بسم ا... الرحمن الرحیم

هست کلید در گنج حکیم

دانشکده علوم پایه

پایان نامه کارشناسی ارشد در رشته آمار

P – مقدار اصلاح شده و کاربردهای آن در آزمون های فرض آماری

توسط :

حمید اسماعیلی

استاد راهنما :

دکتر مینا توحیدی

دی ماه ۱۳۸۸

به نام خدا

اظهار نامه

اینجانب حمید اسماعیلی دانشجوی آمار ریاضی دانشکده‌ی علوم اظهار می‌کنم که این پایان نامه حاصل پژوهش خودم بوده ودرجاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌کنم که تحقیق و موضوع پایان نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده یا در اختیار غیر قرار ندهم . کلیه حقوق این اثر مطابق با آیین نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است .

نام و نام خانوادگی : حمید اسماعیلی

تاریخ و امضا : ۱۳۸۹/۲/۱۰

به نام خدا

P – مقدار اصلاح شده و کاربردهای آن در آزمون های فرض آماری

به وسیله :

حمید اسماعیلی

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی
از فعالیت های لازم برای اخذ درجه کارشناسی ارشد

در رشته:

آمار ریاضی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه: عالی

دکتر مینا توحیدی ، استادیار بخش آمار (رئیس کمیته)..

دکتر عبدالرسول برهانی حقیقی ، استادیار بخش آمار ....

دکتر عبدالرضا بازرگان لاری ، استادیار بخش آمار ....

دی ماه ۱۳۸۸

پیشکش به روح پدر مهربانم

و تقدیم به مادر عزیز و دلسوزم.

سپاسگزاری

من بی تو دمی قرار نتوانم کرد | احسان ترا شمار نتوانم کرد

گربر تن من زبان شود هر مویی | یک شکر تو از هزار نتوانم کرد

حمد و سپاس پروردگاری را که نورش چراغ معرفت در دلها افروخت و پرده های جهل و نـادانی را از چشمهای مشتاقانش برداشته و بوسیله علم آنان را از ظلمات به سمت نور هدایت نمود.

سپاس بیکران نثار یکتا خداوندی که تمام درختان و آبها اگر قلم و مرکب شوند نیز هیچگاه نتواننـد آمار کوچکی از آفریده هایش را بر کاغذ آورند .

اکنون که در آستانه گذر از دومین مقطع دانشجویی خود هستم بر خود لازم می دانـم ، سپاسـگذار اساتید و معلمان عزیزی باشم که ذره ذره معرفت را از کلامشان آموختم ، همانها که تعبیر عظـیم و انسانی از کلمه ایثار و ازخود گذشتگی اند تا آنجا که عمر خود را سراسر صرف خدمت در سنگر علم و دانش نموده اند .

پیشکش به روح مهربان پدری که افقهای فرداو آینده زندگی مرا با دستهای پینه بسته اش به علم و دانش گره زد .

سجده شکر بر آستان مهر مهربانی مادری می سایم که همواره در سردترین و سخت ترین روزهـای زندگی بهترین پشتیبان من بوده است .

مراتب سپاس صمیمانه خود را نثار استاد عزیز و فرهیخته سرکار خانم دکتر مینا توحیدی می نمایم که رهنمودهایشان در تدوین این اثر در جای جای و سطر به سطر آن نمودار است .

قدردان زحمات بی دریغ اساتید بزرگوار جناب آقای دکتـر عبدالرسـول برهـانی حقیقـی و دکتـر عبدالرضا بازرگان لاری هستم که مرا در این راه صمیمانه یاری نمودند .

تقدیم به تمام کسانی که حقیقت و علم و دانش بر تمام وجودشان سیطره افکنده و در دل و جانشان تجلی نموده است .

ای بار خدا به حق هستی | شش چیز به ما عطا فرستی

ایمان و امام و تندرستی | علم و عمل و فراخ دستی

چکیده :

P – مقدار اصلاح شده و کاربردهای آن در آزمون های فرض آماری

به کوشش

حمید اسماعیلی

در بسیاری از آزمون‌هایی که ما انجام می‌دهیم ، برای رد یا تایید فرض صفر از p-value استفاده می‌کنیم . ما می‌توانیم در بسیاری از این آزمون‌ها به راحتی با چند آزمایش اولیه و تجربی ، برای پارامتر مورد علاقه ، یک کران بالا و یک کران پایین به دست آوریم اما با توجه به اینکه p-value تنها به مقدار مشخص شده θ_0 و فرض صفر بستگی دارد ، نمی‌تواند به شایستگی از این اطلاعات برای گرفتن یک تصمیم بهتر و درستتر ، استفاده کند . ما در این پایان نامه ، در ابتدا به بررسی p-value می‌پردازیم و میزان دقت آن برای رد یا تایید فرض صفر بررسی می‌کنیم و همچنین بعضی از معایب p-value را از دو دیدگاه آمار کلاسیک و آمار بیز بیان می‌کنیم . سپس به معرفی معیاری به نام p-value اصلاح شده خواهیم پرداخت و نشان می‌دهیم وقتی که فضای پارامتری کراندار باشد ، p-value اصلاح شده ، که به خوبی از اطلاعات مربوط به فضای پارامتری استفاده می‌کند ، معیار بهتری نسبت به p-value برای رد یا تایید فرض صفر است .

فهرست مطالب

عنوان **صفحه**

فصل اول

تعاریف اولیه

۱-۱) تعاریف اولیه :

از آنجایی که در این پایان نامه از اصطلاحات آمار بیز ، نظریه تصمیم و استنباط آماری بسیار زیاد استفاده کرده‌ایم ، بنابراین در ابتدا تعریف و مفهوم‌های این اصطلاحات را بیان می‌کنیم و همچنین چند قضیه مهم نظریه تصمیم را نیز بیان خواهیم کرد .

تعریف ۱-۱-۱ : تابع تصمیم : تابعی با دامنه $\mathrm{O_n}$ (فضای یافته‌ها) و برد A (فضای کارها) است . بنابراین ، هنگام مشاهده هر $\mathrm{x \in O_n}$ ، به یک کار a می‌پردازیم .

تعریف ۱-۱-۲ : تابع زیان : تابعی دو متغیره ، کراندار و نامنفی با دامنه $\mathrm{T \times A}$ (حاصلضرب دکارتی فضای اوضاع طبیعت و فضای کارها) است که آن را با $\mathrm{L}(\theta,\mathrm{a})$ نشان می‌دهند .

در زیر به معرفی چند تابع زیان معروف و پرکاربرد می‌پردازیم :

تابع زیان درجه دو : تابع زیان $\mathrm{L}(\theta,\mathrm{a})=(\theta-\mathrm{a})^2$ را تابع زیان درجه دوم ، یا توان دوم خطا می‌نامند .

تابع زیان قدرمطلق خطا : تابع زیان $\mathrm{L}(\theta,\mathrm{a})=|\theta-\mathrm{a}|$ را تابع زیان قدرمطلق خطا می‌نامند .

تابع زیان خطی : تابع زیان $\mathrm{c,d}>0$, $\mathrm{L}(\theta,\mathrm{a})=\begin{cases}\mathrm{c}(\theta-\mathrm{a}) & \theta\ge \mathrm{a}\\ \mathrm{d}(\mathrm{a}-\theta) & \theta<\mathrm{a}\end{cases}$ را تابع زیان خطی می‌نامند .

تابع زیان صفر و یک : فرض کنید که $\mathrm{A}=\{\mathrm{a_0,a_1}\}$، فضای کارها باشد . کار $\mathrm{a_0}$ را پذیرفتن فرض $\mathrm{H_0}$ و کار $\mathrm{a_1}$ را پذیرفتن فرض $\mathrm{H_1}$ می‌گیریم .تابع زیر را که برای $\mathrm{a_0}$ و $\mathrm{a_1}$ به صورت زیر تعریف شده است ، تابع زیان صفر و یک می‌نامند :

$$L(\theta,a_0)=\begin{cases}0 & \theta=\theta_0\\ 1 & \theta=\theta_1\end{cases} \quad , \quad L(\theta,a_1)=\begin{cases}1 & \theta=\theta_0\\ 0 & \theta=\theta_1\end{cases}$$

تعریف ۱-۱-۳ : تابع ریسک : برای هر تصمیم مشخص d ، تابع ریسک تابعی است نامنفی و پایاندار با دامنه T که آن را با $R_d(\theta)$ نشان می‌دهیم و به صورت زیر تعریف می‌کنیم :

$$R_d(\theta)=E\left[L\left(\theta,d\left(\underline{X}\right)\right)\right]$$

تعریف ۱-۱-۴ : ریسک بیز : فرض کنیم $R_d(\theta)$ تابع ریسک برای تصمیم d (با تابع زیان از پیش تعیین شده) و $\pi(\theta)$ چگالی پیشین θ باشد . ریسک بیز عبارت است از :

$$r_{d,\pi}=E\left[R_d(\theta)\right]$$

تعریف ۱-۱-۵ : تصمیم با داده : تصمیمی است که به یافته $\underline{X}$ بستگی دارد و با مشاهده $\underline{x}$ باید کار a را انجام داد . اگر $\underline{x}$ تغییر کند ، کار a نیز تغییر خواهد کرد .

تعریف ۱-۱-۶ : تصمیم بی داده : تصمیمی است که به یافته $\underline{X}$ بستگی ندارد یعنی یافته $\underline{X}$ هرچه باشد $d(\underline{X})=a_j$.در حقیقت تصمیم بی داده را می‌توان حالت خاصی از تصمیم با داده در نظر گرفت که در آن داده‌ها نقشی ندارند .

تعریف ۱-۱-۷ : تصمیم بیز : *فرض کنیم* D *مجموعه تصمیم‌ها و* $\pi(\theta)$ چگالی پیشین *مشخصی باشد . تصمیم* $d_b\in D$ *را یک تصمیم بیز نسبت به* $\pi(\theta)$ *می‌نامیم ، هرگاه برای هر* $d\in D$ *داشته باشیم :*

$$r_{d_b,\pi}\le r_{d,\pi}$$

به سخنی دیگر ، d_b *را تصمیم بیز می‌گوییم هرگاه کمترین ریسک بیز را دارا باشد ، یعنی*

$$r_{d_b,\pi}=\min_{d\in D} r_{d,\pi}$$

تعریف ۱-۱-۸ : تصمیم مینیماکس : *فرض کنیم* D *مجموعه تصمیم‌ها و* $R_d(\theta)$ تابع ریسک برای تصمیم $d\in D$ با تابع زیان $L(\theta,d)$ *باشد . تصمیم* $d_m\in D$ *را مینیماکس گوییم ، هرگاه برای هر* $d\in D$ *داشته باشیم :*

$$\max_{\theta\in T} R_{d_m}(\theta) \le \max_{\theta\in T} R_d(\theta)$$

تعریف ۱-۱-۹ : مقایسه دو تصمیم : *فرض کنیم* $R_1(\theta)$ و $R_2(\theta)$ *به ترتیب تابع ریسک دو تصمیم* d_1 و d_2 *باشند :*

الف) d_1 *را به خوبی* d_2 *می‌گوییم ، هرگاه* $R_1(\theta) \le R_2(\theta)$ *برای هر* $\theta \in T$.

ب) d_1 *را بهتر از* d_2 *می‌گوییم ، هرگاه* $R_1(\theta) \le R_2(\theta)$ *برای هر* $\theta \in T$ و $R_1(\theta) < R_2(\theta)$ *برای برخی* $\theta \in T$.

ج) d_1 و d_2 *را هم‌ارز می‌گوییم ، هرگاه* $R_1(\theta) = R_2(\theta)$ *برای هر* $\theta \in T$.

د) d_1 و d_2 *را غیر قابل مقایسه می‌گوییم ، هرگاه هیچکدام از سه حالت بالا برقرار نباشد .*

تعریف ۱-۱-۱۰ : تصمیم روا : *تصمیم* d_a *را روا گوییم ، اگر تصمیم دیگری در* D *نتوان پیدا کرد که از* d_a *بهتر باشد .*

تعریف ۱-۱-۱۱ : *کلاس کامل* *: فرض کنیم* $C \subset D$. *مجموعه* C *را یک کلاس کامل می‌گوییم اگر برای هر* $d' \notin C$ ، *یک تصمیم* $d \in C$ *داشته باشیم به طوری که* d *بهتر از* d' *باشد .*

تعریف ۱-۱-۱۲ : *آزمون مینیماکس :* *آزمونی که بر اساس تصمیم مینیماکس* $d_m \in D$ *باشد ، آزمون مینیماکس نامیده می‌شود . بنابراین* d_m *را باید به گونه‌ای انتخاب کنیم تا داشته باشیم :*

$$\min_d \left\{ \max\left\{ R_d(\theta_0), R_d(\theta_1) \right\} \right\} = \max\left\{ R_{d_m}(\theta_0), R_{d_m}(\theta_1) \right\}$$

تعریف ۱-۱-۱۳ : *آزمون بیز :* *فرض کنیم متغیر تصادفی* X *دارای چگالی* $f(x|\theta)$ *باشد و بخواهیم آزمون* $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$ *را بر اساس نمونه تصادفی از* X *انجام دهیم که یافته آن بردار عددی* $\underline{x} = (x_1, \dots, x_n)$ *است . چگالی پیشین پارامتر* θ *را به صورت* $\pi(\theta) = \begin{cases} \pi_0 & \theta = \theta_0 \\ \pi_1 & \theta = \theta_1 \end{cases}$ *تعریف می‌کنیم . در صورتی که به تابع زیان توجهی نداشته باشیم*

فرض H_0 را وقتی رد می‌کنیم که احتمال پسین در θ_0 کمتر از احتمال پسین در θ_1 باشد یعنی : $\pi(\theta_0|x) < \pi(\theta_1|x)$. اما اگر ما تابع زیانی داشته باشیم که در آن $L(\theta_0,a_1) = A$ و $L(\theta_1,a_0) = B$ باشد ، آنگاه فرض H_0 را وقتی رد می‌کنیم که امید پسین زیان کار a_1 کمتر از امید پسین زیان کار a_0 باشد ، یعنی : $k\pi_0 L(\theta_0|\underline{x})\times A < k\pi_1 L(\theta_1|\underline{x})\times B$

۱-۲) چند قضیه :

در این بخش چند قضیه پرکاربرد و مهم نظریه تصمیم را بیان می‌کنیم .

قضیه ۱-۲-۱(قضیه ریسک بیز پسین) : *فرض کنیم d_b تصمیم بیز با تابع زیان $L(\theta,d)$ در میان تمام تصمیم‌ها باشد . ریسک بیز پسین d_b کمترین مقدار را دارد .*

قضیه ۱-۲-۲(قضیه تصمیم بیز و مینیماکس) : *فرض کنیم d_b تصمیم بیز نسبت به چگالی پیشین $\pi(\theta)$ باشد . اگر تابع ریسک d_b ، یعنی $R_{d_b}(\theta)$ ، روی T (فضای پارامتر و یا مجموعه اوضاع طبیعت) ثابت باشد ، آنگاه d_b یک تصمیم مینیماکس است .*

قضیه ۱-۲-۳(قضیه کلاس کامل و مجموعه روا) : *فرض کنیم C یک کلاس کامل باشد . اگر Qمجموعه تصمیم‌های روا باشد ، آنگاه $Q \subset C$.*

قضیه ۱-۲-۴(قضیه بیز و تصمیم روا) : *اگر d_b تصمیم بیز یکتا با چگالی پیشین $\pi(\theta)$ و تابع زیان $L(\theta,d)$ باشد ، آنگاه d_b تصمیمی رواست .*

قضیه ۱-۲-۵(قضیه تصمیم روا و تصمیم مینیماکس) : *اگر d_a تصمیمی روا با تابع ریسک ثابت روی فضای اوضاع طبیعت یعنی T باشد ، آنگاه d_a یک تصمیم مینیماکس است .*

قضیه ۱-۲-۶(قضیه آزمون مینیماکس) : *فرض کنیم $g(\underline{x}|\theta)$ چگالی نمونه تصادفی $\underline{X} = (X_1,...,X_n)$ باشد ، ناحیه بحرانی $C_m = \left\{\underline{x}\left|\frac{g(\underline{x}|\theta_1)}{g(\underline{x}|\theta_0)} > c\right.\right\}$ را طوری می‌سازیم که مقدار ثابت c در $R_{d_m}(\theta_0) = R_{d_m}(\theta_1)$ صدق کند ، آنگاه تصمیم $d_m = \begin{cases} a_0 & \underline{x} \in \overline{C_m} \\ a_1 & \underline{x} \in C_m \end{cases}$ و آزمون متناظر با آن مینیماکس است . (نسبت بالا همان نسبت راستنمایی است) .*

قضیه ۱-۲-۷ : *اگر d_b یک تصمیم بیز نسبت به توزیع پیشین π باشد به طوری که ریسک بیز آن متناهی باشد ، آنگاه d_b تصمیمی رواست .*

فصل دوم

برآورد درستی و دقت در آزمون فرض

۲-۱)مقدمه :

روش‌های معمولی که ما با آنها آزمون فرض ر ا انجام می‌دهیم ، معمولاً با یک مساله آزمون فرض به عنوان یک مساله انتخاب تصمیم (بین H_0 و H_1) و نه به عنوان یک مساله برآوردیابی برخورد می‌کنند . به طور دقیق تر نتیجه‌ای که از انجام یک آزمون فرض رسمی و متداول گرفته می‌شود ، این است که آیا یک فرض درست است یا خیر ؟ و یک اندازه معیاری برای بررسی دقت آزمون ارائه نمی‌شود . در این فصل فرض می کنیم که آزمون فرض ها ، یک مساله برآوردیابی در نظریه تصمیم باشد و می‌توانیم نتایج جالب و قابل توجه‌ای بدست می-آوریم . مخصوصا تابع زیانهای قابل قبول و خوبی در نظریه تصمیم وجود دارد که تحت آنها برآوردگرهایی به دست می‌آوریم که می‌توان از آنها به عنوان یک اندازه معیار بر علیه فرض صفر استفاده کرد و تحت این تابع زَیان ها ، بعضی نتایج جالب برای p-value بدست می آید .

۲-۱-۱ روش استاندارد : آزمون فرض کلاسیک بر اساس لم نیمن-پیرسون انجام می شود . (Lehman(۱۹۸۶)) و منجر به توابع تصمیمی می شود که تابع تصمیم 0-1 نامیده می شوند (به جز در مورد آزمون های تصادفی شده) . این شیوه متداول و رایج برای انجام دادن آزمون‌ها ، هر چند از نظر آماردانان کلاسیک بسیار بهینه و خوبند ولی از جهات مختلف مورد انتقاد واقع شده اند . اولاً بسیاری از آماردانان بیز مانند (Berger(۱۹۸۵a,b ، Dickey(۱۹۷۷) و Degroot (۱۹۷۳) در تحقیقاتشان مسایل بیزی را مطرح کردند که به اشکالات لم نیمن-پیرسون با یک بحث دقیق اشاره می کند . آنها نشان دادند که برای آزمون دو طرفه $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta \neq \theta_0$ (آزمون‌هایی که فرض صفر در آنها تک نقطه‌ای است) بین لم نیمن-پیرسون و آزمون بیز تناقض وجود دارد . در واقع در حالی که لم نیمن-پیرسون فرض H_0 را رد می کند ، اما احتمال بیز پسین H_0 نزدیک به یک می شد و فرض H_0 را تایید می کرد !! اما (Degroot(۱۹۷۳ و (Casella and Berger (۱۹۸۷ a نشان دادند که این تناقض برای آزمون یک طرفه $H_0 : \theta \leq \theta_0$ در مقابل $H_1 : \theta > \theta_0$ وجود ندارد و ثابت

کردند که برای توزیع های پیشین ناسره ، مقدار p-value (معیار بدست آمده برای رد یا تایید فرض H_0 با لم نیمن-پیرسون) و احتمال بیز پسین H_0 با هم برابرند . دومین ایرادی که به لم نیمن-پیرسون وارد است این است که معیارهایی که برای ارزیابی درستی و برآورد دقت یک آزمون استفاده می شوند ، نوعاً معیارهایی هستند که از قبل مشخص شده است (Predata) که در بیشتر مواقع اندازه آزمون (α) می باشد . (وقتی نیمن و پیرسون ، لم خود را به Fisher ، آماردان برجسته انگلیسی ارائه دادند ، Fisher با بیان همین عیب یعنی از قبل مشخص کردن میزان دقت در آزمون ((α)) ، حاضر به پذیرش لم نیمن-پیرسون نشد .) یک ایرادی که همواره به لم نیمن-پیرسون و شروطش وارد می‌شود این می باشد که وقتی Postdata(داده‌های پسین) مشاهده شود ممکن است این معیارها کاملاً غیر قابل قبول باشد(kiefer(۱۹۷۷)، Robinson(۱۹۷۹a,b)) . راه دیگری که توسط kiefer پیشنهاد شد استفاده از p-value برای ارزیابی درستی و دقت فرض صفر است . در واقع kiefer و بیشتر آماردانان ترجیح می‌دهند p-value به کار ببرند به آن سبب که عقیده دارند p-value دلیل محکمی علیه رد فرض صفر ارائه می‌دهد . این ایده ها هم سو با هدف این فصل است که میزان دقت آزمون فرضها بر اساس معیاری برحسب Postdata ها ارزیابی شود . (از اشکالات لم نیمن-پیرسون این است که برای برآورد دقت در طی روند آزمون از (α) و توان آزمون استفاده می کند که آنها بر اساس Predata هستند نه Postdata .)

موضوع مهم در تحقیقات علمی این است که آنها همگی از p-value برای نتیجه گیری در مورد درستی یا نادرستی فرض صفر استفاده می کنند و با نتایج دیگر آزمون در سطح (α) (مانند ناحیه بحرانی،تابع آزمون،...) کاری ندارند ، در واقع p-value بی قید و شرط به عنوان اندازه ی معیاری برای تایید و یا رد فرض H_0 استفاده می شود . یکی ازدلایل انجام تحقیق اخیر این است که ما به دنبال دلیل های معقولی در آزمون هستیم که p-value را به عنوان یک معیار قابل قبول بپذیریم . استفاده از p-value به طور وسیع در آزمایشات انجام می شود و ما به عنوان آماردانان باید تصمیم بگیریم که p-value ویژگی قابل قبول بودن را دارد یا نه؟

۲-۱-۲ معایب p-value : بیشترین انتقادات از p-value در آمار بیز مطرح می شود ، هر چند این انتقادات در بقیه ی جاها نیز وجود دارد. حتی با وجود اینکه p-value می تواند مشابه احتمالات بیز پسین باشد باز تعداد زیادی نقص آشکار برای انتقاد کردن وجود دارد. از آنجا که در بیشتر موارد ، p-value به شکل $P(x)=P(T(X)>T(x))$ است که $T(x)$

مقدار مشاهده شده ی متغیر تصادفی $T(X)$ است ،این مشکل و اعتراض همواره روی مقادیر مشاهده شده بعید و دور از ذهن نمونه وجود دارد (که هرگز اتفاق نمی افتند) . مثلاً اگر $T(x)$ ما مقدار بسیار بزرگی شود آنگاه p-value به سمت صفر می رود و فرض H_0 همواره تایید می شود . علاوه بر این ، در صورت نادیده گرفتن این مقادیر ،این در تضاد با اصل درست نمایی است که بیان می کند که استنباط ها باید تنها بر اساس مقادیر مشاهده شده باشند . ((۱۹۸۴) Berger and Wolpert) . (۱۹۸۷a) Casella and Berger یک رابطه بین p-value و دلایل قبول یا رد فرض صفر در آزمون بیز را بیان کردند . آنها نشان دادند که دلایلی که برای رد یا قبول فرض صفر در آزمون بیزی به کار می‌رود را می‌توان برحسب آماره-ای از مشاهدات که p-value از روی آنها بدست آورده می‌شود بیان کرد . (۱۹۸۷a) Casella and Berger نشان دادند که برای آزمون یک طرفه $H_0:\theta\le\theta_0$ در مقابل $H_1:\theta>\theta_0$ ، هرگاه تابع چگالی از خانواده پارامتر مکانی باشد ($f_\theta(x)=f_\theta(x-\theta)$) و دارای خاصیت نسبت درستنمایی یکنوا (MLR) و متقارن حول صفر باشد ، آنگاه p-value برابر با احتمال بیز پسین $P(H_0|x\in G_{us})$ است که G_{us} مجموعه توزیع های پیشین تک مدی θ_0 و متقارن حول θ_0 است . ما در این مورد که p-value یک معیار خوب است تردید و نگرانی نداریم و روندها و معادلات بیز نیز ممکن است با هم تفاوت داشته باشند اما تعداد زیادی انتقاد در آمار بیز بر اساس تناقضات نسبت به p-value مطرح می‌شود که ما در زیر به تعدادی از این انتقادات اشاره می‌کنیم :

Lindley در سال ۱۹۵۷ نشان داد که برای آزمون دو طرفه $H_0:\theta=\theta_0$ در مقابل $H_1:\theta\ne\theta_0$ در توزیع نرمال بین استفاده از p-value و استفاده از احتمال بیز پسین تناقض وجود دارد . او نشان داد که ممکن است فرض صفر با p-value رد شود ولی احتمال بیز پسین H_0 مقدار زیادی شود که در نتیجه فرض صفر را تایید کند . بعدها آماردانان نشان دادند که اختلاف بین آزمون معنی داری و آزمون های بیزی در فرض های دو طرفه به توزیع پیشین و یا تعداد نمونه بستگی ندارد و برای نمونه های کوچک نیز وجود دارد . در واقع اختلاف ذکر شده بستگی به مجموعه‌ای دارد که تحت آن بزرگترین کران پایین احتمال پسین فرض صفر بدست آورده می‌شود .

سی سال بعد ، Berger and delampady در سال ۱۹۸۷ ، این تناقض بین p-value و احتمال بیز پسین را در آزمون دو طرفه برای خانواده توزیع دوجمله‌ای نشان دادند . آنها ثابت

کردند که اگر توزیع پیشین ، متقارن حول θ_0 با مد θ_0 و یا مزدوج با میانگین θ_0 و یا هر توزیع ممکن با میانه θ_0 باشد ، بین مقدار p-value و احتمال بیز پسین H_0 ، اختلاف فاحشی وجود دارد . همچنین Berger and sellke در همین سال ، نشان دادند که در آزمونهای دو طرفه برای توزیع نرمال ، حداقل احتمال بیز پسین H_0 (وقتی چگالی پیشین هر توزیع دلخواهی باشد) به مقدار قابل ملاحظه‌ای از p-value بیشتر است .

همان گونه که دیدید این تناقضات همه بر اساس این حقیقت بود که در رخدادها و آزمایشات ، ممکن است p-value بسیار کوچکتر از احتمالات بیز پسین در آزمونهای دو طرفه باشد . (در واقع با احتمال بیز پسین فرض صفر تایید می شود درحالیکه p-value فرض صفر را رد می کند .) اما همان‌طور که قبلاً گفتیم Casella and Berger در سال ۱۹۸۷ نشان دادند که در آزمون های یک طرفه این تناقض وجود ندارد و p-value به صورت یک حدی از برآوردگر بیز است . این که ما بیان می‌کنیم که p-value در آزمون‌های یک طرفه می‌تواند برابر با احتمال بیز پسین باشد ، ممکن است با پاراگراف اول که ما گفتیم p-value در تضاد با اصل درست نمایی است تناقض داشته باشد ، اصلی که با قانون بیز ارتباط چندانی ندارد . اما باید دقت کنیم که برابری بین p-value و احتمالات بیز پسین تنها یک برابری ریاضی یعنی برابری دو انتگرال متفاوت است . اساساً محاسبات که ما برای بدست آوردن این دو مقدار انجام می‌دهیم با هم متفاوت است .

از طرف دیگر معایب بیز و مسائل مربوط به p-value ممکن است در آمار کلاسیک نیز رخ دهد . یک آماردان کلاسیک نیز ممکن است با وجود علاقه بیشترش به انتخاب p-value تا برآوردگر بیز ، p-value را رها کند وقتی با استفاده از قضایای کلاسیک ، p-value یک توجیه منطقی نداشته باشد یعنی درحالیکه داده‌های مشاهده شده از فرض صفر حمایت می-کنند ، p-value بدست آمده فرض صفر را رد می‌کند و یا برعکس . در برخی از موارد تعریف p-value و بدست آوردن آن ممکن است سخت باشد . به طور مثال Berger and delampady نشان دادند که اگر X دارای توزیع دوجمله‌ای با پارامترهای n و θ باشد ($X \sim BIN(n,\theta)$). آنگاه برای آزمون دو طرفه $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta \neq \theta_0$ ، با یک سری محاسبات p-value به صورت زیر خواهد شد :

$$\text{p-value} = P_{\theta=\theta_0}\left(\left\{y : f\left(y|\theta=\theta_0\right) < f\left(x|\theta=\theta_0\right)\right\}\right)$$

اما اگر p-value با یک روش صحیح تعریف شود آن را معمولاً با استفاده از بعضی تابع آزمون های بهینه لم نیمن-پیرسون می توان به دست آورد .

همچنین بعضی تعاریف ارائه شده برای p-value نیز درست نیست که ما در زیر به بعضی از این تعاریف اشاره می‌کنیم :

۱) برخی اوقات p-value را احتمال خطای نوع اول می‌نامند درحالیکه احتمال خطای نوع اول را خود محقق تعیین می‌کند و به توزیع آماره آزمون تحت فرض H_0 بستگی دارد و تابعی از مشاهدات نیست .

۲) برخی p-value را به عنوان احتمال درستی H_0 در نظر می‌گیرند درحالیکه احتمال درستی H_0 کاملاً با p-value متفاوت است زیرا اگر یک توزیع پیشین روی فضای $\{H_0, H_1\}$ و p-value=P درنظر بگیریم آنگاه $p(H_0|P=p)$ برابر است با

$$p(H_0|P=p)=\frac{f(p|H_0)\pi(H_0)}{f(p|H_0)\pi(H_0)+f(p|H_1)\pi(H_1)}$$

اگر $\pi(H_0)=\pi(H_1)=\frac{1}{2}$ درنظر بگیریم آنگاه

$$p(H_0|P=p)=\frac{f(p|H_0)}{f(p|H_0)+f(p|H_1)}$$

چون تحت فرض صفر $P \sim U(0,1)$ است بنابراین $f(p|H_0)=1$ است و در نتیجه

$$p(H_0|P=p)=\frac{1}{1+f(p|H_1)}$$

بنابراین همانطور که می‌بینیم حتی در حالت برابری احتمالات پیشین H_0 و H_1 ، احتمال درستی H_0 کاملاً متفاوت از p-value است .

۳) بعضی نیز p-value را احتمال پذیرش H_0 در نظر می‌گیرند اگر آزمایش تکرار شود که این نیز نادرست است زیرا احتمال پذیرش H_0 حتی اگر آزمایش(یا نمونه‌گیری) تحت همان شرایط اولیه تکرار شود ، همان خواهد بود که دربار اول بوده که اگر H_0 درست باشد برابر $1-\alpha$ و اگر H_0 درست نباشد برابر β است و ربطی به مقدار مشاهدات ندارد .

تعریفی که بیشتر آماردانان برای p-value در نظر می‌گیرند معیاری برای میزان پشتیبانی داده‌ها از فرض H_0 است ولی باز هم دارای یک سری معایبی می‌باشد که هم آماردانان کلاسیک و هم آماردانان بیز بسیار به آنها اشاره کرده‌اند . در زیر چند مورد از این عیب ها را بیان می‌کنیم :

۱) p-value به H_1 بستگی ندارد . به طور مثال اگر $X \sim N(\theta,\sigma^2)$ باشد ، آنگاه p-value برای هر دو آزمون $\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta = 1 \end{cases}$ و $\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta = 100 \end{cases}$ ، با مشاهده x = ۱/۵ برابر با ۰/۰۵ خواهد شد در حالی که اختلاف بین دو فرض H_1 بسیار واضح و زیاد است .

۲) p-value نسبت به H_0 صعودی (و یا نسبت به H_1 نزولی) نیست . اگر p-value برای آزمون $\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta > 0 \end{cases}$ برابر V_1 و برای آزمون $\begin{cases} H_0 : \theta = 0 \\ H_1' : \theta \neq 0 \end{cases}$ برابر V_2 در نظر بگیریم ، از آنجایی که $H_1 \subset H_1'$ ما انتظار داریم که $V_2 \le V_1$ باشد . اما به طور مثال برای مثال نرمال بالا $V_2 = 2V_1$ خواهد شد .

۳) p-value نسبت به H_0 و H_1 نقش متقارنی ندارد یعنی اگر

$$H_0 : A \quad , \; H_1 : B \Rightarrow \text{p-value} = V_1$$

$$H_0 : B \quad , \; H_1 : A \Rightarrow \text{p-value} = V_2$$

در حالی که انتظار داریم $V_2 = 1 - V_1$ باشد ولی اینگونه نیست . به طور مثال ، فرض کنیم $X \sim N(\theta,۱)$ باشد ، آنگاه برای آزمون $H_0 : \theta = 0$ در مقابل $H_1 : \theta = 1$ ، با مشاهده x ، $V_1 = P(X \ge x | \theta = 0) = 1 - \Phi(x)$. حال اگر جای فرض‌های H_0 و H_1 را عوض کنیم ، $V_2 = P(X \ge x | \theta = 1) = \Phi(1 - x)$.

۴) وقتی $H_1 \to H_0$ ، p-value به سمت $\frac{1}{2}$ نمی‌رود زیرا p-value به H_1 بستگی ندارد .

۵) p-value ، کالیبره نیست . وقتی p-value برابر با ۰/۰۳ باشد بدین معنی است که میزان پشتیبانی داده از فرض H_0 (در مقابل فرض H_1) ۳ درصد است . حال اگر p-value برابر ۰/۵ باشد ، آیا صحیح است که بگوییم داده‌ها به یک میزان از H_0 و H_1 پشتیبانی می‌کنند ؟!!!

عیب‌ها و نواقص p-value که بعضی از آنها را بیان کردیم و همچنین رفتار متفاوت p-value برای آزمون‌های یک طرفه و دو طرفه ، یک دلیل برای انجام تحقیق اخیر است . این رفتار متفاوت ، این فکر را مطرح می‌کند که ممکن است خود فرمول بندی مسئله ناقص باشد . به