کاربردهای آزمون χ^2

① آزمون همبستگی بین دو متغیر کیفی:

فرض کنیم متغیر کیفی X دارای r طبقه و متغیر کیفی Y دارای c طبقه باشد به صورت جدول زیر. تعداد مشاهدات در خانه (i,j) را با O_{ij} نمایش می‌دهیم

$X \backslash Y$	1	2	…	j	…	c	جمع
1	O_{11}	O_{12}	…	O_{1j}	…	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	…	O_{2j}	…	O_{2c}	$O_{2.}$
⋮	⋮	⋮	…	⋮	⋮		
i	O_{i1}	O_{i2}	…	O_{ij}	…	O_{ic}	$O_{i.}$
⋮	⋮	⋮	…	⋮	…	⋮	
r	O_{r1}	O_{r2}	…	O_{rj}	…	O_{rc}	$O_{r.}$
جمع	$O_{.1}$	$O_{.2}$		$O_{.j}$		$O_{.c}$	$O_{..}$

در این بخش فرض صفر و فرض مقابل به صورت زیر است

H_0: بین X و Y رابطه وجود ندارد

H_1: 〃 〃 〃 〃 دارد

مثال: تعدادی از دانشجویان یک دانشگاه به تصادف انتخاب و بر حسب دو متغیر وضعیت درآمد و نحوه گذراندن اوقات فراغت رده‌بندی و نتایج آنها در جدول زیر آمده است.

$o_{11} = 5$

$e_{11} = \frac{32(35)}{116}$

درآمد \ گذراندن فراغت	سینما	پارک	دیدار خویشان	مسافرت	جمع
کم درآمد	5 / 9.66	9 / 4.97	12 / 8	6 / 9.38	32
درآمد متوسط	10 / 11.77	4 / 6.05	7 / 9.75	18 / 11.43	39
پردرآمد	20 / 13.57	5 / 6.98	10 / 11.25	10 / 13.19	45
جمع	35	18	29	34	116

آیا در سطح خطای 5% بین وضعیت درآمد و نحوه گذراندن اوقات فراغت دانشجویان رابطه وجود دارد؟

H_0: بین وضعیت درآمد و نحوه گذراندن اوقات فراغت دانشجویان رابطه وجود ندارد

H_1: " " " " " دارد

$$\chi^{2*} = \frac{(5-9.66)^2}{9.66} + \frac{(9-4.97)^2}{4.97} + \frac{(12-8)^2}{8} + \cdots + \frac{(10-13.19)^2}{13.19} = 18.77$$

$$df = (r-1)(c-1) = (3-1)(4-1) = 6$$

$$\chi^2_{0.05,6} = 12.592$$

جواب

$\chi^{2*} = 18.77 > 12.592$

پس H_0 رد می‌شود یعنی بین متغیر وضعیت درآمد و نحوه گذراندن اوقات فراغت دانشجویان رابطه وجود دارد.

براساس استقلال دو متغیر برای هر خانه یک فراوانی مورد انتظار زیر را با این فرمول:

$$e_{ij} = \frac{o_{i.} \cdot o_{.j}}{o_{..}}$$

فراوانی مورد انتظار

$$= \frac{\text{جمع سطر} \times \text{جمع ستون}}{\text{جمع کل}}$$

سپس مقدار

$$\chi^{2*} = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

را بدست می‌آوریم.

H_0 را در سطح خطای α رد می‌کنیم اگر

$$\chi^{2*} > \chi^2_{\alpha\,(r-1)(c-1)}$$

شرایط اعتبار آزمون χ^2

① هیچ خانه‌ای نباید فراوانی مورد انتظار کمتر از 1 داشته باشد

② حداکثر 20٪ از خانه‌ها فراوانی مورد انتظار کمتر از 5 داشته باشند

توجه: در صورت وجود هر یک از دو حالت فوق باید دو سطر یا دو ستون مجاور را ادغام کرد تا مشکل حل شود. به طور درست درجه آزادی نهایی از جدول ادغام شده به دست می‌آید.

x \ y	متغیر 1	متغیر 2	متغیر 3	متغیر 4	جمع
متغیر 1	2 / 2.5	7 / 9	3 / 1	10 / 9.5	
متغیر 2	15 / 7	11 / 10	12 / 10	6 / 16	
متغیر 3	9 / 7.5	3 / 2	11 / 15	12 / 3.5	
جمع					

2×4

$(2-1)(4-1) = 3$

$df = (2-1)(2-1) = 1$

مثال: تعدادی از معلمان شهرستان سبزوار را بر حسب دو متغیر مدرک تحصیلی و مقطع تدریس در جدول زیر آمده است

مدرک \ مقطع تدریس	ابتدایی	راهنمایی	دبیرستان	جمع
دیپلم	15 8.81	3 5.78	1 ~~4.40~~	19
فوق‌دیپلم	10 11.59	12 7.61	3 5.80	25
لیسانس	5 9.27	5 6.09	10 4.64	20
فوق‌لیسانس	2 2.31	1 1.51	2 ~~1.16~~	5
جمع	32	21	16	69

5/12

آیا در سطح خطای ۵٪ بین مدرک تحصیلی و مقطع تدریس رابطه وجود دارد؟

H_0: بین مدرک تحصیلی و مقطع تدریس معلمان رابطه وجود ندارد

H_1: " " " دارد

مدرک / مقطع تحصیلی	ابتدایی	راهنمایی	دبیرستان	جمع
دیپلم	15 (8.81)	3 (5.78)	1 (4.41)	19
فوق دیپلم	10 (11.59)	12 (7.61)	3 (5.80)	25
لیسانس و بالاتر	7 (11.59)	6 (7.61)	12 (5.80)	25
جمع	32	21	16	69

$$\chi^{2*} = \frac{(15-8.81)^2}{8.81} + \cdots + \frac{(12-5.80)^2}{5.80} = 21.21$$

$$df = (3-1)(3-1) = 4 \qquad \chi^2_{0.05,4} = 9.488$$

چون $\chi^2 = 21.21 > 9.488$ ، پس H_0 رد می‌شود.

بین تحصیلات و مقطع تدریس رابطه وجود دارد.

۲) آزمون نیکویی برازش

در این آزمون برازش یک توزیع احتمال را بر روی یک مجموعه داده آزمون می‌کنیم

$$\begin{cases} H_0: & \text{داده‌ها از توزیع مورد نظر پیروی می‌کنند} \\ H_1: & \text{" " " نمی‌کنند} \end{cases}$$

در این حالت نیز جدول به صورت زیر خواهد بود درست مثل قسمت قبل k مقدار داریم

x_1	x_2	.	.	.	.		x_k
O_1	O_2						O_k

o_i: فراوانی مشاهده شده از x_i می‌باشد

بر اساس H_0 یک مقدار مورد انتظار e_i برای o_i به دست آورده سپس مقدار

$$\chi^{2*} = \sum \frac{(o_i - e_i)^2}{e_i}$$ را محاسبه کنیم

H_0 را در سطح خطای α رد می‌کنیم اگر $\chi^{2*} > \chi^2_{\alpha, k-1}$.

مثال ۱: سکه‌ای را 60 بار پرتاب کرده‌ایم که 25 بار شیر و 35 بار خط آمده است. آیا در سطح خطای ۵٪ این سکه سالم است؟

شیر	خط
25 / 30	35 / 30

$$\chi^{2*} = \frac{(25-30)^2}{30} + \frac{(35-30)^2}{30} = \frac{25}{30} + \frac{25}{30} = 1.67$$

$$\begin{cases} H_0: P(H)=P(T)=1/2 & \text{سکه سالم است} \\ H_1: P(H)\neq P(T) & \text{سکه معیوب است} \end{cases}$$

$$\chi^2_{0.05,1} = 3.841$$

چون $\chi^{2*} = 1.67 \ngtr 3.841$ پس H_0 رد نمی‌شود

یعنی سکه سالم است.

مثال ۲: تاسی را 120 بار پرتاب کردیم. نتیجه زیر حاصل شده است

1	2	3	4	5	6
25 / 20	10 / 20	18 / 20	30 / 20	12 / 20	25 / 20

$$\begin{cases} H_0: P(1)=P(2)=\dots=P(6)=\frac{1}{6} & \text{تاس سالم است} \\ H_1 & \text{تاس معیوب است} \end{cases}$$

آیا در سطح خطای ۵٪ این تاس سالم است؟

$$\chi^{2*} = \frac{(25-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(30-20)^2}{20} + \frac{(12-20)^2}{20} + \frac{(25-20)^2}{20}$$

$$= \frac{25+100+4+100+64+25}{20} = \frac{318}{20} = 15.9$$

$\chi^2_{0.05,5}$

چون $15.9 > 11.07$ پس H_0 رد می‌شود

تاس معیوب است

مثال: نمونه‌ای تصادفی از دانشجویان یک درس انتخاب و نمره ریاضی آنها را

ثبت و در جدول زیر گروه‌بندی کرده‌ایم:

نمره	o_i (f_i)	p_i	e_i	x_i	$f_i x_i$	$f_i x_i^2$
0 – 4	7	0.14686	5.87	2	14	28
4 – 8	10	0.29747	11.90	6	60	360
8 – 12	15	0.33897	13.56	10	150	1500
12 – 16	6	0.17122	6.85	14	84	1176
16 – 20	2		1.82	18	36	648
	40		40		344	3702

آیا در سطح خطای ۵٪ می‌توان پذیرفت که توزیع نمرات ریاضی دانشجویان نرمال است؟

$$s^2=\frac{n\sum f_i x_i^2-(\sum f_i x_i)^2}{n(n-1)}=\frac{40(3702)-(344)^2}{40(39)}=19.07 \qquad \bar{x}=\frac{344}{40}=8.6$$

$$s=4.37$$

$$P(X<4)=P\left(\frac{X-\mu}{\sigma}<\frac{4-\mu}{\sigma}\right)=P\left(Z<\frac{4-\mu}{\sigma}\right)=$$

$$=P\left(Z<\frac{4-8.6}{4.37}\right)=P(Z<-1.05)$$
$$=0.14686$$

$$P(4<X<8)=P\left(\frac{4-8.6}{4.37}<Z<\frac{8-8.6}{4.37}\right)=P(-1.05<Z<-0.14)$$
$$=P(Z<-0.14)-P(Z<-1.05)$$
$$=0.44433-0.14686$$
$$=0.29747$$

$$P(8<X<12)=P\left(-0.14<Z<\underbrace{\frac{12-8.6}{4.37}}_{0.78}\right)$$
$$=P(Z<0.78)-P(Z<0.14)$$
$$=0.78230-0.44433=0.33897$$

$$P(12<X<16)=P\left(0.78<Z<\underbrace{\frac{16-8.6}{4.37}}_{1.69}\right)$$
$$=P(Z<1.69)-P(Z<0.78)=0.95352-0.78230=0.17122$$

نمره	f_i / o_i	p_i	e_i	x_i	$f_i x_i$	$f_i x_i^2$
0 – 4	7	0.14686	5.87	2	14	28
4 – 8	10	0.29747	11.90	6	60	360
8 – 12	15	0.33897	13.56	10	150	1500
12 – 16	6	0.17122	6.85	14	84	1176
16 – 20	2		1.82	18	36	648
			40		344	3702

$$\chi^{2*} = \frac{(7-5.87)^2}{5.87} + \frac{(10-11.90)^2}{11.90} + \cdots + \frac{(2-1.82)^2}{1.82} = 0.797$$

چون $\chi^{2*} = 0.797 \not> \chi^2_{0.05,4} = 9.488$ پس H_0 رد نمی‌شود

یعنی توزیع نرمال است

آیا در سطح خطای ۵٪ ... نرمال است؟

آیا در سطح خطای ۵٪ می‌توان گفت توزیع نمرات دارای نرمال با میانگین ۱۰ می‌باشد؟

آیا در سطح خطای ۵٪ ... ۱۰ واریانس ۱۵ دارد؟