آزمون نیکویی برازش :

Goodness of Fit Test

روشی است برای بررسی توزیع احتمالی (متغیر تصادفی مورد بررسی در) جمعیت.

رسته ها	رسته اول	...	رسته kام
مشاهدات (فراوانی ها)	O_1		O_k

* ابتدا باید داده ها را به صورت رسته ای نمایش داد.

* سپس باید فرض ها را به صورت زیر فرمول بندی کرد:

$$\begin{cases} H_0: P_1 = P_{01}, \ \dots, \ P_k = P_{0k} \\ H_1: \text{o.w.} \end{cases}$$

P_{01} ← عدد معلوم ، P_{0k} ← عدد معلوم

$$P_i = P(X \in \text{رسته } i\text{ام})$$

برای انجام آزمون فوق در سطح α، ابتدا $e_i = nP_{0i}$ را محاسبه می کنیم و سپس

رسته ها	رسته اول	...	رسته kام
مشاهدات (فراوانی ها)	O_1	...	O_k
تعداد مورد انتظار برای فراوانی ها	$e_1 = nP_{01}$	...	$e_k = nP_{0k}$

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \overset{H_0}{\approx} \chi^2_{(k-1)}$$

آماره χ^2 پیرسون

$$\text{if } X_0^2 \geq \chi^2_{1-\alpha,(k-1)} \Rightarrow RH_0$$

مقدار مشاهده شده برای X^2

* اگر برای محاسبه P_{0i} ها نیاز به برآورد پارامترهایی داشتیم در محاسبه درجه آزادی، تعداد پارامترهای برآورد شده را نیز از $k-1$ کم می کنیم.

* این آزمون زمانی اعتبار دارد که $e_i \geq 5$ باشد، در غیر اینصورت باید یک یا چند طبقه از جدول را با هم ادغام کرد تا شکل برطرف شود.

مثال (از کتاب ریاضی های آماری باشاچاریا):

فروشی ماشین لباسشویی خانگی در 5 رنگ مختلف به فروش می رسد. بازاریابی می خواهد میزان تقاضا برای رنگ های مختلف را مقایسه کند. نمونه ای 300 تایی از لباسشویی های فروخته شده را انتخاب می کند و نتایج زیر را به دست می آورد. آیا میزان تقاضا برای این 5 رنگ یکسان است؟

سبز	خرمایی	قرمز	آبی	سفید
86	65	52	40	55

ابتدا فرض ها را به صورت مناسب صورتبندی می کنیم :

$H_0: p_1 = \frac{1}{5}, \dots, p_5 = \frac{1}{5}$

H_1 : o.w.

```
data a;
input p obs;
e = 300*p;
chi = (obs-e)**2/e;
cards;
0.2  88
0.2  65
0.2  52
0.2  40
0.2  55
;
proc means;
var chi;
output out=b sum=chisq;
run;
data final;
set b;
df=4;
pvalue = 1-prob(0.95,4);
chisq_table = cinv(0.95,4);
run;
proc print data=final;
var chisq df pvalue chisq_table;
run;
```

$X_0^2 \geq \chi^2_{(4)\,0.95} \Rightarrow RH_0$

$p\text{-value} = P_{H_0}(X^2 \geq X_0^2) = 1 - P_{H_0}(X^2 \leq X$

$X^2 \sim \chi^2_{(4)}$

```
proc freq;
tabels obs/chisq testp=(20 20 20 20 20);
run;
```

مثال: بررسی کنید که آیا مشاهدات زیر از توزیع پواسون پیروی می‌کنند؟

x_i ها	1	2	3	4	5	6	7
فراوانی‌ها	2	14	20	18	16	14	2

ابتدا باید فرض‌ها را به صورت زیر مدل‌بندی کرد:

$$\begin{cases} H_0: p_1 = p_{01}, \dots, p_7 = p_{07} \\ H_1: o.w. \end{cases} \Longleftarrow \begin{cases} H_0: X \sim P(\lambda) \\ H_1: o.w. \end{cases}$$

$$p_i = P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!} = p_{0i}$$

تحت فرض H_0

برای محاسبه p_{0i} ها باید ابتدا λ را به دست آورد.

$$\hat{\lambda} = \bar{X} = \frac{1\times 2 + 2\times 14 + \dots + 7\times 2}{\underbrace{2+14+\dots+2}_{n}}$$

پس p_{0i} ها را می‌یابد و بر اساس آن $e_i = np_{0i}$ را به دست آورد.

بدنامه نویسی آن به کلاس انجام می‌شود: ...

مثال: بررسی کنید که آیا مشاهدات زیر از توزیع $U(1, 11)$ پیروی می‌کنند؟

1.5 , 2 , 3.4 , 9.5 , 11 , 3.5 , ...

ادامه همگی به کلاس ارتباط می‌شود.