ILLESZKEDÉSVIZSGÁLAT

Tiszta illeszkedésvizsgálat

Becsléses illeszkedésvizsgálat

χ^2-próba az eloszlás (illeszkedés) vizsgálatára

Illeszkedésvizsgálat Poisson-eloszlásra

1. példa

(G.E.P. Box, W.G. Hunter, J.S. Hunter: Statistics for experimenters, J. Wiley, 1978, p. 143)

Müzligyárban ellenőrzik a mazsolák számát. Az előírás az, hogy egy mintavevő kanálnyi müzliben 36 szem mazsolának kell lennie. 12 mintát vettek, az ezekben talált mazsola-szemek száma:
43, 46, 50, 40, 38, 29, 31, 35, 41, 52, 48, 37.
Teljesül-e az előírás?

Poisson-eloszlás: ritka események modellezésére használható

Alkalmazásának feltételei:

- ✓ bármely egységben bekövetkező eseménynek függetlennek kell lennie a többi egységbelitől
- ✓ az esemény bekövetkezésének valószínűsége bármely egységben azonos, és arányos az egység méretével
- ✓ annak valószínűsége, hogy két vagy több előfordulás következik be egy egységben, az egység méretének csökkentésével nullához tart

Az adagonkénti mazsolák számának előfordulási valószínűsége Poisson-eloszlással írható le:

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

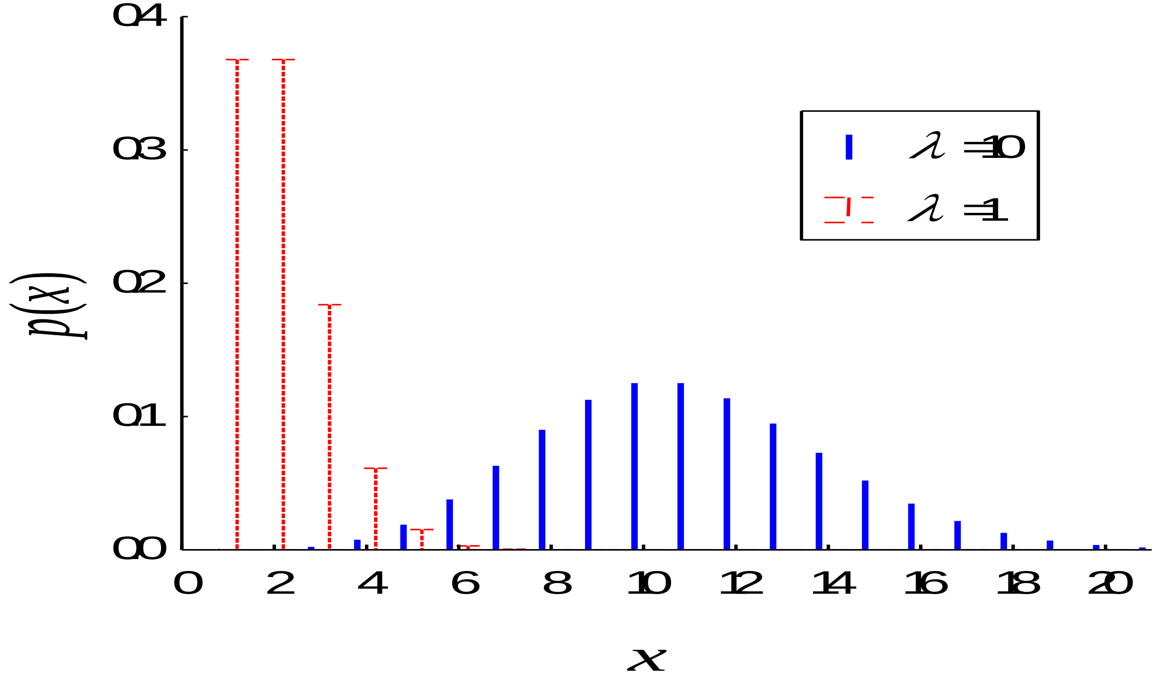
$$E(k) = \lambda \qquad Var(k) = \lambda$$

Nullhipotézis:

$$\mathrm{H}_0 : \lambda = \lambda_0 = 36 \qquad \mathrm{H}_1 : \lambda \neq 36$$

azaz az adatok adott paraméterű Poisson-eloszlást követnek (ún. ***tiszta illeszkedésvizsgálat***)

A Poisson-eloszlás közelíthető normális eloszlással, ha a λ paraméter elég nagy:



$$z \approx \frac{k-\lambda}{\sqrt{\lambda}}$$

az egy adagban található mazsola-szemek száma (k) normális eloszlással közelítve

$$\sum_i {z_i}^2 = \chi^2$$

a több mintára is felírva, ha az adagonként található mazsola-szemek száma független egymástól

Próbastatisztika: $$\chi_0^2 = \sum_i \left(\frac{k_i - \lambda_0}{\sqrt{\lambda_0}} \right)^2 = \sum_i \frac{(k_i - \lambda_0)^2}{\lambda_0}$$

$$\chi_0^2 = \frac{(43-36)^2}{36} + \frac{(46-36)^2}{36} + \ldots\ldots + \frac{(37-36)^2}{36} = 24.06$$

ν=12, p=0.02

Nullhipotézist - miszerint a mazsolák száma λ=36 paraméterű Poisson-eloszlás szerint ingadozik - elutasítottuk.

Két eset lehetséges:

- Poisson-eloszlású, de nem 36 szem mazsola jut egy kanálba
- nem Poisson-eloszlást követ a mazsolák száma (overdispersion)

1. eset: A Poisson-eloszlás λ paramétere nem 36.

a Poisson-eloszlás additív tulajdonságú, így

$$\sum_i k_i \quad \text{~Poisson,} \qquad \sum_i \lambda_i \qquad \text{paraméterrel}$$

A példában: $\sum_i k_i = 490 \qquad \sum_i \lambda_i = 12 \cdot 36 = 432$

$$H_0 : \lambda = \lambda_0 = 432 \qquad H_1 : \lambda \neq 432$$

$$\chi_0^2 = \frac{(490 - 432)^2}{432} = 7.79 \qquad \nu=1,\ p=0.0053$$

Nullhipotézist elutasítottuk, azaz mintavevő-kanalanként a mazsolák száma nem 36.

Adjunk becslést az egy mintavevő kanálra jutó mazsolák számára:

$$43+46+50+40+38+29+31+35+41+52+48+37=490$$

$$\hat{\lambda}=\frac{490}{12}=40.83$$

Végezzünk hipotézisvizsgálatot az egy mintavevő kanálra jutó átlagos mazsola-szám felhasználásával!

2. eset: A mazsola-szemek eloszlása λ=40.83 paraméterű Poisson-eloszlás (*ún. becsléses illeszkedésvizsgálat*)

$$\chi_0^2 = \frac{(43-40.83)^2}{40.83} + ... + \frac{(37-40.83)^2}{40.83} = 14.34$$

$\nu = 12-1=11$ (az eloszlás paraméterét is az adatokból becsültük)

p=0.215

Következtetés:
Az egy mintavevő kanálra jutó mazsola-szám ugyan Poisson-eloszlás szerint ingadozik (az egyes kanalakban a mazsola-szám független), de nem 36 szem mazsola jut átlagosan egy mintavevő kanálra.

2. példa

(G.U. Yule, M.G. Kendall: Bevezetés a statisztika elméletébe, Közgazdasági és Jogi Könyvkiadó, Bp. 1964, p. 207)

A táblázat a Dél-London egyes 0.25 m^2–es négyzeteire esett bombák számát mutatja a II. világháborúban.
Tudtak-e a németek célozni?

bomba	tapasztalt gyakoriság
0	229
1	211
2	93
3	35
4	7
7	1

összesen 576 négyzet

összesen 537 találat

$$\hat{\lambda} = \frac{537}{576} = 0.9323$$

bomba	tapasztalt gyakoriság	várható gyakoriság
0	229	226.74
1	211	211.39
2	93	98.54
3	35	30.62
4	7	7.14
7	1	1.57

$$\hat{\lambda} = \frac{537}{576} = 0.9323$$

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\chi_0^2 = \sum_i \sum_j \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\nu = 6 - 1 = 5$$

$$p(0) = \frac{e^{-0.9322} \cdot 0.9322^0}{0!} = 0.3937$$

$$E_{ij}(0) = N \cdot p(0) = 576 \cdot 0.3937 = 226.76$$

$$\chi_0^2 = \frac{(229 - 226.76)^2}{226.76} + \ldots\ldots + \frac{(1 - 1.57)^2}{1.57} = 1.17$$

$$\chi^2_{\mathrm{krit}}(\nu = 5) = 11.07$$

H_0-t elfogadjuk, azaz véletlen volt, hogy hová esett a bomba

Illeszkedésvizsgálat multinomiális eloszlásra

Binomiális eloszlás: kétféle kimenetel

$$P(k) = \binom{n}{k} \pi^k (1-\pi)^{n-k} = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

Multinomiális eloszlás: többféle (*c* féle) kimenetel

$$P(n_1, n_2, \ldots, n_c) = \frac{n!}{n_1! n_2! \ldots n_c} \pi_1^{n_1} \pi_2^{n_2} \ldots \pi_c^{n_c}$$

$$\chi^2 = \sum_i^c \frac{(n_i - n\pi_i)^2}{n\pi_i} \qquad \nu = c - 1$$

a szumma tagjai között egy összefüggés van: $n = \sum_i n_i$

$$\chi_0^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Observed

Expected

3. példa

(A. C. Wardlaw: Practical statistics for experimental biologists, J. Wiley & Sons, 1985 p. 112)

129 olyan gyermek vércsoportját vizsgálták, akinek mindkét szülője AB vércsoportba tartozott.
28 gyermeknek volt A (AA), 36-nak B (BB) és 65-nek AB a vércsoportja.
A Mendel-féle öröklődési szabályok szerint az esetek ¼-ében kell A, ¼-ében B, ½-ében pedig AB előfordulásnak lennie.
Ellentmondanak az eredmények a Mendel-szabálynak?

genotípus	AA	A0	AB	BB	B0	00
fenotípus	A	A	AB	B	B	0

$$H_0: \pi_{AA} = \frac{1}{4}; \quad \pi_{BB} = \frac{1}{4}; \quad \pi_{AB} = \frac{2}{4}$$

tiszta illeszkedésvizsgálat!

	Talált (O)	Feltételezett (E)
A	28	$\frac{129}{4} = 32.25$
B	36	$\frac{129}{4} = 32.25$
AB	65	$\frac{129}{2} = 64.5$
Σ	129	129

$$\chi^2 = \sum_i^c \frac{(n_i - n\pi_i)^2}{n\pi_i}$$

$$\chi_0^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

ν=3-1=2, kétoldali

$$\chi_0^2 = \frac{\left(28 - \frac{129}{4}\right)^2}{\frac{129}{4}} + \frac{\left(36 - \frac{129}{4}\right)^2}{\frac{129}{4}} + \frac{\left(65 - \frac{129}{2}\right)^2}{\frac{129}{2}} = 1$$

$$\chi_{krit}^2(\alpha = 0.025, \nu = 2) = 7.378$$

$$\chi_{krit}^2(\alpha = 0.975, \nu = 2) = 0.0506$$

H_0-t elfogadjuk.