Curs 14

Testarea ipotezelor statistice

14.1 Testul lui Pearson (χ^2)

Este folosit ca indicator al concordanţei unei repartiţii empirice cu una teoretică. Este unul din cele mai importante teste statistice.

Se consideră o selecţie de volum n, $x_1', x_2', \ldots, x_n'$.

Ipoteza H_0: selecţia făcută provine dintr-o populaţie cu funcţia de repartiţie F complet specificată (de exemplu normală standard). Ipoteza alternativă H_1: selecţia nu provine din populaţia definită de F.

Se consideră o selecţie de volum n, $x_1', x_2', \ldots, x_n'$. Datele se împart în r clase, $x_1 < x_2 < \cdots < x_r$. Obţinem intervalele $[x_1, x_2)\,,\ [x_2, x_3)\,, \ldots,\ [x_r, \infty)\,.$ Fie f_j numărul de observaţii din intervalul j şi p_j probabilitatea teoretică ca v. a. din care provine selecţia să ia valori în intervalul j, dacă admitem ipoteza H_0. Atunci frecvenţa teoretică este np_j.

Statistica utilizată în testarea acestei ipoteze este:

$$\chi^2 = \sum_{j=1}^{r} \frac{(f_j - np_j)^2}{np_j} \tag{14.1}$$

care urmează legea χ^2 cu $r-1$.

14.1.1 Algoritmul testului lui Pearson

Pasul 1. Alegerea lui r: Mann şi Wald au propus formula

$$r = 4\left(\frac{2(n-1)^2}{C^2}\right)^{\frac{1}{5}}$$

unde C este cuantila de ordin α a repartiţiei $N(0,1)$.

Pasul 2. Selecţia de volum n se împarte în r intervale. Se calculează f_k numărul de observaţii în intervalul k iar p_k probabilitatea teoretică ca v. a. în studiu să ia valori în intervalul k în ipoteza că H_0 este adevărată. Pentru aceasta se estimează media (sau alţi parametri) şi se generează acelaşi număr de date repartizate conform legii teoretice alese. Pentru o buna aplicare a testului trebuie ca $np_k \geq 5$ pentru $k = \overline{1, r}$.

Pasul 3. Se calculează

$$\chi^2 = \sum_{j=1}^{r} \frac{(f_j - np_j)^2}{np_j}.$$

Pasul 4. Se stabileşte nivelul de semnificaţie $1-\alpha$, se calculează $\chi^2_{\alpha, r-1}$.

Ipoteza H_0 se respinge dacă $\chi^2 > \chi^2_{\alpha, r-1}$.

Dacă $\chi^2 < \chi^2_{\alpha, r-1}$ nu avem motive să respingem ipoteza H_0.

Observaţia 14.1.1 *În practică se poate da o expresie mai uşor de reţinut pentru (14.1) şi anume*
$$\chi^2 = \sum \frac{(O-E)^2}{E},$$
unde O sunt datele observate şi E cele estimate (pentru care trebuie decisă buna potrivire).

Observaţia 14.1.2 *Dacă se estimează un număr de s parametri, atunci χ^2 urmează o repartiţie χ^2 cu $r-s-1$ grade de libertate.*

Exemplul 14.1.3 La o oră de vârf s-au sondat canalele de televiziune TVR1, PROTV, TVR2, Antena 1, Realitatea, care au avut la 23 martie audienţele 34%, 8%, 12%, 34%, 12% respectiv. La un sondaj printre 500 telespectatori după 6 luni s-au constatat rezultatele următoare: 159 spectatori pentru TVR1, 49, 62, 161, 69 pentru celelalte. Apare o diferenţă semnificativă?

Rezolvare. În acest exemplu avem $r=5$, $X=$ audienţa, $n=500$, $p_1=0.34$, $p_2=0.08$, $p_3=0.12$, $p_4=0.34$, $p_5=0.12$.

Numărul estimat de telespectatori la 23 martie este respectiv
$$np_1 = 500\cdot 0.34 = 170,\quad np_2 = 500\cdot 0.08 = 40,\quad np_3 = 500\cdot 0.12 = 60,$$
$$np_4 = 500\cdot 0.34 = 170,\quad np_5 = 500\cdot 0.12 = 60.$$

Calculăm
$$\chi^2 = \tfrac{(159-170)^2}{170} + \tfrac{(49-40)^2}{40} + \tfrac{(62-60)^2}{60} + \tfrac{(161-170)^2}{170} + \tfrac{(69-60)^2}{60} = 4.629\,9.$$

Datele pot fi organizate conform următorului tabel

Grupa	1	2	3	4	5
f_k	159	49	62	161	69
p_k	0.34	0.08	0.12	0.34	0.12
np_k	170	40	60	170	60
$f_k - np_k$	-11	9	2	-9	9
$(f_k - np_k)^2$	121	81	4	81	81
$\frac{(f_k - np_k)^2}{np_k}$	0.71176	2.025	6.6667×10^{-2}	0.47647	1.35

Alegem $\alpha = 0.1$, $r-1=4$, atunci $\chi^2_{0.1}(4) = 7.779$. Deoarece $\chi^2 < \chi^2_{0.1}(4)$ ($4.629\,9 < 7.7794$) rezultă că nu avem argumente suficiente pentru a respinge ipoteza H_0.

Dacă luăm $\alpha = 0.5$ atunci $\chi^2_{0.5}(4) = 3.36$ şi H_0 se respinge cu o eroare de 50%.

Secvenţa de program Mathematica care realizează această testare este dată mai jos:

Potrivire cu repartiţia binomială

Exemplul 14.1.4 O reclamă a fost difuzată în mass-media. Dintr-un eşalon de 800 de persoane au fost 434 care nu au auzit (văzut) reclama; 329 au auzit o dată; 35 de două ori şi 2 de 3 ori (nimeni mai mult de trei ori). Ne propunem să verificăm la nivel de semnificaţie 95% dacă numărul de dăţi când o persoană a aflat de acea reclamă urmează o repartiţie binomială cu parametru $p=0.2$.

Rezolvare. Paşii care trebuie urmaţi în testarea ipotezei statistice

Pasul 1. Variabila care interesează: X numără de câte ori o persoană a auzit reclama.

Pasul 2. H_0 : variabila X urmează o distribuţie binomială.

Pasul 3. H_1 : variabila X nu urmează o distribuţie binomială.

Pasul 4. $\alpha = 0.05$.

Pasul 5. Testul statistic este

$$\chi^2 = \sum_{k=1}^{r} \frac{(f_k - np_k)^2}{np_k}$$

Pasul 6. Ipoteza H_0 este respinsă dacă $\chi^2 > \chi^2_{0.05}(3) = 7.81473$. Regiunea critică este $(7.81473, \infty)$.

Pasul 7. Calculele: $r = 4$, $k = 1, 2, 3, 4$ şi aceste valori sunt luate de X cu probabilităţile $p_k = C_3^{k-1} p^{k-1} (1-p)^{4-k}$, deci

$$p_1 = C_3^0 (0.2)^0 (0.8)^3 = 0.512, \quad p_2 = C_3^1 (0.2)^1 (0.8)^2 = 3 \cdot 0.128 = 0.384,$$
$$p_3 = 3 \cdot (0.2)^2 (0.8)^1 = 0.096, \quad p_4 = (0.2)^3 (0.8)^0 = 0.008$$

$n = 800$,

Grupa	0	1	2	3
f_k	434	329	35	2
p_k	0.512	0.384	0.096	0.008
np_k	409.6	307.2	76.8	6.4
$f_k - np_k$	24.4	21.8	-41.8	-4.4
$(f_k - np_k)^2$	595.36	475.24	1747.2	19.36
$\frac{(f_k - np_k)^2}{np_k}$	$7.864\,1 \times 10^{-2}$	1.547	22.75	3.025

$$\chi^2 = 28.776, \quad \chi^2_{0.95}(3) = 7.81473.$$

Pasul 8. $28.776 > 7.81473$ rezultă că respingem H_0, deci datele observate nu urmează o lege binomială cu parametru $p = 0.2$.

În ipoteza că variabila studiată ar urma, totuşi, o repartiţie binomială, estimăm parametrul p cu metoda verosimilităţii maxime. Construim funcţia de verosimilitate maximă:

$$p_1 = C_3^0 p^0 (1-p)^3, \; p_2 = C_3^1 p^1 (1-p)^2, \; p_3 = C_3^2 p^2 (1-p)^1, \; p_4 = C_3^3 p^3,$$
$$L(x_1, x_2, x_3, x_4; p) = p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$
$$L(434, 329, 35, 2; p) = (1-p)^{3 \cdot 434} 3p^{329} (1-p)^{2 \cdot 329} 3p^{2 \cdot 35} (1-p)^{35} P^{3 \cdot 2} =$$
$$= 9p^{405} (1-p)^{1995},$$
$$\ln L(434, 329, 35, 2; p) = \ln 9 + 405 \ln p + 1995 \ln(1-p),$$
$$\frac{\partial \ln L(434, 329, 35, 2; \theta)}{\partial p} = \frac{405}{p} - \frac{1995}{1-p},$$
$$\frac{405}{p} - \frac{1995}{1-p} = 0, p = \frac{27}{160} = 0.16875.$$

Potrivire cu repartiţia Poisson

Exemplul 14.1.5 Vrem să stabilim dacă numărul defecţiunilor unui circuit al unei imprimante urmează o repartiţie Poisson. Se studiază comportatea acestui circuit pe un număr de 60 de imprimante şi sunt consemnate numărul defecţiunilor care au apărut.

Nr. defecţiunilor	Frecvenţa
0	32
1	15
2	9
≥3	4

Media distribuţiei Poisson presupuse este necunoscută şi trebuie estimată din datele eşantionului. Estimarea mediei numărului defectelor circuitului o facem printr-un estimator punctual al mediei, $\overline{x} = \frac{32\cdot 0+15\cdot 1+9\cdot 2+4\cdot 3}{60} = 0.75$. Din distribuţia Poisson cu parametrul $\lambda = 0.75$ putem calcula probabilităţile teoretice p_i, $i = 1, 2, 3, 4$,

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472\,367$$
$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354\,275$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.132\,853$$
$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.0405054$$

Aplicăm testul.

Pasul 1. Variabila care interesează: X numărul de defecte care apar la un circuit.

Pasul 2. H_0 : variabila X urmează o distribuţie Poisson.

Pasul 3. H_1 : variabila X nu urmează o distribuţie Poisson.

Pasul 4. $\alpha = 0.05$

Pasul 5. Testul statistic este

$$\chi^2 = \sum_{k=1}^{r} \frac{(f_k - np_k)^2}{np_k}$$

Pasul 6. Determinăm regiunea critică. χ^2 urmează o distribuţie hi-pătrat cu 2 grade de libertate. Sunt patru grade, ar trebui să urmeze o distribuţie cu 3 grade de libertate dar, deoarece s-a estimat parametrul distribuţiei Poisson, numărul gradelor de libertate se consideră cu o unitate mai mic, conform Observaţiei 14.1.2.

Frecvenţele observate

f1 = 32;

f2 = 15;

f3 = 9;

f4 = 4;

Determinăm regiunea critică:

Ipoteza H_0 este respinsă dacă $\chi^2 > \chi^2_{0.05}(2) = 5.99146$. Regiunea critică este $(5.99146, \infty)$.

Pasul 7. Calculele:

```
CHI = (n p1 - f1)^2/(n p1) + (n p2 - f2)^2/(n p2) + (n p3 - f3)^2/(n p3) + (n p4 - f4)^2/(n p4)
```

3.46021

Pasul 8. $\chi^2 = 3.46021 < \chi^2_{0.05}(2) = 5.99146$. Nu avem motive să respingem ipoteza nulă.

Potrivire cu repartiţia uniformă

Exemplul 14.1.6 Un comerciant vrea să vadă dacă un anumit produs se vinde la fel (uniform) în 5 dintre magazinele sale. Prin experiment constată că într-o saptamână s-au realizat vânzări în mii RON de 43, 29, 52, 34, 48. Este această informaţie suficientă pentru a considera că există mari diferenţe între cele 5 magazine?

Pasul 1. Variabila care interesează: X veniturile realizate prin vânzări în 5 magazine, în decursul unei săptămâni.

Pasul 2. Ipoteza H_0: variabila X urmează o distribuţie uniformă. Probabilitatea de a cumpăra de la unul din cele cinci magazine este aceeaşi $p_1 = p_2 = \cdots = p_5 = 0.2$, cu alte cuvinte vânzările sunt aceleaşi.

Observaţia 14.1.7 *În cazul unei repartiţii uniforme a vânzărilor pe cele* 5 *magazine, ele ar trebui să vândă fiecare în valoare de* $206 \cdot 0.2 = 41.2$ *mii RON. Acestea sunt datele estimate dacă ipoteza* H_0 *ar fi adevărată. Datele date în enunţ sunt cele observate.*

Pasul 3. Ipoteza H_1: variabila X nu urmează o distribuţie uniformă.

Pasul 4. $\alpha = 0.1$

Pasul 5. Testul statistic este

$$\chi^2 = \sum_{k=1}^{r} \frac{(f_k - np_k)^2}{np_k}.$$

Pasul 6. Determinăm regiunea critică. χ^2 urmează o distribuţie hi-pătrat cu 4 grade de libertate.

```
p1 = 0.2; p2 = 0.2; p3 = 0.2; p4 = 0.2; p5 = 0.2;
f1 = 43; f2 = 29; f3 = 52; f4 = 34; f5 = 48;
n = f1 + f2 + f3 + f4 + f5
```

206

Regiunea critică este $(7.77944, \infty)$.

Deoarece am estimat probabilităţile teoretice $p_1 = p_2 = \cdots = p_5 = 0.2$ care depind de volumul selecţiei (numărul magazinelor cercetate) specialiştii recomandă să se considere că au rămas $5 - 1 - 1 = 3$ grade de libertate. În acest caz regiunea critică este $(6.25139, \infty)$.

Pasul 7. Calculele:

```
CHI = (n p1 - f1)^2/(n p1) + (n p2 - f2)^2/(n p2) + (n p3 - f3)^2/(n p3) + (n p4 - f4)^2/(n p4) + (n p5 - f5)^2/(n p5)
```

8.90291

Pasul 8. Pentru $r = 5$ la un nivel de semnificaţie de 10% $\chi^2_{0.05}(4) = 7.77944 > 8.902\,91$, deci nu avem motive să respingem ipoteza H_0. La fel şi pentru cazul a trei grade de libertate.

Observăm că dacă considerăm $\alpha = 0.05$, atunci avem două situaţii (3 sau 4 grade de libertate), pentru care regiunile critice sunt $(9.48773, \infty)$ şi respectiv $(7.81473, \infty)$.

Observăm că între cele două situaţii interpretările sunt diferite şi atunci concluzia se trage în funcţie de conjunctură. Pentru $r = 5$ la un nivel de semnificaţie de 5% $\chi^2_{0.05}(4) = 9.48773 > 8.902\,91$, deci nu avem motive să respingem ipoteza H_0. În cazul în care se consideră $5 - 1 - 1 = 3$ grade de libertate, $\chi^2_{0.05}(3) = 7.81473 < 8.902\,9$ şi în acest caz putem zice că sunt diferenţe mari între magazine.

În acest caz putem zice că sunt diferenţe mari între magazine.

Potrivire cu repartiţia normală

Exemplul 14.1.8 Presupunem că timpul mediu de asamblare (în minute) pentru un eşantion de 300 de aparate electronice a fost $m = 84$, cu abaterea medie pătratică $\sigma = 3$. S-a observat că aceste 300 de aparate au fost repartizate astfel: pentru 15 aparate au fost necesare sub 78 de minute, pentru 39 între 78-81 minute, pentru 96 între 81-84 minute, pentru 87 între 84-87, pentru 48 aparate între 87-90 minute şi pentru 15 aparate peste 90 minute. Să se testeze la nivel de semnificaţie de 1% dacă datele anterioare sunt repartizate normal.

Pasul 1. Variabila care interesează: X timpul mediu de asamblare a unor aparate electronice.

Pasul 2. H_0 : variabila X urmează o distribuţie normală.

Pasul 3. H_1 : variabila X nu urmează o distribuţie normală.

Pasul 4. $\alpha = 0.1$

Pasul 5. Testul statistic este

$$\chi^2 = \sum_{k=1}^{r} \frac{(f_k - np_k)^2}{np_k}.$$

Pasul 6. Determinăm regiunea critică. Datele sunt împărţite în 6 grupe. S-a estimat media, $m = 84$ şi abaterea medie pătratică, $\sigma = 3$. Testul statistic urmează o distribuţie $\chi^2_{0.01}(3)$.

m = 84;
σ= 3;
f1 = 15;
f2 = 39;
f3 = 96;
f4 = 87;
f5 = 48;
f6 = 15;
n = f1 + f2 + f3 + f4 + f5 + f6
300

Regiunea critică este: $(11.3449, \infty)$.

Pasul 7. Calculele:Avem $r = 6$ grupe; dacă X = timpul de asamblare şi dacă $X \in N(84, 9)$, atunci

$$\begin{aligned}
p_1 &= P(X < 78) = \Phi\left(\frac{78-84}{3}\right) = \Phi(-2) = 1 - \Phi(2) = \\
&= 1 - 0.977\,25 = 0.022\,75 \\
p_2 &= P(78 \le X < 81) = \Phi\left(\frac{81-84}{3}\right) - \Phi\left(\frac{78-84}{3}\right) = \\
&= \Phi(-1) - \Phi(-2) = 0.977\,25 - 0.841\,34 = 0.135\,91 \\
p_3 &= P(81 \le X < 84) = \Phi\left(\frac{84-84}{3}\right) - \Phi\left(\frac{81-84}{3}\right) = \\
&= \Phi(0) - \Phi(-1) = 0.841\,34 - 0.5 = 0.341\,34 \\
p_4 &= P(84 \le X < 87) = \Phi\left(\frac{87-84}{3}\right) - \Phi\left(\frac{84-84}{3}\right) = \\
&= \Phi(1) - \Phi(0) = 0.841\,34 - 0.5 = 0.341\,34 \\
p_5 &= P(87 \le X < 90) = \Phi\left(\frac{90-84}{3}\right) - \Phi\left(\frac{87-84}{3}\right) = \\
&= \Phi(2) - \Phi(1) = 0.977\,25 - 0.841\,34 = 0.135\,91 \\
p_6 &= P(90 \le X) = 1 - \Phi\left(\frac{90-84}{3}\right) = 1 - \Phi(2) = \\
&= 1 - 0.977\,25 = 0.022\,75
\end{aligned}$$

$n = 300$
$\chi^2 = 23.6597$

Pasul 8. $\chi^2 = 23.6597 > \chi^2_{0.01}(3) = 11.3449$ rezultă că respingem ipoteza H_0.

Potrivire cu repartiţia exponenţială

În general nu avem o densitate de probabilitate pentru X, ci numai un model statistic pe care îl alegem din modelele cunoscute. Considerăm un eşantion $(X_1, X_2, \ldots, X_k) \in Multinomial\,[n, p_1, p_2, \ldots, p_k]$. În mod natural ne punem problema să gasim un estimator de verosimilitate maximă în acest caz. Funcţia de verosimilitate maximă se ia de forma:

$$L(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k) = (p_1)^{x_1} (p_2)^{x_2} \cdots (p_k)^{x_k}.$$

Exemplul 14.1.9 Presupunem că este testată durata de funcţionare a unui tip de becuri (măsurată în ore) pentru un eşantion de 30 de becuri. S-a observat că durata de funcţionare a 9 becuri este sub 1000 de ore, a 12 becuri între 1000 şi 2000 de ore, a 8 becuri între 2000 şi 3000 de ore, iar a unui bec peste 3000 de ore. Să se testeze la nivel de semnificaţie de 0.1% dacă datele prezentate sunt repartizate exponenţial.

Avem nevoie de o estimare a parametrului v. a. X care ia ca valori durata de funcţionare (măsurată în sute de ore) a becurilor, presupusă exponenţială $X \in Exponential\,[\theta]$. Reamintim că pentru repartiţia exponenţială $F(x) = \begin{cases} 0 & \text{dacă } x \leq 0 \\ 1 - e^{-\theta x} & \text{dacă } x > 0 \end{cases}$.

Datele se împart în patru grupe, $(0, 1]$, $(1, 2]$, $(2, 3]$, $(3, \infty)$.

$$\begin{aligned} p_1 &= P(0 \leq X < 1) = F(1) - F(0) = 1 - e^{-\theta}, \\ p_2 &= P(1 \leq X < 2) = F(2) - F(1) = e^{-\theta} - e^{-2\theta}, \\ p_3 &= P(2 \leq X < 3) = F(3) - F(2) = e^{-2\theta} - e^{-3\theta}, \\ p_4 &= P(3 \leq X) = 1 - F(3) = e^{-3\theta}. \end{aligned}$$

Pentru a calcula probabilităţile trebuie estimat parametrul θ. Folosim estimatorul de verosimilitate maximă. Funcţia de verosimilitate maximă este

$$L(x_1, \ldots, x_n; \theta) = \left(1 - e^{-\theta}\right)^9 \left(e^{-\theta} - e^{-2\theta}\right)^{12} \left(e^{-2\theta} - e^{-3\theta}\right)^8 \left(e^{-\theta}\right)^1.$$

$$\begin{aligned} \ln L(x_1, \ldots, x_n; \theta) &= \ln \left(1-e^{-\theta}\right)^9 \left(e^{-\theta}-e^{-2\theta}\right)^{12} \left(e^{-2\theta}-e^{-3\theta}\right)^8 \left(e^{-3\theta}\right)^1 = \\ &= 9\ln(1 - e^{-\theta}) + 12\ln(e^{-\theta} - e^{-2\theta}) + 8\ln(e^{-2\theta} - e^{-3\theta}) - 3\theta, \end{aligned}$$

$$\begin{aligned} &\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \theta)}{\partial \theta} = \\ &= \frac{9e^{-\theta}}{1-e^{-\theta}} + \frac{12}{e^{-\theta}-e^{-2\theta}}\left(2e^{-2\theta}-e^{-\theta}\right) + \frac{8}{e^{-2\theta}-e^{-3\theta}}\left(3e^{-3\theta}-2e^{-2\theta}\right) - 3 = \\ &= \frac{9e^{-\theta}}{1 - e^{-\theta}} + \frac{12}{1 - e^{-\theta}}\left(2e^{-\theta} - 1\right) + \frac{8}{1 - e^{-\theta}}\left(3e^{-\theta} - 2\right) - 3 = \\ &= \frac{1}{e^{-\theta} - 1}\left(31 - 60e^{-\theta}\right), \end{aligned}$$

$$\frac{1}{e^{-\theta} - 1}\left(31 - 60e^{-\theta}\right) = 0 \Rightarrow \hat{\theta} = 0.66036,$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \theta)}{\partial \theta} = 0 \Rightarrow \hat{\theta} = 0.660\,36,$$

$$\begin{aligned} &\frac{\partial^2 \ln L(x_1, x_2, \ldots, x_n; \theta)}{\partial \theta^2} = \\ &= -57\frac{e^{-\theta}}{1 - e^{-\theta}} - 9\frac{e^{2(-\theta)}}{(1 - e^{-\theta})^2} - 12\frac{e^{-\theta}}{(1 - e^{-\theta})^2}\left(2e^{-\theta} - 1\right) - \\ &- 8\frac{e^{-\theta}}{(1 - e^{-\theta})^2}\left(3e^{-\theta} - 2\right) = -29\frac{e^{-\theta}}{(1 - e^{-\theta})^2} < 0, \end{aligned}$$

$$\left.\frac{\partial^2 \ln L(x_1, x_2, \ldots, x_n; \theta)}{\partial \theta^2}\right|_{\theta=0.66036} =$$
$$= -\frac{57e^{-0.66036}}{1-e^{-0.66036}} - \frac{9e^{2(-0.66036)}}{(1-e^{-0.66036})^2} - \frac{12e^{-0.66036}}{(1-e^{-0.66036})^2}\left(2e^{-0.66036}-1\right) -$$
$$- \frac{8e^{-0.66036}}{(1-e^{-0.66036})^2}\left(3e^{-0.66036}-2\right) = -64.137.$$

$\hat{\theta} = 0.660\,36$ este estimator de verosimilitate maximă.

Aplicăm testul.

Pasul 1. Variabila care interesează: X durata de funcţionare măsurată în sute de orea a unui tip de becuri.

Pasul 2. H_0 : variabila X urmează o distribuţie exponenţială.

Pasul 3. H_1 : variabila X nu urmează o distribuţie exponenţială.

Pasul 4. $\alpha = 0.001$

Pasul 5. Testul statistic este

$$\chi^2 = \sum_{k=1}^{r} \frac{(f_k - np_k)^2}{np_k}.$$

Pasul 6. Determinăm regiunea critică.

Regiunea critică este $(13.8155, \infty)$.

Pasul 7. Calculele:

Probabilităţile teoretice sunt

$$\begin{aligned}
p_1 &= 1 - e^{-0.660\,36} = 0.483\,33, \\
p_2 &= e^{-0.660\,36} - e^{-2 \cdot 0.660\,36} = 0.24972, \\
p_3 &= e^{-2 \cdot 0.660\,36} - e^{-3 \cdot 0.660\,36} = 0.12902, \\
p_4 &= e^{-3 \cdot 0.660\,36} = 0.137\,92, \\
\chi^2 &= 2.086\,2 + 2.713\,2 + 4.405\,5 + 2.379\,3 = 11.584
\end{aligned}$$
$$\chi^2_{0.999}(2) = 13.8155$$

f1 = 9;

f2 = 12;

f3 = 8;

f4 = 1;

n = 30;

CHI = (n p1 - f1)^2/(n p1) + (n p2 - f2)^2/(n p2) + (n p3 - f3)^2/(n p3) + (n p4 - f4)^2/(n p4)

11.5837

$\chi^2 < \chi^2_{0.001}(2)$, rezultă că nu avem motive să respingem ipoteza H_0.