Pـ مقدار به عنوان معیاری برای پشتیبانی فرضها؛ محاسن و معایب

نویسنده: مارک جی. شرویش[1]

مترجمان: انوشیروان کاظم نژاد[2]، محمدرضا محبی، معصومه ثناگو[3]

چکیده

Pـ مقدارها (احتمالهای معنی‌داری) در آزمونهای فرض به عنوان وسیله‌ای برای فراهم آوردن اطلاعات بیشتر در مورد رابطهٔ بین آزمون فرض و داده‌ها به کار رفته‌اند و تنها یک معیار ساده برای تصمیم گیری دربارهٔ رد کردن یا رد نکردن فرض نیست. در واقع تمام کتابهای مقدماتی آمار، محاسبات مربوط به Pـ مقدار را در حالت فرضهای یکطرفه، دوطرفه و فرض صفر نقطه‌ای، به خصوص هنگامی که میانگین نمونه‌ای از جامعهٔ نرمال مد نظر باشد، ارائه می‌دهند. اما به حالت سومی که حالت بینابین حالت فرض یکطرفه و دوطرفه است، یعنی فرض فاصله‌ای اشاره نشده است. در این مقاله نشان می‌دهیم که Pـ مقدارها برای داده‌های معین، توابعی پیوسته از فرضها هستند، در این صورت می‌توانیم با هرسه نوع فرض، از نظر ریاضی یکسان برخورد کنیم. همچنین در ادامه نشان می‌دهیم به کارگیری غیر رسمی Pـ مقدار به عنوان معیار ارزشیابی یا شواهد و قرائن به نفع فرضها، دارای مشکلات منطقی و جدی می‌باشد.

کلمات کلیدی: شواهد و قرائن، Pـ مقدار، فرض بازه‌ای، فرض صفر تک نقطه‌ای، فرض یکطرفه، معیار پشتیبانی.

۱. مقدمه

فرض کنیم مشاهدهٔ X دارای توزیع نرمال با میانگین μ و واریانس واحد باشد. پارامتر M را میانگین جامعه فرض می‌کنیم. پس انواع متداول فرضها برای M به صورت $H_1: M=\mu_0$ در برابر فرض $A_1: M\neq\mu_0$ (فرض صفر تک نقطه‌ای)، فرض $H_2: M\leq\mu_0$ در برابر فرض $A_2: M>\mu_0$، فرض $H_3: M\geq\mu_0$ در برابر فرض $A_3: M<\mu_0$ (فرض یکطرفه)، و نوع دیگر آن به عنوان فرض بازه‌ای به شکل $H_4: M\in[\mu_1,\mu_2]$، در برابر فرض $A_4: M\notin[\mu_1,\mu_2]$ است. فرضهای صفر تک نقطه‌ای و یکطرفه، حالت حدی فرض بازه‌ای برای مواقعی است که کرانها به یکدیگر

---

[1] Mark J. Shervish

[2] دانشیار گروه آمار زیستی، دانشگاه تربیت مدرس

[3] دانشجوی کارشناسی ارشد آمار زیستی، دانشگاه تربیت مدرس

نزدیک می‌شوند و یا یک کران به صورت نامحدود افزایش پیدا می‌کند، می‌باشد. در بخش ۲ نشان می‌دهیم که Pـ مقدار تابع پیوسته‌ای از فرضها در ردهٔ همهٔ فرضهای صفر تک نقطه‌ای، یکطرفه و بازه‌ای خواهد بود. این موضوع به ما امکان می‌دهد به تمام انواع فرضها نگرش یکسانی داشته باشیم. پس هر تفسیری که برای Pـ مقدار صورت می‌پذیرد باید با تمام انواع فرضهای ذکر شده سازگار باشد. زیرا به طور واقعی هیچکدام از آنها دارای ماهیت متفاوتی نیستند.

یکی از تفسیرهای متداول P- مقدار، میزان پشتیبانی فرض H و یا ملاکی به نفع H پس از مشاهدهٔ $X=x$ در داده‌هاست (در [۲، ٤ و ۱۰] حالات مختلفی راجع به این تفسیر از Pـ مقدار بحث شده است که با مطالب آورده شده در این مقاله تفاوت دارد). در بخش ۳ شرایط منطقی و ساده‌ای که باید در هر معیار پشتیبانی صدق شود، معرفی می‌کنیم.

اگر فرض H' دربرگیرندهٔ H باشد، آنگاه حداقل به همان اندازه که H پشتیبانی می‌شود، H' نیز باید پشتیبانی شود. سپس نشان می‌دهیم Pـ مقدار دارای چنین شرایطی نیست. در بخش ٤ برخی از روشهای بهبود تفسیر Pـ مقدار به عنوان معیاری در پشتیبانی فرضها ذکر شده است.

۲. پیوستگی تابع Pـ مقدار

فرض می کنیم که همهٔ آزمونهای تحت مطالعه، نااریب و به طور یکنواخت تواناترین آزمون ($UMPU$) باشند. در این صورت برای فرض H و مشاهدهٔ $X=x$، یک احتمال معنی‌داری $P_H(X)$ را داریم. به روشهای مختلفی می‌توان P - مقدار را تعریف کرد. یکی از روشها تعریف $P_H(X)$ به صورت احتمال وقوع مشاهدات X' حداقل به فرینی X[4] در یک تکرار مستقل آزمایشی، وقتی که H درست است، می‌باشد. همچنین می‌توانیم $P_H(X)$ را به عنوان بزرگترین کران پایین مجموعهٔ همهٔ سطوح معنی‌داری α معرفی کنیم که می‌توان H را در سطح α رد کرد. می‌توانیم تصور کنیم به ازای داده‌های معین، نظیر $X=x$، $P_H(X)$ نسبت به انواع مختلف فرضهای H تغییر می‌کند. دربخش ۳ مثالی در این ارتباط و در حالتی که دو فرض مختلف برای یک پارامتر داریم، آورده شده است. برای انواع مختلف فرضها از این به بعد نمادهای زیر را به کار می‌بریم.

فرض کنیم $a\in(-\infty,\infty)$، $b\in[a,\infty)$ و $P_{a,b}(X)$ نشان دهندهٔ P - مقدار باشد، در این صورت حالات $a=b$، ∞، $b=\infty$، $a=-\infty$ و $-\infty<a<b<\infty$ به ترتیب مربوط به مواردی است که فرض مورد نظر H_1، H_2، H_3 یا H_4 است. فرض کنید Φ تابع توزیع نرمال استاندارد و Φ^{-1} تابع چندکی آن باشد، دراین صورت داریم:

$$P_{\mu_0,\mu_0}(x)=2\Phi(-|x-\mu_0|)$$
$$P_{\mu_0,\infty}(x)=\Phi(x-\mu_0)\,,\ P_{-\infty,\mu_0}(x)=\Phi(\mu_0-x)$$

فرض بازه‌ای در آزمون $UMPU$ در سطح α، توسط لهمن [۱۱] به صورت زیر معرفی شده است:

$H_4:\mu\in[\mu_1,\mu_2]$ رد می شود، اگر $|x-0.5(\mu_1+\mu_2)|>c$ باشد که c از رابطهٔ زیر به دست می‌آید:

$$\Phi(0.5[\mu_1-\mu_2]-c)+\Phi(0.5[\mu_2-\mu_1]-c)=\alpha \quad (۱)$$

رابطهٔ (۱) معادل P_{μ_1} (رد H_4) و معادل P_{μ_2} (رد H_4) است که در آن p_μ احتمال شرطی $M=\mu$ خواهد بود. توجه داشته باشید که این رابطه تابعی مشتق پذیر و اکیدا نزولی از c است و وقتی $c=0$ باشد، برابر ۱ است و وقتی $c\to\infty$، به سمت صفر میل می کند. بنابراین رابطهٔ (۱) دارای جواب یکتای $c(\alpha:\mu_2-\mu_1)$ می‌باشد. واضح است که $c(\alpha:\mu_2-\mu_1)$ تابعی پیوسته و اکیدا نزولی از α است وقتی که $\alpha\in(0,1]$ و نیز $c(1:\mu_2-\mu_1)=0$ و $\lim_{\alpha\to} c(\alpha:\mu_2-\mu_1)$. با استفاده از آزمون $UMPU$، Pـ مقدار برای فرض H_4 با داده‌های $X=x$ عبارت است از:

$$P_{\mu_1,\mu_2}(X)=c^{-1}(|x-0.5[\mu_1+\mu_2]|:\mu_2-\mu_1) \quad (۲)$$
$$=\begin{cases}\Phi(x-\mu_1)+\Phi(x-\mu_2) & ;x<0.5[\mu_1+\mu_2]\\ \Phi(\mu_1-x)+\Phi(\mu_2-x) & ;x\ge 0.5[\mu_1+\mu_2]\end{cases}$$

که برای $c^{-1}(y,z)$ عدد d طوری انتخاب می‌شود که $c(d,z)=y$ باشد. معادلهٔ (۲) با جایگزین کردن $c=|x-0.5[\mu_1+\mu_2]|$ در معادلهٔ (۱) و توجه به این نکته که α ی حاصله برابر $P_{\mu_1,\mu_2}(x)$ می‌باشد به صورت زیر قابل محاسبه است:

$$(-\infty,\mu_0]=\bigcup_{a\le\mu_0}[a,\mu_0]$$

---

٤ Exterem

چون وقتی $a \to -\infty$ اجتماع مجموعه‌های $[a,\mu_0]$، برای $a \le \mu_0$ به طور یکنواخت صعودی است. می‌توان نوشت:

$$(-\infty,\mu_0] = \lim_{a\to-\infty}[a,\mu_0]$$

مقادیر احتمال نیز از همین حد پیروی می‌کند یعنی:

$$\lim_{a\to-\infty} P_{a,\mu_0}(x) = P_{-\infty,\mu_0}(x)$$

که به آسانی با جایگزین کردن $\mu_2=\mu_0$ در رابطهٔ (۲) و وقتی $\mu_1 \to -\infty$ حاصل می‌شود. به همین صورت برای همهٔ xها رابطهٔ:

$$\lim_{b\to\infty} P_{\mu_0,b}(x) = P_{\mu_0,\infty}(x)$$

برقرار است.

نقطهٔ حدی دیگر $\lim_{(\mu_1,\mu_2)\to(\mu_0,\mu_0)} P_{\mu_1,\mu_2}(x)$ است که اگر $x>\mu_0$ ، آنگاه سطر بالایی رابطهٔ (۲) مورد استفاده قرار می گیرد و حاصل حد برابر $P_{\mu_0,\mu_0}(x) = 2\Phi(x-\mu_0)$ می‌شود. با دلایل مشابه برای $x\le\mu_0$، مقدار حد برابر $P_{\mu_0,\mu_0}(x)=2\Phi(\mu_0-x)$ خواهد بود. برای $x=\mu_0$ با استفاده از هر دو سطر رابطهٔ (۲)، $P_{\mu_0,\mu_0}(\mu_0)=1$ خواهد بود. چون رابطهٔ (۲) تابعی پیوسته از μ_1 و μ_2 است، برای فواصل بین نقاط بحرانی، به آسانی برابری:

$$\lim_{(\mu_1,\mu_2)\to(a,b)} P_{\mu_1,\mu_2}(x) = P_{a,b}(x)$$

نتیجه می شود. درنهایت اگر هر دو مقدار μ_1 و μ_2 نامتناهی باشند ($\mu_2\to\infty,\ \mu_1\to-\infty$)، دراین صورت هر دو سطر رابطهٔ (۲) برای تمام مقادیر x برابراست با $P_{-\infty,\infty}(x)=1$.

آنچه ما نشان دادیم این است که $P_{\mu_1,\mu_2}(x)$ تابعی پیوسته از (μ_1,μ_2) است، حتی وقتی که یکی از کرانها و یا هر دو کران نامتناهی شود و یا $|\mu_1-\mu_2|\to 0$. همان طور که فرضهای صفر تک نقطه‌ای و یکطرفه، حالتهای حدی فرض بازه‌ای هستند، P- مقدارهای حاصل نیز حد P- مقدار فرض بازه‌ای هستند. این واقعیت باعث می‌شود که فرض صفر تک نقطه‌ای را حالت حدی فرض بازه‌ای در نظر بگیریم. اگر $\mu_1-\mu_2>0$ به اندازهٔ کافی کوچک باشد و $\mu_0=0/5(\mu_1+\mu_2)$، آنگاه مقدار $P_{\mu_0,\mu_0}(x)$ به مقدار $P_{\mu_1,\mu_2}(x)$ بسیار نزدیک خواهد بود. به همین ترتیب فرض یکطرفه را می‌توان تقریبی از یک فرض بازه‌ای بسیار بزرگ در نظر گرفت. پیوستگی P- مقدار تنها به توزیع نرمال محدود نمی‌شود. مثلا با محاسباتی مشابه این پیوستگی برای توزیعهایی مانند نمایی، دوجمله‌ای یا یکنواخت نیز قابل اثبات است.

۳. P- مقدار، معیار مناسبی برای پشتیبانی یا رد فرضها نیست

P- مقدار معمولا برای آزمون فرضها مورد استفاده قرار می گیرد، پس از محاسبهٔ $P_H(x)$، اگر $P_H(x)<\alpha$ باشد، محقق فرض H را در سطح α رد می کند. چون هر قدر $P_H(x)$ بزرگتر باشد، رد کردن H مشکلتر است. $P_H(x)$ غالبا به عنوان معیار پشتیبانی H توسط مشاهدهٔ $X=x$ به کار گرفته می‌شود. توصیهٔ به کارگیری این معیار همیشه غیر رسمی بوده و هیچگاه به صورت تئوری ویژگیهایی که یک معیار پشتیبانی باید داشته باشد، ارائه نشده است. البته این معیار برای مسایل ساده بسیار موفقیت آمیز عمل می کند. مثلا $P_{\mu_0,\mu_0}(x)$ به صورت تابعی از x، هر قدر از μ_0 دورتر باشد، کاهش پیدا می کند که بیان کنندهٔ این خاصیت معیار پشتیبانی است که هر قدر مشاهدات از فرض دورتر باشند، میزان پشتیبانی فرض به وسیلهٔ آنها ضعیفتر است. به طور مشابه $P_{-\infty,\mu_0}(x)$ تابعی صعودی از μ_0 می‌باشد و دارای این خاصیت است که هر قدر فرض مقدار بیشتری از فضای پارامتر را پوشش دهد، پشتیبانی فرض قویتر خواهد بود.

چون از احتمالات پسین به عنوان پشتیبانی فرض، وقتی داده‌ها مشاهده شده‌اند، استفاده می‌شود، بسیاری از آمارشناسان سعی کردند P- مقدار را به عنوان نوعی احتمال پسین تفسیر کنند [۱، ۵، ۶ و۹]. فیشر [۷] نیز با معرفی احتمالات اعتقادی[۵] چنین هدفی را بدون توسل به استدلال بیزی در نظر داشت.

برای هر معیار پشتیبانی خاصیتهای زیر باید برقرار باشد:

- هر قدر داده‌ها از فرض دورتر باشند، میزان پشتیبانی کمتر می‌شود.
- هر قدر داده‌ها به فرض نزدیکتر باشند، میزان پشتیبانی فرض قویتر می‌شود.
- هر قدر فرض گسترده باشد، درجه پشتیبانی آن بیشتر می‌شود.

خاصیت سوم معادل مفهوم انسجام[۶] در مقایسات چندگانه [۸] می‌باشد.

---

۵ *Fiducial Probability*

۶ *coherence*

فرض کنید فرض H جزئی از فرض H' باشد. آزمونهای H و H' دارای انسجام هستند، هرگاه رد H' همواره مستلزم رد کردن H باشد. می‌توان گفت یک ملاک پشتیبانی فرضها، انسجام است. اگرفرض H جزیی از فرض H' باشد، میزان پشتیبانی H' حداقل به بزرگی میزان پشتیبانی H به طریق اولی پشتیبانی H' نیز می‌باشد. برای مثال میزان پشتیبانی برای فرض $H': M \le ۳$، حداقل به اندازهٔ میزان پشتیبانی برای فرض $H: M \le ۲$ است. برای فرضهای یکطرفه، Pـ مقدار در این ملاک صدق می‌کند. اما به طور کلی Pـ مقدار به عنوان ملاکی در پشتیبانی فرضها، دارای انسجام نیست. مثلا فرض کنید $\mu_۰ < x$، در این صورت داریم: $P_{-\infty,\mu_۰}(x) = ۰/۵ P_{\mu_۰,\mu_۰}(x)$، هرچند که فرض $H_۱: M = \mu_۰$ جزیی از فرض $H_۲: M \le \mu_۰$ است.

دلایل زیادی برای این که عدد انسجام[7] باعث مشکل نشده است، را می‌توان تصور کرد. اولا به ندرت فرضهای یکطرفه و صفر تک نقطه‌ای در یک مسئله با هم به کار می‌روند ولی این موضوع کلیت ندارد. مثلا مواردی هست که استفاده از فرض یکطرفه یا فرض صفر تک نقطه‌ای مورد بحث است. در این مسائل انسجام اهمیت پیدا می‌کند. بررسی فرض بازه‌ای، عدم اعتماد به توانایی Pـ مقدار را به عنوان یک معیار پشتیبانی آشکارتر می‌کند. برای وقتی $x > \mu_۲, \mu_۱$، از رابطهٔ (۲) نتیجه می‌شود که $P_{\mu_۱,\mu_۲}(x)$ تابعی اکیدا صعودی برای $\mu_۱$ و $\mu_۲$ است. به طور کاملا اختیاری فرض می‌کنیم $x > \mu_۲ > \mu_۱$ و $\mu_۱ > \mu'_۱$ در این صورت:

$$P_{\mu'_۱,\mu_۲}(x) < P_{\mu_۱,\mu_۲}(x)$$

از آنجا که $P_{\mu'_۱,b}(x)$ تابعی پیوسته در b است، $\mu'_۲$ یی وجود دارد که $\mu'_۲ \in [\mu_۲, x]$، به طوری که:

$$P_{\mu'_۱,\mu'_۲}(x) - P_{\mu'_۱,\mu_۱}(x) < P_{\mu'_۱,\mu_۲}(x) - P_{\mu'_۱,\mu_۲}(x)$$

ازساده کردن معادلهٔ بالا به رابطهٔ

$$P_{\mu'_۱,\mu'_۲}(x) < P_{\mu'_۱,\mu_۱}(x)$$

می‌رسیم. در نتیجه عدم انسجام P – مقدار، در فرض بازه‌ای $[\mu_۱, \mu_۲]$ که زیر مجموعه‌ای از $[\mu'_۱, \mu'_۲]$ است، نشان داده می‌شود. به عنوان مثال عددی، فرض کنید $x = ۲٫۱۸$، $\mu_۱ = -۰/۵$ و $\mu_۲ = ۰/۵$ باشد. از رابطهٔ (۲) نتیجه می‌شود:

$$P_{-۰/۵,۰/۵}(۲/۱۸) = \Phi(-۲/۶۸) + \Phi(-۱/۶۳) = ۰/۰۵۰۲$$

اگر $\mu'_۱ = -۰/۸۲$ و $\mu'_۲ = ۰/۵۲$ باشد، با توجه به رابطهٔ (۲) داریم:

$$P_{-۰/۸۲,۰/۵۲}(۲/۱۸) = \Phi(-۳) + \Phi(-۱/۶۶) = ۰/۰۴۹۸$$

به زبان ساده‌تر اگر Pـ مقدار را به عنوان ملاکی برای پشتیبانی فرضها به کار ببریم، معادل این است که بیان کنیم پشتیبانی بیشتری برای فرض $H: M \in [-۰/۵, ۰/۵]$ نسبت به $H': M \in [-۰/۸۲, ۰/۵۲]$ وجود دارد، در حالی که می‌دانیم H زیر مجموعه‌ای از H' است.

از آنجا که آزمون فرضها رابطهٔ بسیار نزدیکی با P- مقدار دارند این موضوع که عدم انسجام P- مقدار، منجر به عدم انسجام آزمونهای در سطح α برای فرضهای چندگانه به مفهوم گابریل [۸] می‌شود، تعجب انگیز نیست. فرض کنید مایلیم H و H' را درسطح ۵ درصد آزمون کنیم (مثلا ممکن است از روش مقایسه‌های چندگانه بن فرونی با $\alpha = ۰/۰۵$ استفاده کنیم). در این صورت پس از مشاهدهٔ $x = ۲٫۱۸$، H' رد می‌شود. اما قادر به رد H نیستیم. اکنون مطمئن هستیم M خارج از بازهٔ $[-۰/۸۲, ۰/۵۲]$ قرار دارد اما هنوز باید تصور کنیم که M داخل بازهٔ $[-۰/۵, ۰/۵]$ است.

دراین بخش نشان دادیم مقادیر احتمال به عنوان ملاکهایی برای پشتیبانی فرضها نسبت به یکدیگر نامنسجم هستند. همچنین می‌توان نشان داد مقادیر احتمال برای فرضهای یکطرفه و صفرتک نقطه‌ای، به طور همزمان نامنسجم هستند. برای مثال $P_{۰/۵,۰/۵}(۲/۱۸) = ۰/۰۹۳$ که از هر دو مقدار $P_{-۰/۵,۰/۵}(۲/۱۸)$ و $P_{-۰/۸۲,۰/۵۲}(۲/۱۸)$ بزرگتر است. بنابراین این ادعا که مقادیر احتمال ملاکی برای پشتیبانی فرضها با توجه به داده‌های مشاهده شده می‌باشند، بدون در نظر گرفتن محدودیتهای بیشتر، صحیح نیست.

٤. چگونگی تفسیر P ـ مقدار

دو تعریف رسمی P ـ مقدار در بخش ۲ آورده شد. اگر به اندازهٔ کافی به موضوع دقت کنیم، پی خواهیم برد که این تعاریف P ـ مقدار معتبر هستند. آنچه ما نشان دادیم نامعتبر بودن کاربرد احتمال به عنوان ملاکی برای پشتیبانی فرضها است.

موضوعی که پرداختن به آن جالب به نظر می رسد امکان اصلاح یا تفسیر مجدد P ـ مقدار به گونه ای است که به عنوان یک ملاک پشتیبانی برای فرضها قابل قبول باشد. یکی از راهکارهای ساده، تعریف یک ملاک پشتیبانی به صورت زیر است:

---

[7] incoherence

$$P'_{a,b}(x) = \sup_{\theta\in[a,b]} P_{\theta,\theta}(x)$$

این معیار گرچه نامنسجم است ولی بر اساس آن همهٔ فرضهای از نوع $H: M \in [a,b]$ مادامی که $a \le x \le b$، دارای میزان پشتیبانی برابر هستند. چنین تعریفی نمی‌تواند خیلی مفید باشد.

روش دیگر بر مبنای استدلال آزمونهای معنی‌داری، ارائه شده است که در آن سعی می‌شود Pـ مقدار را برای یک فرض معین به صورت تابعی از داده‌ها (احتمالا قبل از مشاهدهٔ داده‌ها) تعریف کنند. در این روش، Pـ مقدار به ازای مقادیر مختلف x به عنوان درجاتی که مجموعه‌های متفاوت داده‌ها، یک فرض خاص را پشتیبانی می‌کنند، درنظر گرفته می‌شود. چنین روشی تا هنگامی که احتمال وجود آزمون فرضهای دیگر در مسئله منتفی باشد، صحیح به نظر می‌رسد اما اشکال کار در آنجاست که نمی‌توان دو نوع مختلف فرض برای یک مجموعه از داده‌ها متصور شد. اشکال جدی دیگر، عدم وجود مقیاسی مطلق برای تفسیر Pـ مقدار است. مقیاس پشتیبانی به نوع فرض بستگی پیدا می‌کند.

برای مثال فرض کنید فردی مایل است معیاری برای پشتیبانی فرض $H: M = ۰/۹$ و در عین حال فرد دیگری مایل است معیاری برای پشتیبانی فرض $H': M \le ۱$ محاسبه کند. پس از این که هر دو دادهٔ یکسان $x = ۲,۷$ را مشاهده کردند، برای فرض اول $P_{۰/۹,۰/۹}(۲/۷) = ۰/۰۷۱۸$ و برای فرض دوم $P_{-\infty,-۱}(۲/۷) = ۰/۰۴۴۶$ به دست می‌آید. مطمئنا داده‌ها، فرض H' را بیشتر پشتیبانی می‌کنند. بنابراین ۰/۷۱۸ باید پشتیبانی کمتری برای فرض نقطه‌ای H باشد تا ۰/۴۴۶ برای فرض یکطرفهٔ H'. به طور خلاصه، تفسیر نقاط خاصی از این مقیاس پشتیبانی (نظیر نقطهٔ متداول ۰/۰۵)، باید از یک آزمون فرض به آزمون فرض دیگر تغییر کند. درمثال عددی بخش ۳، هر یک از دو فرض بازه‌ای، مقیاس خاص خود برای تفسیر Pـ مقدار لازم دارند، گرچه هر دو فرض از نوع بازه‌ای هستند.

یک روش جالب دیگر، بر اساس این استدلال پیشنهاد می‌شود که عدم انسجام ذکر شده در این مقاله، از آنجا ناشی شده است که فرضهای بزرگتر (در مقایسه با فرضهای کوچکتر) شامل پارامترهای دورتر از داده‌ها بودند. برای مثال فرض کنید $H: M \in [-۰/۵,۰/۵]$ و $H': M \in [-۰/۸۲,۰/۵۲]$، با دادهٔ $x = ۲,۱۸$ باشد. در این صورت مقادیر درون بازهٔ $[-۰/۸۲,-۰/۵۲]$ عضو H' هستند ولی عضو H نیستند. به طور یقین از دادهٔ مشاهده شده، دورتر می‌باشند.

ممکن است H' را به علت در برگرفتن پارامترهای اضافی که از داده‌ها دور هستند، جریمه کنیم. این ایده در ابتدا جذاب به نظر می‌رسد. اما توجه به این موضوع که P ـ مقدار برای فرضهای یک طرفه، وقتی مقادیر دورتر از داده‌ها را بردارند، جریمه نمی‌شوند. این مطلب با ایدهٔ قبل متناقض است.

فرض کنید دورتر بودن را به صورت زیر تعریف کنیم:

تعریف ۱ـ فرض کنید H فرض مورد نظر باشد و θ' عضو H نباشد. همچنین فرض کنید برای تمام مقادیر θ متعلق به H، رابطهٔ $P_{\theta',\theta'}(x) < P_{\theta,\theta}(x)$ برقرار باشد. در این صورت θ' را دورتر از H نسبت به x می‌خوانیم.

در توزیع نرمال وقتی H' شامل H باشد و نقاطی که جزو اشتراک H' و H نیستند، نسبت به H دورتر از x باشند، رابطهٔ $P_{H'}(x) < P_H(x)$ برقرار است. اما حتی چنین موضوعی نیز در تمام موارد صادق نیست. مثلا وقتی x دارای توزیع یکنواخت $U(۰,\theta)$ است به سادگی می‌توان نشان داد که برای $H: M \in [a,b]$ داریم:

$$P_{a,b}(x) = \begin{cases} \dfrac{x}{a} & ;x \le a \\ \dfrac{b-x}{b-a} & ;a < x \le b \\ ۰ & ;b < x \end{cases}$$

فرض کنید $H: M \in [a,b]$ و $H': M \in [a,b']$ با شرط $b \le b'$ برقرار باشد، $X = x < a$ مشاهده شود، آنگاه $P_{a,b}(x) = P_{a,b'}(x)$ خواهد بود. حتی این رابطه برای وقتی که تمام $\theta' > b$ دورتر از x نسبت به H باشد نیز صدق می‌کند. توجه کنید که اگر $\theta \in [a,b]$ و $\theta' > b$ باشد، داریم:

$$P_{\theta',\theta'}(x) = \frac{x}{\theta'} < \frac{x}{\theta} = P_{\theta,\theta}(x)$$

حتی اگر قادر به ساختن یک معیار پشتیبانی جریمه شده برای آزمون فرضها شویم، این موضوع به معنی عدم شواهد علیه فرض نخواهد بود. بلکه آنچه قابل استنباط است وجود حداقل یک قسمت از فضای پارامتر است که توسط داده‌ها پشتیبانی نمی‌شود. به طور خلاصه قادر به بیان تعریفی سازگار از Pـ مقدار به عنوان معیاری در پشتیبانی فرضها نیستیم.

۵. بحث

نشان دادیم فرضهای یکطرفه و صفر تک نقطه‌ای موضوعاتی متفاوت از یکدیگر نیستند و تنها حالات حدی فرض دیگر با نام فرض بازه‌ای می‌باشند. برای توزیعهای معمولی، Pـ مقدار تابع پیوسته‌ای از فرضها است. همچنین نشان دادیم Pـ مقدار به عنوان معیاری در پشتیبانی آزمون فرض مربوط به آنها قابل به کارگیری نیستند. می‌توان استدلال کرد که مقدار فرضهایی را که شامل مقادیر دور از داده‌ها هستند، جریمه می‌کند. ولی حتی استدلال نیز در حالت ساده، توزیع یکنواخت برقرار نیست.

اعلام Pـ مقدار امری متداول در آمار کاربردی است، بخصوص در موارد چند پارامتری که مورد بحث این مقاله قرار نگرفته است. با این حال وقتی Pـ مقدار در مسایل ساده‌ای نظیر آنچه در اینجا بحث شد، به عنوان معیاری در پشتیبانی داده‌ها قابل به کارگیری نباشد، کاربرد آن درمسایل پیچیده‌تر بیشتر مورد توجه قرار می‌گیرد.

مراجع

[1] Berger, J.O. and Sellke, T., 1987, *Testing a Point Null Hypothesis: The Irreconcilability of P-Value and Evidence,* (with comments), Journal of the American Statistical Association, 82, 112-122.

[2] Berkson, J., 1942, *Tests of Significance Considered as Evidence,* Journal of the American Statistical Association, 37, 325-335.

[3] Berry, D.A., 1990, *Basic Principles in Designing and Analyzing Clinical Studies,* Statistical Methodology in the Pharmaceutical Sciences. Ed. Berry, D.A., New York: Marcel Dekker, pp. 1-55.

[4] Blyth, C.R. and Staudte, R.G., 1995, *Estimating Statistical Hypothesis, Statistics and Probability Letters,* 23, 45-52.

[5] Casella, G. and Berger, R.L., 1987, *Reconciling Bayesian and Frequentist Evidence in the One-sided Testing Problem,* (with comments), Journal of the American Statistical Association, 82, 106-111.

[6] DeGroot, M.H., 1973, *Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio,* Journal of the American Statistical Association, 68, 966-969.

[7] Fisher, R.A., 1935, *The Fiducial Argument in Statistical Interence,* Annals of Eugenice, 6, 391-398.

[8] Gabriel, K.R., 1969, *Simultaneous Test Procedure–Some Theory of Multiple Comparisons,* The Annals of Mathematical Statistics, 40, 224-250.

[9] Hodges, J., 1992, *Who Knows What Alternative Lurks in the Hearts of Significance Tests? (with comments),* in Bayesian Statistics 4, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford: Clarendon Press, pp. 247-266.

[10] Lehmann, E.L., 1975, *Nonparametrics: Statistical Methods Based on Ranks,* San Francisco: Holden–Day.

[11] Lehmann, E. L., 1986, *Testing Statistical Hypothesis,* 2nd ed., New York: John Wiley.

[12] Russel, D., 1983, Equal Employment Opportunity Commission V. Federal Reserve Bank of Richmond, 698 Federal Reporter 2d series, 633-675, United States Court of Appeal, Fourth Circuit.