

# برآورد فاصله اطمینان برای نسبت‌های نزدیک به صفر و یک: یک مطالعه ثانویه مدل سازی

سحر حمزه<sup>۱</sup>، علیرضا سلطانیان<sup>۲</sup>، جواد فردمال<sup>۲</sup>

<sup>۱</sup> کارشناسی ارشد گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۲</sup> دانشیار آمار زیستی، گروه آمار زیستی، مرکز تحقیقات مدل‌سازی بیماری‌های غیر واگیر، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

نویسنده رابط: علیرضا سلطانیان، نشانی: همدان، خیابان شهید فهمیده، دانشگاه علوم پزشکی همدان، تلفن: ۰۸۱۳۳۸۰۰۲۵، پست الکترونیک: soltani@umsha.ac.ir

تاریخ دریافت: ۹۵/۰۱/۱۹؛ پذیرش: ۹۵/۰۷/۰۳

**مقدمه و اهداف:** فاصله اطمینان توزیع دو جمله‌ای با پارامتر  $p$ ، باید دقیق و دارای احتمال پوششی  $1 - \alpha$  به ازای همه مقادیر  $P$  باشد. در این مطالعه خطای اسمی نوع اول و احتمال پوششی فاصله‌های اطمینان کلوفر-پیرسون، والد، ویلسون و تبدیل آرک سینوس مکرر مقایسه شدند.

**روش کار:** به منظور مقایسه‌ی برآورد فاصله اطمینان حاصل از چهار روش فوق، از شبیه‌سازی استفاده شد. داده‌ها ۱۰۰۰ بار از توزیع دو جمله‌ای با پارامترهای  $p$ ،  $n$  و  $\mu=np$  تولید شدند. برای شبیه‌سازی و تحلیل داده‌ها از نرم‌افزار R نسخه‌ی ۳.۰.۲ استفاده شد.

**یافته‌ها:** یافته‌های حاصل از مطالعه شبیه‌سازی شده نشان داد تبدیل آرک سینوس مکرر حدود اطمینان را بین فاصله  $[0,1]$  نگه می‌دارد و احتمال پوششی مناسبی دارد، اما برای برخی از مقادیر  $p$ ، خطای نوع اول بالا و احتمال پوششی پایینی دارد. فاصله اطمینان دقیق کلوفر-پیرسون حدود اطمینان را بین فاصله  $[0,1]$  نگه می‌دارد و خطای نوع اول یا به‌طور عکس احتمال پوششی قابل قبول و خوبی دارد. سایر روش‌ها حدود اطمینان را به ازای مقادیر غایی  $p$ ، بین نگه  $[0,1]$  نمی‌دارند.

**نتیجه‌گیری:** این مطالعه حاضر داد که برآورد فاصله اطمینان به روش کلوفر-پیرسون نسبت به سایر روش‌ها سطح اسمی خطای نوع اول را ثابت و کوچک‌تر محاسبه می‌کند.

**واژگان کلیدی:** توزیع دو جمله‌ای، نسبت، فواصل اطمینان دقیق و تقریبی، احتمال پوششی، خطای نوع اول

## مقدمه

یکی از اساسی‌ترین تجزیه و تحلیل‌ها در استنباط آماری، تشکیل فاصله اطمینان برای نسبت  $(p)$ ، یعنی پارامتر توزیع دو جمله‌ای است (۱). در برخی مواقع استنباط در این زمینه به دلیل نزدیک بودن مقدار نسبت به یک و صفر مشکلات محاسباتی به وجود می‌آورد (۲)، و تعیین فاصله اطمینان برای نسبت یک صفت دو حالتی به روش کلاسیک ممکن است در چنین حالت‌هایی نامعتبر باشد (۳).

پژوهشگران تاکنون روش‌های مختلفی را برای برآورد فاصله اطمینان پارامتر توزیع دو جمله‌ای  $(p)$  پیشنهاد کرده‌اند، که در یک تقسیم‌بندی ساده می‌توان آن‌ها را به دو دسته مختلف تقسیم کرد، یکی روش‌های دقیق و دیگری روش‌های تقریبی. سطح پوششی روش‌های دقیق برای حداقل است، در حالی که در روش‌های تقریبی این سطح ممکن است کم‌تر باشد (۲).

روش‌های تقریبی و دقیق<sup>۱</sup> بسیاری وجود دارند، که تعدادی از آن‌ها در این مقاله معرفی و مورد مقایسه قرار گرفتند. فاصله اطمینان تقریب نرمال به نام «والد» نام‌گذاری شده است؛ چرا که یافته‌های آن از معکوس شدن آزمون والد برای  $p$  به دست آمده است. در فاصله تقریب نرمال وقتی اندازه نمونه کوچک و نسبت خیلی نزدیک به صفر و یک است، دقت بالایی وجود ندارد (۴). ادوین بیدول ویلسون فاصله اطمینان تقریبی دیگری پیشنهاد کرد، که «فاصله ویلسون» نام دارد. احتمال پوششی فاصله اطمینان ویلسون به‌طور قابل قبولی نزدیک سطح اسمی  $1 - \alpha$  در نوسان است، به استثنای مواقعی که  $p$  خیلی نزدیک به صفر و یک باشد (۵). فاصله اطمینان جفری از پیشین بتا برای استنباط روی  $p$ ، پارامتر توزیع دو جمله‌ای استفاده می‌کند (۶). احتمال پوششی فاصله اطمینان جفری از لحاظ کیفیت در بیش‌تر نقاط فضای

<sup>۱</sup>Exact

همان‌گونه که بیان شد هدف از این مطالعه ارایه‌ی یک برآورد فاصله اطمینان مناسب، در نسبت نمونه است که در آن تحت هر شرایطی فاصله اطمینان حاصل شده در محدوده‌ی [۰,۱] قرارگیرد، و دارای بهترین احتمال پوششی برای هر مقداری از حجم نمونه و نسبت نمونه‌ای باشد. در واقع برای مواقعی که نسبت دوجمله‌ای نزدیک به صفر است، حد پایین منفی نشود و مواقعی که نسبت دوجمله‌ای نزدیک به یک است؛ حد بالا بزرگ‌تر از یک نشود.

## روش کار

در این مطالعه از داده‌های شبیه‌سازی که از توزیع دوجمله‌ای و پواسن تولید شده‌اند؛ استفاده شده است. در این بخش ابتدا ۴ روش استفاده شده برای برآورد فاصله اطمینان، والد، ویلسون، کلوپر-پیرسون و تبدیل آرک سینوس مکرر، به‌طور مختصر معرفی می‌شوند و سپس با استفاده از شبیه‌سازی به مقایسه این روش‌ها پرداخته می‌شود. در نهایت هم برای فهم بهتر مطلب یافته‌هایی که از داده‌های شبیه‌سازی شده روی انواع فاصله اطمینان‌ها به‌دست آمد، روی داده‌های واقعی به عنوان یک مثال اعمال شد. داده‌های واقعی، از یک مطالعه‌ی مرور سیستماتیک و متآنالیز استخراج شده که در آن مطالعه‌هایی که به میزان کشف ژن HER2 با استفاده از دو روش ایمونوهیستوشیمی (IHC)<sup>۱</sup> و فلورسانس در محل هیبریداسیون (FISH)<sup>۲</sup> توأمان پرداخته‌اند، مورد بررسی قرار گرفته‌اند (۲۰).

ابتدا فاصله اطمینان والد که در بیش‌تر متون آماری براساس نرمال بودن مجانبی نسبت نمونه و برآورد خطای استاندارد ارایه شده است؛ معرفی می‌شود و در ادامه به معرفی فواصل اطمینان ویلسون، کلوپر-پیرسون و تبدیل آرک سینوس مکرر پرداخته می‌شود.

فاصله اطمینان تقریبی والد:

فاصله اطمینان والد  $100(1-\alpha)\%$  برای  $p$   $p \pm z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{n}}$  است که نسبت موفقیت در یک توزیع برنولی است،  $z_c$  چارک  $1-c$  از توزیع نرمال استاندارد معرفی می‌شود، درصد خطا و اندازه نمونه است. این فاصله مجموعه‌ای از مقادیر است، که سطح معنی‌داری آن‌ها از در آزمون در مقابل، با استفاده از آزمون آماری تجاوز می‌کند (۴). در بسیاری از موارد، احتمال پوشش واقعی برای  $p$ ‌های نزدیک به صفر و یک، توسط فاصله مذکور

پارامتر، شبیه به فاصله ویلسون است و به سطح اسمی کاملاً نزدیک است (۳). هم‌چنین مطالعه‌ای در سال‌های اخیر انجام شده که از دو تبدیل لوجیت و آرک سینوس مکرر استفاده کرده است (۷). تعداد دیگری از فواصل تقریبی نیز وجود دارند، که در این مقاله به آن‌ها اشاره نشده است (۱۰-۸).

فاصله اطمینان‌های دقیق، برای همه مقادیر ممکن پارامتر، احتمال پوششی بزرگ‌تر یا برابر با سطح اسمی دارند (۱۱). فاصله اطمینان دقیق به محافظه‌کاری و پهنای زیاد، گرایش دارد و ضریب اطمینان در این روش قویاً بزرگ‌تر از سطح اسمی است. در روش‌های تقریبی فاصله‌ها کوتاه‌تر هستند، اما احتمال پوششی آن‌ها برای همه مقادیر پارامتر بالاتر از سطح اسمی نیست (۳). روش‌های دقیقی توسط استرن در سال ۱۹۵۴ میلادی (۱۲)، کرو در سال ۱۹۵۶ (۱۳)، و بلیس و استیل در سال ۱۹۸۳ (۱۱) پیشنهاد شد، اما این فواصل به سبب پیچیدگی در محاسبه‌ها پذیرفته نشدند. در سال ۱۹۳۴ میلادی کلوپر و پیرسون روشی معرفی کردند، که در بیش‌تر نوشته‌های پزشکی، از آن به عنوان یک فاصله اطمینان دقیق برای یک نسبت دوجمله‌ای استفاده می‌شود (۱۴). چندین دلیل برای استفاده وسیع‌تر فاصله دقیق فاصله کلوپر-پیرسون نسبت به سایر فواصل دقیق وجود دارد. یکی این‌که این روش رایج‌تر است و به سادگی در دسترس است. روش حل این فاصله در کتاب‌های قدیمی آماری پیدا می‌شود و تقریباً در همه‌ی نرم‌افزارهای آماری برنامه آن وجود دارد و اجرا می‌شود. دیگر فواصل دقیق که کوتاه‌تر از فاصله کلوپر-پیرسون هستند؛ احتمال پوششی این فواصل به کم‌تر از  $1-\alpha$  نوسان می‌کند (۲).

مرور مطالعه‌های پیشین که در این زمینه انجام شده است؛ نیز درستی استفاده از فاصله‌های دیگر را به جای استفاده از فاصله والد به دلیل محدودیت‌های این روش، به اثبات می‌رساند (۱۹-۱۵، ۲۰، ۱).

در هنگام انتخاب بین دو روش دقیق و تقریبی، مهم است تا از منافع و هزینه‌های مربوط به دو نوع روش آگاهی وجود داشته باشد. همان‌طور که در پاراگراف‌های پیشین مشاهده می‌شود؛ نوسان‌های پوششی فاصله‌های تقریبی در چندین مطالعه مقایسه شده است. در این مطالعه تلاش شد تا ارزش استفاده از فواصل دقیق به صورت صریح و روشن مشخص شود. ارزش استفاده از فاصله کلوپر-پیرسون به جای این فاصله‌ها این است که از نظر طول مورد انتظار و حجم نمونه مورد نیاز به  $p$  و  $\alpha$  غیر حساس است.

<sup>۱</sup>Immunohistochemistry; IHC

<sup>۲</sup>Fluorescence in situ hybridization; FISH

است و  $(n, \theta)$  متغیر تصادفی دوجمله‌ای با  $n$  آزمایش و احتمال موفقیت  $\theta$  است. به دلیل رابطه‌ای که بین توزیع تجمعی دو جمله‌ای و توزیع بتا وجود دارد، گاهی اوقات فاصله کلپر-پیرسون با استفاده از چارک‌های توزیع بتا نوشته می‌شود:

$$(P_L, P_U) = (\text{Beta}(\alpha/2, x, n-x+1), \text{Beta}(1-\alpha/2, x+1, n-x))$$

که  $x$  تعداد موفقیت‌ها،  $n$  تعداد آزمایش‌ها و  $\text{Beta}(p; v, w)$  - $p$  امین چارک از توزیع بتا با پارامترهای شکل  $v$  و  $w$  است. وقتی  $x$  هریک از مقادیر  $0$  یا  $n$  را بگیرد، حدود فاصله قابل محاسبه است. برای  $x=0$  و  $x=n$  کران فاصله اطمینان برابر یک می‌شود (۲). فاصله‌ی اطمینان کلپر-پیرسون احتمال پوشش واقعی را برابر یا بالای سطح اطمینان اسمی تضمین می‌کند، حتی برای  $n$ ‌های بسیار بزرگ (۳).

دو بل آرک سینوس مکرر: برای جلوگیری از به دست آمدن حدودی که خارج از محدوده  $[0, 1]$  باشد و واریانسی که تثبیت شده نیست، از تبدیلی روی نسبت استفاده می‌شود که تقریبی از توزیع نرمال است. این تبدیل به صورت زیر است:

$$(t = \frac{\sin^{-1} \sqrt{x/(n+1)} + \sin^{-1} \sqrt{(x+1)/(n+1)}}{2})$$

که  $x$  تعداد موفقیت‌ها و  $n$  تعداد آزمایش‌ها است. تبدیل آرک سینوس مکرر، به صورت زیر قابل محاسبه است،

$$\{t - 1/\sin t\} / n \approx 0.5 \{t + (\sin t)\} [1 - \sin t] p = 0.5 \{1 - \text{sgn}(\cos t)\}$$

در اینجا "sgn" نشان‌گر علامت است.

سناریوی شبیه‌سازی: داده‌های لازم برای انجام شبیه‌سازی از توزیع دوجمله‌ای تولید شدند. شبیه‌سازی توزیع دوجمله‌ای با پارامترهای  $n$  و  $p$  ۳۶۰ بار و هر بار ۱۰۰۰ مرتبه انجام شد. برای این‌که تغییرات  $n$  و  $P$  را بر احتمال پوششی و خطای نوع اول مشاهده شوند،  $n$  با ۸ مقدار مختلف ۱۰، ۵۰، ۱۵۰، ۲۵۰، ۵۰۰، ۱۰۰۰، ۱۵۰۰، ۲۰۰۰ و  $P$  با مقادیر ۰/۴، ۰/۵، ۰/۶ و  $[0.01, 0.1]$  و  $[0.9, 0.999]$  که به صورت تصاعد حسابی افزایش می‌یابند؛ ۰/۰۱، ۰/۰۰۵، ۰/۰۱، ۰/۰۱۵، ۰/۰۲۰، ۰/۰۲۵، ۰/۰۳۰، ۰/۰۳۵، ۰/۰۴۰، ۰/۰۴۵، ۰/۰۵۰، ۰/۰۵۵، ۰/۰۶۰، ۰/۰۶۵، ۰/۰۷۰، ۰/۰۷۵، ۰/۰۸۰، ۰/۰۸۵، ۰/۰۹۰، ۰/۰۹۵، ۰/۱۰۰ و ۰/۹۹۹، ۰/۹۹۵، ۰/۹۹۰، ۰/۹۸۵، ۰/۹۸۰، ۰/۹۷۵، ۰/۹۷۰، ۰/۹۶۵، ۰/۹۶۰، ۰/۹۵۵، ۰/۹۵۰، ۰/۹۴۵، ۰/۹۴۰، ۰/۹۳۵، ۰/۹۳۰، ۰/۹۲۵، ۰/۹۲۰، ۰/۹۱۵، ۰/۹۱۰، ۰/۹۰۵ و ۰/۹۰۰ در نظر گرفته شد. در این مطالعه از شبیه‌سازی داده‌ها، مقادیری از  $P$  که نزدیک به حدود صفر و یک بودند، داده تولید شد. زیرا برای سایر مقادیر  $P$ ، در فواصل اطمینان‌های

ضعیف است و فقط برای مواقعی که  $(P, 1-p) \times \min$  حداقل ۵ یا ۱۰ باشد، استفاده می‌شود. در سطح اسمی ۹۵ درصد، احتمال پوششی فاصله تقریب نرمال رفتار ضعیف و نامنظمی برای  $n$ ‌های کوچک حتی وقتی که  $p$  نزدیک به صفر و یک نباشد، نشان می‌دهد (۳).

فاصله اطمینان تقریبی ویلسون:

فاصله اطمینان ویلسون بهبود یافته فاصله اطمینان والد است، احتمال پوشش واقعی فاصله اطمینان ویلسون نسبت به احتمال پوشش واقعی فاصله‌ی اطمینان والد به مقدار اسمی نزدیک‌تر است. فاصله اطمینان ویلسون با استفاده از فرمول زیر به دست می‌آید.

$$\hat{p} + \frac{1}{2n} z_{1-\alpha/2}^2 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}} \div \left( 1 + \frac{1}{n} z_{1-\alpha/2}^2 \right)$$

مرکز فاصله اطمینان ویلسون  $(\hat{p} + 1/2n z_{(1-\alpha/2)}^2) / (1 + 1/n)$  است.  $p \approx x/n$  نسبت نمونه‌ای است،  $x$  تعداد موفقیت‌ها،  $n$  اندازه نمونه و  $[z_{(1-\alpha/2)} - \Phi(\alpha/2)] / [100(1-\alpha/2, z)]$  درصد از توزیع نرمال استاندارد است (۵). فاصله ویلسون با مشتق از در رابطه زیر به دست می‌آید: آزمونی که در میان نابرابری قرار گرفته آزمون امتیاز است، بنابراین فاصله ویلسون را برخی مواقع فاصله امتیاز ویلسون می‌نامند (۵). فاصله‌ی امتیاز ویلسون برای نمونه‌های کوچکی از آزمایش‌ها یا احتمال حداکثری دارای ویژگی‌های خوبی است. احتمال پوششی فاصله اطمینان ویلسون به‌طور قابل قبولی نزدیک سطح اسمی  $1-\alpha$  در نوسان است، اما برای نسبت‌هایی که خیلی نزدیک به صفر و یک باشند؛ احتمال پوششی ضعیفی دارد (۳).

فاصله‌ی اطمینان دقیق کلپر-پیرسون یک روش قدیمی و خیلی رایج برای محاسبه‌ی فاصله اطمینان‌های دوجمله‌ای است. این فاصله اغلب روش «دقیق» نامیده می‌شود، زیرا بر مبنای احتمال تجمعی توزیع دوجمله‌ای است. فاصله‌ی کلپر-پیرسون می‌تواند به‌عنوان

$$\langle \theta | P[\text{Bin}(n, \theta) \leq X] \geq \alpha/2 \rangle \cap \langle \theta | P[\text{Bin}(n, \theta) \leq X] \geq \alpha/2 \rangle$$

نوشته شود که  $x$  تعداد موفقیت‌های مشاهده شده در نمونه

سال‌های ۲۰۰۱ تا ۲۰۱۱ استخراج شده و نسبت یافته‌های مثبت (تأیید بیان HER2 در بیماران سرطان پستان) در روش FISH نسبت به روش IHC محاسبه شد (۲۰). پس از تعیین نسبت یافته‌های مثبت در روش FISH نسبت به روش IHC، با در نظر گرفتن توزیع دو جمله‌ای برای نسبت توافق بین دو روش FISH و IHC مقادیر p نسبت دو جمله‌ای برای همه مطالعه‌ها به دست آمد.

### یافته‌ها

برای تصمیم‌گیری در مورد بهترین فاصله اطمینان در میان فاصله اطمینان‌های دقیق و تقریبی، احتمال پوششی این فواصل مقایسه شد و فاصله اطمینانی که در سطح خطای  $\alpha$ ، احتمال پوششی حداقل  $1-\alpha$  داشته باشد و برای تمام مقادیر p، به کم‌تر از  $1-\alpha$  تنزل پیدا نکند؛ فاصله اطمینان مناسبی است.

پس از محاسبه فواصل اطمینان به روش‌های والد، ویلسون و کلوپر-پیرسون و تبدیل آرک سینوس مکرر برای همه مقادیر p و n، احتمال پوششی نیز برای همه مقادیر و برای هر ۴ روش محاسبه شد. برای مشخص شدن بهترین فاصله اطمینان، در یک حالت مقدار حجم نمونه ثابت و مقادیر p متغیر قرار داده و در حالت بعدی مقدار p ثابت و حجم نمونه متغیر قرار داده شد.

همان‌طور که در شکل شماره یک مشاهده می‌شود، با تغییر p وقتی حجم نمونه (n) ثابت است بیش‌ترین احتمال پوششی مربوط به روش دقیق کلوپر-پیرسون است؛ زیرا برای همه‌ی مقادیر p این فاصله اطمینان در سطح ۰/۰۵، احتمال پوششی برابر و بیش‌تر از ۰/۹۵ دارد. در روش تبدیل آرک سینوس مکرر روند احتمال پوششی نامنظم است برای مقادیری از p که در نزدیکی به صفر هستند؛ روند نزولی و از سطح ۰/۹۵ پایین‌تر است، و فاصله‌ی اطمینان والد و ویلسون هر دو احتمال پوششی ضعیف و نامنظمی دارند (شکل‌های ۱-الف، ۱-ب، ۱-ج و ۱-د). احتمال پوششی والد با افزایش حجم نمونه از ۵۰ (شکل ۱-الف) به ۵۰۰ (شکل ۱-د) هموارتر شده، به استثنای مواردی که p به صفر یا یک خیلی نزدیک است، تقریباً  $1-\alpha$  را نشان می‌دهد. در روش تبدیل آرک سینوس مکرر نیز با افزایش حجم نمونه احتمال پوششی بهتر می‌شود، اما همچنان برای برخی از مقادیر نسبت نمونه احتمال پوششی کم‌تر از ۰/۹۵ است، اما همان‌طور که در شکل ۱-الف تا ۱-د مشاهده می‌شود، برای روش ویلسون برای همه مقادیر p، احتمال پوششی ضعیف بوده و به کم‌تر از  $1-\alpha$  تنزل پیدا می‌کند (شکل‌های ۱-الف تا ۱-د).

در این مطالعه ۴ روش یاد شده در حجم نمونه‌های ۱۰ و ۲۰۰۰

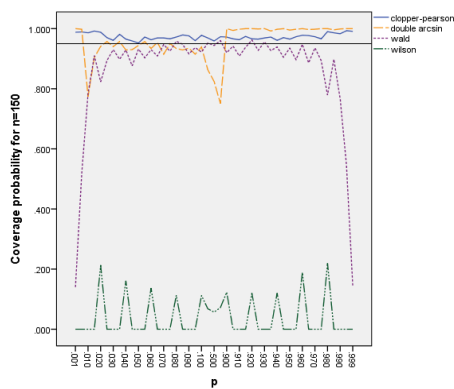
مختلف روند تقریباً یکسان است و هدف، بررسی مقادیر P نزدیک به صفر و یک است. پس برای مقادیر P بین [۰/۰۱، ۰/۱] و [۰/۹۹، ۰/۹] که دارای بیش‌ترین خطر برای ضعیف بودن احتمال پوششی هستند؛ به ۴ روش فوق فاصله اطمینان ساخته شد، البته برای صحیح‌تر و متقارن شدن شکل نمودارها برای p مقادیر ۰/۴، ۰/۵ و ۰/۶ هم در نظر گرفته شد. برای مشخص شدن تأثیر تغییرات حجم نمونه، مقادیر n نیز تغییر داده شد. P، ۴۵ مقدار دارد. ۳۶۰ ترکیب از (n, p) به دست آمد، که هر ترکیب هزارمرتبه تکرار شد. بنابراین ۳۶۰ ماتریس، که هر ماتریس هزار ردیف و n ستون با درایه‌های صفر و یک دارد، ساخته شد. برای هر ردیف از ماتریس‌ها  $\hat{P}$  به دست آمد،  $\hat{P}$  میانگین یک‌ها در هر ردیف است و در نهایت فواصل اطمینان از ۳ روش والد، ویلسون و کلوپر-پیرسون برای هر ردیف به دست آورده شد و مقایسه‌ها انجام شد. در تبدیل آرک سینوس مکرر داده‌ها از توزیع پواسن با پارامتر  $\mu=np$  تولید شدند و برای نسبت نمونه‌های مختلف و حجم نمونه‌های متفاوت شبیه‌سازی انجام شد. در این مطالعه برای شبیه‌سازی و تحلیل داده‌ها از نرم‌افزار R نسخه‌ی ۳.۰.۲ و نرم‌افزار SPSS نسخه‌ی ۲۱ استفاده شده است.

در این مطالعه برای ارزیابی این روش‌ها در یک وضع واقعی از داده‌های مربوط به مطالعه‌ی بحرینی و همکاران که در سال ۲۰۱۴ میلادی انتشار یافت، استفاده شد. مطالعه بحرینی و همکاران یک مطالعه متاآنالیز به منظور کشف ژن HER2 با استفاده از دو روش IHC و FISH است (۲۰). ژن HER2 ۲۰-۱۵ درصد تقویت کننده و افزایش دهنده سرطان پستان است (۲۱، ۲۲). بیان پروتئین HER2 به روش ایمونوهیستوشیمی (IHC) و روش گسترش ژن HER2 توسط فلورسانس در محل هیبریداسیون (FISH) نیز تشخیص داده می‌شود، که پیش‌بینی کننده قابل اعتمادتری از پاسخ بالینی به درمان سرطان پستان نسبت به IHC است (۲۳، ۲۴). بسیاری از مطالعه‌ها، توافق بین IHC و FISH را بررسی کرده‌اند (۲۷-۲۳، ۲۵). برخی از مطالعه‌ها رنگ‌آمیزی IHC با نمره ۳+ را که عمدتاً در توافق با FISH در تعیین پروتئین HER2 است در مقایسه با IHC با نمره‌ی رنگ‌آمیزی ۲+، ۱+ و ۰ در عدم تعیین پروتئین HER2 نشان می‌دهند (۲۷-۲۵). بنابراین، داده‌های مورد استفاده در این مقاله بررسی اثر IHC در مقایسه با FISH در تشخیص تغییرات ژن HER2 در سرطان پستان انسان است. در مطالعه‌ی بحرینی و همکاران، تمام مقادیر مربوط به نتایج روش‌های IHC (تعداد ۱/۰، ۲+ و ۳+) و FISH (تعداد amplification و non-amplification) در مطالعات مختلف از

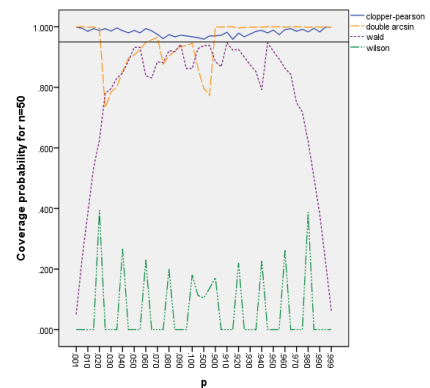
دلیل روش کلپر- پیرسون به منظور برآورد فواصل اطمینان در یک مطالعه متآنالیز به کار برده می‌شود.

جدول شماره (۱) نسبت توافق بین دو روش FISH و IHC را در یافتن ژن HER2 در بیماران مبتلا به سرطان پستان را نشان می‌دهد و همان‌طور که مشاهده می‌شود؛ مقادیر  $p$  برای برخی از مطالعه‌ها برابر با صفر یا یک شده است. پس از محاسبه فاصله اطمینان با استفاده از روش مرسوم و رایج والد فواصل اطمینان بین صفر و یک به دست نمی‌آید و برای مقادیر کوچک  $p$ ، دارای حد پایین منفی و برای مقادیر بزرگ  $p$ ، دارای حد بالای بیش‌تر از یک است، که نامعتبر هستند. بنابراین برپایه یافته‌هایی مطالعه شبیه‌سازی شده؛ مطالعه‌هایی که دارای فاصله اطمینان‌های نامعتبر هستند؛ مجدداً مورد ارزیابی قرار گرفتند.

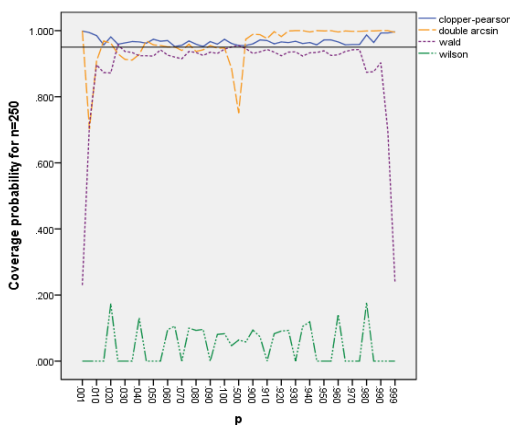
تایی (حجم نمونه متناسب با داده‌های واقعی) مجدداً شبیه‌سازی شد. همان‌طور که در شکل دو-الف مشاهده می‌شود؛ دو روش ویلسون و والد وقتی حجم نمونه بسیار کم است؛ احتمال پوششی مناسبی ندارند و روش‌های کلپر- پیرسون و تبدیل آرک سینوس مکرر در حجم نمونه کم احتمال پوششی را تقریباً ثابت نگه داشته‌اند، اما در حجم نمونه بسیار بزرگ ( $n=2000$ ) شکل دو - (ب) روش ویلسون باز هم رفتار پوششی مناسبی نداشته و تبدیل آرک سینوس مکرر نیز گودی عمیقی برای برخی از مقادیر  $p$  در احتمال پوششی خود دارد، اما روش‌های کلپر- پیرسون و والد تقریباً شبیه به هم و احتمال پوششی بالایی دارند. بر اساس یافته‌های مطالعه شبیه‌سازی شده مشاهده می‌شود که روش کلپر- پیرسون در تمامی حالات یعنی در حجم نمونه‌های مختلف ( $n$ ) و نسبت‌های مختلف ( $p$ ) احتمال پوششی را در حد  $1-\alpha$  نگه می‌دارد. به همین



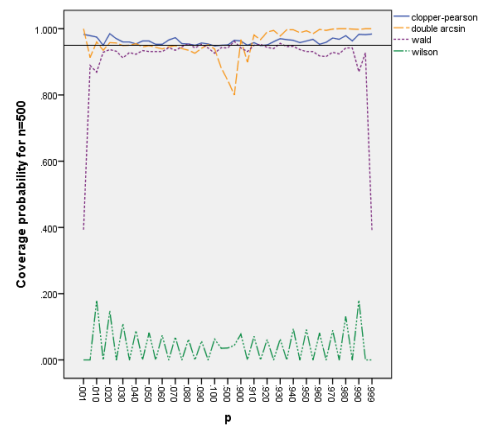
شکل ۱- الف



شکل ۱- ب

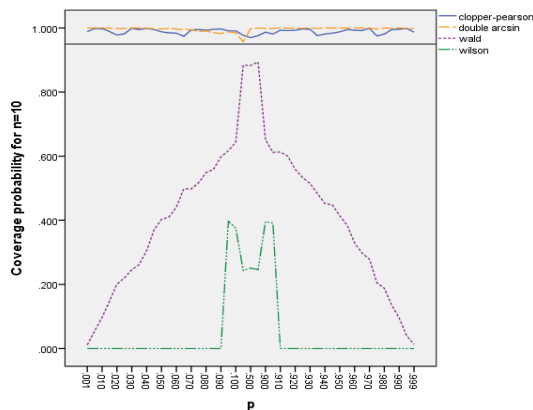


شکل ۱- ج

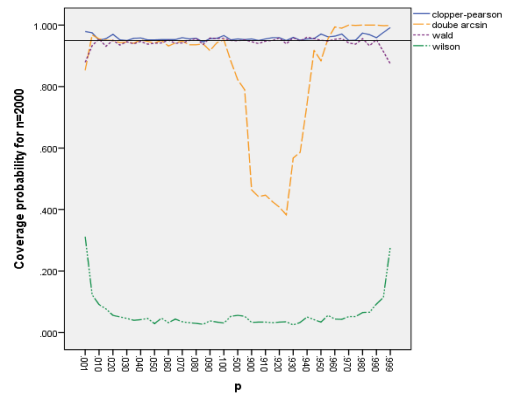


شکل ۱- د

شکل شماره ۱- نمودار احتمال پوششی ۴ روش فاصله اطمینان کلپر پیرسون، تبدیل آرک سینوس مکرر، والد و ویلسون در سطح خطای ۰/۰۵ برای تغییرات  $p$  و چهار مقدار  $n=50, 150, 250, 500$



شکل ۲- الف



شکل ۲- ب

شکل شماره ۲- نمودار احتمال پوششی برای ۴ روش فاصله اطمینان کلوپرپیرسون، تبدیل آرک سینوس مکرر، والد و ویلسون برای تغییرات  $p$  در حجم نمونه‌های  $n=10, 2000$  در سطح خطای  $0.05$

جدول شماره ۱- محاسبه فاصله اطمینان با استفاده از روش کلوپر- پیرسون در سطح خطای  $0.05$  مطالعات<sup>●</sup> روش IHC بانمره  $1+0$  در پیرسون در سطح خطای  $0.05$

| روش IHC با نمره $3+$ در مقایسه با روش FISH (+) |                | روش IHC بانمره $2+$ در مقایسه با روش FISH (+) |                | روش IHC بانمره $1+0$ در مقایسه با روش FISH (-) |                | مطالعات          |
|--|----------------|---|----------------|--|----------------|------------------|
| فاصله اطمینان کلوپرپیرسون                      | نسبت توافق (p) | فاصله اطمینان کلوپرپیرسون                     | نسبت توافق (p) | فاصله اطمینان کلوپرپیرسون                      | نسبت توافق (p) |                  |
| (0/814-1)                                      | ۱              | (0/094-0/991)                                 | 0/666          | -  | -              | Lebeau et al.    |
| (0/781-1)                                      | ۱              | (0/118-0/881)                                 | 0/5            | -  | -              | Lebeau et al.    |
| (0/814-1)                                      | ۱              | (0/028-0/600)                                 | 0/222          | -  | -              | Lebeau et al.    |
| (0/814-1)                                      | ۱              | (0/016-0/404)                                 | 0/133          | -  | -              | Lebeau et al.    |
| (0/715-1)                                      | ۱              | (0/482-0/977)                                 | 0/818          | -  | -              | Lebeau et al.    |
| (0/814-1)                                      | ۱              | (0/017-0/428)                                 | 0/142          | -  | -              | Lebeau et al.    |
| (0/913-1)                                      | ۱              | (0/276-0/578)                                 | 0/422          | (0/875-0/974)                                  | 0/937          | McCormick et al. |
| (0/876-0/978)                                  | 0/941          | (0/343-0/621)                                 | 0/481          | (0/973-0/999)                                  | 0/992          | Dowsett et al.   |
| (0/683-0/987)                                  | 0/9            | (0/145-0/569)                                 | 0/333          | (0/843-0/976)                                  | 0/929          | Ellis et al.     |
| (0/615-0/997)                                  | 0/916          | (0/001-0/722)                                 | 0/076          | (0/855-0/954)                                  | 0/914          | Ainsworth et al. |
| (0/701-0/963)                                  | 0/87           | (0-0/675)                                     | 0              | (0/942-1)                                      | ۱              | Loring et al.    |
| (0/739-0/990)                                  | 0/92           | (0/142-0/451)                                 | 0/277          | (0/968-0/999)                                  | 0/994          | Sapino et al.    |
| (0/751-0/998)                                  | 0/95           | (0/167-0/766)                                 | 0/454          | -  | -              | Bergqvist et al. |
| (0/942-1)                                      | ۱              | (0/296-0/600)                                 | 0/444          | (0/950-1)                                      | ۱              | Umemura          |
| (0/110-0/586)                                  | 0/312          | (0/036-0/709)                                 | 0/285          | (0/615-0/997)                                  | 0/916          | Gilbert et al.   |
| (0/917-1)                                      | ۱              | (0/448-0/922)                                 | 0/733          | (0/928-0/971)                                  | 0/953          | Vegt et al.      |
| (0/909-1)                                      | ۱              | (0/408-0/729)                                 | 0/833          | (0/947-0/995)                                  | 0/969          | Vegt et al.      |
| (0/630-1)                                      | ۱              | -   | -              | (0/283-0/994)                                  | 0/8            | Vogel et al.     |
| (0/735-1)                                      | ۱              | (0/004-0/641)                                 | 0/166          | (0/950-1)                                      | ۱              | Rhodes et al.    |
| (0/587-0/997)                                  | 0/91           | (0/050-0/538)                                 | 0/230          | (0/947-1)                                      | ۱              | Rhodes et al.    |
| (0/630-1)                                      | ۱              | (0/292-1)                                     | ۱              | (0/913-0/996)                                  | 0/975          | Rhodes et al.    |
| (0/715-1)                                      | ۱              | (0/031-0/650)                                 | 0/250          | (0/950-1)                                      | ۱              | Rhodes et al.    |

## بحث

۰/۵ یا ۰/۶ احتمال پوششی با افزایش حجم نمونه کاهش می‌یابد. هم‌چنین می‌توان از روش ویلسون به دلیل نامناسب بودن احتمال پوششی و عدم ثبات در تغییرات احتمال پوششی به ازای حجم نمونه‌های متفاوت و نسبت نمونه‌ای (p)، صرف‌نظر کرد.

در این مطالعه از تبدیل آرک سینوس مکرر روی نسبت نمونه‌های استفاده شد که این تبدیل برای نسبت‌های نزدیک و برابر با صفر و یک، فاصله اطمینان را بین [۰, ۱] نگه می‌دارد و دارای احتمال پوششی بالایی است اما نامنظمی‌هایی در احتمال پوششی آن مشاهده شد، که برای برخی از مقادیر نسبت نمونه روند نزولی می‌شود. البته باید گفت که این روش هنگامی که p نزدیک به صفر یا یک است؛ می‌تواند به‌عنوان یک روش مناسب به‌کار گرفته شود.

فاصله اطمینان دقیق کلورپ- پیرسون با فرض ثابت بودن حجم نمونه، برای تغییرات p آزموده شد و نتایج قابل قبول و شگرفی مشاهده شد. در این روش برای همه مقادیر p، حتی مقادیر بسیار نزدیک و برابر با صفر و یک هم احتمال پوششی برابر سطح اطمینان اسمی به‌دست آمد. روش دقیق کلورپ پیرسون را برای حالتی که حجم نمونه متغیر باشد و p ثابت فرض شود هم آزموده و بازمه نتایج قابل قبول مشاهده شد. در این حالت هم برای p نزدیک به صفر و یک با افزایش حجم نمونه به تدریج از ۱۰ به ۲۰۰۰، احتمال پوششی تغییری نکرد و به ازای همه مقادیر حجم نمونه‌ها، برابر و بزرگ‌تر از سطح اطمینان اسمی بود.

بنابراین در این مطالعه فاصله اطمینان دقیق کلورپ- پیرسون به‌عنوان بهترین روش برآورد فاصله‌ای برگزیده شد. به منظور بررسی دقیق‌تر روش‌های فوق پیشنهاد می‌شود توان آزمون هر یک محاسبه و مورد مقایسه قرار گیرد. علاوه بر این می‌توان در یک مطالعه مستقل بررسی کرد که کدام‌یک از روش‌های فوق دارای کم‌ترین واریانس است.

## نتیجه‌گیری

در این مطالعه مشاهده شد که فاصله اطمینان رایج و مرسوم والد همیشه برای استفاده در مطالعات مناسب و صحیح نیست. فاصله اطمینان والد برای مواقعی که نسبت نمونه‌ای به صفر و یک نزدیک باشد نتایج کاملاً نامعتبری به همراه دارد و حدود اطمینانی خارج از صفر و یک به‌دست می‌آورد. تبدیل آرک سینوس مکرر تنها در مواقعی که p به صفر یا به یک بسیار نزدیک است؛ از احتمال پوششی خوبی برخوردار هست، اما با این وجود می‌تواند به‌عنوان روشی مناسب اختیار شود. استفاده از فاصله اطمینان

همان‌طور که گفته شد به دست آوردن فاصله اطمینان برای p، نسبت توزیع دو جمله‌ای با استفاده از روش رایج والد دارای محدودیت‌ها و نواقصی است که در سال‌های اخیر با استفاده از سایر روش‌های دقیق و تقریبی موجود تلاش شده تا این نواقص و محدودیت‌ها از بین برود.

در این مقاله سعی شد تا به خواص پوششی نامنظم فاصله اطمینان والد پرداخته شود. با استفاده از شبیه‌سازی مونت کارلو مشاهده شد که احتمال پوششی در روش دقیق کلورپ- پیرسون، به ازای همه pها، احتمال پوششی حداقل  $1-\alpha$  است و برای همه مقادیر نسبت نمونه‌ای فاصله اطمینان در محدوده [۰, ۱] قرار می‌گیرد. در سایر مطالعه‌ها هم این موضوع و نتایج تا حد زیادی به اثبات رسیده است (۲،۹،۱۰،۱۵،۱۹،۲۸).

این مطالعه با توجه به نتایج شبیه‌سازی به این مهم دست یافت که فاصله اطمینان والد با فرض ثابت بودن حجم نمونه در نسبت‌های مختلف دارای احتمال پوششی مناسبی نیست. احتمال پوششی این فاصله اطمینان برای  $n \geq 50$  قابل قبول است به استثناء برای مقادیری از p که نزدیک و برابر با صفر و یک باشند. هم‌چنین فاصله اطمینان والد با فرض ثابت بودن p و تغییرات حجم نمونه مورد بررسی قرار گرفت و این نتیجه حاصل شد که برای مقادیری از p که نزدیک به صفر و یک هستند به تدریج با افزایش حجم نمونه از ۱۰ تا ۲۰۰۰ احتمال پوششی به سطح اسمی نزدیک می‌شود.

احتمال پوششی برای مقادیری از p که به حدود صفر و یک نزدیک نیستند؛ مثلاً برای p برابر با ۰/۴، ۰/۵ و ۰/۶ به تغییرات حجم نمونه تقریباً بی‌تفاوت است و با افزایش حجم نمونه تغییری نمی‌کند و برای  $n \geq 50$  احتمال پوششی قابل قبول و برابر با سطح اسمی دارد. پس بنابراین از فاصله اطمینان تقریبی دیگری که اصلاح شده فاصله اطمینان والد است استفاده شد، این فاصله اطمینان به روش ویلسون است. برای فاصله اطمینان ویلسون نیز تغییرات p و تغییرات حجم نمونه بررسی شد. در فاصله اطمینان ویلسون با فرض ثابت بودن حجم نمونه و تغییرات مقادیر p مشاهده شد که احتمال پوششی برای این روش برای همه مقادیر p بسیار ضعیف است و هم‌چنین برای حالتی که مقدار p ثابت فرض شود و مقادیر حجم نمونه تغییر داده شود نیز مشاهده شد که برای حالتی که p نزدیک به صفر و یک باشد. برای  $n \geq 500$  به‌صورت نوسانی افزایش می‌یابد، اما برای pهای میانی مثل ۰/۴،

این مقاله برگرفته از پایان‌نامه کارشناسی ارشد در رشته آمار زیستی بود، که نویسندگان مراتب تشکر و قدردانی خود را از حمایت مالی معاونت پژوهشی دانشگاه علوم پزشکی همدان و مرکز تحقیقات مدل سازی بیماری های غیر واگیر دانشکده بهداشت دانشگاه علوم پزشکی همدان اعلام می‌دارند...

دقیق کلپر- پیرسون به علت محاسبه حدود اطمینان صحیح و مجاز برای همه مقادیر نسبت نمونه‌ای و داشتن احتمال پوششی برابر و بزرگ‌تر از سطح اطمینان اسمی برای انتخاب فاصله اطمینان به ازای  $p$  های نزدیک به صفر یا یک است.

## تشکر و قدردانی

### منابع

- Agresti Alan, Coull Brent A. Approximate is Better than "exact" for interval estimation of binomial proportions. *The American Statistician*. 1998; 52: 119-26.
- Thulin M. The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics*. 2014; 8: 817-40.
- Lawrence D, Brown T, Cai T, DasGupta A. Interval Estimation for a Binomial Proportion. *Statistical Science*. 2001; 16: 101-33.
- Laplace PS. *Theorie Analytique des Probabilites*. Paris, Courcier. 1812.
- Wilson EB. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*. 1927; 22: 209-12.
- Berger JO. *Statistical Decision Theory and Bayesian Analysis*, 2<sup>nd</sup> ed. New York: Springer, 1985.
- Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. *Journal of epidemiology and community health*. 2013;67:974-8.
- Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*. 1998; 17: 857-72.
- Pires AM. Interval estimators for a binomial proportion: Comparison of twenty methods. *Statistical Journal*. 2008; 6: 165-97.
- Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statistical Science*. 2001; 16: 101-33.
- Blyth CR, Still HA. Binomial confidence intervals. *Journal of the American Statistical Association*. 1983; 78: 108-16.
- Sterne TH. Some remarks on confidence or fiducial limits. *Biometrika*. 1954; 275: 8-41.
- Crow EL. Confidence intervals for a proportion. *Biometrika*. 1956; 43: 423-35.
- Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26: 404-13.
- Brown LD, Cai TT, DasGupta A. Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*. 2002; 30: 160-201.
- Cai TT. One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference*. 2005; 131: 63-88.
- Vos PW, Hudson S. Problems with binomial two-sided tests and the associated confidence intervals. *Australian & New Zealand Journal of Statistics*. 2008;50:81-9.
- Staicu AM. Higher-order approximations for interval estimation in binomial settings. *Journal of Statistical Planning and Inference*. 2009; 139: 3393-404.
- Newcombe RG, Nurminen MM. In defence of score intervals for proportions and their differences. *Communications in Statistics -Theory and Methods*. 2011; 1271: 40-82.
- Bahreini F, Soltanian AR, Mehdipour P. A meta-analysis on concordance between immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) to detect HER2 gene overexpression in breast cancer. *Breast Cancer*. 2014.
- Piak S, Bryant J, Park C, Fisher B, Tan-Chiu E, Hyams D, et al. ErbB-2 and response to doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breast cancer. *J Natl Cancer Inst*. 1998; 90: 1361-70.
- Eichhorn Pieter JA, Jose' B. HER2 Signatures in breast cancer: ready to go to print? *Am Soc Clin Oncol*. 2010; 28: 1809-10.
- Pauletti G, Dandekar S, Rong H, Ramos L, Peng H, Seshadri R, et al. Assessment of methods for tissue-based detection of the HER-2/neu alteration in human breast cancer: a direct comparison of fluorescence in situ hybridization and immunohistochemistry. *J Clin Oncol*. 2000; 18: 3651-64.
- Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keit DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*. 1989; 244: 707-12.
- Dowsett M, Bartlett J, Ellis IO, Salter J, Hills M, Mallon E, et al. Correlation between immunohistochemistry (HerceptTest) and fluorescence in situ hybridization (FISH) for HER-2 in 426 breast carcinomas from 37 centers. *J Pathol*. 2003; 199: 418-23.
- Perez EA, Roche CP, Jenkins RB, Reynold CA, Halling KC, Ingle JN, et al. HER2 testing in patients with breast cancer: poor correlation between weak positivity by immunohistochemistry and gene amplification by fluorescence in situ hybridization. *Mayo Clin Proc*. 2002; 77: 148-54.
- Rhodes A, Sarson J, Assam EE, Dean SJR, Cribb EC, Parker A. The reliability of rabbit monoclonal antibodies in the immunohistochemical assessment of estrogen receptors, progesterone receptors, and HER2 in human breast carcinomas. *Am Soc Clin Pathol*. 2010; 134: 621-32.
- Newcombe RG. *Confidence Intervals for Proportions and Related Measures of Effect Size*. Chapman & Hall. 2012



# Confidence Interval Estimation of Proportion near Zero or One: A Modeling Secondary Study

Hamzeh S<sup>1</sup>, Soltanian AR<sup>2</sup>, Faradmal J<sup>2</sup>

1- MSc of Biostatistics, Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

2- Associate of Biostatistics, Department of Biostatistics, School of Public Health, Modeling of Non-Communicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

Corresponding author: Soltanian AR, soltanian@umsha.ac.ir

**Background and Objectives:** When computing a confidence interval for a binomial proportion  $p$ , one must choose an exact interval that has a coverage probability of at least  $1-\alpha$  for all values of  $p$ . In this study, we compared the confidence intervals of Clopper-Pearson, Wald, Wilson, and double ArcSin transformation in terms of maintaining a constant nominal type I error.

**Methods:** Simulations were used to compare four methods of estimating a confidence interval, including the Clopper-Pearson, Wald, Wilson, and double ArcSin. The data were generated from the binomial and Poisson distribution with parameters  $p$ ,  $n$  and  $\mu=np$ , 1000 were produced. Type I error of each method was calculated per simulation. The above methods were used to estimate confidence intervals in a meta-analysis study.

**Results:** The results of the simulation study showed that double ArcSin keep confidence interval at  $[0,1]$ , but for some proportion has high type I error or low coverage probability. The Clopper-Pearson interval guarantees that the coverage probability is always equal to or above the nominal confidence level for any fixed  $p$ .

**Conclusion:** This study showed that confidence interval estimations the Clopper-Pearson than other methods of calculating the type I error fixed and smaller.

**Keywords:** Binomial distribution, Proportion, Exact confidence interval, Approximate confidence interval