



از سطح معنی داری تا مقدار احتمال^۱

مریم قدسی نرگس عباسی^۲

چکیده

در این مقاله آزمونهای معنی داری را با مثالهای مربوط در دو بخش ارایه می دهیم [۲]. سپس ساختار ابتدایی و مفهومی های کلیدی آزمونهای فرض را تا جایی که بتوان کمیت مقدار احتمال را تعریف کرد، عنوان کرده [۷] و در پایان با ارایه بخش کوچکی به تفاوت سطح معنی داری و مقدار احتمال و اینکه مقدار احتمال کمیتی است که می تواند در بخشهای گوناگون، شکلهای متفاوتی را به خود بگیرد، می پردازیم.

۱ مقدمه

آزمونها با مثالی که اربوتنت^۳ در سال ۱۷۱۰ داد شروع شده ولی گسترش واقعی آن توسط کارل پیرسون^۴ در سال ۱۹۰۰ بود که آزمون خوبی برازندگی^۵ نامیده شد، بعداً و. س. گاست^۶ در سال ۱۹۰۸ تحت عنوان استیودنت^۷ کارش را منتشر ساخت، و در حدود بیست سال بعد آر. ا. فیشر^۸ آزمونهای معنی داری را با انواع مثالها، ارایه داد. البته حقیقتاً

فیشر، مبدع بزرگ آزمونهای معنی داری، ماهیت آنها را به طرز مبهمی توضیح نداده بود. نیمن و پیرسون^۱ در سال ۱۹۲۸ نظریه ریاضی را برای این روش ابداع کردند و بعداً یک نظریه کلی تر از آزمونها را با نام آزمون فرض به وجود آوردند که این نظریه تحت تأثیر تفکری است که بر مبنای آن آزمونهای آماری، عناصر نظریه تصمیم می باشند.

From Level of Significance Up P-Value^۱
 مریم قدسی و نرگس عباسی، گروه آمار - دانشکده علوم، دانشگاه شیراز^۲
 Arbuthnot^۳
 K. Pearson^۴
 Goodness of fit test^۵
 W. S. Gosset^۶
 Student^۷
 R. A. Fisher^۸
 Neyman and Pearson^۹

و ارایه آن بایستی مراحل زیر را طی کرد:

(۱) فرضیه آماری H_0 : H_0 حکمی است درباره اینکه توزیع Y ، $f_Y(y)$ می باشد، بنابراین H_0 توضیحی در ارتباط با چگالی $f_Y(y)$ می باشد که به دو صورت زیر ممکن است وجود داشته باشد:

— H_0 چگالی $f_Y(y)$ را کاملاً مشخص^{۱۶} کند که در آن صورت، H_0 را ساده^{۱۷} می نامیم.

— H_0 چگالی $f_Y(y)$ را مشخص^{۱۸} کند اما به مقادیری از پارامترهای نامعلوم ولی معین بستگی داشته باشد که H_0 را در این حالت مرکب^{۱۹} می نامیم.

(۲) آماره آزمون: $t = t(y)$ را تابعی از مشاهدات و $T = t(y)$ را متغیر تصادفی مربوط به آن در نظر می گیریم، T را آماره آزمون^{۲۰} می نامند، اگر دارای دو شرط زیر باشد:

— توزیعش تحت فرض H_0 کاملاً مشخص باشد و در حالتی که فرضیه H_0 مرکب است، اگر آنرا به فرضیه های ساده تجزیه کنیم، همگی جزء یک خانواده توزیع باشند.

— بزرگترین مقدار t ، قویترین دلیل انحراف از H_0 از نوعی که برای آزمون لازم است را بدهد.

(۳) تعیین سطح معنی داری و اعلام نتیجه: برای مشاهده موجود، $t \equiv t_{obs} \equiv t(y)$ را محاسبه و سطح معنی داری^{۲۱} را به صورت زیر تعریف می کنیم:

$$p_{obs} = Pr(T \geq t_{obs} ; H_0)$$

در این مقاله سعی ما بر این است که ساختار این دو نوع آزمون را جداگانه معرفی نماییم و تا آن جایی ادامه دهیم که سطح معنی داری را در آزمون معنی داری و مقدار احتمال را در آزمون فرض معرفی کرده، سپس بحث را به اتمام می رسانیم. و از شما خواننده عزیز درخواست می کنیم که ابتدا مقاله را تا به انتها خوانده، سپس قضاوت کنید.

۲ آزمونهای معنی داری

در این فصل از اصطلاحات خاصی مانند آزمون معنی داری خلص^{۱۰}، اکسترم^{۱۱}، آزمون سازگاری^{۱۲}، نوع انحراف از H_0 ^{۱۳}، دلیل انحراف^{۱۴} استفاده شده است. شایان ذکر است که این اصطلاحات را کاکس و هینکلی^{۱۵} در کتاب خود به کار برده اند و نویسندگان دیگر ممکن است از اصطلاحات دیگری استفاده نمایند.

۱.۲ آزمون معنی داری خلص

در این قسمت ابزار کار ما دو چیز است: فرضیه ای آماری به نام H_0 ، که تنها توزیع داده ای است که به وضوح فرمول بندی شده، و جهت انحراف از H_0 به این صورت است که همه مجموعه داده ها در یک جهت به طور یکتایی اکسترم هستند و اگر این جهت انحرافات پرمعنی از H_0 را ندانیم آزمون سازگاری داده ها با فرضیه آماری بی معنی است.

یک آزمون سازگاری داده ها با فرضیه H_0 را تحت ابزار خیلی محدود بالا، آزمون معنی داری خلص می نامند که برای تعریف

Pure Significance test	۱۰
extreme	۱۱
test of consistency	۱۲
type of departure from H_0	۱۳
Evidence of Departure	۱۴
Cox and Hinkley	۱۵
Specify	۱۶
Simple	۱۷
Specify	۱۸
Composiee	۱۹
Test Statistic	۲۰
Level of Significance	۲۱

مثال ۱ (Numerical law)

اغلب یک دنباله از اعداد از یک قانون ساده پیروی می کنند؛ حتی اگر هیچ طرح نظری برای آن پیدا نشود. یک مثال کلاسیک قانون باد^{۲۲} است، که می گوید فاصله خورشید تا هر سیاره را به وسیله فرم هندسی ساده زیر می توان نشان داد:

$$b_n = a + b^{2^n}; n = 1, 2, \dots$$

به طوری که n ترتیب هر سیاره از خورشید است (مثلاً برای عطارد $n = 1$ ، ونوس $n = 2$ ، ...)، a و b ثابت هایی هستند که به وسیله قرار دادن اعداد واقعی در این رابطه به دست می آیند، مشروط بر اینکه فاصله خورشید تا زمین، برابر یک فرض شود در این صورت ثابت ها برابرند با

$$a = 0/4, \quad b = 0/075$$

جدول زیر فواصل واقعی (d_n) را با اعداد باد (b_n) مقایسه می کند:

سیاره	n	فاصله باد b_n	فاصله واقعی d_n
Mercury	۱	۰/۵۵	۰/۳۸۷
Venus	۲	۰/۷	۰/۷۲۳
Earth	۳	۱	۱
Mars	۴	۱/۶	۱/۵۲۴
Asteroids	۵	۲/۸	۲/۹
Jupiter	۶	۵/۲	۵/۲۰۳
Saturn	۷	۱۰	۹/۵۴۶
Uranus	۸	۱۹/۶	۱۹/۲۰
Neptune	۹	۳۸/۸	۳۰/۰۹
Pluto	۱۰	۷۷/۲	۳۹/۵

می بینیم که اعداد واقعی با اعداد باد، متفاوتند؛ پس سؤالی که پیش می آید این است که: آیا قانون باد واقعی است یا اینکه ساخته ذهن و به طور مصنوعی می باشد؟ بنابراین با یک بحث آماری روبرو می شویم: رابطه کلی $y_i = a + bz_i$ (*) برای $i = 1, \dots, n$ به طوری که $z_{i+1} - z_i = C$ را در نظر بگیرید. اگر داده ها از فرم فوق تبعیت نکنند، به صورت تصادفی می باشند. بنابراین فرض H_0 به صورت زیر در می آید:

مدلی به داده ها برازش نشود: H_0

متغیر تصادفی P به طوری که p_{obs} یک مقدار مشاهده شده آن باشد، در حالت پیوسته دارای توزیع یکنواختی، تحت فرض H_0 است. می توان p_{obs} را به صورت فاصله داده ها از توزیع فرض شده برای مجموعه های داده ممکن که بر روی مقیاس $[0, 1]$ برده شده اند، در نظر گرفت. مقادیر کوچک، فاصله زیاد و مقادیر بزرگ (یعنی مقادیر نزدیک به یک)، فاصله را کم نشان می دهند [۵]. یا اینکه p_{obs} را به عنوان یک اندازه سازگاری داده ها، با H_0 با تفسیر فرضی زیر در نظر گرفت: فرض کنید قرار باشد که داده های موجود را به عنوان دلیل علیه H_0 قبول کنیم در آن صورت ناچار خواهیم بود که حتی تمام داده های با مقدار بزرگتر را به عنوان دلیل قوی تر بپذیریم. بنابراین اگر قرار باشد که داده های تحت تجزیه و تحلیل را به عنوان تصمیم گیرنده منصف، علیه H_0 در نظر بگیریم p_{obs} احتمال آن است که ما اشتباهاً اعلام کنیم دلیلی علیه H_0 وجود دارد.

توجه کنید که ما در اینجا p_{obs} را به عنوان شاخص فاصله و یا شاخص سازگاری داده ها با H_0 در نظر گرفته ایم و صریحاً درگیر تصمیم های ناظر بر قبول یا رد H_0 نبوده، خصوصاً به اینکه آیا p_{obs} از یک مقدار از پیش تعیین شده ای مثل پنج درصد یا یک درصد بیشتر یا کمتر است، علاقه مند نیستیم. یعنی به دنبال این نمی باشیم که خطی مرزی بین داده ها ایجاد نماییم؛ به طوری که $p_{obs} > 0/05$ برقرار باشد. با این حال، سطح های استاندارد مانند یک درصد و پنج درصد و ... برای گزارش اینکه داده ها در سطح پنج درصد معنی دار نبوده، اما در سطح یک درصد معنی دار می باشند خیلی مناسب هستند یعنی شرط آشیان گزینی - که آن را بررسی خواهیم کرد - برقرار باشد. در واقع ما فقط وقتی می توانیم تصمیم رد و یا قبول H_0 را اتخاذ نماییم که آزمون ما براساس نظریه تصمیم پایه ریزی شده باشد.

پس $p_{obs} = Pr(T \geq t_{obs} ; H_0)$ که اگر p_{obs} کوچک باشد سازگاری داده‌ها با H_0 کم و یا فاصله داده‌ها از H_0 زیاد است و داده‌های ما مدل (*) را حمایت می‌کنند.

۲.۲ آزمونهای معنی‌داری بهبودیافته

آزمون معنی‌داری عبارت است از محاسبه یک آماره و سطح معنی‌داری بودن حاصله از روی داده‌ها. (کاکس و هینکلی [۲]) مشکل بزرگی که در حل مسائل فوق وجود داشت، انتخاب آماره آزمون و سپس پیدا کردن توزیع آن بود. در این قسمت روشی را ارائه می‌دهیم که در رفع مشکل راه‌گشاست.

فرض می‌کنیم علاوه بر فرضیه H_0 ، فرضیه مخالف دیگری به نام H_A که جهت انحراف از H_0 را نشان می‌دهد نیز داریم. توجه کنید که در این مرحله مستقیماً در مورد درستی یا نادرستی H_0 یا H_A ، یا برآورد اندازه انحرافات ممکن از H_0 ، پاسخی نمی‌دهیم. تنها سؤالی که پرسیده می‌شود این است که :

Is there evidence of inconsistency with H_0 ?

آیا دلیل ناسازگاری با H_0 وجود دارد؟

فرمول‌بندی: در (۲.۱)، برای هر نقطه نمونه‌ای، یک سطح معنی‌داری را تعریف کردیم، بدین صورت که احتمال مقدار مشاهده شده یا مقدار اکسترم‌تر آماره آزمون را به دست آوردیم. حال در این جا می‌خواهیم دیدگاه دیگری را برای آزمون تعریف کنیم که گرچه با تعاریف قبل مختلف است اما ذاتاً با یکدیگر هم‌ارز می‌باشند.

فرض کنید برای هر $0 < \alpha < 1$ ، مجموعه‌ای از نقاط در فضای نمونه‌ای باشند به طوری که از H_0 در سطح α اختلاف معنی‌داری دارند. w_α را ناحیه بحرانی اندازه α می‌نامیم و

چون ما به دنبال رابطه بین اعداد هستیم پس آماره‌ای را که تعریف می‌کنیم نباید تحت تغییر مبدأ و واحد اندازه‌گیری، تغییر کند و به عبارتی باید پایا^{۲۲} باشد. پس داریم:

$$d_i = \frac{y_{i+1} - y_i}{y_n - y_1} ; i = 1, \dots, n-1$$

d_i آماره مورد نظر است و فرض H_0 را به این صورت تعریف می‌کنیم که d_i ها به صورت تصادفی می‌باشند. قرار می‌دهیم:

$$\frac{y_i - y_1}{y_n - y_1} = u_{i-1} ; i = 2, \dots, n-1$$

و

$$u_{n-1} = 1$$

u_i ها را مشاهده‌هایی از آماره‌ها ترتیبی توزیع یکنواخت استاندارد در نظر می‌گیریم چون اولاً $u_{i+1} > u_i$ و ثانیاً $0 < u_i < 1$ برای $i = 1, \dots, n-2$. پس خواهیم داشت :

$$d_1 = u_1$$

$$d_2 = u_2 - u_1$$

:

$$d_{n-2} = u_{n-2} - u_{n-3}$$

$$d_{n-1} = 1 - u_{n-2}$$

و بالاخره

$$H_0 : f_D(d_1, \dots, d_{n-1}) = (n-2)!$$

$$(d_i \geq 0, \sum_{i=1}^{n-1} d_i = 1)$$

که D_i ها متغیر تصادفی مربوط به d_i های مشاهده شده می‌باشند برای $i = 1, \dots, n-1$ ، اما تحت مدل (*) داریم:

$$d_i = \frac{1}{n-1} ; i = 1, \dots, n-1$$

در پایان آماره مورد نظر را به صورت زیر تعریف می‌کنیم :

$$T = \sum_{i=1}^{n-1} (D_i - \frac{1}{n-1})^2$$

Invariant^{۲۲}

آزمون را به وسیله این نواحی بحرانی تعریف می کنیم در صورتی که در دو شرط زیر صدق کنند:

$$w_{\alpha_1} \subset w_{\alpha_2} \quad \text{if} \quad \alpha_1 < \alpha_2 \quad (1)$$

این شرط را شرط آشیان گزینی (Nesting Condition) می نامند و یک شرط اساسی است چون ممکن است کسی ادعا کند داده ها در سطح یک درصد انحراف معنی دار از H_0 دارند اما در سطح پنج درصد اینگونه نمی باشند.

و شرط دوم برای تضمین این که سطح معنی داری، یک معنی داری فیزیکی فرضی همانند آنچه که در (۲.۱) گفته شد دارد، به ازای هر α قرار می دهیم :

$$Pr(Y \in w_{\alpha} ; H_0) = \alpha, \quad 0 < \alpha < 1 \quad (2)$$

برای کامل شدن بحث در موازی با (۲.۱)، سطح معنی داری p_{obs} از داده y را به صورت زیر تعریف می کنیم :

$$p(y) \equiv p_{obs} = \inf\{\alpha ; y \in w_{\alpha}\} \quad (3)$$

هر آزمونی به روش بالا می تواند تعریف شود اما قصد ما ارایه بهترین آزمون در انواع مختلف w_{α} می باشد.

برای این کار فرض می کنیم α ، دلخواه اما ثابت بوده، شرط آشیان گزینی (۱) حتماً برقرار است. برای راحتی فرض می کنیم که H_0 و H_A هر دو ساده می باشند و

$$H_0 : f_0(y), \quad H_A : f_A(y)$$

و همچنین w_{α} و w'_{α} دو ناحیه بحرانی اندازه α می باشند. با استفاده از (۲) داریم:

$$Pr(Y \in w_{\alpha} ; H_0) = Pr(Y \in w'_{\alpha} ; H_0) \quad (4)$$

برای فرضیه مخالف H_A ، w_{α} را به w'_{α} ترجیح می دهیم هرگاه

$$Pr(Y \in w_{\alpha} ; H_A) = Pr(Y \in w'_{\alpha} ; H_A) \quad (5)$$

ناحیه بحرانی w_{α} را بهترین ناحیه بحرانی اندازه α ی آزمون می نامیم؛ اگر (۵) برای هر w'_{α} ایی که در شرط (۴) صدق کند برقرار باشد.

نسبت درستنمایی $\frac{f_A(y)}{f_0(y)} = Lr_{A_0}(y)$ یک آماره بسنده می باشد، چون :

$$\begin{aligned} & Pr\{Y = y | Lr_{A_0}(Y) = l ; H_0\} \\ &= \frac{Pr\{Y = y \cap \frac{f_A(y)}{f_0(y)} = l ; H_0\}}{Pr\{\frac{f_A(y)}{f_0(y)} = l ; H_0\}} \\ &= \frac{f_0(y)}{\sum_l f_0(y)} = \frac{l f_0(y)}{\sum_l l f_0(y)} \\ &= \frac{f_A(y)}{\sum_l f_A(y)} = P\{Y = y | Lr_{A_0}(Y) = l ; H_A\} \end{aligned}$$

پس قابل قبول است که بگوییم فقط از طریق نسبت درستنمایی، بهترین ناحیه بحرانی، به داده ها بستگی خواهد داشت و افزون بر این، با یک بیان ناقص و اجمالی، بزرگترین مقدار $Lr_{A_0}(y)$ بدترین برازش (Fit) برای H_0 است. برای سادگی کار فرض می کنیم که $Lr_{A_0}(y)$ تحت H_0 ، یک متغیر تصادفی پیوسته است؛ به طوری که برای هر α ، یک C_{α} ی یکتایی وجود دارد که

$$Pr\{Lr_{A_0}(Y) \geq C_{\alpha} ; H_0\} = \alpha$$

یعنی، اگر نواحی بحرانی را به صورت همه مقادیر بزرگ نسبت درستنمایی در نظر بگیریم، آنگاه C_{α} ی یکتایی وجود دارد که ناحیه بحرانی نسبت درستنمایی، به صورت $Pr\{Lr_{A_0}(Y) \geq C_{\alpha}\}$ تعریف می گردد. این ناحیه بحرانی در شرط آشیان گزینی (۱) صدق می کند.

برای هر اندازه α ، ناحیه بحرانی نسبت درستنمایی، بهترین ناحیه بحرانی است که این نتیجه اصلی، لم نیمن - پیرسن نامیده می شود.

اثبات: فرض کنید w_{α} ، ناحیه بحرانی نسبت درستنمایی و w'_{α} هر ناحیه بحرانی دیگری باشد، که اندازه هر دو آنها α است. پس :

$$\int_{w_{\alpha}} f_0(y) dy = \int_{w'_{\alpha}} f_0(y) dy = \alpha$$

و در نتیجه

$$\int_{w_\alpha - w'_\alpha} f_0(y) dy = \int_{w_\alpha - w'_\alpha} f_0(y) dy \quad (6)$$

اما در مجموعه $w_\alpha - w'_\alpha$ ، اعضای w'_α وجود ندارند پس در این ناحیه $f_A(y) \geq C_\alpha f_0(y)$ ؛ و در مجموعه $w'_\alpha - w_\alpha$ ، اعضای w_α وجود ندارند پس در این ناحیه $f_A(y) < C_\alpha f_0(y)$. و از (6) نتیجه می شود که

$$\int_{w_\alpha - w'_\alpha} f_A(y) dy \geq \int_{w'_\alpha - w_\alpha} f_A(y) dy$$

نابرابری اکید است، مگر اینکه نواحی w_α و w'_α هم ارز باشند. اگر به هر دو طرف انتگرالها، انتگرال روی $w_\alpha \cap w'_\alpha$ اضافه کنیم، خواهیم داشت:

$$\int_{w_\alpha} f_A(y) dy \geq \int_{w'_\alpha} f_A(y) dy$$

و این، اثبات را تمام خواهد کرد.

توجه کنید که اگر w'_α دارای اندازه کمتر از α باشد، باز هم آخرین نامساوی برقرار است؛ یعنی امکان ندارد که بتوانیم توان ناحیه بحرانی نسبت درستنمایی اندازه α را به وسیله ناحیه دیگری که دارای اندازه های کمتر از α است، بهبود بخشیم.

مثال ۲ - میانگین با واریانس معلوم

فرض کنید $Y_1, \dots, Y_n$ نمونه تصادفی i.i.d از $N(\mu, \sigma^2)$ باشند، به طوری که σ^2 معلوم است و اگر فرضیه های فرضیه های صفر و مخالف را با توجه به اینکه $\mu_A > \mu_0$ است، به صورت زیر مشخص کنیم:

$$H_0: \mu = \mu_0 \quad H_A: \mu = \mu_A$$

آنگاه یک محاسبه ساده نشان می دهد که

$$\begin{aligned} Lr_{A,0}(y) &= \frac{\frac{1}{(\sqrt{2\pi})^{\frac{n}{2}} \sigma^{\frac{n}{2}}} \exp\{-\frac{\sum (y_i - \mu_A)^2}{2\sigma^2}\}}{\frac{1}{(\sqrt{2\pi})^{\frac{n}{2}} \sigma^{\frac{n}{2}}} \exp\{-\frac{\sum (y_i - \mu_0)^2}{2\sigma^2}\}} \\ &= \exp\{-\frac{n\bar{y}(\mu_A - \mu_0) - \frac{n}{2}\mu_A^2 + \frac{n}{2}\mu_0^2}{\sigma^2}\} \end{aligned}$$

اما در بسط بالا همه کمیتها به جز $\bar{y}$ ثابت هستند و چون $\mu_A - \mu_0 > 0$ است، ناحیه بحرانی هم ارز با C_α $Lr_{A,0}(y) \geq C_\alpha$

$\bar{y} \geq d_\alpha$ می باشد، این عمل، قدم اول را کامل می کند. در قدم دوم، ما d_α را به گونه ای انتخاب می کنیم که شرط آشیان گزینی برقرار باشد. $\bar{Y}$ ، تحت H_0 ، دارای توزیع $N(\mu_0, \frac{\sigma_0^2}{n})$ است، بنابراین

$$d_\alpha = \mu_0 + k_\alpha^* \frac{\sigma_0}{\sqrt{n}}$$

به طوری که $\Phi(-k_\alpha^*) = \alpha$ و بهترین ناحیه بحرانی به فرم زیر می شود:

$$w_\alpha = \{\bar{y}; \frac{\bar{y} - \mu_0}{\sigma_0/\sqrt{n}} \geq k_\alpha^*\}$$

سطح معنی داری تعریف شده در (۳) دقیقاً همانی است که به وسیله به کار بردن عملگر قسمت (۲.۱) برای آماره آزمون $\bar{Y}$ به دست آورده شده است. اما نکته قابل توجه این است که ناحیه بحرانی ما به μ_A بستگی ندارد، اما اگر شرط $\mu_A < \mu_0$ را قرار می دادیم، ناحیه بحرانی به صورت

$$w_\alpha = \{\bar{y}; \frac{\bar{y} - \mu_0}{\sigma_0/\sqrt{n}} \leq k_\alpha^* \quad 0 < \alpha < 1\}$$

می شد و این بدان معناست که جهت انحراف از H_0 را باید به خوبی بدانیم و جهت اکسترم داده ها را تشخیص دهیم.

۳ آزمون فرض

اولین نکته مطرح شده در آزمون فرض این است که کسی می خواهد تصمیم بگیرد که آیا یک فرضیه فرمول بندی شده صحیح است یا نه؟ انتخاب بین دو تصمیم است، رد یا قبول فرضیه. یک عملگر تصمیم برای چنین مسأله ای آزمون فرض نامیده می شود. تصمیم براساس مقدار یک متغیر تصادفی Y پایه ریزی می شود. چگالیهای در H_A را، مخالفهای H_0 می نامیم. به طوری که H_A کلاسی از مخالفهاست. فرض کنید تصمیم های مربوط به رد یا قبول H_0 را به ترتیب با d_1 و d_0 تعریف می کنیم. یک آزمون غیر تصادفی هر مقدار y از Y را به یکی از تصمیم های d_1 و d_0 نسبت می دهد و بنابراین فضای

معمولاً نواحی متفاوت با سطح‌های معنی داری مختلف، در دسترس است، پس مناسب است که علاوه بر اعلام نتیجه رد یا قبول فرضیه در سطح معنی داری داده شده، کوچکترین سطح معنی داری $\hat{\alpha} = \hat{\alpha}(y)$ که احتمال معنی داری یا مقدار احتمال (p-value)، نامیده می‌شود نیز ارائه گردد.

۴ چند سخن

در مقدمه مقاله گفتیم که آزمون معنی داری را تا سطح معنی داری پیش می‌بریم و دیدیم که با به دست آوردن سطح معنی داری بلافاصله توانستیم نتیجه را اعلام نماییم و مثالی را ارائه دهیم.

اما در آزمون فرض مقدار احتمال فقط یک کمیت اسکار است که اضافه بر روند اصلی آزمون، تعریف می‌شود و در بحثها و جایگاههای مختلف اهمیت و شکل آن متفاوت می‌باشد.

مثلاً در نظریه بیزی احتمال‌های کلاسیک خطای نوع اول و دوم معمولاً هیچ ارتباط نزدیکی با احتمال‌های پسین فرضیه ندارند اما مقدار احتمال با احتمال پسین در بعضی حالات ارتباط نزدیکی دارد که این موضوع ممکن است کار کلاسیکها را برای استفاده از مقدار احتمال توجیه کند [۱]. و یا با دانستن خاصیت MLR، آزمون UMP و در کل با پیشروی در مبحث آمار، مقدار احتمال را می‌توان به صورتهای جدیدتری تعریف کرد. به مثال زیر که مسأله ۹ فصل ۳ از کتاب آزمون فرضیه‌های آماری است [۷] توجه فرمایید:

فرض کنید تابع چگالی احتمال X ، با پارامتر θ ، دارای خاصیت MLR در $T(y)$ باشد و مسأله آزمون $H_0: \theta \leq \theta_0$ در برابر $H_A: \theta > \theta_0$ را در نظر بگیرید. اگر توزیع T پیوسته باشد، مقدار احتمال، آزمون UMP به صورت $p\text{-value} = P_{\theta_0}(T \geq t)$ که مقدار مشاهده T است. اگر برای آزمونهای تصادفی، $\hat{\alpha}$ را به عنوان

نمونه هم به دو ناحیه مکمل S_0 و S_1 تقسیم می‌شوند. اگر Y در S_0 قرار گیرد فرضیه پذیرفته می‌شود و در غیر این صورت رد می‌شود. مجموعه S_0 ناحیه پذیرش و مجموعه S_1 را ناحیه رد یا ناحیه بحرانی می‌نامیم.

در هنگام انجام آزمون، ممکن است دو خطا انجام دهیم، رد فرضیه وقتی که آن درست است (خطای نوع اول) و یا پذیرش آن وقتی غلط است (خطای نوع دوم). نتیجه‌های این دو خطا کاملاً متفاوت است. برای مثال اگر شخصی وجود یک بیماری را آزمایش کند، تصمیم غلط موجب ناراحتی بیمار و ضرر مالی می‌شود در حالی که نقص در تشخیص وجود درد، ممکن است موجب مرگ بیمار شود.

این مثال به نوعی بیان می‌کند که این دو خطا را باید بسیار کاهش دهیم. متأسفانه وقتی تعداد مشاهدات داده می‌شود، دو احتمال را همزمان نمی‌توانیم کنترل کنیم. عموماً یک باند برای رد به غلط H_0 وقتی که آن درست است می‌گذاریم و تلاش برای کاهش احتمال دیگر را مشروط به این عمل می‌کنیم. بنابراین ابتدا $0 < \alpha < 1$ را انتخاب نموده، و آن را سطح معنی داری [۷] می‌نامیم که عبارتست از:

$$P(\delta(Y) = d_1; H_0) = P(Y \in S_1; H_0) = \alpha \quad (7)$$

آنگاه به شرط معلوم بودن α عبارت :

$$Pr(\delta(Y) = d_1; H_A) = P(Y \in S_1; H_A) \quad (8)$$

را حداکثر نموده، داریم:

$$\sup_{H_0} Pr(Y \in S_1) = \alpha \quad (9)$$

طرف چپ (۹)، اندازه آزمون نامیده می‌شود. شرط (۷) توجه ما را به آزمونهایی جلب می‌کند که اندازه آنها از سطح معنی داری داده‌شده تجاوز نمی‌کنند. مقدار به دست آمده از رابطه (۸) را برای همه فرضیه‌های مخالف H_A ، توان آزمون می‌نامند.

انتخاب یک سطح معنی داری α ، معمولاً اختیاری است و این نشان‌دهنده مطالعه بیشتر در این زمینه است. در عمل

کوچکترین سطح معنی داری که در آن فرضیه با احتمال یک رد می شود، تعریف کنیم این تساوی بدون فرض پیوستگی نیز برقرار است. پس می توان گفت که در این مقاله در مورد مقدار احتمال فقط مقدمات را گفتیم و برای اینکه مقدار احتمال را بیشتر بشناسیم احتیاج به بحثهای بیشتری است که به خواننده محترم واگذار می نمایم.

مراجع

- [1] J. Breger, (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.
- [2] D. R. Cox and D. V. Hinkley, (1974), *Theoretical Statistics*, Chapman & Hall, London..
- [3] B. Efron, (1971), *Dose an observed sequence of numbers follow a simple rule? (Another look at bode's law)*, Journal of the American Statistical Association, 66(335), 552-559.
- [۴] حاجزولهمن، مفاهیم پایدای احتمال و آمار، مترجم دکتر سیامک نوریلوچی، انتشارات آوای نور، تهران ۱۳۷۲.
- [۵] کمپتورن، اسکار، آزمونهای معنی دار بودن و آزمونهای فرض چه مصرفی دارند؟، مترجم دکتر محمدرضا مشکانی، اندیشه ریاضی.
- [6] O. Kempthorn and J. L. Folks, (1971), *Probability, Statistics, and Data Analysis*, The Iowa state University Press.
- [7] E. L. Lehman, (1993), *Testing Statistical Hypotheses*, Chapman & Hall.