

آزمون کای دو (χ^2)

توزیع کای دو

آزمون کای دو یک آزمون ناپارامتریک است که توسط فیشر ارائه شده و کار اصلی آن، بررسی معناداری تفاوت بین فراوانی‌های مشاهده شده و مورد انتظار است. فرمول کلی کای دو به شکل زیر می‌باشد.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_o : فراوانی مشاهده شده

f_e : فراوانی مورد انتظار

توزیع کای دو از توزیع نرمال بدست می‌آید. اگر متغیر X متعلق به جامعه نرمال با میانگین μ و انحراف معیار σ باشد و از این جامعه نمونه تصادفی به حجم n انتخاب شود، سپس هریک از اعضای نمونه با استفاده از فرمول

$Z = \frac{X - \mu_x}{\sigma_x}$ استاندارد شوند، متغیر تصادفی زیر از توزیع کای دو با n درجه آزادی پیروی می‌کند.

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2$$

توزیع کای دو دارای رابطه ای به شکل زیر می باشد که در آن k درجه آزادی را نشان می‌دهد و میانگین و واریانس آن به ترتیب k و $2k$ می‌باشد.

$$F(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} * \frac{1}{2^{k/2}} * x^{\left(\frac{k}{2}-1\right)} * e^{-\frac{x}{2}}$$

نحوه عملکرد آزمون کای دو

جمع‌آوری اطلاعات، مقداری را برای فراوانی در سلول‌ها نشان می‌دهد. به این فراوانی‌ها، فراوانی مشاهده شده گویند. از طرفی در صورتی که طبق فرض صفر استقلال بین دو متغیر وجود داشته باشد، انتظار داریم فراوانی‌ها به شکل دیگری باشد. طبیعی است که بین این دو نوع فراوانی اختلاف وجود داشته باشد. در صورتی که اختلاف قابل توجه نباشد، می‌توانیم فرض مورد نظر را بپذیریم.

در غیر این صورت فرض صفر رد می‌شود. علت تقسیم کسر $(f_o - f_e)$ بر f_e نرمالایز کردن یا بی‌مقیاس کردن داده‌هاست.

شرایط استفاده از آزمون کای دو

- 1- تصادفی بودن داده‌ها
 - 2- مستقل بودن نمونه‌ها
 - 3- بزرگ بودن نمونه به اندازه کافی
- گفته می‌شود فراوانی سلول‌ها نباید کمتر از 10 (به عقیده بعضی‌ها 5) باشد. در صورتی که این اتفاق بیافتد باید طبقات ادغام شود. به همین دلیل بعضی از محققین اعتقاد دارند برای انجام آزمون کای دو تعداد نمونه‌ای بزرگتر از 50 مورد نیاز است. در نرم افزار SPSS در صورت وجود سلول‌هایی با فراوانی کمتر از 5 به مقدار بیش از 20٪ سلول‌ها هشدار داده می‌شود، این بدین معنی است که تست کای دو، قدرت لازم را جهت انجام آزمون ندارد.
- 4- آزمون کای دو برای متغیرهای رتبه‌ای و اسمی استفاده می‌شود.

مراحل آزمون کای دو

- 1- فراوانی‌های مورد انتظار را با استفاده از توزیع مورد آزمون تعیین کنید.
- 2- آماره χ^2 را تعیین کنید.
- 3- مقدار بحرانی را از جدول کای دو با سطح خطای مورد نظر و درجه آزادی $N-K-1$ محاسبه نمایید. که در آن N تعداد طبقات و K تعداد متغیرهای برآورد شده می‌باشد.
- 4- تصمیم‌گیری کنید.

آزمون نیکویی برازش، میزان تناسب و ارزش یکسری از داده‌ها را با یک توزیع خاص مثل توزیع نرمال یا پواسون بررسی می‌کند. در صورتی که داده‌ها از شکل توزیع فاصله زیادی نداشته باشند برازش داده‌ها با آن توزیع تعیین و در غیر این صورت رد می‌شود. روش آزمون در این جا نیز مانند 2 کاربرد دیگر آزمون کای دو است بدین معنا که با محاسبه مقادیر فراوانی مورد انتظار این مقادیر در قالب رابطه توزیع χ^2 با فراوانی‌های واقعی یا مشاهده شده مقایسه می‌شود و در نهایت با مقایسه آماره حاصل با حدود بحرانی برازش داده‌ها مشخص می‌شود.

مثال: فرض کنید برای بررسی سالم بودن یک تاس 300 بار آن را پرتاب کرده که توزیع نتایج این

300 بار بین اعداد 1 تا 6 به صورت جدول زیر است. (فقط قسمت سفید رنگ جدول داده مسئله است

بقیه جدول را باید محاسبه کنیم)

عدد \ فراوانی	1	2	3	4	5	6	جمع
F_o	45	55	50	62	40	45	300
F_e	50	50	50	50	50	50	300
$\frac{(F_o - F_e)^2}{F_e}$	0/5	0/5	0	2/88	2	0/5	6/38

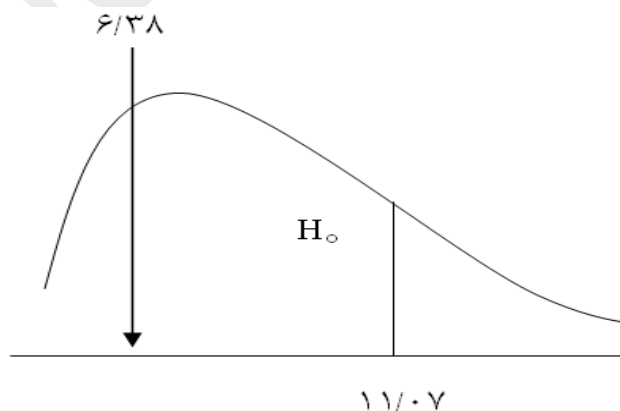
این تاس در صورتی سالم است که دارای توزیع یکنواخت باشد. یعنی فراوانی تمام اعداد یکسان باشد.

$$1- \text{توزیع یکنواخت} = \frac{300}{6} = 50$$

مفروضات:

- 2- H_o = تاس سالم است = توزیع داده‌ها یکنواخت است
- 3- H_1 = تاس سالم نیست = توزیع داده‌ها یکنواخت نیست

تعداد طبقات $N =$ تعداد پارامترهای برآورد شده $K =$ $df = N - 1 - K$



بنابراین فرض صفر مبنی بر یکنواخت بودن داده‌ها رد نمی‌شود.