باسمه تعالی

مجموعه مقالات

اولین سمینار

استنباط شواهدی

گروه آمار دانشگاه فردوسی مشهد

۲۲ خرداد ۱۳۹۸

مقدمه

واژه‌های "استنباط شواهدی" و "شواهد آماری" که اخیراً در بعضی از متون روش‌شناسی آمار به چشم می‌خورد، اشاره به مکتب نوینی از استنباط آماری دارد که در آن استنباط صرفاً متکی به داده‌ها و الگوی احتمالی است و از مولفه‌های ذهنی، شخصی و موردی نظیر باورهای پیشین و توابع زیان تأثیر نمی‌پذیرد. جلودار این حرکت جدید پروفسور ریچارد رویال استاد (بازنشسته) دانشگاه جان‌هاپکینز آمریکا است که کتاب وی تحت عنوان "شواهد آماری" اصول اولیه این شیوه جدید استنباط را در بردارد.

محور اصلی این روش، تابع درستنمایی است، ولی علاوه بر اتکا به اصل درستنمایی (مانند بسیاری از روش‌های کلاسیک) این شیوه استنباط به شدت بر اصل مشابه (ولی متفاوت) دیگری به نام "قانون درستنمایی" متکی است. آزمون‌ها و بازه‌های اطمینان (اگر استفاده از این نام‌ها در شیوه جدید مناسب باشد) به گونه‌ای دیگر تعریف و اجرا می‌شوند.

گرچه تاکنون پژوهش‌های زیادی درباره‌ی این مکتب جدید انجام نشده است ولی به نظر می‌رسد که این شیوه تحلیل داده‌های آماری در آینده نزدیک جایی در کنار مکاتب تحلیل بیزی و نظریه نیمن‌ـپیرسون خواهد داشت.

با آرزوی توفیق

مهدی عمادی (دبیر)

خرداد ۱۳۹۸

محورهای سمینار

۱. نقش معیارهای اطلاع در اندازه‌گیری میزان شواهد
۲. وجوه تفاوت و تشابه مکتب بیز و مکتب شواهدی
۳. معیارهای پشتیبانی داده‌ها از فرضیه‌های آماری
۴. مشکلات و معایب مربوط به p-value
۵. اصل درستنمایی و قانون درستنمایی
۶. اشتباهات رایج در استنباط آماری
۷. اندازه‌گیری قوت شواهد آماری
۸. قوانین استنباط شواهدی
۹. الگوهای فکری در آمار

اعضای کمیته علمی (به ترتیب حروف الفبا):

۱. دکتر محمد آرشی، دانشگاه صنعتی شاهرود (نماینده انجمن آمار ایران)
۲. دکتر احد جمالی‌زاده، دانشگاه شهید باهنر کرمان
۳. دکتر رحیم چینی‌پرداز، دانشگاه شهید چمران اهواز
۴. دکتر آرزو حبیبی‌راد، دانشگاه فردوسی مشهد
۵. دکتر علی دست‌برآورده، دانشگاه یزد
۶. دکتر مهدی دوست‌پرست، دانشگاه فردوسی مشهد
۷. دکتر عباس رسولی، دانشگاه زنجان
۸. دکتر احسان زمان‌زاده، دانشگاه اصفهان
۹. دکتر سید محمود طاهری، دانشگاه تهران (دبیر کمیته علمی)
۱۰. دکتر مهدی عاشوری، موسسه پژوهشی حکمت و فلسفه
۱۱. دکتر مهدی عمادی، دانشگاه فردوسی مشهد (دبیر سمینار)
۱۲. دکتر محمدرضا مشکانی، دانشگاه شهید بهشتی تهران
۱۳. دکتر فاطمه یوسف‌زاده، دانشگاه بیرجند

اعضای کمیته برگزار کننده (به ترتیب حروف الفبا):

۱. دکتر حسینعلی آذرنوش، دانشگاه فردوسی مشهد (بازنشسته)
۲. دکتر جعفر احمدی، دانشگاه فردوسی مشهد
۳. دکتر محمد امینی، دانشگاه فردوسی مشهد
۴. دکتر سیمیندخت براتپور، دانشگاه فردوسی مشهد (بازنشسته)
۵. دکتر ابوالقاسم بزرگ‌نیا، دانشگاه فردوسی مشهد (بازنشسته)
۶. دکتر احمدرضا بهرامی، دانشگاه فردوسی مشهد (معاون پژوهش و فناوری دانشگاه)
۷. دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد
۸. دکتر مهدی جباری نوقابی، دانشگاه فردوسی مشهد
۹. دکتر آرزو حبیبی‌راد، دانشگاه فردوسی مشهد
۱۰. دکتر کاظم خشیارمنش، دانشگاه فردوسی مشهد (معاون پژوهشی دانشکده)
۱۱. دکتر مهدی دوست پرست، دانشگاه فردوسی مشهد
۱۲. دکتر مصطفی رزمخواه، دانشگاه فردوسی مشهد (معاون آموزشی دانشکده)
۱۳. دکتر عبدالحمید رضایی رکن آبادی، دانشگاه فردوسی مشهد
۱۴. دکتر علیرضا سهیلی، دانشگاه فردوسی مشهد (رئیس دانشکده)
۱۵. دکتر بهرام صادقپور گیلده، دانشگاه فردوسی مشهد (مدیر گروه آمار)
۱۶. دکتر حسن صادقی، دانشگاه فردوسی مشهد (بازنشسته)
۱۷. دکتر محمد صال مصلحیان، دانشگاه فردوسی مشهد (مدیر پژوهشی دانشگاه)
۱۸. دکتر محمد مهدی طباطبایی، دانشگاه فردوسی مشهد (بازنشسته)
۱۹. دکتر مهدی عمادی، دانشگاه فردوسی مشهد (دبیر سمینار)
۲۰. دکتر معصومه فشندی، دانشگاه فردوسی مشهد
۲۱. دکتر وحید فکور، دانشگاه فردوسی مشهد
۲۲. دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد
۲۳. دکتر حسینعلی نیرومند، دانشگاه فردوسی مشهد (بازنشسته)

کادر اجرایی سمینار (به ترتیب حروف الفبا):

۱. عادل احمدی نادی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۲. امیرحسین اسکندانی، دانشجوی کارشناسی ریاضی، دانشگاه فردوسی مشهد
۳. مجتبی اصفهانی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۴. محمد برات‌نیا، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۵. فاطمه حوتی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۶. مریم خوبی، کارشناس عکاسی، دانشگاه تهران
۷. سید حسین رسولی، دانشجوی کارشناسی ریاضی، دانشگاه فردوسی مشهد
۸. محمد جواد سبک‌خیز، دانشجوی کارشناسی ارشد آمار، دانشگاه فردوسی مشهد
۹. طاهره عالمی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۱۰. جابر کاظم‌پور، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۱۱. مرتضی محمدی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۱۲. زهرا محمدیان، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد

حامیان سمینار:

۱. دانشگاه فردوسی مشهد
۲. دانشکده علوم ریاضی دانشگاه فردوسی مشهد
۳. گروه آمار دانشگاه فردوسی مشهد
۴. انجمن‌های علمی دانشجویی دانشکده علوم ریاضی دانشگاه فردوسی مشهد
۵. انجمن آمار ایران
۶. قطب علمی داده های ترتیبی و فضایی
۷. پایگاه استنادی علوم جهان اسلام
۸. مرکز اطلاع‌رسانی علوم و فناوری

فهرست مقالات فارسی

جایگاه استنباط شواهدی در فلسفهٔ آمار

طاهری، س.م. ۱۰

نیاز توجه به مبانی آمار در آمار کاربردی

عاشوری، م. ۱۷

نگاه کردن به نمودار های توابع درستنمایی

عمادی، م. ۲۴

برآورد معیار وابستگی شواهدی با استفاده از تابع چگالی مفصل

محمدی، م.، عمادی، م.، امینی، م. ۳۰

انعکاس‌پذیری توان‌هایی از عملگر ضربی روی فضاهای توابع خاص

باقری نژاد، س. ح.، ایلون کشکولی، ع. ۳۹

مروری بر مفهوم فرض و فرضیه

شرفی، م.، قهرمانی، م. ۴۵

کنکاشی در اندازه‌های اطلاع از دیدگاه بیزی

رحمانپور، آ. ۵۰

مدل انتقال واریانس در مدل رگرسیونی تحت محدودیت خطی تصادفی

رحمی، ف.، بابادی، ب.، راسخ، ع. ۵۸

برآورد لیو جک نایف تحت محدودیتهای خطی تصادفی در مدلهای رگرسیونی

طلادزفولی، م.، راسخ، ع.، بابادی، ب. ۶۴

مقایسه توان آزمون‌های آنالیز واریانس، کروسکال-والیس و ولش در داده‌های غیر نرمال

طالبی، س.، جام بر سنگ، س. ۷۲

انتخاب توزیع پیشین در استنباط بیزی
اروجی، آ.، اکبری، ن.، فکور، و. ۸۱

خطاهای آماری مقالات منتشر شده در مجلات دندان پزشکی سال ۱۳۹۷
حیدری بهنوئیه ، ا.، جام بر سنگ ، س.، احمدی نیا گدنه ، ح. ۹۰

اشتباهات رایج آماری
جباری نوقابی، م.، تلخی، ن. ۹۶

تفسیر نتایج آزمونهای کلاسیک آماری
زمان زاده، ا.، دست برآورده، ع.، ارقامی، ن.ر. ۱۱۱

برخی راه حل ها برای مشکل p ـ مقدارها
اتله خانی، م. ۱۱۷

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

جایگاه استنباط شواهدی در فلسفهٔ آمار

طاهری، س.م [۱]

[۱]دانشکده فنی، دانشگاه تهران.

چکیده

فلسفهٔ آمار به مباحث بنیادین مرتبط با علم آمار می‌پردازد، از جمله: مبانی منطقی علم آمار، اصول استنباط آماری، هدف یا اهداف آمار، روش‌شناسی علم آمار، چیستیِ مدل‌های آماری، و الگوهای فکری در استنباط آماری. یکی از الگوهای فکری آمار، که چند دهه است بیشتر به آن توجه شده، الگوی فکری شواهدی است. در این مقاله، نخست بیان می‌شود که علم چیست؟ فلسفه چیست؟ و فلسفهٔ علم چیست؟ آنگاه توضیحاتی دربارهٔ فلسفهٔ آمار و مباحث اصلی مربوط ارائه می‌شود. سپس به جایگاه الگوی فکری شواهدی در فلسفهٔ آمار و اهمیت و نقش آن در مطالعات علمی پرداخته می‌شود.

کلمات کلیدی: الگوی فکری شواهدی، شواهد آماری، فلسفه‌ی علم.

۱ علم چیست؟ فلسفه چیست؟

علم معانی اصطلاحی بسیار دارد. یک تعریف رایج/رسمی علم این است: مجموعه گزاره‌ها/قضایا دربارهٔ یک موضوع که با هدف خاص تنظیم شده باشند. گفتنی است معنای اصلی و نخستین علم، دانستن در برابر ندانستن است. این معنا، که به کلیت علم (علم به معنای عام/وسیع) نظر دارد، معادل واژهٔ *knowledge* است و در برابر جهل قرار می‌گیرد. تا چند سده پیش به هر دانشی در هر زمینه‌ای علم گفته می‌شد. ولی از زمان رنسانس، کم‌کم و در برخی مجامع، علم منحصراً به دانستنی هایی اطلاق می‌شود که از طریق تجربهٔ حسی به‌دست آمده باشند. در این معنا علم در برابر جهل قرار نمی‌گیرد بلکه در برابر همهٔ دانستنی‌های غیرتجربی/غیرحسی، مانند اخلاق، فلسفه، منطق و ادبیات، قرار می‌گیرد. واژهٔ *science* معادل این معنای علم است. آشکار است که علم در این معنای دوم، بخشی از علم در معنای اول است.

[۱] Sm-taheri@ut.ac.ir

فلسفه، از اقسام علم به‌معنای وسیع آن است. فلسفه، به‌کوتاهی، عبارت است از هستی‌شناسی (وجود/بودن و احکام و عوارض آن)، و با کمی توضیح عبارت است از تلاش برای آگاهی از حقیقتِ هستی‌ها، به‌ویژه حقیقتِ آگاهی، ذهن، زبان و حیات. تعبیری عام‌تر از فلسفه نیز رایج است که طبق آن فلسفه عبارت است از هر نوع فعالیت خردورزانه، و لذا فلسفیدن به خردورزی/استدلال‌ورزی معنی می‌شود.

۲ فلسفهٔ علم چیست؟

از حدود یک سده پیش، مطالعات دربارهٔ چیستی و چگونگی علم بیشتر و متمرکزتر شد. این مطالعات رفته‌رفته عنوان فلسفهٔ علم ($Philosophy\ of\ Science$) به‌خود گرفت. این شاخه از دانش به مبانی و ماهیت علم، معرفت‌بخشی علم، اصول روش‌های علمی و سنجش این روش‌ها می‌پردازد. به سخن دیگر، فلسفهٔ علم عبارت است از علمِ علم. مهم‌ترین مباحث و پرسش‌هایی که در فلسفهٔ علم، به معنای عام، مطرح است عبارتند از:

۱) چه ویژگی‌هایی بررسی علمی را از دیگر بررسی‌ها متمایز می‌کند؟

۲) ملاک اعتبار روش‌های علمی چیست؟

۳) تبیین علمی ($Scientific\ Explanation$) چیست؟ و ملاک‌های درستی آن چیست؟

۴) قوانین علمی چه جایگاهی در معرفت‌بخشی دارند؟ این قواعد چه چیزی و چگونه امور واقع را بازنمایی می‌کنند؟

۵) جایگاه مشاهده ($Observation$) و فرضیه ($Hypothesis$) در علم چیست؟

۶) استقراء ($Induction$) چه نقشی در علم دارد؟

۷) آیا علم ماهیتی فراتاریخی دارد یا اینکه متناسب با ارزش‌ها و تاریخ اقوام می‌توان علوم متفاوتی داشت؟

۸) آیا تحول علم تدریجی و گام به گام است یا انقلابی و کل‌گرایانه؟

در مباحث بالا، علم خاصی منظور نیست بلکه پرسش‌ها دربارهٔ علم به‌طور کلی است.

از سوی دیگر، به‌مرور زمان شاخه‌هایی از فلسفهٔ علم سامان گرفت که پرسش‌هایی همانند پرسش‌های بالا را دربارهٔ علم‌هایی خاص مدنظر قرار می‌دهند. مانند فلسفهٔ علم ریاضی (یا، به‌کوتاهی، فلسفهٔ ریاضی)، فلسفهٔ تاریخ، فلسفهٔ فیزیک، فلسفهٔ اخلاق و فلسفهٔ هنر. برای مثال، در فلسفهٔ ریاضی ($Philosophy\ of\ Mathematics$) به چنین مباحثی پرداخته می‌شود:

۱) حقیقت مفاهیم و اشیاء ریاضی چیست؟

۲) رابطهٔ بین منطق و ریاضیات چیست؟

۳) ریاضیات چه نوع آگاهی به انسان می‌بخشد؟ و معرفت‌بخشی ریاضیات چگونه است؟

۴) چه رابطه‌ای بین جهان انتزاعی ریاضیات و جهان واقع وجود دارد؟ مدل‌ها و روابط ریاضی چه اموری از عالم واقع را بازنمایی می‌کنند؟

۵) ذهن چگونه با مفاهیم و قضایای ریاضی آشنا می‌شود؟

۶) آیا درستی احکام ریاضی صرفاً وابسته به ماهیت صوری آنها است یا اینکه این احکام دارای محتوای

خبری هستند؟

۷) کاربردپذیری (*Applicability*) ریاضیات به چه معنی است و چگونه است؟

۸) فرایند رشد نظریه‌های ریاضی چگونه است؟

در پاسخ به پرسش‌های بالا، اتفاق‌نظر بین اندیشمندان ریاضی نیست. در این باره مکاتب مختلف فلسفهٔ ریاضی پاسخ‌های متفاوتی دارند. مهم‌ترین این مکاتب عبارتند از منطق‌گرایی، شهودگرایی (مفهوم‌گرایی/ساخت‌گرایی)، صورت‌گرایی و، از چند دههٔ پیش، انسان‌گرایی و طبیعی‌گرایی. علاقمندان به آشنایی با این مکاتب و پاسخ‌هایی به پرسش‌های بالا به (۱) و (۴) مراجعه کنند. همچنین برای آشنایی با کلیاتی دربارهٔ علم، فلسفه و فلسفهٔ علم، و تفاوت‌ها و تمایزهای آنها، به (۵) و (۲) مراجعه شود.

۳ فلسفهٔ علم آمار (به‌کوتاهی: فلسفهٔ آمار) چیست؟

همچون هر شاخهٔ دیگری از علم، علم آمار نیز در دو جهت متفاوت می‌تواند گسترش یابد. جهت شناخته‌شده‌تر، ساختاری است. مطالعاتی که رو به این جهت دارند عمدتاً مبتنی بر این پرسش‌اند: "بر پایهٔ آنچه داریم چه می‌توان ساخت؟" (/ "چه می‌توان به‌دست آورد؟") جهت دوم به ژرفا توجه دارد. در این سو پرسش کلیدی این است: "آنچه داریم بر چه بنیان‌هایی قرار دارد؟" مطالعاتی که معطوف به این جهت است فلسفهٔ آمار (*Philosophy of Statistics*) نام دارد. در فلسفهٔ آمار به پرسش‌هایی ازین دست پرداخته می‌شود:

۱) بنیان‌های منطقی علم آمار چیست، و استنباط آماری بر پایهٔ چه اصولی استوار است؟ آیا این اصول سازگار و پذیرفتنی هستند؟

۲) آیا علم آمار روش‌شناسی (*Methodology*) خاص خود را دارد؟ اگر آری، این روش‌شناسی چگونه است و چه مؤلفه‌هایی دارد؟

۳) هدف یا اهداف علم آمار و روش‌های آماری چیست؟ آیا هدف از روش‌های آماری ـ برای مثال آزمون فرضیه‌های آماری (*Testing Statistical Hypotheses*) ـ تصمیم‌گیری در مورد پذیرش یا رد یک فرضیه است؟ آیا هدف، رسیدن به یک باور (*Belief*) در مورد فرضیه است؟ یا اینکه هدف، سنجش میزان تأیید فرضیه به‌وسیلهٔ مشاهدات است؟

۴) مدل‌های آماری چه اموری را از عالم واقع بازنمایی (*Representation*) می‌کنند؟ و به این بازنمایی‌ها چه مقدار و چگونه می‌توان اطمینان داشت؟

۵) مدل‌های آماری چه هنگام و چگونه و چه نوع از مفهوم علیّت (*Causality*) را بازنمایی می‌کنند؟

۶) در بررسی‌های آماری، مشاهدات را چگونه تفسیر کنیم؟ ابزارهای مناسب برای تفسیر مشاهدات (/ اطلاعات /داده‌ها) در مقام شواهد آماری (*Statistical Evidence*) چیست؟ قوّت داده‌های مشاهداتی به‌نفع یک فرضیه یا مدل را چگونه بسنجیم؟

۷) آیا مدل‌ها و تحلیل‌های آماری جنبهٔ تبیینی (*Explanatory*) دارند یا جنبهٔ تأییدی (*Confirmatory*)؟

۸) الگوهای فکری استنباط آماری ـ به‌ویژه فراوانی‌گرایی (*Frequentism*) و بیزگرایی (*Bayesianism*) ـ چه تمایزات و تفاوت‌هایی با هم دارند و هر یک چه برجستگی‌ها و ضعف‌هایی دارند؟ این ضعف‌ها چه تأثیراتی بر تفکر رایج آماری گذاشته است؟

این پرسش‌ها از سویی مرتبط با استنباط آماری و از سویی به مباحثی از روش‌شناسی و فلسفهٔ علم مربوط است. و البته کاربران آمار در هر شاخه از علم و فناوری، هنگام تحلیل‌های آماری و تفسیر نتایج، با چنین پرسش‌ها و چالش‌هایی روبرو هستند.

الگوهای فکری در آمار

در پاسخ به پرسش‌های بالا، الگوهای فکری (پارادایم‌ها/مکاتب فکری) آمار پاسخ‌های متفاوتی مطرح کرده‌اند. تا حدود دو دهه پیش، الگوهای فکری در فلسفهٔ آمار به دو ردهٔ کلی فراوانی‌گرا و بیزگرا (منسوب به تامس بیز کشیش و ریاضیدان انگلیسی سدهٔ هجدهم م.) تقسیم می‌شدند. مشهور چنین است که تمایز اصلی این دو الگوی فکری به عینی بودن فراوانی‌گرایی و ذهنی بودن بیزگرایی باز می‌گردد، هرچند برخی بیزگرایان منکر این نوع نگرش به مکتب بیز هستند. در هر حال، پاسخ‌های این دو الگوی فکری به بسیاری از پرسش‌های بالا متفاوت است. گفتنی است در الگوی فکری فراوانی‌گرا (در برخی متون: مکتب کلاسیک) دو گرایش عمده وجود دارد: الگوی فکری نیمن‌ـ‌پیرسن $(Neyman-Pearson\ Paradigm)$ و الگوی فکری فیشری $(Fisherian\ Paradigm)$، که تمایز اصلی آنها در این است که الگوی فکری نیمن‌ـ‌پیرسن بیشتر شبیه نوعی الگوی تصمیم‌گیری است، در حالی‌که روش کار در الگوی فکری فیشری مبتنی بر آزمون‌های رد است. به‌نظر می‌رسد که مکتب فکری نیمن‌ـ‌پیرسن هم‌راستا با (/ متأثر از (؟)) گسترش شاخه‌های بهینه‌سازی و نظریهٔ تصمیم در ریاضیات پس از جنگ جهانی دوم شکل گرفته است؛ و همچنین توجه به الگوی فکری فیشری، به‌ویژه آزمون‌های رد / آزمون‌های معنی‌داری، هم‌راستا با (/ متأثر از (؟)) مکتب ابطال‌پذیری پوپر بوده است که در اواسط سدهٔ بیستم رواج یافته بود.

۴ استنباط شواهدی: یک الگوی فکری جدید مبتنی بر درست‌نمایی

الگوی فکری فراوانی‌گرا، که الگوی فکری مسلط بر آموزش و پژوهش‌های آماری بوده و هست، از اواسط سده بیستم م. مورد نقادی بیزگرایان قرار گرفت. بیزگراها توجه مجامع علمی را به نقاط ضعف منطقی و ناسازگاری‌های روش‌های آماری فراوانی‌گرا (در هر دو الگوی فکری نیمن‌ـ‌پیرسن و فیشری) جلب کردند. از سوی دیگر، کاربران آمار نیز هنگام تفسیر نتایج روش‌های آماری فراوانی‌گرا با دشواری‌هایی روبرو می‌شدند. بیزگراها تلاش کردند تا کاستی‌های فراوانی‌گرایی را در چارچوب الگوی فکری بیزی رفع کنند. ولی آمار بیز یک مؤلفهٔ اصلی، یعنی توزیع پیشین، دارد که ذهنی است، و لذا اقبال مجامع علمی به آمار بیز همواره محل مناقشه بوده است. مجادله میان فراوانی‌گراها و بیزگراها درازدامن است. برای مطالعهٔ نمونه‌هایی کوتاه از استدلال‌های دو رقیب، (۷) و (۹) را ببینید. یکی از جدیدترین بررسی‌ها در مقایسهٔ روش‌های این دو الگوی فکری در (۶) ارائه شده است.

از اواخر سدهٔ بیستم م. مبانی الگوی فکری جدیدی در آمار شکل گرفت که به الگوی فکری شواهدی (نیز، الگوی فکری درست‌نمایی) نامدار شد. این الگوی فکری مبتنی بر اصلی به نام اصل درست‌نمایی $(The\ Likelihood\ Principle)$ و قاعده‌ای به نام قانون درست‌نمایی $(The\ Low\ of\ Likelihood)$ است. گفتنی است که، از لحاظ تاریخی، تابع درست‌نمایی به‌وسیلهٔ فیشر در سال ۱۹۲۱ م. معرفی شد. با معرفی اصل

درست‌نمایی به‌وسیلهٔ برنباوم در سال ۱۹۶۲ م. و همچنین قضیهٔ برنباوم دربارهٔ معادل‌بودن اصل درست‌نمایی با مجموع اصول شرطی‌سازی (/ شرطی‌بودن) و بسندگی، توجه به رویکردهای درست‌نمایی جدی‌تر شد. تا اینکه با انتشار مقالات ریچارد رویال در دههٔ ۱۹۸۰ و سپس انتشار کتاب وی در سال ۱۹۹۷ ((۱۰) و (۳)) الگوی فکری شواهدی در مقام یک الگوی فکری جانشین (/ بدیل / رقیب) برای دو الگوی فکری فراوانی‌گرا و بیزگرا مطرح شد، (نیز (۸) را ببینید).

در الگوی فکری شواهدی، شواهد آماری و سنجش عینی آنها نقش کلیدی دارند. بررسی شواهد آماری و سنجش آنها برپایهٔ درست‌نمایی و مفاهیم مرتبط، مانند تابع درست‌نمایی، اصل درست‌نمایی، قانون درست‌نمایی و نسبت درست‌نمایی، انجام می‌شود. برجستگی‌های این الگوی فکری را می‌توان چنین برشمرد:

۱) الگوی فکری شواهدی چارچوبی علمی برای تفسیر داده‌های آماری به منزلهٔ شواهد فراهم می‌آورد. این نکته از این رو مهم است که یکی از اهداف اساسی در هر زمینهٔ علمی، تولید شواهد دربارهٔ پدیدهٔ تحت بررسی است، و علم آمار نقش و وظیفه‌ای اساسی در فراهم آوردن روش‌های عینی برای توصیف، تفسیر و کمّی‌سازی این شواهد دارد.

۲) مبانی نظری این الگوی فکری، اصل درست‌نمایی است که یک اصل مورد پذیرش غالب آماردانان است.

۳) استنباط شواهدی به تناقض‌ها و کاستی‌هایی که الگوی فکری فراوانی‌گرا دچار آنهاست نمی‌انجامد. این موضوع همان اندازه که از دید نظری مهم است از لحاظ کاربردی نیز اهمیت دارد، زیرا کاربران روش‌های آماری فراوانی‌گرا در به‌کارگیری این روش‌ها و تفسیر نتایج آنها با چالش‌های متعددی روبرو هستند.

۴) این الگوی فکری، روش‌هایی ساده و عینی برای اندازه‌گیری قوت شواهد آماری، به‌نفع یک فرضیه یا یک مدل، پیش‌رو می‌گذارد.

۵) الگوی فکری شواهدی فارغ از مؤلفه‌های ذهنی و شخصی است. این نکته تمایز و برتری ویژهٔ این الگوی فکری در مقایسه با الگوی فکری بیزگرا (شامل توزیع پیشین و تابع زیان) است و از این جهت مهم است که در شاخه‌های مختلف دانش بر موضوع عینی‌گرایی در ارزیابی مشاهدات تأکید می‌شود. عبارت معروف در این زمینه بیانگر محور اصلی این نگرش است: "ببینیم داده‌ها خود چه می‌گویند؟"

بحث و نتیجه‌گیری

سه مکتب (الگوی فکری) در فلسفهٔ آمار را به کوتاهی توضیح دادیم و جایگاه الگوی فکری شواهدی را تشریح کردیم. ولی یک پرسش کلیدی همچنان پابرجاست: **"کدام الگوی فکری مناسب است؟"** در پاسخ به اینکه کدام الگوی فکری آماری مناسب است، باید نخست به این پرسش پاسخ دهیم که: هدف (از دیدگاهی دیگر: وظیفهٔ) علم آمار چیست؟ و آیا اساساً علم آمار یک هدف و وظیفه دارد؟ یا می‌توان اهداف و وظایفی چندگانه برای آن تصور کرد؟ برای تشریح مطلب، مثالی کلی را، با الهام از مثالی دربارهٔ تشخیص پزشکی از کتاب رویال، طرح می‌کنیم:

فرض کنید یک بررسی مشاهداتی دربارهٔ یک مدل آماری یا یک فرضیهٔ آماری انجام شده است. سه پرسش پیش‌روی ما قرار دارد:

۱) برپایهٔ این مشاهدات، چه کاری باید انجام دهیم؟ (پذیرش مدل/فرضیه یا رد آن؟)
۲) برپایهٔ این مشاهدات، چه باوری دربارهٔ درستی مدل/فرضیه می‌توانیم داشته باشیم؟
۳) نتایج حاصل، چه چیزی دربارهٔ مدل/فرضیه به ما می‌گوید؟ به سخن دیگر، چگونه مشاهدات حاصل را همچون شواهدی در پشتیبانی از مدل/فرضیه تفسیر کنیم؟

هر یک از سه الگوی فکری مورد بحث، به یک پرسش از سه پرسش بالا می‌پردازند. بدین‌ترتیب که

*۱) الگوی فکری فراوانی‌گرا، به تدارک پاسخ به پرسش نخست می‌پردازد. گویی هدف و وظیفهٔ آمار تصمیم‌گیری است.

*۲) برپایهٔ الگوی فکری بیزگرا، پرسش دوم پرسش اصلی علم آمار است، زیرا در مکتب بیز آنچه مهم است عبارت است از تصحیح باور در پرتو شواهد.

*۳) طبق مکتب شواهدی، این پرسش سوم است که پرسش کلیدی است، زیرا وظیفه و هدف علم آمار پاسخ به این پرسش است.

پیش‌تر گفتیم که سه الگوی فکری فراوانی‌گرا، بیزگرا و شواهدی (شاهدگرا)، مبانی مختلف و روش‌های متفاوتی دارند. اکنون آشکار است که این مبانی و این روش‌های متفاوت، در تدارک پاسخ به پرسش‌هایی هستند که اساساً خود پرسش‌ها متفاوت‌اند. درواقع، باید گفت:

سه پاسخ رویکردهای فراوانی‌گرا، بیزگرا و شاهدگرا، سه پاسخ به سه پرسش متفاوت‌اند، نه سه پاسخ به یک پرسش.

اما کدام پرسش، پرسش کلیدی و اساسی علم آمار است؟ بررسی این پرسش و پاسخ به آن در این مختصر نمی‌گنجد. شاید وقتی دیگر.

مراجع

[۱] براون، ج.ر. (۱۳۹۴). *فلسفهٔ ریاضیات* (*ترجمهٔ محمدقاسم وحیدی‌اصل*)، انتشارات نوشتگان.

[۲] شیخ‌رضایی، ح. و کرباسی‌زاده، ا. (۱۳۹۵). *آشنایی با فلسفهٔ علم*، انتشارات هرمس.

[۳] رویال، ر.م. (۱۳۹۸). *شواهد آماری: رویکردی مبتنی بر درست‌نمایی* (*ترجمهٔ ناصررضا ارقامی و سید محمود طاهری*)، مرکز نشر دانشگاهی.

[۴] کولی‌ون، م. (۱۳۹۶). *درآمدی بر فلسفهٔ ریاضی معاصر* (*ترجمهٔ کامران شهبازی*)، انتشارات نقد فرهنگ.

[۵] گیلیس، د. (۱۳۹۴). *فلسفهٔ علم در قرن بیستم* (*ترجمهٔ حسن میانداری*)، انتشارات سمت.

[6] Dienes, Z. and Mclatchie N. (2018), Four reasons to prefer Bayesian analyses over significance testing, *Psychonomic Bulletin and Review* 25, 207 - 218.

[7] Efron, B. (1986), Why isn't everyone a Bayesian?, *The American Statistician* 40, 1-5.

[8] Evans M. (2015), *Measuring Statistical Evidence Using Relative Belief,* Chapman and Hall/CRC.

[9] Gustafson P. (2019), *Paul's Top Ten Reasons To Be A Bayesian*, https://www.stat.ubc.ca/ gustaf/530/topten.pdf

[10] Royall, R.M. (1997), *Statistical Evidence, A likelihood paradigm,* Chapman and Hall.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

نیاز توجه به مبانی آمار در آمار کاربردی

عاشوری، م [۱]

[۱]موسسه پژوهشی حکمت و فلسفه

چکیده

آیا مبانی آمار تنها مورد نیاز پژوهشگران آمار نظری و آمار ریاضی است و روش‌شناسان و پژوهشگران تنها از تکنیک‌های آمار کاربردی استفاده می‌کنند و نیازی به آشنایی با مبانی آمار ندارند؟ در این مجال با دو مثال از روان‌سنجی و اقتصادسنجی، نشان داده می‌شود که توجه به مبانی آمار چگونه در آمار کاربردی نقشی اساسی دارد. آماره پایایی ملاکی بسیار رایج برای ارزیابی ابزار اندازه‌گیری سازه‌های روانی است. این آماره مبتنی بر محاسبه همبستگی میان نمره‌های اندازه‌گیری شده است. اما اندیشه اصلی این است که نتایج اندازه‌گیری شده متشکل از دو بخش است، بخش ثابت که بازنمای نمره واقعی و نمره مشاهده شده که حاصل جمع نمره واقعی و خطا است. خطا نیز نوسان تصادفی پیرامون نمره واقعی در نظر گرفته شده است. مثال دیگر رواج استفاده از استراتژی «گرنجر» برای «شناسایی علّی» در تحلیل سری‌های زمانی است. این روش به ما می‌گوید که اگر دخیل کردن مقدار گذشته متغیر مستقل در پیش‌بینی متغیر وابسته باعث بهبود در دقت پیش‌بینی شود، می‌توانیم نتیجه بگیریم که X عامل علّی برای متغیر Y بوده است.

کلمات کلیدی: فلسفه آمار، چیستی احتمال، متغیر تصادفی، روانسنجی، اقتصادسنجی.

۱ پیش‌گفتار

پرسش اصلي اين است كه آيا مباني آمار تنها مورد نياز پژوهشگران آمار نظري و آمار رياضي است و پژوهشگران از تكنيكهاي آمار كاربردي استفاده ميكنند و نيازي به آشنايي با مباني آمار ندارند؟ برای پاسخ به این پرسش به کاربرد آمار در مدلسازی علمی میپردازیم. مدلها در علوم اهمیتی اساسی دارند. بسیاری از پیشرفتهای علمی مرهون قدرتی است که مدلهای علمی در شناخت پدیدههای طبیعی و انسانی در اختیار ما قرار دادهاند.

[۱] Ashoori425@yahoo.com

«مدلسازی علمی» از روشهای رایج در علوم پایه، مهندسی، علوم زیستی و علوم اجتماعی است و نقش استنباط آماری در این امر انکارناپذیر است. بررسی نحوه کاربردپذیری «استنباط آماری» در مدلسازی، نشان میدهد که چگونه توجه به مبانی آمار در آمار کاربردی نقشی اساسی دارد و عدم توجه به آن میتوان سبب خطا در کاربست آمار در فهم پدیدهها گردد.

۲ چیستی مدل علمی در دیدگاه ساختاری

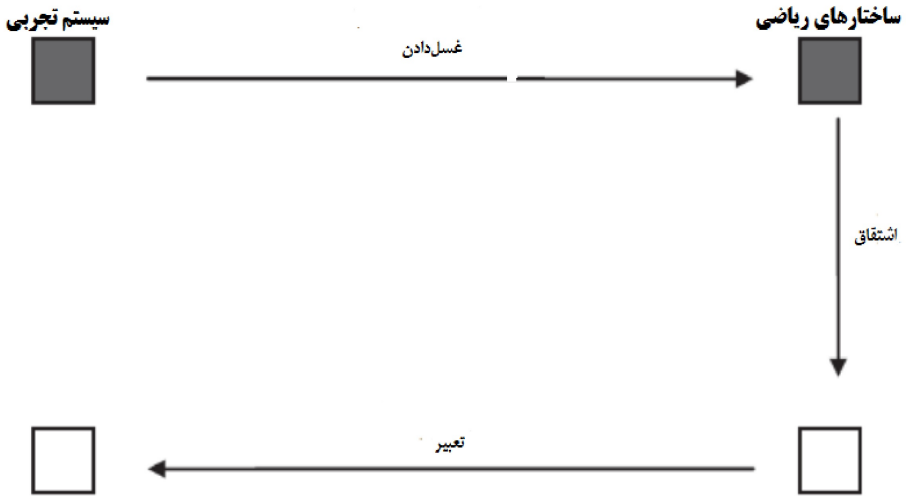
در دیدگاهی واقعگرایانه، مدلها چیزی را در مورد واقعیت «بازنمایی» میکنند. در اصطلاح آنچه را بازنما است، «منبع بازنمایی»[۱] یا «مدل» و واقعیتی را که بازنمایی میشود «هدف بازنمایی»[۲] گویند (سوارز[۳]، ۲۰۰۹). یعنی، رابطهٔ بازنمایی دو طرف دارد: منبع و هدف (بنگرید به شکل ۱).

شکل ۱: طرفین رابطهٔ بازنمایی

از این منظر «مدل» بیانگر و بازنماییکننده تمام واقعیت نیست و تنها قرار است که «بخشی انتخابی» از واقعیت را نشان دهد. در کاربرد ریاضیات و آمار، از معادلات و ساختارهای ریاضی به عنوان منبع بازنمایی و مدل استفاده میشود. در دیدگاههای ساختاری، ریاضیات منبعی غنی از «ساختار»ها است و مدل کمّی، یک ساختار نظریهٔ مجموعهای[۴] است. در این دیدگاه، مدل M=<A،R> با مجموعه اجزاء ساختاری A=a۱، a۲، a۳،... و رابطه R میان این اجزاء است؛ هدف بازنمایی نیز واقعیت T=<A’،L> با مجموعه اجزاء A’=a۱’،a۲’،a۳’،... و رابطه L میان این اجزاء است (عاشوری و طاهری، ۱۳۹۶). مدلهای ریاضی و مدلهای آماری هر دو در این وِیژگی که روابط میان مفاهیم را با استفاده از متغیرها صورتبندی و قابل اندازهگیری میکنند، مشترک هستند ولی مدل آماری این اختلاف را با مدل ریاضی دارد که در آن مؤلفه یا مؤلفههایی وجود دارد که بیانگر «عدم اطمینان احتمالاتی»[۵] درباره آن متغیرها است. در مدل آماری این فرض وجود دارد که برخی از متغیرهای مشخصکننده مفاهیم علمی را باید به عنوان متغیری تصادفی[۶] در نظر گرفت؛ در نتیجه باید از توابع توزیع[۷] و پارامترهای آن در مدلسازی روابط میان وضعیتهای امور استفاده نمود.

۳ کاربردپذیری ریاضی و آمار در مدلسازی علمی

هنگامی یک نظریه ریاضیاتی در علوم کاربرد پیدا میکند، مدل ریاضی یا آماری باید روابط ساختاری اصلیِ سیستم مورد مطالعه را پوشش دهد، ولی کارکرد ریاضیات و آمار صرفاً این نیست که نقشهای ساختاری از هدف بازنمایی تهیه کند. عملیاتهای تبدیل ساختارها در «آمار و ریاضی کاربردی» اهمیت بسیاری دارند. ساختارها به یکدیگر تبدیل میشوند و قدرت اصلی نظریههای ریاضیات نیز در همین «تبدیلات» است. در این دیدگاه، برای ریاضیات کاربردی سه گام طرح میشود: (۱) یک نگاشت از «سیستم مورد مطالعه یا تجربی»[۸] به یک ساختار مناسب ریاضی که آن را «غسلدادن»[۹] مینامند؛ (۲) «اشتقاق» یا «استنتاج» ساختار ریاضی جدید از ساختار ریاضی اولی با استفاده از روابط تبدیل ریاضی؛ (۳) یک نگاشت از ساختار ریاضی جدید به سیستم مورد مطالعه که آن را نگاشت «تعبیر» مینامند(بوئنو و کولیوان[۱۰]، ۲۰۱۱).



شکل ۲: نگاشت میان دامنه امور فیزیکی و امور ریاضی و استنتاج ساختارها در جهان ریاضی و آمار

نقش اصلی آمار و ریاضی کاربردی، مربوط به گام دوم است. به وسیله این تبدیلات، ویژگیهایی که در ساختار اولیه پنهان بودند، در ساختار تبدیل یافته آشکار میگردند. یعنی اجزائی در ساختار جدید ظاهر میشوند که باعث کشف ویژگیهایی بدیع در سیستم هدف میشوند.

۴ دو مبحث از مبانی آمار

فلسفه و مبانی آمار و احتمال مباحث بسیاری دارد، مانند «تعابیر فلسفی احتمال»، «اندازه گیری عدم قطعیت»، «احتمال رویدادهای تک موردی»، «شانس و تصادف»، «اختلاف نظرهای فلسفی درباره احتمال شرطی»، «تفاوت آمار و نظریه احتمال»، «الگوهای فکری (پارادایمهای) استنباط آماری»، معیارهای انتخاب مدل»، «پارادکس های آمار و احتمال»، «معیارهای انتخاب براوردگر مانند دقت، صحت و سادگی» و غیره (باندی اوپادیهای،). دو بحث از مبانی آمار در کاربرد استنباط آماری اهمیت اساسیتری دارد: (۱)

تعابیر فلسفي از احتمال و (۲) پارادایمهای استنباط آماری. بحث اول در چیستی متغیر تصادفی و بحث دوم در رابطه دادهها و مدل آماري نقش دارد.

۱.۴ چیستی متغیر تصادفی

مسئله مفاد جملات احتمالاتی، با عنوان «تعابیر فلسفی احتمال» شناخته میشود (گیلیس[۱۱]، ۲۰۰۰). مسئله این است که آیا احتمالها «امر و ویژگی عینی» مانند بسامد[۱۲]، گرایش[۱۳] و تمایل[۱۴] هستند(فون میزس،[۱۵] ۱۹۵۷؛ رایشنباخ،[۱۶] ۱۹۴۹)، یا آنکه احتمالها «امری ذهنی» مانند درجه باور (رمزی[۱۷]، ۱۹۳۱؛ دِفینتی[۱۸]، ۱۹۳۱) میباشند؟ از تعابیر بسیار مطرح درباره احتمال «تعبیر بسامدی» است. تعبیر بسامدی میگوید که احتمال یک نتیجه برابر با نسبت «تعداد آزمایشهایی که در آن نتیجه حاصل میشود» به «تعداد آزمایشهایی که صورت گرفته است» مفهوم «آزمایش» در معنایی بسیار گسترده مدّ نظر است (فون میزس، ۱۹۸۱). در مقابل، بیزگرایی دیدگاهی ذهنیگرا درباره تعبیر احتمال است که مدعی است اعتقادات (باورهای) یک شناسا باید در همه زمانها از اصول احتمال متابعت کند و تنها زمانی میباید تغییر کند که شناسا اطلاعات جدید به دست میآورد. این مبحث در مسئله ما خود را به صورت چیستی متغیر تصادفی نشان میدهد.

۲.۴ رابطه دادهها و مدل آماری

در یک تقسیمبندی کلّی، سه پارادایم اصلی در حوزه استنباط آماری وجود دارد: (۱) استنباط بسامدی (فراوانیگرا)، (۲) استنباط بیزی و (۳) استنباط شواهدی (رویال[۱۹]، ۱۹۹۷؛ ارقامی، ۱۳۹۱). رویال اختلاف این سه رویکرد را در هدفی میدانند که از استنباط دنبال میشود. هر هدف در قالب یک پرسش قابل طرح است، پرسشهای اساسی که میتواند هدف استنباط آماری باشد از قرار زیر است: الف. با توجه به دادهها چه عملی باید انجام دهیم؟ ب. با توجه به شواهد باید به چه فرضیهای و به چه میزان باور داشتهباشیم؟ پ. شواهد در مورد فرضیه H و نقیض آن چه میگویند؟ (سوبر[۲۰]، ۲۰۰۸: ۵-۴؛ عمادی، ۱۳۸۴ :۱۴) خواه آنچنانچه رویال مدعی است که میان پارادایمهای استنباط آماری و این پرسشها تناظر برقرار است یا خیر، این پرسشها مشخص میکند که عرض اصلی استنباط چیست و دادهها نقش شواهد را دارد یا خیر و رابطه آن با فرضیه رابطهای منطقی است یا صرفا رابطهای ذهنی میباشد؟

۵ دو مثال از لزوم توجه به مبانی آمار

این دو بحث فوق چگونه خود را در آمار کاربردی نشان میدهد؟ برای روشنشدن به دو مثال از روانسجی و اقتصادسنجی اشاره میکنیم که از مصداقهای پرکاربرد کاربرد آمار در علوم است.

۱.۵ روانسمجی و محاسبه پایایی در ارزیابی ابزارهای اندازهگیری

از کاربردهای روشهای کمّی تحقیق در علوم رفتاری، اندازهگیری سازههای[۲۱] روانی است و برای اندازهگیری این سازهها از تعاریف عملیاتی و ابزارهایی چون خودگزارشی یا مشاهده محقق استفاده میشود. در روانسنجی استانداردبودن ابزار اندازهگیری اهمیت بسیار دارد و روایی و پایایی دو ملاک اصلی برای ارزیابی ابزار اندازهگیری است. پایایی به قابلیت تکرارپذیری اندازهگیری اشاره دارد و هرچه همسانی نتایج اندازهگیری بیشتر باشدف پایایی بیشتر خواهد بود. بسنجش پایایی از آمارههای مختلفی استفاده میشود که مبتنی بر محاسبه همبستگی میان نمرههای اندازهگیریشده است. اما اندیشه اصلی این است که نتایج اندازهگیری شده متشکل از دو بخش است، یک قسمت ثابت که نمره واقعی است و نمره مشاهده شده حاصل جمع نمره واقعی و خطا است که به عنوان نوسان تصادفی پیرامون نمره واقعی در نظر گرفته شده است. اندیشه فوق نشان میدهد که تعبیری عینی از نمره مشاهده شده و نتایج اندازهگیری لحاظ شده است و آماره پایایی نیز به ما کمک میکند که تصمیم بگیریم، آیا استفاده از ایزار اندازهگیری معقول است. حال آیا میتوان از سویی تعبیری عینی از احتمال داشت و از آمارهها پایایی استفاده نمود و از سوی دیگر تعریف عملیاتی را برای اندازهگیری سازه استفاده نمود که بیشتر مناسب تعبیری ذهنی از متعیر تصادفی است.

۲.۵ اقتصادسنجي و نگاه احتمالاتی به پدیدههای اقتصادی

از آعاز تکوین و بسط روشهای اقتصادسنجی، این مسئله مطرح بوده است که آیا میتوان پدیدههای افتصادی را تصادفی و احتمالاتی دانست یا خیر؟ هاولمو[۲۲] (۱۹۴۳) برای پاسخ مثبت به این مسئله تلاش بسیار نمود. این مسئله از آن جهت اهمیت داشت که در نیمههای قرن بیستم، آمار کلاسیک و بسامدگرا رایج بود و بر اساس آن پدیده مورد مطالعه با مدل آماری باید تکرارپذیر باشد. در چنین دیدگاهی نیز انتظار محققان این است که شاخصهای اقتصادی و روابط میان متغیرها، بازنمایی کننده وضعیتهای واقعی مالی یا پولی است و به همین دلیل اغلب در پژوهشهای کمّی و آماری مسئله تحقیق بدین صورت تقریر میشود که آیا دادهها از فرضیه تحقیق حمایت میکند یا خیر؟ به عبارتی پزسش اصلی از سنخ «پرسش شواهد» است و دادههای اقتصادی به عنوان «شواهد» مؤید یا ناقض فرضیه ترسیم میشود. مسئله دیگر در اقتصادسنجی، بررسی فرضیههای علّی است. در بسیاری از پدیدههای افتصادی، تحلیل دادههایی اهمیت دارد که بر اساس «سری زمانی»[۲۳] مرتب شدهاند و پیشبینی متغیر وابسته نیاز به دانستن مقدار «آینده» متغیرهای مستقل دارد. یکی از استراتژیهای رایج در «شناسایی علّی» در تحلیلهای سری زمانی استفاده از روشی است که به ما میگوید که اگر دخیل کردن مقدار گذشته متغیر X در پیشبینی متغیر Y باعث بهبود در دقت پیشبینی شود، میتوانیم نتیجه بگیریم که X عامل علّی برای متغیر Y بوده است. این معیار مبتنی بر این ایده اساسی است که اگر دخیل کردن تغییرات یک متغیر (پس از امتحان کردن سایر متغیرها) باعث ارتقاء پیشبینی یک متغیر در دورههای بعدی شود، رابطهای باید میان این دو متغیر باشد. ولی این استراتژی به معنای دقیق فلسفی واقعا معیاری برای علیت نیست، زیرا در بسیاری از مواردی که این معیار بهکار میرود مانند اقتصاد، عاملها «پیشنگر»[۲۴] هستند و در نتیجه خیلی محتمل است که از دید اقتصادسنجی متغیر X همیشه قبل از متغیر Y حرکت کند ولی در واقع حرکت X به خاطر «انتظاراتی» باشد که از رفتار Y دارد. برای مثال

با دیدن خرید لباس گرم توسط مردم می‌توانیم انتظار داشته باشیم که دمای هوا در ماه آینده پایین خواهد رفت؛ پس خرید لباس قدرت پیش‌بینی ما برای دوره بعد را ارتقاء می‌دهد ولی واقعا خرید لباس علتی برای سردی هوا نیست. پیشنگر بودن و تأثیر انتظارات بر وضعیتهای اقتصادی، عینیبودن و تصادفیبودن واقعیتهای اقتصادی را مورد سؤال قرار میدهد.

۶ جمعبندی

هنگام کاربرد آمار در علوم، علاوه بر مبانی و قضایای آمار، ایدههایی نیز در مورد پدیدههای مورد مطالعه وجود دارد که باید میان این دو دسته مبانی سازگاری وجود داشته باشد. استفاده از آمار بسامدی یا آمار بیز باید متناسب با خصوصیت آنچه به عنوان متعیر تصادفی و هدف از تحقیق سازگاری داشته باشد.

مراجع

[۱] ارقامی، ناصر رضا ۱۳۹۱، «گفتگو در باب مکاتب آماری»، گفتگو کننده: سید محمود طاهری، در کتاب نکوداشت دکتر ناصر رضا ارقامی، انتشارات یازدهمین کنفرانس آمار ایران، تهران: دانشگاه علم و صنعت ایران و انجمن آمار ایران.

[۲] عاشوری، مهدی و سیدمحمود طاهری ۱۳۹۶، «کاربردپذیری استنباط آماری از منظر بازنمایی مدلهای علمی»، روششناسی علوم انسانی (۹۱)، صص ۸۳-۶۸

[۳] عمادی، مهدی ۱۳۸۴، استنباط شاهدگرا (پایاننامه دکترا آمار ریاضی)، مشهد: دانشگاه فردوسی مشهد، دانشکده علوم ریاضی، گروه آمار.

[4] Bueno, O., and Colyvan, M. 2011. "An Inferential Conception of the Application of Mathematics". Nous. 45: 345–374.

[5] de Finetti, B. 1931. "On the Subjective Meaning of Probability". In P. Monari and D. Cocchi (eds.), Induction and Probability. Bologna: Clueb, 1993.

[6] Gillies, D. 2000. Philosophical Theories of Probability. Routledge.

[7] Haavelmo, T. 1943. "The Statistical Implications of a System of Simultaneous Equations". Econometrica. 11: 1–12. Reprinted in D.F. Hendry and M.S.

[8] Morgan (Eds.) 1995. The Foundations of Econometric Analysis: 477–490. Cambridge University Press.

[9] Ramsey, F. 1931. "Truth and Probability". In R.B. Braithwaite (ed.), Foundations of Mathematics and Other Essays. London: Routledge Kegan Paul,

[10] 156–98.

[11] Reichenbach, H. 1949. The Theory of Probability. Berkeley: University of California Press.

[12] Royall, R. 1997. Statistical Evidence. A Likelihood Paradigm. Chapman and Hall Press.

[13] Sober, E. 2008. Evidence and Evolution. Cambridge University Press.

[14] Suárez, M. (ed.) 2009. Fictions in Science. Philosophical Essays on Modelling and Idealisation. Routledge.

[15] von Mises, R. 1957. Probability, Statistics and Truth. Macmillan.

[16] von Mises, R. 1981. Probability, Statistics and Truth, 2nd Edition, Dover Press.

[17] Wakefield, J. 2013. Bayesian and Frequentist Regression Methods. Springer.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

نگاه کردن به نمودار های توابع درستنمایی

عمادی، م [۱]

[۱] گروه آمار، دانشگاه فردوسی مشهد

چکیده

هر دو دیدگاه نیمن –پیرسنی و فیشری به آمار دارای سه بخش اصلی است : برآورد ـ آزمون فرضیه و بازه های اطمینان . این بخش ها بیانگر یک رویکرد مقطعی و تدریجی به تحلیل و تفسیر شواهد آماری اند. هریک از این سه بخش مجموعه ای پیچیده از مفاهیم و روش هایی هستند که باید کارشناسانه به کار گرفته شوند. برای مثال در یک مسیله ی آزمون فرضیه معنای رد نشدن یک فرضیه باید بر حسب : اندازه و توان آزمون – تمایز بین رد نکردن یک فرضیه و پذیرش آن – تمایز بین معنی داری آماری و معنی داری عملی – تاثی اندازه نمونه بر معنای مقدار-p و مانند این ها درک شوند. ولی دیده ایم که چنین موضوع هایی آنقدر پیچیده اند که گاه خبرگان آمار نیز در آنها اختلاف نظر دارند. از سوی دیگر قبول این واقعیت که توابع درستنمایی ابزار مناسبی برای نمایش و ارایه شواهد اماری اند تحلیل های آماری را ساده و هماهنگ می سازند.

کلمات کلیدی: آزمون فرضیه، بازه های اطمینان، تفسیر شواهد آماری، توابع درستنمایی، مقدار-p

۱ پیش‌گفتار

یکی از مهمترین تکنیک های آمار کاربردی استفاده از فاصله اطمینان است. متاسفانه در این روش به این نکته توجه نمی کنیم که امکان تکرار در آزمایشات وجود ندارد. زیرا از لحاظ نظری قضایای مربوطه متکی بر استنباط کلاسیک یا فراوانی گرا است. فاصله بیزی یا فاصله باور نیز به توزیع پیشین وابسته بوده که آن هم به اعتقادات شخصی تکیه می کند و نمی تواند نظر آمار دان یا محقق را جلب کند. براین اساس فاصله شواهدی که مبتنی بر تابع درستنمایی است را پیشنهاد می کنیم که هیچکدام از معایب روشهای فوق را ندارد.

[۱] emadi@um.ac.ir

توابع درستنمايی

فرض كنيد متغير تصادفي X داراي توزيع نرمال با ميانگين μ و واريانس معلوم يك باشد. دو فرضية رقيب زير را در نظر مي‌گيريم.

$$\begin{cases} H_1 : \mu = 0 \\ H_2 : \mu = 1 \end{cases}$$

براي بررسي دو فرضية فوق به روش شواهدي نمونه‌اي تصادفي به حجم n اخذ مي‌كنيم. در اينصورت داريم

$$r = \frac{L_1}{L_2} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2} x_i^2\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}(x_i - 1)^2\right)}$$

$$= \prod_{i=1}^{n} \exp\left(\frac{1}{2} - x_i\right) = \exp\left(n\left(\frac{1}{2} - \bar{x}\right)\right),$$

بررسي بهتر استنباط شواهدي را *Royall* به طريق ديگري بيان مي‌كند كه مي‌گويد "بجاي آنكه نسبت درستنمايي را براي دو فرضية رقيب بدست آوريم، تابع درستنمايي را (باتقسيم بر ماكسيمم مقدار آن) طوري تعديل مي‌كنيم كه بيشترين مقدار آن يك باشد. سپس آن تابع را بر حسب پارامتر مورد نظر رسم كرده و در نهايت با توجه به منحني تابع درستنمايي استنباط شواهدي را نه تنها براي دو فرضية رقيب مورد بحث بلكه براي هر دو فرضية سادة رقيب انجام مي‌دهيم". اين روش براي مثال ۲-۱ به شكل زير است:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}(x_i - \mu)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

براي اينكه بيشترين مقدار $L(\mu)$ يك شود كافيست ماكسيمم $L(\mu)$ را محاسبه كرده و $L(\mu)$ را بر آن تقسيم كنيم.

$$\max_{\mu\in R} L(\mu) = L(\bar{x})$$

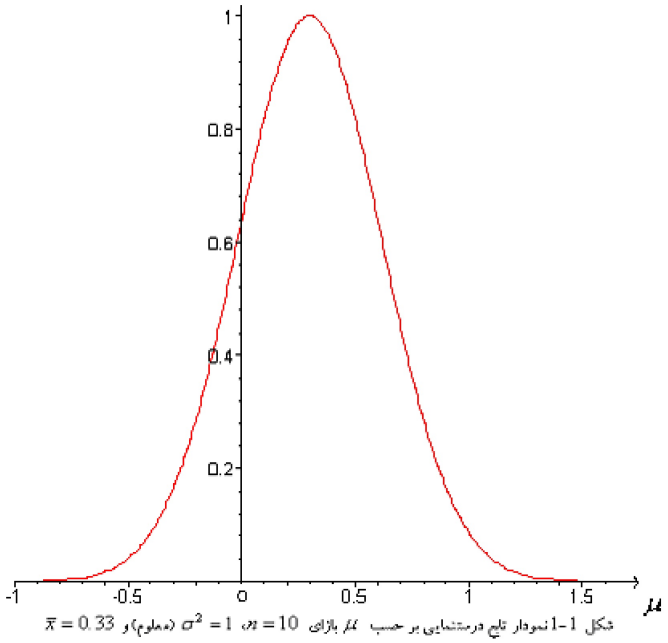
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_i^2+\frac{n}{2}\bar{x}^2\right),$$

در نتیجه داریم

$$L(\mu) \propto \exp\left(-\frac{n(\mu-\bar{x})^2}{2}\right).$$

حال تابع فوق را بر حسب μ رسم مي‌كنيم.



شكل ۱-۱ نمودار تابع درستنمايي بر حسب μ بازاي $n=10$، $\sigma^2=1$ (معلوم) و $\bar{x}=0.33$

با توجه به شكل ۱ - ۱ مشاهده مي‌شود $L(\cdot)$ بزرگتر از $L(1)$ بوده پس داده‌ها از H_1 پشتيباني بيشتري مي‌كنند و اگر k را برابر ۸ اختيار كنيم پشتيباني از H_1 قوي محسوب مي‌شود.

فواصل آماری

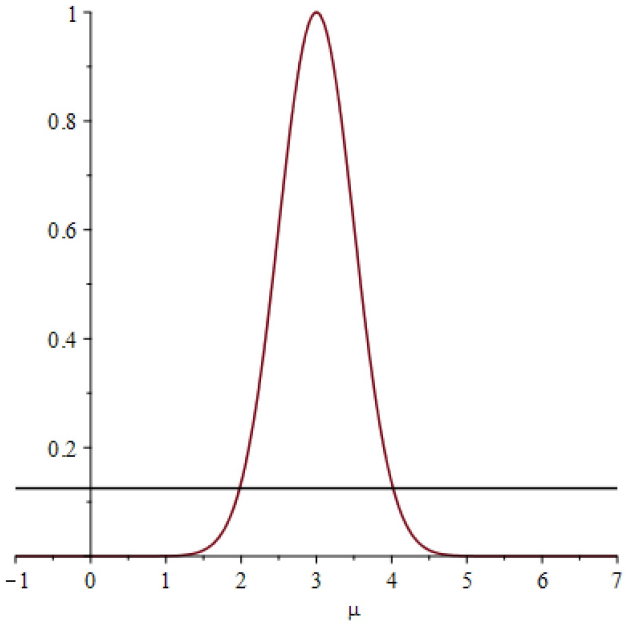
آمار كلاسيک: فاصله اطمينان

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{0.025}$$

آمار بیز: فاصله اعتبار یا باور

$$\frac{n\bar{x}}{n+1} \pm \frac{Z_{0.025}}{\sqrt{n+1}}$$

آمار شواهدی: فاصله شواهدی



ـ تابع درستنمايي در قبال وجود پارامتر مزاحم
روش متعامدسازي

فرض كنيد X_1 و X_2 متغيرهاي تصادفي مستقل از توزيع پواسن به ترتيب با پارامترهاي λ_1 و λ_2 باشند. تابع درستنمايي متعامد را براي پارامتر $\theta = \frac{\lambda_1}{\lambda_2}$ به دست مي‌آوريم براي اين منظور پارامتر مزاحم λ را به صورت زير تعريف مي‌كنيم

$$\gamma = \lambda_1 + \lambda_2.$$

با توجه به پارامتر مورد علاقة θ و پارامتر مزاحم λ داريم

$$\lambda_1 = \frac{\theta\gamma}{1+\theta}, \quad \lambda_2 = \frac{\gamma}{1+\theta},$$

تابع درستنمايي بر حسب پارامترهاي جديد عبارتست از

$$L(\theta,\gamma) \propto \frac{\theta^{x_1}}{(1+\theta)^{x_1+x_2}} \gamma^{x_1+x_2} e^{-\gamma}.$$

ـ روش درستنمايي كناري

فرض كنيد $X_1, \cdots, X_n$ نمونه‌اي تصادفي از توزيع نرمال با پارامترهاي مجهول μ (ميانگين) و σ^2 (واريانس) باشند، و فرض كنيد كه پارامتر مورد علاقه باشد. پارامترهاي μ و σ^2 متعامد نيستند، زيرا تابع درستنمايي آنها به صورت زير است

$$L(\mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right).$$

ـ روش درستنمايي نيمرخ

در اين روش ماكسيمم تابع درستنمايي را زماني كه مقدار پارامتر مورد توجه ثابت fixed)) در نظر گرفته شود، بدست مي‌آوريم. در اين حالت تابع درستنمايي مرتبط با پارامتر مورد علاقه را با نماد نمـايش مي‌دهيم (علامت Profile است). مثال ۵-۱ فرض كنيد متغير تصادفي X داراي توزيع نرمال با ميانگين مجهول μ و واريانس σ^2 مجهول باشد. براي استنباط شواهدي دربارة پارامتر مورد علاقة μ بر اساس نمونه‌اي تصادفي به حجم n از توزيع فوق داريم.

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

ـ روش درستنمايي شرطي

فرض كنيد متغيرهاي تصادفي مستقل X و Y داراي توزيع دوجمله‌اي به صورت زير باشند.

$$X \sim Bin(m, p_x), \qquad Y \sim Bin(n, p_y),$$

$$\psi = \frac{\frac{p_x}{1-p_x}}{\frac{p_y}{1-p_y}},$$

که m و n معلوم هستند. مي‌خواهيم استنباط شواهدي را براي نسبت بختها يعني بررسي كنيم. براي اين منظور توزيع شرطي متغير تصادفي X به شرط متغير تصادفي $X+Y$ را بدست مي‌آوريم.

$$L_C(\psi) \propto \mathrm{P}\left(X = x \mid X+Y = x+y;\ p_x, p_y\right)$$

$$\propto \frac{1}{\sum_{j=a}^{b} \binom{m}{j}\binom{n}{x+y-j}\psi^{j-x}},$$

ـ روش درستنمايي برآورد شده

$$L(\mu, \sigma^2) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$

$$L_E(\mu) \propto c(\underline{x}) s^{-n} \exp\left(-\frac{1}{2s^2}\sum(x_i-\mu)^2\right)$$

مراجع

[۱] پایان نامه دکتری مهدی عمادی دی ۱۳۸۲، دانشگاه فردوسی مشهد

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

برآورد معیار وابستگی شواهدی با استفاده از تابع چگالی مفصل

محمدی، م [۱] عمادی، م [۲] امینی، م [۳]

[۱،۲،۳] گروه آمار، دانشگاه فردوسی مشهد

چکیده

اندازه واگرایی بین دو توزیع نسبت به دور شدن دو توزیع از یکدیگر صعودی است. یکی از کاربردهای اندازه‌های واگرایی در حالت دو بعدی پی‌بردن به ساختار وابستگی داده‌ها است. هدف ما در این مقاله بیان ارتباطی جدید بین تابع مفصل و اندازه‌های واگرایی با قابلیت اندازه‌گیری شواهد آماری است. در این مقاله معیاری برای اندازه‌گیری وابستگی بین دو متغیر براساس اندازه‌های واگرایی و تابع چگالی مفصل معرفی خواهیم کرد. این معیار وابستگی مورد تائید دیدگاه شواهدی است و قدرت اندازه‌گیری شواهد آماری را دارد. برای برآورد تابع چگالی مفصل از روش هسته بتا استفاده خواهیم کرد. جهت نمایش دقت برآورد اندازه‌های وابستگی شواهدی، نتایج شبیه‌سازی ارائه می‌گردد.

کلمات کلیدی: اندازه واگرایی، وابستگی، شواهد آماری، تابع چگالی مفصل.

۱ مقدمه

یکی از مهمترین نکات مورد توجه هنگام کار با داده‌ها، بررسی وابستگی بین متغیرها است. برای بررسی وابستگی بین داده های از معیارهای مختلفی استفاده می شود. پرکاربردترین معیارهای شامل ضریب همبستگی خطی پیرسون، τـ کندال و ρـ اسپیرمن است. این معیارها قدرت اندازه‌گیری وابستگی در دم‌ها و یا وابستگی‌های ضعیف را ندارند و برآوردگرهای سازگاری نیستند (بلوم و همکاران (۲)). برای رفع این مشکلات جو (۸) اطلاع متقابل را برای اندازه‌گیری وابستگی بین متغیرها در حالت چند متغیره معرفی کرد. ما و سان (۱۰) در سال ۲۰۱۱ برای اولین بار رابطه بین اطلاع متقابل و تابع مفصل را با عنوان اندازه

[۱] mortezamohammadi408@mail.um.ac.ir

آنتروپی مفصل معرفی کردند. بلومنتریت و اسمیت (۱) این اندازه را به عنوان معیاری برای بیان وابستگی در نظر گرفتند و روش‌های ناپارامتری مختلفی را برای برآورد آن ارائه کردند. ما در این مقاله معیار جدیدی برای اندازه‌گیری وابستگی براساس اندازه‌های واگرایی متقارن و تابع چگالی مفصل که مورد تائید دیدگاه شواهدی است، معرفی می‌کنیم.

مطالعه تابع مفصل[۱] و کاربردهای آن یکی از پدیده‌های اخیر علم آمار است. تاریخچه تابع مفصل با کارهای فرشه (۶) آغاز شده است. نظریه مفصل در سال ۱۹۵۹ به‌وسیله قضیه اسکلار[۲] (۱۵) بدین‌گونه مطرح گردید که توزیع توام را می‌توان به دو بخش توزیع‌های حاشیه‌ای و تابع مفصل، که ساختار وابستگی متغیرها را نشان می‌دهد، تقسیم نمود. این قضیه بیان می‌کند که اگر X و Y متغیرهای تصادفی به ترتیب با توابع توزیع F_X و F_Y و تابع توزیع توأم F باشند، آنگاه تابع مفصلی مانند C وجود دارد به طوری‌که،

$$F(x,y) = C(F_X(x), F_Y(y)). \qquad (۱)$$

اگر F_X و F_Y پیوسته باشند، آنگاه C منحصر به فرد است، در غیر این صورت C روی $RanF \times RanG$ یکتا خواهد بود، که $RanF_X$ و $RanF_Y$ به ترتیب برد توابع توزیع حاشیه‌ای F_X و F_Y می‌باشند. بالعکس اگر C یک مفصل و F_X و F_Y توابع حاشیه‌ای باشند، آنگاه تابع F تعریف شده در رابطه (۱) یک تابع توزیع توأم با حاشیه‌ای های F_X و F_Y است.

مفصل‌های بیضوی یکی از مهمترین خانواده توابع مفصل هستند. این خانواده مفصل شامل مفصل نرمال و تی می‌باشد که در آمار، اقتصاد و بیمه کاربردهای فراوانی دارد. مزیت اصلی خانواده مفصل بیضوی این است که سطوح متفاوت همبستگی میان توزیع‌های حاشیه‌ای را به روشنی مشخص می‌کند. مفصل نرمال به صورت (۲) تعریف می‌شود:

$$\begin{aligned} C^{Gu}(u,v,\rho) &= \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \, exp\Big\{ -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \Big\} \, dx \, dy \end{aligned} \qquad (۲)$$

به‌طوری‌که Φ_ρ تابع توزیع نرمال استاندارد دو متغیره با ضریب همبستگی ρ می‌باشد.

در صورتی که X و Y دو متغیر تصادفی به ترتیب با توابع چگالی حاشیه‌ای $f_X(x)$ و $f_Y(y)$ و توابع توزیع حاشیه‌ای $U = F_X(X)$ و $V = F_Y(Y)$ و تابع مفصل C باشند، تابع چگالی مفصل بدین صورت تعریف می‌شود،

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \, \partial v},$$

یا به عبارت دیگر، اگر $f(x,y)$ تابع چگالی توأم دو متغیر تصادفی X و Y باشد، آنگاه

$$f(x,y) = c(F_X(x), F_Y(y)).f_X(x).f_Y(y).$$

کتاب نلسن (۱۳)یک منبع مطالعاتی بسیار خوب در زمینه تابع مفصل است.

[۱] Copula
[۲] Sklar

مقدار اطلاع متقابل [۳] (MI) برای دو متغیر تصادفی پیوسته X و Y با تابع چگالی توام $f(x,y)$ و به ترتیب توزیع‌های حاشیه‌ای $f_X(x)$ و $f_Y(y)$ به صورت زیر تعریف می‌شود:

$$MI(X;Y) = \int_Y \int_X f(x,y)\ \log\left(\frac{f(x,y)}{f_X(x).f_Y(y)}\right) dx\ dy.$$

اندازه واگرایی کولبک‌ـلیبلر [۴] (KL) بین دو توزیع، اولین بار توسط کولبک و لیبلر (۹) درسال ۱۹۵۱ مطرح شد. این اندازه نسبت به دور شدن دو توزیع از یکدیگر صعودی است. فاصله کولبک‌ـلیبلر جهت اندازه‌گیری فاصله میان دو تابع چگالی $f_1(x)$ و $f_2(x)$ به صورت زیر تعریف می‌شود:

$$KL(f_1|f_2) = \int_X f_1(x)\ \log\frac{f_1(x)}{f_2(x)}\ dx.$$

این فاصله برابر صفر است اگر و فقط اگر $f_1(x) = f_2(x)$.

فاصله کولبک‌ـلیبلر بین تابع چگالی توام دو متغیر تصادفی پیوسته X و Y و حاصلضرب توابع چگالی حاشیه‌ای آن‌ها برابر اطلاع متقابل است. این معیار را می‌توان به عنوان ابزاری برای اندازه‌گیری وابستگی بین متغیرهای تصادفی در نظر گرفت. اندازه‌های واگرایی متقارن جفری[۵] و هلینجر [۶] جهت اندازه‌گیری فاصله میان دو تابع چگالی $f_1(x)$ و $f_2(x)$ به ترتیب توسط روابط (۷) و (۸) تعریف می‌شود:

$$Je(f_1|f_2) = \int_X (f_1 - f_2)(log f_1(x) - log f_2(x))dx \qquad (۳)$$

$$He(f_1|f_2) = \int_X (\sqrt{f_1(x)} - \sqrt{f_2(x)})^2 dx. \qquad (۴)$$

این اندازه‌ها برابر صفر هستند اگر و فقط اگر $f_1(x) = f_2(x)$.

اندازه های واگرایی جفری و هلینجر بین تابع چگالی توام دو متغیر تصادفی پیوسته و حاصلضرب توابع چگالی حاشیه‌ای آن‌ها را به عنوان اندازه‌های وابستگی جفری و هلینجر معرفی می‌کنیم. برای برآورد اندازه وابستگی جفری و هلینجر بایستی تابع چگالی توام و توابع چگالی حاشیه‌ای را برآورد کرد. جهت سهولت افزایش دقت برآورد، این اندازه‌های وابستگی را ابتدا برحسب تابع چگالی مفصل نوشته و سپس فقط تابع چگالی مفصل را برآورد می‌کنیم.

برای برآورد تابع چگالی مفصل روش‌های مختلفی از جمله مستطیلیHistogram، نزدیکترین همسایگیN‌ـ Neighbour، earest انعکاس آینهMirror-Reflection، تبدیل پروبیت Transformation Probit و هسته بتا kernel Beta توسط چارپنتیر (۴) جمع‌آوری شده است. اخیرا ناگلر (۱۱) یک مقایسه جامع بین روش‌های معمول برآورد تابع چگالی مفصل انجام داده است و کدهای لازم را در بسته **kdecopula** (۱۲) نرم افزار R قرار داده است. در این مقاله برای برآورد تابع چگالی مفصل از روش ناپارامتری هسته بتا در حالت دو متغیره استفاده خواهیم کرد.

استنباط شواهدی اشاره به مکتب نوینی از استنباط آماری دارد که در آن استنباط صرفا براساس داده‌ها بوده و از مولفه‌های ذهنی و شخصی نظیر اطلاعات پیشین تاثیر نمی‌پذیرد. ریچارد رویال (۱۴) اصول اولیه

[۳] Mutual Information
[۴] Kullback-Leibler
[۵] Jeffrey
[۶] Hellinger

استنباط شواهدی را معرفی کرد. محور اصلی استنباط شواهدی، تابع درستنمایی است و این روش استنباط متکی بر اصل درستنمایی و قانون درستنمایی است.
هدف ما در این مقاله بیان ارتباطی جدید بین تابع مفصل و اندازه‌های واگرایی و معرفی معیاری است که قابلیت اندازه‌گیری وابستگی و شواهد آماری را دارند. جهت سهولت کار فقط حالت دو متغیره پیوسته را بررسی خواهیم نمود. در بخش دوم مقاله روش برآورد تابع چگالی مفصل ارائه می‌شود. در بخش سوم به معرفی اندازه‌های واگرایی متقارن برحسب تابع چگالی مفصل می‌پردازیم و روش برآورد آنها بیان می‌شود. در بخش چهارم نتایج شبیه سازی جهت مقایسه دقت برآوردها ارائه می‌گردد.

۲ برآورد تابع چگالی مفصل

فرض کنید $(U_j, V_j)_{j=1,\dots,n}$ یک نمونه تصادفی از مفصل دو متغیر $C(u,v)$ باشند و هدف برآورد تابع چگالی متناظر یعنی $c(u,v)$ باشد. با استفاده از روش‌های معمولی هسته می‌توان تابع چگالی مفصل را به روش زیر برآورد کرد،

$$\hat{c}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} K_{b_n}(u-U_i)K_{b_n}(v-V_i),\ (u,v)\in[0,1]^2,$$

که در آن $K_{b_n}(.) = \frac{K(.)}{b_n}$، $K(.)$ تابع هسته و b_n پهنای باند است. به طور معمول، تابع هسته یک تابع چگالی احتمال کراندار و متقارن در نظر گرفته می‌شود. یکی از مشکلات این برآوردگر اریبی مرزی است به این معنی که مقدار قابل توجهی از جرم احتمال را خارج از مربع واحد قرار می دهد. یا به عبارت دیگر ممکن است $\hat{c}_n$ یک تابع چگالی در بازه $[0,1]^2$ نباشد.

ابتدا چن (۵) در سال ۱۹۹۹ برآوردگر هسته بتا را برای تابع چگالی یک متغیره با تکیه‌گاه فشرده معرفی کرد. بوزمارنی و رولین (۳) نشان دادند که برآوردگر هسته بتا در نقاط مرزی مشکل اریبی ندارد و دارای مرتبه همگرایی $O(n^{-\frac{4}{5}})$ است. همچنین اثبات کرند که این برآوردگر برای چگالی‌های پیوسته برآوردگری سازگار است. برآوردگر هسته بتا براساس تابع چگالی بتا با پارامترهای α و β به صورت زیر تعریف می‌شود:

$$K(u,\alpha,\beta) = \frac{u^{\alpha-1}(1-u)^{\beta-1}}{B(\alpha,\beta)}\ ,\quad u\in[0,1],\quad \alpha,\beta>0$$

که در آن $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

برآوردگر هسته بتا برای برآورد تابع چگالی مفصل را به صورت ۵ معرفی می‌گردد (۴)،

$$\hat{c}_\beta(u,v) = \frac{1}{n}\sum_{i=1}^{n} K(U_i, \frac{u}{b_n}+1, \frac{1-u}{b_n}+1)K(V_i, \frac{v}{b_n}+1, \frac{1-v}{b_n}+1),\ (u,v)\in[0,1]^2. \qquad (۵)$$

با توجه به اینکه تکیه‌گاه تابع چگالی مفصل و هسته بتا یکسان هستند، مشکل اریبی مرزی در برآورد تابع چگالی مفصل برطرف خواهد شد. یکی دیگر از مزایای برآوردگر هسته بتا برای برآوردتابع چگالی مفصل، انعطاف‌پذیری و تغییر شکل در مرزها با توجه به پارامتر پهنای باند است. جهت برآورد پهنای باند بهینه از مینیم کردن میانگین جمع‌بسته توان دوم خطاها Mean Squared Integrated Error (MISE) استفاده خواهیم کرد.

۳ اندازه‌های وابستگی شواهدی برحسب تابع مفصل

محور اصلی استنباط شواهدی، تابع درستنمایی است و این روش استنباط متکی بر اصل درستنمایی و قانون درستنمایی است. اصل درستنمایی بیان می‌کند که هرگونه استنباط آماری باید بر مبنای تابع درستنمایی باشد. قانون درستنمایی بیان می‌کند که براساس نسبت درستنمایی $R(x)=\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$، میزان پشتیبانی داده‌ها از $\theta=\theta_1$ بیشتر(کمتر) از $\theta=\theta_0$ است هرگاه $R(x)>1$ $(R(x)<1)$ و میزان پشتیبانی داده‌ها از دو مقدار پارامتر یکسان است هرگاه $R(x)=1$.

نسبت درستنمایی $R(x)$، میزان شواهد آماری در پشتیبانی $\theta=\theta_1$ در مقابل $\theta=\theta_0$ را اندازه می‌گیرد(۱۴). حال فرض کنید آزمایشگری بخواهد بین دو آزمایش با هزینه‌های تقریبا یکسان، آزمایشی را که دارای شواهد آماری بالقوه بیشتری باشد، برای کسب داده‌های خود انتخاب کند. در نظر بگیرید x مشاهدات مربوط به آزمایش E از بردار تصادفی X با چگالی $f_\theta(x)$ باشد. حبیبی و همکارن (۷) رابطه (۶) را به عنوان معیاری برای اندازه‌گیری شواهد آماری بالقوه در آزمایش E معرفی کردند.

$$S_\psi(E)=E_{\theta_1}\Big[\psi(R_E(X))\Big]+E_{\theta_0}\Big[\psi(\frac{1}{R_E(X)})\Big], \qquad (۶)$$

که در آن ψ تابعی صعودی است. اگر $\psi(t)=log(t)$، آنگاه

$$\begin{aligned} S_1(E) &= E_{\theta_1}\Big[log(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)})\Big]+E_{\theta_0}\Big[log(\frac{f_{\theta_0}(X)}{f_{\theta_1}(X)})\Big] \\ &= KL(f_{\theta_1},f_{\theta_0})+KL(f_{\theta_0},f_{\theta_1}) \\ &= Je(f_{\theta_1},f_{\theta_0}) \end{aligned}$$

که در آن $KL(f_{\theta_1},f_{\theta_0})$ و $Je(f_{\theta_1},f_{\theta_0})$ به ترتیب فاصله کولبک‌ـلیبلر و فاصله جفری بین دو توزیع f_{θ_1} و f_{θ_0} است.

عبارت $S_\psi(E)$ در رابطه (۶) یک معیار متقارن است، یعنی با تعویض دو تابع f_{θ_1} و f_{θ_0} مقدار $S_\psi(E)$ تغییر نمی‌کند. لذا برای اینکه یک اندازه واگرایی بتواند شواهد آماری را اندازه‌گیری کند، بایستی متقارن باشد. بنابراین اندازه واگرایی جفری قابلیت اندازه‌گیری شواهد آماری موجود در یک آزمایش را دارند، در حالی اندازه واگرایی کولبک‌ـلیبلر این قابلیت را ندارد.

با در نظر گرفتن $\psi(t)=1-\frac{1}{\sqrt{t}}$، برای $t>0$ می‌توان نوشت،

$$\begin{aligned} S_2(E) &= E_{\theta_1}\Big[1-\frac{1}{\sqrt{\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}}}\Big]+E_{\theta_0}\Big[1-\frac{1}{\sqrt{\frac{f_{\theta_0}(X)}{f_{\theta_1}(X)}}}\Big] \\ &= \int_\chi \Big[(f_{\theta_0}-\sqrt{f_{\theta_0}f_{\theta_1}})+(f_{\theta_1}-\sqrt{f_{\theta_0}f_{\theta_1}})\Big]d\mu \\ &= D_H(f_{\theta_1},f_{\theta_0}) \end{aligned}$$

یعنی اندازه واگرایی هلینجر می‌تواند شواهد آماری موجود در یک آزمایش را اندازه‌گیری کند.

اندازه‌های وابستگی جفری و هلینجر را با استفاده از تابع چگالی مفصل می‌توان به صورت زیر نوشت،

$$Je^c(u,v) \equiv Je(f|f_1 f_2) = \int_Y \int_X (f(x,y) - f_1(x)f_2(y)) log\Big(\frac{f(x,y)}{f_1(x)f_2(y)}\Big)dxdy$$
$$= \int_0^1 \int_0^1 (c(u,v) - 1)\ log(c(u,v))dudv, \quad (7)$$
$$He^c(u,v) \equiv He(f|f_1 f_2) = \int_Y \int_X (\sqrt{f(x,y)} - \sqrt{f_1(x)f_2(y)})^2 dxdy$$
$$= \int_0^1 \int_0^1 (\sqrt{c(u,v)} - 1)^2 dudv. \quad (8)$$

حال اندازه‌های وابستگی جفری و هلینجر را برای توزیع نرمال دومتغیره بدست می‌آوریم و در بخش بعد دقت برآورد این اندازه‌های وابستگی را با استفاده از روش هسته بتا گزارش می‌کنیم. فرض کنید $(X,Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho^2)$. با توجه به تعریف مفصل نرمال(۲)، به راحتی می‌توان نتایج زیر را به دست آورد.

$$Je^c(u,v) = -\log(1-\rho^2), \quad (9)$$
$$He^c(u,v) = 2\Big(1 - 2\frac{(1-\rho^2)^{\frac{1}{2}}}{(4-\rho^2)^{\frac{1}{2}}}\Big) \quad (10)$$

همانطور که در روابط (۹) و (۱۰) مشاهده می‌کنیم، مقادیر اندازه‌های وابستگی جفری و هلینجر نسبت به ρ در بازه $[0,1]$صعودی است. در صورتی‌که $\rho = 0$ باشد، اندازه‌های وابستگی برابر صفر هستند. در نتیجه اندازه‌های وابستگی می‌توانند ساختار وابستگی بین دو متغیر تصادفی را بیان کنند.

نکته ۳.۱. اندازه‌های وابستگی عددی در بازه $[0,+\infty)$ هستند. می‌توان با اعمال تبدیل $T : [0,\infty] \longrightarrow [0,1]$ اندازه‌های وابستگی استاندارد را تعریف کرد. تبدیل T را به صورت زیر در نظر می‌گیریم:

$$T(x) = \sqrt{1 - e^{-x}} \quad (11)$$

همانطور که قبلا اشاره کردیم، برای برآورد اندازه وابستگی جفری و هلینجر که بعد از این اندازه‌های وابستگی شواهدی نامیده می‌شوند، بایستی تابع چگالی توام و توابع چگالی حاشیه‌ای را برآورد کرد. جهت افزایش دقت برآورد و سهولت بیشتر، ابتدا این اندازه‌های وابستگی را ابتدا برحسب تابع چگالی مفصل نوشته و سپس فقط تابع چگالی مفصل را برآورد می‌کنیم.

۴ نتایج عددی

در این بخش به کمک شبیه‌سازی مونت‌ـ کارلو، بسته نرم افزاری تابع مفصل در نرم‌افزار R و مراحلی که توسط نویسندگان مقاله ارائه می‌گردد، می‌توان اندازه‌های وابستگی شواهدی را برآورد کرد. فرض کنید نمونه‌ای تصادفی به حجم n از دو متغیر تصادفی دلخواه X و Y در اختیار داریم. برای برآورد اندازه‌های وابستگی شواهدی آنها به ترتیب مراحل زیر را اجرا می‌کنیم:

۱) مقادیر $\hat{U}_i = \frac{R_i}{n+1}$ و $\hat{V}_i = \frac{S_i}{n+1}$ رابه عنوان برآورد رتبه‌ای تابع توزیع تجمعی X و Y در نظر می‌گیریم که در آن R_i و S_i به ترتیب رتبه‌های مربوط به X_i و Y_i هستند.

۲) تابع چگالی مفصل $(U_i, V_i)_{i=۱,\cdots,n}$را به روش ناپارامتری هسته بتا برآورد می‌کنیم.

۳) به کمک شبیه‌سازی مونت‌ـ کارلو اندازه‌های وابستگی شواهدی جفری و هلینجر را بدین صورت برآورد می‌کنیم:

$$\hat{Je}^{c}(u,v) = \frac{۱}{n}\sum_{i=۱}^{n}\left(۱ - \frac{۱}{\hat{c}_{\beta}(\hat{u}_i,\hat{v}_i)}\right) log(\hat{c}_{\beta}(\hat{u}_i,\hat{v}_i))$$

$$\hat{He}^{c}(u,v) = \frac{۱}{n}\sum_{i=۱}^{n}\left(۱ - \frac{۱}{\sqrt{\hat{c}_{\beta}(\hat{u}_i,\hat{v}_i)}}\right)^{۲}.$$

که در آن n حجم نمونه است.

فرض کنید (X,Y) دارای توزیع نرمال دومتغیره استاندارد با ضریب همبستگی ρ باشند. می‌دانیم که به ازای مقادیر ۰، ۰/۲، ۰/۵ و ۰/۹ برای ρ به ترتیب مقادیر دقیق اندازه وابستگی جفری برابر با ۰، ۰/۰۲۰۴، ۰/۱۴۳۸ و ۰/۸۳۰۴ و مقادیر دقیق اندازه وابستگی هلینجر برابر با ۰، ۰/۰۱۰۳، ۰/۰۷۷۷ و ۰/۵۲۱۴ است. مشاهده می‌کنیم با افزایش شدت وابستگی(ρ) مقادیر اندازه‌های وابستگی افزایش می‌یابد. جهت مقایسه دقت برآورد اندازه‌های وابستگی جفری و هلینجر مقادیر مجذور میانگین توان‌دوم خطاها Mean Root Error Square (RMSE)

و میانگین قدرمطلق انحرافات Mean Absolute Deviation (MAD) را برای تابع مفصل نرمال تحت مقادیر مختلف حجم نمونه (n) و پارامتر ρ بدست می‌آوریم. نتایج در جداول ۱ و ۲ برای ۱۰۰۰ مرتبه تکرار شبیه‌سازی ارائه گردیده است.

جدول ۱: مقادیر $RMSE$ اندازه‌های واگرایی شواهدی برای مقادیر مختلف حجم نمونه (n) و پارامتر شدت وابستگی (ρ)

ρ	اندازه وابستگی	n				
		۱۰	۲۰	۵۰	۱۰۰	۲۰۰
۰	Je	۰,۲۰۶۹	۰,۰۴۸۲	۰,۰۲۲۶	۰,۰۱۶۱	۰,۰۱۱۹
	He	۰,۰۴۸۱	۰,۰۲۲۱	۰,۰۱۱۷	۰,۰۰۷۹	۰,۰۰۵۹
۰,۲	Je	۰,۱۳۷۵	۰,۰۵۶۰	۰,۰۲۷۰	۰,۰۱۶۹	۰,۰۱۲۹
	He	۰,۰۵۴۸	۰,۰۲۸۴	۰,۰۱۳۴	۰,۰۰۸۴	۰,۰۰۵۹
۰,۵	Je	۰,۱۷۹۸	۰,۰۹۵۵	۰,۰۶۹۶	۰,۰۵۴۵	۰,۰۴۳۵
	He	۰,۰۶۹۱	۰,۰۵۲۱	۰,۰۳۸۱	۰,۰۳۱۳	۰,۰۲۵۰
۰,۹	Je	۰,۳۷۹۸	۰,۳۱۶۷	۰,۲۶۷۷	۰,۲۴۷۰	۰,۲۲۳۴
	He	۰,۲۷۷۱	۰,۲۶۵۵	۰,۲۴۶۳	۰,۲۳۸۷	۰,۲۲۹۷

نتایج جدول ۱ و ۲ نشان می‌دهد که با افزایش حجم نمونه مقادیر RMSE و MAD کاهش یافته و در نتیجه میزان دقت برآوردها افزایش می‌یابد. همچنین نتایج نشان می‌دهد که دقت برآوردها در وابستگی‌های ضعیف بیشتر از وابستگی های قوی است. هچمنین نتایج نشان می‌دهد که با افزایش شدت وابستگی (پارمتر

جدول ۲: مقادیر MAD اندازه‌های واگرایی شواهدی برای مقادیر مختلف حجم نمونه (n) و پارامتر شدت وابستگی (ρ)

ρ	اندازه وابستگی	n				
		۱۰	۲۰	۵۰	۱۰۰	۲۰۰
۰	Je	۰,۰۷۴۷	۰,۰۳۲۰	۰,۰۲۰۳	۰,۰۱۳۹	۰,۰۱۰۴
	He	۰,۰۲۶۴	۰,۰۱۶۶	۰,۰۰۹۵	۰,۰۰۷۱	۰,۰۰۵۳
۰,۲	Je	۰,۰۷۳۳	۰,۰۲۷۵	۰,۰۱۶۴	۰,۰۱۱۶	۰,۰۰۹۳
	He	۰,۰۴۰۴	۰,۰۱۸۷	۰,۰۰۸۰	۰,۰۰۶۰	۰,۰۰۴۶
۰,۵	Je	۰,۱۳۷۳	۰,۰۸۳۰	۰,۰۶۰۳	۰,۰۴۷۶	۰,۰۳۴۸
	He	۰,۰۵۷۶	۰,۰۴۶۰	۰,۰۳۰۷	۰,۰۲۵۸	۰,۰۲۱۱
۰,۹	Je	۰,۳۳۶۱	۰,۲۹۲۷	۰,۲۴۵۰	۰,۲۲۹۱	۰,۲۱۱۹
	He	۰,۲۵۷۳	۰,۲۴۴۵	۰,۲۳۱۰	۰,۲۱۶۷	۰,۱۹۷۷

ρ)، دقت هر دو برآورد اندازه‌های وابستگی افزایش می‌یابد. با مقایسه مقادیر RMSE و MAD در شدت‌های وابستگی مختلف و حجم نمونه‌های متفاوت می‌توان به این نتیجه رسید که دقت برآورد اندازه وابستگی شواهدی هلینجر بیشتر از جفری است.

۵ بحث و نتیجه‌گیری

در این مقاله ارتباطی بین تابع مفصل و اندازه واگرایی جفری و هلینجر که قابلیت اندازه‌گیری شواهد آماری را دارد، بیان کردیم. همچنین یک روش جدید برای برآورد اندازه واگرایی جفری و هلینجر بر اساس برآورد تابع چگالی مفصل به روش هسته بتا در بخش نتایج عددی معرفی نمودیم. روش ارائه شده در این مقاله برای یافتن ساختار وابستگی ضعیف بسیار دقیق‌تر از روش‌های معمول همبستگی از قبیل همبستگی پیرسون است و محدود به بررسی روابط خطی نمی‌باشد. نتایج مقایسه اندازه‌های وابستگی شواهدی جفری و هلینجر نشان داد که در شدت‌های وابستگی مختلف و حجم نمونه‌های متفاوت، دقت برآورد اندازه وابستگی شواهدی هلینجر بیشتر از جفری است.

مراجع

[1] Blumentritt, T., & Schmid, F. (2012). *Mutual information as a measure of multivariate association: analytical properties and statistical estimation.* Journal of Statistical Computation and Simulation, 82(9), 1257-1274.

[2] Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). *Distribution free tests of independence based on the sample distribution function*. The Annals of mathematical statistics, 485-498.

[3] Bouezmarni, T., & Rolin, J. M. (2003). *Consistency of the beta kernel density function estimator*. Canadian Journal of Statistics, 31(1), 89-98.

[4] Charpentier, A., Fermanian, J., & Scaillet, O. (2006). *Copulas: From theory to application in finance, chapter The Estimation of Copulas: Theory and Practice*. Risk Books, Torquay, UK, 1, 35-62.

[5] Chen, S. X. (1999). *Beta kernel estimators for density functions*. Computational Statistics & Data Analysis, 31(2), 131-145.

[6] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. Ann. Univ. Lyon, $3^e serie, Sciences, Sect.A, 14, 53-77$.

[7] Habibi, A., Arghami, N. R., Ahmadi, J. (2006). Statistical evidence in experiments and in record values. Communications in Statistics-Theory and Methods, 35(11), 1971-1983.

[8] Joe, H. (1989). *Relative entropy measures of multivariate dependence*. Journal of the American Statistical Association, 84(405), 157-164.

[9] Kullback, S., & Leibler, R. A. (1951). *On information and sufficiency*. The Annals of mathematical statistics, 22(1), 79-86.

[10] Ma, J., Sun, Z. (2011). Mutual information is copula entropy. Tsinghua Science & Technology, 16(1), 51-54.

[11] Nagler, T. (2018). *kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities*. Journal of Statistical Software 84(7), 1-22.

[12] Nagler, T., Wen, K., & Nagler, M. T. (2018). *kdecopula: Kernel Smoothing for Bivariate Copula Densities*. R package version 0.9.2, URL https://CRAN.R-project.org/package=kdecopula.

[13] Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

[14] Royall, R. (1997). Statistical Evidence: A Likelihood Paradigm(Vol. 71). Chapman & Hall.

[15] Sklar, M. (1959). *Fonctions de repartition an dimensions et leurs marges*. Publ. inst. statist. univ. Paris, 8, 229-231.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

انعکاس‌پذیری توان‌هایی از عملگر ضربی روی فضاهای توابع خاص

باقری نژاد، س. ح [۱] ایلون کشکولی، ع [۲]

[۱،۲] دانشکده ریاضی، دانشگاه یاسوج.

چکیده

در این مقاله به بررسی و استخراج شرایط کافی برای انعکاس پذیری هر توانی از عملگر ضربی روی فضاهای با ناخ سری های معمول لوران پرداخته شده است.

کلمات کلیدی: عملگر ضربی، فضای با ناخ وابسته به سری های لارنت β، عملگر انعکاسی.

۱ پیش‌گفتار

تعریف ۱.۱. اگر E فضای با ناخ جدایی‌پذیر و $A \in B(E)$، آنگاه کوچکترین زیرجبر $B(E)$ که شامل A و همانی I بوده و با T. W (توپولوژی عملگر ضعیف) بسته است، را با نماد $W(A)$ نمایش می‌دهیم.

تعریف ۲.۱. اگر E فضای با ناخ جدایی‌پذیر و $A \in B(E)$، آنگاه یک عملگر $A \in B(E)$ را انعکاسی گوییم هرگاه: $A \lg lat(a) = W(A)$.
به عبارت دیگر دقیقا انعکاسی است وقتی به اندازه کافی زیرفضای پایا داشته باشد که $W(A)$ را مشخص می‌کند. برای شناخت بیشتر به (۵) مراجعه کنید.

تذکر ۱.۱. اگر $A \in B(E)$ ، آنگاه $lat(A)$ عبارت است از شبکه تمام زیرفضاهای پایای A و $A \lg lat(A)$ عبارت است از جبر تمام عملگرهای B در E به طوریکه $lat(A) \subset lat(B)$.

گزاره ۱.۱. فرض کنید $T_۱ = z \in C : |z| = r_۱ ۱ - \epsilon_۱ \subset \omega_۱$ آنگاه؛

[۱] Parizan2010@yahoo.com

$L_۱ = f \in L^(\rho)(\beta) : \int_{T_۱} Z^n f(z)dz = ۰, n = ۱, ۲, ...$،

۱. زیرفضایی از $L^(\rho)(\beta)$ است.

۲. در $L^(\rho)(\beta)$ بسته است.

۳. تحت M_z پایا بوده و شامل ثابت‌ها نیز هست. برای اثبات رجوع کنید به (۶).

لم ۱.۱. اگر $\varphi \in H(\omega_۱ ۱) \bigcap l^\rho_\infty(\beta)$، آنگاه دنباله‌ای از چندجمله‌ای های r_n وجود دارند به طوریکه؛

$$M_{r_n} \xrightarrow{W.T} M_\varphi$$

برای اثبات رجوع کنید به (۳).

نتیجه ۱.۱. نتیجه‌ی لم ۱.۱ :
اگر $\varphi \in H(\omega_۱ ۱) \bigcap l^\rho_\infty(\beta)$، آنگاه؛ $M_\varphi \in W(M_z)$

برهان. فرض کنید؛

$\varphi \in H(\omega_۱ ۱) \bigcap l^\rho_\infty(\beta)$.

در این صورت بنابر لم ۱.۱ ؛

$M_{r_n} \xrightarrow{W.T} M_\varphi$.

که در آن $r_n = \rho_n(\varphi)$. چون r_n یک چندجمله‌ای است و $M_{r_n} = r_n(M_x)$ نتیجه می‌گیریم که؛ $M_\varphi \in W(M_z)$.

□

قضیه ۱.۱. برای هر $K \geqslant ۱$، عمگر M_{z^K} روی $l^\rho(\beta)$ انعکاسی است.

برهان. اولا می‌دانیم که $\omega_۱$ ناتهی است، چون $r_۱ ۱ > r_۰ ۱$. مقادیر ثابت و مثبت $\epsilon_۱$ و $\epsilon_۰$ را چنان انتخاب می‌کنیم که دایره‌های:

$T_۰ = z \in C : |z| = r_۰ ۱ + \epsilon_۰$،

$T_۱ = z \in C : |z| = r_۱ ۱ + \epsilon_۱$،

درون $\omega_۱$ واقع شده و هیچ نقطه‌ی مشترکی نداشته باشند. طبق گزاره ۱.۱ ؛

$L_۱ = f \in l^\rho(\beta) : \int_{T_۱} Z^n f(z)dz = ۰, n = ۱, ۲, ...$،

زیرفضایی بسته در $l^\rho(\beta)$ بوده و تحت M_z پایا است.
به ازای $K \in N$، می‌دانیم که؛

$W(M_{z^K}) \subset A\lg lat(M_{z^k})$

حال فرض کنید $A \in A\lg lat(M_{z^k})$. چون؛

$$lat(M_z) \subset A\lg lat(M_{z^k}),$$

پس؛

$$lat(M_z) \subset lat(A)$$

و در نتیجه؛

$$A \in A\lg lat(M_z).$$

با توجه به آنکه به ازای هر $\lambda \in \Omega_1$؛

$$M_z^* e(\lambda) = \overline{\lambda} e(\lambda),$$

بنابراین به ازای $f \in l^\rho(\beta)$ داریم؛

$$\begin{aligned} < f, M_\varphi^* e_\lambda > =< M_\varphi f, e_\lambda > &=< \varphi f, e_\lambda > \\ &= e_\lambda(\varphi f) \\ &= e_\lambda(\varphi) e_\lambda(f) \\ &= \varphi(\lambda) f(\lambda) \\ &= \varphi(\lambda) < f, e_\lambda > \\ &=< f, \overline{\varphi}(\lambda) e_\lambda >, \end{aligned}$$

پس؛

$$M^*_{\varphi^e(\lambda)} = \overline{\varphi}(\lambda) e_\lambda.$$

در این صورت پیمای یک‌بعدی $e(\lambda)$ تحت M_z^* و در نتیجه تحت A^* پایا بوده و برای هر $\lambda \in \Omega_1$ می‌توان نوشت؛

$$A^* e(\lambda) = \overline{\varphi}(\lambda) e_\lambda,$$

و لذا:

$$< Af, e(\lambda) >=< f, A^* e(\lambda) >= \varphi(\lambda) e(\lambda), \forall \lambda in \omega_1, f \in L^\rho(\beta),$$

و این نتیجه می‌دهد که؛

$$A = M_\varphi, \varphi \in L^\rho_\infty(\beta),$$

بنابراین؛

$.\varphi \in H^{\infty}(\omega_1)$

چون $L_1 \in lat(M_z)$، داریم $AL_1 \subset L_1$، لذا

$.A_1 = \varphi \in L_1$

با بکارگیری فرمول انتگرال کوشی می‌توان نوشت؛

$،\varphi = \varphi_0 + \varphi_1$

به طوری که؛

$،\varphi_0 \in H_0(\omega_0 1), \varphi_1 \in H(\omega_1 1)$

و منظور از $H_0(\omega_0 1)$ فضای تمام توابعی در $H(\omega_0 1)$ است که در ∞ به صفر می‌گراید. اکنون دایره‌ی T_2 را چنان مناسب (بسته) انتخاب می‌کنیم تا دایره T_1 با کوچکترین شعا در ناحیه $ext(T_2)$ قرار گیرد. بنابراین φ_0 در $ext(T_2)$ تحلیلی است. حال می‌توانیم بنویسیم؛

$$،\varphi_0(z) = \sum_{n=-\infty}^{-1} \widehat{\rho}_0(n) z^n, Z \in ext(T_2)$$

که در آن:

$$،\widehat{\varphi}_0(n) = \frac{1}{2\pi i} \int_{T_1} \frac{\varphi_0(z)}{Z^{n+1}} !zmn < 0$$

و چون $\varphi_1 \in H(\omega_1 1)$، داریم؛

$$.\int_{T_1} z^n \varphi_1(z) dz = 0, n = 0, 1, 2, ...$$

اما از آنجا که $\varphi \in L_1$، برای آنکه تساوی

$$\varphi = \varphi_0 + \varphi_1$$

برقرار باشد، باید داشته باشیم؛

$$.\int_{T_1} z^n \varphi_0(z) dz = 0, n = 0, 1, 2, ...$$

از این عبارت نتیجه می‌شود که به ازای هر عدد صحیح منفی n، $\widehat{\varphi}_0(n) = 0$ بنابراین:

$$،\varphi_0(z) = 0, z \in ext(T_2)$$

از اینکه $\varphi_0 \equiv 0$ نتیجه می‌شود؛

$$.\varphi = \varphi_0 \in H(\omega_1 1)$$

بنابراین؛

$$،\varphi \in H(\omega_1 1) \cap L^{\rho}_{\infty}(\beta)$$

و لذا با توجه به لم ۱.۱، دنباله‌ای از چندجمله‌ای‌های r_n وجود دارند که؛

$$.M_{r_n} \xrightarrow{W.T} M_{\varphi}$$

اکنون فرض کنید S_K پیمای خطی بسته‌ای از مجموعه‌ی $\{f_{nk} : n \geqslant ۰\}$ باشد. در این صورت خواهیم داشت؛

$$.M_{z^k} f_{nk} = f_{(n+۱)k} \in S_k, \forall n \geqslant ۰$$

بنابراین؛

$$،S_k \in lat(M_{z^K})$$

و در نتیجه؛

$$.S_k \in lat(M_\varphi)$$

اکنون فرض کنید؛

$$.\varphi(z) = \sum_{n=۰}^{\infty} \hat{\varphi}(n) z^n$$

چون $l \in S_k$، بنابراین؛

$$،M_\varphi l = \varphi \in S_k$$

پس به ازای $n \geqslant ۰$ و هر $i \neq n_k$ داریم؛

$$.\hat{\varphi}(i) = ۰$$

حال با توجه به نتیجه‌ای از ساختار ویژه‌ی r_n در لم ۱.۱، هر r_n باید چندجمله‌ای به صورت Z^k باشد. به عبارت دیگر؛

$$،r_n(z) = \varphi_n(z^k)$$

به ازای هر چندجمله‌ای φ_n.
بنابراین؛

$$.M_{r_n} = r_n(M_z) = \varphi_n(M_{z^K}) \xrightarrow{W.T} A$$

و در نهایت بنابر نتیجه‌ی ۱.۱ داریم، $A \in W(M_{z^k})$. پس M_{z^k} انعکاسی بوده و لذا اثبات تمام است. □

۲ دست‌آوردهای پژوهش

۱ ـ قضیه‌ی ۱.۱ صورت قضیه‌ی (۲,۱) در (۴) را اصلاح می‌کند و بیان می‌دارد که رابطه‌ی $\| M_s \| \leqslant C \| S \|$ در این قضیه ضروری نیست. ۲. در قضیه‌ی ۱.۱ چنانچه علاوه بر فرض $r_{۱} < r_{۱}$.۱ عملگر M_z نیز روی $L^p(\beta)$ معکوس‌پذیر باشد، آنگاه: M_{z^K} به ازای هر عدد صحیح K، انعکاسی است. برای اثبات رجوع کنید به (۲).

مراجع

[۱] جان بی کان وی، مینا اتفاق، سینا اعتماد، *مقدمه‌ای بر آنالیز تابعی*، انتشارات آرنا، (۱۳۹۲).

[۲] والتر رو دین، علی‌اکبر عالم‌زاده، *آنالیز تابعی*، انتشارات میقان، (۱۳۸۹).

[3] B. Yousefi, On the Space $L_p(\beta)$, *rendiconti del circolo matematico di palermo*(in Italic)**Vol. 49** (2000), 115-120.

[4] Author's name(s), On the Eighteens Questions of Allen Shields , *International Journal of Mathematics*(in Italic) **Vol. 16** (2005), 1-6.

[5] H. Radjavi and P. Rosenthal , *Invariant Subspaces*(in Italic), Springer, 1973

[6] T. W. Gamelin , *Uniform Algebras*(in Italic), AMS Chelsea, 1984

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

مروری بر مفهوم فرض و فرضیه

شرفی، م [۱] قهرمانی، م [۲]

[۱]گروه آمار، دانشگاه رازی
[۲]گروه آمار، دانشگاه پیام نور، تهران

چکیده

تاکنون اختلاف نظرها و بحثهای زیادی در مورد استفاده و کاربرد دو واژه فرض و فرضیه در بین محققان آماری وجود داشته است. فرض ایدهای است که انسان از چیزهای اطراف برای اثبات درستی یا نادرستی چیزی برداشت میکند. اما فرضیه یک پرسش جدید است که بنابر دلایل موجود به ذهن انسان (محقق) خطور میکند. در این مقاله سعی بر تشریح و تفسیر دو واژه فرض و فرضیه در آمار شده است.

کلمات کلیدی: فرض، فرضیه، نظریه.

۱ پیش‌گفتار

برای روشن شدن مفهوم فرضیه از مثال‌های ساده و روزمره شروع می‌کنیم. فرض کنید برق خانه‌ای قطع شده و صاحبخانه با مشکل روبرو گردیده است. مسئله او این است که علت قطع برق چیست؟ در عالم تفکر و با استفاده از معلومات کلی و شناخت‌های قبلی چند حدس یا گمان به ذهن او خطور می‌کند؛ مثلاً امکان دارد فیوز کنتور دچار مشکل شده باشد (پریدن یا سوختن). امکان دارد عیب از اتصال سیم‌های برق باشد. امکان دارد از کارخانه برق قطع شده باشد و نظایر آن. اینها همه تصورات ذهنی هستند که برای او بوجود می‌آید و منشأ آنها نیز معلومات قبلی و قضایای کلی هستند که نسبت به آنها آگاهی دارد. تمام این تصوّرات ذهنی که برای او بوجود می‌آید، در واقع فرض‌هایی هستند که در ذهن او نقش می‌بندند و ذهن او را به سمت

[۱] m.sharafi@razi.ac.ir
[۲] m.ghahramani.stat@gmail.com

آن جهت می‌دهند تا تلاش کاوشگرانهٔ خودش را برای حل مسئله قطع برق در راههای معدودی به کار گیرد. او بلافاصله شروع به کاوش و تحقیق در مورد هر یک از آنها می‌کند و ابتدا سهل‌ترین و محتمل‌ترین راه را انتخاب و آزمایش می‌کند؛ مثلاً ممکن است ابتدا به سراغ کنتور برود و با مشاهدهٔ آن از وضعیت فیوز آگاهی یابد. اگر مشکل از آن باشد اقدام به رفع اشکال نموده جریان برق را برقرار می‌کند، در این صورت مسئله حل می‌شود. ولی اگر قطع برق از فیوز نبود، به سراغ احتمال بعدی می‌رود. اگر این بار نیز به نتیجه نرسید، به سراغ موارد بعدی می‌رود و یکی پس از دیگری آنها را آزمایش می‌کند تا راه‌حل پیدا شود. این مثال ساده حامل پیام‌های خوبی برای کسی است که می‌خواهد روش تحقیق علمی را یاد بگیرد؛ زیرا به زبان ساده او را با مفهوم مسئله، فرضیه‌سازي، روش جمع‌آوری اطلاعات و آزمایش فرضیه و نتیجه‌گیری آشنا می‌کند و نقش فرضیه را در وصول به هدف تحقیق واضح می‌نماید. در تعریف فرضیه می‌توان گفت: فرضیه عبارت است از حدس یا گمان اندیشمندانه درباره ماهیت، چگونگی و روابط بین پدیده‌ها، اشیاء و متغیرها، که محقق را در تشخیص نزدیکترین و محتمل‌ترین راه برای کشف مجهول کمک می‌نماید؛ بنابراین، فرضیه گمانی است موقتی که درست بودن یا نبودنش باید مورد آزمایش قرار گیرد. فرضیه براساس معلومات کلی و شناخت‌های قبلی یا تجارب محقق پدید می‌آید. این شناخت‌ها ممکن است براساس تجارب یا مطالعات قبلی باشد، از منابع شفاهی بدست آمده باشد، یا در جریان مطالعهٔ ادبیات تحقیق حاصل شده باشد.

۲ مفهوم فرض و فرضیه

فرض ایده‌ای است که انسان از چیزهای اطراف برای اثبات درستی یا نادرستی چیزی برداشت می‌کند. اما فرضیه یک پرسش جدید است که بنابر دلایل موجود به ذهن انسان (محقق) خطور می‌کند. درواقع فرضیه چیزی است که تعدادی فرض به همراه دارد، یعنی تا فرض نباشه فرضیه هم داده مطرح نمی‌شود. فرضیه یا انگاره یک توضیح پیشنهادی برای یک پدیده یا رخداد است؛ به گونه‌ای دیگر باید گفت که فرضیه یک حدس منطقی و آزمایش‌پذیر مبتنی بر پدیده‌هایی است که در جهانِ طبیعت مشاهده می‌کنید. در تعریفی دیگر، فرضیه، به فرضی گفته می‌شود که به عنوان یک توضیح قابل آزمایش مطرح می‌شود و پایهٔ تحقیقات بعدی را تشکیل می‌دهد. معمولاً تشکیل یک فرضیه، نخستین گام در حل مسئله و شرح یک پدیده است. یک فرضیهٔ علمی چه اثبات شده باشد و چه نباشد، نباید عنوان شود که تنها یک فرضیه است زیرا فرضیه‌های علمی پایه و اساس روش‌های علمی هستند.

برای درکی بهتر از فرضیهٔ علمی، باید فهمید که عملکرد روش علمی چیست و چگونه کار می‌کند: پس از تولید مشاهدات گوناگون در ارتباط با یک پدیده طبیعی و فرمول‌بندی و استخراج یک سؤال دربارهٔ چگونگی (و نه چیستی) آن پدیده، دانشمندان باید بتوانند فرضیه‌ای برای آن تولید کنند که به‌صورت بالقوه، پاسخی برای آن سؤال و مشاهدات باشد. سپس آنها پیش‌بینی‌های آزمایش‌پذیری را از داخل آن فرضیه که پاسخ احتمالیِ چگونگی آن پدیده است استخراج کرده و آن را بارها و بارها آزمایش می‌کنند و داده‌ها را تحلیل و آنالیز کرده و می‌سنجند. پس از انجام این مراحل، آنها در نهایت می‌توانند اعلام کنند که آن فرضیه درست است یا نادرست. حتی پس از آن نیز آن فرضیه نیاز دارد که بارها و بارها آزمایش شود و آزمایشات مکرر توسط دانشمندان مختلف روی آن انجام شود تا بتواند بصورت عمومی توسط مجامع علمی به عنوان یک

فرضیهٔ قابل قبول پذیرفته شود. شاید یک مثال بتواند شرحی کامل‌تر از فرضیهٔ علمی ارائه دهد. تصور کنید که هر روز که از خواب بیدار می‌شوید، مشاهده می‌کنید که زباله‌های شما سرنگون شده و ناخواسته در اطراف حیاط پخش شده‌اند. شما فرضیه‌ای می‌سازید که گربه‌ها مسئول این پدیده هستند. برای آزمایش این فرضیه، شاید نیاز باشد که شما یک شب یا چندین شب را بیدار بمانید تا گربه‌ها را زیر نظر بگیرید تا در نهایت نتیجه‌گیری کنید که آیا فرضیهٔ شما درست است یا نادرست. اما شاید با وجود بیدار ماندن شما و مشاهده گربه‌ها طی هزار شب، در شب هزار و یکم توفان عامل این پدیده باشد. یا ممکن است عامل این پدیده تنها در خانهٔ شما گربه‌ها باشند. پس نیاز به مشاهده‌ها و آزمایش‌های گوناگون فارغ از زمان و مکان مشخص و توسط افراد مختلف است تا صحت یک فرضیه تایید یا رد شود. مثال فوق نشان می‌دهد که چرا فرضیات مبتنی بر شبه علم، فرضیات علمی محسوب نمی‌شوند (و قطعاً نظریه‌های علمی هم نیستند). زیرا چیزی برای مشاهده آنها وجود ندارد. و همچنین چیزی برای آزمایش هم وجود ندارد. بطور مثال عقیده وجود خدا یا جاودانگی روح ، خارج از محدودهٔ طبیعی و در نتیجه خارج از محدوده علم است (۷). فرضیه معمولا با اگر و آنگاه بیان می شود. مثل : اگر برای گاو ها موسیقی پخش کنیم آنگاه شیر بیشتری می دهند. در تفاوت پیش فرض و فرضیه های تحقیق در پایان نامه نویسی باید گفت که پیش فرض ها شرایط را برای سوال ها و فرضیه ها مهیا می کنند. در این بخش با پیش فرض ها یا مفروضات و نیز فرضیه های تحقیق آشنا می شوید. پیش فرض‌ها یا مفروضات، مجموعه ای از اصول مسلم ، قضایای بنیادین و گزاره‌هایی هستند که به طور عملیاتی در راستای اهداف تحقیق، مورد پذیرش قرار گرفته‌اند. مفروضات تحقیق، ماهیت داده‌ها، تجزیه و تحلیل و تعبیر و تفسیر آنها را تحت تأثیر قرار می‌دهد. برای مثال، «این نمونه به نوعی مربوط به کل جامعه آماری است» (ماهیت)؛ «این معیار برای استفاده در این یافته‌ها، معتبر و با ارزش تلقی می‌شود» (تفسیر). مفروضات، کل تلاشهای مربوط به تحقیق را تحت تأثیر قرار می دهد. به ویژه می‌توان گفت که شرح مفروضات به عنوان مبنای تنظیم سؤالات، بیان فرضیه‌ها و تعبیر و تفسیر داده‌های حاصل از تحقیق محسوب می شود؛ مفروضات تحقیق، به نتایج به دست آمده از پژوهش، معنا و مفهوم می‌بخشد و از پیشنهادهای محقق پشتیبانی می‌کند. در هر صورت، مجموعه منابع مزبور می‌تواند ذهن محقق را آماده نماید تا دربارهٔ چگونگی متغیرها و روابط آنها در تحقیق مورد نظر حدس بزند و پیش‌بینی بعمل آورد و حاصل آن را در قالب قضایای حدسی و خبری تدوین نماید. نکته‌ای که باید دربارهٔ تفاوت فرضیه با نظریه و قوانین یا معلومات کلی بیان شود این است که نظریه و قوانین عمدتاً مشتمل بر قضایای کلی و عمومی هستند و به مورد خاصی تعلق ندارند و می‌توانند مصادیق زیادی داشته باشند، درحالی‌که فرضیه حالت کلی ندارد و مختص مسئله تحقیق است که از قضایای کلی ناشی می‌شود ولی در قلمرو یک تحقیق خاص شکل می‌گیرد؛ به همین دلیل، یک محقق نمی‌تواند فرضیهٔ خود را در تحقیق مورد نظرش بصورت قضیهٔ کلی بیان نماید. هرچند از قضایای کلی به روش قیاسی و از کل به جزء فرضیه‌سازی نماید و فرضیه‌های خود را با قضایای کلی استنتاج نماید، باید مفاهیم و اصطلاحات مربوط به فرضیه را به مسئلهٔ تحقیق خود محدود کند.

۳ ارتباط آزمون آماری با فرضیه آماری

آزمون آماری تکنیکی برای گرفتن تصمیم های کمّی درباره یک یا چند فرآیند است. این آزمون تعیین می کند که آیا شواهد کافی برای «رد» فرضیه مطرح شده وجود دارد یا نه. در این حالت به فرضیه در نظر گرفته شده در اصطلاح فرضیه صفر گفته می شود. رد نشدن یک فرضیه، نتیجه ای مطلوب و مناسب است. اما گاهی نتایج بدست آماده از آزمون آماری می تواند برای ما ناامید کننده باشد. از اینرو رد شدن فرضیه صفر احتمالاً نشان دهده آن است که ما داده‌های کافی برای «اثبات» موضوع مورد نظر خود را در اختیار نداریم. فرضیه تحقیق حاکی از وجود رابطه یا اثر و یا تفاوت بین متغیرها است و یا در واقع وجود این حالات را تایید نموده و آن را واقعی و حقیقی می‌داند. این فرضیه شامل دو نوع جهت دار و بدون جهت است. فرضیه جهت‌دار به فرضیه‌ای گفته می‌شود که در آن جهت ارتباط یا جهت تأثیر متغیر مستقل بر متغیر وابسته مشخص و معیّن است. از این فرضیه هنگامی استفاده می‌شود که پژوهشگر دلایل مشخصی برای پیش‌بینی رابطه معینی داشته باشد. همچنین فرضیه صفر و فرضیه مقابل آن می تواند یک طرفه باشد. بنابراین، آزمون آماری نیازمند یک جفت فرضیه است که به نام های زیر شناخته می شوند: فرضیه صفر عبارتی است که درباره یک فرضیه کلی بیان می شود. ما ممکن است شک داشته باشیم که فرضیه صفر درست باشد، به این معنی که نیاز است درستی این فرضیه مورد «آزمون» قرار گیرد. در حقیقت، فرضیه مقابل ممکن است همان عبارتی باشد که ما اعتقاد به درست بودن آن داریم. به این ترتیب روش آزمون فرضیه به صورتی طراحی می شود که ریسک رد کردن فرضیه صفر در زمان درست بودن آن، حداقل باشد. این ریسک که در اصطلاح به آن یا خطای نوع اول گفته می شود اشاره به سطح اطمینان در نظر گرفته شده برای آزمون دارد. با آزمون فرضیه کردن یک فرضیه با یک مقدار کم ، مشخص می شود که در واقع با رد کردن فرضیه صفر، مفهوم در نظر گرفته شده در فرضیه مقابل یا جایگزین «تایید» شده است. ریسک پذیرفتن فرضیه صفر زمانی است که این فرضیه در واقع نادرست است، خطای نوع دوم یا "بتا" نامیده می شود. اختلاف زیاد بین واقعیت و فرضیه صفر ساده تر قابل شناسایی است و خطای نوع دوم کمتری در آن مشاهده می شود؛ در حالی که شناسایی اختلاف‌های کوچک به دشواری امکان پذیر است و این موضوع منجر به خطای زیاد نوع دوم می‌شود. علاوه بر این، ریسک با افزایش ریسک کاهش پیدا می کند. ریسک خطای نوع دوم معمولاً با یک منحنی ویژگی عملیاتی برای آزمون خلاصه می شود.

۴ تفاوت فرضیه علمی با نظریه علمی

بر اساس یک باور غلط، بسیاری از مردم به اشتباه تصور می‌کنند که نظریه علمی همان فرضیه علمی است که پیشرفت کرده و بهتر شده و تبدیل به نظریه شده است وجه تمایز نظریه‌های علمی و فرضیه‌های علمی در آن است که فرضیه‌های علمی، برآورد و تخمین حاصله از یک پدیده تجربیِ آزمایش‌پذیر و محدود هستند و اگر چه که امری علمی و قدرتمند می‌باشند اما توضیح جامع و ذهنی ارائه نمی‌کنند. همچنین وجه تمایز نظریهٔ علمی با قانون علمی در آن است که قوانین علمی، توضیحی محدود و نه جامع از نحوهٔ رفتار طبیعت در شرایط خاص ارائه می‌کنند (۹). طبق گفتهٔ دانشگاه کالیفرنیا، "فرضیه‌ها، نظریه‌ها و قوانین همانند سیب،

پرتقال و گلابی‌ها هستند. نمی‌توانند به یکدیگر رشد یابند و تبدیل شوند، مهم نیست که چه مقدار کود و آب به پای آنها داده شود." یک فرضیهٔ علمی، شرح و برآورد و تخمینی محدود از یک پدیده است بدون توضیح دربارهٔ علت و چرایی آن؛ یک نظریهٔ علمی توضیحی عمیق و ذهنی از مجموعه‌ای از پدیده‌های مشاهده شده و مرتبط است که به علت و چرایی آنها می‌پردازد (۸). بنابراین یک نظریهٔ علمی شامل یک یا چند فرضیه هستند که این فرضیه‌ها توسط آزمایش‌های مکرری پشتیبانی می‌شوند. نظریه‌ها، قله‌های علوم هستند و صحت آنها به طور گسترده در مجامع علمی پذیرفته شده‌اند (۷). اما نظریه‌ها به طور مستقیم آزمایش‌پذیر نیستند، پس اثبات‌پذیر نیستند. نظریه‌ها تنها تأیید یا رد و ابطال می‌شوند. آنچه که یک نظریه را تأیید یا رد می‌کند، فرضیه‌هایی است که از پس آن نظریه تولید می‌شوند و به طور مستقیم مورد آزمایش قرار می‌گیرند. ابطال یا اثبات یک فرضیه، موجب رد یا تأیید یک نظریه می‌شود.

بحث و نتیجه‌گیری

یک نظریه که فرضیه های بسیاری بر اساس آن ساخته شده و مورد آزمایش قرار گرفته، به طور کل ابطال نمی شود زیرا بخشی از عملکرد طبیعت را به درستی نشان می‌دهد. همچنین فرضیه در علم آمار روشی برای بررسی ادعاها یا فرض‌ها دربارهٔ پارامترهای توزیع در جوامع آماری است. در این مقاله سعی شده است که به طور ساده مفهوم واژه فرض و فرضیه بیان شود و تفاوت فرضیه با نظریه هم آمده است. برای بحث‌های بیشتر می‌توان از لحاظ آمار تخصصی یا سایر علوم مفهوم فرضیه در مسائل استنباطی (مثلا انتخاب مدل) مورد تجزیه و تحلیل قرار داد.

دلاور، ع. (۱۳۹۰). روش تحقیق در روان‌شناسی و علوم تربیتی، *متهران: نشر ویرایش*، **ص ۶۱**.

حسن‌زاده، ر. (۱۳۹۵). روشهای تحقیق در علوم رفتاری: راهنمای عملی تحقیق، *تهران: نشر ویرایش*، **ص ۴۰**.

Washington (2008), *Medicine, National Academy of Sciences*. National Academies Press. p. 11.

https://www.livescience.com/21457-what-is-a-law-in-science-definition-of-scientific-law.html.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

کنکاشی در اندازه‌های اطلاع از دیدگاه بیزی

رحمانپور، آ [۱]

[۱]شرکت داده پردازان نهاد کویر

چکیده

اندازه های اطلاع به دلیل کاربرد فراوانی که در علوم مختلف دارد همواره مورد توجه بوده است. بر این اساس ابتدا به معرفی اندازه های اطلاع پرداخته شده و سپس اندازه های اطلاع از دیدگاه کلاسیک و بیزی مورد بررسی قرار گرفته شده و به مقایسه ی آنها پرداخته ایم.همچنین برآوردگرهای اندازه های اطلاع را معرفی کرده و در ادامه نکاتی از مفاهیم قابلیت اعتماد بر اساس اندازه های اطلاع را مورد مطالعه قرار داده ایم. در نهایت نیز به طور مختصر به معرفی اندازه های اطلاع دینامیکی از دیدگاه بیزی پرداخته ایم.

کلمات کلیدی: اندازه های اطلاع ،بیزی ،آنتروپی.

۱ پیش‌گفتار

نظریه اطلاع را می‌توان به عنوان دست آورد جنگ جهانی دوم تلقی کرد. در طول جنگ جهانی دوم، گروهی در شرکت بل به فرآیند کدسازی پیام های میان روزولت و چرچیل برای حفظ امنیت اطلاعات مبادله شده می‌پرداختند. یکی از اعضای این گروه شانون [۱] نام داشت. وی در آوریل ۱۹۱۶ در میشیگان متولد شد و در سال ۱۹۴۰ درجه‌ی استادی را در رشته های مهندسی الکترونیک و ریاضیات دریافت کرد. شانون در سال ۱۹۴۱ به عنوان یک مهندس مخابرات به عضویت شرکت سیستم بل [۲] درآمد و در سال ۱۹۴۸ با پایه گذاری نظریه‌ی اطلاع ، تحول مهمی در دنیای مخابرات و ارتباطات ایجاد کرد. متن اصلی مقاله‌ی او در همان سال در

[۱]data.kavir@gmail.com

[۱] *Shannon*
[۲] *Bell System*

مجله‌ی "تکنیکی سیستم بل" [۳] منتشر و بعد از آن در کتابی با عنوان "نظریه‌ی ریاضی ارتباطات" [۴] چاپ شد.

۲ اندازه های اطلاع از دیدگاه کلاسیک

آنتروپی چیست؟
آنتروپی در علوم گوناگون مانند شیمی، فیزیک، آمار و ... به کار می‌رود، اما از نگاه علم آمار به صورت زیر تعریف می‌شود:
آنتروپی، شگفتی مورد انتظاریست که از وقوع پیشامدی با احتمال رخداد p حاصل می‌شود که این مقدار یک عدد حقیقی است.میزان شگفتی، وابسته به احتمال وقوع پیشامد است، زیرا هرچه احتمال رخداد پیشامدی کمتر باشد و آن پیشامد رخ دهد، شگفتی بیشتر می شود.اکنون ضابطه‌ی آنتروپی را با توجه به تعریفی که ارائه دادیم به شکل زیر بیان می‌کنیم:

$$H(X) = E(s(P(X))),$$

که در حالت گسسته برابر با $H(X) = -\sum_{i=۱}^{n} p_i \log p_i$ است.

قضیه ۲,۱ فرض کنید $p_۱...p_N$ و$q_۱...q_N$ همگی نامنفی باشند و همچنین داشته باشیم:

$$\sum_{i=۱}^{N} p_i = ۱, \sum_{i=۱}^{N} q_i = ۱.$$

در این صورت رابطه‌ی :

$$-\sum_{i=۱}^{N} p_i \log(p_i) \leq -\sum_{i=۱}^{N} p_i \log(q_i),$$

راخواهیم داشت و تساوی برقرار می‌شود اگر و فقط اگر $p_i = q_i$.
قضیه ۲,۲ حال اگر $p_۱...p_N$ همگی مثبت باشند، طوری که :

$$\sum_{i=۱}^{N} p_i = ۱,$$

نتیجه می‌گیریم که:

$$-\sum_{i=۱}^{N} p_i \log(p_i) \leq \log N.$$

در اینجا نیز تساوی برقرار خواهد شد، اگر و فقط اگر$i = ۱...N$ ، $p_i = \frac{۱}{N}$.
برای حالت پیوسته، آنتروپی شانون برابر $H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx$ است، که این مقدار گاهی منفی می‌شود.

[۳] *Bell System Technical*
[۴] *The Mathematical Theory of Communication*

۱.۲ آنتروپی توأم و آنتروپی شرطی

پس از بیان آنتروپی برای یک متغیر تصادفی، در اینجا آنتروپی توأم را ارائه می‌دهیم، اما قبل از آن لازم است آنتروپی شرطی را معرفی کنیم. آنتروپی شرطی Xبه شرط Y برابر است با:

$$H(X \mid Y) = E(S(P(X \mid Y))) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i \mid y_j) = -\sum_j P_Y(y_j) \sum_i P(x_i \mid y_j) \log P(x_i \mid y_j),$$

که در آن $P(x_i, y_j)$ چگالی تؤام دو متغیر تصادفی X وY، $P(x_i \mid y_j)$ چگالی X به شرط Y و $P_Y(y_j)$ چگالی حاشیه ای Y می‌باشند. حال آنتروپی توأم دو متغیر تصادفی را بیان می‌کنیم:

$$H(X,Y) = H(X \mid Y) + H(Y) = H(Y \mid X) + H(X). \qquad (۱)$$

این عبارت به قاعده‌ی زنجیره ای آنتروپی معروف است.
دیدگاه کلاسیک هرچند در بسیاری از مسائل راهگشا است و به کارگیری آن آسان تر می‌باشد اما در پاسخگویی به برخی از مسائل نامناسب عمل می‌کند. بنابراین با معرفی اندازه های اطلاع از دیدگاه کلاسیک، زمینه را برای ورود به مبحث اندازه های اطلاع از دیدگاه بیزی فراهم آوردیم تا علی رغم محاسبات دشوارتر، پاسخ مسائلی را که ناگزیر به حل آن هستیم، به دست آوریم .
در روش‌های کلاسیک که قبلا مطالعه کردیم، فرض می‌کردیم پارامتر θ یک ثابت نامعلوم است، آنگاه از جامعه ای که دارای تابع چگالی $f_\theta(x)$ می‌باشد نمونه‌ی تصادفی $X_۱...X_n$ ، را انتخاب کرده و با توجه به این نمونه در مورد پارامتر θ تصمیم گیری می‌شود. اما در روش بیزی فرض می‌کنیم که پارامتر θ یک متغیر تصادفی است که درباره‌ی آن یک سری اطلاعات در اختیار داریم و می‌توانیم بر اساس آن، یک توزیع احتمالی را به θ نسبت بدهیم که توزیع پیشین نام دارد. توزیع پیشین بر اساس تجربیات گذشته و قبل از مشاهده‌ی داده‌ها تعیین می‌گردد. پس از نمونه گیری از جامعه‌ی مورد نظر این توزیع تصحیح می‌شود. توزیع پیشین تصحیح شده را توزیع پسین می‌نامند و بر اساس آن تصمیم گیری درباره‌ی θ انجام می‌گیرد.

اندازه‌های اطلاع از دیدگاه بیزی مورد توجه چندانی قرار نگرفته است. زلنر[۵](۱۹۷۱) تابع اطلاع برای داده‌های x با تابع پیشین $p(\theta)$ را به صورت :

$$G(p(\theta)) = E_X(H(p(\theta)) - H(f(x \mid \theta))) = E_X(K(f(x \mid \theta)) : p(\theta))),$$

تعریف کرد. گیل [۶]و جینز (۱۹۷۹) نیز روش محاسباتی را برای استنباط بیزی آنتروپی بدست آوردند. اما رابطه‌ی مستقیم بین اندازه‌های اطلاع و روش بیزی برای اولین بار توسط زلنر در سال ۱۹۸۸ ارائه شد، وی این نظریه را بر اساس قوانین بیزی بیان کرد.
با توجه به مباحثی که مطرح شد، سعی داریم اندازه‌های اطلاع را از نگاه بیزی مورد مطالعه قرار دهیم، زیرا همانطور که می‌دانیم توزیع پسین در روش بیزی، اثر عوامل گوناگون روی متغیرهای تصادفی را در نظر می‌گیرد.

[۵] $Zellner$
[۶] $Gill$

۳ نقش نظریه اطلاع از دیدگاه بیزی

نظریه اطلاع در دهه‌های گذشته در علوم گوناگونی همچون ارتباطات، اقتصاد و آمار پیشرفت بسیاری داشته است. تعاریف زیادی از نظریه اطلاع در مباحث مختلف ارائه شده است که ماهیت همه‌ی آنها یکسان است. گاهی اطلاع جلوه‌ی فیزیکی پیدا می‌کند. مثلا در بازی بیس بال مربی ممکن است با حرکات دست پیامی را برای یکی از بازیکنان ارسال کند.گاهی نیز اطلاع به عنوان اندازه‌ی کاهش عدم قطعیت در نظر گرفته می‌شود. (شانون (۱۹۴۸)، آیریز [۷](۱۹۹۴)). به عنوان مثال اگر بخواهیم از روی رفتار و افکار فرد رأی دهنده حدس بزنیم که وی به چه کسی رأی می‌دهد. کار را با پرسش مجموعه ای از سؤالات که پاسخ آنها بله و خیر می‌باشد، شروع می‌کنیم. با پرسش این سؤالات و تحلیل و تفسیر پاسخ‌ها می‌توان به خواسته‌ی فرد رأی دهنده پی برد. فرض می‌کنیم این رأی گیری در ایالت متحده انجام شود. ۱۳ درصد افراد ایالت متحده، ساکن کالیفرنیا هستند. اولین سؤال این است: آیا رأی دهنده ساکن کالیفرنیاست؟ اگر پاسخ مثبت باشد اطلاعی که بدست می‌آید با پاسخ منفی متفاوت است، در واقع وقتی پاسخ مثبت باشد عدم قطعیت بیشتر کاهش می‌یابد. زیرا جواب مثبت ۸۷ درصد و جواب منفی ۱۳ درصد از انتخاب‌ها را حذف می‌کند.

در این مثال که سؤالها دو پاسخی می‌باشد H_i برابر است با:

$$H_i = f_i \log_۲ \frac{۱}{f_i} = -f_i \log_۲ f_i,\ i = ۱, ۲...n \qquad (۲)$$

و برای سری‌های n تایی از این سؤالها داریم:

$$\sum_{i=۱}^{n} H_i = K \sum_{I=۱}^{n} f_i \log_۲ \frac{۱}{f_i} = -K \sum_{i=۱}^{n} f_i \log_۲ f_i, \qquad (۳)$$

که در آن f_i فراوانی i امین پاسخ بله و K واحد مقیاس گذاری، مثبت و دلخواه است و با توجه به واحد اندازه گیری که در نظر می‌گیریم، می‌تواند هرعددی را بپذیرد. لگاریتم بر مبنای عدد ۲ است زیرا پاسخ سوالها دو جواب دارد که یا بله و یا خیر است. این تابع برای اولین بار توسط اسلاتر [۸] (۱۹۳۸) بیان شد و پس از مطالعاتی که شانون بر روی این تابع انجام داد، آنتروپی شانون (۱۹۴۸) بدست آمد.

در ادامه‌ی مبحث آنتروپی از دیدگاه بیزی، آنتروپی پارامتر θ با تابع توزیع پیشین مطلقا پیوسته‌ی $P(\theta)$ روی فضای Θ را معرفی می‌کنیم که این مقدار برابر:

$$H(\Theta) \equiv H(p(\theta)) = -\int_{\Theta} p(\theta) \log p(\theta) d(\theta), \qquad (۴)$$

خواهد شد که در آن $p(\theta)$ تابع چگالی احتمال P تابع توزیع پیشین θ است.

با توجه به مطالب مطرح شده پیرامون توزیع پیشین بیزی نشان دادیم که در بعضی مواقع واریانس و آنتروپی همسو با هم و برخی اوقات برخلاف هم عمل می‌کنند.

---

[۷] *Ayres*

[۸] *Slater*

همچنین ممکن است آنتروپی و واریانس ارزیابی متناقضی درباره‌ی اطلاع داده‌های نمایی ارائه دهند، در واقع ممکن آنتروپی نشان دهد که با استفاده از داده‌ها عدم قطعیت کاهش می‌یابد اما واریانس خلاف این را نشان دهد.

در مورد داده‌های نرمال با پارامتر مقیاس معلوم، با توزیع پیشین آمیخته، آنتروپی و واریانس درباره‌ی میانگین، همسو با هم عمل می‌کنند و نشان می‌دهند که همه‌ی نمونه‌ها درباره‌ی میانگین اطلاع دهنده هستند، اما در مورد پارامتر مقیاس و میانگین نامعلوم ممکن است این دو مشخصه مخالف هم عمل کنند.
در پایان و پس از بیان روش بیزی و بررسی اندازه‌های اطلاع از دیدگاه بیزی و دیدگاه کلاسیک میتوان دریافت نتایجی که از روش بیزی بدست می‌آید، در صورتی که توزیع پیشین مناسبی توسط آماردان در نظر گرفته شود، دقیق تر است. اما در حالت کلی روش بیزی برای پیش بینی یک پارامتر مفیدتر از روش کلاسیک است. به عنوان مثال در پیش بینی وضعیت هوای یک ناحیه‌ی خاص استفاده از روش بیزی مفیدتر است، زیرا اطلاع از موقعیت مکانی، کوهستانی بودن، خشک بودن و یا معتدل بودن آن ناحیه در پیش بینی وضعیت هوا تأثیر دارد که در صورت استفاده از روش کلاسیک نمی توان از این اطلاعات کمک گرفت. اما چون استفاده از این روش دشوار است، کمتر بکار گرفته می‌شود و در بیشتر موارد روش کلاسیک استفاده می‌شود.
قابلیت اعتماد به معنای عملکرد رضایت بخش یک سیستم در شرایط کاری مشخص و در مدت زمان معین می‌باشد.
این مبحث در صنایع هوافضا، صنایع هسته ای و همچنین صنایع فرآیندهای پیوسته ای مانند: صنایع فولاد و صنایع شیمیایی کاربرد دارد.
قابلیت اعتماد را با دو روش تحلیلی و مشابه سازی مورد بررسی قرار می‌دهند که در روش تحلیلی از مدل های ریاضی و در روش مشابه سازی از یک سری آزمایشات واقعی بهره می‌گیرند.
اکنون با توجه به اهمیت قابلیت اعتماد و کاربردهای آن در صنایع گوناگون برآنیم تا کاربرد اندازه های اطلاع را در مفاهیم قابلیت اعتماد مورد تجزیه و تحلیل قرار دهیم. اصل آنتروپی ماکسیمم را قبلا معرفی کرده ایم. نقش این اصل در مدل های قابلیت اعتماد همانند رده ای از آماره‌های بسنده در تابع چگالی احتمال با پارامتر نامعلوم است.
بحث کاربرد اندازه‌های اطلاع در قابلیت اعتماد را با پیش بینی طول عمر اقلام تولیدی یک کارخانه آغاز می‌کنیم.
به این منظور آزمایش طول عمر را انجام می‌دهیم و با استفاده از تجزیه و تحلیل اقلام معیوب (شکست) و اقلام سالم (بقا)، طول عمر اقلام تولید شده را تخمین میزنیم.
سؤالی که در اینجا مطرح می‌شود این است که در آزمون طول عمر به دنبال بقا هستیم یا شکست، در واقع می‌خواهیم بدانیم که اقلام سالم به پیش بینی طول عمر بیشتر کمک می‌کند یا اقلام معیوب!
بیشتر مهندسان معتقدند اطلاعی که از مشاهده‌ی اقلام معیوب بدست می‌آید، مفیدتر است و در پیش بینی طول عمر می‌تواند بیشتر کمک کند.
معمولا آزمون طول عمر را برای پیش بینی دقیق طول عمر اقلام تولیدی جدید انجام می‌دهیم. اطلاعاتی که از این آزمون بدست می‌آید، توزیع پارامترها و پیش بینی آنها را تحت تأثیر خود قرار می‌دهد، اما برای بدست

آوردن توزیع های پیش بینی کننده ، ابتدا لازم است توزیع پسین پارامترها را بیابیم. در مفاهیم قابلیت اعتماد و تحلیل بقا، سن معمول یک سیستم با استفاده از اندازه‌های اطلاع توزیع طول عمر، مورد مطالعه و بررسی قرار می‌گیرد. در چنین شرایطی، اطلاع کولبک_لایبلر و آنتروپی شانون به زمان بستگی پیدا می‌کند و بنابراین دینامیکی (پویا) خواهند بود. در این قسمت سعی می‌کنیم به بحث و بررسی درباره‌ی این اندازه‌های اطلاع بپردازیم.

موضوع اصلی که به دنبال معرفی آن هستیم، اندازه‌های اطلاع بیزی دینامیکی می‌باشد. اندازه‌های اطلاع بیزی برای مقایسه‌ی طرح ها، مدل ها و ارزیابی داده‌ها به کار می‌روند. بحث دینامیکی این اندازه‌ها توسط دانشمندانی چون: لیندلی (۱۹۵۶)، زلنر (۱۹۷۱،۱۹۷۷و۱۹۹۷) ، برناردو [۹] (۱۹۷۹)، پلسون [۱۰] (۱۹۹۲)، ابراهیمی و همکاران (۲۰۰۶) ... مورد بررسی قرار گرفته است.
در بحث اندازه‌های اطلاع بیزی دینامیکی، توزیع پیشین $p(\theta)$ را برای پارامتر طول عمر مدل $F_{Y|\theta}(y \mid \theta)$ در نظر می‌گیریم. ممکن است y و θ هر کدام یا هر دو بردار باشند، اما برای سهولت کار، حالتی که هر دوی آنها یک بعدی باشند را مورد مطالعه قرار می‌دهیم. ابتدا اندازه‌های اطلاع دینامیکی در حالت کلاسیک را که قبلا معرفی کردیم ، مورد کنکاش بیشتر قرار داده و سپس این اندازه‌ها را از دیدگاه بیزی بررسی می‌کنیم.

۴ اندازه‌های اطلاع دینامیکی

فرض می‌کنیم متغیر تصادفی نامنفی X در مسائل قابلیت اعتماد و تحلیل بقا، نشان دهنده‌ی طول عمر باشد.
توزیع طول عمر باقی مانده‌ی F و G به ترتیب $F_t(x) = P_F(X - t \leq x \mid X > t)$ و $G_t(x) = P_G(X - t \leq$ $x \mid X > t)$ با تابع چگالی های $f_t(x) = \frac{f(x)}{\bar{F}(t)}$ و $g_t(x) = \frac{g(x)}{\bar{G}(t)}$ می‌باشد که در آن $\bar{F}(t) = ۱ - F(t)$ و $\bar{G}(t) = ۱ - G(t)$ نشان دهنده‌ی تابع بقا هستند.
در این صورت اطلاع کولبک_لایبلر برای دو تابع چگالی f_t و g_t برابر است با:

$$K(f, g, t) \equiv K(f_t, g_t) = \int_t^\infty f_t(x) \log \frac{f_t(x)}{g_t(x)} dx = E_t \log \frac{f(X)}{g(X)} - \log \frac{\bar{F}(t)}{\bar{G}(t)},$$

که در آن E_t امید ریاضی نسبت به چگالی f_t می‌باشد.
تا اینجا با اندازه‌های اطلاع پویا از دیدگاه کلاسیک آشنا شدیم. اکنون اندازه‌های اطلاع از دیدگاه بیزی، وقتی تکیه گاه توزیع داده‌ها بریده شده باشد را مورد بحث و بررسی قرار می‌دهیم .
این مبحث را ابراهیمی و همکاران (۲۰۰۶) برای اولین بار به طور جدی مورد مطالعه قرار دادند. تمرکز در این بخش بر روی توزیع طول عمرهای $t \geq ۰$ است. همانطور که در مقدمه‌ی فصل بیان کردیم، توزیع پیشین $p(\theta)$ را برای پارامتر طول عمر مدل $F_{Y|\theta}(y \mid \theta)$ در نظر می گیریم.
در ابتدا تابع توزیع $f_{Y_t|\theta}(y_t \mid \theta, t)$ را به شکل زیر تعریف می کنیم:

[۹] Bernardo
[۱۰] Polson

$$f_{Y_t|\theta}(y_t \mid \theta, t) = \frac{f_{Y|\theta}(y_t \mid \theta)}{\bar{F}_{Y|\theta}(t \mid \theta)}, y_t > t \qquad (۵)$$

حال با استفاده از تابع چگالی $f_{Y_t|\theta}(y_t \mid \theta, t)$ ، توزیع پیشین را تصحیح می‌کنیم، بنابراین خواهیم داشت:

$$p_{\Theta|\ y_t}(\theta \mid\ y_t, t) \propto \frac{f_{Y|\theta}(y_t \mid \theta) p_\Theta(\theta)}{\bar{F}_{Y|\theta}(t \mid \theta)}, y_t > t, \theta \in \theta, \qquad (۶)$$

که این تابع چگالی، چگالی پسین پارامترθ است.

واضح است که $p_{\Theta|\ y_t}(\theta \mid\ y_{\cdot}, \cdot) = p_{\Theta|\ y}(\theta \mid\ y)$.

آینده‌ی تحقیق

آنچه در این مقاله از نظر گذشت، حاصل مطالعه‌ی مقالات گوناگون و همچنین کتاب های متعدد مرتبط با مبحث اندازه‌های اطلاع و روش بیزی بود. اما مسلما این مطالب تمام جنبه ها را در برنمی گیرد و مفاهیم فوق قابل بررسی و گسترش هستند و ادامه‌ی تحقیق توقف نمی‌پذیرد. لذا موارد پیشنهادی زیر می‌تواند ادامه‌ی این مسیر باشد:

۱ـ بررسی نظریه‌ی اطلاع بیزی بر مبنای تابع مفصل .

۲ـ مطالعه و کنکاش بیشتر اندازه‌های بیزی دینامیکی.

۳ـ محاسبه‌ی آنتروپی ماکسیمم از دیدگاه بیزی در حالت تک متغیره و چند متغیره و کاربرد آن در بحث برآورد.

Abel, p. s. and Singpurwalla, N. D. (1994).To survive or to fail: that is the question. Amer. Statist., 48, 18–21.

Abraham, B. and Sankaran, P. G. (2004). Renyi's entropy for residual lifetime distribution. Stat. Paper 46, 17-30.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions On Automatic Control, 19, 716-723.

Basharin, G. P. (1959). On a statistical estimate for the entropy of a sequence of independent random variables, Theory of Prob. and Its Appl., 4, 333-336.

Bernardo, J. M. (1979). Expected information as expected utility. The Annals of Statistics, 7, 686-690.

Cover, T. M. and Thomas, j. A. (2006). Elements of information theory. Second Edition, Wiley InterScience.

Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. J. Appl. Probab. 39, 434–440.

Ebrahimi, N., Soofi, E. S. , Soyer, R. (2010). On the sample information about parameter and prediction. Statist. Sci, Volume 25, 348-367.

Yousefzadeh, F. and Arghami, N. R. (2008), Testing Exponentiality Based on Type II Censored Data and a New cdf Estimator, Communications in Statistics-Simulation and Computation, 37, 1479–1499.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

مدل انتقال واریانس در مدل رگرسیونی تحت محدودیت خطی تصادفی

رحمی، ف [۱] بابادی، ب [۲] راسخ، ع [۲]

[۱،۲]گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز

چکیده

در این مقاله مدل انتقال واریانس را برای یک مدل رگرسیونی تحت محدودیت خطی تصادفی ارائه می‌دهیم. این مدل بر این پایه استوار است که یک نقطه پرت در اثر تورم در واریانسش رخ می‌دهد. برآورد پارامترها در این مدل به‌دست آمده و برای شناسایی نقطه پرت از آماره آزمون امتیاز استفاده می‌شود. در بخش آخر نیز بحث و نتیجه‌گیری ارائه می‌گردد.

کلمات کلیدی: مدل انتقال واریانس، محدودیت تصادفی، نقاط پرت.

۱ پیش‌گفتار

نقاط پرت مشاهداتی هستند که به نظر می‌رسد با سایر داده‌های مجموعه سازگار نیستند و می‌توانند تاثیر مخربی بر تحلیل آماری داشته باشند. برای تشخیص این نوع مشاهدات در مدل‌های خطی، روش‌های مختلفی پیشنهاد شده است. در میان این روش‌ها می‌توان به حذف موردی و مدل‌های انتقال واریانس اشاره کرد. اساس روش اول مبتنی بر این فرض است که نقاط پرت ناشی از تغییر در میانگین این مشاهدات می‌باشند و دومین روش این فرض را مطرح می‌کند که یک نقطه‌ی پرت از یک عبارت خطا با واریانس متورم به‌دست می‌آید. (۲) نشان دادند که برآوردهای ماکسیمم درستنمایی برای موقعیت نقطه پرت تحت این دو روش می‌تواند متفاوت باشد؛ مگر این که قدرمطلق باقیمانده استیودنت شده، مربوط به قدرمطلق باقیمانده‌ی همان مشاهده باشد.

[۱] fatemerahmi3@gmail.com

با استفاده از باقیمانده ماکسیمم درستنمایی (REML)، (۵) نشان داد که واریانس باقیمانده و موقعیت نقطه پرت تحت دو روش فوق یکسان است. از طرفی در صورتی که بین متغیرهای توضیحی هم‌خطی وجود داشته باشد، این اتفاق موجب برآورد با واریانس‌های بزرگ برای ضرایب رگرسیونی می‌شود. بنا به گفته‌ی (۵) یک روش برای کاهش مشکلات هم‌خطی، استفاده از اطلاعات کمکی در خصوص پارامترهای مدل رگرسیونی است که از پژوهش‌های گذشته به‌دست آمده و از آن تحت عنوان محدودیت خطی یاد می‌شود. در این مقاله یک مدل رگرسیونی تحت محدودیت خطی تصادفی با یک نقطه پرت در نظر گرفته می‌شود. فرض شده است که نقطه پرت از یک خطای با واریانس متورم به‌وجود می‌آید. در بخش ۲ ابتدا مدل رگرسیونی تحت محدودیت خطی تصادفی معرفی و برآورد پارامترها ارائه می‌شود. سپس مدل انتقال واریانس در این مدل در بخش سوم معرفی شده و خواص مجانبی برآورد پارامترها مطرح می‌شوند. در بخش ۴ آماره ی آزمون امتیاز که قبلا توسط (۱) برای مدل‌های آمیخته مورد استفاده قرار گرفته است برای این مدل ارائه می‌گردد، در بخش آخر یک بحث و نتیجه‌گیری در رابطه با این موضوع صورت می‌گیرد.

۲ مدل رگرسیونی تحت محدودیت تصادفی

فرض کنید y یک بردار m بعدی از مشاهدات به فرم زیر باشد:

$$y = X\beta + e \qquad (۱)$$

که X یک ماتریس معلوم $m \times p$، β یک بردار مجهول از پارامترها و e یک بردار تصادفی m بعدی از خطاها و دارای توزیع نرمال با میانگین صفر و ماتریس واریانس کوواریانس $\sigma^2 I_m$ می‌باشد. برآورد حداقل مربعات برای β و σ^2 به صورت زیر به دست می‌آید:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\sigma}^2 = m^{-1}(y'y - y'X\hat{\beta})$$

در مدل (۱) برای برآورد پارامترها، محدودیت خطی تصادفی زیر را در نظر می گیریم:

$$r = R\beta + v \qquad (۲)$$

جایی که R یک ماتریس $j \times p$ معلوم با رتبه‌ی j، v بردار خطای $j \times 1$ با امید یاضی صفر و ماتریس واریانس‌ـ کوواریانس $\sigma^2 I_j$ و r بردار $j \times 1$ به عنوان متغیری تصادفی با امید ریاضی $E(r) = R\beta$ است. برآورد پارامترها تحت محدودیت را با نمادهای $\hat{\beta}_r$ و $\hat{\sigma}_r^2$ نشان داده و به فرم زیر نمایش می‌دهیم:

$$\hat{\beta}_r = (X'X + R'R)^{-1}(X'Y + R'r)$$

$$\hat{\sigma}_r^2 = \frac{1}{n}\left[(y'y - y'X\hat{\beta}_r) + (r'r + r'R\hat{\beta}_r)\right]$$

که $n = m + j$.

۳ مدل انتقال واریانس تحت محدودیت خطی تصادفی

یک مدل انتقال واریانس در صورتی که iامین مشاهده پرت باشد به صورت زیر تعریف می‌شود:

$$\begin{pmatrix} Y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \beta + f_i d_i + \begin{pmatrix} e \\ v \end{pmatrix} \qquad (۳)$$

در واقع عبارت $f_i d_i$ به مدل (۱) اضافه می گردد که d_i یک بردار ستونی با مقدار یک در iامین عنصر و مقدار صفر در بقیه عناصراست و f_i یک ضریب تصادفی مجهول با توزیع $f_i \sim N(۰, \alpha_i)$ برای $\alpha_i \geq ۰$ می‌باشد. ماتریس واریانس کوواریانس برای داده‌ها تحت مدل انتقال واریانس برای یک داده پرت به صورت زیر نوشته می‌شود:

$$Var(y) = \alpha_i d_i d_i' + \sigma^۲ I_m$$

فرم ماتریس واریانس کوواریانس داده‌ها را می‌توان به صورت زیر بازنویسی کرد:

$$Var(y) = \sigma^۲ \left[(w_i - ۱) d_i d_i' + I_m \right] = \sigma^۲ H_i$$

که $w_i = \alpha_i / \sigma^۲ + ۱$ واریانس برای واحد iام به صورت متورم و به فرم $\sigma^۲ w_i$ می‌باشد. برآورد β و $\sigma^۲$ در مدل انتقال واریانس تحت محدودیت تصادفی (۸) به‌صورت زیر به‌دست می‌آیند:

$$\hat{\beta}_r(w_i) = (X' H_i^{-۱} X + R' R)^{-۱} (X' H_i^{-۱} y + R' r)$$

$$\hat{\sigma}_r^۲ (w_i) = \frac{۱}{n} \left[y' H_i^{-۱} y - y' H_i^{-۱} X \hat{\beta}_r(w_i) \right.$$

$\hat{\beta}_r (w_i)_r(w_i)$+r'r-r'R و $\hat{\sigma}_r^۲ (w_i)$ را می‌توان به عنوان تابع‌هایی از برآوردهای ماکسیمم درستنمایی β و $\sigma^۲$ زمانی که هیچ نقطه‌ی پرتی وجود ندارد، درنظر گرفت.

برای هر i مقدار $v_i = x_i' (X'X + R'R)^{-۱} x_i$ را تعریف می کنیم ($۰ \leq v_i \leq ۱$)، به طوری که بردار x_i' سطر iام ماتریس X است. سپس، $\hat{y}_i = x_i' \hat{\beta}_r$ را درنظر می‌گیریم. با شرط $v_i \neq ۱$، iامین باقیمانده استیودنت شده را به صورت زیر درنظر می‌گیریم:

$$t_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_r \sqrt{(۱ - v_i)}} \sqrt{\frac{n-p}{n}}$$

در نتیجه:

$$\hat{\beta}_r (w_i) = \hat{\beta}_r - \frac{w_i - ۱}{w_i - (w_i - ۱) v_i} \left[(X'X + R'R)^{-۱} X' d_i (y_i - \hat{y}_i) \right]$$

$$\hat{\sigma}_r^۲ (w_i) = \hat{\sigma}_r^۲ \left[۱ - \frac{(w_i - ۱)(۱ - v_i)}{w_i - (w_i - ۱) v_i} \frac{t_i^۲}{n-p} \right]$$

ویژگی‌های این توابع را می‌توان به فرم زیر بیان کرد:

۱. زمانی که $۱ - v_i = ۰$ باشد، آنگاه $\hat{\beta}_r(w_i) = \hat{\beta}_r$ و $\hat{\sigma}_r^۲(w_i) = \hat{\sigma}_r^۲$.

۲. اگر $۱ - v_i > ۰$ باشد، آنگاه:

$$\lim_{w \to +\infty} \hat{\beta}_r (w_i) = \hat{\beta}_r - (y_i - \tilde{y}_i) (۱ - v_i)^{-۱} (X'X + R'R)^{-۱} x_i$$

که برآورد معمول ماکسیمم درستنمایی β تحت محدودیت خطی تصادفی اگر مشاهده‌ی iام خارج شود را به ما می‌دهد. همچنین

$$\lim_{w\to+\infty}\hat{\sigma}_r^{۲}(w_i)=\hat{\sigma}_r^{۲}\left(۱-t_i^{۲}(n-p)^{-۱}\right)=(n-۱)\,n^{-۱}\hat{\sigma}_{r(i)}^{۲}$$

که برآورد معمول ماکسیمم درستنمایی $\sigma^{۲}$ تحت محدودیت تصادفی هنگامی که مشاهده‌ی iام خارج می‌شود را به ما نشان می‌دهد.

با جایگذاری این برآوردها در تابع درستنمایی و حذف مقادیر ثابت، تابعی برحسب w به فرم زیر به‌دست می‌آید:

$$h_r(w_i)=-nlog\left[\hat{\sigma}_r^{۲}(w_i)\right]-\log(w_i)$$

$h_r(w_i)$ را با حذف مقادیر ثابت می‌توان به فرم زیر نوشت:

$$\tilde{h}_r(w_i)=-log\left[\hat{\sigma}_r^{۲}-\frac{w_i-۱}{w_i-(w_i-۱)\,v_i}\frac{t_i^{۲}}{n-p}\right]-log(w_i)$$

مقدار مشاهده شده‌ی w_i را با $\tilde{w}_i$ نشان داده که در بازه‌ی $[۱,+\infty)$ قرار گرفته و $\tilde{h}(w_i)$ را ماکسیمم می‌کند.

مشتق‌گیری از $\tilde{h}_r(w_i)$ دشوار می‌باشد در نتیجه ماکسیمم این تابع را می‌توان با استفاده از محاسبات نرم‌افزاری محاسبه کرد، که وجود این مقدار توسط (۲) اثبات شده است. در بخش بعدی با استفاده از آزمون امتیاز w_i را مورد آزمون قرار خواهیم داد.

۴ آزمون امتیاز

در این بخش ما آماره آزمون امتیاز را برای آزمون کردن فرض صفر که مشاهده غیرعادی نیست ($w_i=۱$) در مقابل فرض مخالف که مشاهده دارای واریانس متورم است ($w_i>۱$) به‌دست می‌آوریم: **آماره آزمون امتیاز:** به منظور به‌دست آوردن آماره آزمون امتیاز، ما تنها به برآورد پارامترها تحت فرض صفر احتیاج خواهیم داشت. با استفاده از قضیه زیر براساس ماتریس اطلاعات مشاهده آماره آزمون را تحت فرض صفر ($w_i=۱$) به‌دست می‌آوریم.

قضیه ۱.۴. آماره آزمون امتیاز برای iامین مشاهده (SC_i)، براساس ماتریس اطلاعات مشاهده شده برای آزمون کردن فرض $H_o: w_i=۱$ به صورت زیر داده شده:

$$SC_i(w_i=۱)=\frac{n\left(t_i^{۲}(۱-v_i)-۱\right)^{۲}}{۲\left\{-n+۲nt_i^{۲}(۱-v_i)-t_i^{۴}(۱-v_i)^{۲}\right\}} \qquad (۴)$$

برهان. فرض کنید که ماتریس اطلاعات مشاهده شده‌ی y برای $\sigma^{۲}$ و w_i به صورت $J(\sigma^{۲},w_i)$ باشد. سپس آماره آزمون امتیاز را برای آزمون کردن فرض $H_.: w_i=۱$ در مقابل فرض $H_۱: w_i>۱$ به فرم زیر به‌دست می‌آید:

$$SC_i=\left[\frac{\partial}{\partial w_i}l_i^{*}\left(\beta,\sigma^{۲},w_i;X,y\right)\right]^{۲}J^{\omega\omega} \qquad (۵)$$

که $J^{\omega\omega}$ درایه‌ی پایین گوشه‌ی سمت راست ماتریس $J^{-1}(\sigma^2, w_i)$ می‌باشد. با قرار دادن مقادیر $w_i = 1$، $\beta = \hat{\beta}_r$ و $\sigma^2 = \hat{\sigma}_r^2$ در عناصر عبارت (۵) داریم:

$$\frac{\partial}{\partial w_i} l_i^* \left(\beta, \sigma^2, w_i; X, y\right) = -\frac{1}{2} + \frac{1}{2\hat{\sigma}_r^2}\hat{e}_i^2 = \frac{1}{2}\left(t_i^2\left(1 - v_i\right) - 1\right)$$

سپس

$$J\left(\sigma^2, w_i\right) = \begin{bmatrix} \frac{n}{2\hat{\sigma}_r^4} & \frac{1}{2\hat{\sigma}_r^4}\hat{e}_i^2 \\ \frac{1}{2\hat{\sigma}_r^4}\hat{e}_i^2 & -\frac{1}{2} + \frac{1}{2\hat{\sigma}_r^2}\hat{e}_i^2 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{n}{2\hat{\sigma}_r^4} & \frac{1}{2\hat{\sigma}_r^2} t_i^2\left(1 - v_i\right) \\ \frac{1}{2\hat{\sigma}_r^2} t_i^2\left(1 - v_i\right) & -\frac{1}{2} + \frac{1}{2} t_i^2\left(1 - v_i\right) \end{bmatrix}$$

که $J^{\omega\omega} = 2n/\left\{-n + 2nt_i^2\left(1 - v_i\right) - t_i^4(1 - v_i)^2\right\}$. با قرار دادن این عبارت در (۵) آماره آزمون حاصل می‌شود. □

از آنجایی که فرض صفر در بازه‌ی فضای پارامتر w قرار دارد ($w_i \geq 1$)، نظریه‌ی استاندارد نسبیت در این مورد صدق نمی‌کند. از این رو، می‌توان از یک برنامه پارامتری بوت استرپ برای تقریب توزیع آماره آزمون امتیاز استفاده کرد.

بحث و نتیجه‌گیری

در این مقاله ابتدا مدل انتقال واریانس را برای یک مدل رگرسیونی ساده بیان کردیم. در بخش بعد این مدل را برای مدل‌های رگرسیونی خطی تحت محدودیت تصادفی معرفی و سپس برآورد پارامترها را تحت این مدل محاسبه کردیم. در بخش آخر نیز با استفاده از تابع امتیاز و ماتریس اطلاعات مشاهده شده، آماره آزمون امتیاز معرفی شد.

مراجع

[1] Babadi, B., et al. (2016). A variance shift model for detection of outliers in the linear mixed measurement error models. *Communications in Statistics-Theory and Methods*, **45(24)**, 7350-7366.

[2] Cook, R.D., Holschuh, N. and Weisberg, S. (1982), A note on an alternative outlier model, *J. R. Statist. Soc. Ser.* B **44**, pp. 370-376.

[3] Harville, D. A. (1977), Maximum likelihood approaches to variance component estimation and to related problems.*J. Amer. Statist. Ass.*, **72**, 320-340.

[4] Rao, C.R., Toutenburg, H., Shalabh and Heumann, C. (2008), *Linear Models and Generalizations: Least Squares and Alternatives*. Third ed, Springer Berlin.

[5] Thompson, R. (1985), A note on restricted maximum likelihood estimation with an alternative outlier model, *J. R. Statist. Soc. Ser.* B **47**, pp. 53-55.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

برآورد لیو جک نایف تحت محدودیتهای خطی تصادفی در مدلهای رگرسیونی

طلادزفولی، م [۱] راسخ، ع [۲] بابادی، ب [۳]

[۱،۲،۳] دانشگاه شهید چمران اهواز، دانشکده علوم ریاضی و کامپیوتر، گروه آمار

چکیده

با وجود مسئله‌ی هم‌خطی در مدل‌های رگرسیونی، برآورد کمترین توان‌های دوم کارایی خود را از دست می‌دهد. راهکارهایی برای کاهش اثرات هم‌خطی پیشنهاد شده که از آن جمله میتوان به استفاده از برآوردگرهای اریب مانند برآورد لیو و یا برآورد ضرایب رگرسیونی تحت محدودیت‌های خطی اشاره کرد. اما با توجه به اریب بودن این برآوردگرها، روش جک نایف جهت کاهش میزان اریبی معرفی شده است. ما در این مقاله برآورد لیو جک نایف تحت محدودیت‌های خطی تصادفی را مورد بررسی قرار خواهیم داد. در ادامه با استفاده از یک مطالعه‌ی شبیه‌سازی عملکرد این برآورد جدید با توجه به دو معیار میانگین مربعات خطا و قدرمطلق اریبی سنجیده می‌شود.

کلمات کلیدی: هم‌خطی، برآورد لیو، روش جک نایف، محدودیت‌های خطی و مدل کانونی.

۱ پیش‌گفتار

هم‌خطی زمانی اتفاق می‌افتد که دو یا چند متغیر توضیحی نسبت به یکدیگر از همبستگی بالایی برخوردار باشند. این موضوع منجر به ایجاد برآوردی با واریانس‌های بزرگ برای ضرایب رگرسیونی می‌شود. برآوردگر لیو یکی از مشهورترین برآوردگرها در مواجه با هم‌خطی است. این برآوردگر با پذیرفتن مقداری اریبی، به‌طور چشم‌گیری واریانس برآورد حاصل را کاهش می‌دهد. لذا با توجه به اینکه این برآوردگر اریب است، اما

[۱] taladezfoli-mahtab@yahoo.com

نتایج حاصل از برازش آن قابل استنادتر و پایدارتر است. بنابراین تنها مسئله ی نگران کننده در ارتباط با به کارگیری این برآوردگر، اریب بودن آن است. از این رو محققان جهت کاهش اریبی برآوردگر موردنظر از روش‌هایی استفاده می کنند که روش جک‌نایف یکی از این روش‌ها است.

از سوی دیگر، گاهی اوقات مجموعه‌ای از اطلاعات کمکی در مورد ضرایب رگرسیونی در اختیار است که به‌کارگیری آنها در تحلیل‌های رگرسیونی و روند برآورد، می‌تواند به‌طور قابل ملاحظه‌ای اثرات ناشی از هم‌خطی را کاهش دهد. معمولا این اطلاعات کمکی به فرم محدودیت‌های خطی بیان می‌شوند. اعمال این محدودیت‌ها بر بردار ضرایب رگرسیونی موجب کاهش واریانس و به عبارتی بهبود دقت برآورد و در نتیجه کاهش اثرات هم‌خطی می‌شود.

در این تحقیق سعی شده که با به‌کارگیری روش جک‌نایف بر روی برآورد لیو، و همچنین استفاده از اطلاعات کمکی به فرم محدودیت‌های خطی تصادفی، برآوردگر جدیدی جهت کاهش هر چه بیش‌تر اثرات هم‌خطی معرفی گردد.

۲ معرفی مدل و برآوردگرها

مدل رگرسیونی زیر را در نظر بگیرید:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(۰, \sigma^۲ \boldsymbol{I}_n) \tag{۱}$$

که در آن $\boldsymbol{X}$ یک ماتریس $n \times p$ از متغیرهای توضیحی، $\boldsymbol{Y}$ یک بردار $n \times ۱$ از مقادیر متغیر پاسخ، $\boldsymbol{\beta}$ یک بردار $p \times ۱$ از پارامترهای نامعلوم، $\boldsymbol{\varepsilon}$ یک بردار $n \times ۱$ از خطاها با امید ریاضی $E(\boldsymbol{\varepsilon}) = ۰$ و ماتریس واریانس‌ـ کوواریانس $Var(\boldsymbol{\varepsilon}) = \sigma^۲ \boldsymbol{I}_n$ است. برآورد کمترین توان‌های دوم معمولی برای مدل (۱) به صورت زیر تعریف می‌شود:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-۱}\boldsymbol{X}'\boldsymbol{Y}. \tag{۲}$$

برای مواجه با مسئله‌ی هم‌خطی، (۲) برآوردگر اریبی را تحت عنوان برآورد لیو، به صورت زیر معرفی کرد:

$$\hat{\boldsymbol{\beta}}(d) = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I}_p)^{-۱}(\boldsymbol{X}'\boldsymbol{Y} + d\hat{\boldsymbol{\beta}}) \tag{۳}$$

در اینجا $۰ < d < ۱$ پارامتر لیو نامیده می‌شود. همچنین برآورد لیو را می‌توان به صورت زیر نوشت:

$$\hat{\boldsymbol{\beta}}(d) = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I}_p)^{-۱}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I}_p)\hat{\boldsymbol{\beta}} \tag{۴}$$

در ادامه برای سهولت محاسبات، مدل (۱) را به فرم کانونی بازنویسی می‌کنیم. فرض کنید $\boldsymbol{T}$ یک ماتریس متعامد باشد به‌طوری که:

$$\boldsymbol{T}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T} = \boldsymbol{\Lambda}$$

که $\boldsymbol{\Lambda}$ ماتریس قطری $p \times p$ است، درایه‌های قطری آن برابر $\lambda_۱, \lambda_۲, \ldots, \lambda_p$ مقادیر ویژه‌ی ماتریس $\boldsymbol{X}'\boldsymbol{X}$ هستند. با استفاده از ماتریس $\boldsymbol{T}$، فرم کانونی مدل (۱) را به صورت زیر بدست می‌آوریم:

$$(۵)\varepsilon+\gamma Z=Y$$

که $Z = XT$، $\gamma = T'\beta$ و $Z'Z = \Lambda$ است. بنابراین برآورد کمترین توان‌های دوم برای مدل (۲) به صورت زیر حاصل می‌شود:

$$\hat{\gamma} = (Z'Z)^{-1}Z'Y = \Lambda^{-1}Z'Y \qquad (۶)$$

همچنین برآورد لیوی بردار γ در مدل (۲) برابر است با:

$$\hat{\gamma}(d) = (\Lambda + I_p)^{-1}(Z'Y + d\hat{\gamma}) = (\Lambda + I_p)^{-1}(\Lambda + dI_p)\hat{\gamma} = F_d\hat{\gamma} \qquad (۷)$$

که

$$F_d = (\Lambda + I_p)^{-1}(\Lambda + dI_p) = diag\left\{\frac{\lambda_1 + d}{\lambda_1 + 1}, ..., \frac{\lambda_p + d}{\lambda_p + 1}\right\}$$

با توجه به اینکه، برآورد لیو یک برآورد اریب است، (۱) به منظور کاهش اریبی برآورد لیو، از روش جک‌نایف بهره برده و برآورد لیو جک‌نایف را معرفی نمودند. برآورد لیو جک‌نایف به صورت زیر حاصل می‌شود:

$$\tilde{\gamma}(d) = (2I_p - F_d)F_d\hat{\gamma}$$

همچنین (۶) یک نسخه‌ی اصلاح شده‌ی برآورد لیو جک‌نایف را پیشنهاد داد و عملکرد این برآوردگر را با توجه به معیار میانگین مربعات خطا با برخی برآوردهای دیگر مورد مقایسه قرار داد.

با توجه به اینکه $\gamma = T'\beta$ است، در نتیجه برآوردهای کمترین توان‌های دوم، لیو و لیو جک‌نایف برای مدل (۱) به ترتیب برابر $\hat{\beta} = T\hat{\gamma}$، $\hat{\beta}(d) = T\hat{\gamma}(d)$ و $\tilde{\beta}(d) = T\tilde{\gamma}(d)$ می‌باشند.

۳ برآورد لیو جک‌نایف تحت محدودیت‌های خطی تصادفی

اگر اطلاعات کمکی در مورد بردار ضرایب رگرسیونی در دسترس باشد، معمولا آنها را به فرم محدودیت‌های خطی بیان می‌کنند که به دو صورت محدودیت‌های خطی دقیق و تصادفی بر روی بردار ضرایب اعمال می‌شوند. در این بخش تنها به بیان محدودیت‌های خطی تصادفی اکتفا می‌کنیم. محدودیت‌های خطی تصادفی به صورت زیر بیان می‌شوند:

$$r = R\beta + \phi, \qquad \phi \sim (0, \sigma^2 I_m) \qquad (۸)$$

که r یک بردار $m \times 1$ از مقادیری معلوم، R یک ماتریس $m \times p$ از اطلاعات کمکی در ارتباط با ضرایب رگرسیونی با رتبه‌ی $m \leq p$ و معلوم و ϕ یک بردار تصادفی و مستقل از بردار ε با بردار میانگین صفر و ماتریس واریانس‌ـ کوواریانس $\sigma^2 I_m$ است.

برای به‌دست آوردن برآورد ضرایب رگرسیونی تحت محدودیت‌های خطی تصادفی،مشابه به روش (۳) عمل می‌کنیم. در این روش، برآورد ضرایب به‌وسیله‌ی یکی کردن اطلاعات نمونه و اطلاعات کمکی (۸) حاصل می‌شوند.

با استفاده از نمادگذاری جدید، $\tilde{Y} = \begin{pmatrix} Y \\ r \end{pmatrix}$ ، $\tilde{X} = \begin{pmatrix} X \\ R \end{pmatrix}$ و $\tilde{\varepsilon} = \begin{pmatrix} \varepsilon \\ \phi \end{pmatrix}$ به مدلی به شکل زیر دست می‌یابیم:

$$\begin{pmatrix} Y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \phi \end{pmatrix} \quad \Longrightarrow \quad \tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim (۰, \sigma^۲ I_{n+m}), \qquad (۹)$$

با به‌کارگیری روش کمترین توان‌های دوم برای مدل (۹)، برآورد بردار β به‌صورت زیر به‌دست می‌آید:

$$\hat{\beta}_r = (\tilde{X}'\tilde{X})^{-۱}\tilde{X}'\tilde{Y} = (X'X + R'R)^{-۱}(X'Y + R'r).$$

برآورد $\hat{\beta}_r$ را برآوردگر آمیخته می‌نامند که برآوردی ناریب است. برای جزییات بیش‌تر به (۵) رجوع شود.

حال با ادغام محدودیت $d\hat{\beta}_r = \beta + \eta$ با مدل (۹) می‌توان به برآورد لیو دست یافت. بنابراین داریم:

$$\begin{aligned} \hat{\beta}_r(d) &= (X'X + R'R + I)^{-۱}(X'Y + R'r + d\hat{\beta}_r) \\ &= (X'X + R'R + I)^{-۱}(X'X + R'R + dI)\hat{\beta}_r = G_d\hat{\beta}_r \end{aligned}$$

که برآورد لیو تحت محدودیت‌های خطی تصادفی (۸) برای مدل (۱) است.

می‌دانیم که مدل (۹) را می‌توان به فرم کانونی نوشت. از این رو ماتریس $P_{p\times p}$ وجود دارد به طوری که

$$P'\tilde{X}'\tilde{X}P = \Pi, \qquad P'P = PP' = I_p$$

که Π یک ماتریس قطری است که مولفه‌های روی قطر آن، $\pi_۱, \ldots, \pi_p$، مقادیر ویژه‌ی ماتریس $\tilde{X}'\tilde{X} = X'X + R'R$ هستند. با استفاده از ماتریس P، مدل (۹) به صورت زیر حاصل می‌گردد:

$$\tilde{Y} = \tilde{X}PP'\beta + \tilde{\varepsilon} \quad \Rightarrow \quad \tilde{Y} = U\theta + \tilde{\varepsilon} \qquad (۱۰)$$

که $U = \tilde{X}P$ و $\theta = P'\beta$. از این رو برآورد آمیخته برای مدل (۱۰) به صورت زیر حاصل می‌شود

$$\hat{\theta}_r = (U'U)^{-۱}U'\tilde{Y} = \Pi^{-۱}U'\tilde{Y}$$

همچنین برآورد لیو برای مدل (۱۰) برابر است با:

$$\hat{\theta}_r(d) = (U'U + I)^{-۱}(U'U + dI)\hat{\theta}_r = (\Pi + I)^{-۱}(\Pi + dI)\hat{\theta}_r = G_d\hat{\theta}_r$$

که در آن G_d یک ماتریس قطری است و اعضای روی قطر آن به صورت زیر تعریف می‌شوند:

$$G_d = diag\left\{\frac{\pi_۱ + d}{\pi_۱ + ۱}, \ldots, \frac{\pi_p + d}{\pi_p + ۱}\right\}.$$

از سوی دیگر، به دلیل اینکه برآورد لیو تحت محدودیت‌های خطی تصادفی، یک برآورد اریب است، لذا با اعمال روش جک‌نایف بر روی برآورد لیو تحت محدودیت، برآورد لیو جک‌نایف تحت محدودیت‌های خطی تصادفی را به‌دست می‌آوریم. برای به‌دست آوردن این برآوردگر از روش ارائه شده توسط (۱) استفاده می‌کنیم.

برآورد $\hat{\theta}_r(d)$ را برای مدل کانونی (۱۰) درنظر بگیرید. روش جک‌نایف بر پایه‌ی حذف مشاهدات اجرا می‌شود در نتیجه برآورد $\hat{\theta}_r(d)$ را بعد از حذف مشاهده‌ی iام، یعنی $\hat{\theta}_{r(-i)}(d)$، به صورت زیر بدست می‌آوریم. با تعریف $\boldsymbol{XP}=\tilde{\boldsymbol{Z}}$ و $\boldsymbol{RP}=\tilde{\boldsymbol{R}}$ داریم:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\boldsymbol{r}(-i)}(d) &= (\boldsymbol{B}-\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i')^{-1}(\tilde{\boldsymbol{Z}}'\boldsymbol{Y}-\tilde{\boldsymbol{z}}_i y_i+\tilde{\boldsymbol{R}}'\boldsymbol{r}) = \left(\boldsymbol{B}^{-1}+\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}\right)(\tilde{\boldsymbol{Z}}'\boldsymbol{Y}+\tilde{\boldsymbol{R}}'\boldsymbol{r}-\tilde{\boldsymbol{z}}_i y_i)\\
&= \boldsymbol{B}^{-1}(\tilde{\boldsymbol{Z}}'\boldsymbol{Y}+\tilde{\boldsymbol{R}}'\boldsymbol{r})-\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i y_i+\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}(\tilde{\boldsymbol{Z}}'\boldsymbol{Y}+\tilde{\boldsymbol{R}}'\boldsymbol{r})-\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i y_i}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}\\
&= \hat{\boldsymbol{\theta}}_r(d)-\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i y_i\left(1+\frac{\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}\right)+\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'\hat{\boldsymbol{\theta}}_r(d)}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}\\
&= \hat{\boldsymbol{\theta}}_r(d)-\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\left(y_i-\tilde{\boldsymbol{z}}_i'\hat{\boldsymbol{\theta}}_r(d)\right)}{1-\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i}\\
&= \hat{\boldsymbol{\theta}}_r(d)-\frac{\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i\tilde{e}_i(d)}{1-\tilde{w}_i}\end{aligned}$$

$\hat{\theta}_{r(-i)}(d)$ را برآورد لیو تحت محدودیت وقتی که iامین مشاهده از مجموعه داده‌ها حذف می‌شود، می‌نامیم. $\tilde{\boldsymbol{Z}}_{(-i)}$ و $\boldsymbol{Y}_{(-i)}$ به ترتیب ماتریس $\tilde{\boldsymbol{Z}}$ و بردار $\boldsymbol{Y}$ بدون مشاهده‌ی iام هستند و $\tilde{w}_i=\tilde{\boldsymbol{z}}_i'\boldsymbol{B}^{-1}\tilde{\boldsymbol{z}}_i$ برابر iامین درایه روی قطر اصلی ماتریس $\tilde{\boldsymbol{W}}=\tilde{\boldsymbol{Z}}\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'$ است و $\tilde{e}_i(d)$ برابر iامین باقیمانده‌ی حاصل از برازش برآورد لیو تحت محدودیت است. همچنین معکوس ماتریس $\boldsymbol{B}$ به صورت زیر تعریف می‌شود:

$$\boldsymbol{B}^{-1}=(\boldsymbol{\Pi}+\boldsymbol{I})^{-1}(\boldsymbol{I}+d\boldsymbol{\Pi}^{-1})=\boldsymbol{G}_d\boldsymbol{\Pi}^{-1}$$

همچنین شبه مقادیر وزنی به صورت زیر تعریف می‌شوند:

$$Q_i^*=\hat{\boldsymbol{\theta}}_r(d)+n(1-\tilde{w}_i)(\hat{\boldsymbol{\theta}}_r(d)-\hat{\boldsymbol{\theta}}_{r(-i)}(d)) \qquad (11)$$

بنابراین، برآوردگر لیو جک‌نایف تحت محدویت‌های خطی تصادفی (۸) برای مدل (۱) به صورت زیر تعریف می‌شود:

$$\tilde{\boldsymbol{\theta}}_r(d)=\frac{1}{n}\sum_{i=1}^{n}Q_i^*$$

که با توجه به این نکته که $\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i'=\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}$ و $\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i y_i=\tilde{\boldsymbol{Z}}'\boldsymbol{Y}$، برآورد لیو جک‌نایف تحت محدودیت برای $\boldsymbol{\theta}$ به صورت زیر بدست می‌آید:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_r(d) &= \hat{\boldsymbol{\theta}}_r(d)+\boldsymbol{B}^{-1}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\tilde{e}_i(d)\\
&= \hat{\boldsymbol{\theta}}_r(d)+\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{Y}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}\boldsymbol{B}^{-1}\boldsymbol{U}'\tilde{\boldsymbol{Y}}=(\boldsymbol{I}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})\hat{\boldsymbol{\theta}}_r(d)+\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{Y}\end{aligned}$$

با توجه به رابطه‌ی اخیر، بردار اریبی برآورد لیو جک‌نایف تحت محدودیت عبارت است از:

$$Bias(\tilde{\boldsymbol{\theta}}_r(d))=E(\tilde{\boldsymbol{\theta}}_r(d))-\boldsymbol{\theta}=(\boldsymbol{I}-\boldsymbol{G}_d)(\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}-\boldsymbol{I})\boldsymbol{\theta}$$

همچنین ماتریس واریانس‌ـ کواریانس برآورد لیو جک‌نایف تحت محدودیت‌های خطی تصادفی به صورت

زیر حاصل می‌گردد:

$$\begin{aligned} Var\left[\tilde{\boldsymbol{\theta}}_r(d)\right] =& \sigma^2(\boldsymbol{I}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})\boldsymbol{G}_d\boldsymbol{\Pi}^{-1}\boldsymbol{G}_d'(\boldsymbol{I}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})' \\ &+\sigma^2(\boldsymbol{I}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})\boldsymbol{G}_d\boldsymbol{\Pi}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}\boldsymbol{B}^{-1} \\ &+\sigma^2\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}\boldsymbol{\Pi}^{-1}\boldsymbol{G}_d'(\boldsymbol{I}-\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})' \\ &+\sigma^2\boldsymbol{B}^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}}\boldsymbol{B}^{-1} \end{aligned}$$

همچنین با توجه به اینکه $\boldsymbol{\theta}=\boldsymbol{P}'\boldsymbol{\beta}$ است، در نتیجه برآوردهای آمیخته، لیو تحت محدودیت و لیو جک‌نایف تحت محدودیت برای مدل (۱) به ترتیب برابر $\hat{\boldsymbol{\beta}}_r=\boldsymbol{P}\hat{\boldsymbol{\theta}}_r$، $\hat{\boldsymbol{\beta}}_r(d)=\boldsymbol{P}\hat{\boldsymbol{\theta}}_r(d)$ و $\tilde{\boldsymbol{\beta}}_r(d)=\boldsymbol{P}\tilde{\boldsymbol{\theta}}_r(d)$ می‌باشند.

۴ مطالعه‌ی شبیه‌سازی

در این بخش به منظور ارزیابی عملکرد برآوردگر جدید لیو جک‌نایف تحت محدودیت، از یک مطالعه‌ی شبیه سازی استفاده می‌کنیم، که طی آن عملکرد برآوردهای کمترین توان‌های دوم معمولی، برآورد آمیخته، لیو، لیو تحت محدودیت، لیو جک‌نایف و لیو جک‌نایف تحت محدودیت را با توجه به دو معیار $MSEM$ و $ABIAS$ مقایسه خواهیم کرد. در این مطالعه داده‌های مربوط به متغیرهای توضیحی را با دنبال کردن کار (۴) تولید می‌کنیم.

در این آزمایش، برای پارامتر σ مقدار ۰٫۵ در نظر گرفته می‌شود. همچنین برای n دو مقدار ۳۰ و ۱۰۰ و برای p مقدار ۳ را در نظر می‌گیریم و آزمایش را ۱۰۰۰ بار تکرار کرده‌ایم. در مطالعه‌ی خود از معیارهای میانگین مربعات خطا و قدرمطلق اریبی برای بررسی عملکرد برآوردهای مختلف استفاده می‌کنیم، که به ترتیب به صورت زیر محاسبه می‌شوند:

$$MSE\left(\boldsymbol{b}\right)=\frac{1}{1000}\sum_{r=1}^{1000}\left(\boldsymbol{b}_r-\boldsymbol{\beta}\right)'\left(\boldsymbol{b}_r-\boldsymbol{\beta}\right)$$

$$ABIAS\left(\boldsymbol{b}\right)=\sum_{j=1}^{p}\left|\bar{b}_j-\beta_j\right|,\qquad \bar{b}_j=\frac{1}{1000}\sum_{r=1}^{1000}b_{j(r)}$$

که $\boldsymbol{b}_r$ برآورد $\boldsymbol{\beta}$ و $b_{j(r)}$ برآورد jامین مولفه‌ی بردار $\boldsymbol{\beta}$ در rامین تکرار آزمایش است. نتایج این شبیه‌سازی در جدول (۱) ارائه شده است.

در بیش‌تر موارد به‌ازای ترکیبات مختلف پارامترهای درنظر گرفته شده در این شبیه‌سازی، برآورد لیو جک‌نایف تحت محدودیت دارای کمترین مقدار قدرمطلق اریبی است. لذا می‌توان گفت روش جک‌نایف و اعمال محدودیت‌های خطی تصادفی برای برآورد لیو، موجب کاهش هر چه بیشتر اریبی برآورد موردنظر می‌شود. همچنین برآورد پیشنهادی عملکرد بهتری با توجه به معیار میانگین مربعات خطا نسبت به برآوردهای دیگر دارد.

باید اشاره کرد که به ازای مقادیر دیگر σ نیز شبیه‌سازی انجام شده اما به دلیل محدودیت در ارائه‌ی مطالب، نتایج تنها برای یک مقدار از σ آورده شده است.

جدول ۱: مقادیر MSE و $ABIAS$ برآوردهای مختلف به ازای $\sigma = ۰/۵$

n	برآورد	$\rho = ۰/۸$		$\rho = ۰/۹$		$\rho = ۰/۹۹$	
		MSE	$ABIAS$	MSE	$ABIAS$	MSE	$ABIAS$
$n = ۳۰$	$\hat{\boldsymbol{\beta}}$	۰/۰۴۵۳۲	۰/۰۱۰۰۸	۰/۱۲۴۵۹	۰/۰۰۵۱۲	۰/۹۴۸۹۵	۰/۰۱۵۱۵
	$\hat{\boldsymbol{\beta}}_r$	۰/۰۳۷۷۱	۰/۰۱۱۹۵	۰/۰۸۲۳۱	۰/۰۰۵۳۱	۰/۴۵۶۰۶	۰/۰۳۴۱۲
	$\hat{\boldsymbol{\beta}}(d)$	۰/۰۴۰۶۴	۰/۰۱۶۰۷	۰/۰۹۰۱۶	۰/۰۱۶۵۴	۰/۲۹۸۹۸	۰/۰۱۳۵۹
	$\hat{\boldsymbol{\beta}}_r(d)$	۰/۰۳۴۶۶	۰/۰۱۸۵۶	۰/۰۶۷۱۸	۰/۰۱۳۶۹	۰/۱۹۶۲۱	۰/۰۳۵۴۰
	$\tilde{\boldsymbol{\beta}}(d)$	۰/۰۴۱۰۰	۰/۰۱۰۰۶	۰/۰۹۰۰۷	۰/۰۰۵۱۰	۰/۲۸۵۸۱	۰/۰۰۹۹۷
	$\tilde{\boldsymbol{\beta}}_r(d)$	۰/۰۳۴۹۴	۰/۰۰۹۲۱	۰/۰۶۵۹۱	۰/۰۰۴۴۵	۰/۱۲۰۵۳	۰/۰۰۷۲۹
$n = ۱۰۰$	$\hat{\boldsymbol{\beta}}$	۰/۰۱۴۸۱	۰/۰۰۶۷۶	۰/۰۳۱۲۹	۰/۰۰۱۷۸	۰/۲۳۴۱۷	۰/۰۱۱۹۰
	$\hat{\boldsymbol{\beta}}_r$	۰/۰۱۴۳۶	۰/۰۰۵۳۱	۰/۰۲۹۷۵	۰/۰۰۱۲۴	۰/۱۶۷۷۸	۰/۰۱۱۸۳
	$\hat{\boldsymbol{\beta}}(d)$	۰/۰۱۴۲۵	۰/۰۰۸۷۱	۰/۰۲۸۸۰	۰/۰۰۴۳۳	۰/۱۳۶۵۰	۰/۰۰۷۸۶
	$\hat{\boldsymbol{\beta}}_r(d)$	۰/۰۱۳۸۱	۰/۰۰۷۲۴	۰/۰۲۷۵۴	۰/۰۰۴۲۸	۰/۱۰۸۳۳	۰/۰۰۸۹۰
	$\tilde{\boldsymbol{\beta}}(d)$	۰/۰۱۴۲۰	۰/۰۰۶۷۶	۰/۰۲۱۱۷	۰/۰۰۱۷۶	۰/۱۰۷۰۲	۰/۰۰۷۱۸
	$\tilde{\boldsymbol{\beta}}_r(d)$	۰/۰۱۳۷۳	۰/۰۰۵۷۳	۰/۰۲۰۴۴	۰/۰۰۱۱۸	۰/۱۰۳۴۷	۰/۰۰۶۹۹

بحث و نتیجه‌گیری

در این مقاله، برآورد لیو در مدل‌های رگرسیونی خطی مورد مطالعه قرار گرفت. این برآوردگر برای کاهش اثرات هم‌خطی در مدل‌های رگرسیونی به‌کار می‌رود. سپس برآورد لیو جک‌نایف بررسی شد. از سوی دیگر، تحقیقات انجام شده نشان می‌دهد که به‌کارگیری اطلاعات کمکی افزون بر نمونه‌ی مورد مطالعه، موجب بهبود روند برآورد می‌شود. از این رو با استفاده از این اطلاعات کمکی به فرم محدودیت‌های خطی تصادفی، برآوردگر جدیدی تحت عنوان برآورد لیو تحت محدودیت معرفی گردید. در پایان با اعمال روش جک‌نایف بر برآورد لیو تحت محدودیت، به برآورد لیو جک‌نایف تحت محدودیت دست یافتیم. همچنین با استفاده از یک مطالعه‌ی شبیه‌سازی نشان دادیم که برآوردگر پیشنهادی این مقاله عملکرد بهتری را با توجه به دو معیار میانگین مربعات خطا و قدرمطلق اریبی دارد.

مراجع

[1] Akdeniz Duran, E., Akdeniz, F. (2012). Efficiency of the modified jackknifed Liu-type estimator. *Stat Papers*, **53**, 265-280.

[2] Liu, K. (1993). A new class of biased estimate in linear regression, *Communications in Statistics - Theory and Methods*, **22**(2) 393–402.

[3] Li, Y. and Yang, H. (2010). A new ridge-type estimator in stochastic restricted linear regression. *Statistics: A Journal of Theoretical and Applied Statistics*, **45:2**, 123-130.

[4] McDonald, G.C. and Galarneau, D.I. (1975). A monte carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, **70** 407–416.

[5] Rao, C. R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives*, 3rd, Berlin: Springer.

[6] Yildiz, N. (2017). On the performance of the Jackknifed Liu-type estimator in linear regression model. *Communications in Statistics - Theory and Methods*, **47**, 2278-2290.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

مقایسه توان آزمون‌های آنالیز واریانس، کروسکال-والیس و ولش در داده‌های غیر نرمال

طالبی، س [۱] جام بر سنگ، س [۲]

[۱،۲] دانشگاه علوم پزشکی شهید صدوقی یزد

چکیده

در این مقاله مقایسه آزمون‌های آنالیز واریانس، کروسکال-والیس، ولش با استفاده از توان و خطای نوع اول زمانیکه فرض نرمال بودن توزیع متغیرها نقض می‌شود، صورت می‌گیرد. از شبیه‌سازی مونت کارلو در سناریوهای مختلف برای بررسی و مقایسه توان آزمون‌ها در شرایط مختلف استفاده شده است و به نظر می رسد که برای توزیع‌های نامتقارن آزمون ناپارامتری کروسکال والیس از معادل پارامتری آن بهتر عمل می‌کند.

کلمات کلیدی: آنالیز واریانس، کروسکال-والیس، ولش، توان آزمون .

۱ پیش‌گفتار

تجزیه و تحلیل واریانس (ANOVA) یکی از روش‌های آماری است که بیشترین استفاده را در تحقیقات پزشکی دارد (۳)، که برای مقایسه میانگین یک صفت کمی در بیش از دو گروه استفاده می‌شود. یک مدل ANOVA فرض می‌کند که:

۱. از هر جامعه نمونه‌های تصادفی مستقل گرفته شده باشند.

۲. صفت مورد بررسی در هریك از جامعه‌ها دارای توزیع نرمال باشند. یعنی اندازه های متغیر وابسته در هر یك از سطوح متغیر فاکتور، دارای توزیع نرمال باشند.

[۱] Samanehtalebi1396@gmail.com

۳. واریانس صفت مورد بررسی در همه جامعه‌ها برابر باشند. یعنی واریانس متغیر وابسته در همه سطوح متغیر فاکتور، با هم مساوی باشند.

توجه داشته باشید که با نرمال بودن توزیع‌های احتمالی و ثابت بودن واریانس‌ها ، توزیع‌های احتمالی تنها با توجه به میانگین آنها متفاوت هستند. آزمون آنالیز واریانس زمانیکه تمام پیش‌نیازها برآورده شوند دارای توان بالایی است. مدل‌های آنالیز واریانس نسبت به انواع خاصی از انحراف از مدل مقاوم هستند، مانند زمانی‌که داده‌ها دارای توزیع نرمال نیستند(۴). با این حال، ناهمگونی واریانس‌ها در داده‌هایی به طور نرمال توزیع شده اند ممکن است منجر به افزایش میزان خطای نوع اول شود (۹).
روش‌های تحلیل و استنباط آماری به دو دسته‌ی آمار پارامتری و آمار ناپارامتری تقسیم می‌شوند. روش‌های آمار پارامتری روش‌هایی هستند که مستلزم معلوم بودن توزیع جامعه هستند. بنابراین در صورتی که توزیع جامعه از الگوی خاصی مثل توزیع نرمال پیروی کند، می‌توان از این روش‌ها استفاده کرد. اما روش‌های آمار ناپارامتری آزاد توزیع بوده و به الگوی جامعه بستگی ندارند.
نتر و همکارانش به‌طور خلاصه بیان می‌کنند که اگر داده‌ها به صورت نرمال توزیع شده باشند اما واریانس ها ثابت نباشند، معیار اصلاحی استاندارد استفاده از روش حداقل مربعات وزنی است. اگر داده‌ها دارای توزیع نرمال نباشند و واریانس‌ها ثابت نباشند، استفاده از تبدیل باکس_کاکس مناسب است. وقتی تبدیلات در تثبیت واریانس‌ها و نزدیک کردن توزیع داده‌ها به توزیع نرمال موفق نیستند، روش‌های ناپارامتری برای مقایسه میانگین ها استفاده می‌شوند (۴).
هدف ما ارزیابی این است که کدام روش‌ها زمانیکه با ناهمگونی واریانس مواجه هستیم توان بالاتری دارند. بنابراین در این مطالعه ما روش‌های ANOVA، Welch ANOVA و Kruskal-Wallis را به‌وسیله‌ی محاسبه میزان خطای نوع اول و توان هریک تحت حالت‌های زیر با هم مقایسه می‌کنیم:

[:=label]وقتی داده‌ها دارای توزیع نرمال (توزیع غیر نرمال) باشند. وقتی حجم نمونه‌های گروه‌های مقایسه مساوی باشند. (طرح متعادل باشد)

۲ روش‌های آماری:

این سه روش در صورت ناهمگونی واریانس سه گروه مورد بررسی قرار می گیرند.

۱.۲ آزمون F در تجزیه و تحلیل واریانس

الف) آنالیز واریانس یك طرفه وقتی حجم نمونه‌ها مساوی باشد:
در اینجا می‌خواهیم آزمون فرضیهی تساوی میانگینهای بیش از دو جامعه را در مقابل فرضیه‌ی جانشین «عدم تساوی میانگین‌ها» انجام دهیم. یعنی برای $k\geq 1$ (k تعداد جامعه‌ها) داریم :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

$$H_a : \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_k$$

در این روش، ابتدا واریانس کل را به دو مؤلفه‌ی پراکندگی بین گروه‌ها و داخل گروهها تجزیه می‌کنم. بنابراین به این روش آنالیز (تحلیل) واریانس گفته می‌شود. روش آنالیز واریانس را به اختصار با ANOVA نمایش میدهند که مخفف عبارت Analyze Of Variance است.
F-test در آنالیز واریانس به شرح زیر است:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

$$j = 1, 2, ..., n_i \; ; \; i = 1, 2, \ldots, k$$

Y_{ij} مقدار متغیر پاسخ برای j امین مشاهده در i امین گروه است.
μ_i میانگین مشاهدات i امین گروه است.
$\varepsilon_{ij} \sim N(۰, \sigma^۲)$ و مستقلند.
اگر برای مقادیر مختلف i ، n_i ثابت باشد، در این صورت طرح را طرح متعادل می‌نامند.
میانگین نمونه برای گروه i با $\bar{Y}_{i.} = \frac{\sum_{j=۱}^{n_i} Y_{ij}}{n_i}$ نشان داده شده می‌شود.
مجموع همه مشاهدات با $Y_{..} = \sum_{i=۱}^{k} \sum_{j=۱}^{n_i} Y_{ij}$ نشان داده می‌شود.
تحلیل واریانس به معنای تفکیک کل تغییرات به اجزایی است که این تغییرات از آنها ناشی میشود. به منظور تجزیه و تحلیل تفاوت بین گروه‌ها کل مجموع مربعات تصحیح شده(SSTO) محاسبه وبه دو جزء SSTR مجموع مربعات مربوط به گروه‌ها (بین گروه‌ها) و SSE مجموع مربعات مربوط به خطا (درون گروه‌ها) تقسیم می‌شود.

$$\text{SSTO} = \text{SSTR} + \text{SSE}$$

$$\sum_i \sum_j \left(Y_{IJ} - \bar{Y}_{..}\right)^2 = \sum_i n_i \left(\overline{Y_{i.}} - \bar{Y}_{..}\right)^2 + \sum_i \sum_j \left(Y_{ij} - \bar{Y}_{i.}\right)^2$$

که SSTO ، $n_t - 1$ درجه آزادی دارد. k گروه داریم بنابراین SSTR، k-1 درجه آزادی دارد و بلاخره درون هر گروه n_i تکرار موجود است که هر کدام $n_i - 1$ درجه‌ی آزادی را برای برآورد خطای آزمایش می‌دهند و چون k گروه وجود دارند پس جمعا $n_t - k$ درجه‌ی آزادی برای خطا داریم. و آماره‌ی آزمون عبارت است از:

$$F^* = \frac{MSTR}{MSE} = \frac{SSTR/_{k-1}}{SSE/_{n_t-k}} \sim F_{\text{k}-1 \; n_t-k}$$

اگر $F^* > F_{1-\alpha;\text{k}-1;\; n_t-k}$ آنگاه Ha نتیجه می‌شود در غیر اینصورت H0 نتیجه می‌شود. که در آن α بعنوان خطای نوع اول نامیده می‌شود و $F_{1-\alpha;\text{k}-1;\; n_t-k}$ نقطه $(1-\alpha)$ درصد توزیع فیشر است (۴)(۱).

F-test Welch's ۲.۲

این آزمون طراحی شده برای مقایسه میانگین‌های بیش از دو گروه به ویژه در مواردی که حجم نمونه‌ها کوچک هستند و واریانس گروه‌ها ناهمگن می باشند، در حالیکه مفروضات نرمال بودن و مستقل بودن برقرار هستند

(۸). در آزمون Welch برای کاهش اثر ناهمگونی از وزن $w_i = \frac{n_i}{s_i^2}$ استفاده می‌شود. بنابراین میانگین کل تعدیل شده و بصورت زیر محاسبه می‌شود:

$$\bar{Y}_{welch} = \frac{\sum_{i=1}^{\mathrm{k}} w_i \bar{Y}_{i\cdot}}{\sum_{i=1}^{\mathrm{k}} w_i}$$

که $\bar{Y}_i$ میانگین نمونه‌ای برای i امین گروه است. مجموع مربعات تیمار و میانگین مربعات تیمار عبارتند از:

$$SSTR_{welch} = \sum_{i=1}^{k} w_i \left(\bar{Y}_{i\cdot} - \bar{Y}_{welch}\right)$$

$$\mathrm{MSTR}_{welch} = \frac{SSTR_{welch}}{k-1}$$

سپس باید عبارت لامبدا براساس فرمول زیر محاسبه گردد:

$$\Lambda = \frac{3\sum_{i=1}^{k} \frac{\left(1 - \frac{w_i}{\sum_{i=1}^{k} w_i}\right)^2}{n_i - 1}}{k^2 - 1}$$

و آماره‌ی آزمون مربوطه به صورت زیر است:

$$F^*_{welch} = \frac{SSTR_{welch}/k-1}{1 + \frac{2\,\Lambda(\mathrm{k}-2)}{3}} \sim F_{\mathrm{k}-1\ \frac{1}{\Lambda}}$$

قاعده تصمیم گیری با در نظر گرفتن سطح معنی داری α، اگر $F^* > F_{1-\alpha;\mathrm{k}-1;\ \frac{1}{\Lambda}}$ آنگاه Ha نتیجه می‌شود در غیر اینصورت H0 نتیجه می‌شود (۶).

۳.۲ Kruskal-Wallis

آزمون کروسکال-والیس یك آزمون ناپارامتری است و برای مقایسه‌ی میانگین بیش از دو جامعه‌ی مستقل استفاده می‌شود، از آنجایی که این آزمون یک آزمون ناپارامتری است، فرض نمی‌کند که متغیر پاسخ به طور نرمال توزیع شود. و مانند بسیاری از آزمون‌های ناپارامتری، بر روی داده‌های با مقیاس رتبه‌ای یا زمانی که حجم نمونه کوچک باشد انجام می‌شود (۷)،(۲).

برای انجام این آزمون ابتدا تمامی داده‌های اخذ شده از جوامع مختلف به عنوان یک نمونه در نظر گرفته سپس رتبه بندی می‌شوند، در صورت وجود مشاهدات تکراری میانگین رتبه‌های آن مشاهدات باید در نظر گرفته شود.

R_{ij}رتبه j امین مشاهده از i امین گروه است که به عنوان متغیر پاسخ در نظر گرفته می‌شود. آماره‌ی این آزمون عبارت است از:

$$K = (n_t - 1)\ \frac{\sum_{i=1}^{k} n_i \left(\bar{R}_{i\cdot} - \bar{R}\right)^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(R_{ij} - \bar{R}\right)^2} \sim F_{k-1.\ n_t-1}$$

k : تعداد گروه ها

n_t : تعداد کل مشاهدات

n_i : تعداد مشاهدات i امین گروه

$\bar{R}_{i\cdot} = \frac{\sum_{j=1}^{n_i} R_{ij}}{n_i}$ میانگین رتبه‌های گروه i ام

$\bar{R} = \frac{1}{2}(n_t + 1)$ میانگین رتبه همه مشاهدات

قاعده تصمیم گیری با در نظر گرفتن سطح معنی داری α، اگر $K > F_{1-\alpha;k-1;\, n_t-k}$ آنگاه Ha نتیجه می‌شود در غیر اینصورت H0 نتیجه می‌شود.(Liu، ۲۰۱۵)

۴.۲ مواد و روش‌ها:

دراین مطالعه ۱۰۰۰۰ عدد تصادفی با تکنیک شبیه سازی مونت کارلو با استفاده از نرم افزار $R_{3.5.1}$ از سناریوهای زیر زمانی که تعداد گروه‌ها (k=۳) تولید شده‌اند.

$$\mathrm{N}(\mu_1{:}\mu_2{:}\mu_3{=}0{:}1{:}2,\sigma_1^2{:}\sigma_2^2{:}\sigma_3^2{=}1{:}1{:}2);$$

$$\mathrm{lnN}(\mu_1{:}\mu_2{:}\mu_3{=}0{:}1{:}2,\sigma_1^2{:}\sigma_2^2{:}\sigma_3^2{=}1{:}1{:}2);$$

$$\text{Beta type I}(\alpha_1{:}\alpha_2{:}\alpha_3{=}0.25,\ 0.5,\ 0.75\ ,\ \beta_1{:}\beta_2{:}\beta_3{=}0.5{,}0.5,\ 0.5);$$

$$\text{Beta type II}(\alpha_1{:}\alpha_2{:}\alpha_3{=}0.5,\ 1.5,\ 2{,}\beta_1{:}\beta_2{:}\beta_3{=}5,\ 5,\ 5);$$

$$\text{Beta type III}(\alpha_1{:}\alpha_2{:}\alpha_3{=}5,\ 5,\ 5\ {,}\beta_1{:}\beta_2{:}\beta_3{=}0.25{,}0.5,\ 0.75).$$

حجم نمونه در طرح متعادل به صورت زیر بوده است.

$$(n_1 : n_2 : n_3 = \ 5:5:5,\ 10:10:10,\ 20:20:20,\ 80:80:80,$$

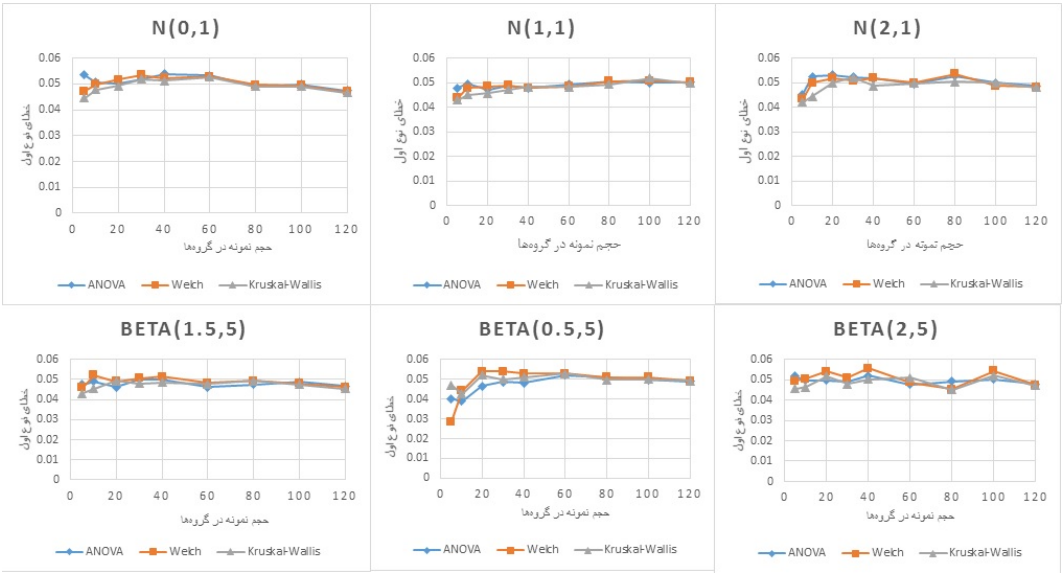
$$30:30:30,\ 40:40:40,\ 60:60:60,\ 100:100:100,\ 120:120:120)$$

۳ یافته ها:

شکل‌های ۱ـ۴ نمودارهای مربوط به مقایسه میزان خطای نوع اول و توان آزمون برای حالت‌های مختلف نمایش می‌دهند. زمانی‌که فرضیه صفر درست است، میزان رد فرضیه صفر به عنوان خطای نوع اول تجربی در نظر گرفته می‌شود و هر روشی که میزان خطای نوع اول تجربی آن به $\alpha = 0.05$ نزدیک تر باشد، بهترین روش است. زمانی‌که فرضیه مقابل درست است، میزان رد فرضیه صفر توان آزمون را نشان می‌دهد. روشی که بالاترین توان را داشته باشد بهترین روش است. در شکل ۱ برای توزیع نرمال زمانی‌که طرح متعادل و حجم نمونه گروه‌ها کوچک $(n_1 : n_2 : n_3 = \ 5:5:5)$ است، آزمون کروسکال‌ـوالیس انحراف بیشتری از خطای نوع اول (۰,۰۵) دارد و مقادیر خطای نوع اول تجربی برای آزمون آنالیز واریانس حول ۰,۰۵ است. در شکل ۴ برای توزیع نرمال در حجم نمونه کوچک $(n_1 : n_2 : n_3 = \ 5:5:5)$ آزمون آنالیز واریانس از توان بیشتری در مقایسه با دو آزمون دیگر برخوردار است اما از حجم نمونه $(n_1 : n_2 : n_3 = \ 20:20:20)$ به بعد

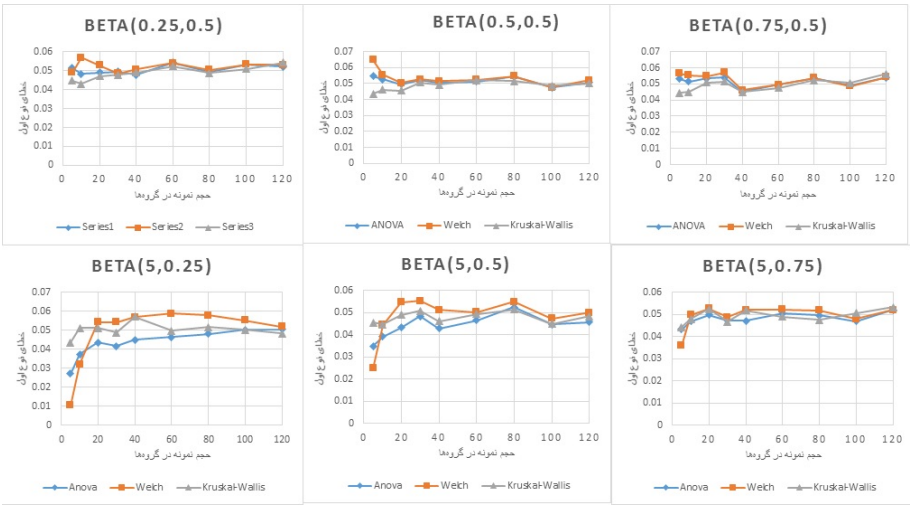
توان آزمون‌ها برابر هستند.
برای توزیع Beta(0.5,5) که چولگی آن ۱,۹ است زمانی‌که حجم نمونه گروه‌ها برابر $(n_1 : n_2 : n_3 = 5 : 5 : 5)$ ، آزمون ولش بیشترین انحراف را از خطای نوع اول (۰,۰۵) دارد و مقدار خطای نوع اول تجربی برای آزمون کروسکال‌ـوالیس نزدیک ۰,۰۵ است، و در حجم نمونه بالاتر از $(n_1 : n_2 : n_3 = 60 : 60 : 60)$ رفتار آزمون‌ها مشابه یکدیگر است. اما برای توزیع‌های Beta(۱,۵،۵) و Beta(۲،۵) که چولگی آنها به ترتیب ۰,۹۵ و ۰,۵۸ می‌باشد مقادیر خطای نوع اول تجربی برای آزمون آنالیز واریانس نزدیک ۰,۰۵ است. همچنین در شکل ۴برای مقایسه توان آزمون‌ها در این سه توزیع نمودار Beta type I رسم شده که آزمون کروسکال‌ـوالیس تا حجم نمونه $(n_1 : n_2 : n_3 = 60 : 60 : 60)$ دارای توان بالاتری است و از حجم نمونه بالاتر از ۶۰ رفتار آزمون‌ها از نظر توان مشابه یکدیگر است و نتیجه قبلی کمتر نمود دارد.
برای توزیع‌های شکل ۵ زمانیکه $n_۱ : n_۲ : n_۳$ کوچکتر از ۳۰ هستند تفاوت در میزان خطای نوع اول تجربی نمود بیشتری دارد. همانطور که در شکل ۲ وقتی $(n_1 : n_2 : n_3 = 5 : 5 : 5)$ در توزیع Beta(۰,۲۵،۰,۵) مقدار خطای نوع اول تجربی آزمون کروسکال‌ـوالیس انحراف بیشتری از خطای نوع اول (۰,۰۵) دارد و زمانیکه حجم نمونه گروه‌ها $(n_1 : n_2 : n_3 = 10 : 10 : 10)$ مقادیرخطای نوع اول تجربی آزمون‌های کروسکال‌ـوالیس و ولش از خطای نوع اول (۰,۰۵) انحراف دارند. برای توزیع Beta(0. 5,0.5) در $(n_1 : n_2 : n_3 = 10 : 10 : 10,5 : 5 : 5)$ مقادیر خطای نوع اول تجربی آزمون ولش انحراف بیشتری دارد و در $(n_1 : n_2 : n_3 = 20 : 20 : 20)$ مقدار خطای نوع اول تجربی آزمون کروسکال‌ـوالیس از خطای نوع اول (۰,۰۵) انحراف دارد. برای توزیع Beta(0.7 5,0.5) وقتی $(n_1 : n_2 : n_3 = 5 : 5 : 5)$ مقادیرخطای نوع اول تجربی آزمون‌های کروسکال‌ـوالیس و ولش از خطای نوع اول (۰,۰۵) انحراف دارند. با توجه به شکل ۴ که نمودار مربوط به مقایسه توان این سه توزیع تحت عنوان Beta type II رسم شده در حجم نمونه کمتر از ۲۰ آزمون کروسکال‌ـوالیس توان بالاتری دارد اما بعد از آن توان سه آزمون روند یکسانی دارند، که با توجه به دو قله‌ای بودن نمودار تابع چگالی توزیع‌ها میانگین چولگی‌های هر توزیع به ترتیب مقادیر ۰,۶۷، ۰ ، ۰,۴۳۶ـ بدست آمدند و از آنجایی که این مقادیر نزدیک به صفر هستند، توان آزمونها در این حالت تفاوت قابل توجهی ندارند.
در شکل ۲ برای توزیع Beta(5,0.25) که مقدار چولگی آن ـ۲,۷۴ بدست آمده، در $(n_1 : n_2 : n_3 = 5 : 5 : 5)$ مقادیر خطای نوع اول تجربی آزمون‌های ولش و آنالیز واریانس بیشترین انحراف از خطای نوع اول (۰,۰۵) دارند و مقادیر خطای نوع اول تجربی آزمون کروسکال‌ـوالیس تقریبا در تمامی حجم‌ها به ۰,۰۵ نزدیک‌تر است. برای توزیع Beta(5,0.5) با چولگی ـ۲,۰۲ ، مقادیر خطای نوع اول تجربی آزمون کروسکال‌ـوالیس تقریبا نزدیک به ۰,۰۵ هستند. برای توزیع Beta(5,0.75) با چولگی ـ۱,۵۵ ، آزمون ولش بیشترین انحراف را از خطای نوع اول (۰,۰۵) دارد. با توجه به شکل ۴ در نمودار Beta type III آزمون کروسکال‌ـوالیس توان بیشتری در مقایسه با دو آزمون دیگر بخصوص در حجم نمونه زیر ۶۰ داشته است.
در شکل ۳ برای توزیع lognormal(0,1) که در مقایسه با توزیع نرمال چگالی آن در دنباله‌ها بیشتر است، مقادیر خطای نوع اول تجربی آزمون آنالیز واریانس در حجم نمونه‌های مختلف در گروه‌ها بیشترین انحراف از خطای نوع اول (۰,۰۵) دارد اما در $(n_1 : n_2 : n_3 = 5 : 5 : 5)$آزمون ولش بیشترین انحراف از خطای نوع اول داشته است. برای lognormal(1,1) انحراف مقدار خطای نوع اول تجربی آزمون ولش برای

$(n_۱ : n_۲ : n_۳ = ۵ : ۵ : ۵)$ در مقایسه با lognormal(0,1) کاهش پیدا کرده است اما همچنان مقادیر خطای نوع اول تجربی آزمون آنالیز واریانس از خطای نوع اول (۰،۰۵) انحراف دارند. برای توزیع lognormal(2,2) مقادیر خطای نوع اول تجربی آزمون کروسکال-والیس مشابه حالت‌های قبل کمترین انحراف از خطای نوع اول (۰،۰۵) داشته‌اند. همچنین در شکل ۴ آزمون کروسکال-والیس بیشترین توان آزمون را در حجم نمونه‌های کمتر از ۴۰ در گروه‌ها داشته است و بعداز $(n_1 : n_2 : n_3 = 40 : 40 : 40)$ نمودار توان آزمون‌های کروسکال-والیس و ولش روند مشابهی را دنبال می‌کنند.

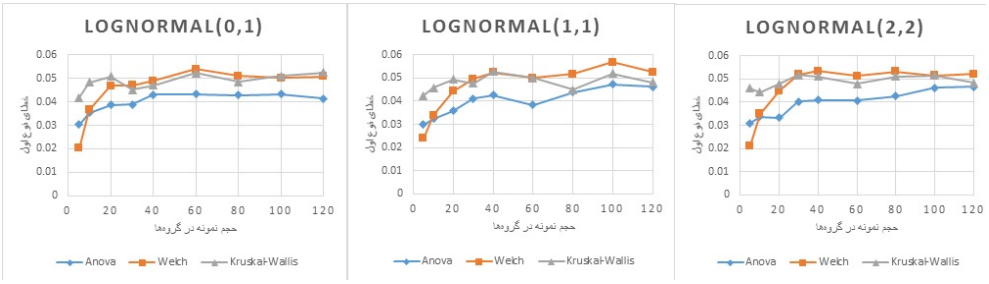


۲.

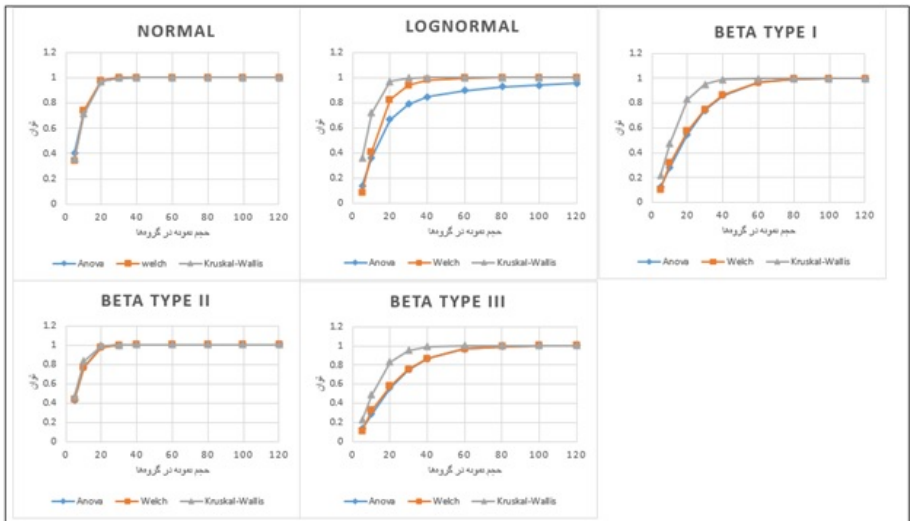
شکل ۱: مقایسه میزان خطای نوع اول تجربی برای توزیع‌های مختلف در طرح متعادل



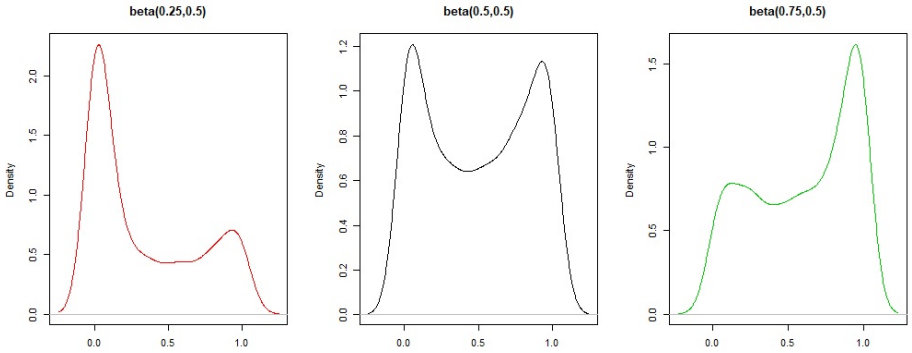
شکل ۲: مقایسه میزان خطای نوع اول تجربی برای توزیع‌های مختلف در طرح متعادل



شکل ۳: مقایسه میزان خطای نوع اول تجربی برای توزیع‌های مختلف در طرح متعادل



شکل ۴: مقایسه میزان توان آزمون‌ها برای توزیع‌های مختلف در طرح متعادل



شکل ۵: توزیع بتا دو قله‌ای

بحث و نتیجه‌گیری

امروز استفاده از توزیع‌های چوله که توانایی بیشتری در بیان رفتار داده‌های مشاهده شده دارند، افزایش چشمگیری یافته است. این توزیع‌ها در علوم مختلفی مانند علوم پزشکی و علوم رفتاری که مقادیر واقعی

متغیرهای تصادفی دارای توزیعی غیر متقارن است مورد استفاده قرار می‌گیرد. بنابراین استفاده از آزمون‌های با توان بالا برای چنین داده‌هایی اهمیت دارد. همانطور که در قسمت قبل نتیجه شد آزمون کروسکال‌ـ والیس زمانی که حجم نمونه در گروه‌های مقایسه کوچک است و توزیع داده‌ها یک توزیع چوله است دارای بالاترین توان می‌باشد حتی برای توزیعهایی مانند توزیع لگ نرمال که مانند توزیع نرمال متقارن اما با چگالی بیشتر در دنباله‌ها هستند.

مراجع

[1] Daniel, W. W. and Cross, C. L. (2018) Biostatistics: a foundation for analysis in the health sciences. Wiley.

[2] Hecke, T. Van (2012) 'Power study of anova versus Kruskal-Wallis test', Journal of Statistics and Management Systems. Taylor & Francis, 15(2–3), pp. 241–247.

[3] Kim, T. K. (2017) 'Understanding one-way ANOVA using conceptual figures', Korean journal of anesthesiology. Korean Society of Anesthesiologists, 70(1), p. 22.

[4] Kutner, M. H. et al. (2005) Applied linear statistical models. McGraw-Hill Irwin Boston.

[5] Levy, K. J. (1978) 'An empirical comparison of the ANOVA F-test with alternatives which are more robust against heterogeneity of variance', Journal of statistical computation and simulation. Taylor & Francis, 8(1), pp. 49–57.

[6] Liu, H. (2015) 'Comparing Welch's ANOVA, a Kruskal-Wallis test and traditional ANOVA in case of Heterogeneity of Variance'.

[7] McKight, P. E. and Najab, J. (2010) 'Kruskal-wallis test', The corsini encyclopedia of psychology. Wiley Online Library, p. 1.

[8] Mendeş, M. and Akkartal, E. (2010) 'Comparison of ANOVA F and WELCH tests with their respective permutation versions in terms of type I error rates and test power', Kafkas Univ Vet Fak Derg, 16(5), pp. 711–716.

[9] Moder, K. (2007) 'How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic', Austrian Journal of Statistics, 36(3), pp. 179–188.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

انتخاب توزیع پیشین در استنباط بیزی

اروجی، آ [۱] اکبری، ن [۲] فکور، و [۳]

[۱،۲]گروه اپیدمیولوژی و آمارزیستی ، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد
[۳] گروه آمار، دانشگاه فردوسی مشهد

چکیده

همواره یکی از نکات مهم در استنباط بیزی، انتخاب توزیع پیشین مناسب برای یک مدل بیزی است. در این مقاله به بررسی این موضوع می پردازیم. هدف این مقاله تقسیم بندی پیشین ها به دو دسته و بررسی نقش و کاربرد هر دسته در آمار بیز است. همچنین با ارائه چند مثال به پیامدهای انتخاب پیشین نامناسب می پردازیم.

کلمات کلیدی: استنباط بیزی، پارادوکس لیندلی-جفریز، پیشین‌های مرجع، توزیع پیشین.

۱ پیش‌گفتار

برای یک مدل آماری که برای داده‌ها در نظر گرفته می‌شود، رهیافت بیزی یک مدل احتمالی اضافی دیگری را برای پارامترهای مجهول در مدل داده‌ها، اختیار می‌کند. یک رهیافت مدل‌بندی عدم قطعیت درباره پارامترها با استفاده از اطلاعات متخصص است که این اطلاع را *اطلاع پیشین* می‌نامند.

اطلاع متخصص باید مستقل از داده‌ها به دست آمده و تحلیل شود. یک راه تضمین این مساله، جمع‌آوری اطلاع پیشین قبل جمع‌آوری داده‌ها است، هرچند کلمه پیشین لزوماً اشاره به این مطلب نیست و مستقل بودن نظر متخصص و تحلیل آن از جمع‌آوری داده‌ها مهم است.

با الگو گرفتن از منبع (۳) به غیر از اطلاع پیشین، دسته دیگری از مدل‌های احتمالاتی را برای پارامترهای مجهول توزیع داده‌ها در نظر می‌گیریم که به آنها *پیشین‌های مرجع* می‌گویند. این پیشین‌های مرجع اساساً پایه‌ای

[۱] arezoo.orooji23@gmail.com

برای بحث و مقایسه هستند و نوعاً، اطلاعات پیشین کمی را مشارکت می‌دهند تا حدی که استفاده از آنها را می‌توان معادل در نظر گرفتن رهیافت‌های غیر بیزی دانست.

هدف این مقاله بررسی این دو دسته از مدل‌های احتمالاتی یعنی *اطلاع پیشین* و *پیشین‌های مرجع* در استنباط بیزی است.

۲ انواع پیشین

قبل از بیان دو نوع دسته‌بندی این نکته را یادآور می‌شویم که از دید منتقدین، شیوه گزینش توزیع‌هاي پیشین به عنوان یک نقطهٔ ضعف استنباط بیزي قلمداد می‌شود. از طرف دیگر، از دید طرفداران این مکتب آماری، در استنباط بیزی قابلیت در نظر گرفتن توزیع‌های پیشین به این معنی است که اطلاعات بیشتري را می‌توان در استنباط لحاظ کرد. در ادامه به تقسیم‌بندی پیشین‌ها بر اساس نقش و کاربرد آنها می‌پردازیم.

۱.۲ پیشین‌های مرجع

همانطور که بیان شد این نوع پیشین‌ها فاقد اطلاعات در مدل‌بندی هستند. برای اطلاع بیشتر درباره این پیشین‌ها می‌توان به منبع (۳) مراجعه کرد.

۱.۱.۲ پیشین تخت

فرض کنید y دارای توزیع $\mathrm{Bin}(n.\theta)$ باشد. گاهی تصور می‌شود که در نظر گرفتن توزیع یکنواخت برای θ نشان از نادیده گرفتن مقدار θ است. با این حال، همانطور که توسط رایفا[۱] و شلایفر[۲] (۱۹۶۱) (۱) بحث شده است، اگر شما در مورد θ ناآگاه باشید در مورد $\theta^{۲}$ هم ناآگاه خواهید بود و نمی‌توانید توزیعی پیدا کنید که هم روی θ و هم روی $\theta^{۲}$ یکنواخت باشد.

با این وجود برای پارامتر یک متغیره θ که می‌تواند مقادیر روی کل خط حقیقی را اختیار کند، اغلب پیشین‌های یکنواخت را در نظر می‌گیرند یعنی $p(\theta) = ۱$. این مهم نیست که شما $p(\theta) = ۱$ را درنظر بگیرید یا $p(\theta) = ۱۲۳۴$، نکته اینجاست که پیشین تخت است و پیشین‌های تخت روی $\mathbb{R}$ ناسره هستند، یعنی $\int p(\theta)\, d\theta = \infty$. در چگالی‌های ناسره اگر حدود انتگرال را بجای $(-\infty, +\infty)$ یک مجموعه کراندار مانند $(-M, M)$ درنظر بگیریم آنگاه انتگرال روی این مجموعه کمتر از ∞ می‌شود. پیشین‌های تخت اغلب با در نظر گرفتن یک پیشین با واریانس بزرگ تقریب زده می‌شود.

یکی از مشکلات پیشین تخت این است که اگر توزیع نمونه‌گیری $f(y \mid \theta)$ باشد و از یک پیشین تخت یا هر پیشین ناسره‌ای استفاده کنیم، چگالی حاشیه‌ای داده‌ها $f(y) = \int f(y \mid \theta) d\theta$ اغلب وجود ندارد.

حسن پیشین‌های تخت این است که در بسیاری از مسایل پیشین تخت توسط داده‌ها منکوب می‌شود.

[۱] Raiffa

[۲] Schlaifer

به عنوان مثال با فرض $p(\theta) = K$، برای ثابتی مانند K، قضیه بیز به صورت زیر نوشته می‌شود:

$$\begin{aligned} p(\theta \mid y) &= \frac{L(\theta \mid y)p(\theta)}{\int L(\theta \mid y)p(\theta)d\theta} \\ &= \frac{L(\theta \mid y)K}{\int L(\theta \mid y)Kd\theta} \\ &= \frac{L(\theta \mid y)}{\int L(\theta \mid y)d\theta}, \end{aligned}$$

همانطور که دیده می‌شود اطلاعات پیشین هیچ نقشی در پسین ندارد. پسین یک درستنمایی نرمال شده برای θ است (همه درستنمایی‌ها نمی‌توانند نرمال شوند چون همه آن‌ها دارای انتگرال محدود نسبت به θ نمی‌باشند). با استفاده از پیشین‌های تخت و تقریب‌های بزرگ نمونه‌ای، استنباط‌های بیزی شبیه استنباط‌های ماکسیمم درستنمایی در روش‌های کلاسیک بزرگ نمونه‌ای می‌شود.

۲.۱.۲ درستنمایی‌های تبدیل‌یافته داده‌ها[۳] (DTL)

باکس و تیائو (۱۹۷۳) (۱) در مورد روشی برای انتخاب پیشین‌های ناآگاهی بخش بحث می‌کنند که مقدار نسبی اطلاعات موجود در داده‌ها را با مقدار نسبی اطلاعات موجود در پیشین مقایسه می‌کند. استدلال آن‌ها می‌گوید که اگر تابع درستنمایی $L(y \mid \theta)$، دارای خاصیت DTL باشد، آنگاه استفاده از پیشین "تخت" برای θ معقول است.

در درستنمایی داده تبدیل‌یافته، اگر تابع درستنمایی را برای مجموعه داده‌های متمایز رسم کنیم، شکل تابع درستنمایی تغییر نمی‌کند. فقط مکان آن بوسیله داده‌ها تغیر داده می‌شود. به عنوان مثال، در داده‌های $N(\theta, ۱/\tau_۰)$، درستنمایی به صورت $L(\theta \mid \overline{y}_۰) \propto \exp[-n\tau_۰(\theta - \overline{y}_۰)^۲/۲]$ می‌باشد. با تغییر $\overline{y}_۰$ و رسم آن به صورت تابعی از θ، منحنی زنگوله‌ای شکل یکسانی به دست می‌آید که فقط مد منحنی حرکت می‌کند.

باکس و تیائو استدلال کردند که برای یک DTL اگر یک پیشین ثابت $p(\theta) = k$ در نظر بگیریم، صرف نظر از این که داده‌ها چه هستند، این پیشین در مقایسه با داده‌ها اطلاعات اضافه‌ای به ما نمی‌دهد. پیشنهاد باکس و تیائو برای پیشین میانگین توزیع نرمال وقتی دقت معلوم است $p(\theta) = k$ می‌باشد.

مثال زیر نشان می‌دهد اگر تابع درستنمایی یک توزیع DTL نباشد می‌توان با تبدیل زدن روی متغیر به DTL رسید.

تابع درستنمایی برای یک تک مشاهده از توزیع نمایی با پارامتر θ به صورت $L(y \mid \theta) \propto \theta e^{-\theta y} \propto \theta y e^{-\theta y}$ است. با پارامتری کردن مجدد θ به صورت $\gamma = \log(\theta)$ می‌توان تابع درستنمایی را به صورت زیر نوشت:

$$L(\gamma \mid y) \propto \exp[\gamma + \log(y) - e^{\gamma + \log(y)}]$$

که DTL است. پیشین باکس‌ـتیائو $p(\gamma) = k$ است و با یک تبدیل معمول، پیشین θ به صورت زیر می‌شود:

$$q(\theta) = p(\log(\theta)) \left| \frac{d\log(\theta)}{d\theta} \right| = k/\theta \propto ۱/\theta$$

[۳] Data translated likelihood

۳.۱.۲ پیشین‌های جفریز

جفریز (۱۹۶۱) رده‌ای از پیشین‌ها را پیشنهاد کرد که بسیاری از آن‌ها به عنوان پیشین‌های "ناآگاهی بخش" یاد می‌کنند. پیشین جفریز به عنوان ریشه دوم اطلاع فیشر تعریف می‌شود. پیشین جفریز یک پیشین مرجع است.

استفاده از پیشین‌های جفریز، اصل قاعده توقف یا اصل درستنمایی را پوشش نمی دهند (۱). پیشین‌های جفریز برای داده‌های دوجمله‌ای و دوجمله‌ای منفی متفاوت است، یعنی استفاده از پیشین‌های جفریز برای این دو نوع داده با تعداد یکسانی از موفقیت و شکست منجر به نتایج متفاوت می‌شود. در واقع منفی مشتق دوم لگاریتم درستنمایی در هر دو مورد یکسان است اما در محاسبه اطلاع فیشر مقدار امید ریاضی آن از دو توزیع متفاوت آمده است.

اگر $y \mid \theta$ دارای توزیع $\mathrm{Bin}(n.\theta)$ باشد تابع درستنمایی به صورت $l(\theta \mid y) \propto \theta^{y}(۱-\theta)^{n-y}$ و لگاریتم درستنمایی به صورت $y\log(\theta) + (n-y)\log(۱-\theta)$ خواهد بود. بعد از گرفتن مشتق لگاریتم درستنمایی و محاسبه ریشه دوم اطلاع فیشر، در نهایت پیشین جفریز به صورت زیر به دست می‌آید

$$p(\theta) \propto \theta^{-\frac{۱}{۲}}(۱-\theta)^{-\frac{۱}{۲}}$$

که توزیع $\mathrm{Beta}(۰/۵, ۰/۵)$ است. در این توزیع احتمال بیشتری نزدیک نقاط صفر و یک قرار می‌گیرد.

استفاده از پیشین جفریز برای مدل‌های تک پارامتری پذیرفته شده است اما استفاده از آن در مدل‌های چند پارامتری اغلب بحث برانگیز است و محاسبه آن می‌تواند سخت باشد.

در توزیع های چند پارامتری اگر $\ell(\theta) \equiv \log[L(\theta \mid y)]$ باشد و مشتق دوم ماتریس $\ddot{\ell}(\theta)$ را به عنوان مشتق جزیی مرتبه دوم ماتریس $\ell(\theta)$ تعریف کنیم که یک ماتریس متقارن از مقادیر $\{\frac{\partial^{۲}}{\partial\theta_j\partial\theta_k}\ell(\theta)\}$ است، اطلاع فیشر به صورت

$$\mathrm{I}(\theta) \equiv \mathrm{E}[-\ddot{\ell}(\theta)]$$

تعریف می‌شود و پیشین جفریز $p(\theta) \propto \sqrt{\det[I(\theta)]}$ خواهد بود.

۴.۱.۲ پیشین‌های مزدوج

هنگامی که چگالی پیشین $p(\theta)$ دارای شکل تابعی مشابه‌ای با چگالی نمونه‌گیری $f(y \mid \theta)$ باشد بعد از اعمال قضیه بیز، توزیع پسین دارای شکل تابعی مشابهی با پیشین می‌شود. چنین جفت‌هایی از چگالی‌های نمونه و چگالی‌های پیشین خانواده مزدوج نامیده می‌شوند. جدول زیر برخی از خانواده های مزدوج را نشان می‌دهد.

۵.۱.۲ پیشین‌های مرجع ناسره استاندارد در خانواده مکان و مقیاس

در خانواده مکان‌ـ‌مقیاس $\frac{۱}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$ برای پارامترهای μ و σ می‌توان پیشین زیر را در نظر گرفت.

$$p(\mu,\tau) = p(\mu)\,p\left(\frac{۱}{\tau}\right) = ۱ \times \frac{۱}{\tau} = \frac{۱}{\tau}$$

که در آن $\sigma = \frac{۱}{\tau}$.

$f(y \mid \theta)$	$p(\theta)$	$p(\theta \mid y)$
$\text{Beta}(n, \theta)$	$\text{Beta}(a, b)$	$\text{Beta}(a + y, b + n - y)$
$\text{Pois}(\theta)$	$\text{Gamma}(a, b)$	$\text{Gamma}(\sum y_i + a, n + b)$
$\text{Exp}(\theta)$	$\text{Gamma}(a, b)$	$\text{Gamma}(a + n, b + \sum y_i)$
$N(\theta, ۱/\tau_*)$	$N(\theta., ۱/\tau.)$	$N\left(\frac{\tau.}{\tau. + n\tau_*}\theta. + \frac{n\tau_*}{\tau. + n\tau_*}\overline{y}., ۱/(\tau. + n\tau_*)\right)$
$N(\mu_*, ۱/\theta)$	$\text{Gamma}(a, b)$	$\text{Gamma}(a + n/۲, b + n[s^۲ + (\overline{y}. - \mu_*)^۲]/۲)$ where $s^۲ = n^{-۱}\sum(y_i - \overline{y}.)^۲$

۲.۲ اطلاع پیشین

زمانی که هدف به دست آوردن توزیع‌های پیشین برای داده‌های دوجمله ای است، خانواده توزیع های بتا یک رده انعطاف پذیر و مناسب برای مدل‌بندی عدم قطعیت در مورد احتمالات می‌باشد. همانطور که می‌دانیم انتخاب‌های متفاوت برای پارامترهای توزیع بتا منجر به اشکال متنوعی از جمله متقارن تک مدی، چوله به راست تک مدی، چوله به چپ تک مدی و توزیع یکنواخت استاندارد می‌شوند. هیچ یک از خانواده‌های تک مدی آنقدر گسترده نیستند که تمام احتمالات برای توزیع پیشین در مورد پارامتر مجهول را دربر بگیرند. نکته حایز اهمیت این است که پیدا کردن پیشینی که موافق با اطلاعات کارشناس است خیلی مهم تر از راحتی استفاده از توزیع بتاست. برای مثال، اگر پیشین کارشناس دو مدی باشد (تجربه نشان داده است که به ندرت چنین حالتی رخ می‌دهد) نمی‌توانیم از توزیع‌های بتا استفاده کنیم زیرا آن‌ها تنها زمانیکه هر دو پارامتر کمتر از یک باشند دومدی هستند. در این حالت می‌توان از توزیع بتا آمیخته استفاده کرد. در اکثر مواقع اطلاعات پیشین را با استفاده از نظر محقق که به دو قسم تقسیم می‌شود به دست می‌آوریم:

۱. بهترین حدس برای احتمال یا پارامتر مجهول θ

۲. بزرگترین و یا کوچکترین مقدار احتمالی که می‌تواند منطقی باشد. به طور مثال، کمیتی که تنها ۵ درصد شانس وجود دارد که مقدار احتمال یا پارامتر مجهول θ از آن بیشتر باشد و یا بالعکس. کمیتی که تنها ۵ درصد شانس وجود دارد که مقدار احتمال یا پارامتر مجهول θ از آن کمتر باشد.

به طور کلی استنباط ما شامل حدسی است که کارشناس در مورد پارامتر مجهول و صدک توزیع پارامتر مجهول می‌زند. به طور معمول بهترین حدسی که کارشناس می‌زند به عنوان مد در نظر گرفته می‌شود. اگر حدسی که در مورد مقدار پیشین پارامتر مجهول θ می‌زند کمتر از ۰٫۵ باشد معمولاً از محقق در مورد کران بالای ۰٫۹۵ یا ۰٫۹۹ برای θ که مقادیر θ با احتمال ۰٫۹۵ یا ۰٫۹۹ کمتر از آن است پرسش می‌شود و بالعکس اگر حدس محقق بیشتر از ۰٫۵ باشد در مورد کران پایین ۰٫۹۵ یا ۰٫۹۹ از او پرسش می‌شود.

لازم به ذکر است اطلاعات پیشین آگاهی بخش باید مستقل از داده‌هایی که در حال تحلیل هستند استخراج شوند در صورتی که اغلب بعد از این که داده‌ها جمع‌آوری شد در مورد مقدار پیشین تصمیم‌گیری می‌شود و این مسأله باعث می‌شود که نتوانیم اطلاعات را نادیده بگیریم. علاوه بر آن گاهی به جای اینکه از نظر محقق در مورد مقادیر پیشین استفاده کنیم می‌توانیم براساس مقادیری که در مقالات معتبر منتشر شده (داده‌های تاریخی) استفاده کنیم.

شرکت پاراید تولید کننده واشرهای توپ هایی است که در دوره‌های آموزشی بازی‌های گلف مورد استفاده قرار می‌گیرد. در شرکت از ۲۴۳۰ قطعه‌ای که تولید می‌شود ۲۲۱۱ قطعه واقعاً ارسال می‌شود. احتمال ارسال یک قطعه معیوب چقدر است؟

فرض برآن است که تعداد قطعات معیوب (y) دارای توزیع دوجمله ای با پارامترهای ۲۴۳۰ و θ است. علاوه بر آن ، مدیر شرکت ادعا دارد نسبت ضایعات بین ۵ تا ۱۵ درصد است. یعنی احتمال اینکه θ بیشتر از ۰/۱۵ باشد برابر با ۰/۰۲۵ و احتمال اینکه θ کمتر از ۰/۰۵ باشد هم برابر با ۰/۰۲۵ است. با فرض اینکه θ با پیشین بتا مدل‌بندی شده است و با استفاده از این دو مقدار احتمال می‌توان توزیع پیشین θ را مشخص کرد.

$$y \mid \theta \sim \text{Bin}(۲۴۳۰, \theta)$$

$$\theta \sim \text{Beta}(۱۲/۰۵, ۱۱۶/۰۶)$$

$$y = ۲۱۹ = ۲۴۳۰ - ۲۲۱۱$$

با توجه به اطلاعات فوق توزیع پسین θ مشخص می‌شود:

$$\theta \mid y \sim \text{Beta}(۲۱۹ + ۱۲/۰۵, ۲۴۳۰ - ۲۱۹ + ۱۱۶/۰۶)$$

حال با استفاده از میانه ۰/۰۹ (به عنوان برآورد نقطه‌ای) و فاصله احتمال ۰/۹۵، احتمال ارسال یک قطعه معیوب برابر با (۰/۱ – ۰/۰۸) برآورد می‌گردد. یک آلیاژ برنجی که در برابر خردگی مقاوم است و به طور گسترده‌ای در وسایل لوله کشی استفاده می‌شود، ازمس، قلع، سرب، روی، نیکل و آهن تشکیل شده است. زمانی که بین ۴ تا ۶ درصد روی به آلیاژ افزوده می‌شود مقاومت فلز افزایش می‌یابد. روی دارای نقطه ذوب پایین تر از مس است، بنابراین مقدار مشخصی از روی اضافه شده تا انتهای فرایند که به آلیاژ حرارت می‌دهند کاهش می‌یابد. به طور معمول مقدار واقعی روی موجود در آلیاژ را قبل از ریختن آلیاژ به قالب اندازه می گیرند. اگر مقدار روی موجود در آلیاژ کمتر از ۴/۴٪ بود مقداری به آن اضافه می‌شود تا درصد آن اصلاح شود.

۱۲ نمونه آلیاژ در ژوئن سال ۲۰۰۳ مورد آزمایش قرار گرفت. فرض می کنیم که y_i نشان دهنده مقدار روی موجود در نمونه iام باشد. علاوه بر آن فرض می‌کنیم مشاهدات از توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ پیروی می‌کنند. احتمال اینکه در نمونه بعدی (نمونه سیزدهم) نیاز به اضافه کردن روی وجود داشته باشد چقدر است؟

فرض کنید

$$y_۱, y_۲, \cdots, y_{۱۲} \mid \mu.\sigma^۲ \sim^{iid} N(\mu.\sigma^۲)$$

درصد روی ۱۲ نمونه عبارت‌اند از ۴/۲۰، ۴/۳۶، ۴/۱۱، ۳/۹۶، ۵/۶۳، ۴/۵۰، ۵/۶۴، ۴/۳۸، ۴/۴۵، ۳/۶۷، ۴/۶۶، ۵/۲۶

برای انجام تحلیل بیزی باید توزیع پیشین پارامترهای ناشناخته μ و $\sigma^۲$ را مشخص کنیم. رئیس کارخانه با ۰/۹۵ اطمینان ادعا می‌کند میانگین درصد روی قبل از ریختن به قالب بین ۴/۵ تا ۵ درصد می‌باشد و به طور میانگین مقدار درصد روی ۴/۷۵ درصد می‌باشد. بنابراین با حل معادله $p(۴/۵ < \mu < ۵) = ۰/۹۵$ مقدار واریانس محاسبه و داریم:

$$\mu \sim N(۴/۷۵, ۰/۰۱۶۳)$$

در مورد پارامتر $\sigma^۲$ هیچ اطلاعات مفیدی نداریم لذا از پیشین مرجع استفاده می کنیم. از آنجایی که شکل تابع چگالی برای $\frac{۱}{\sigma^۲}$ زیباتر است توزیع پیشین را برای $\frac{۱}{\sigma^۲}$ با پارامترهای کوچک به صورت زیر مشخص می‌کنیم.

$$\sigma^{-۲} \sim \text{Gamma}(۰/۰۰۱, ۰/۰۰۱)$$

در مورد پارامتر μ، با استفاده از شبیه‌سازی‌های کامپیوتری میانه پسین برابر با ۴/۶۹ درصد می‌شود و با احتمال ۹۵ درصد مقدار میانگین بین ۴/۴۹ و ۴/۹۰ درصد قرار دارد. یعنی:

$$\Pr(۴/۴۹ < \mu < ۴/۹۰ \mid y_۱, y_۲, \dots y_{۱۲}) = ۰/۹۵$$

برای σ میانه توزیع پسین برابر با ۰/۶۴ درصد و با احتمال ۹۵ درصد بین ۰/۴۴ درصد و ۱/۰۳ درصد قرار دارد. یعنی:

$$\Pr\left(۰/۴۴ < \sigma < ۱/۰۳ \mid y_۱, y_۲, \dots y_{۱۲}\right) = ۰/۹۵$$

اما ما به دنبال $\Pr(y_{۱۳} < ۴/۴ \mid \mu, \sigma)$ هستیم. شبیه‌سازی‌های کامپیوتری نشان داده است که احتمال نیاز به اضافه کردن روی ۰/۳۳ است و فاصله احتمال ۰/۹۵ مربوطه (۰/۴۵ − ۰/۲) می‌باشد.

۳ انتخاب نامناسب پیشین

در این بخش با ارایه مثالی معروف، به نقش انتخاب پیشین نامناسب در استنباط بیزی می پردازیم.

۱.۳ پارادوکس لیندلی-جفریز

پارادوکس لیندلی-جفریز بیان می‌کند که اگر در تحلیل بیز، توزیع پیشین نامناسبی انتخاب شود، این امکان که پسین نامناسبی به وجود آید وجود دارد.

فرض کنید $y \mid \theta \sim N(\theta, ۱)$ و هدف آزمون فرضیه $H_. : \theta = ۰$ در مقابل $H_۱ : \theta > ۰$ باشد. نخست نیاز است توزیع احتمال پیشین برای مقادیر ممکن θ یعنی $\theta \geq ۰$ مشخص شود. اگر از توزیعی پیوسته برای توزیع احتمال پیشین استفاده کنیم، با توجه به تعریف، احتمال پیشین $\theta = ۰$ برابر با صفر است. بنابراین هیچ شانسی برای پذیرش فرض صفر وجود ندارد. لذا، برای غلبه بر این مشکل در نقطه $\theta = ۰$ مقدار احتمال ۰/۵ را در نظر می گیریم و آن را به صورت $q_. = \Pr(\theta = ۰) = ۰/۵$ نشان می دهیم و بقیه احتمال را در $\theta > ۰$ توزیع می کنیم. واضح است برای $\theta > ۰$ مقدار احتمال ۰/۵ می‌باشد و آن را به صورت $۱ - q_. = \Pr(\theta > ۰) = ۰/۵$ نشان می دهیم. حال باید برای $\theta > ۰$ توزیعی را در نظر بگیریم و البته نیاز به توزیعی داریم که بسیار پراکنده است. برای راحتی محاسبات از توزیع نرمال برای تقریبی از توزیع شرطی استفاده می کنیم. از طرفی برای اینکه پراکندگی زیادی وجود داشته باشد نیاز است مقدار انحراف معیار بزرگ باشد. و از طرف دیگر برای اینکه این تقریب منطقی باشد نیاز است مقدار میانگین به طور قابل ملاحظه ای بزرگتر از انحراف معیار باشد به طوری که احتمال خیلی کمی برای مشاهده $\theta < ۰$ وجود داشته باشد. فرض کنید y دارای توزیع پیشین نرمال با واریانس معلوم است لذا، توزیع حاشیه‌ای آن به صورت

$$y \mid \theta > ۰ \sim N(\mu_., ۱ + \sigma_.^۲)$$

است. برای تصمیم به پذیرش یا رد فرض صفر نیاز به محاسبه احتمال پسین H_0 است.

$$\Pr(\theta = 0 \mid y) = \frac{q_0 f(y \mid \theta = 0)}{q_0 f(y \mid \theta = 0) + (1 - q_0) f(y \mid \theta > 0)}$$

در عبارت فوق با توجه به اینکه $q_0 = 0/5$ است، احتمال مذکور بزرگتر از ۰/۵ می‌شود اگر و فقط اگر $f(y \mid \theta = 0) > f(y \mid \theta > 0)$ باشد و یا معادل آن زمانیکه $y^2 < \log(1 + \sigma_0^2) + (y - \mu_0)^2/(\sigma_0^2 + 1)$ فرض صفر پذیرفته می‌شود. برای شفاف سازی این پارادوکس فرض کنید $\sigma_0^2 = 1097$ و $\mu_0 = 8103$ است. در این وضعیت فرض صفر پذیرفته می‌شود اگر و فقط اگر $y \leq 2/646$.

اما پارادوکس موجود در این مساله این است که اگر ۲/۶ مشاهده شود می پذیریم که این متغیر از $N(0, 1)$ آمده است در حال که اگر واقعاً متغیری دارای توزیع نرمال استاندارد باشد، مشاهده ۲/۶ بسیار غیر محتمل است و علت این مشکل انتخاب پیشین نامناسب و توزیع نامناسب احتمال بر مقادیر بزرگ θ می‌باشد.

۴ نتیجه گیری

روش‌های بیزی از تمام اطلاعات موجود استفاده می‌کند. این اشاره به این واقعیت دارد که رویکرد بیزی شامل اطلاعات پیشین است. از آنجا که اطلاعات پیشین باید تمام اطلاعات موجود را جدا از خود داده‌ها نشان دهند، بنابراین هیچ اطلاعاتی در تجزیه و تحلیل بیزی حذف نمی‌شود. تکنیک‌های بیزی بطور خاص برای تصمیم‌گیری مناسب هستند. که این از طبیعت احتمالاتی و پارامترها به دست می‌آید. چیزی که تصمیم‌گیری را سخت می‌کند عدم اطمینان است. عدم اطمینان در مورد عواقب هر تصمیم، ناشی از عدم دانش در مورد برخی واقعیات یا پارامترهای موجود می‌باشد. روش‌های بیزی می‌توانند این عدم قطعیت را با استفاده از احتمال شخصی تعیین کنند. در حقیقت یک توزیع پسین برای پارامترهای مجهول براساس شواهد موجود استنتاج می‌شود. تعیین عدم قطعیت‌ها در تصمیم‌گیری جز حیاتی تصمیم‌گیری منطقی و مبتنی بر شواهد محسوب می‌شود.

تمایز اصلی بین روش‌های غیر بیزی و روش‌های بیزی این است که رویکرد بیزی از اطلاعات پیشین در مورد پارامترها استفاده می‌کند که این یک منبع قدرت و همچنین یک منبع اختلاف می‌باشد. اختلاف موجود به این علت است که اطلاعات پیشین در اصل وابسته به نظر شخصی می‌باشد و این منجر به توزیع های پسین متفاوت می‌شود. در این حالت ادعا می‌شود که کلا رویکرد بیزی ذهنی است . بسیاری از منتقدان براین باورند که ذهنی بودن نقطه ضعف اصلی رویکرد بیزی است.

اما این یک اعتراض غیرقابل توجیه است زیرا علم حقیقتاً نمی‌تواند عینی باشد. در عمل، روش بیز ماهیت واقعی روش علمی را در موارد زیر نشان می‌دهد.

۱. ذهنیت در توزیع پیشین از طریق قرار دادن پیشین پایه بر روی شواهد قابل دفاع و استدلال به حداقل می‌رسد.

۲. از طریق جمع‌آوری داده‌ها، اختلاف در پیشین‌ها حل می‌شود و توافق حاصل می‌شود.

اگر اطلاعات پیشین مبهم و غیرمستقیم باشد، در ترکیب با داده‌ها وزن ناچیزی می‌گیرد و اثر پسین به طور کامل بر اساس اطلاعات داده‌ها (مانند تابع درستنمایی) بیان می‌شود. این ویژگی قضیه بیز، روند علم را

نشان می‌دهد، که در آن تجمع شواهد عینی، فرآیند اصلی است که به موجب آن اختلاف نظر ها حل می‌شود. هنگامی که داده‌ها شواهد قطعی را ارائه می دهند، اساسا هیچ جایگاهی برای دیدگاه ذهنی باقی نمی ماند.

مراجع

[1] G.E.P. Box, and G.C. Tiao, (1973), *Bayesian Inference in Statistical Analysis*, JohnWiley and Sons, New York.

[3] Ronald Christensen, Wesley Johnson, Adam Branscum and Timothy E. Hanson. (2010), *Bayesian ideas and data analysis; an introduction for scientists and statistician*, Second Edition (Chapman & Hall/CRC Texts in Statistical Science).

[2] A. O'Hagan. (2004), *Bayesian statistics: principles and benefits,* Frontis, 3:31–45.

[1] H. Raiffa, and R. Schlaifer, (1961), *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

خطاهای آماری مقالات منتشر شده در مجلات دندان پزشکی سال ۱۳۹۷

حیدری بهنوئیه ، ا [۱] جام بر سنگ ، س [۲] احمدی نیا گدنه ، ح [۳]

[۱،۲] دانشگاه شهید صدوقی یزد
[۳] دانشگاه ولی عصر (عج) رفسنجان

چکیده

هدف از انجام این تحقیق بررسی میزان خطاهای آماری مقالات منتشر شده در مجلات دندان پزشکی سال ۱۳۹۷ می‌باشد. بدین منظور ۱۰۰ مقاله از مقالات چاپ شده مجلات دندان پزشکی در سال ۱۳۹۷ به صورت به صورت تصادفی انتخاب شدند و خطاهای آماری موجود در آن ها مورد بررسی قرار گرفته است، نتایج نشان داد که در ۵۰ مقاله بررسی شده موارد خطای عدم گزارش میزان معناداری ۱۵، اشتباه معناداری ۱۰، اشتباه گزارش معناداری ۱۴، اسم آزمون گفته نشده ۱۸، آزمون اشتباه نسبت به نوع مطالعه استفاده شده است ۴، نشانه گذاری انحراف معیار و میانگین اشتباه ۱۰، تعابیر غلط ۹، آزمون در نتایج نیست ۱۶، برای داده‌ها میانگین گزارش نشده ۱۱ مورد بوده است.

کلمات کلیدی: اخطاهای آماری ، مجلات دندان پزشکی، آزمون فرضیه، سطح معناداری.

۱ پیش‌گفتار

از علوم آماری در تمام فرایندهای مطالعات علمی از برنامه‌ریزی تا مراحل گزارش استفاده شده‌است. این حقیقت که امروزه آمار در مطالعات علمی ضروری است نشان می‌دهد که استفاده صحیح از روش‌های آماری نیز بسیار مهم است. مراجع پزشکی همچنین بر اهمیت آمار تاکید دارند که پزشکان باید حداقل خوانندگان

[۱] Ahmadheydari37@yahoo.com

خوبی از آمار باشند. محققان در مجلات علمی، که دانش کافی از آمار ندارند، ممکن است در استفاده از دانش آماری در هر گام، از جمله برنامه‌ریزی، طراحی، اجرا، تحلیل و ارائه داده‌ها دچار اشتباه شوند. اگرچه بسیاری از خوانندگان به دنبال این نوشته‌ها شناخته نمی‌شوند، اما اکثر مقالات در ادبیات پزشکی حاوی خطاهای آماری هستند. برخی از این خطاها به طور مستقیم بر نتایج تاثیر می‌گذارند در حالی که دیگران خطاهای ارایه در نمایش یا اصطلاحات را ارایه می‌کنند و تاثیر عمده‌ای بر نتیجه ندارند (۱). در هر دو شرایط باید از این اشتباهات اجتناب کرد. از دهه ۱۹۶۰ و ۱۹۷۰ بسیاری از پژوهشگران، که تمایل به جلب توجه به خطاها و حذفیات در آمار و روش‌شناسی دارند، رایج‌ترین روش‌های آماری مورد استفاده در مجلات پزشکی را مورد بررسی قرار داده و بر اهمیت استفاده صحیح از آمار در نشریات علمی تاکید کرده‌اند و تحقیقات و طرح‌های خود را در مورد این موضوع منتشر کرده‌اند. برخی نویسندگان طرح مورد استفاده در آزمایش‌ها، حذفیات و اشتباهات در طراحی و استفاده نامناسب از طراحی در نشریات پزشکی را مورد مطالعه قرار دادند (۶ ـ ۶). چندین نویسنده انواع آنالیز انجام‌شده و آزمون‌های آماری مورد استفاده، روش‌های آماری نادرست، سو استفاده از آزمون‌های آماری و کاربرد آماری نامناسب، و شکست در فهرست آزمون‌های آماری مورد استفاده در نشریات پزشکی را مورد مطالعه قرار دادند (۱۳ ـ ۱۳). برخی نویسندگان ارایه داده‌ها، استفاده نادرست در ارایه تحلیل توصیفی، اشتباهات در خلاصه کردن داده‌ها، و استفاده نادرست از مقیاس‌های مکان و پراکندگی در نشریات پزشکی را مورد مطالعه قرار دادند (۳، ۵، ۱۱ ـ ۱۶). برخی نویسندگان بر دانش آماری و آموزش آماری برای متخصصین بالینی در مطالعات خود تاکید کردند (۸، ۱۶ ـ ۱۹). برخی نویسندگان در مطالعات خود بر اهمیت مشاوره آماری، اهمیت مرور آماری و ارزیابی کیفیت آماری قبل از انتشار نسخه‌های خطی تاکید کردند و اثر داوری آماری در فرآیند مرور (۴، ۶، ۸، ۲۰، ۲۱). مجلات رادیولوژی به مرور مقالات، سرمقاله و بررسی کتاب برای مسائل آماری در طول ۱۰۰ سال گذشته یا به همین ترتیب، برای جلب توجه و آموزش پژوهشگران به منظور اجتناب از خطاهای آماری، منتشر کرده‌اند. اشتباهات در استفاده از آمار ممکن است، در هر مرحله از یک مطالعه تحقیقاتی رخ دهد. یک مطالعه علمی ممکن است طراحی و اجرا شود، اما اگر به درستی تحلیل نشده و ارائه شده‌باشد، حتی یک اشتباه تنها می‌تواند باعث از دست دادن اهمیت خود شود (۱۸). استفاده نادرست از آمار منجر به نتایج اشتباه و نیز از دست دادن نیروی کار، زمان و هزینه می‌شود (۱۱). همچنین رابطه روشنی بین آمار و اخلاق وجود دارد. انتشار نتایج گمراه‌کننده که منعکس‌کننده حقیقت نیستند، یک مساله اخلاقی بالقوه در همان زمان است. انتشار یافته‌های نادرست ممکن است یک مرجع گمراه‌کننده برای مطالعات بیشتر باشد. علاوه بر این، حذف اشتباهات تنها برای محققانی که درگیر مطالعات علمی هستند، مهم نیست، بلکه برای پزشکانی که می‌توانند مستقیماً از نتایج این مطالعات در مطب بالینی خود استفاده کنند، مهم نیست. در این زمینه، هم مجلات علمی و هم محققانی که وظیفه انتقال دانش علمی را انجام می‌دهند، مسیولیت بزرگی برای اجتناب از اشتباهات انجام می‌دهند. هدف این پژوهش، بررسی مقالات مجلات دندان پزشکی به لحاظ خطاهای آماری است. بنابراین، هدف ما کمک به تولید نشریات علمی با کیفیت بالا از طریق توانمند سازی دانشمندان، ویراستار مجله و افراد درگیر در فرآیند ارزیابی مقاله در مجلات دندان پزشکی برای حساس و دقیق در مورد خطاهای آماری است. پیشینه تحقیق:

رضائیان (۱۳۹۵) در مقاله ای تحت عنوان باز پسگیري مقالات منتشر شده به دلیل اشتباهات آماري، با

توجه به وجود اشتباهات آماری در مقالات پیشنهاد می دهد که بهتر است یک متخصص آمار از ابتداي طراحی مطالعه در تیم پژوهشی عضو بوده و نظرات وي مورد توجه قرار گیرد و در مقاله منتشر شده باید آنالیزهاي آماري بکار رفته با دقت هر چه تمامتر شرح داده شوند. Guigan Mc گزارش داد که ۴۰

نتایج تجربی:

شاخص		فراوانی	درصد
خطای وابسته به معناداری	عدم گزارش میزان معناداری	۱۵	۳۰
	اشتباه معناداری	۱۰	۲۰
	اشتباه گزارش معناداری	۱٤	۲۸
خطای مربوط به آزمون	اسم آزمون گفته نشده	۱۱	۲۲
	اسم آزمون در روش آمده ولی استفاده نشده	۱۸	۳٦
	آزمون اشتباه نسبت به نوع مطالعه استفاده شده است	٤	۸
	آزمون لازم انجام نشده	٤	۸
نشانه گذاری انحراف معیار و میانگین اشتباه		۱۰	۲۰
تعابیر غلط		۹	۱۸
آزمون در نتایج نیست		۱٦	۳۲
برای داده ها میانگین گزارش نشده		۱۱	۲۲
مجموع		۵۰	۱۰۰

شکل ۱: توزیع خطاهای آماری در مطالعات مشابه

نتایج بررسی مقالات نشان داد که موارد خطای عدم گزارش میزان معناداری ۱۵ ، اشتباه معناداری۱۰ ، اشتباه گزارش معناداری ۱۴، اسم آزمون گفته نشده ۱۱، اسم آزمون در روش آمده ولی استفاده نشده ۱۸، آزمون اشتباه نسبت به نوع مطالعه استفاده شده است ۴، آزمون لازم انجام نشده ۴، نشانه گذاری انحراف معیار و میانگین اشتباه ۱۰، تعابیر غلط ۹، آزمون در نتایج نیست ۱۶، برای داده ها میانگین گزارش نشده ۱۱مورد بوده است. بحث و نتیجه گیری:

به منظور دستیابی به کیفیت بالا در هر جنبه از یک مطالعه علمی، باید توجه خاصی به نماد، ارایه، و بیان علمی اصلاح شود. توجه به نتایج این تحقیق، خطاهای آماری در مجلات دندان پزشکی نیز مشاهده می‌شوند. در بین دلایل اصلی برای این اشتباهات که توسط پژوهشگران انجام‌شده، با یک متخصص biostatistics در مورد موضوع مشورت نمی‌کنند، با فرض اینکه آن‌ها در مورد آمار بسیار خوب می‌دانند، اما در واقع به اندازه کافی دانش کافی ندارند و بی‌دقتی می‌کنند (۱۲، ۳۲). پیش‌گیری از اشتباهات آماری در نسخه‌های خطی، مسئولیت هر دو محققانی است که مطالعات علمی را انجام می‌دهند که این مطالعات را در مجلات خود منتشر می‌کنند. محقق باید دانش آماری اساسی داشته باشد تا روش‌های آماری را در یک مطالعه علمی بخواند و تفسیر کند. برای کسب اطمینان از کسب سواد آماری باید در نظر گرفته شود که آمار باید به دقت و به اندازه کافی به دانشجویان پزشکی و کسانی که آموزش اقامت دارند آموزش داده شود. علاوه بر این، آلتمن در زمینه توسعه آموزش استاندارد در آمار، پیشنهاداتی ارائه کرده‌است. دادن اهمیت لازم برای آموزش آمار، مانع از اشتباهات جدی در تحقیقات علمی آینده خواهد شد و تضمین خواهد کرد که

دانش آموزان سوادآموزی آماری کافی خواهند داشت. پیشنهاد دیگر تشویق یادگیری این موضوعات از طریق سمینارها درباره مهارت‌های تفکر انتقادی و روش تحقیق در جلسات علمی در نواحی غیر از آموزش پزشکی است. بسیار مهم است که فرضیه یک مطالعه طوری طراحی شود که به اندازه کافی ارزیابی شود و داده‌ها به درستی جمع‌آوری شوند و داده‌های جمع‌آوری‌شده به درستی تحلیل شوند. در این زمینه، دریافت مشاوره آماری در تمام مراحل این مطالعه مانند برنامه‌ریزی، اجرا، جمع‌آوری داده‌ها و مراحل آنالیز مفید خواهد بود (۹). محققان باید متخصصان biostatistics را در مطالعه علمی خود داشته باشند و باید قبل از شروع یک مطالعه علمی و در تمام مراحل زیر، از آن‌ها مشاوره بگیرند (۲، ۳۳). بیش‌ترین مسیولیت برای ویراستاران مجلات این است که نسبت به آمار موجود در روند بازبینی مقاله حساسیت بیشتری داشته باشند و درخواست کنند. کمک متخصصان biostatistics و همچنین متخصصان در موضوع مربوطه در فرآیند ارزیابی نشریات علمی ارسال‌شده به مجلات آن‌ها. reviewers که مقالات مجلات علمی را مورد انتقاد قرار می‌دهند معمولاً با توجه به تخصص آن‌ها در موضوع پزشکی مربوطه انتخاب می‌شوند، و در نتیجه، روش آماری مورد استفاده در اکثر تحقیقات، ممکن است به اندازه کافی بررسی و ارزیابی دقیقی نداشته باشد (۳۳). گودمن و همکاران (۳۴) پیشنهاد کردند که یک مخزن خبره می‌تواند برای ارزیابی روش در مطالعات علمی ایجاد شود و این نشریه می‌تواند داوران را از این استخر انتخاب کند. آلتمن (۶) همچنین اثرات و اهمیت داشتن یک reviewer آماری برای مطالعات علمی را مورد بحث قرار داد. ایده خوبی است که متخصصان biostatistics را در این نوع استخر برای روش‌شناسی در پرتو این دو نکته بگنجانید. امروزه، روزنامه‌هایی که از biostatistics reviewers استفاده می‌کنند و متخصصان biostatistics در صفحه سرمقاله دارند در عدد (۳۳) رو به افزایش هستند. از همه مجلات علمی انتظار می‌رود که این عمل را انجام دهند. نباید فراموش کرد که سو استفاده از آمار ممکن است منجر به نتایج گمراه‌کننده، علوم ضعیف، و در نهایت منجر به مراقبت نامناسب از بیمار شود (۳۵). روند بررسی این است که انجام یک بررسی آماری قبل از بررسی‌های دیگر متخصصان رشته مربوطه انجام شود (۱۳). برای مثال، یک اشتباه که در نتیجه بررسی آماری یافت می‌شود و نیاز است اصلاح شود، نتایج را تحت‌تاثیر قرار خواهد داد و در نتیجه بررسی آماری باید قبل از مرور توسط کارشناسان مربوطه انجام شود تا به طور غیر ضروری فرآیند بازبینی مقاله را طولانی‌تر کند. علاوه بر همه اینها، استفاده از دستورالعمل‌هایی که در مورد جلوگیری از اشتباه‌ات آماری در مقالات به توافق رسیده‌اند ممکن است گسترده‌تر شوند. برخی از مجلات دستورالعمل‌های آماری را در این زمینه منتشر می‌کنند، در حالی که دیگران فهرست‌های آماری برای داوران تولید می‌کنند. علاوه بر این، مجله ممکن است شامل بخش‌های آماری ویژه و مجموعه برای جلب توجه و آموزش پژوهشگران به منظور جلوگیری از خطاهای آماری باشد. چنین روش‌هایی ممکن است برای نویسندگان مفید باشد، در طراحی مطالعه و تحلیل داده‌های آن‌ها؛ برای بررسی کنندگان، در ارزیابی و انتقاد از نسخه‌های خطی؛ و برای خوانندگان در فهم و تفسیر مقالات منتشر شده (۲۵). این روش‌ها می‌توانند به افزایش کیفیت علمی و علمی نشریات کمک کنند. مطالعه ما محدودیت‌هایی دارد. مقالاتی که ما بررسی کردیم در مجلات دندان پزشکی هستند. این مطالعه می‌تواند شامل سایر مجلات باشد. با این حال، نتایج این مطالعه نشان می‌دهد که خطاهای آماری حتی در مجلات مشهور نیز دیده می‌شوند. موضوعات مانند طراحی مطالعه و نمونه‌برداری از این مطالعه حذف شدند. علاوه بر این، طبقه‌بندی خطای اماری، شدت این خطاها و پیامدهای بالقوه آن‌ها

را در نظر نمی‌گیرد.

مراجع

[۱] رضائیان، محسن (۱۳۹۵) باز پسگیري مقالات منتشر شده به دلیل اشتباهات آماري، مجله دانشگاه علوم پزشکی رفسنجان، دوره پانزدهم، ۵۹۲-۵۹۱

[2] Demirtas H. Statistical errors in medical publication. Biom Biostat Int J 2015; 2:00021.[Cross-Ref]

[3] Schor S, Karten I. Statistical evaluation of medical journal manuscripts. JAMA 1966; 195:1123–1128. [CrossRef]

[4] Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articlesin BMJ from January to March 1976. Br Med J1977; 1:85–87. [CrossRef]

[5] Glantz SA. Biostatistics: How to detect, correctand prevent errors in the medical literature.Circulation 1980; 61:1–7. [CrossRef]

[6] MacArthur RD, Jackson GG. An evaluation ofthe use of statistical methodology in the Journal of Infectious Diseases. J Infect Dis 1984;149:349–354. [CrossRef]

[7] Altman DG. Statistical reviewing for medical jour-nals. Stat Med 1998; 17:2661–2674.

[8] Gardner MJ, Bond J. An exploratory study ofstatistical assessment of papers publishedin the British Medical Journal. JAMA 1990;263:1355–1357. [CrossRef]

[9] Welch GE 2nd, Gabbe SG. Review of statisticsusage in the American Journal of Obstetricsand Gynecology. Am J Obstet Gynecol 1996;175:1138–1141. [CrossRef]

[10] Levine D, Bankier AA, Halpern EF. Submissionto Radiology: our top 10 list of statistical errors. Radiology 2009; 253:288–290. [CrossRef]

[11] Šimundić AM, Nikolac N. Statistical errors inmanuscripts submitted to Biochemia Medica-Journal. Biochem Med (Zagreb) 2009; 19:294–300. [CrossRef]

[12] Ercan I, Ocakoglu G, Sigirli D, Ozkaya G. Assessment of submitted manuscripts in medical sciences according to statistical errors. Turk J Med Sci 2012; 32:1381– 1387. [CrossRef]

[13] Ercan I, Karadeniz PG, Cangur S, Ozkaya G,Demirtas H: Examining of published articleswith respect to statistical errors in medical sciences. UHOD 2015; 25:130–138. [CrossRef]

[14] Ercan I, Kaya MO, Uzabacı E, Mankır S, Can FE,Bashir Albishir M. Examination of published articles with respect to statistical errors in veterinary sciences. Acta Vet (Beogr) 2017; 67:33–42. [CrossRef]

[15] McGuigan S. The use of statistics in the British Journal of Psychiatry. Br J Psychiatry 1995;167:683–688. [CrossRef]

[16] Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research-a review of common pitfalls. Swiss Med Wkly2007; 137:44–49.

[17] Feinstein AR. A survey of the statistical procedures in general medical journals. Clin Pharmacol Ther 1974; 15:97–107. [CrossRef]

[18] Gardenier J, Resnik D. The misuse of statistics:concepts, tools, and a research agenda. Account Res 2002; 9:65–74. [CrossRef]

[19] Altman DG. Statistics and ethics in medical research, Misuse of statistics is unethical. Br Med J 1980; 281:1182–1184. [CrossRef]

[20] Goldin J, Zhu W, Sayre JW. A review of the statistical analysis used in papers published inClinical Radiology and British Journal of Radiology. Clin Radiol 1996; 51:47–50. [CrossRef]

[21] Bhattacharyya T, Bhattacharjee A, Balasubramanian S. Bridging the gap between bio-statisticians and oncologists: Need of the hour incomprehensive cancer research. Indian J Cancer. 2015; 52:561–562. [CrossRef]

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

اشتباهات رایج آماری

جباری نوقابی، م [۱] تلخی، ن [۲]

[۱،۲]گروه آمار، دانشگاه فردوسی مشهد

چکیده

امروزه محققان شاهد وجود اشتباهات آماری رایجی در بسیاری از مقالات علمی بوده‌اند. به منظور اینکه نتایج ارائه شده در مقاله‌های علمی صحیح و قابل اتکا باشد، پیش از چاپ در نشریه‌های تخصصی مورد داوری قرار می‌گیرند اما با این حال در برخی از مقاله‌های پذیرفته و چاپ شده در نشریه‌های علمی- پژوهشی اشتباهات مختلفی مشاهده شده است که باعث می‌شود از اعتبار نتایج گزارش شده در مقالات کاسته شود. در این مقاله به معرفی برخی از اشتباهات رایج آماری که محققان در تجربه و تحلیل و نوشتن مقالات علمی دچار می‌شوند، پرداخته شده است. امید است نتایج حاصل از این مقاله در به کار بردن صحیح روش‌های آماری در مقالات علمی نقش مهمی داشته باشد. توصیه می‌شود افرادی که قصد دارند در این موارد مطالعه بیشتری داشته باشند، از متون مفصل‌تر و همچنین سایر منابع ذکر شده کمک بگیرند.

کلمات کلیدی: اشتباهات آماری، اعتبار نتایج، مقالات علمی.

۱ پیش‌گفتار

محققان نرخ بالایی از اشتباهات آماری را در تعداد زیادی از مقالات علمی حتی در بهترین ژورنال ها مشاهده کردند. در واقع، مشکل گزارش‌های ضعیف آماری، بسیارگسترده و به صورت بالقوه، جدی و شناخته نشده هستند، با وجود اینکه بیشتر اشتباهات، مربوط به مفاهیم پایه آماری هستند اما می‌توان با داشتن درک صحیحی از روش‌های آماری و مفاهیم مربوطه، از به کار بردن نادرست آن‌ها در تحقیقات علمی اجتناب

[۱] jabbarinm@um.ac.ir

ورزید. به منظور اینکه نتایج ارائه شده در مقاله‌های علمی صحیح و قابل اتکا باشد، پیش از چاپ در نشریه‌های تخصصی مورد داوری قرار می‌گیرند اما با این حال در برخی از مقاله‌های پذیرفته شده و چاپ شده در نشریه‌های علمی ـ پژوهشی مشاهده شده است که اشتباهات مختلفی وجود دارد که باعث می‌شود از اعتبار نتایج گزارش شده در مقالات کاسته شود. بنابراین لازم هست تا محققین مبتدي از اشتباهات رايجي كه در زمان برنامه‌ريزي، تجزيه و تحليل داده‌ها، روش تحقيق و نوشتن يك مقاله علمي ممكن است با آن مواجه شوند، آگاهي يابند كه اين امر مستلزم اين است كه قبل از شروع تحقيق و نگارش مقاله خود بر تمامی مراحل مطالعه تسلط کافی داشته باشند و همچنین با فنون روش تحقیق و مفاهیم آن آشنایی کامل داشته باشند. محققان می‌توانند در دو زمینه مرتکب اشتباه شوند، هر دو زمینه‌ای که مطرح خواهند شد، می‌توانند صحت نتایج گزارش شده در متون را تحت تأثیر قرار دهند. اولین زمینه، اشتباهاتی هستند که در طول فرایند تحقیق رخ می‌دهند (مانند اشتباهات در برنامه‌ریزی و یا اجرای مطالعه) و دومین زمینه، مواردی هستند که محقق به هنگام تجزیه و تحلیل داده‌ها، تفسیر و ارائه نتایج (در قالب یک مقاله یا پوستر و...) مرتکب می‌شود.
نکته‌ای که در اینجا حائز اهمیت می‌باشد و می‌بایست بدان اشاره نمود، این هست که اغلب اوقات خطاهای برنامه‌ریزی و اجرای مطالعه توسط یک ویرایشگر مجله که فقط نسخه نهایی متن نگارش شده را مشاهده می‌کند، قابل شناسایی و تشخیص نمی‌باشند. از این رو هر مطالعه بایستی با بیشترین دقت و توجه به جزئیات انجام شود، زیرا خطاهای عمدی و غیرعمدی که در طول فرآیند تحقیق اتفاق می‌افتند، زیان‌بار خواهند بود و از این رو بیشترین اهمیت را دارا می‌باشند. آمار را می‌توان به دو حوزه گسترده تقسیم نمود : آمار توصیفی که مربوط به نحوه توصیف نمونه‌هایی از داده‌های جمع‌آوری شده در یک تحقیق است و آمار استنباطی که مربوط به نحوه برآورد (یا استنباط) ویژگی‌های جامعه از نمونه‌ای که انتخاب شده است. در روش‌های آمار استنباطی (کلاسیک یا بیزی) که مبتنی بر منطق لم نیمن پیرسون هستند نمی‌توان هر نوع تفسیری که محقق می‌خواهد را از نتایج آمار استنباطی داشت و استفاده نمود. بنابراین اگر محقق بخواهد تفسیرهای متفاوتی در مورد استنباط آماری داشته باشد بهتر است که از روش مناسب دیگری تحت عنوان استنباط شواهدی استفاده نماید. استنباط شواهدی در عرصه تفاسیر منطقی از آمار کلاسیک پیش است و در مقابل تفسیر با نظام استقرایی آمار کلاسیک ابزاری قوی در اختیار دارد که مشکل آن را حل می‌کند. مقاله حاضر به بررسی اشتباهات رایجی که به واسطه‌ی زمینه دوم اتفاق خواهند افتاد می‌پردازد که در ادامه مورد بررسی قرار می‌گیرند.

۲ روش های آمارگیری

روش‌های آمارگیری به دو دسته سرشماری و نمونه‌گیری تقسیم می‌شوند. به منظور برآورد یک مشخصه در جامعه می‌توان از سرشماری یا نمونه‌گیری استفاده کرد. سرشماری از جامعه متناهی، بررسی‌ای است که تمام واحدهای جامعه را در بر می‌گیرد و باید گفت در بسیاری از موارد، اجرای سرشماری در یک جامعه متناهی، کاری شدنی است. به طور مثال در موسسه‌ای که ۵۰ نفر کارمند دارد می‌توان با بررسی پرونده‌های استخدامی آن‌ها سنوات خدمت تمام آن‌ها را فهرست کرد که در واقع یک سرشماری انجام شده است. وقتی حجم جامعه متناهی و بزرگ باشد انجام سرشماری به دلیل وقتگیر و پرهزینه بودن غالبا عملی نیست. بنابراین برای بررسی

یک مشخصه به نمونه‌گیری متوسل می‌شوند. نمونه بخشی از جامعه تحت بررسی است که با روشی که از پیش تعیین شده است انتخاب می‌شود، به قسمی که می‌توان از این بخش، استنباط‌هایی درباره کل جامعه بدست آورد. تعیین روش آمارگیری سرشماری یا نمونه‌گیری و یا انواع مختلف روش‌های نمونه‌گیری می‌تواند بر داده‌های بدست آمده و نتایج تجزیه و تحلیل تاثیر بسزایی بگذارد، به عبارتی دیگر اگر در یک مطالعه شرایط نمونه‌گیری تصادفی طبقه‌ای برقرار باشد بدیهی است که انتخاب نمونه از هر روش دیگری مثلا خوشه‌ای نماینده واقعی جامعه را بدست نخواهد داد، به عنوان مثال اگر بخواهیم در مورد متوسط درآمد شهروندان شهر مشهد تحقیقی انجام بدهیم و دسته‌بندی افراد جامعه در شهر مشهد را بر حسب مناطق شهرداری درنظر بگیریم، در هر منطقه شهرداری افرادی با کمترین درآمد و بیشترین درآمد زندگی می‌کنند، بنابراین شرایط نمونه‌گیری تصادفی خوشه‌ای برقرار هست، واضح است که در اینجا اگر به روش نمونه‌گیری تصادفی طبقه‌ای نمونه انتخاب شود، نماینده واقعی جامعه را به دست نخواهد داد.

۳ روش های تعیین حجم نمونه

در انجام هر تحقيق علمي، موضوع تعيين حجم نمونه بسيار اهميت دارد و نتايج بدست آمده از تحقيق بستگی به انتخاب يک نمونه مناسب با حجم كافي خواهد داشت. حداقل حجم نمونه مورد نياز بايد براساس اهداف اصلي تحقيق برآورد شود. پژوهش‌ها و مطالعات كاربردي بر اساس اهداف اصلي به دو دسته، مطالعات معطوف به مقدار (مطالعات توصيفی) و مطالعات معطوف به تصميم (مطالعات توصيفی-تحليلی) تقسيم می‌شوند. يكي از راه‌هاي برآورد حجم نمونه در پژوهش‌هاي معطوف به مقدار، استفاده از روش فاصله اطمينان است. در اين روش به دليل اينكه در بسياري از موارد از تقريب نرمال براي برآورد حداقل حجم نمونه مورد نياز استفاده مي‌شود، طبيعي است كه حجم نمونه كافي بزرگ باشد. در مطالعات معطوف به تصميم، علاوه بر توصيف، بدنبال استنتاج در مورد اهداف، فرضيه‌ها يا سئوالات مطرح شده نيز هستيم. در اين مطالعات حداقل حجم نمونه مورد نياز، بايستی براساس آزمون فرضيه‌های مورد استفاده تعيين شود. در اين صورت، چنانچه با توجه به مقادير مفروض احتمال خطاهاي نوع اول و دوم و ساير اطلاعات مربوط به جامعه و آزمون فرضيه‌ها، حداقل حجم نمونه مورد نياز برآورد گردد، امكان بزرگ بودن احتمال رخ دادن يكي از خطاهاي تصميم گيري نسبت به ديگري وجود نخواهد داشت. چنانچه حجم نمونه كمتر از ميزان لازم در نظر گرفته شود، ممكن است نتايج استنباط شده از آن در مورد جامعه از دقت كافي برخوردار نباشد و امكان تعميم نتايج حاصل از نمونه به جامعه وجود نداشته باشد. بسیاری از پژوهشگران در علوم مختلف جهت تعیین حجم نمونه مورد نیاز برای انجام تحقیق‌های خود از جدول مورگان استفاده می‌کنند. باید گفت استفاده از این جدول برای هر هدفی نادرست است. در واقع برای هر روش آماری تجزیه و تحلیل داده‌ها، یک روش تعیین حجم نمونه وجود دارد. کتاب تکنیک‌های نمونه‌گیری از ویلیام جی.کوکران دارای فرمول‌های زیادی جهت محاسبه حجم نمونه مورد نیاز و هرکدام مختص موضوع خاصی می‌باشد و فرمول کوکران که بسیار مشهور است و اکثرا استفاده می‌کنند تنها یکی از فرمول‌های این کتاب است، که این فرمول مربوط به برآورد حجم نمونه در برآورد نسبت می‌باشد. بنابراین نمی‌توان برای هر هدفی در مطالعات از فرمول کوکران استفاده نمود. توصیه می‌شود با توجه به هدف و روش آماری موردنیاز جهت پاسخ به فرضیه‌های

تحقیق از نرم افزارهای R، NCSS & PASS یا G-Power برای تعیین حجم نمونه مناسب استفاده گردد.

۴ تناسب حجم نمونه با حجم جامعه

معمولا تصور می‌شود که هر چه حجم جامعه بزرگتر باشد حجم نمونه نیز باید به همان نسبت بیشتر و متناسب با بزرگی جامعه باشد. ولی البته چنین نیست. یکی از تصورات شهودی نادرست این است که اگر (به عنوان مثال) از جامعه‌ای دارای ۱۰,۰۰۰واحد، نمونه‌ای به حجم ۵۰ لازم است، از جامعه‌ای دارای ۱۰۰,۰۰۰ واحد، نمونه‌ای به حجم ۵۰۰ نیاز هست. در واقع جزء اولین دستآوردهای آمار استنباطی (آمار مبتنی بر حساب احتمالات) این بود که این درک شهودی نادرست است. به عنوان یک مثال ساده فرض کنید می‌خواهیم نسبت P را با دقت d و با ضریب اطمینان $(۱-\alpha)$ از جامعه‌ای که دارای N واحد است، برآورد کنیم. حجم نمونه لازم از فرمول

$$n = \frac{z^۲ P(۱-P)/d^۲}{۱ + (z^۲ P(۱-P)/d^۲)/N},$$

به دست می‌آید، که در آن z عددی از جدول توزیع نرمال می‌باشد. جدول زیر مقادیر حجم نمونه لازم را به ازای $P = ۰/۵$، $d = ۰/۷$ و مقادیر مختلف حجم جامعه به دست می‌دهد: با استفاده از نتایج بدست آمده

N	۱۰	۱۰۰	۵۰۰	۱۰۰۰	۱۰۰۰۰	۱۰۰۰۰۰	۱۰۰۰۰۰۰	∞
n	۱۰	۶۶	۱۴۱	۱۶۴	۱۹۲	۱۹۶	۱۹۶	۱۹۶
$۱۰۰ * (n/N)$	۱۰۰	۶۶	۲۸,۲	۱۶,۴	۱,۹۲	۰,۱۹۶	۰,۰۱۹۶	۰

مشاهده می‌شود که هر چه حجم جامعه بزرگتر می‌شود حجم نمونه مورد نیاز به همان نسبت بیشتر نخواهد شد و از یک مرحله‌ای به بعد حجم نمونه لازم به ازای بزرگتر شدن حجم جامعه، ثابت باقی می‌ماند. بنابراین عباراتی از قبیل "نمونه ای به حجم ۱۰% جامعه برای دقت مورد نظر کافی است" بدون اینکه اطلاعی از حجم جامعه داشته باشیم بسیار گمراه کننده است و نباید بکار برد.

۵ مشکل حجم نمونه بسیار زیاد

در میان پژوهشگراني که از روش‌هاي آماري براي رفع نيازهاي علمي حيطه تخصصي خود بهره مي‌گيرند، اين برداشت نادرست وجود دارد که هر چه حجم نمونه بزرگتر باشد (يا به حجم جامعه نزديکتر باشد)، نتايج بهتري حاصل خواهد شد. اما بايد توجه داشت که همواره در نمونه‌هاي با حجم بزرگ، تضميني براي اتخاذ تصميم منطبق بر واقعيت وجود ندارد. اکثر کسانی که تجربه‌ای در تحلیل داده‌های آماری دارند مي‌دانند يکي از مسائل اساسي که به دنبال انتخاب نمونه‌اي با حجم بزرگ، ممکن است حادث شود ، اين است که توان آزمون‌هاي آماري مورد استفاده در مطالعات تحليلي، بسيار افزايش مي‌يابد، که اين موضوع در عمل باعث مي‌شود فرضيه صفر به نادرستي رد شود و شانس اين که نتايج " معنی‌دار " شوند بيشتر است. به عبارتی دیگر با توجه به اینکه معمولاً سطح معنی‌داری برابر ۰,۰۵ و گاهی هم (به ندرت) برابر ۰,۰۱ در نظر

گرفته می‌شود، زمانی که حجم نمونه بسیار بزرگ است، احتمال پذیرش نادرست فرضیه H_0 به مراتب از۰٫۰۱ و امثال آن کمتر است. پس راه حل چیست؟ راه حل معقول این است که هیچگاه به بدست آوردن یک نتیجه معنی‌دار اکتفا نکنیم بلکه همراه با انجام آزمون، تفاوت میانگین دو جامعه (یا مقدار ضریب همبستگی) را نیز برآورد کنیم و در صورتی که آن را ناچیز تشخیص دهیم، نتیجه را غیر معنی‌دار اعلام نماییم. به عنوان مثال محققی، زمانی نمرات ریاضی دانش‌آموزان دو ناحیه آموزش و پرورش را مقایسه کرده و تفاوت میانگین نمرات ریاضی دو ناحیه را بسیار بسیار معنی‌دار یافته و از ناهماهنگی آموزش ریاضی در دو ناحیه به شدت نگران شده بود. اما بعد از اینکه متوجه شد تفاوت میانگین دو نمونه کمتر از ۰٫۰۳ بوده است و معنی‌داری فوق العاده این تفاوت به خاطر وجود چند هزار دانش‌آموز در هر گروه بوده است، نگرانی‌اش برطرف شد. زمانی که از p-مقدار جهت رد یا پذیرش فرضیه صفر استفاده می‌شود، باید گفت p-مقدار در واقع بیان می‌کند که آیا نتایج در اثر شانس رخ داده است یا خیر و به خودی خود نمی‌تواند ملاک خوبی برای نشان دادن معنی‌دار بودن یا نبودن رابطه باشد و همیشه باید با اندازه اثر (بزرگی و جهت تفاوت، زمانی که دو گروه مقایسه می‌شوند) و فاصله اطمینان ۹۵% برای تفاوت (فاصله اطمینان حدودی را می‌دهد که ما انتظار داریم مقدار واقعی جامعه در آن محدوده قرار بگیرد) همراه باشد. بنابراین p-مقدار، اندازه اثر و فواصل اطمینان برای اندازه اثر باید در کنار هم، برای رسیدن به نتایج معنی‌دار، گزارش شوند. همچنین زمانی که حجم نمونه بسیار زیاد باشد، ممکن است سوالاتی از جمله اینکه ۱ـ آیا وجود یا عدم وجود همبستگی بین دو متغیر یا وجود یا عدم وجود تفاوت بین دو میانگین جامعه به این بستگی دارد که ما چه حجم نمونه‌ای به کار ببریم؟ ۲ـ آیا اگر حجم نمونه ما به اندازه کافی بزرگ باشد، می‌توانیم از معنی‌دار شدن نتایجمان اطمینان داشته باشیم؟ در ذهن محقق شکل بگیرد که در پاسخ به سوالات مطرح شده باید گفت گرچه جواب سوال اول مسلماً (و عقلاً) منفی می‌باشد ولی جواب سوال دوم متاسفانه مثبت است. اما باید دید که چرا چنین است؟ دلیل آن بسیار ساده است و آن این واقعیت است که در عمل (در مورد متغیرهای پیوسته) هیچگاه تفاوت دو میانگین جامعه (یا ضریب همبستگی دو متغیر تصادفی) صفر نیست. حجم نمونه کوچک قادر به کشف تفاوت‌های کوچک (ضریب همبستگی‌های نزدیک صفر) نیست، در صورتی که حجم نمونه بزرگ به راحتی کوچک‌ترین تفاوت‌ها (کمترین ضرایب همبستگی) را ظاهر می‌سازد. البته باید به این موضوع نیز اشاره نمود که یک اندازه نمونه کوچک لزوما باعث یک مطالعه ضعیف نمی‌شود بلکه توانایی تعمیم دادن و کشف نتیجه درباره جمعیت مورد نظر دشوارتر می‌شود. همچنین، با اندازه نمونه‌های کوچک، باید همواره مراقب بود که در مورد قدرت شواهد موجود اغراق نشود یا نتیجه‌ای فراتر از آنچه که مشاهده شده، گرفته نشود.

۶ جامعه هدف، جامعه در دسترس

جامعه هدفpopulation target جامعه‌ای است که نتایج به دست آمده از نمونه بایستی به آن تعمیم داده شود. باید اذعان داشت دسترسی به جامعه هدف به صورت مستقیم امکان‌پذیر نخواهد بود. در واقع محقق زیرگروهی از جامعه هدف را که به آن جامعه مورد مطالعهpopulation study یا جامعه منبعsource population گفته می‌شود، در اختیار دارد و در این راستا بایستی نماینده‌representative گویایی از

جامعه هدف باشد. نمونه‌ایpopulation sample که جهت انجام تحقیقات جمع‌آوری می‌شود، ازجامعه مورد مطالعه حاصل می‌شود که این نمونه بایستی بگونه‌ای انتخاب شود که وضعیت جامعه مورد مطالعه را به خوبی به تصویر بکشد. موردی که بایستی بدان اشاره نمود این هست که گرفتن نمونه از بخشی از جامعه و تعمیم آن به کل جامعه بسیار متداول است که این کار اغلب نادرست بوده و اگر آگاهانه انجام شود، تقلب محسوب می‌شود. ولی تحت بعضی شرایط نه تنها اشکالی ندارد بلکه تنها راه انجام دادن پژوهش آماری موردنظر است. مثلاً گرفتن نمونه از شرکتهای پذیرفته شده در بورس اوراق بهادار تهران که در مشهد مستقرند، برای برآورد میزان سود کل شرکت‌های عضو بورس، کار درستی نیست. یا به عنوان مثالی دیگر گرفتن نمونه از کارمندان، برای برآورد میزان محبوبیت کاندیدایی که مدافع حقوق زنان است، کار درستی نیست. ولی گرفتن نمونه از دانش‌آموزان ابتدایی شهر مشهد و بررسی تاثیر تنبیه بدنی بر پیشرفت تحصیلی و تعمیم نتایج به کل کلان شهرهای ایران ممکن است قابل توجیه باشد. در مثال فوق، نتایج را می‌توان به دانش‌آموزان ابتدایی کلان شهرها در سال‌های آینده نیز تعمیم داد. در واقع دلیل انجام چنین تحقیقی فقط می‌تواند امکان همین تعمیم اخیر باشد. تحت شرایط مناسب، حتی می‌توان نتایج پژوهشی را که در مورد کارگران یک کارخانه انجام شده است، به همه کارخانه‌های مشابه در کشور تعمیم داد. معمولاً پژوهش در مورد پارامترهای دو متغیری (نظیر استقلال، همبستگی و رگرسیون) از یک زیرجامعه به کل جامعه بلا اشکال است، چون، برخلاف پارامترهای یک متغیری، روابط بین متغیرها از زیرجامعه به جامعه خیلی تغییر نمی‌کنند. یکی از اشتباهات دیگری که خیلی رایج است وقتی پیش میآید که زیرجامعه (جامعه در دسترس) کاملاً سرشماری میشود. در این موارد پژوهشگر با استناد به این که تمام جامعه سرشماری شده است، از هر گونه تحلیل آماری سر باز می‌زند و به آمار توصیفی بسنده می‌کند. در صورتی که جامعه اصلی (جامعه هدف) بسیار بزرگ‌تر است و برای آن نیاز به استنباط داریم.

۷ عدم استفاده از میانگین در شرایطی که صحیح نیست

میانگین راحت‌ترین معیاری است که مرکزیت داده‌ها را نشان می‌دهد ولی همواره مناسب‌ترین شاخص نیست. می‌دانیم که در جامعه‌های با توزیع شدیداً چوله، استفاده از میانگین برای نشان دادن متوسط جامعه مناسب نیست. برای مثال، جمله"میانگین پاداش آخر سال کارمندان دولت ۲۹۰ هزارتومان است" ، با توجه به این که تعداد کمی وزیر و مدیر ارشد با پاداش بسیار زیاد و تعداد زیادی کارمند جزء با پاداش بسیار کم وجود دارد، تصویر نادرستی از توزیع پاداش‌ها به خواننده می‌دهد. اشتباه دیگری که در پژوهش‌های آماری رایج‌تر است این است که بعضی اقدام به میانگین‌گیری از مقادیر متغیرهایی می‌کنند که در مقیاس ترتیبی یا اسمی هستند. می‌دانیم که در مقیاس ترتیبی، بر خلاف مقیاس‌های فاصله‌ای و نسبتی، فواصل مساوی دارای ارزش یکسان نیستند و لذا میانگین‌گیری از مقادیر آن‌ها منطقی نیست. البته در این موارد (مقیاس رتبه‌ای) استفاده از میانه اشکال ندارد.

۸ همبستگی های گمراه کننده

یکی دیگر از اشتباهات رایجی که در استفاده از مفاهیم آماری وجود دارد و بایستی بدان پرداخت، مبحث همبستگی است که در این قسمت به مفاهیمی چون همبستگی جعلی، استنباط علیت معکوس از همبستگی، همبستگی صوری و همبستگی بسیط یا تعدیل نشده پرداخته می‌شود.

۱.۸ همبستگی جعلی

به عنوان مثال فرض کنید در یک پژوهش، محققی همبستگی بین "عیار چغندرهای کرت A " و " عیار چغندرهای کرت B " را محاسبه و اعلان کرده است. برای کرت A از بذر (الف) و برای کرت B از بذر (ب) استفاده کرده است. تنها وجه تشابه (ارتباط) دو کرت در این است که هر یک دارای ۳۶ بوته است. مثال دیگر، محاسبه همبستگی بین "درآمد حسابداران مرد " و " درآمد حسابداران زن " در شرکت‌های خصوصی است. چنین همبستگی‌هایی نه تنها قابل محاسبه نیستند، بلکه غیرمنطقی بوده و قابل تعریف نیز نمی‌باشند. برای تعریف (و محاسبه) همبستگی دو متغیر، لازم است آن دو متغیر روی واحدهای آماری یک مجموعه (نمونه آماری) تعریف شده باشند. مثل همبستگی بین وزن و عیار چغندرهای یک کرت. (واحد آماری = چغندر) یا مثل همبستگی بین درآمد حسابداران یک بیمارستان با درآمد همسران‌شان. (واحد آماری = خانواده)

۲.۸ استنباط علیت معکوس از همبستگی

در بحث مباحث همبستگی، گرفتن نتیجه علیتی به راحتی امکان‌پذیر نیست به طور مثال، آیا از این که " رتبه گروه‌های آمار دانشگاه‌ها با رتبه دانشگاه‌ها همبستگی قابل توجهی دارد "، می‌توان نتیجه گرفت که کیفیت یک دانشگاه به کیفیت گروه آمار آن بستگی دارد؟ یا این که درست برعکس است و واقعیت این است که دانشگاه‌های خوب، گروه‌های آمار با کیفیت دارند؟ متاسفانه باید بیان نمود آمار هیچ راهی برای پاسخ دادن به این سوال ندارد. تنها زمانی که دو متغیر تقدم و تاخر زمانی داشته باشند تشخیص علت و معلول ساده است در غیر این صورت نمی‌توان از رابطه همبستگی بین دو متغیر استنباط علیتی داشت.

۳.۸ همبستگی صوری

بسیار پیش می‌آید که همبستگی نسبتاً بالایی بین دو متغیر مشاهده می‌شود، در حالی که هیچ یک بر دیگری تاثیر ندارد، بلکه وجود متغیر سومی در میان است. برای مثال همبستگی بالایی بین جویدن آدامس و ارتکاب تخلف‌های رانندگی مشاهده می‌شود، در صورتی که هیچ یک علت یا معلول دیگری نیست، بلکه بی‌قراری و عصبی مزاج بودن فرد، علتی برای هردو محسوب می‌شود و با هر دو همبستگی واقعی دارد.

۴.۸ همبستگی بسیط یا تعدیل نشده

گاهی تاثیر چندین متغیر مستقل بر متغیر وابسته، به خاطر وجود همبستگی بین خود آن‌ها، به سادگی قابل تفکیک نیست. مثال واقعی زیر از مجله Statistician چاپ آمریکا سال ۱۹۸۲ نقل می‌شود. بر اساس ارقام مربوط به X= فروش سالانه مواد دخانی (سیگار، سیگار برگ و توتون پیپ) و Y= تعداد موارد گزارش شده از سرطان ریه در سال. رشد پیوسته و یکنواخت فروش سالانه مواد دخانی از سال ۱۸۸۰ تا سال ۱۹۸۰ میلادی، و همراه با آن رشد پیوسته و یکنواخت تعداد سالانه موارد سرطان ریه و در نتیجه همبستگی بسیار بالایی (برابر با ۰,۹۸) بین X و Y مشاهده می‌شود. اما این یک همبستگی ساده (بسیط) است و همبستگی واقعی (خالص) زمانی به دست می‌آید که تاثیر سایر متغیرهایی که در همین بازه زمانی دارای رشد بوده، و بر Y نیز موثر بوده‌اند، را حذف کنیم. این کار به کمک محاسبه همبستگی جزئی (Partial Correlation) X و Y امکان‌پذیر است. در طول دوره فوق، متغیرهای دیگری نیز بر Y موثر بوده‌اند که همگی نیز در این مدت در حال رشد بوده‌اند، نظیر افزایش فاضلاب شیمیایی کارخانه‌جات، تشعشعات رادیویی، تلویزیونی، میکروویو، آزمایش‌های هسته‌ای، استرس حاصل از افزایش شهرنشینی، افزایش استفاده از مواد افزودنی و نگه‌دارنده و بسیاری از متغیرهای سرطان‌زای دیگر، که متاسفانه ارایه فهرست کاملی از آن‌ها، امکان‌پذیر نیست. پس از تعدیل ضریب همبستگی به وسیله حذف اثرات متغیرهای نام برده شده، ضریب همبستگی جزئی (واقعی، خالص) بین X و Y برابر ۰,۴۶ به دست می‌آید، که البته باز هم کم نیست ولی با مقدار اولیه ۰,۹۸ تفاوت زیادی دارد.

۹ بروز اشتباه در اثر ادغام داده ها

این مطلب را با ذکر یک مثال شروع می‌کنیم، بدین صورت که در اواسط جنگ جهانی دوم، کارمندی در وزارت دفاع آمریکا متوجه شد که سهم زنان در تک تک صنایع آمریکا، نسبت به قبل از جنگ، افزایش یافته است، در حالی که سهم زنان در کل صنایع آمریکا، نسبت به قبل از جنگ، کاهش یافته است ! مثال عددی زیر امکان بروز تعارض‌هایی از این دست را ثابت می‌کند:
گرچه تفاوت درصد باکیفیت بالا در دو گروه (سطح محافظه‌کاری) معنی‌دار نیست، ولی ارقام فوق با تصور قبلی در مورد اثر مثبت داشتن سطح کیفیت اطلاعات حسابداری بر سطح محافظه‌کاری مغایرت دارد. اما زمانی‌که اطلاعات به تفکیک نوع شرکت (خدماتی و غیرخدماتی) جمع‌آوری شود،
"تغییر مقدار یا تغییر جهت همبستگی در اثر ادغام داده‌ها" را تعارض سیمپسون (Simpson's paradox) می‌نامند. توالی همبستگی صوری، همبستگی تعدیل نشده و تعارض سیمپسون در این متن تصادفی نبوده است، زیرا دو مورد قبلی حالات خاصی از مورد سوم هستند.

۱۰ اطمینان از درستی نتیجه یک آزمون مبتنی بر p-مقدار

اگر آزمونی را در سطح (مثلاً) ۰,۰۵ انجام دهید و فرضیه $H_.$ رد نشود (پذیرفته شود)، نمی‌توانید ادعا کنید که ۹۵%به نتیجه آزمون اطمینان دارید. ببینیم چرا؟ عدد ۰,۰۵ در بالا احتمال خطای نوع اول است، یعنی

آزمون طوری انجام شده است که احتمال رد کردن H_0، در صورتی که H_0 درست باشد، ۰٫۰۵ است. لذا اگر H_0 رد می‌شد، یک نوع "اطمینان" نسبی به درستی H_1 وجود داشت، ولی H_0 رد نشده (پذیرفته شده) است و میزان "اطمینان" به درستی H_0 به احتمال خطای نوع دوم بستگی دارد که معمولاً معلوم نیست و در مواردی که فرضیه H_1 مرکب است، مقدار مشخصی ندارد. لذا در این شرایط، اطمینانی به صحت پذیرش H_0 نمی‌توان داشت. حتی زمانی که H_0 رد می‌شود، گرچه یک "اطمینان" نسبی به درستی H_1 وجود دارد، اما نمی‌توان گفت که "احتمال درستی H_1 برابر ۹۵% است" زیرا احتمال درستی H_1 روی فضای H_0, H_1 تعریف می‌شود، در صورتی که احتمال رد H_0 روی فضای متغیر تصادفی مشاهده شده (X) تعریف می‌شود. و این دو فضای احتمال، کاملاً متفاوت‌اند. در واقع اگر پژوهشگر به دنبال این باشد که ببیند داده‌ها چقدر از فرضیه H_0 یا از H_1 پشتیبانی می‌کنند، باید از روش‌های استنباط شواهدی استفاده نماید.

۱۱ نداشتن مشاهده در تمام سطوح یک متغیر

محققی ضمن تحقیق در مدارس کودکان استثنایی اصفهان، مشاهده کرده بود که والدین ۳۱% از کودکان عقب مانده ذهنی، خویشاوند هستند و نتیجه گرفته بود که این بررسی آماری آنچه را که علم ژنتیک به آن رسیده است (یعنی تاثیر سوء ازدواج خویشاوندی بر سلامت ذهنی کودکان) تایید می‌کند. این دامی است که خیلی‌ها در امثال آن گرفتار می‌شوند (غفلت از دیدن روی دیگر سکه). این محقق چون ۳۱% به نظرش درصد بالایی از ازدواج خویشاوندی آمده است، چنین استنباطی کرده است. در صورتی که اگر می‌دانست درصد ازدواج خویشاوندی در کل جامعه نیز نزدیک به همین مقدار است (صرف نظر از این که چنین تاثیری واقعاً وجود دارد یا نه) چنین نتیجه‌ای نمی‌گرفت. یا (مثلاً) اگر درمی‌یافت که در کل جامعه این درصد برابر ۳۵% است نتیجه‌ای کاملاً برعکس می‌گرفت. لذا برای نتیجه‌گیری‌هایی از این قبیل، باید به درصد ازدواج خویشاوندی در هر دو گروه (کسانی که ازدواج خویشاوندی داشته‌اند و کسانی که نداشته‌اند) دسترسی داشته باشیم. به طور کلی، استنباط‌هایی که بر اساس یک یا بخشی از سطوح یک متغیر انجام می‌شود ناقص بوده و می‌تواند گمراه کننده باشد.

۱۲ استفاده نکردن از روش‌های چند متغیری

در بسیاری از مطالعات، چندین متغیر روی هر واحد نمونه، اندازه‌گیری می‌شود. این گونه داده‌ها را داده‌های چند متغیری می‌نامند. مقایسه میانگین‌های این متغیرها در دو (یا چند) نمونه را باید با روش‌های تحلیل چند متغیری انجام داد. گاهی داده‌های چند متغیری را یکی‌یکی و با استفاده از روش‌های یک متغیری بررسی می‌کنند. این کار به معنی دور ریختن بخش مهمی از اطلاعات، که در ساختار کوواریانس متغیرها نهفته است، می‌باشد، که گاه منجر به نتایج نادرست می‌شود. برای مثال فرض کنید اندازه‌گیری قد و وزن ۶۴ نفر به نتایج زیر منجر شده است و هدف این است که میانگین قد و وزن افراد به ترتیب با اعداد ۱۷۲ و ۶۸

مقایسه شود:

$$\overline{X} = (۶۸, ۱۷۲), \quad S = \begin{bmatrix} ۱۱۲ & ۵۵ \\ ۵۵ & ۱۱۴ \end{bmatrix}, \quad \begin{cases} H_۰ : (\mu_۱, \mu_۲) = (\mu_{۱۰}, \mu_{۲۰}) = (۶۶, ۱۷۴) \\ H_۱ : (\mu_۱, \mu_۲) \neq (۶۶, ۱۷۴) \end{cases}$$

از آزمون چند متغیری $T^۲$ی هتلینگ استفاده می‌کنیم، که نتیجه می‌شود:

$$T^۲ = n(\overline{X} - \mu_۰)' S^{-۱} (\overline{X} - \mu_۰) = ۱۰/۸۴ > ۱۰/۰۹ * \frac{(n-۱)p}{(n-p)} F_{۰/۰۱}(۲/۶۴).$$

مشاهده می‌شود که فرضیه صفر آزمون رد خواهد شد. حال فرض کنید که کوواریانس مشاهده شده را ندیده گرفته و به جای آزمون چند متغیری، دو آزمون یک متغیری برای فرضیه‌های $H_۰ : \mu_۱ = ۶۶$ و $H_۰ : \mu_۲ = ۱۷۴$ انجام دهیم. نتیجه می‌شود:

$$|T^۱| = \frac{|۶۸ - ۶۶|}{\sqrt{\frac{۱۱۲}{۶۴}}} = ۱/۴۲ < ۲, \quad |T^۲| = \frac{|۱۷۲ - ۱۷۴|}{\sqrt{\frac{۱۱۴}{۶۴}}} = ۱/۸۸ < ۲.$$

که بر اساس این محاسبات هیچ یک از دو فرضیه صفر رد نمی‌شوند. بنابراین نباید آزمون‌های تک متغیری را جایگزین آزمون‌های چند متغیره کرد و به جای هم به کار برد.

۱۳ عدم تشخیص لانه‌ای بودن طرح مطالعه

طرح آزمایشی در مطالعات آزمایشی فوق العاده اهمیت دارد. برای اینکه بخواهیم در هر طرح اندازه‌گیری روش آماری مناسب را انتخاب کنیم بایستی به ماهیت آن طرح توجه زیادی داشته باشیم. چون در خیلی از موارد ممکن است یک طرح در واقع لانه‌ای باشد ولی با استفاده از روش‌های آنالیز واریانس بلوکی اثر تیمار بررسی شود، یا ممکن است طرح لانه‌ای باشد اما از طرح کرت خرد شده استفاده شود و یا بالعکس. به عنوان مثال برای مقایسه سود ۴ محصول شیمیایی (a، b، c، d) از ۳ کارخانه (A، B، C)، در هر کارخانه از هر یک از محصولات ۳ نمونه تصادفی گرفته و در هر نمونه میزان سود اندازه‌گیری شده است. اگر دقت کافی لحاظ نشود، ممکن است برای تحلیل داده‌ها از یک طرح دو عاملی استفاده شد، که به جدول آنالیز واریانس زیر منجر می‌شود:

ملاحظه می‌شود که سود مواد به طور معنی‌داری متفاوت است. به علاوه اثر متقابل (معنی‌دار) به عامل

منبع تغییرات	SS	df	MS	F	*sig*
کارخانه‌ها	۱۵,۱	۷,۵	۲	۱,۰۲	۰,۳۸
مواد	۲۵,۶	۳	۸,۶	۳,۲۴	۰,۰۳۸
اثر متقابل	۴۴,۳	۶	۷,۴	۲,۸۰	۰,۰۳۳
خطا	۴۳,۳	۲۴	۲,۶		

غیر معنیدار کارخانه‌ها ربط داده شده و میتواند تحلیل‌گر را به این نتیجه‌گیری سوق دهد که کارخانه‌ها واقعاً متفاوت‌اند ولی اثر آن‌ها توسط اثر متقابل پنهان شده است. اما با کمی دقت متوجه می‌شویم که باید داده‌ها

را حاصل یک طرح لانه‌ای ببینیم. ماده شیمیایی تولید شده توسط کارخانه A همان ماده شیمیایی تولید شده توسط کارخانه B یا C نیست. بنابراین اثر متقابل بین کارخانه و مواد معنی ندارد. مواد شیمیایی هر کارخانه منسوب به همان کارخانه است و نه به کارخانه‌های دیگر. اصطلاحاً، عامل مواد در داخل عامل کارخانه لانه کرده است. با استفاده از طرح لانه‌ای به جدول آنالیز واریانس زیر می‌رسیم:
ملاحظه می‌شود که با اعمال روش صحیح تفاوت بین سود مواد شیمیایی معنی‌دار نیست. بنابراین بایستی

منبع تغییرات	SS	df	MS	F	sig
کارخانه‌ها	۱۵,۱	۷,۵	۲	۱,۰۲	۰,۳۸
مواد	۶۹,۹	۹	۷,۸	۲,۲۴	۰,۰۵۶
خطا	۴۳,۳	۲۴	۲,۶		

در انتخاب و تشخیص طرح اندازه‌گیری دقت شود و نتایج واقعیت وضعیت موجود را نشان دهند.

۱۴ عدم تشخیص طرح اندازه‌گیری‌های مکرر

به طور کلی می‌توان گفت هرگاه اثر K تیمار روی یک متغیر پاسخ، مورد تحقیق باشد، باید توجه داشت که آیا آن K تیمار در K نوبت متوالی روی یک مجموعه از افراد (یا اشیاء) اعمال می‌شود مثل اجرای یک طرح کلان اقتصادی (مثلا هدفمندی یارانه‌ها) در کشور در ۳ دوره زمانی مختلف و مجزا یا روی K مجموعه مختلف از افراد (یا اشیاء) اعمال می‌شود. مثل اعمال سه طرح کلان اقتصادی برای ۳ دسته روستاها، شهرها و کلان شهرها. در مورد اول یک طرح اندازه‌گیری‌های مکرر داریم و در مورد دوم با طرح آنالیز واریانس یک‌طرفه مواجه هستیم. حالت خاص اندازه‌گیری‌های مکرر وقتی است که K=۲ باشد، که در آن صورت نام طرح "مقایسه میانگین‌های زوج شده" است، که آن نیز گاهی با طرح "مقایسه میانگین‌ها بر اساس دو نمونه مستقل" اشتباه می‌شود. زمانی که متغیر پاسخ نیز یک متغیر کیفی است، باید داده‌های طرح مقایسه‌های زوج شده را با "آزمون تقارن در جدول‌های توافقی (مک نمار، کاپا یا کاپای وزنی یعنی آزمون Bowker)" تحلیل کنیم. این طرح نیز اغلب با طرح "آزمون استقلال در جدول‌های توافقی" اشتباه می‌شود. در این حالت وقتی K=۲ است، باید از آزمون مک نمار استفاده شود.

۱۵ استفاده از روش آماری نامنطبق بر طرح اندازه‌گیری

اغلب محققان در مطالعاتی که می‌خواهند تاثیرگذار بودن یا نبودن یک مداخله را روی یک صفت بررسی نمایند و اندازه‌گیری‌ها در قبل و بعد از مداخله انجام می‌شود، از آنالیز کواریانس استفاده می‌کنند، بدین صورت که مقادیر صفت در قبل از مداخله را به عنوان کووریت و بعد از مداخله را به عنوان متغیر پاسخ در نظر می‌گیرند. باید گفت این مطالعات نوع خاصی از آنالیز واریانس اندازه‌های مکرر نوع II (هم اثر درون‌گروهی و هم اثر بین‌گروهی وجود داشته باشد) می‌باشد و استفاده از آنالیز کواریانس در چنین مسائلی صحیح نمی‌باشد. برای روشن شدن این موضوع و دلایل نادرست بودن استفاده از آنالیز کواریانس به جای

طرح آنالیز واریانس اندازه‌های مکرر نوعII، سه طرح زیر را در نظر بگیرید، فرض کنید تعدادی از افراد را براساس حجم نمونه‌ای که تعیین شده است در اختیار دارید، این افراد را می‌توان به طور تصادفی به ۲ گروه یکی گروه کنترل و دیگری گروه آزمایشی تقسیم نمود. بر روی افراد درگروه کنترل هیچ مداخله‌ای انجام نمی‌شود اما در گروه آزمایش مداخله انجام می‌شود. سپس صفت موردنظر (که باید مستقیما اندازه‌گیری شود) مثلا میزان HDL افراد در هر دو گروه کنترل و آزمایش اندازه‌گیری می‌شود. این طرح را طرح شماره ۱ مینامیم. این دو گروه جدا و مستقل از هم می‌باشند. اگر هدف بررسی تاثیر مداخله بر میزان HDL افراد باشد، باید آزمونی را که برای مقایسه میانگین گروه‌های مستقل (دو یا چند گروه) استفاده می‌شود، بکار برد. گاهی شرایطی پیش می‌آید که نمی‌توان افراد را به دو گروه تقسیم نمود مثلا محدودیتی که وجود دارد این است که تعداد افراد زیادی در دسترس نیست (مثلا دو یا سه نفر در دسترس هستند)، طرح شماره ۲ به این صورت طراحی می‌شود که قبل از اینکه افراد در یک دوره تمرینی شرکت کنند تست HDL از آن‌ها گرفته شود سپس مداخله روی افراد انجام شود و پس از انجام مداخله مجدد از آنها تست HDL گرفته شود. در این طرح باید از آزمون‌هایی برای مقایسه دو گروه زوجی استفاده شود، چون افراد در دو گروه یکی هستند. اگر دو طرح شماره ۱ و ۲ را با هم مقایسه کنیم خواهیم دید منطق حاکم بر مقایسه این دو طرح با هم فرق می‌کند. در طرح ۱ باید همه افراد گروه کنترل را با همه افراد گروه آزمایش مقایسه شوند اما در طرح شماره ۲ پس آزمون هر فرد با پیش آزمون خودش مقایسه می‌گردد. البته طرح شماره ۱ حالت کلی‌تری نسبت به طرح شماره ۲ می‌باشد. مشکل بزرگی که در طرح شماره ۱ وجود دارد این است که ممکن است افرادی که به تصادف در گروه کنترل نسبت به گروه آزمایش قرار گرفتند وضعیت HDL بهتری را قبل از اعمال مداخله داشتند. پس یکی از دلایلی که ممکن است گفته شود مداخله بر HDL تاثیر داشته یا نداشته در واقع همگن نبودن دو گروه می‌باشد. در طرح شماره ۲ هم گفته می‌شود که هر فردی با خودش مقایسه شده و ویژگی‌های فردی در این کار دخالت داشته است و ممکن است به این دلایل تاثیر مداخله معنی‌دار شده یا نشده است. بنابراین هریک از این دو طرح یک مشکل اساسی دارند. حال ترکیبی از این دو طرح را در نظر بگیرید و طرح شماره ۳ بنامید. در طرح شماره ۳ ابتدا افراد به طور تصادفی به دو گروه تقسیم می‌شوند و از هر دوی این گروه‌ها پیش آزمون گرفته می‌شود و سپس در گروه آزمایش مداخله موردنظر و در گروه کنترل مداخله‌نما (طرح‌های یک‌سو کور و دوسو کور و سه‌سو کور هم در اینجا مطرح هست که در اینجا وارد بحث‌های کارآزمایی بالینی نمی‌شویم) اعمال می‌گردد. سپس از هر دو گروه پس آزمون گرفته می‌شود، حال در اینجا از یک دیدگاه دو گروه زوجی داریم و از یک دیدگاه دو گروه مستقل داریم. در این طرح اگر بخواهیم اثر مداخله را بررسی نماییم، باید چکار کرد و چه آزمونی انجام داد؟ ۵ آزمون می‌توان انجام داد، پیش آزمون دو گروه را با هم و پس آزمون دو گروه با هم مقایسه شود (دو آزمون تی مستقل)، پیش آزمون و پس آزمون گروه کنترل با هم و پیش آزمون و پس آزمون گروه آزمایش با هم مقایسه شود (دو آزمون تی زوجی)، آزمون دیگر اینکه اختلاف پیش آزمون و پس آزمون در دو گروه با هم مقایسه شود (یک آزمون تی مستقل). واضح است که انجام این ۵ آزمون درست نمی‌باشند، یک دلیل این می‌باشد که سطح معنی‌داری (سطح خطا) زیاد می‌شود (البته این مورد قابل رفع می‌باشد یعنی به جای ۰,۰۵ با ۰,۰۱ مقایسه شود). اما ایراد اساسی که وجود دارد این است که در هر آزمون تی که انجام می‌شود اثر گروه‌های غایب در نظر گرفته نمی‌شود. زمانی که اختلاف پیش آزمون و پس آزمون با هم مقایسه می‌شود در واقع اثر پیش آزمون یا به عبارتی اثر مداخله از بین می‌رود. در

طرح شماره ۳ اثرات باید همزمان در نظر گرفته شود. در این طرح وقتی اثر مداخله معنی‌دار است که گروه کنترل و گروه آزمایشی در پیش آزمون با هم فرق نکنند اما در پس آزمون با هم فرق کنند. پس آزمون نسبت به پیش آزمون در گروه کنترل فرق نکند اما در گروه آزمایشی فرق کند. که البته اگر چند گروه و چند زمان وجود داشته باشد تفاسیر به مراتب متفاوت خواهد بود. اما در اینجا حالت ساده با دو گروه و دو زمان در نظر گرفته شده است. عده‌ای از افراد از آنالیز کواریانس برای تحلیل چنین مسائلی استفاده می‌کنند که باید اظهار نمود استفاده از این روش کاملا نادرست می‌باشد. در اینجا لازم است قبل از بیان دلایل نامناسب بودن استفاده از آنالیز کواریانس، ابتدا به تعریف متغیر هم‌عامل پرداخته شود. هم‌عامل متغیری کمی است که متفاوت از متغیر وابسته و همجنس نیستند و از طرفی با افزایش یا کاهش متغیر هم‌عامل متغیر وابسته نیز افزایش یا کاهش می‌یابد. دلیل اول اینکه استفاده از آنالیز کواریانس بجای آنالیز واریانس اندازه‌های مکرر اشتباه می‌باشد این است که متغیر پیش آزمون و پس آزمون از یک جنس هستند و این با تعریف متغیر هم‌عامل در تناقض است. دلیل دوم اینکه که در آنالیز واریانس واریانس متغیر وابسته به دو یا چند مولفه تجزیه می‌شود اما در آنالیز کواریانس، کواریانس بین متغیر وابسته و هم‌عامل تجزیه می‌شود یعنی در واقع ماتریس کواریانس به دو یا چند مولفه تجزیه می‌شود. در طرح شماره ۳ همبستگی بین پیش آزمون و پس آزمون اصلا مهم نیست، بلکه اختلاف پیش آزمون و پس آزمون مهم است که باید در گروه آزمایشی و نه در گروه کنترل معنی‌دار باشد. دلیل سوم اینکه طرح‌ها یا آینده‌نگر هستند یا گذشته‌نگر. طرح شماره ۳ یک طرح آینده‌نگر میباشد بنابراین باید به این نکته هم توجه کرد که وقتی طرح موردنظر آینده‌نگر باشد روش آماری نمی‌تواند گذشته‌نگر باشد در حالی‌که آنالیز کواریانس برای یک مطالعه گذشته‌نگر می‌باشد و متغیر هم‌عامل و متغیر وابسته باید همزمان اندازه‌گیری شده باشند نه اینکه یکی در یک زمان و دیگری در یک زمان دیگر یا چند روز بعد اندازه‌گیری شده باشند. بنابراین وقتی طرح موردنظر یک طرح آینده‌نگر باشد روش آماری هم باید قطعا آینده‌نگر باشد و روش‌های آینده‌نگر هم مطالعات طولی هستند. آنالیز کواریانس هم طولی نیست بلکه مقطعی است. آنالیز واریانس اندازه‌های تکراری حالت خاص وساده‌ای از مطالعات طولی است.

۱۶ نتیجه‌گیری

در نهایت چنین نتیجه می‌شود که وجود برخی اشتباهات رایج آماری در مقالات علمی حتی بهترین مجله‌ها به دلیل عدم تسلط کافی محققان بر مراحل انجام تحقیق و عدم آشنایی کامل آنها با فنون و مفاهیم مربوط به روش‌های آماری می‌باشد و باعث می‌گردد نتایج گزارش شده مورد اعتماد و قابل اتکا نباشد، از این رو به محققان توصیه می‌شود خطاها و اشتباهاتی که ممکن است در طی فرآیند تحقیق و یا به هنگام تجزیه و تحلیل، تفسیر و ارائه نتایج رخ دهند را تا حد ممکن درنظر گرفته و در جهت رفع آن‌ها بکوشند.

۱. اولین و مهمترین بخش فرآیند علمی آمار، موضوع تعیین حجم نمونه می‌باشد که از اهمیت بسیار بالایی برخوردار می‌باشد. به تبع تعیین حجم نمونه، روش‌های آمارگیری سرشماری و انواع روش‌های نمونه‌گیری مطرح می‌شوند که در نتایج حاصل تاثیر بسزایی خواهند داشت. طرح آمارگیری نمونه‌ای باید با توجه به خصوصیات، چارچوب جامعه آماری و همچنین اهداف تحقیق تعیین شود و هر طرح آمارگیری نمونه‌ای را نمی‌توان برای هر جامعه آماری به کار برد. طبیعتا باید دید که شرط‌های مربوط به طرح‌های آمارگیری نمونه‌ای

چه هستند و اگر آن شرایط بر اساس هدف وجود داشته باشند از طرح مناسب استفاده کرد و در شرایطی که لازم هست سرشماری نیز صورت گیرد. پس از مشخص نمودن طرح آمارگیری، در صورت نیاز به نمونه‌گیری، بایستی حجم نمونه لازم براساس نوع مطالعه و اهداف تحقیق تعیین شود زیرا حجم نمونه خیلی بزرگ یا خیلی کوچک در صحت نتایج تاثیر گذاشته و ما را از واقعیت دور می‌کند.

۲. مسئله دیگری که باید به آن توجه نمود این است که روش‌های آمار توصیفی و آمار استنباطی به صورت نادرست، به جای هم استفاده نشوند. در برخی شرایط در صورتی که جامعه در دسترس، کاملا سرشماری می‌شود، تنها به ارائه آمار توصیفی بسنده نشود و از روش‌های آمار استنباطی جهت پی بردن به اطلاعات و شناخت جامعه سرشماری شده استفاده شود، زیرا جامعه هدف بسیار بزرگتر می‌باشد و دسترسی به آن امکانپذیر نمی‌باشد. جامعه‌ای که کاملا سرشماری شده است زیرگروهی از جامعه هدف می‌باشد که در دسترس محقق بوده است و در مرحله بعد برای جامعه هدف نیاز به استنباط داریم.

۳. در تفسیر نتایج به دست آمده از روشهای آمار استنباطی دقت لازم به عمل آید و از تفسیر نادرست آنها اجتناب شود یا به عبارتی دیگر تفاوت بین آزمون فرضیه‌های آماری و آزمون‌های معنی‌داری به خوبی درک و ارتباط آنها با استنباط شواهدی در نظر گرفته شود. به عنوان مثال در حجم نمونه‌های بالا همواره فرضیه صفر در مقابل فرضیه مقابل رد می‌شود و از این رو بایستی p ـ مقدار همراه با اندازه اثر و فاصله اطمینان برای اندازه اثر جهت رد یا عدم رد فرضیه صفر گزارش شود. ضمن این که وقتی گفته می‌شود فرضیه صفر رد می‌شود یا رد نمی‌شود فقط یک اصطلاح است و هیچ آزمون آماری درستی یا نادرستی فرضیه صفر را تضمین نمی‌کند. در آمار کلاسیک از p ـ مقدار برای استنباط شواهدی استفاده می‌شود که برای این کار مناسب نیست و باید از روش‌های آمار شواهدی استفاده نمود زیرا هیچ یک از معایبی که p ـ مقدار به آن دچار است را ندارد، بلکه دارای مزیت‌هایی است که p ـ مقدار چنین مزایایی را ندارد.

۴. در تشخیص طرح اندازه‌گیری و روش آماری مبتنی بر آن طرح، تلاش لازم صورت پذیرد و از به کار بردن روش آماری نامناسب یا نامنطبق بر طرح اندازه‌گیری اجتناب گردد. به عنوان مثال، دقت شود که طرح‌های لانه‌ای، آنالیز واریانس بلوکی، طرح‌های کرت خرد شده و یا ... به اشتباه به جای هم به کار برده نشوند. زیرا در نتایجی که به دست می‌آید به شدت تاثیرگذار و این نتایج فاقد اعتبار خواهند بود. موضوع دیگری که بسیار مهم است و اغلب محققین در استفاده از آن دچار اشتباه می‌شوند مربوط به استفاده از آنالیز کواریانس به جای طرح آنالیز واریانس اندازه‌های مکرر در مطالعاتی که مقادیر یک صفت به صورت قبل و بعد از مداخله جمع‌آوری می‌شود، می‌باشد. باید تعریف دقیق متغیر هم عامل را بشناسند و از پیش آزمون به عنوان متغیر هم عامل استفاده نکنند. خصوصا اینکه مطالعات قبل و بعد حالت خاصی از آنالیز واریانس اندازه‌های مکرر (یک مطالعه طولی آینده‌نگر) می‌باشد و نباید به جای آن از آنالیز کواریانس که یک مطالعه مقطعی گذشته‌نگر می‌باشد، استفاده نمود.

مراجع

[۱] ارقامی، ن. ر. ؛ ۱۳۷۴، بعضی اشتباهات رایج در تحقیقات علوم تربیتی و اجتماعی، *گردهم‌آیی اعضاء شوراهای تحقیقات کشور*، تبریز

[۲] برومیده، ع. ا. ؛ ۱۳۸۴، با برخی از اشتباهات رایج در تحلیل های آماری آشنا شویم، *اندیشه آماری*، سال نهم شماره اول

[۳] . ارقامی، ن. ر. ؛ ۱۳۸۰، سنجری فارسی پور، ن. و بزرگنیا، ا. ، *مقدمه ای بر بررسی های نمونه ای*، انتشارات دانشگاه فردوسی

[۴] اسماعیلی، ح. ؛ ۱۳۷۸، اشتباهات رایج آماری در پروژه های تحقیقاتی، *اولین سمینار دانشجویی آمار ایران*، دانشگاه شهید بهشتی

[۵] خردمندنیا، م. ؛ ۱۳۷۷، عواقب نادیده گرفتن کوواریانس ها، *اندیشه آماری*، سال سوم شماره اول

[۶] مشکانی، م. ر. ؛ ۱۳۷۰، *آمار مقدماتی*، جلد اول، چاپ دوم، مرکز نشر دانشگاهی

[۷] موسوی، س. ج. ؛ سرمد، م. ؛ ۱۳۸۰، فریب های آماری، پروژه کارشناسی آمار

[۸] ارقامی، ن. ر. ؛ زمانزاده، ا. ؛ راحتی، س. ؛ ۱۳۸۸، شواهدآماری (۱)، *اندیشه آماری*، سال ۱۳ شماره اول

[۹] ارقامی، ن. ر. ؛ دست برآورده، ع. ؛ زمانزاده، ا. ؛ ۱۳۸۹، شواهدآماری (۲)، *اندیشه آماری*، سال ۱۳، شماره دوم

[۱۰] جباری نوقابی، ه؛ جباری نوقابی، م، نکاتی در برآورد حجم نمونه و معرفی نرم افزار مربوطه، *ندا*، شماره دوم، سال چهارم، ۱۳ - ۲۱

[۱۱] عمیدی، ع. ؛ ۱۳۸۱، *روشهای نمونه گیری۱* ، چاپ چهارم، انتشارات دانشگاه پیام نور

[۱۲] ذوق دار مقدم، م. ؛ ۱۳۸۹، نقش استنباط شواهدی در تشخیص توزیعهای نرمال آمیخته، پایان نامه کارشناسی ارشد، رشته آمار ریاضی

[۱۳] ارقامی، ن. ر. ؛ زمانزاده، ا. ؛ ۱۳۸۹، آزمون فرضیه، آزمون معنی داری و اندازه گیری شواهد آماری، *دهمین کنفرانس آمار ایران*

[14] Zinsmeister AR, Connor JT. Ten common statistical errors and how to avoid them. *The American journal of gastroenterology.* 2008;103(2):262-6.

[15] Lang, Tom. "Twenty statistical errors even you can find in biomedical research articles." *Croatian medical journal* 45 4 (2004): 361-70 .

[16] Lang T. Common statistical errors even you can find, I errors in descriptive statistics and in interpreting probability values2012.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

تفسیر نتایج آزمونهای کلاسیک آماری

زمان زاده، ا [۱] دست برآورده، ع [۲] ارقامی، ن.ر [۳]

[۱]گروه آمار، دانشگاه اصفهان
[۲]گروه آمار، دانشگاه یزد
[۳]گروه آمار، دانشگاه فردوسی مشهد

چکیده

در این مقاله قصد داریم که به بررسی درستی یا نادرستی تفسیر های متداول از آزمون های کلاسیک اماری که غالبا بر مبنای لم نیمن ـ پیرسون ساخته می شود بپردازیم و سپس اینکه چگونه روش درست تفسیر آزمون ها باید بااستفاده از هدف محقق تفسیر شود را مورد بحث قرار می دهیم.

کلمات کلیدی: لم نیمن- پیرسون، خطای نوع اول، خطای نوع دوم، قانون درستنمایی، شواهد آماری و استنباط شواهدی

۱ پیش‌گفتار

جملات زیر نمونه ای از تفسیرهای متداول و رایج از نتایج آزمون های کلاسیک آماری هستند.

۱. فرضیه H_0 در سطح ۰/۰۵ رد می شود پس ۰/۹۵ درصد مطمئن هستیم که این فرضیه نادرست است.

۲. فرضیه صفر پذیرفته می شود و چون آزمون در اندازه $\alpha = 0.01$ بوده است به درستی این فرضیه که میانگین درآمد خانم ها و آقایان باهم برابر است اطمینان زیادی داریم.

۳. چون مقدار $p - value$ آزمون برابر با ۰/۰۳ شده است پس ۰/۹۷ اطمینان داریم که احتمال ابتلا به سرطان خون به جنسیت بستگی دارد.

[۱] ehsanzamanzadeh@yahoo.com

۴. چون مقدار $p-value$ آزمون همبستگی برابر با ۰/۰۰۰ شده است، درآمد شدیدا به ضریب هوشی بستگی دارد.

در این مقاله می خواهیم به بررسی این مساله بپردازیم که چرا تفسیرهایی از جنس آنچه که در فوق ارائه شد، نادرست است و چگونه نتیجه یک آزمون آماری باید با استفاده از هدف محقق تفسیر شود.

شایان ذکر است که تمام مطالب ارائه شده در این مقاله، برگرفته شده از مقاله (۱) می باشد که با تغییرات اندک و به طور خلاصه شده در نخستین سمینار استنباط شواهدی، ارائه شده است.

۲ هدف از انجام آزمون های نیمن- پیرسونی

نظریه نیمن- پیرسون نظریه ای است که بسیاری از آزمون های آماری بر مبنای آن ساخته شده اند. در روش نیمن-پیرسون آزمون به نحوی انجام می شود که در آن احتمال خطای نوع اول برابر با مقدار کوچک و از قبل تعیین شده ای مثل α باشد. این نظریه به آزمونی منجر می شود که هرچند مقدار خطای نوع دوم درآن از قبل تعیین نمی شود اما این خطا کمترین مقدار ممکن را اختیار می کند و با افزایش حجم نمونه به سمت صفر میل می کند. از سویی دیگر با حجم نمونه ثابت هر چندقدر خطای نوع اول (α) را کوچکتر اختیار کنیم مقدار خطای نوع دوم (β) بزرگتر خواهد شد. اینکه محقق چه مقدار را برای α و در نتیجه برای β بایستی انتخاب کند، بستگی به این دارد که چه زیان هایی را هنگام ارتکاب این خطاها متحمل خواهد شد. مساله ای که در نظریه نیمن- پیرسون به آن پاسخ داده می شود این هست که وقتی خطای نوع اول و حجم نمونه مشخص است چگونه باید آزمون را انجام داد که خطای نوع دوم آن می نیموم شود. چون توان آزمون برابر با $\beta - ۱$ است، آزمونی که از لم نیمن- پیرسون نتیجه می شود پرتوان ترین آزمون نامیده می شود. مساله مهمی که باید به آن توجه کرد آن است که هرچند احتمالات خطاهای آزمون نیمن-پیرسون قابل تفسیر هستند (به این مفهوم که ۱۰۰٪ α موارد تحت $H_۰$ و ۱۰۰٪ β موارد تحت $H_۱$ منجر به نتیجه اشتباه می شوند) اما نتیجه یک آزمون به این صورت قابل تفسیر نمی باشد. نتیجه یک آزمون نیمن-پیرسونی چیزی غیر از یکی از دو جمله "رد $H_۰$" یا "قبول $H_۰$" نیست و هیچگونه تفسیری بر آن وارد نمی باشد. واضح است که ردفرضیه صفر به معنی نادرست بودن آن و پذیرفتن فرضیه صفر به معنی درستی آن نیست. لذا تفسیر صحیح یک آزمون نیمن-پیرسونی چیست؟ نظر خود نیمن (۲) در مورد تفسیر یک آزمون به صورت زیر است

- مساله آزمون فرضیه زمانی پیش می آید که شرایط ما را وادار کرده است که از بین دو عمل A و B یکی را انتخاب کنیم.
- بسیار مهم است که معنی دقیق پذیرفتن فرضیه صفر، که عبارت است از این که عمل A را و نه عمل B باید انجام دهیم، به خاطر بسپاریم.

در واقع در نظریه نیمن- پیرسون فرض بر آن است که محقق مردد است که از بین دو عمل A و B کدام را انجام دهد ولی می داند که اگر $H_۰$ درست باشد بهتر است عمل A و اگر $H_۱$ درست باشد بهتر است عمل B را انجام دهد. لذا برای تصمیم گیری و خروج از تردید اقدام به جمع آوری داده و انجام دادن ازمون می کند. مثال های زیر را درنظر بگیرید:

- بیمار را تحت درمان قراردهیم یا نه؟
- اجازه پخش فلان دارو را صادر کنیم یا نه؟
- نسبت به حفر چاه آب در محل مورد نظر اقدام کنیم یا نه؟
- محموله رسیده را بپذیریم یا نه؟

در مثال های فوق، محقق فرضیه صفر (مفید بودن درمان، مفید بودن دارو، وجودآب در محل، و قابل قبول بودن محموله) را در مقابل نقیض آن $H_۱$ آزمون می کند. در صورت پذیرش $H_۰$ عمل A (تحت درمان قرار دادن، صدور اجازه پخش، حفر چاه و پذیرش محموله) و در صورت رد آن، عمل B (نقیض عمل A) را انجام خواهد داد. در اینجا هدف محقق (که آن را هدف ۱ می نامیم) پاسخ به سوال زیر است:

- با توجه به داده ها کدامیک از دو عمل A و B را باید انجام داد؟

در جهت می نیموم کردن رسیکی که از این طریق متوجه محقق است، محقق می خواهد که احتمالات خطاهای نوع اول و دوم تا حد امکان کمترین باشد.

مجددا این نکته را تاکید می کنیم که ویژگی هایی که بهینه بودن آزمون های کلاسیک بر اساس آنها استوار است، یعنی اندازه و توان آزمون، قبل از انجام آزمون و حتی قبل از آنکه به جمع آوری داده ها اقدام کنیم، معلوم و مشخص اند و لذا این آزمون ها غیر از اینکه "$H_۰$ رد شد" یا "$H_۰$ پذیرفته شد" تفسیر دیگری ندارد. بنابراین هیچ کدام از چهار تفسیری که در مقدمه ذکر شد در نظریه نیمن-پیرسون وجود ندارد.

در اینجا باید به نکته مهم دیگری نیز در مورد نقش $p-value$ در پاسخ به سوال یک اشاره کنیم این است که برای رسیدن به جواب نهایی آزمون (یعنی رد یا قبول $H_۰$) از مراجعه به جداول آماری بی نیاز باشیم و با توجه به آنچه که از نیمن نقل کردیم، در جهت نیل به هدف یک، $p-value$ هیچ گونه تفسیر دیگری ندارد.

۳ احتمال درستی فرضیه صفر

فرض کنید هدف از انجام آزمون نیمن-پیرسونی، پاسخ به سوال زیر باشد

- با توجه به داده ها احتمال درستی فرضیه $H_۰$ و یا $H_۱$ چقدر است؟

استفاده از آزمون های نیمن -پیرسونی برای هدف فوق، نادرست و یا دست کم ناکارآمد می باشد، زیرا برای نیل به این هدف باید احتمال درستی $H_۰$ و $H_۱$ را از نظر محقق، قبل از انجام آزمون بدانیم، زیرا با استفاده از قانون بیز می توان نوشت

$$P(H_۰|AH_۰) = \frac{(۱-\alpha)P(H_۰)}{(۱-\alpha)P(H_۰)+\beta P(H_۱)}$$

که در آن $AH_۰$ به معنی پذیرش فرضیه صفر است.

ملاحظه می شود که برای یافتن احتمال درستی فرضیه صفر، باید نسبت $O=\frac{P(H_۰)}{P(H_۱)}$ را بدانیم و این در صورتی است که در نظریه نیمن-پیرسون صحبتی از احتمالات پیشین نشده است. از این گذشته حتی اگر این

بخت را برابر با یک درنظر بگیریم، هنوز استفاده از آزمون های نیمن-پیرسونی برای پاسخ به سوال مطرح شده در این بخش ناکارآمد است زیرا معمولا تابع آزمون یک آماره بسنده نیست و استفاده از تابع آزمون بجای کل داده ها باعث عدم استفاده از کلیه اطلاعات موجود در داده ها می شود و لذا نتیجه یک آزمون کلاسیک به هر صورت که ارائه شود مقدار احتمال مطلوب در فوق را در برندارد.

در ادامه به استدلال نادرستی که متداول ترین نوع استفاده از آزمون های نیمن-پیرسون هست اشاره می کنیم. چنین استدلال می شود که

- چون اگر H_0 درست باشد، احتمال رد H_0 کوچک می باشد پس وقتی H_0 رد می شود، به احتمال زیاد H_0 نادرست است.

استدلال فوق که قانون احتمال کم نام دارد، در واقع تعمیم غیرمجاز اصل عکس نقیض در منطق است که می گوید: "اگر وقوع مشاهده D در صورت درست بودن فرضیه H امکان پذیر نباشد، در این صورت مشاهده D فرضیه H را باطل می کند".

قانون احتمال کم، که تعمیم نادرست اصل فوق است، می گوید:

"اگر $P(D|H) = \epsilon$ و ϵ عدد کوچکی باشد، آنگاه مشاهده D باعث می شود که اطمینان زیادی به نادرست بودن فرضیه H داشته باشیم."

۴ شواهد آماری به عنوان هدف

در بسیاری از پژوهش های کاربردی که منجر به چاپ مقاله در نشریات علمی می شود، هدف محقق، که آن را هدف سوم می نامیم، پاسخ به سوال زیر است:

- داده ها از کدام فرضیه (فرضیه H_0 یا H_1) بیشتر پیشتبانی می کنند و به چه میزان؟

برای بررسی نادرستی استفاده از قانون احتمال کم، حتی وقتی هدف ۳ مدنظر باشد، یعنی برای اینکه ببینیم صرفا کوچک بودن $P(D|H_0)$ شواهد قوی بر علیه H_0 وقتی که D مشاهده می شود نیست، به مثال های زیر توجه کنید.

۱. فرض کنید متغیر تصادفی X دارای توزیع دوجمله ای $B(۱۲۰, p)$، $H_0 : p = \frac{۱}{۲}$ و $D : X = ۵۰$ باشد. آنگاه $P(X = ۵۰|H_0) = ۰/۰۱۳$ ولی علی رغم کوچکی این احتمال، نه تنها مشاهده $X = ۵۰$ شاهدی بر نادرست بودن H_0 نیست بلکه از آن حمایت نیز می کند.

۲. اگر متغیر تصادفی X دارای توزیع $N(\theta, ۱)$ باشد و $H_0 : \theta = ۰$ و $D : X = ۰/۳۸۶$ باشد که در آن D یک مشاهده از توزیع $N(\theta, ۱)$ است که تا سه رقم اعشار گرد شده است. آنگاه $P(D|H_0) = ۰/۰۰۰۳۷۰$ می باشد. چون این احتمال کوچک است، آیا می توان گفت مشاهده فوق شاهدی قوی برعلیه H_0 است؟ مسلما خیر.

۳. فرض کنید یکی از دو جعبه A و B به تصادف انتخاب می شود و یک مهره از ان به تصادف انتخاب می شود. اگر مهره انتخاب شده قرمز باشد و بدانیم جعبه A دو مهره قرمز و ۹۸ مهره سیاه دارد. آیا

می توان گفت که قرمز بودن مهره انتخاب شده قویا بر رد فرضیه صفر "مهره از جعبه A انتخاب شده است" دلالت می کند؟ مسلما نمی توان به این سوال پاسخ داد مگر انکه بدانیم ترکیب مهره ها در جعبه دیگر چیست. اگر جعبه B دارای ۵۰ مهره سفید و ۵۰ مهره قرمز باشد، آنگاه قرمز بودن مهره انتخاب شده، شاهدی قوی بر علیه $H_.$ است. اما اگر جعبه B دارای یک مهره قرمز و ۹۹ مهره سفید باشد، مسلما قرمز بودن مهره انتخاب شده نه تنها شاهدی بر علیه $H_.$ نیست بلکه از آن حمایت نیز می کند. اگر جعبه B دارای هیچ مهره قرمز نباشد، بی شک قرمز بودن مهره انتخاب شده، فرضیه صفر را نه تنها رد نمی کند، بلکه اثبات هم می کند.

۵ نسبت درستنمایی

از مثال های فوق می توان چنین نتیجه گرفت که برای اینکه D شاهدی قوی بر علیه $H_.$ باشد صرفا کوچک بودن $P(D|H_.)$ کافی نیست بلکه باید دید که $P(D|H_۱)$ چقدر است. $H_۱$ فرضیه ای است که قوت D را به عنوان شاهدی بر علیه $H_.$ در مقابل آن باید سنجید. به $H_۱$ فرضیه جانشین نیز گفته می شود.

در واقع، میزان پشتیبانی مشاهده D از فرضیه $H_.$ فقط در مقابل یک فرضیه دیگری مانند $H_۱$ معنی دارد و شدت آن را می توان با مقدار

$$r = \frac{P(D|H_.)}{P(D|H_۱)}$$

که برابر با نسبت درستنمایی است، اندازه گیری کرد.

در مثال نخست، هیچ فرضیه جانشینی وجود ندارد که به ازای آن $D : X = ۵۰$ شاهدی قوی بر علیه $H_.$ باشد.

در مثال دوم، اگر $H_۱ : \theta = a$ فرضیه جانشین باشد، آنگاه مقدار نسبت درستنمایی به ازای $a = ۰/۱, ۱, ۲$ به ترتیب برابر است با

$$r = \frac{P(D|H_.)}{P(D|H_۱)} = ۰/۹۵, ۱/۱۱, ۳/۴۲$$

یعنی میزان پشتیبانی مشاهده $D = ۰/۳۸۶$ شدیدا به فرضیه جانشین بستگی دارد.

حال با توجه به مطالب بیان شده در فوق، چنانچه بخواهیم بر اساس رد یا قبول فرضیه صفر، میزان پشتیبانی داده ها از $H_.$ در مقابل $H_۱$ تعیین کنیم، نه تنها مقدار خطای نوع اول، بلکه میزان خطای نوع دوم را نیز باید بدانیم و تنها بخاطر کوچک بودن خطای نوع اول، نمی توان رد فرضیه صفر را شاهدی قوی بر علیه آن و یا پذیرش فرضیه صفر را به عنوان شاهدی قوی بر علیه $H_۱$ تفسیر کرد.

حال اگر بخواهیم از نسبت درستنمایی به عنوان میزان پشتیبانی از یک فرضیه، در مقابل فرضیه دیگر استفاده کنیم، پشتیبانی داده ها از فرضیه $H_۱$ در مقابل $H_.$ هنگامی که $H_.$ رد می شود، برابر است با

$$r_۱ = \frac{P(RH_.|H_۱)}{P(RH_.|H_.)} = \frac{۱-\beta}{\alpha}$$

و میزان پشتیبانی داده ها از $H_.$ در مقابل $H_۱$ زمانی که $H_.$ پذیرفته می شود برابر است با

$$r_۲ = \frac{P(AH_.|H_.)}{P(AH_.|H_۱)} = \frac{۱-\alpha}{\beta}$$

جدول ۱، مقادیر مختلف r_1 و r_2 را به ازای $\alpha = ۰/۰۵$ و مقادیر مختلف β نشان می دهد. چنانکه از این جدول واضح است، هنگامی که فرضیه صفر پذیرفته می شود، در مقایسه با هنگامی که این فرضیه رد می شود، تفسیر نتیجه حاصل خیلی بیشتر متاثر از مقدار β است. به عبارت دیگر زمانی که فرضیه صفر پذیرفته می شود، در بیان اینکه "داده ها از فرضیه صفر قویا پشتیبانی می کنند" باید بسیار احتیاط کرد، زیرا میزان شواهد پشتیبان فرضیه صفر می تواند از ۱/۹ تا بی نهایت تغییر کند.

اما وقتی که فرضیه صفر رد می شود، میزان پشتیبانی داده ها بر علیه آن از ۱۰ تا ۱۲۰ (به ازای $\beta = ۰/۵$ تا $\beta = ۰$) تغییر می کند، که در همه موارد نسبت درستنمایی بر علیه فرضیه صفر، مقدار قابل توجه است.

جدول ۱: مقادیرمختلف r_1 و r_2 به ازای $\alpha = ۰/۰۵$ و مقادیر مختلف β

β	۰	۰/۰۱	۰/۰۵	۰/۰۱	۰/۲	۰/۵	۰/۹۵
r_2	۱۲۰	۱۹۸	۱۹	۱۸	۱۶	۱۰	۱
r_1	∞	۹۵	۱۹	۹/۵	۴/۷۵	۱/۹	۱

۶ نتیجه گیری

مطالب ارائه شده در این مقاله را می توان به صورت زیر خلاصه کرد:

- وقتی که فرضیه صفر رد می شود و مقدار خطای نوع اول کوچک است، شواهد قوی بر علیه فرضیه صفر وجود دارد و صحت این تفسیر خیلی به مقدار خطای نوع دوم بستگی ندارد.
- هنگامی که فرضیه صفر پذیرفته می شود، شدت شواهد بر علیه این فرضیه، حتی هنگامی که مقدار خطای نوع اول کوچک است، شدیدا به مقدار خطای نوع دوم بستگی دارد. پس اگر از مقدار خطای نوع دوم اطلاع نداریم بهتر است به گفتن اینکه "شواهد موجود در داده ها بر علیه $H_.$ قاطع نیست، اکتفا کنیم."

مراجع

[۱] ارقامی ن.ر، زمان زاده، الف. و راحتی، س. (۱۳۸۷). شواهد آماری، تفسیر نتایج آزمون های آماری کلاسیک، *اندیشه آماری*، ۱، ۱۰-۳.

[2] Neyman. J. (1950), *First course in probability and statistics*, Henry & Holt Company, New York.

اولین سمینار
استنباط شواهدی
۲۲ خرداد ۱۳۹۸

برخی راه حل ها برای مشکل p ـ مقدارها

اتله خانی، م [۱]

[۲]گروه آمار، دانشکده علوم ریاضی و آمار، دانشگاه ملایر، ملایر

چکیده

این مقاله به چند راه حل که برای رفع مشکل استفاده از p ـ مقدارها پیشنهاد شده است می پردازد و اینکه این راه حل ها نمی توانند جایگزین خوبی برای p ـ مقدارها باشند. در انتها نیز پیشنهاداتی برای حل مشکل p ـ مقدارها ارایه می شود. از جمله اینکه بجای استفاده از آزمون فرضیه ها از روشهای بیزی استفاده شود و در تحلیل های آماری سعی شود از نگاه مطلق نگری دور شده و تحلیل هایی که همراه با عدم قطعیت است مورد توجه قرار گیرد.

کلمات کلیدی: p- مقدارها، فاکتور بیز، معنی‌داری آماری، فاصله اطمینان.

۱ پیش‌گفتار

در بسیاری از موارد مشاهده می‌شود که حتی دانشمندان مجرب به طور مرتب از pـ مقدارها استفاده نموده و آنها را بد تفسیر می‌کنند. استفاده از کلمه پذیرش بجای عدم رد فرضیه صفر و تفسیر pـ مقدار به عنوان یک نوع احتمال پشتیبانی از فرضیه صفر. از طرفی رایج است که رد فرضیه صفر را به عنوان شاهدی برای پذیرش فرضیه مقابل در نظر می گیرند (۱). به عنوان نمونه در برخی موارد $p-value < ۰/۰۵$ را به عنوان شاهدی قوی در رد فرضیه صفر در نظر می‌گیرند و $p-value > ۰/۱۵$ را به عنوان شاهدی به نفع فرضیه صفر و p نزدیک ۰/۱ را شاهدی ضعیف برای یک اثر یا شاهدی از یک اثر ضعیف در نظر می‌گیرند. متاسفانه هیچکدام از این نتیجه‌گیری‌ها در حالت کلی درست نیست. یک p ـمقدار کم لزوما شاهدی قوی در رد فرضیه صفر نیست (۱۱) و یک pـمقدار بالا لزوما به نفع فرضیه صفر نیست. لذا pـمقدارها معیار سنجش

[۱] m-akhani@yahoo.com

اندازه اثرهای اصلی نیستند. این خطاها می‌توانند دارای دو منشا باشند: اول اینکه مشکلات اجتناب ناپذیری در ریاضیات و منطق pـ مقدارها وجود دارد. دوم اینکه محققان تمایل دارند تا از داده‌هایشان نتیجه‌گیری‌های قوی‌تری داشته باشند. البته مفاهیم دیگری از آمار مانند "درجه آزادی " نیز بطور روزمره توسط محققان استفاده می‌شود که نتایج آن دارای اختلال است و همچنان منتشر می‌شوند (۶). شاید به نظر برسد که این مشکلات از کمبود آموزش آمار است و با اضافه کردن یک پاراگراف به کتابهای درسی و متقاعد کردن متخصصان کاربردی به مشورت بیشتر با آماردانان مشکل حل می‌شود. اما به نظر می‌رسد که مشکل اصلی تفکر قطعی‌نگری است و اینکه محققان کاربردی و آماردانان عادت کرده‌اند که اطمینان بیشتر را از داده‌هایی که در دست دارند بدست آورند. مثلا در مورد اینکه pـ مقدارها نمی‌توانند برای دسته‌بندی شواهد به روش دلخواه بکار گرفته شوند می‌توان به (۸) مراجعه نمود. مشکل این است که از تحلیل آماری چیزی خواسته می‌شود که قادر نیست به سادگی آن را انجام دهد، در حالی که داده‌ها چنین اطلاعاتی را ندارند. لذا بهتر است تا آماردانان در برخی از ادعاهایشان در مورد اثر بخشی روش‌هایی که ساخته‌اند تجدید نظر کنند.

۲ برخی از راه حل‌ها و اشکالات آنها

۱.۲ آموزش آمار و توجه بیشتر به آماردانان

اگر آمار بتواند یک راه‌حل خوب برای مشکل معنی‌داری ارایه نماید بسیار زیباست و فقط نیاز است که این راه‌حل با وضوح بیشتری آموزش داده شود. در صورتیکه این طور نیست، بطور مثال (۱۰) نشان دادند که بسیاری از آماردانان ایده‌های اصلی را به درستی درک نکرده‌اند. لذا آموزش بیشتر فایده‌ای در بر ندارد اگر معلم نداند که چه چیزی را آموزش می‌دهد. مسئله این نیست که چیز درست را بصورت ضعیف آموزش می‌دهیم، متاسفانه ما چیز غلط را خیلی خوب آموزش می‌دهیم.

۲.۲ فواصل اطمینان جایگزین آزمون فرضیه‌ها

به طور معمول از فاصله اطمینان برای دو منظور استفاده می‌شود. اول این که آیا شامل صفر هست یا خیر؟ که این همان آزمون فرضیه است با نام دیگر. دوم، استفاده از فاصله اطمینان به عنوان بیانی از اطمینان در برآورد پارامتر مورد نظر. یعنی چند درصد از این بازه‌ها پارامتر را در بر می‌گیرند. اما فاصله اطمینان می‌تواند علی رغم اینکه از نظر آماری معنا دار است در واقع کاملا بی‌معنا باشد. برای مثال در امریکا بر اساس داده‌های یک نظرسنجی نسبت زنانی که به باراک اوباما رای می‌دهند را ۲۰ درصد براورد نمودند و انحراف معیار را ۰٫۰۸ گزارش کردند (۴) . یک فاصله اطمینان ۹۵ درصد برای این نسبت برابر با (۰٫۰۴, ۰٫۳۶) بدست می‌آید. این براورد از نظر آماری معنی دار است اما با توجه به (۵) هر دو کران بازه بی‌معنا است زیرا در کمپین‌های انتخاباتی تغییرات نظرات کم است. بنابراین اگرچه فواصل اطمینان شامل اطلاعاتی فراتر از pـ مقدارها هستند اما آنها قادر نیستند آن اطمینان بخشی که خارج از بحث براورد مد نظر محققان است را حاصل نمایند.

۳.۲ تمرکز بر "معنی‌داری کاربردی" بجای "معنی‌داری آماری"

در نگاه واقع بینانه همه فرضیه های آماری نادرست هستند: اثرات دقیقا صفر نیستند، گروهها دقیقا همتوزیع نیستند، توزیع‌ها واقعا نرمال نیستند، اندازه‌گیریها کاملا نااریب نیستند و ... بطور مثال در مطالعه اثر یک داروی ضد فشار خون مطالعه آماری نشان داد که این دارو فشار خون را به میزان $۰/۳mmHg$ با انحراف معیار ۰/۱ کاهش می‌دهد. به وضوح این برآورد یک تفاوت معناداری با صفر دارد. اما این کاهش در مقیاس واقعی بسیار کم است. لذا در یک مطالعه، مقایسه‌ها می‌توانند از لحاظ آماری معنی‌دار باشند اما از لحاظ اهمیت کاربردی و عملی کاملا بی‌معنا باشند. بنابراین تمایز بین معنی‌داری آماری و کاربردی حلال مشکل pـ مقدارها نیست.

۴.۲ فاکتور بیز

یکی از راههای که برای اصلاح pـمقدار پیشنهاد می‌شود این است که ایده آزمون فرضیه را حفظ کنید اما احتمالهای دمی را در نظر نگیرید و بجای آن از احتمالهای پسین برای فرضیه‌های صفر و مقابل با پیشین جفریز استفاده کنید (۱۲). مشکل این نظریه این است که درستنمایی‌های کناری از مدلها (و بنابراین فاکتور بیز و احتمال‌های پسین متناظر) به شدت به شکل توزیع پیشین وابسته است و بطور معمول به صورت اختیاری و توسط محقق انتخاب می‌شود. برای مثال: مسئله‌ای را در نظر بگیرید که توزیع پیشین یک توزیع نرمال با پارمتر مکان صفر و مقیاس ۱۰ باشد و پارامتری که می‌خواهد براورد شود دارای دامنه $(-۱, ۱)$ باشد. این توزیع پیشین اساسا بی‌معنا است. اطلاعات بیشتر در (۷) و (۲) آمده است.

بحث و نتیجه‌گیری

بهترین جایگزین برای آزمون فرضیه‌ها و pـمقدارها، استفاده از استنباط بیزی است. مثلا (۹) نحوه جایگزینی استنباط بیزی با آزمون t را بیان نموده است. اما در انتخاب توزیع پیشین باید دقت نمود زیرا استفاده از توزیع‌های پیشین با اطلاع می‌تواند دردسر ساز باشد. البته روشهای غیر بیزی هم وجود دارند که اطلاعات قبلی را در محاسبه احتمالهای پسین وارد نمایند که برای اطلاعات بیشتر به (۳) می‌توان مراجعه نمود. بطور خلاصه ما ترجیح می‌هیم که بطور کامل از آزمون فرضیه‌ها اجتناب شود. این کار باعث انکار یکی از اهداف اصلی علوم آماری خواهد شد، که طرفداران زیادی در علوم مختلف دارد و اغلب علاقه‌مند به رد فرضیه صفر و پذیرش فرضیه مقابل هستند. اما مشکل بزرگتری که در آموزش آمار وجود دارد ارتباط بین فرضیه‌های آماری با فرضیه‌ها و نظریه‌های علمی است که تقریبا همیشه یک مسئله باز است. را حل مشکل pـمقدارها حرکت به سمت پذیرش عدم قطعیت و پذیرفتن تغییرات است. لذا توصیه می‌شود که در همکاری و مشاوره پروژه‌ها به نتیجه‌های دودویی نه گفته شود و وقتی که داده‌ها واقعا هیچ ضمانتی به ما نمی‌دهند، در مقابل پاسخهای شفاف مقاومت شود. لذا لازم است که روی نمایش نتیجه‌گیری‌های آماری با عدم قطعیت کار شود. همچنین باید دقت شود که بیشتر "اثرات" نمی‌توانند صفر باشند (حداقل در علوم اجتماعی و پزشکی) و بحث در مورد وجود یا عدم وجود یک اثر، کار بیهوده‌ای به نظر می‌رسد.

مراجع

[1] Gelman, A. (2014), Confirmationist and falsificationist paradigms of science. Statistical Modeling, Causal Inference, and Social Science blog, 5 Sept. http://andrewgelman.com/2014/09/05/confirmationist-falsificationist-paradigms-science/

[2] Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., and Rubin D. B. (2013), *Bayesian Data Analysis*, third edition. London: Chapman and Hall.

[3] Gelman A., Carlin J. B. (2014), Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science.* 9, 641-651.

[4] Gelman A., Carlin J. (2017), Some Natural Solutions to the p-Value Communication Problem—and Why They Won't Work, *J. Amer. Statist. Assoc.* 112, 899-901.

[5] Gelman A., Goel S., Rivers D., and Rothschild D. (2016). The mythical swing voter, *Quarterly Journal of Political Science.* 11, 103-130.

[6] Gelman A., Loken E. (2014), The statistical crisis in science, *Amer. Scient* 102, 460–465.

[7] Gelman A., Rubin D. B. (1995). Avoiding model selection in Bayesian social research, *Sociological Methodology.* 165–173.

[8] Gelman A., Stern H. (2006), The difference between "significant" and "not significant" is not itself statistically significant, *Amer. Scient.* 60, 328–331.

[9] Kruschke J. K. (2013), Bayesian estimation supersedes the t test.*Journal of Experimental Psychology: General.* 142, 573-603.

[10] McShane B. B, Gal D. (2017), Statistical Significance and the Dichotomization of Evidence, *J. Amer. Statist. Assoc.* 112, 885-895.

[11] Morris, C. N. (1987), Comment, *J. Amer. Statist. Assoc.* 82, 131–133.

[12] Rouder J. N., Speckman P. L., Sun D., Morey R. D., and Iverson G. (2009), Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review.* 16, 225-237.