ تساوی کوچکتر از هریک از عناصر طرف راست تساوی است یک الگوی معین علی نه تنها براساس نبودن همبستگی یا همبستگی جزئی استوار است، بلکه تحت شرایطی معین براساس این ادعا است که یک همبستگی معین حاصل دو یا چند همبستگی دیگر است برای آن که نشان دهیم چگونه ازاین روش درالگوهای پیچیده تر استفاده میشود، ازیک مثال همبستگی شش عاملی استفاده می کنیم که نمودار آن بشکل زیر است (تصویر ۳)

از میان ۱۵ همبستگی دو بعدی ممکن بین شش متغیر[۲]، در الگوی بالا بنابر آن چه که گذشت فقط ۷ همبستگی مورد توجه است. بنابراین الگوی فوق ازطریق ۸ فرضیه همبستگی جزئی که باید معادل صفر باشند مشخص میشود. برای آنکه الگوی فوق را با ارقام بدست آمده تجربی قیاس کرد، باید تساویهای زیربرقرار باشند[۳]:

---

۱) *Koenig,R: Handbuch der Empirischen Sozial forschung 1973,3.Band*

۲ تعداد همبستگی های ممکن از فرمول $\frac{n(n-1)}{2}$ بدست میآید که در آن n تعداد متغیرهای الگوست: $\frac{۶(۶-۱)}{۲} = ۱۵$

۳ این اصول را در کتب بیشماری که در زمینهٔ آمار موجود است می توان یافت.

$r_{AB} = 0$ $r_{AD} = 0$

$r_{CB} = 0$ $r_{CD} = 0$

$r_{AE.BCD} = 0$ $r_{BF.ACDE} = 0$

$r_{DE.ABC} = 0$ $r_{CF.ABDE} = 0$

علاوه بر این در این الگو باید $r_{AE.B} = r_{AC} . r_{CE}$ باشد.

با استفاده از یک مثال عددی (همبستگی ۴ عاملی) موضوع را روشن میکنیم :

فرض کنیم که برای توجیه علی سطح عملکرد دانش‌آموز در مدرسه (Z) متغیرهای زیر معتبر باشند :

۱ــ موقعیت اقتصادی ــ اجتماعی پدر (W)

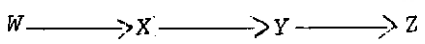
نحوه تربیت پدر (X)

۳ــ حیثیت گروهی دانش‌آموز در کلاس مدرسه (Y)

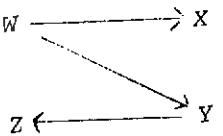
حال بر اساس فرضیه های ممکن می‌توان دو الگو برای این مطلب ارائه داد .

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

الگوی اول



الگوی دوم



برای الگوی اول می‌توان تساویهای زیر را بر قرار کرد :

$r_{XZ} + r_{XY} . r_{YZ}$

$r_{WY} = r_{WX} . r_{XY}$

$r_{WZ} = r_{WX} . r_{XY} . r_{YZ}$

برای الگوی دوم فقط میتوان دو تساوی ارائه داد که یکی از آنها رانمی‌توان درالگوی اول یافت :

$$r_{XZ+}{}^{r}{}_{XY}.{}^{r}{}_{YZ}$$

$$r_{WZ}={}^{r}{}_{WY}.{}^{r}{}_{YZ}\ ;$$

حال فرض میکنیم که از ارقام جمع آوری شده بطریق تجربی (empirie) همبستگی های زیر محاسبه شده اند (جدول ۱)

	W	X	Y	Z
W	—	۰/۴۹	۰/۵۳	۰/۳۹
X	۰/۴۹	—	۰/۶۱	۰/۵۱
Y	۰/۵۳	۰/۶۱	—	۰/۸۰
Z	۰/۳۹	۰/۵۱	۰/۸۰	—

جدول ۱ ــ ماتریس همبستگی

حال میتوان ارزش های موجود ماتریس راباهمبستگی‌هائی‌که از فرضیه الگوها بدست‌آمده‌اند مقایسه کرد. نتیجه این مقایسه در جدول ۲ نشان‌داده‌شده‌است. درست است که نمی‌توان اختلاف معنی‌دار بین همبستگی‌ها رادراین‌حالت‌محاسبه‌کرد، امااین مقایسه نشان میدهد که الگوی دوم با ارقام جمع‌آوری شده هماهنگی بیشتری دارد.

جدول ۲ ــ مقایسه ارزش های انتظاری وارزش های واقعی

فرضیه‌های‌الگو	ارزش های واقعی	ارزش های انتظاری
	الگوی اول	
$r_{XZ}={}^{r}{}_{XY}.{}^{r}{}_{YZ}$ ؟	۰/۵۱	(۰/۶۱)×(۰/۸۰) = ۰/۴۹
$r_{WY}={}^{r}{}_{WZ}.{}^{r}{}_{XY}$	۰/۵۳	(۰/۴۹)×(۰/۶۱)= ۰/۳۰
$r_{WZ}={}^{r}{}_{WX}.{}^{r}{}_{XY}$	۰/۳۹	(۰/۴۹)×(۰/۶۱)×(۰/۸۰)= ۰/۲۴
$r_{.YZ}$		
	الگوی دوم	
$r_{XZ}={}^{r}{}_{XY}.{}^{r}{}_{YZ}$	۰/۵۱	(۰/۶۱)×(۰/۸۰)= ۰/۴۹
$r_{WZ}={}^{r}{}_{WY}.{}^{r}{}_{YZ}$	۰/۳۹	(۰/۵۳)×(۰/۸۰)= ۰/۴۲

در الگوی اول همانطوریکه از جدول پیداست همبستگی بین موقعیت اقتصادی ـ اجتماعی پدر و موقعیت گروهی پسر در کلاس درس (r_{WX}) و همبستگی بین موقعیت اقتصادی ـ اجتماعی پدر و سطح عملکرد دانش آموز در مدرسه (r_{WZ}) بزرگتر از ارزش های انتظاری است که در جدول نشان داده شده است . در الگوی دوم که موقعیت اقتصادی اجتماعی پدر مستقیما" با حیثیت گروهی دانش آموز در رابطه است همبستگی علّی بین موقعیت اقتصادی ـ اجتماعی و عملکرد دانش آموز کمتر از طریق همبستگی بین موقعیت اقتصادی ـ اجتماعی و عملکرد او افزایش میابد. این امر با آزمون تجربی نیز هماهنگی دارد . البته باید یادآور شویم که در این مثال نباید تصور کرد که با ایجاد الگوی دوم فرضیه تأئید شده است. ولی می‌دانیم که الگوی دوم بهتر از الگوی اول با واقعیت تجربی سازگار است . بدیهی است که امکان وجود یک الگوی دیگر نیز که نتایج مشابهی را ارائه دهد موجود است ولی در اینجا بمنظور رعایت اختصار از ذکر آن صرف نظر شده است چنانچه دو الگو به نتیجه مساوی برسند ، یا باید انتخاب یکی از این الگوها بر اساس تعمق نظری باشد یا متغیرهای جدیدی را بهر دو الگو اضافه کرده و روش را دوباره تکرار کنیم . یکی از موارد استفاده و طرز آزمون همبستگی های چند عاملی تجزیه و تحلیل گروههای ناهمگن است که سان کوئیست و مورگان Songist & morgan آنرا تکمیل کرده اند. در این روش سعی بر آنست که با استفاده از ترکیبی از صفات (ضرایب ویژگی) متغیرهای غیر وابسته را برحسب (سن ، جنس ، وابستگی به حزب معین . . . . .) برای نمونه‌های مورد مطالعه ارائه دهیم برای این منظور پدیدار یا جامعه اصلی را ابتدا بطور کامل بدو گروه تقسیم میکنیم بطوری که هر گروه از طریق صفت (ضریب ویژگی) یک متغیر وابسته مشخص شود . سپس هریک از گروه‌ها باز هم بر اساس یک متغیر وابستهٔ دیگر ، گروه بندی میشوند ، تا آخر . . . . .) این کار تا جائی ادامه می‌یابد که دیگر تقسیم گروهها بنا بقاعده ای معین امکان‌پذیر نیست . لازمه این گروه بندی آنستکه اولا هر گروهی با توجه به متغیر وابسته بطور درونی از جامعه اصلی همگن تر باشد و ثانیا" آنکه گروهها تا حدممکن بطور برونی ناهمگن (هتروگن Heterogen) باشند . با توجه به تلفیق صفات (ضرایب ویژگی) در گروههای نهائی می‌توان تحت شرایط معینی ضریب متغیر وابسته را توجیه نمود . برای مثال میتوان در تجزیه و تحلیل در آمد متوسط، خرده گروههای زیر را تشکیل داد :

یک خرده گروه بسیار همگن (تعیین این گروه از طریق انحراف بسیار اند ک در آمد فردی از میانگین حسابی در آمد امکان پذیر است) با درآمد زیاد که از طریق صفات انتخابی فی المثل تحصیل دانشگاهی ، جنس مرد و سن بین ۳۵ تا ۵۵ سال مشخص میشود (مانند

استادان مرد دانشگاه)

خرده گروه همگن دوم متوسط درآمد اندک از طریق صفات انتخابی دیگر، مانند تحصیل دوره ابتدائی، جنس زن و سن کمتر از ۳۵ سال مشخص میشود. در این میان ممکن است گروهی نیز وجود داشته باشد که درآمد متوسط آن بین بسیار کم و بسیار زیاد قرار گرفته باشد و زیاد هم همگن نباشد و مثلا" از طریق صفات زیر: داشتن دیپلم متوسطه جنس مرد و کمتر از ۳۵ سال مشخص شود، باید دانستکه در گروههای همگن صفت مشخصهٔ متغیر غیر وابسته داری یک ارزش توجیهی دقیق برای مقدار درآمد متوسط گروه است توزیع گروه اصلی به خرده گروهها بر اساس اصل پراکندگی آماری است. همانطورکه می دانیم پراکندگی معیاری برای همگن بودن یا ناهمگن بودن هر گروه است و آنرا مربع انحراف هر نمونه مورد مطالعه از میانگین گروه مربوط گویند. بنابراین هر چه ارزش های انفرادی ازحد متوسط بیشتر منحرف شده باشند بهمان اندازه نیز گروه ناهمگن تر بوده و پراکندگی بیشتر است. ولی باید توجه داشت که در روش گروههای ناهمگن فرمول پراکندگی تغییر داده شده و بجای:

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_1-\bar{x})^2$$

باضرب کردن دوطرف تساوی در n،

از فرمول $n.s^2=\sum_{i=1}^{n}(x_1-\bar{x})^2$

استفاده میشود. این اصطلاح را مجموع مربعات QS (Total sum of squares) مینامند. اصل قاعده پراکندگی یا در این مورد بخصوص مجموع مربعات میگوید: کل مجموع مربعات $QS(n.s^2)$ همه واحدها (n) در گروه اصلی را می توان به مجموع مربعات درونی و برونی قسمت کرد. مجموع مربعات درونی $intern\ QS$ مجموع مربعات گروه پائین (در اینجا ۲ یعنی: $(n_1s_1^2+n_2s_2^2$ است) و مجموع مربعات برونی (با تعداد واحدهای هر گروه جزئی یا خرده گروه $(n_1,n_2$ مربع انحراف میانگین توازنی خرده گروه (x_1,x_2) از میانگین گروه اصلی است یعنی:

$$ns^2=n_1s_1^2+n_2s_2^2+n_1(\bar{x}_1-\bar{x})^2+n_2(\bar{x}_2-\bar{x})^2$$

$$ns^2=\underbrace{\sum_{i=1}^{2}n_i\,s_i^2}_{\text{مجموع مربعات درونی}}+\underbrace{n_i\sum_{i=1}^{2}(\bar{x}_i-\bar{x})^2}_{\text{مجموع مربعات برونی}}$$

عبارت اول این تساوی در قسمت راست $\sum_{i=1}^{2} n_i s_i^2$ معیاری

برای همگن بودن درون هر خرده گروه و عبارت دوم $n_i \sum_{i=1}^{2} (\bar{x}_1 - \bar{x})^2$

معیاری برای ناهمگن بودن خرده گروهها با هم است. حال میتوان این روش را دقیق تر تفسیر کرد:

هدف ما اینست که از میان کلیه متغیرهای غیر وابسته، متغیری را بیابیم که گروه‌بندی‌ها را بطوری مجاز میدارد که مجموع مربعات درونی *Within sum of squares* بحداقل رسیده و مجموع مربعات برونی *between sum of squares* بحدا کثر برسد. مجموع مربعات برونی معیاری برای توجیه و اریانس متأثر از گروه‌بندی یا بعبارت دیگر کاهش ناهمگنی گروه اصلی است بعکس مجموع مربعات درونی معیاری برای واریانس غیر قابل توجیه و یا بسخن دیگر ناهمگنی موجود در گروههای زیرین است به حداکثر رساندن مجموع مربعات درونی در هر مرحلهٔ تجزیه انجام شده و خود معیاری برای انتخاب متغیر غیروابسته در مرحله بعدی گروه‌بندی ست. در اینکه کدامیک ازمتغیرها به نتیجه مطلوب میرسد از طریق آزمایش امکان پذیر میشود. چنانچه یک متغیر غیروابسته بیش از دو ضریب داشته باشد، در اینصورت برای دو گانگی این متغیر نیزترکیب مطلوبی از ضرایب بطریق آزمایش محاسبه میشود.

روشی را که شرح دادیم اکنون مختصراً با یک مثال روشن می‌کنیم:

می‌خواهیم بررسی کنیم که درآمد متوسط شاغلین در صنعت بچه چیز وابستگی دارد. (متغیر وابسته درآمد معمولاً یک درجه دوره‌ایست،)* متغیرهای غیر وابسته درآمد فرضاً موقعیت شغلی (کارمند ــ کارگر**)، جنس (زن ــ مرد)** و گروههای متفاوت از نظر کار کرد عملی (گروه شغلی) باشند (گروه شغلی ۱ــ گروه شغلی ۲ــ گروه شغلی ۳) *** ن. ک. به تصویر ۵

---

* درجه دوره ای *Interval scale*

** درجه اسمی *nominal scale*

*** درجه نظمی *ordinal scale*

نمونه‌های مورد مطالعه n=۴۰۰ است که‌بصورت جدول‌زیرتوزیع شده‌است (۱) .

شغل	کارمند						کارگر					
جنس	مرد			زن			مرد			زن		
گروه	۱	۲	۳	۱	۲	۳	۱	۲	۳	۱	۲	۳
$\bar{x}$	۱۴	۹	۷	۱۰	۸	۶	۹	۸	۷	۷	۶	۵	
n	۱۰	۱۰	۲۰	۱۰	۱۰	۴۰	۴۰	۴۰	۸۰	۱۰	۵۰	۸۰	۴۰۰
$n.\bar{x}$	۱۴۰	۹۰	۱۴۰	۱۰۰	۸۰	۲۴۰	۳۶۰	۳۲۰	۵۶۰	۷۰	۳۰۰	۴۰۰	۲۸۰۰

جدول شماره ۳ ــ درآمد متوسط شاغلین درصنعت، $\bar{X}$ درآمدمتوسط × ۱۰۰ واحد پولی nفراوانی

در مرحله اول میانگین کلیه ارقام محاسبه میشود . این میانگین‌معادل میانگین‌توازنی کلیه میانگین های انفرادی است : $\frac{1}{n}\sum \bar{x}n = \frac{1}{۴۰۰}\times ۲۸۰۰ = ۷$

مجموع مربعات QS (یعنی مجموع مربع انحراف های درآمد ۴۰۰ نفر مزبور از میانگین توازنی درآمد (۷×۱۰۰=۷۰۰) معادل ۲۷۴۰ است . اکنون باید بیآزمائیم که کدامیک از سه متغیر غیر وابسته یعنی موقعیت شغلی ، جنس و گروه شغلی بهترین‌عامل گروه‌بندی‌است یا بعبارت دیگر دو گروه زیرین را با کوچکترین مجموع مربعات درونی و بزرگترین مجموع مربعات برونی ارائه میدهد . معیار آزمون ، مجموع مربعات برونی است زیرا ساده ترقابل محاسبه است . اولین متغیر جنس است بدین ترتیب که گروه اصلی بدو گروه زیرین تقسیم میشود که عبارتند از شاغلین زن‌و شاغلین مرد، حال برای هرگروهی‌میانگین‌درآمدمحاسبه میشود :

درآمد متوسط شاغلین مرد $\bar{x}_m = \frac{1}{n_m}\sum_{i \varepsilon m} x_i = \frac{۱۶۱۰}{۲۰۰} = ۸/۰۵$

درآمد متوسط شاغلین زن $\bar{x}_f = \frac{1}{n_f}\sum_{i \varepsilon f} x_i = \frac{۱۱۹۰}{۲۰۰} = ۵/۹۵$

مجموع مربعات برونی (توجیه شده) عبارت است

$$ns^2_{ext} = (\bar{x}_m - \bar{x})^2 . n_m + (\bar{x}_f - \bar{x})^2 . n_f$$

$$= (۸/۰۵ - ۷)^۲ \times ۲۰۰ + (۵/۹۵ - ۷)^۲ \times ۲۰۰ = ۴۴۱$$

می‌بینیم که توزیع گروه اصلی بر حسب جنس، دو گروه ارائه میدهد که در میانگین با هم اختلاف آشکاری دارند، مجموع مربعات درونی گروه های جدید دقیقا" بمقدار ۴۴۱ (قدر مطلق مجموع مربعات درونی) کوچکتر ازمجموع‌مربعات‌گروه‌اصلی‌یا ۱۶/۱% = ۱۰۰×$\frac{۴۴۱}{۲۷۴۰}$ آنست. توجیه دقیق این مطلب آنست که ناهمگنی گروه اصلی (کل جامعه نمونه) در اثر تقسیم بدوگروه‌زیرین‌برحسب‌جنس، بمقدار ۱۶/۱%کاهش‌یافته‌است. درمرحله دوم موقعیت شغلی بررسی میشود. در اینجا گروه اصلی بر حسب موقعیت. شغلی بدو گروه زیرین کار مند و کارگر توزیع میشود. ارزش های بدست آمده عبارتند از.

$ns^2_{ext} = ۱۰۸$ $\quad \bar{x} = ۶/۷$ کارگر $\quad \bar{x} = ۷/۹$ کارمند

عدد بدست آمده برای مجموع مربعات برونی در اینجافقط ۳/۹درصدناهمگنی‌گروه اصلی را کاهش میدهد و بنابراین نامناسب بر از توزیع گروه ها بر حسب جنس است.

بدیهی است که گروه های شغلی نیز باید بهمین ترتیب محاسبه شود ولی‌دراین جا همانطور که اشاره کردیم متغیر دارای سه ضریب یعنی گروه شغلی ۱و ۲و ۳ است بنابراین امکانات گروه بندی به بیش از دو گروه افزایش مییابد یعنی:

گروه شغلی ۱ با گروه شغلی ۲و۳

گروه شغلی ۱و۲ با گروه شغلی ۳

گروه شغلی ۱و۳ با گروه شغلی ۲

باتوجه به روشی‌که ارائه میشود لازم نیست که ترکیب های مختلف گروهی همگی محاسبه شود: کافیست که تعداد ضرایب (در این جا معادل ۳) را طوری تنظیم‌کنیم‌که‌ارزش‌متغیر بزرگتر در جلو قرار گیرد (در اینجا درآمد متوسط‌گروه شغلی ۱ بیشتر از درآمدمتوسط گروه شغلی۲ و درآمد متوسط‌گروه شغلی ۲ بیشترازدرآمد متوسط‌گروه شغلی ۳ است‌یعنی:

$$I_1 > I_2 > I_3$$

بنابراین اصل توزیع گروه ها بصورت زیر کافیست

k_1	با	$k_2 + k_3 + \cdots + k_n$
$k_1 + k_2$	با	$k_3 + k_4 + \cdots + k_n$
$\vdots$	$\vdots$	$\vdots$
$k_1 + k_2 + \cdots + k_{n-1}$	با	k_n

در مثال بالا کافیست که گروههای زیر را تشکیل دهیم :

گروه شغلی ۱ با گروه شغلی ۲و۳

گروه شغلی ۱و۲ با گروه شغلی ۳

اعدادی که پس از محاسبه بدست میآید عبارتند از :

$\overline{x} = ۶/۴۵$ $ms^2_{ext} = \%۲۰/۵$

(۲+۳) ـ گروه‌شغلی

$\overline{x} = ۹/۵۷$

(۱) ـ گروه‌شغلی

$\overline{x} = ۶/۱$ $ns^2_{ext} = \%۱۴/۷$

(۳) گروه‌شغلی

$\overline{x} = ۸/۱$

(۱+۲) گروه‌شغلی

در اینجا معلوم میشود که توزیع گروه اصلی بر حسب متغیر غیر وابسته‌گروه‌شغلی‌زیر یعنی شاغلین گروه شغلی ۱ و شاغلین گروه شغلی ۲و۳ بهترین نوع توزیع ممکن است زیرا ناهمگنی گروه اصلی را به ۲۰/۵ در صد تقلیل میدهد .

پس از توزیع گروهها به گروه های زیرین مرحله ۱ بایدتوزیع‌راادامه‌داده وگروه های زیرین مرحله ۲ و غیره‌بهمین‌ترتیب تشکیل گردیده‌ومحاسبه‌شوند .این‌کارمنجربه‌یک تصویر شاخه مانند میگردد که نام این روش نیز ازاین‌تصویرگرفته‌شده‌است(Tree Analysis)

ادامه گروه بندی تابع قواعدی است که در زیر بانها اشاره می‌کنیم :

۱ـ زمانی که تعداد واحدهای مورد مطالعه در یک گروه خیلی کوچک شود (اشتباه متوسط نمونه خیلی زیاد شود) باید از ادامه گروه بندی خودداری شود .

۲ـ زمانی که مجموع مربعات گروه زیرین خیلی کوچک بشود (وقتی مجموع مربعات خیلی کوچک بشود ، در اینصورت ادامه گروه های زیرین‌تغییرمحسوسی درکاهش‌ناهمگنی‌گروه‌های بالاتر نخواهد داشت) گروه بندی متوقف میگردد .

۳ـ وقتی که در صد مجموع مربعات برونی QS_{ext} گروه زیرین نسبت به‌مجموع‌مربعات گروه اصلی خیلی کوچک بشود یعنی گروه‌بندی توجیه قانع‌کننده‌ای‌نداشته باشد باز هم گروه بندی متوقف میگردد .

۴ـ وقتی که تعداد گروه های زیرین خیلی زیادبشود ،دراینصورت توجیه‌نظری آنها کار بسیارمشکل و درعین حال بیهوده ای است .

تعیین ارزشهائی که گروه بندی را متوقف میکنند یک عمل تجربی است : معمولا" وقتی که گروههای زیرین ۱ تا ۲ درصد QS اصلی را بیان میکنند و تعداد نمونه ها به زیر ۱۰میرسد ادامه گروه بندی جایز نیست . چنانچه متغیر وابسته دو گونه باشد *Dichotomic*

محاسبه بسیار ساده میگردد ، مثلا " چنانچه متغیر وابسته بصورت سئوال ، آیا بدبیرستان میروید ؟ و جواب بلی " و "نه" باشد در اینصورت میتوان دو گونگی متغیروابسته‌راباتوجه به متغیر غیر وابسته (کارگر بودن یا کارگر نبودن پدر) بصورت جدول زیر نوشت :

جدول ۴ ــ رفتن به دبیرستان دررابطه باشغل پدر

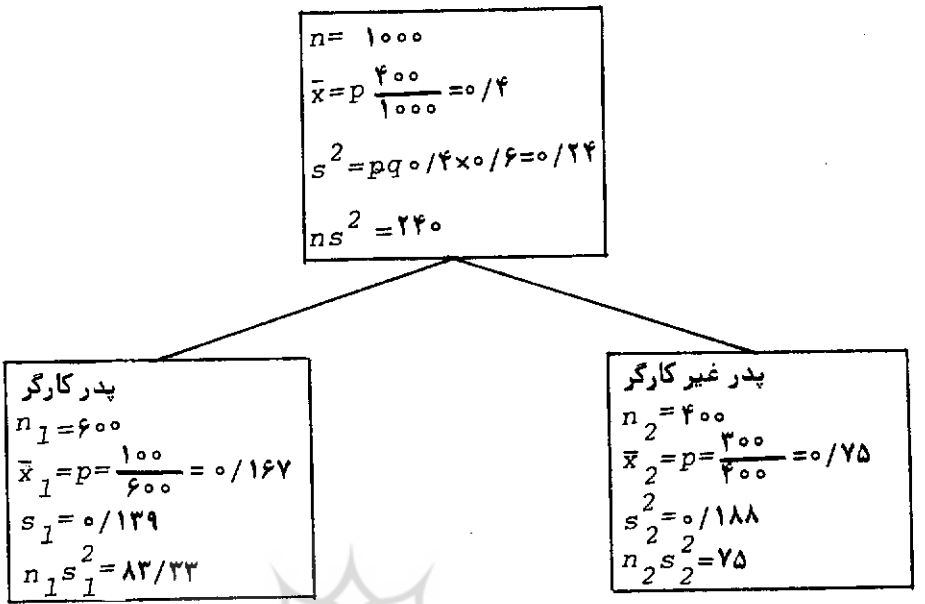
متغیر غیر وابسته / متغیر وابسته		شغل پدر کارگر: خیر	شغل پدر کارگر: بلی	
متغیر وابسته	x_i \ n_i	n_1	n_2	Σ
دوره‌دبیرستان	۱ بلی	۱۰۰	۳۰۰	۴۰۰
	۰ خیر	۵۰۰	۱۰۰	۶۰۰
Σ		۶۰۰	۴۰۰	۱۰۰۰

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

چنانچه ضرایب متغیر وابسته را با ارزش های ۱و۰ مشخص کنیم ، داریم :

$$\bar{x}=p=\frac{1}{n}\ (x_1 n_1+x_2\ n_2)=\frac{1}{1000}\ (1\times 400+\ 0\times 600)=\ 0/4$$

در اینجا (P) ارزش جزئی آندسته از نمونه‌ها را نشان میدهد که دارای ضریب ۱ برحسب‌متغیر وابسته هستند (دراینجا : بدبیرستان میروند) . محاسبه پراکندگی درحالت دو گونگی متغیر وابسته با این فرمول محاسبه میشود $s^2=p(1-q)=pq$ که q ارزش جزئی آندسته ازنمونه‌هائیست‌که‌دارای‌ضریب‌صفر برحسب متغیروابسته‌هستند (در اینجا . بدبیرستان نمی‌روند) .بنابراین برای کلیه ارقام داریم :



تصویر ۵ بریده‌ای از یک گروه بندی

$Qs=(\bar{x}_1-\bar{x})^2 n_1+(\bar{x}_2-\bar{x})^2 n_2=۸۱/۶۷$

مجموع مربعات :

$ns^2-(n_1s_1^2+n_2s_2)^2=۲۴۰-(۷۵+۸۳/۳۳)=۸۱/۶۷$

اکنون با استفاده از یک مثال (گروه بندی) در تصویر۵ - بطورمختصر به نحوه‌توجیه متغیرهای مربوط می‌پردازیم :

اولین توجیه این مثال آنست که گروههای نهائی بطور درونی همگن‌هستند (هرگروه فقط جزء کوچکی از مجموع مربعات گروه اصلی است) . در مقابل گروههای نهائی آشکارابا هم اختلاف دارند (ارزش های متوسط هر یک ، اختلاف شدیدی با هم دارند) . از طرف دیگر بطریق گروه بندی بر حسب متغیرهای غیر وابسته ۳۷٪ پراکندگی گروه اصلی قابل توجیه است .

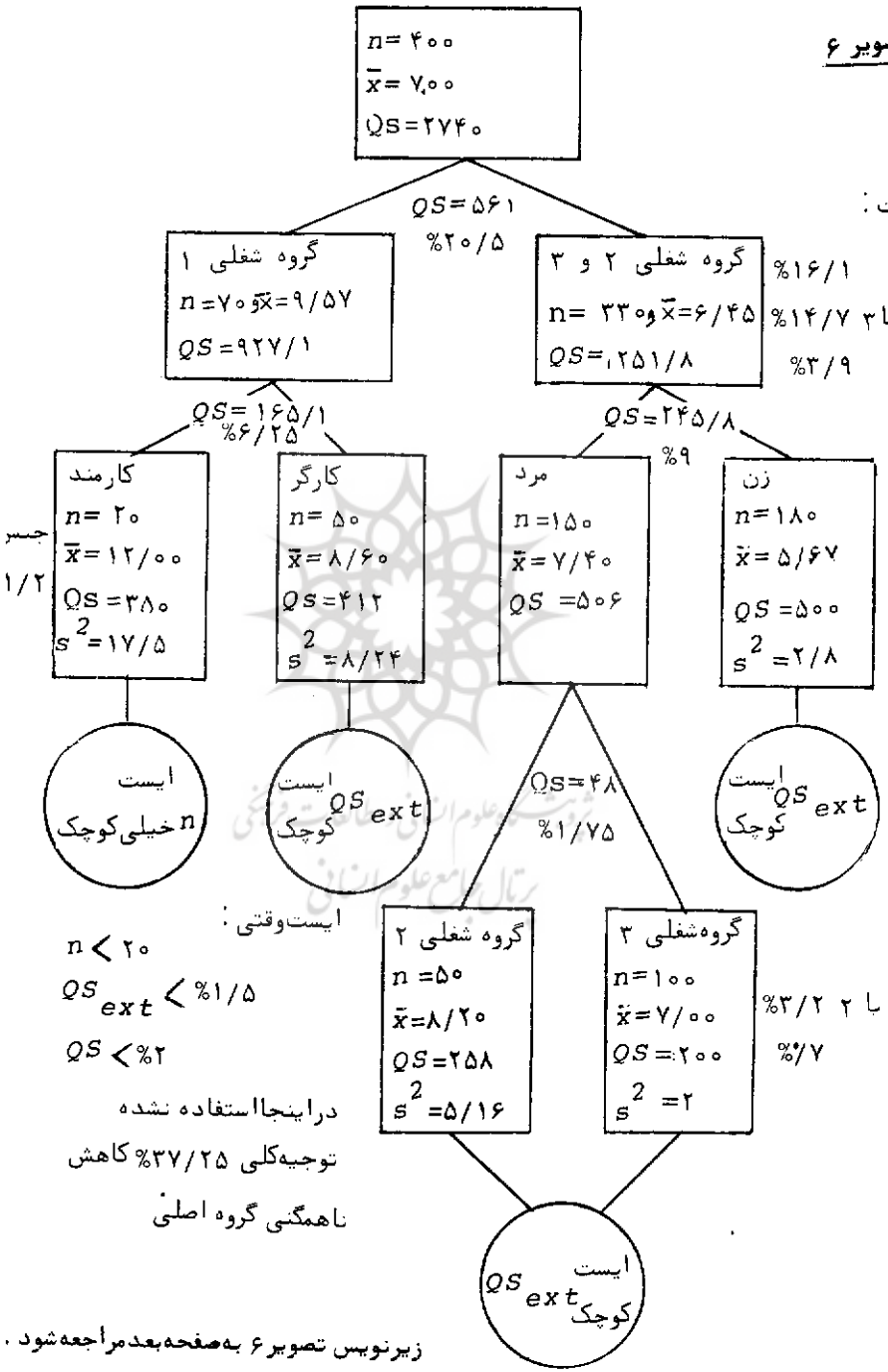
عدم تقارن موجود در گروه بندی (یک متغیر فقط در یک شاخه‌موردتوجه‌قرار گرفته یا در یک شاخه زودتر بررسی شده است) اجازه توجیهات عطفی در گروههای بالا تر را می‌دهد . بدین ترتیب که في المثل متغیر جنس، اختلاف متوسط در آمد را در گروههای شغلی پائین نشان می‌دهد ولی نه در گروه شغلی یک ، این بدان معنی است‌که‌زنهابخصوص در موقعیت های شغلی نامناسب تر بیشتر تحت فشار هستندودراین‌مشاغل جنسیت بیش

از عملکرد (گروه شغلی) بر درآمد تأثیر می‌گذارد . بعقیده سان‌کوئیست و مورگان Sonquist & Morgan این نوع عدم تقارن‌ها معرفهائی برای کنش‌های متقابل (interaction) یا همبستگی‌های متقابل بین چند متغیر هستند . امکان یافتن این همبستگی‌ها ارزش عالی این روش را بیان می‌کند ولی ارزش واقعی این روش در کاهش همبستگی‌های بیشمار و پیچیده بصورت ساده و قابل فهم است .

یکی از خواص عمده این روش آنست که از پیش هیچگونه قیدوشرط یا فرضیه‌ای مربوط به نوع همبستگی ها (مثلا " همبستگی خطی و شبیه حرف U و غیره موجود نیست) ، مثلا" رابطه متغیر غیر وابسته با مقوله های " خیلی زیاد و خیلی کم " با "نسبتا "زیاد "یک رابطه دو گونه بصورت U است . آزمون معنی دار بودن رابطه بین همبستگی ها در این روش هنوز تکامل نیافته است و این نیز نقص نیست ، زیرا این روش بیشتر برای توجیه نظری همبستگی ها معتبر است تا اثبات وجود همبستگی از طریق آماری .

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

تصویر ۶



امکانات :

جنس %۱۶/۱

شغلی ۱و۲با۳ %۱۴/۷

شغل %۳/۹

جنس ۱/۲

گروه شغلی ۳ با ۲ %۳/۲

شغل %۷

ایست‌وقتی :

$n < ۲۰$

$QS_{ext} < \%۱/۵$

$QS < \%۲$

دراینجااستفاده نشده

توجیه‌کلی ۳۷/۲۵% کاهش

ناهمگنی گروه اصلی

زیرنویس تصویر۶ به‌صفحه‌بعدمراجعه‌شود .

توضیحات مربوط به تصویر ـ۶ـ

Qs= مجموع مربعات

Qs_{ext}= مجموع مربعات محاسبه شده برونی دریک ستون

%= درصدتوجیهی مجموع مربعات برونی به‌مجموع مربعات

مثلا" در ستون ۱ $\frac{۵۶۱}{۲۷۴۰} = \%۲۰/۵$

برای آزمون صحت می‌توان ازفرمول زیراستفاده کرد :

$$Qs = Qs_{int} + Qs_{ext}$$

$$۲۷۴۰ = (۹۲۷/۱ + ۱۲۵۱/۸) + ۵۶۱$$

s^2= پراکندگی گروههای نهائی $\frac{Qs}{n}$ معرف همگنی گروههای نهائی است .

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع و ماخذ:

Blalock, H.M., Causal Interferences in Nonexperiemental Research, Chapel Hill, 1964

Boudon, R., A New Look at Correlational Analysis, Methodology in Social Research, N.Y. 1968

Hyman, H.H., Survey Design and Analysis, Glencoe, Ill., 1966

Koenig, R. Handbuch der empirischen Sozialforschung, 1974

Mayntz, R. Holm, K. Huebner, P., Einfuehrung in die Methoden der Empirischen Soziologie, 3. Aufl. Westdeutscher Verlag/Opladen 1972

Sonquist J.A. Morgan, J.N. The Detection of Interaction Effects, Monog. No. 35, Survey Research Center, Michigan, 1964

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی