روش‌های چندمتغیره گسسته

امین روشنی

گروه آمار - دانشگاه لرستان

پاییز ۱۴۰۴

منابع درس

فهرست منابع اصلی درس روش‌های چندمتغیره گسسته به شرح زیر می‌باشد:

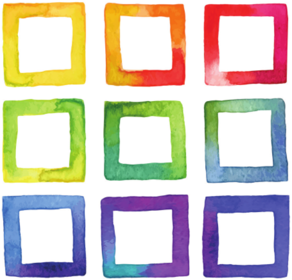
- Agresti, A. *An Introduction to Categorical Data Analysis*, 3^{th} Edition, Wiley, 2019.
- Agresti, A. *Foundations of Linear and Generalized Linear Models*, Wiley, 2015.
- Bilder, C.R. and Loughi, T.M. *Analysis of Categorical Data with R*, CRC Press, 2014.
- Bishop, Y.M.M. Fienberg, S.E. and Holland, P. W. *Discrete Multivariate Analysis*, Springer, 2007.
- Stokes, M.E. Davis, C.A. and Koch, G.G. *Categorical Data Analysis Using SAS*, 3rd Edition, SAS Institute, 2012.

- گنجعلی، م. و رضایی قهرودی، ز.، تحلیل چندمتغیره گسسته در مطالعات طولی و مقطعی، پژوهشکده آمار، ۱۳۸۹.
- شاهکار، غ. روش‌های چندمتغیره گسسته، دانشگاه پیام نور، ۱۳۸۶.

AN INTRODUCTION TO

CATEGORICAL DATA ANALYSIS

THIRD EDITION



ALAN AGRESTI

with website

WILEY

سرفصل‌های کلاس

- **فصل اول**: مقدمه
- **فصل دوم**: تحلیل جدول‌های توافقی
- **فصل سوم**: مدل‌های خطی تعمیم‌یافته
- **فصل چهارم**: رگرسیون لوجستیک
- **فصل پنجم**: مدل‌های لُگ-خطی

فصل اول: مقدمه

داده‌های با پاسخ رسته‌ای

متغیر رسته‌ای

یک متغیر رسته‌ای، رده‌ای یا طبقه‌ای (Categorical Variable) نوعی متغیر است که مقادیر آن در قالب گروه‌ها یا رده‌های مشخصی تعریف می‌شوند.

- **ایدئولوژي سیاسی:** آزادی‌خواه، میانه‌رو یا محافظه‌کار
- **محل زندگی:** خانه، آپارتمان یا واحد مسکونی (condominium)
- **نوع ایمیل:** هرزنامه (spam) یا قانونی (legitimate)

متغیرهای رسته‌ای اغلب به عنوان متغیر کیفی (Qualitative) شناخته می‌شوند تا آن‌ها را از متغیرهای کمی (Quantitative) مانند درآمد و وزن، متمایز کرد.

کاربرد متغیرهای رسته‌ای در علوم مختلف

- **علوم اجتماعی:** سنجش نگرش‌ها و نظرات با رده‌هایی مانند موافق و مخالف؛ بله و خیر؛ موافق، مخالف و نامشخص (دو دِل)
- **علوم بهداشتی و پزشکی:** موفقیت آمیز بودن درمان پزشکی (بله، خیر)، تشخیص سرطان سینه (طبیعی، خوش‌خیم، احتمالاً خوش‌خیم، مشکوک به بدخیم، بدخیم)
- **علوم رفتاری:** تشخیص نوع بیماری روانی (اسکیزوفرنی، افسردگی، اختلال روانی)
- **اکولوژی (بوم‌شناسی):** کاربری زمین در تصاویر ماهواره‌ای (جنگل، باتلاق، علفزار، کشاورزی، شهری)
- **بازاریابی:** تلفن همراه مصرف کننده (سامسونگ، اپل، نوکیا، ال‌جی)

متغیرهای پاسخ و متغیرهای توضیحی

در اکثر تحلیل‌های آماری بین متغیر پاسخ (Response Variable) و متغیرهای توضیحی (Explanatory Variables) تفاوت‌هایی وجود دارد. به عنوان مثال در مدل‌های رگرسیون معمولی، چگونگی تغییر در میانگین یک متغیر پاسخ کمی مانند درآمد سالانه را با توجه به سطوح متغیرهای توضیحی مانند سابقه تحصیلی و سابقه شغلی توصیف می‌کنند. متغیر پاسخ گاهی متغیر وابسته و متغیر توضیحی گاهی متغیر مستقل نامیده می‌شود.

هنگامی که می‌خواهیم تاکید کنیم متغیر پاسخ یک متغیر تصادفی است، از حروف بزرگ برای نمایش آن استفاده می‌شود (مانند Y). برای اشاره به یک مقدار مشاهده شده نیز ار حروف کوچک استفاده می‌شود (مانند $y = ۰$).

در این درس، مدل‌های آماری را ارائه می‌دهیم که **یک متغیر پاسخ رسته‌ای را به چند متغیر توضیحی** مرتبط می‌سازد. برای مثال فرض کنید در یک مطالعه هدف بررسی وابستگی نظرات در مورد قانونی بودن ازدواج همجنس‌گرایان (با رده‌های بله و خیر) به متغیرهای توضیحی مانند سابقه تحصیلی، درآمد سالانه، وابستگی به حزب سیاسی، وابستگی مذهبی، سن، جنسیت و نژاد باشد.

انواع متغیرهای رسته‌ای

- **دودویی (Binary):** متغیرهای رسته‌ای که تنها دو رده دارند مانند (بله و خیر) برای داشتن بیمه درمانی یا (موافق و مخالف) برای قانونی کردن ماری جوانا
- **ترتیبی (Ordinal):** متغیرهای رسته‌ای که بیش از دو رده داشته باشند و این رده‌ها یک ترتیب طبیعی داشته باشند. به عنوان مثال، فراوانی احساس اضطراب (هرگز، گاهی اوقات، اغلب، همیشه)، درد سردرد (خفیف، متوسط، شدید)
- **اسمی (Nominal):** متغیرهای رسته‌ای که بیش از دو رده داشته باشند ولی رده‌ها هیچ ترتیب طبیعی نداشته باشند. به عنوان مثال، وابستگی مذهبی (مسیحی، یهودی، مسلمان، بودایی، هندو، بی‌دین، هیچ‌کدام)، نحوه رسیدن به محل کار (خودرو شخصی، دوچرخه، اتوبوس، مترو، پیاده)

توزیع‌های احتمال برای متغیرهای رسته‌ای

همانطور که می‌دانیم در مدل‌های آماری مانند رگرسیون معمولی، برای متغیر پاسخ کمی، توزیع نرمال نقش اساسی را ایفا می‌کند. اما در مدل‌های آماری برای داده‌های رسته‌ای، در حالت کلی نمی‌توان از این توزیع استفاده کرده و باید از توزیع‌های گسسته مانند توزیع **دو جمله‌ای** (Binomial) و **چند جمله‌ای** (Multinomial) که از مهمترین و پراستفاده‌ترین توزیع‌ها در این زمینه هستند، استفاده کرد.

آزمایش برنولی: یک آزمایش تصادفی است که دو برآمد مجزا دارد. در اصطلاح برآمدی را که بیش‌تر مورد نظر است موفقیت و برآمد دیگر را شکست در نظر میگیرند. احتمال موفقیت در یک آزمایش برنولی را پارامتر آزمایش می‌نامند و با π نمایش می‌دهند (احتمال شکست نیز برابر است با $1-\pi$)

متغیر تصادفی برنولی: اگر در یک آزمایش تصادفی برنولی با پارامتر π، Y مقدارهای ۱ و ۰ را ترتیب در صورت موفقیت و شکست اختیار کند، در آن صورت Y یک متغیر تصادفی برنولی با تابع جرم احتمال به صورت زیر است

$$f(y) = \pi^y (1-\pi)^{1-y}, \qquad y = 0, 1$$

متغیر تصادفی دوجمله‌ای: اگر Y نمایانگر تعداد موفقیت در یک دنباله nتایی از آزمایش‌های مستقل برنولی با پارامتر π باشد، آن‌گاه Y یک متغیر دوجمله‌ای با تابع جرم احتمال به صورت زیر است

$$f(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \qquad y = 0, 1, \ldots, n$$

که در آن

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

مثال

فرض کنید یک امتحان با ۱۰ سوال چند جوابی داریم که هر سوال دارای ۵ گزینه است. دانشجویی که برای امتحان آماده نیست به تصادف به این سوال‌ها پاسخ می‌دهد. اگر احتمال پاسخ صحیح به هر سوال ۰٫۲ باشد و متغیر Y تعداد پاسخ‌های صحیح به این سوال‌ها باشد، آنگاه این متغیر دارای توزیع دوجمله‌ای با $n = ۱۰$ و $\pi = ۰٫۲$ خواهد بود. احتمال آنکه این دانشجو به هیچ سوالی پاسخ درست ندهد برابر است با

$$\mathbb{P}(0) = \mathbb{P}\{Y = 0\} = \frac{10!}{0!(10-0)!} 0{,}2^0 (0{,}8)^{10-0} = 0{,}8^{10} = 0{,}107$$

احتمال اینکه تنها به یک سوال پاسخ درست بدهد

$$\mathbb{P}(1) = \mathbb{P}\{Y = 1\} = \frac{10!}{1!(10-1)!} 0{,}2^1 (0{,}8)^{10-1} = 10(0{,}2) 0{,}8^9 = 0{,}268$$

جدول زیر مقدار تابع جرم احتمال توزیع دوجمله‌ای را برای تمام مقادیر ممکن $y = ۰, ۱, \ldots, ۱۰$ و مقادیر مختلف $\pi = ۰/۲, ۰/۵, ۰/۸$ نشان می‌دهد.

Table 1.1 Binomial distributions with $n = 10$ and $\pi = 0.20, 0.50$, and 0.80. The binomial distribution is symmetric when $\pi = 0.50$.

y	$P(y)$ when $\pi = 0.20$ ($\mu = 2.0, \sigma = 1.26$)	$P(y)$ when $\pi = 0.50$ ($\mu = 5.0, \sigma = 1.58$)	$P(y)$ when $\pi = 0.80$ ($\mu = 8.0, \sigma = 1.26$)
0	0.107	0.001	0.000
1	0.268	0.010	0.000
2	0.302	0.044	0.000
3	0.201	0.117	0.001
4	0.088	0.205	0.005
5	0.027	0.246	0.027
6	0.005	0.205	0.088
7	0.001	0.117	0.201
8	0.000	0.044	0.302
9	0.000	0.010	0.268
10	0.000	0.001	0.107

میانگین، واریانس و انحراف معیار توزیع دوجمله‌ای برای n آزمایش و با پارامتر π به صورت بدست می‌آید

$$\mu_Y = \mathbb{E}(Y) = n\pi, \qquad \mathbb{V}ar(Y) = n\pi(۱ - \pi), \qquad \sigma_Y = \sqrt{n\pi(۱ - \pi)}$$

به عنوان مثال، برای جدول قبل، در حالتی که $\pi = ۰/۲$ داریم

$$\mathbb{E}(Y) = ۱۰ \times ۰/۲ = ۲, \qquad \mathbb{V}ar(Y) = ۱۰ \times ۰/۲ \times ۰/۸ = ۱/۶, \qquad \sigma_Y = \sqrt{۱۰ \times ۰/۲ \times ۰/۸} = ۱/۲۶۵$$

- شکل توزیع دو جمله‌ای به ازای $\pi = ۰/۵$ متقارن است.
- برای n ثابت، هر چقدر π به ۰/۵ نزدیک‌تر باشد، شکل توزیع دو جمله‌ای زنگوله‌ای تر (مشابه توزیع نرمال) خواهد شد.
- برای π ثابت، هر چقدر n افزایش یابد، شکل توزیع دو جمله‌ای زنگوله‌ای تر خواهد بود.
- برای nهای بزرگ، می‌توان مقدار احتمال توزیع دو جمله‌ای را با استفاده از توزیع نرمال با میانگین $\mu = n\pi$ و واریانس $\sigma^۲ = n\pi(۱-\pi)$ تقریب زد. قاعده‌ای که در مورد مقدار n وجود دارد آن است که بایستی مقادیر $n\pi$ و $n(۱-\pi)$ هر دو حداقل مقدار ۵ داشته باشند. به عنوان مثال، برای $\pi = ۰/۱$ بایستی $n \geq ۵۰$ باشد.

آزمایش چندجمله‌ای: یک آزمایش تصادفی که c ($c \geq ۲$) برآمد ممکن داشته باشد را یک آزمایش c-جمله‌ای می‌نامند و به هر یک از نتایج این آزمایش چند جمله‌ای، یه رسته یا رده می‌گویند.

مثال‌هایی از آزمایش‌های تصادفی چندجمله‌ای:

- تولد یک نوزاد که نتیجه آن پسر یا دختر است ($c = ۲$).
- پرتاب یک تاس که می‌تواند یک عدد بین ۱ تا ۶ باشد ($c = ۶$).
- وضعیت تأهل و تجرد افراد بالغ یک جامعه که می‌تواند متأهل، مجرد، مطلقه یا بیوه باشد ($c = ۴$).

یک آزمایش c-جمله‌ای را در نظر بگیرید که هر برآمد آن بتواند تنها و تنها در یکی از c رسته ممکن آزمایش قرار بگیرد. احتمال تعلق هر برآمد به رده iام، ($i = ۱, ۲, \ldots, c$) را با π_i نشان می‌دهیم ($۰ < \pi_i < ۱$). اعداد $\pi_۱, \pi_۲, \ldots, \pi_c$ را **پارامترهای این آزمایش** می‌گوییم. این پارامترها در شرط $\sum_{i=۱}^{c} \pi_i = ۱$ صدق می کنند.

متغیر تصادفی چند جمله‌ای: یک آزمایش چندجمله‌ای با پارامترهای $\pi_۱, \pi_۲, \ldots, \pi_c$ را در نظر بگیرید. فرض کنید این آزمایش را در شرایط یکسان مستقلاً n بار تکرار کنیم. اگر متغیر تصادفی Y_i، $(i = ۱, ۲, \ldots, c)$، تعداد دفعاتی باشد که برآمد آزمایش در رسته i قرار می گیرد، در این حالت بردار تصادفی $\mathbf{Y} = (Y_۱, \ldots, Y_c)$ را یک متغیر تصادفی چند جمله‌ای با پارامترهای n و $\pi_۱, \pi_۲, \ldots, \pi_c$ می‌گویند.

توزیع متغیر تصادفی چندجمله‌ای: اگر $\mathbf{Y} = (Y_۱, \ldots, Y_c)$ یک متغیر تصادفی چندجمله‌ای با پارامترهای n و $\pi_۱, \pi_۲, \ldots, \pi_c$ باشد، در این صورت می‌توان ثابت نمود که:

$$\begin{aligned} f(y_۱, y_۲, \ldots, y_c) &= \mathbb{P}\{Y_۱ = y_۱, Y_۲ = y_۲, \ldots, Y_c = y_c\} \\ &= \left(\frac{n!}{y_۱! \ldots y_c!}\right) \pi_۱^{y_۱} \ldots \pi_c^{y_c}, \qquad ۰ < \pi_i < ۱, \sum_{i=۱}^{c} \pi_i = ۱, \sum_{i=۱}^{c} y_i = n \end{aligned}$$

ویژگی‌های توزیع چندجمله‌ای

اگر $\mathbf{Y} = (Y_۱, \dots, Y_c)$ یک متغیر تصادفی چندجمله‌ای باشد، در این صورت:

- با فرض $c = ۲$، $\pi_۱ = \pi$، $\pi_۲ = ۱ - \pi$ توزیع دو جمله‌ای حالت خاصی از توزیع چند جمله‌ای است.
- هر Y_i خود یک متغیر تصادفی دوجمله‌ای با پارامترهای n و π_i است. در نتیجه

$$\mathbb{E}(Y_i) = n\pi_i, \quad \mathbb{V}ar(Y_i) = n\pi_i(۱ - \pi_i)$$

- Y_iها به هم وابسته‌اند.
- تابع مولد گشتاورهای توأم Y_iها چنین است

$$M(t_۱, t_۲, \dots, t_c) = (\pi_۱ e^{t_۱} + \dots + \pi_c e^{t_c})^n$$

به علاوه

$$\mathbb{C}ov(Y_i, Y_j) = -n\pi_i\pi_j \quad , \quad \rho(Y_i, Y_j) = -\sqrt{\frac{\pi_i\pi_j}{(۱ - \pi_i)(۱ - \pi_j)}}$$

تمرین

درستی ویژگی‌های بالا را اثبات کنید.

در عمل مقدار پارامترهای توزیع دو جمله‌ای و چند جمله‌ای نامعلوم است و با استفاده از مشاهدات، این پارامترها برآورد می‌شوند. یکی از روش‌ها پر کاربرد برآوردیابی، روش حداکثر درستنمایی (MLE) است که در ادامه نحوه بکارگیری آن را برای توزیع دوجمله‌ای شرح می‌دهیم.

تابع درستنمایی

در رویکرد آمار پارامتری، خانواده‌ای از توزیع‌های احتمال با پارامتر نامعلوم برای متغیر پاسخ در نظر گرفته می‌شود. برای یک خانواده توزیع خاص، می‌توانیم داده‌های مشاهده شده را در رابطه تابع جرم احتمال قرار دهیم و سپس بررسی کنیم که چگونه مقدار چگالی احتمال به مقدار پارامتر(ها) نامعلوم وابسته است.

احتمال داده‌های مشاهده شده که به عنوان تابعی از پارامتر نامعلوم بیان می‌شود، در اصطلاح تابع درستنمایی (Likelihood Function) نامیده می‌شود.

مثال سوال‌های چندگزینه‌ای

در مثال سوال‌های چندگزینه‌ای، فرض کنید $y = ۰$ مشاهده شود. در این حالت تابع درستنمایی به صورت زیر تعریف می‌شود

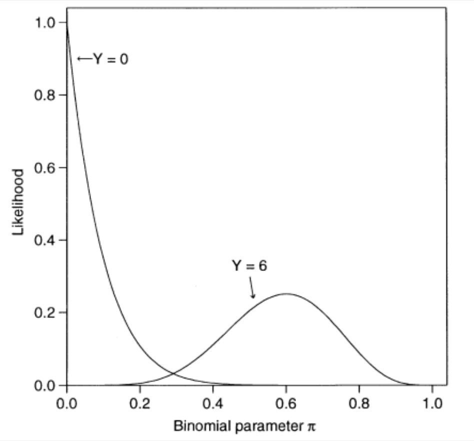
$$\ell(\pi) = \mathbb{P}(۰) = \frac{۱۰!}{۰!(۱۰)!}\pi^{۰}(۱-\pi)^{۱۰-۰} = (۱-\pi)^{۱۰}, \quad ۰ \le \pi \le ۱$$

برآورد حداکثر درستنمایی

برآورد حداکثر درستنمایی یک پارامتر، مقدار پارامتری است که به ازای آن، تابع درستنمایی **بیشترین** مقدار خود را اختیار کند. یعنی، مقدار پارامتری است که به ازای آن، احتمال مشاهده داده‌ها (داده‌های در دسترس) بزرگترین مقدار خود را داشته باشد.

برای مثال قبل، اگر $\pi = ۰/۴, ۰/۲, ۰$ باشد،

$$\ell(۰/۴) = (۱-۰/۴)^{۱۰} = ۰/۰۰۶$$
$$\ell(۰/۲) = (۱-۰/۲)^{۱۰} = ۰/۱۰۷$$
$$\ell(۰) = (۱-۰)^{۱۰} = ۱$$



شکل بالا نشان می‌دهد که برای $y = 0$ تابع درستنمایی بیشترین مقدار خود را در $\pi = 0$ اختیار می‌کند. بنابراین زمانی که $n = 10$ آزمایش $y = 0$ موفقیت دارد، برآورد حداکثر درستنمایی پارامتر π که با $\hat{\pi}$ نمایش داده می‌شود، برابر با صفر است.

یعنی $y = 0$ موفقیت در $n = 10$ آزمایش، زمانی که $\pi = 0$ باشد بیشتر است از زمانی که π برابر با هر مقدار دیگری باشد، رخ می‌دهد.

در حالت کلی برآورد ML پارامتر π در توزیع دو جمله‌ای از رابطه $\pi = \frac{y}{n}$ بدست می‌آید که در واقع همان **نسبت نمونه‌ای** برای n آزمایش است.

در مثال قبل اگر $y = ۶$ موفقیت در $n = ۱۰$ آزمایش مشاهده کنیم، برآورد ML پارامتر π برابر است با $\hat{\pi} = \frac{۶}{۱۰} = ۰/۶$.
یعنی، مشاهده مقدار $y = ۶$ در $n = ۱۰$ آزمایش، زمانی که $\pi = ۰/۶$ باشد بیشتر از زمانی که π برابر هر مقدار دیگری باشد، رخ می‌دهد.

تا قبل از مشاهده داده‌ها، مقدار برآورد حداکثر درستنمایی نامعلوم است. بنابراین یک متغیر تصادفی است و دارای یک توزیع احتمال است. در اصطلاح به آن برآوردگر گفته می‌شود و مقدار آن به ازای داده‌های مشاهده شده یک برآورد نامیده می‌شود.

برآوردگرهای حداکثر درستنمایی جزو محبوب‌ترین برآوردگرها هستند، زیرا رفتار بزرگ نمونه‌ای خوبی دارند. توزیع نمونه‌ای برآوردگر حداکثر درستنمایی تقریباً نرمال است.

- قضیه حد مرکزی: توزیع نمونه‌ای نسبت نمونه‌ای $\hat{\pi}$ برای nهای بزرگ تقریباً نرمال است.
- قانون اعداد بزرگ: با افزایش n، $\hat{\pi}$ به سمت نسبت جامعه (یعنی π) میل می‌کند.

آزمون معناداری در مورد پارامتر توزیع دو جمله‌ای

توزیع نمونه‌ای برآوردگر $\hat{\pi}$ دارای میانگین و انحراف استاندارد (Standard Error) به صورت زیر است

$$\mathbb{E}(\hat{\pi}) = \pi, \qquad SE = \sigma(\hat{\pi}) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

فرض صفر (Null Hypothesis)

$$H_0 \; : \; \pi = \pi_0$$

را در نظر بگیرید که در آن π_0 یک مقدار ثابت مانند ۰٫۵ است. تحت فرض صفر، خطای استاندارد $\hat{\pi}$ برابر است با

$$SE_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

که به آن خطای استاندارد صفر گویند.

در این حالت آماره آزمون به صورت زیر تعریف می‌شود

$$z = \frac{\hat{\pi} - \pi_0}{SE_0} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

برای nهای بزرگ، توزیع نمونه‌ای این آماره، نرمال استاندارد است. برای آزمون دوطرفه

$$\begin{cases} H_0 \ : \ \pi = \pi_0 \\ H_a \ : \ \pi \neq \pi_0 \end{cases}$$

فرض صفر رد می‌شود اگر $|z| > Z_{1-\frac{\alpha}{2}}$ که در آن $Z_{1-\frac{\alpha}{2}}$، چندک توزیع نرمال استاندارد است.

مثال: بررسی نظرات درباره سقط جنین قانونی

آیا اکثریت یا اقلیت افراد بزرگسال در ایالات متحده آمریکا معتقدند که یک زن باردار می‌تواند سقط جنین کند؟

فرض کنید π نشان دهنده نسبت جمعیت بزرگسالان آمریکا باشد که به این سوال پاسخ مثبت می‌دهد. می‌خواهیم فرض $\begin{cases} H_۰ \ : \ \pi = ۰/۵ \\ H_a \ : \ \pi \neq ۰/۵ \end{cases}$ را آزمون کنیم. اگر از میان ۱۸۱۰ پاسخ‌دهنده به این سوال، ۸۳۷ نفر پاسخ مثبت و ۹۷۳ نفر پاسخ منفی داده باشند، نسبت نمونه‌ای پاسخ‌های مثبت $\hat{\pi} = \frac{۸۳۷}{۱۸۱۰} = ۰/۴۶۲۴$ است.

آماره آزمون

$$z = \frac{\hat{\pi} - \pi_۰}{\sqrt{\frac{\pi_۰(۱-\pi_۰)}{n}}} = \frac{۰/۴۶۲۴ - ۰/۵}{\sqrt{\frac{۰/۵(۱-۰/۵)}{۱۸۱۰}}} = -۳/۲۰$$

با فرض $\alpha = ۰/۰۵$، داریم $Z_{۰/۹۷۵} = ۱/۹۶$. بنابراین چون $|z| > ۱/۹۶$ فرض صفر رد می‌شود.

محاسبه مقدار P-Value

$$\mathbb{P}(|Z| > |z|) = \mathbb{P}(|Z| > ۳٫۲۰) = \mathbb{P}(Z < -۳٫۲۰) + \mathbb{P}(Z > ۳٫۲۰) = ۰٫۰۰۱۳۷۴ \approx ۰٫۰۰۱۴ \quad < ۰٫۰۵$$

پس شواهد بسیار قوی وجود دارد که $\pi < ۰٫۵$، یعنی کمتر از نیمی از آمریکایی‌ها طرفدار سقط جنین قانونی هستند.

مقدار P-Value را می‌توان از روابط زیر نیز بدست آورد

$$\begin{aligned} ۲\min\{\mathbb{P}(Z > z), \mathbb{P}(Z < z)\} &= ۲\min\{\mathbb{P}(Z > -۳٫۲۰), \mathbb{P}(Z < -۳٫۲۰)\} \\ &= ۲(۰٫۰۰۰۶۸۷) = ۰٫۰۰۱۳۷۴ \simeq ۰٫۰۰۱۴ \quad < ۰٫۰۵ \end{aligned}$$

$$۲\,\mathbb{P}(Z > |z|) = ۲\mathbb{P}(Z > ۳٫۲) = ۲(۰٫۰۰۰۶۸۷) = ۰٫۰۰۱۳۷۴ \simeq ۰٫۰۰۱۴ \quad < ۰٫۰۵$$

فاصله اطمینان برای پارامتر π

یک آزمون معناداری صرفاً نشان می‌دهد که آیا یک مقدار خاص برای پارامتر قابل قبول است یا خیر؟ اما با استفاده از فاصله اطمینان برای تعیین محدوده مقادیر قابل قبول، اطلاعات بیشتری در مورد پارامتر بدست می‌آوریم.

فرض کنید $SE = \sqrt{\frac{\hat{\pi}(۱-\hat{\pi})}{n}}$ خطای استاندارد برآورد شده برای $\hat{\pi}$ باشد. این رابطه در واقع با جایگزاری مقدار برآورد حداکثر درستنمایی پارامتر (یعنی $\hat{\pi}$) در رابطه $\sqrt{\frac{\pi(۱-\pi)}{n}}$ بدست آمده است. یک فاصله اطمینان $۱۰۰(۱-\alpha)٪$ برای π از رابطه زیر بدست می‌آید

$$\hat{\pi} \pm Z_{۱-\frac{\alpha}{۲}} \mathrm{SE}, \qquad \mathrm{SE} = \sqrt{\frac{\hat{\pi}(۱-\hat{\pi})}{n}}$$

این نوع فاصله اطمینان را، فاصله اطمینان **والد** (Wald) می‌نامند.

در مثال سقط جنین قانونی، با $\hat{\pi} = ۰/۴۶۲$ و $n = ۱۸۱۰$، یک فاصله اطمینان ٪۹۵ عبارت است با

$$۰/۴۶۲ \pm ۱/۹۶\sqrt{\frac{۰/۴۶۲(۱-۰/۴۶۲)}{۱۸۱۰}} \Rightarrow \quad ۰/۴۶۲ \pm ۰/۰۲۳ \Rightarrow \quad (۰/۴۳۹, ۰/۴۸۵)$$

پس ما می‌توانیم ۹۵ درصد مطمئن باشیم نسبت جمعیت آمریکایی‌هایی که طرفدار سقط جنین قانونی بوده‌اند بین ۰/۴۳۹ تا ۰/۴۸۵ است.

- رابطه‌ای که در بخش قبل برای پیدا کردن فاصله اطمینان پارامتر π ذکر شد، با اینکه ساده است اما وقتی π نزدیک صفر یا یک باشد عملکرد ضعیفی دارد، مگر اینکه n خیلی بزرگ باشد.
- احتمال پوشش (Coverage Probability) واقعی آن، یعنی احتمال اینکه این روش بازه‌ای ایجاد کند که مقدار پارامتر واقعی را در بر گیرد، ممکن است بسیار کمتر از مقدار اسمی (مثلا ۰٫۹۵) باشد.

یک راه بهتر برای ساخت فواصل اطمینان، استفاده از رابطه آماره آزمون

$$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

است. این فاصله اطمینان را، فاصله اطمینان امتیاز (Score Confidence Interval) یا ویلسون (Wilson) می‌نامند.

یک فاصله اطمینان ۹۵٪ شامل تمام مقادیر $\pi_\circ$ است که به ازای آن P-Value آزمون دوطرفه بیشتر از ۰٫۰۵ باشد. به عبارت دیگر این فاصله اطمینان شامل تمام مقادیری از $\pi_\circ$ است که در سطح معناداری ۰٫۰۵ فرض صفر آزمون دو طرفه **پذیرفته** می‌شود. این مقادیر همان مقادیری هستند که به ازای آن‌ها قدرمطلق آماره آزمون کمتر از $Z_{1-\frac{\alpha}{2}}$ است ($|z| < 1{,}96 \Rightarrow -1{,}96 < z < 1{,}96$).

مثال

فرض کنید که یک کارآزمایی بالینی، برای ارزیابی یک درمان جدید، ۹ موفقیت در ۱۰ کارآزمایی داشته باشد. برای نسبت نمونه‌ای $\hat{\pi} = \frac{9}{10} = 0{,}9$ و $n = 10$، مقدار آماره آزمون به ازای $\pi_\circ = 0{,}596$ برابر است با ۱٫۹۶

$$z = \frac{\hat{\pi} - \pi_\circ}{\sqrt{\frac{\pi_\circ(1-\pi_\circ)}{n}}} = \frac{0{,}9 - 0{,}596}{\sqrt{\frac{0{,}596(1-0{,}596)}{10}}} = 1{,}96$$

مقدار آماره آزمون به ازای $\pi_0 = ۰/۹۸۲$ برابر است با $-۱/۹۶$

$$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{۰/۹ - ۰/۹۸۲}{\sqrt{\frac{۰/۹۸۲(۱-۰/۰۹۸۲)}{۱۰}}} = -۱/۹۶$$

همه مقادیر π_0 بین ۰/۵۹۶ و ۰/۹۸۲ مقادیر $|z| < ۱/۹۶$ و $\mathrm{P-Value} > ۰/۰۵$ را نتیجه می‌دهند.
بنابراین،فاصله اطمینان امتیاز ۹۵٪ برای π برابر است با $(۰/۵۹۶, ۰/۹۸۲)$

در حالت کلی این فاصله اطمینان از حل معادله زیر بر حسب π_0 بدست می‌آید

$$\frac{|\hat{\pi} - \pi_0|}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = ۱/۹۶$$

تمرین ۲

برای مثال کارآزمایی بالینی، فاصله اطمینان والد را محاسبه کنید.

فرض کنید در ۱۰ آزمایش بالینی، ۱۰ موفقیت رخ دهد. در این حالت $\hat{\pi} = ۱$ و در نتیجه $\mathrm{SE} = \sqrt{\frac{۱(۱-۱)}{۱۰}} = ۰$.
بنابراین فاصله اطمینان والد مقدار غیر واقعی

$$۱ \pm ۱/۹۶ \times SE \Rightarrow (۱,۱)$$

را ارائه می‌دهد. در مقابل فاصله اطمینان امتیاز، بازه $(۰/۷۲,۱)$ که مقادیر معقول‌تری دارد را ارائه می‌دهد.
با این حال پوشش واقعی فاصله امتیاز نیز زمانی که π بسیار به ۰ یا ۱ نزدیک باشد، کمی از مقدار اسمی آن (مثلا ۹۵ درصد) کمتر خواهد بود.
با اضافه کردن ۲ واحد به تعداد موفقیت‌ها و ۲ واحد به تعداد شکست‌ها (جمعا ۴ واحد به n افزوده می‌شود) **فاصله اطمینان اگرستی-کول** معرفی می‌شود. این فاصله اطمینان برای نمونه‌های کوچک و نسبت‌های نزدیک ۰ یا ۱ پوشش اطمینان بهتری نسبت به والد و حتی امتیاز دارد.

برای مثال بالا، فاصله اطمینان اگرستی-کول برابر است با

$$\hat{\pi} = \frac{۱۲}{۱۴} = ۰/۸۵۷۱ \Rightarrow \mathrm{SE} = \sqrt{\frac{۰/۸۵۷۱(۱-۰/۸۵۷۱)}{۱۴}} \Rightarrow ۰/۸۵۷۱ \pm ۱/۹۶\sqrt{۰/۰۰۸۷} \Rightarrow (۰/۶۷۳۸,۱)$$

فرض کنید β یک پارامتر دلخواه باشد (مانند پارامتر نسبت π). برای آزمون معناداری $H_0 : \beta = \beta_0$ که در آن β_0 یک عدد ثابت است، ساده‌ترین آماره آزمون استفاده از توزیع بزرگ-نمونه‌ای نرمال برآوردگر حداکثر درستنمایی پارامتر β (یعنی $\hat{\beta}$) است. این آماره آزمون به صورت زیر است

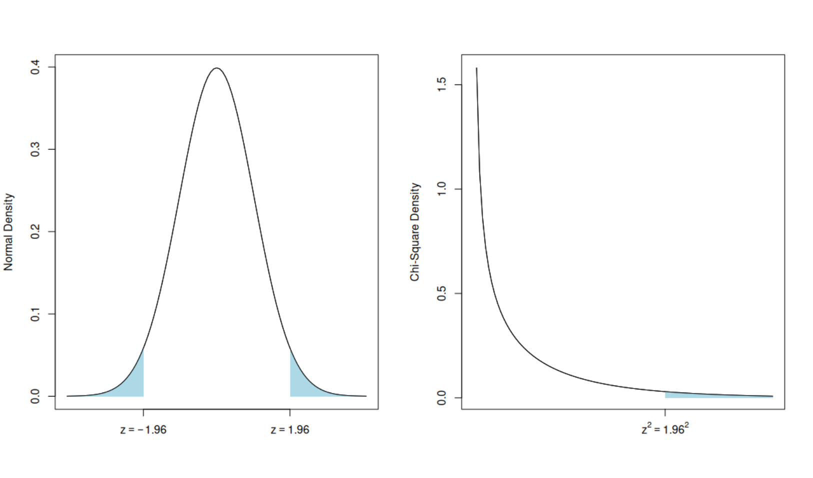
$$z = \frac{\hat{\beta} - \beta_0}{SE}$$

که در آن SE خطای استاندارد $\hat{\beta}$ است.

آزمون والد (Wald test)

اگر در آماره آزمون فوق، برای محاسبه مقدار SE از برآورد حداکثر درستنمایی پارامتر استفاده شود (به عنوان مثال در توزیع دو جمله ای $\mathrm{SE} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$) آماره z را **آماره والد** می‌نامند. تحت فرض صفر، توزیع این آماره تقریباً نرمال است. به طور معادل z^2 تقریباً دارای توزیع کای-دو با $df = 1$ درجه آزادی است. آزمون z که ازین آماره آزمون استفاده می‌کند یا به طور معادل آزمون کای-دو که از z^2 استفاده می‌کند، **آزمون والد** نامیده می‌شود.

برای به دست آوردن P-Value در آزمون دو طرفه $H_a : \beta \neq \beta_0$، می‌توان از احتمال دم‌های توزیع نرمال و یا دم سمت راست نقطه z^2 در توزیع کای-دو با $df = 1$ درجه آزادی استفاده کرد.



آزمون امتیاز (Score test)

آزمونی که از خطای استاندارد صفر $\mathrm{SE}_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$ در محاسبه آماره آزمون z استفاده کند، آزمون امتیاز (نمره) نامیده می‌شود.

آزمون نسبت درستنمایی (Likelihood-Ratio test)

آماره آزمون نسبت درستنمایی از رابطه زیر محاسبه می‌شود

$$\mathrm{LR} = 2\log\left(\frac{\ell_1}{\ell_0}\right)$$

که در آن ℓ_0 مقدار تابع درستنمایی به ازای فرض صفر است و ℓ_1 مقدار ماکسیمم تابع درستنمایی روی کل فضای پارامتر است. تحت فرض صفر، این آماره آزمون دارای توزیع بزرگ-نمونه‌ای کای-دو با $df = 1$ درجه آزادی است.

فرض صفر در آزمون نسبت درستنمایی رد می‌شود اگر مقدار آماره LR بزرگتر از چندک توزیع کای-دو در نقطه $1-\alpha$ با یک درجه آزادی باشد.

- برای مدل‌های رگرسیون معمولی که توزیع متغیر Y نرمال فرض می‌شود، هر سه آزمون والد، امتیاز و نسبت درستنمایی، آماره آزمون و $\mathrm{P\text{-}Value}$ یکسان تولید می‌کنند.
- در سایر مدل‌های آماری، زمانی که n بزرگ باشد و فرض صفر درست باشد، این سه آزمون رفتار یکسانی دارند.
- زمانی که n کوچک (تا متوسط) باشد، آزمون والد اعتبار کمتری نسبت به دو آزمون دیگر دارد.

مثال

در مثال کارآزمایی بالینی، اگر آزمون مد نظر

$$\begin{cases} H_۰ : \pi = ۰/۵ \\ H_a : \pi \neq ۰/۵ \end{cases}$$

باشد، با فرض ۹ موفقیت در ۱۰ آزمایش، داریم $\hat{\pi} = ۰/۹$

آزمون والد

$$\mathrm{SE} = \sqrt{\frac{\hat{\pi}(۱-\hat{\pi})}{n}} = \sqrt{\frac{۰/۹(۰/۱)}{۱۰}} = ۰/۰۹۵ \Rightarrow z = \frac{\hat{\pi}-\pi_۰}{SE} = \frac{۰/۹-۰/۵}{۰/۰۹۵} = ۴/۲۲$$

آماره کای-دو برابر است با $(۴/۲۲)^۲ = ۱۷/۷۸ > \chi^۲_{۰/۹۵}(۱) = ۳/۸۴$ و $\mathrm{P-Value} < ۰/۰۰۱$.

آزمون امتیاز

$$\mathrm{SE}_\circ = \sqrt{\frac{\pi_\circ(۱-\pi_\circ)}{n}} = \sqrt{\frac{۰/۵(۰/۵)}{۱۰}} = ۰/۱۵۸ \Rightarrow z = \frac{\hat{\pi}-\pi_\circ}{SE_\circ} = \frac{۰/۹-۰/۵}{۰/۱۵۸} = ۲/۵۳$$

آماره کای-دو برابر است با $(۲/۵۳)^۲ = ۶/۴۰ > \chi^۲_{۰/۹۵}(۱) = ۳/۸۴$ و $\mathrm{P-Value} = ۰/۰۱۱$.

آزمون نسبت درستنمایی

تابع درستنمایی

$$\ell(\pi) = \frac{۱۰!}{۹!۱!}\pi^۹(۱-\pi)^۱ = ۱۰\pi^۹(۱-\pi)$$

محاسبه مقادیر $\ell_\circ$ و $\ell_۱$

$$\ell_\circ = \ell(۰/۵) = ۱۰(۰/۵)^۹(۰/۵) = ۰/۰۰۹۷۷$$

$$\ell_۱ = \ell(\hat{\pi}) = \ell(۰/۹) = ۱۰(۰/۹)^۹(۰/۱) = ۰/۳۸۷۴$$

در نتیجه مقدار آماره آزمون نسبت درستنمایی برابر است با $۲\log(\ell_۱/\ell_\circ) = ۷/۳۶$. بر اساس توزیع کای-دو با یک درجه آزادی، $\mathrm{P-Value} = ۰/۰۰۷$.

استنباط کوچک-نمونه‌ای

در مثال قبل مقدار سه آماره بدست آمده با یکدیگر اختلاف زیادی داشتند. این وضعیت زمانی که n مقدار کوچکی دارد و مقدار برآورد حداکثر درستنمایی مقداری نزدیک مرز فضای پارامتر دارد، رخ می‌دهد. در این حالت می‌توان گفت که توزیع برآوردگر حداکثر درستنمایی احتمالاً از نرمال بودن دور است. در این وضعیت، روش‌های ویژه کوچک-نمونه‌ای قابل اعتمادتر هستند.

همانطور که گفته شد، در استنباط‌های مربوط به پارامتر دو جمله‌ای، آزمون‌های نسبت درستنمایی و امتیاز و همچنین فواصل اطمینانی که بر اساس این آزمون‌ها شکل می‌گیرند، زمانی معتبر هستند که $n\pi \geq ۵$ و $n(۱-\pi) \geq ۵$. در غیر این صورت بهتر است مستقیماً از توزیع دو جمله‌ای استفاده شود.

فرض کنید بخواهیم آزمون

$$\begin{cases} H_۰ : \pi = ۰/۵ \\ H_a : \pi > ۰/۵ \end{cases}$$

را برای مثال کارآزمایی بالینی با ۹ موفقیت در ۱۰ آزمایش انجام دهیم. مقدار P-Value برابر است با

$$\mathbb{P}(Y \geq ۹) = \mathbb{P}(۹) + \mathbb{P}(۱۰) = \frac{۱۰!}{۹!۱!}(۰/۵)^{۹}(۰/۵)^{۱} + \frac{۱۰!}{۱۰!۰!}(۰/۵)^{۱۰}(۰/۵)^{۰} = ۰/۰۱۱$$

برای فرض مقابل دو طرفه $H_a : \pi \neq ۰/۵$

$$\mathbb{P}(Y \geq ۹ \text{ or } Y \leq ۱) = ۲\mathbb{P}(Y \geq ۹) = ۰/۰۲۱$$

در توزیع‌های گسسته استنباط کوچک-نمونه‌ای با استفاده از P-Value مقداری محافظه‌کار است. یعنی تمایل دارد که مقدار P-Value بزرگتر باشد (در نتیجه کمتر باعث رد فرض صفر شود). در این حالت از معیار دیگری به نام mid P-Value استفاده می‌شود. در این معیار، نصف احتمال مشاهده شده استفاده می‌شود.

برای مثال قبل، در حالتی که فرض مقابل به صورت $H_a : \pi > ۰/۵$ باشد

$$\mathrm{mid\ P-Value} = \frac{\mathbb{P}(۹)}{۲} + \mathbb{P}(۱۰) = ۰/۰۰۶$$

و در حالتی که فرض مقابل به صورت $H_a : \pi \neq ۰/۵$ باشد مقدار این معیار برابر است با ۰/۰۱۲.