

۱) آزمون فرض داده کمی

h.farhadi 17 @ yahoo.com

۱- مقایسه نسبت متغیر وابسته کیفی در دو گروه مستقل

۲- مقایسه نسبت متغیر وابسته کیفی در دو گروه وابسته

۳- مقایسه نسبت متغیر وابسته کیفی در چند گروه مستقل

آزمون فرضیه برای داده های کیفی

## مقایسه نسبت متغیر وابسته کیفی در دو گروه وابسته (آزمون مک نمار)

طرح داده ها

جدول ۸-۲

| پاسخ بعد از آموزش |   |   |
|-------------------|---|---|
| پاسخ قبل از آموزش | + | - |
| +                 | a | b |
| -                 | c | d |

جدول ۸-۱

| کنترل |   |   |
|-------|---|---|
| مورد  | + | - |
| +     | a | b |
| -     | c | d |

بین عامل در معرض و متغیر پاسخ ارتباطی وجود ندارد:  $H_0$

بین عامل در معرض و متغیر پاسخ ارتباطی وجود دارد:  $H_0$

$$n = b + c$$

$$x = b$$

$$\pi_0 = 0.5$$

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/(b + c)}} = \frac{[b/(b + c)] - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/(b + c)}} = \frac{b - c}{\sqrt{b + c}}$$

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

مقایسه با مقدار بحرانی  
توزیع نرمال

مقایسه با مقدار بحرانی توزیع کای دو

## مثال

۸-۱- در یک مطالعه مورد-شاهدی، ۳۱۷ بیمار مبتلا به سرطان داخل زهدان، با ۳۱۷ بیمار مبتلا به انواع سرطانیهای دیگر همسان شدند و استفاده از استروژن، ۶ ماه قبل از تشخیص بیماری برای آنها تعیین شده بود. نتایج در جدول ۸-۲ نشان داده شده است. با استفاده از آزمون McNemar توان دوم کای، ارتباط بین استفاده از استروژن و سرطان داخل زهدان را بیابید. فرض صفر و فرض مقابل رانیز بیان کنید.

جدول ۸-۲

| مورد               | کنترل   |              |
|--------------------|---------|--------------|
|                    | استروژن | بدون استروژن |
| استفاده از استروژن | ۳۹      | ۱۱۳          |
| بدون استروژن       | ۱۵      | ۱۵۰          |



$$Z = \frac{113 - 15}{\sqrt{113 + 15}} = 8.66$$

$$> 1.64$$

$$\alpha = 0.05$$

$$\chi^2 = \frac{(113 - 15)^2}{113 + 15} = 75.03 > 3.84$$

## موارد استفاده آزمون $\chi^2$

- آزمون استقلال در جدول های توافق (آزمون استقلال دو متغیر کیفی)
- مقایسه نسبت متغیر وابسته کیفی در دو یا چند گروه مستقل

## مفروضات:

۱. داده های نمونه بطور تصادفی انتخاب شده اند.

۲. ما قصد داریم این فرض را آزمون کنیم که برای یک جدول توافقی، متغیر سطر و متغیر ستون مستقلند.

۳. برای هر خانه جدول توافقی، فراوانی مورد انتظار ( $E_i$ ) حداقل ۵ است.

جدول ۷-۸

| $X_1$    | $X_2$ | ۱        | ۲        | ... | J        | جمع سطر  |
|----------|-------|----------|----------|-----|----------|----------|
| 1        |       | $O_{11}$ | $O_{12}$ | ... | $O_{1J}$ | $O_{1.}$ |
| 2        |       | $O_{21}$ | $O_{22}$ | ... | $O_{2J}$ | $O_{2.}$ |
| .        |       | .        | .        | ... | .        | .        |
| .        |       | .        | .        | ... | .        | .        |
| .        |       | .        | .        | ... | .        | .        |
| I        |       | $O_{I1}$ | $O_{I2}$ | ... | $O_{IJ}$ | $O_{I.}$ |
| جمع ستون |       | $O_{.2}$ | $O_{.2}$ | ... | $O_{.J}$ | $O_{..}$ |

فرضیات:  $\left. \begin{array}{l} H_0 : \text{دو متغیر سطر و ستون مستقل اند} \\ H_1 : \text{دو متغیر سطر و ستون وابسته اند.} \end{array} \right\}$

---

$o_{ij}$ : فراوانی مشاهده شده

$e_{ij}$ : فراوانی مورد انتظار

$$\pi_{ij} = \pi_{i+} \cdot \pi_{+j}$$

$$\hat{\pi}_{ij} = \hat{\pi}_{i+} \cdot \hat{\pi}_{+j} = p_{i+} \cdot p_{+j} = \frac{o_{i+}}{n} \cdot \frac{o_{+j}}{n}$$

$$e_{ij} = n\hat{\pi}_{ij} = \frac{o_{i+} \cdot o_{+j}}{n} = \frac{\text{جمع ستون}(\text{جمع سطر})}{\text{حجم نمونه}}$$

آماره آزمون عبارتست از:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

درجه آزادی مورد نظر در این حالت برابر است با:

$$df = (1 - \text{تعداد ستونها}) (1 - \text{تعداد سطرها})$$

---

ناحیه بحرانی:

فرض صفر (استقلال بین دو متغیر) رد می شود  $\Rightarrow \chi^2 > \chi^2_{1-\alpha, [(r-1)(c-1)]}$

r: تعداد سطرها

c: تعداد ستونها

## مثال:

در تحقیقی از ۱۰۰۰ مورد مرگ مردان بین سنین ۴۵ تا ۶۴ ساله، علت مرگ همراه عادت سیگار کشیدن آنها ثبت شده است.

پژوهشگر به دنبال پاسخ این سوال است که آیا براساس داده های حاصل می توان نتیجه گرفت که رابطه ای بین علت مرگ و سیگار کشیدن افراد وجود دارد یا نه؟

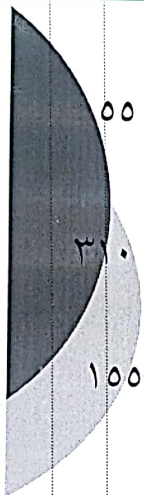
از داده های جدول زیر برای آزمون این ادعا که علت مرگ مستقل از سیگار کشیدن است استفاده می کنیم.

$\left\{ \begin{array}{l} H_0: \text{اعتیاد به سیگار و علت مرگ مستقلند} \\ H_1: \text{اعتیاد به سیگار و علت مرگ مستقل نیستند} \end{array} \right.$

|           | علت مرگ |                |            |      |
|-----------|---------|----------------|------------|------|
|           | سرطان   | بیماریهای قلبی | سایر امراض | جمع  |
| سیگاری    | ۱۳۵     | ۳۱۰            | ۲۰۵        | ۶۵۰  |
| غیرسیگاری | ۵۵      | ۱۵۵            | ۱۴۰        | ۳۵۰  |
| جمع       | ۱۹۰     | ۴۶۵            | ۳۴۵        | ۱۰۰۰ |

$$E_1 = \frac{650 \times 190}{1000} = 123.5$$

(خانه با فراوانی مشاهده شده ۱۳۵)



$$E_2 = \frac{190 \times 350}{1000} = 66.5$$

خانۀ با فراوانی مشاهده شده ۵۵

---


$$E_3 = \frac{465 \times 650}{1000} = 302.25$$

خانۀ با فراوانی مشاهده شده ۳۱۰

$$E_4 = \frac{465 \times 350}{1000} = 162.75$$

خانۀ با فراوانی مشاهده شده ۱۵۵

$$E_5 = \frac{345 \times 650}{1000} = 224.25$$

خانۀ با فراوانی مشاهده شده ۲۰۵

$$E_6 = \frac{345 \times 350}{1000} = 120.75$$

خانۀ با فراوانی مشاهده شده ۱۴۰

$$\chi^2_{1-\alpha, [(r-1)(c-1)]} = \chi^2_{(0.95, 2)} = 5.991$$

نقطه بحرانی

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - E_i)^2}{E_i} = \left[ \frac{(135 - 123.5)^2}{123.5} + \frac{(310 - 302.25)^2}{302.25} + \dots + \frac{(140 - 120.75)^2}{120.75} \right] = 8.349$$

if  $\chi^2 > \chi^2_{1-\alpha, df} \Rightarrow H_0$  رد می شود  $8.349 > 5.991 \Rightarrow H_0$  رد می شود

فرض صفر استقلال بین دو متغیر را رد می کنیم سپس به نظر می رسد که مصرف سیگار و علت مرگ وابسته اند.

## ۲. آزمون نیکویی برازش

در این آزمون به تطابق توزیع نمونه با توزیع نظری با استفاده از ملاک (کای دو) می پردازیم.

هدف ما آزمون معنی دار بودن اختلاف بین فراوانی های مشاهده شده و فراوانی هایی است که از نظر تئوری انتظار داریم. به عبارتی آزمون می کنیم که تا چه اندازه توزیع فراوانی مشاهده شده بر توزیع فراوانی نظری منطبق می شود. (یا برازنده است)

### فراوانی مشاهده شده:

منظور تعداد افرادی از نمونه که در یک گروه خاص قرار گرفته اند.

### فراوانی مورد انتظار:

فراوانی براساس قبول فرضیه صفر (تطابق نمونه با توزیع نظری) را فراوانی مورد انتظار گویند که برای محاسبه فراوانی منتظره این گروه باید احتمال مربوط به آن گروه را که از توزیع نظری براساس فرضیه صفر محاسبه می شود در تعداد مشاهدات ( $n$ ) ضرب کنیم.

فرضیات:

$H_0$  : توزیع نمونه با توزیع موردنظر تطابق دارد (مثلاً نرمال است)

$H_1$  : توزیع نمونه با توزیع موردنظر تطابق ندارد.

آماره آزمون عبارتست از:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i}$$

$n_i$ : فراوانی مشاهده شده

$e_i$ : فراوانی مورد انتظار

$k$ : تعداد رسته های مختلف یا تعداد گروههای مختلف

درجه آزادی موردنظر برای این آزمون به صورت  $df = m - k - 1$  محاسبه می شود. ( $m$  تعداد پارامترهای جامعه است.)

ناحیه بحرانی برای این آزمون عبارتست از:

$$\chi^2_{(محاسبه شده)} > \chi^2_{1-\alpha, (k-1)} \Rightarrow \text{فرضیه } H_0 \text{ رد می شود}$$

## مثال :

اطلاعات جدول زیر که مربوط به فشارخون سیستمولیک نمونه ای  
از مردان ۳۵ سال به بالای روستایی است رادرنظر بگیرید.  
اگر میانگین و انحراف معیار نمونه به ترتیب برابر ۲۵/۱۳۳ و ۲۷/۲۱  
میلی متر جیوه است .  
\_ تطابق توزیع صفت فشارخون رادراین جامعه باتوزیع نظری  
نرمال آزمون کنید .

$H_0$  : توزیع داده هانرمال است  
 $H_1$  : توزیع داده هانرمال نیست

| گروه | فشارخون<br>سیستولیک | فراوانی مشاهده<br>شده ( $n_i$ ) | احتمال مربوط<br>به هر گروه | فراوانی<br>مورد انتظار<br>( $e_i$ ) |
|------|---------------------|---------------------------------|----------------------------|-------------------------------------|
| ۱    | کمتر از ۹۰          | ۰                               | ۰.۲۱۲/۰                    | ۱۲/۲                                |
| ۲    | ۹۰-۱۲۰              | ۱۵                              | ۰.۲۴۶۴/۰                   | ۶۴/۲۴                               |
| ۳    | ۱۲۰-۱۵۰             | ۵۷                              | ۰.۵۱۴۷/۰                   | ۴۷/۵۱                               |
| ۴    | بالاتر از ۱۵۰       | ۲۸                              | ۰.۲۱۷۷/۰                   | ۷۷/۲۱                               |
| جمع  |                     | ۱۰۰                             | ۱.۰۰                       |                                     |

احتمال مربوط به گروه ۱

$$p(X < 90) = p\left(Z < \frac{90 - 133.25}{21.27}\right) =$$

$$p(Z < -2.03) = 0.5 - 0.4788 = 0.0212$$



احتمال مربوط به گروه ۲

$$p(90 < X < 120) =$$

$$p\left(\frac{90-133.25}{21.27} < Z < \frac{120-133.25}{21.27}\right) =$$

$$p(-2.03 < Z < -0.62) = 0.4788 - 0.2324 = 0.2464$$

احتمال مربوط به گروه ۴

$$p(X > 150) = p\left(Z > \frac{150-133.25}{21.27}\right) =$$

$$1 - p(Z < 0.787) = 1 - (0.5 + 0.2823) = 0.2177$$

فراوانی مورد انتظار گروه ۱

$$E_1 = np = 100 \times 0.2464 = 24/64$$

فراوانی مورد انتظار گروه ۴

$$E_4 = np = 100 \times 0.2177 = 21/77$$


$$\chi^2 = \sum_{i=1}^4 \frac{(n_i - e_i)^2}{e_i} = 2.12 + 3.77 + 0.594 + 1.78 = 8.264$$

$$df = k - 1 - m = 4 - 1 - 2 = 1$$

K: تعداد گروهها  
m: تعداد پارامترهای مستقلی است که توسط نمونه برای توزیع نظری برآورد شده است.

$$\chi_{0.99}(1) = 6.63, \quad 8.264 > \chi_{0.99}^2(1) \Rightarrow \text{reject } H_0$$

نتیجه: توزیع فشار خون در جامعه مورد مطالعه از نوع توزیع نرمال نیست.

مثال :

اطلاعات جدول زیر متضمن مطالعه ای از ۱۴۷ حادثه صنعتی است که مراقبتهای پزشکی لازم دارند.

این ادعا را آزمون کنید که حوادث در روزهای هفته به صورت زیر توزیع شده اند .

۳۰٪ در روز شنبه ، ۱۵٪ در روز یکشنبه ، ۱۵٪ در روز دوشنبه ،  
۲۰٪ در روز سه شنبه و ۲۰٪ در روز چهارشنبه .

| روزها |        |        |         |          | حوادث<br>مشاهده شده |
|-------|--------|--------|---------|----------|---------------------|
| شنبه  | یکشنبه | دوشنبه | سه شنبه | چهارشنبه |                     |
| 31    | 42     | 18     | 25      | 31       |                     |

فرض صفر این ادعاست که درصدهای بیان شده درست هستند،  
پس فرض صفر و فرض مقابل به صورت زیر است.

$$H_0: p_1=0.03, p_2=0.15, p_3=0.15, p_4=0.2, p_5=0.2$$

$H_1$ : حداقل یکی از نسبتهای قبلی مساوی مقدار ادعا شده نیست.

محاسبه فراوانیهای مورد انتظار:

$$E_1 = np = 147 \times 0.03 = 44.1 \quad \text{فراوانی مورد انتظار روز شنبه}$$

$$E_2 = np = 147 \times 0.15 = 22.05 \quad \text{فراوانی مورد انتظار روز یکشنبه}$$

▪  
▪  
▪

$$E_5 = np = 147 \times 0.2 = 29.4 \quad \text{فراوانی مورد انتظار روز چهارشنبه}$$

| روزها |        |        |         |          |                         |
|-------|--------|--------|---------|----------|-------------------------|
| شنبه  | یکشنبه | دوشنبه | سه شنبه | چهارشنبه |                         |
| 31    | 42     | 18     | 25      | 31       | تعداد حوادث مشاهده شده  |
| 44.1  | 22.05  | 22.05  | 29.4    | 29.4     | تعداد حوادث مورد انتظار |

$$\chi^2 = \sum_{i=1}^4 \frac{(n_i - e_i)^2}{e_i} = \frac{(31 - 44.1)^2}{44.1} + \frac{(42 - 22.05)^2}{22.05} + \dots + \frac{(31 - 29.4)^2}{29.4} = 23.431$$

$$\alpha = 0.05, \quad df = k - 1 = 5 - 1 = 4$$

$$\chi_{0.05}^2(4) = 9.488 \quad 23.431 > \chi_{0.05}^2(4) \Rightarrow \text{reject } H_0$$

شواهد کافی برای رد این ادعا که حوادث مطابق درصد های داده شده توزیع شده اند وجود دارد.