

Seminarul 7

1. i) Estimați parametrul necunoscut $p \in (0, 1)$ pentru distribuția binomială a unei caracteristici cercetate: $X \sim Bino(N, p)$, unde $N \in \mathbb{N}^*$ este cunoscut, cu metoda momentelor, respectiv metoda verosimilității maxime. Sunt estimatorii obținuți nedeplasați, respectiv consistenți?

ii) Într-o urnă sunt bile albe și negre. Proportia de bile albe $p \in (0, 1)$ este necunoscută. În urma a $n = 6$ serii a câte $N = 5$ extrageri cu returnarea bilei extrase în urnă s-au obținut: 3, 4, 2, 0, 2, respectiv 1, bile albe. Estimați valoarea lui p cu metoda momentelor, respectiv metoda verosimilității maxime.

2. i) O caracteristică cercetată X are funcția de densitate

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

unde $\lambda > 0$ este fixat. Estimați parametrul necunoscut λ cu metoda momentelor, respectiv metoda verosimilității maxime. Sunt estimatorii obținuți consistenți?

ii) Durata culorii roșii (în minute) X a unui anumit semafor are funcția de densitate f_X dată mai sus, cu parametrul $\lambda > 0$ necunoscut. Un taximetrist (curios din fire) a observat următoarele durate (în minute) ale culorii roșii pentru acest semafor: 1, $\frac{3}{2}$, 3, 2, 3, $\frac{5}{2}$, 1, 2. Aplicați metoda momentelor, respectiv metoda verosimilității maxime, pentru a estima valoarea lui λ , folosind datele furnizate de taximetrist.

Pentru rezolvarea următoarelor probleme de statistică, folosiți următoarele valori numerice calculate în Python:

```
[1]: from scipy.stats import norm, t, chi2
```

```
[2]: norm.ppf([0.025, 0.05, 0.95, 0.975])
```

```
[2]: array([-1.95996398, -1.64485363,  1.64485363,  1.95996398])
```

```
[3]: t.ppf([0.025, 0.05, 0.95, 0.975], 4), t.ppf([0.025, 0.05, 0.95, 0.975], 99)
```

```
[3]: (array([-2.77644511, -2.13184678,  2.13184678,  2.77644511]),  
      array([-1.98421695, -1.66039116,  1.66039116,  1.98421695]))
```

```
[4]: chi2.ppf([0.025, 0.05, 0.95, 0.975], 2), chi2.ppf([0.025, 0.05, 0.95, 0.975], 4)
```

```
[4]: (array([0.05063562, 0.10258659, 5.99146455, 7.37775891]),  
      array([ 0.48441856,  0.71072302,  9.48772904, 11.14328678]))
```

3. Considerăm următoarele date statistice pentru masa corporală a persoanelor dintr-o anumită populație: 71 kg; 68 kg; 77 kg; 69 kg; 65 kg. Presupunem că masa corporală este o caracteristică ce urmează distribuția normală.

a) Știind că varianța/dispersia masei corporale este 20, determinați un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.

b) Știind că varianța/dispersia masei corporale este necunoscută, determinați un interval de încredere bilateral cu nivelul de încredere 95% pentru media masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că media masei corporale este 75 kg.

c) Determinați un interval de încredere bilateral cu nivelul de încredere 95% pentru varianța masei corporale, apoi testați, cu probabilitatea de risc 5%, ipoteza că abaterea standard a masei corporale este 10 kg.

4. Într-un sondaj de opinie, suntem interesați de proporția p a persoanelor dintr-un anumit oraș care ar vota candidatul A împotriva candidatului B . 576 din participanții la sondaj au declarat că ar vota candidatul A , iar 324 din participanți au declarat că ar vota cu candidatul B .

a) Determinați un interval de încredere bilateral cu nivelul de încredere 95% pentru p .

b) Testați cu probabilitatea de risc 5% ipoteza că $p = 0,6$.

5. Un provider de internet își asigură clienții că viteza conexiunii la internet este în medie 250 Mbps între orele 20:00 și 22:00. Pe de altă parte, providerul susține că în acest interval orar conexiunea nu este stabilă, având o abatere standard de 40 Mbps. În urma unei selecții de 100 de clienți s-a constatat că valoarea mediei de selecție este 242 Mbps pentru viteza conexiunii între orele specificate.

i) Să se construiască un interval de încredere unilateral stâng cu nivelul de încredere 95% pentru media vitezei conexiunii.

ii) Să se testeze, cu nivelul de semnificație 5%, dacă media vitezei este cea pretinsă de provider.

6. O companie dorește înlocuirea unui sistem de frânare pentru un anumit tip de mașină cu unul nou, care să reducă semnificativ distanța de frânare. Media distanței de frânare pentru vechiul sistem este mai mare sau egală decât 50 m, pentru o viteză de 80 km/h pe ploaie. În urma testării a 100 de mașini cu noul sistem de frânare instalat, pentru o viteză de 80 km/h pe ploaie, s-a constatat că valoarea mediei de selecție este 49 m și că valoarea abaterii standard de selecție este 1 m pentru distanța de frânare a acestui eșantion.

i) Să se construiască un interval de încredere unilateral drept cu nivelul de încredere 95% pentru media distanței de frânare a noului sistem.

ii) Să se testeze cu probabilitatea de risc 5% dacă noul sistem de frânare este diferit de cel vechi.

7. Un sondaj de opinie are în vedere studiul a două caracteristici: F , genul de film preferat, și M , genul de muzică preferat. S-au obținut următoarele date statistice:

| Film \ Muzică | Rock | Pop |
|---------------|------|-----|
| | | |
| Acțiune | 25 | 45 |
| Comedie | 26 | 39 |
| Dramă | 39 | 26 |

Testați, cu nivelul de semnificație 5%, dacă F și M sunt independente.

8. Într-o urnă sunt 100 de bile. Fiecare bilă este colorată cu una din culorile: roșu, albastru, verde. În urma extragerii a 120 bile, cu returnare, s-au obținut 34 de bile roșii, 55 de bile albastre și 31 de bile verzi. Folosind un test statistic cu nivelul de semnificație 5%, se poate afirma că în urnă sunt 25 de bile roșii, 50 de bile albastre și 25 de bile verzi?