

بسم الله الرحمن الرحيم

ن والقلم و ما يسطرون

قسم به قلم و آنچه می بخارند



دانشگاه شهید بهشتی

دانشکده علوم ریاضی

گروه آمار

پایان نامه دوره کارشناسی ارشد آمار اقتصادی و اجتماعی

اثر علی در مطالعات مشاهده ای

استاد راهنما

دکتر محمدرضا فرید روحانی

استاد مشاور

دکتر مسعود البرز

نگارنده

مرضیه رعیتی

دی ۱۳۸۸

۱۴۲۶۶۸

۱۳۸۹ / ۷ / ۲۴

کتابخانه دانشگاه شهید بهشتی
تاسیس ۱۳۵۷

کلیه حقوق اعم از چاپ و تکثیر، نسخه برداری، ترجمه، اقتباس و

... از این پایان نامه برای دانشگاه شهید بهشتی محفوظ است. نقل

مطالب با ذکر مأخذ بلامانع است.

تقدیم به

دو گوهر گرانبهای گیتی، پدر و مادر عزیزم

و تقدیم به آنانی که قلبشان برای اعتلای ایران می‌تپد و خالصانه برای

پیشرفت آن می‌کوشند. زنده باد دکتر قریب‌ها و پروفسور حسابی‌ها...

سپاس‌گزاری

منت‌خدای را عزوجل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت.

سپاس‌خدای را که به من نعمت تفکر آموخت و مرا در طریق تحصیل و علم آموزشی یاری فرمود. و سپاس‌فزون‌تر از اینکه در مسیر پایان‌نامه، مرا رهنمایی شایسته و استادی بزرگوار، همچون جناب آقای دکتر فریدروحانی قرار داد تا با همراهی‌های بی‌دریغ ایشان، موفق به طی این طریق شوم و علاوه بر علم، اخلاق عملی را بیاموزم. از ایشان صمیمانه تشکر و قدردانی می‌نمایم. باشد که همواره در پناه ایزد منان و عنایات خاص حق تعالی باشند.

سعی‌نابرده، در این راه به جایی نرسی مزد اگر می‌طلبی، طاعت استاد بپر

(حضرت حافظ)

از استاد ارجمند جناب آقای دکتر مسعود البرز، مشاور محترم پایان‌نامه کمال تشکر را دارم. همچنین از استاد فرزانه جناب آقای دکتر محمد قاسم وحیدی اصل و استاد گرانقدر جناب آقای دکتر محمدرضا فقیهی که به من افتخار داوری این پایان‌نامه را فرمودند بسیار متشکرم. از تمامی اساتید گروه آمار دانشکده‌ی علوم ریاضی که در این دوره‌ی تحصیلی، فرصت بهره‌مندی از محضرشان را یافتم صمیمانه تشکر می‌نمایم. همچنین از جناب آقای دکتر یدالله محرایی استاد محترم دانشکده‌ی بهداشت، که در بررسی وجود داده‌های پزشکی مناسب، بویژه در پژوهشکده‌ی علوم غدد درون‌ریز و متابولیسم، به بنده مشاوره دادند سپاسگزارم.

مرضیه‌ریعی

Raeqati.2007@yahoo.com

زمستان ۸۸

چکیده

در مطالعات مشاهده‌ای، تخصیص تیمار به واحدهای مورد مطالعه، تحت کنترل پژوهشگر نیست. از این رو گروه‌های تیماری و غیر تیماری (شاهد) ممکن است از نظر متغیرهای کمکی مشاهده شده تفاوت زیادی داشته باشند، و این تفاوت‌ها منجر به برآوردهای اریب از اثر تیمار شود. در این پایان نامه به بیان روش‌های گوناگون برای برآورد اثر تیمار در مطالعات مشاهده‌ای با هدف کاهش اریبی و با رهیافت ایجاد تعادل بر حسب متغیرهای کمکی مشاهده شده در گروه تیمار و شاهد، می‌پردازیم. بررسی اثر تیمار و یا استنباط علی با دو رویکرد آینده‌نگر - گذشته‌نگر و مقطعی - طولی انجام شده است.

در نهایت برخی روش‌های برآورد اثر تیمار را با معرفی بسته‌های نرم‌افزاری موجود در R روی یک داده‌ی واقعی بکار می‌بریم.

واژه‌های کلیدی : مطالعات مشاهده‌ای، اثر علی، نمره‌ی تمایل، جورسازی، زیررده‌بندی، تعدیل رگرسیونی، اریبی انتخاب، برآوردگر تفاضل در تفاضل.

پیشگفتار

روش استاندارد طلایی برای برآورد اثر علی تیمار، آزمایش‌های کنترل شده‌ی تصادفیده است که در آن‌ها تیمار مورد نظر به صورت تصادفی با نظارت و تحت کنترل پژوهشگر، به واحدهای آماری تخصیص می‌یابد. اما در اغلب زمینه‌های علمی که هدف مطالعه جوامع انسانی است، اعم از علوم اقتصادی، اجتماعی و پزشکی، بدلیل محدودیت‌های اخلاقی، قانونی و سیاسی، امکان تخصیص تصادفی تیمار به واحدهای مورد مطالعه وجود ندارد که در نتیجه این مطالعات را در زمره‌ی مطالعات مشاهده‌ای قرار می‌دهند. در این گونه مطالعات، مقایسه‌ی مستقیم پاسخ گروه‌های تیماری برای برآورد اثر تیمار، در معرض آریبی قابل ملاحظه‌ای قرار دارد که می‌تواند نتایج گمراه‌کننده یا فاقد اعتبار را به همراه داشته باشد. روش‌هایی وجود دارد که با رهیافت ایجاد هم توزیعی برحسب متغیرهای کمکی در گروه‌های تیماری سعی در کاهش آریبی و برآورد صحیحی از اثر تیمار دارند.

حتی در مواردی که امکان اجرای یک آزمایش کنترل شده‌ی تصادفیده وجود داشته باشد طراحی و تحلیل صحیح یک مطالعه‌ی مشاهده‌ای معادل آن، می‌تواند از نظر هزینه و زمان اجرای یک آزمایش به صرفه باشد و نتایج نسبتاً مطلوبی در برداشته باشد.

برآوردگرهای اثر علی را می‌توان از حیث به کارگیری متفاوت اطلاعات متغیر پاسخ به دو دسته تقسیم کرد: الف) برآوردگرهای مقطعی: این برآوردگرها، اثر تیمار را با استفاده از مقادیر متغیر پاسخ در یک نقطه یا یک مقطع خاصی از زمان محاسبه می‌کنند که در فصل دوم پایان‌نامه به آن پرداخته خواهد شد. ب) برآوردگرهای طولی: این دسته از برآوردگرها اثر تیمار را به کمک صورت پانلی متغیر پاسخ برآورد می‌کنند. یکی از مهم‌ترین برآوردگرهای طولی در فصل سوم معرفی خواهد شد.

همچنین در این پایان نامه به استنباط علی، از دو دیدگاه آینده‌نگر و گذشته‌نگر پرداخته شده است که با این رویکرد فصل‌های پایان‌نامه به دو دسته تقسیم می‌شوند. فصل‌های اول، دوم، چهارم و پنجم رویکرد آینده‌نگر و فصل سوم رویکرد گذشته‌نگر دارد.

در فصل اول، تعریف سه مفهوم اساسی در این پایان‌نامه یعنی اثر علی، تیمار و مطالعات مشاهده‌ای بیان می‌شود. یک مدل برای استنباط علی، مسئله‌ی اساسی آن و راه‌های استنباط علی مطرح می‌شود. سپس انواع مطالعات مشاهده‌ای و همچنین پارامترهای مورد توجه در این نوع مطالعات، معرفی می‌شوند.

فصل دوم دربردارنده‌ی روش‌های گوناگون برای کاهش اربیی در این نوع مطالعات است. این فصل را می‌توان در دو بخش اصلی یک متغیره و چندمتغیره در نظر گرفت که در حالت چندمتغیره از کمیت مهمی به نام نمره‌ی تمایل استفاده خواهد شد.

در فصل سوم با دیدگاه گذشته‌نگر، به استنباط علی پرداخته می‌شود.

در فصل چهارم، برای حالتی که داده‌های طولی در دست باشد، برآوردگری موسوم به تفاضل در تفاضل متناسب به حکمن و همچنین فرض مورد نیاز برای این برآوردگر معرفی می‌شود. اربیی ارزیابی به مؤلفه‌هایی، تجزیه می‌شود و نتیجه‌ی برآورد آن‌ها با استفاده از داده‌های *JTPA* بیان می‌شود. در انتها یک صورت عمومی برای برآوردگرهای جورسازی ارائه می‌شود که در همین بخش به چند روش جورسازی دیگر نیز اشاره می‌شود.

در فصل پنجم به معرفی بسته‌ی نرم افزاری، نحوه‌ی استفاده از آن در مطالعات مشاهده‌ای و محاسبه‌ی اثر تیمار پرداخته شده و روی یک داده‌ی واقعی اعمال می‌شود.

فهرست مندرجات

۱	۱ اثر علی
۱-۱-۱-۱	علیت.....
۱-۱-۱-۲	مدل استنباط پیوند.....
۲-۱-۱-۱	مدل علی روبین (RCM).....
۱-۲-۱-۱-۱	فرض مقدار واحد- تیمار پایدار (SUTVA).....
۳-۱-۱-۱	مسئله‌ی اساسی استنباط علی.....
۴-۱-۱-۱	راه حل آماری.....
۱-۴-۱-۱-۱	اهمیت و نقش تصادفیدن.....
۲-۱-۱-۱	راه‌های استنباط علی.....
۱-۲-۱-۱	آزمایش‌های کنترل شده‌ی تصادفیده (RTC).....
۲-۲-۱-۱	مطالعات مشاهده‌ای.....
۱-۲-۲-۱-۱	مسئله‌ی آشکار مطالعات مشاهده‌ای.....
۲-۲-۲-۱-۱	دلایل اهمیت مطالعات مشاهده‌ای.....
۳-۲-۲-۱-۱	فرض تخصیص تیمار به طور قوی قابل اغماض (SITA).....
۳-۲-۱-۱	آزمایش‌های طبیعی.....
۳-۱-۱-۱	انواع مطالعات مشاهده‌ای.....
۱-۳-۱-۱	مطالعات آینده‌نگر.....

- ۱۵ مطالعات گذشته‌نگر ۲-۳-۱
- ۱۶ مطالعات آینده‌نگر تاریخی ۳-۳-۱
- ۱۷ پارامترهای دیگر اثر علی ۴-۱
- ۱۷ متوسط اثر تیمار روی تیمار شده‌ها ۱-۴-۱
- ۱۹ نسبت بخت‌ها ۲-۴-۱

۲۰ روش‌های تعدیل اثر علی با داده‌های مقطعی

- ۲۱ روش‌های یک متغیره ۱-۲
- ۲۱ زیررده‌بندی ۱-۱-۲
- ۳۰ تعدیل رگرسیونی ۲-۱-۲
- ۳۱ جورسازی ۳-۱-۲
- ۳۴ جورسازی پرگاری ۱-۳-۱-۲
- ۳۵ جورسازی (نزدیکترین در دسترس) ۲-۳-۱-۲
- ۳۷ جورسازی متوسط ۳-۳-۱-۲
- ۳۹ روش‌های چند متغیره ۲-۲
- ۳۹ رگرسیون خطی موازی ۱-۲-۲
- ۴۲ رگرسیون خطی ناموازی ۲-۲-۲
- ۴۴ درصد کاهش آریبی مساوی (EPBR) ۳-۲-۲
- ۴۵ جورسازی با متر ماها لانوبیس ۴-۲-۲
- ۴۷ جورسازی تشخیصی ۵-۲-۲
- ۴۹ روش‌های مبتنی بر نمره‌ی تمایل ۳-۲

- ۱-۳-۲ نمره‌ی متعادل سازی و نمره‌ی تمایل ۵۰
- ۲-۳-۲ برآورد نمره‌ی تمایل ۶۰
- ۳-۳-۲ کاهش اریبی در روش‌های مبتنی بر نمره‌ی تمایل ۶۲
- ۴-۳-۲ کارایی زیررده‌بندی روی نمره‌ی تمایل در حذف اریبی ۶۵

۳ استنباط علی در مطالعات گذشته‌نگر ۷۰

- ۱-۳ استنباط علی در مطالعات آینده‌نگر ۷۱
- ۱-۱-۳ سطوح استنباط علی ۷۱
- ۲-۳ استنباط علی در مطالعات گذشته‌نگر ۷۳
- ۱-۲-۳ جدول دوطرفه‌ی استاندارد ۷۳
- ۲-۲-۳ رابطه‌ی احتمال‌های آینده‌نگر و گذشته‌نگر ۷۴
- ۳-۲-۳ چرا جدول دو طرفه‌ی استاندارد گمراه کننده است؟ ۷۶
- ۴-۲-۳ نقش متغیرهای کمکی در مطالعات گذشته‌نگر ۷۷
- ۳-۳ نکاتی چند پیرامون استنباط علی گذشته‌نگر ۸۰

۴ اثر علی در داده‌های طولی ۸۱

- ۱-۴ مسئله‌ی ارزیابی ۸۲
- ۲-۴ جورسازی به عنوان جایگزینی برای آزمایش‌ها ۸۵
- ۳-۴ بسط جورسازی و برآوردگر تفاضل در تفاضل هکمن ۸۷
- ۴-۴ داده‌های *JTPA* ۹۰

- ۹۳ تجزیه‌ی معیار اریبی ارزیابی ۵-۴
- ۹۷ بررسی فرض‌های تأییدکننده‌ی جورسازی و برآوردگر تفاضل در تفاضل ۶-۴
- ۹۸ ۱-۶-۴ آماره‌های آزمون ۶-۴
- ۱۰۳ فرم عمومی برآوردگرهای جورسازی ۷-۴

۱۰۷ ۵ برنامه‌های کاربردی در یک مثال واقعی

- ۱۰۷ داده‌های *NSW* ۱-۵
- ۱۱۰ جورسازی ژنتیک ۲-۵
- ۱۱۱ *Matching* و توابع آن ۳-۵
- ۱۱۱ تابع *Match* ۱-۳-۵
- ۱۱۲ تابع *GenMatch* ۲-۳-۵
- ۱۱۳ تابع *MatchBalance* ۳-۳-۵
- ۱۱۵ *MatchIt* بسته‌ی ۳-۵
- ۱۱۶ اعمال برنامه‌ها روی داده‌های *NSW* و تحلیل نتایج ۴-۵

۱۲۴ A رگرسیون ناپارامتری ۱۲۴

۱۳۰ B برنامه‌های محاسباتی مورد استفاده ۱۳۰

۱۳۷ C واژه‌نامه‌ی فارسی به انگلیسی ۱۳۷

۱۴۱..... D نام نامه

۱۴۳..... کتاب نامه

فصل اول

اثر علی

« یکی از اهداف اصلی آمار، همیشه تعیین روابط علی از داده‌های آزمایشی و ناآزمایشی بوده است.»

گلیمور (۱۹۸۶)

۱-۱ علیّت

پیشینه‌ی موضوع علیّت به فلسفه برمی‌گردد. فلاسفه این بحث را حول جستجوی علت یک پدیده یا پدیده دنبال می‌کرده‌اند. به طوری که کاکس (۱۹۹۲) مشهورترین تعریف فلسفی علیّت را به این صورت بیان می‌کند: اگر C علت اثر E باشد، آنگاه اگر E مشاهده شود C باید رخ داده باشد. اما آمارشناسان عملاً زمانی وارد عرصه علیّت شدند که مسئله‌ی اثر یک علت و اندازه‌گیری آن مطرح و مورد توجه قرار گرفت. این رویکرد تحولی در تحلیل سنتی علیّت محسوب می‌شود. در این رویکرد تأکید بر «اثر یک علت» است و نه «علت یک اثر» (هلند ۱۹۸۶).

قبل از معرفی مدلی برای علیّت، مدل استنباط پیوند را مرور کرده و سپس تفاوت این دو مدل را مورد

بررسی قرار می‌دهیم.

۱-۱-۱ مدل استنباط پیوند

مدل مناسب برای استنباط پیوند، مدل آماری استاندارد است که دو متغیر را در کل جامعه به هم مربوط می‌سازد. اینک به معرفی اجمالی این مدل می‌پردازیم.

فرض کنید U معرف جامعه و u نشانگر واحد آماری در این جامعه باشد. متغیرهای تعریف شده روی این واحدها، عناصر اصلی مدل‌های ارائه شده برای پیوند و علیت هستند. از آنجا که نقش زمان نیز باید مورد توجه قرار گیرد، جامعه‌ی واحدها داخل یک چارچوب زمانی مشخص تعریف می‌شود. البته نقش زمان در استنباط علی اهمیت بیشتری در مقایسه با استنباط پیوند دارد.

در نظر بگیرید $Y(u)$ و $A(u)$ به ترتیب متغیرهای پاسخ و تبیینی واحد u ام جامعه باشند. پارامترهای پیوندی توسط توزیع توأم و شرطی تعیین می‌شوند. یک نوع پارامتر پیوندی، رگرسیون Y روی A ، یعنی امید شرطی $E(Y | A = a)$ است. استنباط پیوند شامل استنباط‌های آماری (برآورد، آزمون و ...) پارامترهای پیوندی مربوط به Y و A بر اساس داده‌های جمع‌آوری شده در مورد Y و A از واحدهای جامعه است.

۲-۱-۱ مدل علی روبین (RCM)

هدف، طراحی مدلی است که امکان فرمول‌بندی شواهد اولیه در خصوص علت و اثر را فراهم کند. این مدل با جامعه‌ای از واحدها آغاز می‌شود. علت‌ها یا به طور معادل تیمارهای مورد نظر، روی واحدهای جامعه عمل می‌کنند. واقعیت این است که اثر یک علت همیشه مرتبط با علت دیگری است. برای مثال عبارت « A سبب B است» تقریباً همیشه به این معنی است که A سبب B است در مقایسه با سایر علت‌ها (یعنی هر علتی به جز A). و اگر برای سهولت از زبان آزمایش‌ها استفاده کنیم، به معنی تفاوت تیمار (یک علت) و شاهد (علت دیگری) است.

تعریف تیمار: به مداخلاتی اطلاق می‌شود که هر یک از واحدهای جامعه‌ی تحت مطالعه، به طور بالقوه بتوانند در معرض آن قرار گیرند یا در معرض آن قرار نگیرند (روزنباوم ۱۹۸۴).

¹ Rubin Causal Model

بنابراین بایستی امکان در معرض قرارگرفتن یا قرارنگرفتن هر واحد تحت علت یا تیمار (صرف‌نظر از اینکه آیا عملاً چنین شده است) وجود داشته باشد. برای مثال، نوع آموزشی که یک دانش‌آموز دریافت می‌کند می‌تواند علت عملکرد او در یک آزمون باشد درحالی‌که نژاد یا جنسیت نمی‌تواند به عنوان یک علت قلمداد شود. پس یک صفت کیفی نمی‌تواند علت باشد زیرا حتی به طور بالقوه نمی‌تواند تغییر یابد. تنها راهی که امکان دارد یک صفت کیفی مقدارش تغییر کند آن است که واحد آماری تغییر ماهیت یابد که در این صورت دیگر همان واحد نیست (هلند ۱۹۸۶).

از این رو گزاره‌های علی که شامل صفت کیفی به عنوان علت هستند، صرفاً در قالب پیوند (همبستگی) بین صفت کیفی و متغیر پاسخ روی واحدهای جامعه بیان می‌شوند و در حوزه‌ی استنباط علی قرار نمی‌گیرند. همانطور که در ابتدای فصل اشاره شد، تحلیل سستی علیت با جستجو درباره‌ی این سؤال آغاز می‌شود که چه علتی موجب اثر مشاهده شده است. به وضوح این یک تلاش علمی ارزشمند است اما رویکرد صحیحی در تحلیل نظری علیت نیست. به علاوه علت یک اثر همیشه با افزایش دانش ما درباره‌ی آن پدیده می‌تواند تغییر یافته و در آن تجدیدنظر صورت گیرد. در مقابل، اندازه‌گیری اثر یک علت اگرچه در مواردی می‌تواند مشکل باشد اما حداقل به دقت قابل تعریف است (هلند ۱۹۸۶).

فرض کنید دو علت یا یک تیمار دو سطحی وجود دارد که با تیمار و شاهد بیان می‌شوند. به علاوه فرض کنید D متغیر نشانگری است که نشان می‌دهد هر واحد جامعه در معرض چه علتی قرارگرفته است. یعنی $D=1$ معرف آن است که واحد تحت تیمار است و $D=0$ نشان می‌دهد که واحد در معرض شاهد است. در یک مطالعه‌ی کنترل شده، D توسط آزمایشگر تعیین می‌شود. در یک مطالعه‌ی کنترل نشده، D تا اندازه‌ای توسط عواملی خارج از کنترل آزمایشگر تعیین می‌شود. در هر دو مورد ویژگی اساسی مفهوم علیت این است که مقدار $D(u)$ برای هر واحد می‌تواند متفاوت باشد (مقدار یکتا نیست، بلکه به طور بالقوه قابل تغییر است). متغیر D مشابه متغیر A در بخش ۱-۱-۱ است اما تفاوت اساسی آن در این است که $D(u)$ نشان می‌دهد u در معرض علت خاصی است در حالی‌که $A(u)$ علاوه بر آن می‌تواند یک ویژگی یا مشخصه‌ی واحد u باشد که در این صورت نمی‌تواند تغییر کند.

اینک به بررسی نقش زمان در استنباط علی می‌پردازیم. اهمیت بررسی نقش زمان از آن رو است که وقتی یک واحد در معرض تیمار یا علی قرار می‌گیرد طبیعتاً این واقعه در یک دوره‌ی زمانی خاصی رخ داده است. از این نقطه نظر متغیرهای تحت بررسی را می‌توان به دو گروه تقسیم بندی کرد: متغیرهای پیش‌معرضی، که مقادیرشان قبل از در معرض علی قرار گرفتن واحد، تعیین می‌شوند و متغیرهای پس‌معرضی که مقادیر آنها پس از در معرض علی قرار گرفتن واحد، اندازه‌گیری می‌شوند.

بنابراین از آنجا که نقش متغیر پاسخ اندازه‌گیری اثر علی است، پس در رده‌ی متغیرهای پس‌معرضی قرار می‌گیرد. رکن اساسی دیگر مدل روبین این است که مقادیر متغیرهای پس‌معرضی به طور بالقوه تحت تأثیر علی خاص ۱ یا ۰ هستند. برای اینکه مدل درست و کاملی داشته باشیم فقط به یک متغیر Y نیاز نداریم بلکه با دو متغیر Y_0 و Y_1 که نماینده‌ی دو پاسخ‌اند مواجه هستیم. $Y_1(u)$ مقدار پاسخ واحد u هنگامی که در معرض تیمار قرار گیرد و $Y_0(u)$ مقدار پاسخ واحد u وقتی تحت تیمار نباشد.

اثر علی تیمار روی واحد u (یا اثر تیمار) به صورت زیر تعریف می‌شود:

$$Y_1(u) - Y_0(u) \quad (1-1)$$

۱-۱-۲-۱ فرض مقدار واحد-تیمار پایدار (SUTVA)^۱

ساختار معرفی شده برای اثر تیمار، ممکن است برای هر وضعیتی کفایت نکند مثلاً ممکن است پاسخ واحد u به تیمار، وابسته به این باشد که واحد u' چه تیماری دریافت کرده است. چنین حالتی معمولاً وقتی رخ می‌دهد که آنها برای رسیدن به منابع در حال رقابت باشند. با این توضیح، مدل روبین تحت فرضی به نام «فرض مقدار واحد-تیمار پایدار»، برای داشتن جواب‌های علی، کفایت می‌کند. SUTVA پیش فرضی است که مقدار پاسخ Y برای واحد u وقتی در معرض تیماری قرار می‌گیرد، یکسان خواهد بود صرف‌نظر از اینکه چه مکانیسمی در تخصیص آن تیمار به واحد u به کار رفته و صرف‌نظر از اینکه واحدهای دیگر چه تیماری دریافت کرده‌اند. و این ویژگی برای همه‌ی واحدها و تیمارها برقرار است (روبین ۱۹۸۶).

^۱ Stable Unit-Treatment Value Assumption

برای بحث مربوط به تخطی‌های ممکن از این فرض به کاکس (۱۹۵۸، فصل دوم) و روبین (۱۹۷۸، بخش ۲.۳) مراجعه کنید.

۱-۱-۳ مسئله‌ی اساسی استنباط علی

بدیهی است که برای هر واحد u در آن واحد تنها یکی از دو مقدار $Y_1(u)$ یا $Y_0(u)$ مشاهده می‌شود و مشاهده‌ی همزمان آنها غیر ممکن است بنابراین مشاهده‌ی اثر تیمار روی u غیر ممکن است. از این رو، برآورد اثر علی، یک مسئله‌ی داده‌های گم شده است زیرا یکی از مقادیر $Y_1(u)$ و $Y_0(u)$ گم شده است.

مسئله‌ی اساسی استنباط علی، در حوزه‌ی ارزیابی‌های اقتصادی، به عنوان مسئله‌ی اساسی ارزیابی شناخته شده و با اصطلاح «برآمد بالقوه»^{*} یاد می‌شود. برای درک مسئله فرض کنید بخواهیم اثر یک برنامه کارآموزی اختیاری را روی شانس پیدا کردن کار، اندازه‌گیری کنیم. در سطح افراد، برآمد بازار کار آنها که دوره‌ی آموزشی خاصی را گذرانیده‌اند و آنها که چنین دوره‌ای را نگذرانیده‌اند را مشاهده می‌کنیم. برای اینکه بدستی اثر آموزش را روی افراد شرکت‌کننده بدانیم باید برآمد مشاهده‌شده‌ی آنها را با برآمد بالقوه آنها در صورتی که در برنامه شرکت نمی‌کردند، مقایسه کنیم. با این وجود فقط یک برآمد، واقعاً مشاهده شده است، که به آن برآمد واقعی می‌گوییم. «برآمد بالقوه» برآمدی است که افراد شرکت‌کننده در صورتی که در برنامه شرکت نمی‌کردند می‌توانستند داشته باشند (یا برآمدی که افراد شرکت نکرده در صورت شرکت در برنامه می‌داشتند). این برآمد قابل مشاهده نیست و به عنوان مسئله‌ی ارزیابی مورد توجه است. همه‌ی رهیافت‌های ارزیابی تلاش می‌کنند تا برآوردی برای «برآمد بالقوه» به دست آورند و از آن برای تشخیص اثر برنامه استفاده کنند (بریسون و همکاران ۲۰۰۲).

غیر ممکن بودن مشاهده‌ی همزمان $Y_1(u)$ و $Y_0(u)$ در بسیاری از کاربردها بدیهی به نظر می‌رسد اما این به معنی غیر ممکن بودن استنباط علی یا عدم وجود دانش و اطلاع مرتبط با این مقادیر نیست، بلکه به

* این اصطلاح در نوشتگان منطق و فلسفه، «خلاف واقع» و «فرضی» ترجمه شده است اما در حوزه‌ی ارزیابی، «برآمد بالقوه» ترجمه‌ی مناسب‌تری به نظر می‌رسد.

موقعیت مورد بررسی بستگی دارد.

اگرچه مدل شامل سه متغیر D ، Y_1 و Y_0 است، فرایند مشاهده تنها شامل دو متغیر D و Y_D است. به عبارت دیگر پاسخ مشاهده شده را بر اساس سه متغیر D ، Y_1 و Y_0 می‌توان به صورت زیر نمایش داد:

$$Y = DY_1 + (1-D)Y_0$$

در تحلیل علیت، تمایز بین سه مقوله‌ی زیر بسیار مهم است اما در مطالعه‌ی پیوند مطرح نیست. الف) مکانیسم اندازه‌گیری Y ، که متغیر پاسخ را ایجاد می‌کند. ب) دو حالت متغیر پاسخ، Y_1 و Y_0 که مرتبط با علیت است که واحد در معرض آن قرار گرفته است. ج) متغیر پاسخ مشاهده شده Y_D . برای مثال در نظر بگیرید بخواهیم اثر آموزش را روی درآمد به دست آوریم: Y متغیر درآمد است، Y_1 و Y_0 به ترتیب درآمد تحت آموزش دیدن و آموزش ندیدن است و Y_D بیانگر درآمد مشاهده شده است.

برای تأکید روی تفاوت‌های مدل استنباط پیوند و مدل روبین آنها را مرور می‌کنیم: هر دو شامل جامعه واحدها و دو متغیر قابل مشاهده هستند، (A, Y) برای پیوند و (D, Y_D) برای علیت. استنباط پیوندی شامل توزیع توأم یا شرطی مقادیر A و Y است و استنباط علی روی مقادیر $Y_1(u) - Y_0(u)$ برای واحدهای جامعه متمرکز است.

۱-۱-۴ راه حل آماری

این مسئله با انتقال از سطح فردی به سطح جامعه آغاز می‌شود.

متوسط اثر علی تیمار (نسبت به شاهد) در کل جامعه، برابر مقدار امیدریاضی تفاضل $Y_1(u) - Y_0(u)$

تعریف می‌شود، یعنی:

$$\begin{aligned} \tau &= E(Y_1 - Y_0) \\ &= E(Y_1) - E(Y_0) \end{aligned} \quad (2-1)$$

حال مسئله به برآورد $E(Y_1)$ و $E(Y_0)$ تبدیل می‌شود.

پارامتر زیر را در نظر بگیرید:

$$\begin{aligned} \tau_{pf} &= E(Y_D | D=1) - E(Y_D | D=0) \\ &= E(Y_1 | D=1) - E(Y_0 | D=0) \end{aligned} \quad (3-1)$$

این پارامتر موسوم به اثر علی ظاهری است. نکته‌ی مهم توجه به تمایز $E(Y_1)$ و $E(Y_1|D=1)$ است. $E(Y_1)$ ، امیدریاضی پاسخ در کل جامعه است وقتی همه واحدهای جامعه تحت تیمار قرار بگیرند در حالی که $E(Y_1|D=1)$ امید ریاضی پاسخ به تیمار در زیر جامعه‌ای است که تحت تیمار قرار گرفته است. این دو در حالت کلی، لزوماً مقادیر یکسانی ندارند (به طور مشابه می‌توان به تمایز $E(Y_0)$ و $E(Y_0|D=0)$ اشاره کرد).

۱-۴-۱-۱ اهمیت و نقش تصادفیدن

اگر تخصیص تیمار به واحدها، با یک مکانیسم تصادفیده انجام شود، در این صورت تخصیص علت به یک واحد مستقل از هر ویژگی مرتبط با آن واحد، از جمله مقادیر Y_1 و Y_0 است، یعنی $(Y_1, Y_0) \perp D$ که در اینصورت $\tau = \tau_{pf}$ ، یعنی:

$$(Y_1, Y_0) \perp D \Rightarrow \tau = \tau_{pf}$$

که در آن نماد \perp معرف استقلال است. به این ترتیب برآورد متوسط اثر علی تیمار با داشتن نمونه‌ای از زیرجامعه‌های تیماری و شاهد، امکان‌پذیر خواهد بود.

۲-۱ راه‌های استنباط علی

استنباط علی به سه طریق صورت می‌گیرد (فریدمن ۲۰۰۵):

- آزمایش‌های کنترل شده تصادفیده
- مطالعات مشاهده‌ای
- آزمایش‌های طبیعی

که به تفصیل در ادامه بیان می‌شود.