از سطح معنی‌داری تا مقدار احتمال [۱]

مریم قدسی   نرگس عباسی [۲]

چکیده

در این مقاله آزمونهای معنی‌داری را با مثالهای مربوط در دو بخش ارایه می‌دهیم [۲]. سپس ساختار ابتدایی و مفهوم‌های 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 معنی‌داری و مقدار احتمال و اینکه مقدار احتمال کمیتی است که می‌تواند در بخشهای گوناگون، شکلهای متفاوتی را به خود بگیرد، می‌پردازیم.

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 معنی‌داری را با انواع مثالها، ارایه داد. البته حقیقتاً فیشر، مبدع بزرگ آزمونهای معنی‌داری، ماهیت آنها را به طرز مبهمی توضیح نداده بود. نیمن و پیرسون[۹] در سال ۱۹۲۸ نظریهٔ ریاضی را برای این روش ابداع کردند و بعداً یک نظریه کلی‌تر از آزمونها را با نام آزمون فرض به وجود آوردند که این نظریه تحت تأثیر تفکری است که بر مبنای آن آزمونهای آماری، عناصر نظریهٔ تصمیم می‌باشند.

---

۱ From Level of Significance Up P-Value

۲ مریم قدسی و نرگس عباسی، گروه آمار– دانشکده علوم، دانشگاه شیراز

۳ Arbuthnot

۴ K. Pearson

۵ Goodness of fit test

۶ W. S. Gosset

۷ Student

۸ R. A. Fisher

۹ Neyman and Pearson

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 معنی‌داری را در آزمون معنی‌داری و مقدار احتمال را در آزمون فرض معرفی کرده، سپس بحث را به اتمام می‌رسانیم. و از شما خواننده عزیز درخواست می‌کنیم که ابتدا مقاله را تا به انتها خوانده، سپس قضاوت کنید.

۲ آزمونهای معنی‌داری

در این فصل از اصطلاحات خاصی مانند آزمون معنی‌داری خلص[۱۰]، اکسترمم[۱۱]، آزمون سازگاری [۱۲]، نوع انحراف از $H_\circ$[۱۳]، دلیل انحراف[۱۴] استفاده شده است. شایان ذکر است که این اصطلاحات را کاکس و هینکلی[۱۵] در کتاب خود به کار برده‌اند و نویسندگان دیگر ممکن است از اصطلاحات دیگری استفاده نمایند.

۱.۲ آزمون معنی‌داری خلص

در این قسمت ابزار کار ما دو چیز است: فرضیه‌ای آماری به نام $H_\circ$، که تنها توزیع داده‌ای است که به وضوح فرمول‌بندی شده، و جهت انحراف از $H_\circ$ به این صورت است که همهٔ مجموعهٔ داده‌ها در یک جهت به طور یکتایی اکسترمم هستند و اگر این جهت انحرافات پرمعنیٰ از $H_\circ$ را ندانیم آزمون سازگاری داده‌ها با فرضیه آماری بی‌معنی است.
یک آزمون سازگاری داده‌ها با فرضیه $H_\circ$ را تحت ابزار خیلی محدود بالا، آزمون معنی‌داری خلص می‌نامند که برای تعریف و ارایه آن بایستی مراحل زیر را طی کرد:

۱) **فرضیه آماری** $H_\circ$: $H_\circ$ حکمی است دربارهٔ اینکه توزیع Y، $f_Y(y)$ می‌باشد، بنابراین $H_\circ$ توضیحی در ارتباط با چگالی $f_Y(y)$ می‌باشد که به دو صورت زیر ممکن است وجود داشته باشد:

- $H_\circ$ چگالی $f_Y(y)$ را کاملاً مشخص[۱۶] کند که در آن صورت، $H_\circ$ را ساده[۱۷] می‌نامیم.
- $H_\circ$ چگالی $f_Y(y)$ را مشخص[۱۸] کند اما به مقادیری از پارامترهای نامعلوم ولی معین بستگی داشته باشد که $H_\circ$ را در این حالت مرکب [۱۹] می‌نامیم.

۲) **آمارهٔ آزمون**: $t = t(y)$ را تابعی از مشاهدات و $T = t(y)$ را متغیر تصادفی مربوط به آن در نظر می‌گیریم، T را آمارهٔ آزمون[۲۰] می‌نامند، اگر دارای دو شرط زیر باشد:

- توزیعش تحت فرض $H_\circ$ کاملاً مشخص باشد و در حالتی که فرضیهٔ $H_\circ$ مرکب است، اگر آنرا به فرضیه‌های ساده تجزیه کنیم، همگی جزء یک خانواده توزیع باشند.
- بزرگترین مقدار t، قویترین دلیل انحراف از $H_\circ$ از نوعی که برای آزمون لازم است را بدهد.

۳) **تعیین سطح معنی‌داری و اعلام نتیجه**: برای مشاهدهٔ موجود، $t \equiv t_{obs} \equiv t(y)$ را محاسبه و سطح معنی‌داری[۲۱] را به صورت زیر/تعریف می‌کنیم:

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

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

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. می‌توان p_{obs} را به صورت فاصله داده‌ها از توزیع فرض شده برای مجموعه‌های دادهٔ ممکن که بر روی مقیاس [۰, ۱] برده شده اند، در نظر گرفت. مقادیر کوچک، فاصله زیاد و مقادیر بزرگ (یعنی مقادیر نزدیک به یک)، فاصله را کم نشان می‌دهند [۵]. یا اینکه p_{obs} را به عنوان یک اندازهٔ سازگاری داده‌ها، با H_0 با تفسیر فرضی زیر در نظر گرفت:

فرض کنید قرار باشد که داده‌های موجود را به عنوان دلیل علیه H_0 قبول کنیم در آن صورت ناچار خواهیم بود که حتی تمام داده‌های با مقدار بزرگتر را به عنوان دلیل قوی‌تر بپذیریم. بنابراین اگر قرار باشد که داده‌های 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 داده‌ها با H_0 در نظر گرفته‌ایم و صریحاً درگیر تصمیم‌های ناظر بر قبول یا رد H_0 نبوده، خصوصاً به اینکه آیا p_{obs} از یک مقدار از پیش تعیین شده‌ای مثل پنج درصد یا یک درصد بیشتر یا کمتراست، علاقه‌مند نیستیم. یعنی به دنبال این نمی‌باشیم که خطی مرزی بین داده‌ها ایجاد نماییم؛ به طوری که $p_{obs} < ۰/۰۵$ برقرار باشد. با این حال، سطح‌های استانداردی مانند یک‌درصد و پنج‌درصد و ... برای گزارش اینکه داده‌ها در سطح پنج‌درصد معنی‌دار نبوده، اما در سطح یک‌درصد معنی‌دار می‌باشند خیلی مناسب هستند یعنی شرط آشیان‌گزینی — که آن را بررسی خواهیم کرد — برقرار باشد. در واقع ما فقط وقتی می‌توانیم تصمیم رد و یا قبول H_0 را اتخاذ نماییم که آزمون ما براساس نظریه تصمیم پایه‌ریزی شده باشد.

مثال ۱ (Numerical law)

اغلب یک دنباله از اعداد از یک قانون ساده پیروی می‌کنند؛ حتی اگر هیچ طرح نظری برای آن پیدا نشود. یک مثال کلاسیک قانون باد[۲۲] است، که می‌گوید فاصله خورشید تا هر سیاره را به وسیله فرم هندسی ساده زیر می‌توان نشان داد:

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

به طوری که n ترتیب هر سیاره از خورشید است (مثلاً برای عطارد $n = ۱$، ونوس $n = ۲$، ...)، a و b ثابتهایی هستند که به وسیله قرار دادن اعداد واقعی در این رابطه به دست می‌آیند، مشروط بر اینکه فاصله خورشید تا زمین، برابر یک فرض شود در این صورت ثابتها برابرند با

$$a = ۰/۴ \quad , \quad b = ۰/۰۷۵$$

جدول زیر فواصل واقعی (dn) را با اعداد باد (b_n) مقایسه می‌کند:

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

می‌بینیم که اعداد واقعی با اعداد باد، متفاوتند؛ پس سؤالی که پیش می‌آید این است که: آیا قانون باد واقعی است یا اینکه ساخته ذهن و به طور مصنوعی می‌باشد؟ بنابراین با یک بحث آماری روبرو می‌شویم : رابطه کلی $y_i = a + bz_i$ (⋆) برای $i = 1, \ldots, n$ به طوری که $z_{i+1} - z_i = C$ را در نظر بگیرید. اگر داده‌ها از فرم فوق تبعیت نکنند، به صورت تصادفی می‌باشند. بنابراین فرض H_0 به صورت زیر در می‌آید:

H_0 : مدلی به داده‌ها برازش نشود

[۲۲] Bode's Law

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[۲۳] باشد. پس داریم:

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

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:

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

و

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

u_i ها را مشاهده‌هایی از آماره‌ها ترتیبی توزیع یکنواخت استاندارد درنظر می‌گیریم چون اولاً $u_{i+1} > u_i$ و ثانیاً $0 < u_i < 1$ برای $i = 1, \ldots, n-2$.

پس خواهیم داشت :

$$\begin{aligned} d_1 &= u_1 \\ d_2 &= u_2 - u_1 \\ &\vdots \\ d_{n-2} &= u_{n-2} - u_{n-3} \\ d_{n-1} &= 1 - u_{n-2} \end{aligned}$$

و بالاخره

$$H_0 \quad : \quad f_D(d_1, \ldots, d_{n-1}) = (n-2)!$$
$$(d_i \geq 0 \; , \; \sum_{i=1}^{n-1} d_i = 1)$$

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

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

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

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

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 $(\star)$ را حمایت می‌کنند.

۲.۲ آزمونهای معنی‌داری بهبودیافته

آزمون معنی‌داری عبارت است از محاسبهٔ 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 ارایه 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 اکسترمم‌تر 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 $0 < \alpha < 1$، w_α 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

[۲۳] Invariant

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

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

این شرط را شرط آشیان‌گزینی (Nesting Condition) می‌نامند و یک شرط اساسی است چون ممکن است کسی ادعا کند داده‌ها در سطح یک‌درصد انحراف معنی‌دار از $H_\circ$ دارند اما در سطح پنج‌درصد اینگونه نمی‌باشند.

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

$$Pr(Y \in w_\alpha \; ; \; H_\circ) = \alpha \;\; , \;\; 0 < \alpha < 1 \qquad (2)$$

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

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

هر آزمونی به روش بالا می‌تواند تعریف شود اما قصد ما ارایه بهترین آزمون در انواع مختلف w_α می‌باشد.

برای این کار فرض می‌کنیم α، دلخواه اما ثابت بوده، شرط آشیان‌گزینی (۱) حتماً برقرار است. برای راحتی فرض می‌کنیم که $H_\circ$ و H_A هر دو ساده می‌باشند و

$$H_\circ \; : \; f_\circ(y) \quad , \quad H_A \; : \; f_A(y)$$

وهمچنین w_α و w'_α دو ناحیه بحرانی اندازه α می‌باشند. با استفاده از (۲) داریم:

$$Pr(Y \in w_\alpha \; ; \; H_\circ) = Pr(Y \in w'_\alpha \; ; \; H_\circ) \qquad (4)$$

برای فرضیه مخالف H_A، w_α را به w'_α ترجیح می‌دهیم هرگاه

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

ناحیه بحرانی w_α را بهترین ناحیه بحرانی اندازه α ی آزمون می‌نامیم؛ اگر (۵) برای هر w'_α ایی که در شرط (۴) صدق کند برقرار باشد.

$Pr(Y \in w_\alpha \; ; \; H_A)$ را توان اندازه α ی آزمون می‌نامیم. پس از این به بعد هدف ما تعریف یک آزمون معنی‌داری به وسیله بهترین ناحیه بحرانی اندازه α می‌باشد.

نسبت درستنمایی $\frac{f_A(y)}{f_\circ(y)} = Lr_{A_\circ}(y)$ یک آماره بسنده می‌باشد، چون :

$$\begin{aligned} & Pr\{Y = y \mid Lr_{A_\circ}(Y) = l \; ; \; H_\circ\} \\ = & \frac{Pr\{Y = y \cap \frac{f_A(y)}{f_\circ(y)} = l \; ; \; H_\circ\}}{Pr\{\frac{f_A(y)}{f_\circ(y)} = l \; ; \; H_\circ\}} \\ = & \frac{f_\circ(y)}{\sum_l f_\circ(y)} = \frac{l f_\circ(y)}{\sum_l l f_\circ(y)} \\ = & \frac{f_A(y)}{\sum_l f_A(y)} = P\{Y = y \mid Lr_{A_\circ}(Y) = l \; ; \; H_A\} \end{aligned}$$

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

$$Pr\{Lr_{A_\circ}(Y) \geq C_\alpha \; ; \; H_\circ\} = \alpha$$

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 $Pr\{Lr_{A_\circ}(Y) \geq C_\alpha\}$ تعریف می‌گردد. این ناحیه بحرانی در شرط آشیان گزینی (۱) صدق می‌کند.

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

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 :

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

و در نتیجه

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

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 (٦) نتیجه می‌شود 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 هم‌ارز 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، بازهم آخرین نامساوی برقرار است؛ یعنی امکان ندارد که بتوانیم توان ناحیه بحرانی نسبت درستنمایی اندازه α را به وسیله ناحیه دیگری که دارای اندازه‌ای کمتر از α است، بهبود بخشیم.

مثال ۲ — میانگین با واریانس معلوم

فرض کنید $Y_1, Y_2, \ldots, Y_n$ نمونه تصادفی i.i.d از $N(\mu, \sigma_0^2)$ باشند، به طوری‌که σ_0^2 معلوم است و اگر فرضیه‌های فرضیه‌های صفر و مخالف را با توجه به اینکه $\mu_A > \mu_0$ است، به صورت زیر مشخص کنیم:

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

آنگاه یک محاسبه ساده نشان می‌دهد که

$$\begin{aligned} Lr_{A_0}(y) &= \frac{\frac{1}{(2\pi)^{\frac{n}{2}}\sigma_0^2}\exp\{-\frac{\sum(y_i-\mu_A)^2}{2\sigma_0^2}\}}{\frac{1}{(2\pi)^{\frac{n}{2}}\sigma_0^2}\exp\{-\frac{\sum(y_i-\mu_0)^2}{2\sigma_0^2}\}} \\ &= \exp\{-\frac{n\bar{y}(\mu_A-\mu_0) - \frac{n}{2}\mu_A^2 + \frac{n}{2}\mu_0^2}{\sigma_0^2}\} \end{aligned}$$

اما در بسط بالا همه کمیتها به جز $\bar{y}$ ثابت هستند و چون $\mu_A - \mu_0 > 0$ است، ناحیه بحرانی هم‌ارز با $Lr_{A_0}(y) \geq C_\alpha$، $\bar{y} \geq d_\alpha$ می‌باشد، این عمل، قدم اول را کامل می‌کند. در قدم دوم، ما d_α را به گونه‌ای انتخاب می‌کنیم که شرط آشیان‌گزینی برقرار باشد. $\bar{Y}$، تحت H_0، دارای توزیع $N(\mu_0, \frac{\sigma_0^2}{n})$ است، بنابراین

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

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

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

سطح معنی‌داری 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 می‌دادیم، ناحیه بحرانی به صورت

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

می‌شد و این بدان معناست که جهت انحراف از H_0 را باید به خوبی بدانیم و جهت اکسترمم داده‌ها را تشخیص دهیم.

۳ آزمون فرض

اولین نکته مطرح شده د ر آزمون فرض این است که کسی می‌خواهد تصمیم بگیرد که آیا یک فرضیه فرمول‌بندی شده صحیح است یا نه؟ انتخاب بین دو تصمیم است، رد یا قبول فرضیه. یک عملگر تصمیم برای چنین مسألهای آزمون فرض نامیده می‌شود. تصمیم براساس مقدار یک متغیر تصادفی Y پایه‌ریزی می‌شود. چگالیهای در H_A را، مخالفهای H_0 می‌نامیم. به طوری که H_A کلاسی از مخالفهاست. فرض کنید تصمیم‌های مربوط به رد یا قبول H_0 را به ترتیب با d_0 و d_1 تعریف می‌کنیم. یک آزمون غیر تصادفی هر مقدار y از Y را به یکی از تصمیم‌های d_0 و d_1 نسبت می‌دهد و بنابراین فضای

نمونه هم به دو ناحیه مکمل $S_{\circ}$ و S_{1} تقسیم می‌شوند. اگر Y در $S_{\circ}$ قرار گیرد فرضیه پذیرفته می‌شود و در غیر این صورت رد می‌شود. مجموعه $S_{\circ}$ 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 $H_{\circ}$ وقتی که آن درست است می‌گذاریم و تلاش برای کاهش احتمال دیگر را مشروط به این عمل می‌کنیم. بنابراین ابتدا $1 > \alpha > \circ$ را انتخاب نموده، و آن را سطح معنی‌داری [۷] می‌نامیم که عبارتست از:

$$P(\delta(Y) = d_{1}; H_{\circ}) = P(Y \in S_{1}; H_{\circ}) = \alpha \qquad (۷)$$

آنگاه به شرط معلوم بودن α عبارت :

$$Pr(\delta(Y) = d_{1}; H_{A}) = P(Y \in S_{1}; H_{A}) \qquad (۸)$$

را حداکثر نموده، داریم:

$$sup_{H_{\circ}} \; Pr(Y \in S_{1}) = \alpha \qquad (۹)$$

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

انتخاب یک سطح معنی‌داری α، معمولاً اختیاری است و این نشان‌دهنده مطالعه بیشتر در این زمینه است. در عمل معمولاً نواحی متفاوت با سطح‌های معنی‌داری مختلف، در دسترس است، پس مناسب است که علاوه بر اعلام نتیجه رد یا قبول فرضیه در سطح معنی‌داری داده شده، کوچکترین سطح معنی‌داری $\hat{\alpha} = \hat{\alpha}(y)$ که احتمال معنی‌داری یا مقدار احتمال (p-value)، نامیده می‌شود نیز ارایه گردد.

۴ چند سخن

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

اما در آزمون فرض مقدار احتمال فقط یک کمیت اسکالر 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 نزدیکتری 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، Pr_{θ} با پارامتر θ، دارای خاصیت MLR در $T(y)$ باشد و مسأله آزمون $H_{\circ} \; : \; \theta \leq \theta_{\circ}$ در برابر $H_{A} \; : \; \theta > \theta_{\circ}$ را در نظر بگیرید. اگر توزیع T پیوسته باشد، مقدار احتمال، آزمون UMP به صورت $\hat{\alpha} = p-value = P_{\theta_{\circ}}(T \geq t)$ داده می‌شود. که t مقدار مشاهده T است. اگر برای آزمونهای تصادفی، $\hat{\alpha}$ را به عنوان

کوچکترین سطح معنی‌داری که در آن فرضیه با احتمال یک رد می‌شود، تعریف کنیم این تساوی بدون فرض پیوستگی نیز برقرار است.
پس می‌توان 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 می‌نماییم.

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] هاجز و لهمن، مفاهیم پایه‌ای احتمال و آمار، مترجم دکتر سیامک نوربلوچی، انتشارات آوای نور، تهران ۱۳۷۲.

[۵] کمپتورن، اسکار، آزمونهای معنی‌داربودن 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.