Várható értékre és szórásra vonatkozó próbák

3. rész

– egy MATLAB® alapú megközelítés –

Baja Zsolt, Vas Orsolya

Matematika és Informatika Intézet, Babeş–Bolyai Tudományegyetem, Kolozsvár, Románia

(bajazsolt98@gmail.com, vas.orsolya@yahoo.com)

11. labor / 2022. december 12–16.

- Az egymintás U- és T-tesztek esetén valamely normális eloszlású valószínűségi változó ismeretlen elméleti várható értékére fogalmaztunk meg hipotéziseket, amelyeket vagy intervallumbecsléssel, vagy a szignifikanciapróba alkalmazásával ellenőriztünk.
- A mostani célunk olyan statisztikai eszközt bíztosítani, amellyel az adott normális eloszlású valószínűségi változó ismeretlen elméleti szórására (vagy szórásnégyzetére) szerkeszthetünk megbízhatósági intervallumokat.
- Ennek érdekében tekintsük az alábbi tételt!

1. Tétel (Egymintás χ^2-próba valószínűségi változója)

Tekintsük az $X \sim \mathcal{N}(\mu, \sigma)$ valószínűségi változónak az n-elemű ($n \geq 2$) független $\{X_j\}_{j=1}^{n}$ mintavételét, ahol $\mu \in \mathbb{R}$ és $\sigma > 0$! Ekkor a

$$V_n = \frac{1}{\sigma^2} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2 = \frac{(n-1)\,\overline{\sigma}^2}{\sigma^2} \tag{1}$$

valószínűségi változó $\chi^2(n-1, 1)$-eloszlású, ahol $\overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$ továbbra is a becsült várható értéket, míg

$$\overline{\sigma}^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2$$

a becsült korrigált szórásnégyzetet jelöli.

- Az egymintás χ^2-próba alkalmazását az alábbi feladat megoldásából sajátíthatjuk el.

1. Feladat (Egymintás χ^2-próba)

Tekintsük az $X \sim \mathcal{N}(\mu, \sigma)$ valószínűségi változónak az n-elemű független $\{X_j\}_{j=1}^n$ mintavételét ($n \geq 2$), ahol a $\sigma > 0$ elméleti szórást ismeretlen paraméternek feltételezzük! Az $\alpha \in (0,1)$ szignifikanciaszint mellett döntsük el a

$$H_0 : \sigma^2 = \sigma_0^2$$

nullhipotézis helyességét a

$$H_k : \sigma^2 \neq \sigma_0^2 \text{ (kétoldali próba)},$$
$$H_j : \sigma^2 > \sigma_0^2 \text{ (jobb oldali próba)},$$
$$H_b : \sigma^2 < \sigma_0^2 \text{ (bal oldali próba)}$$

lehetséges alternatív hipotézisekkel szemben, ahol $\sigma_0 > 0$ egy általunk megfogalmazott tipp az ismeretlen σ szórásra. Ugyanakkor szerkesszünk mindhárom esetben megbízhatósági intervallumot az ismeretlen σ paraméterre, valamint alkalmazzuk a megfelelő szignifikanciapróbákat és implementáljuk is az így kapott hipotézisellenőrző módszert!

Megoldás

- A feladat megoldását az (1)-es $\chi^2(n-1,1)$-eloszlású valószínűségi változóra vezethetjük vissza.
- Ekkor a 2. labor anyaga alapján a $\chi^2(n-1,1)$-eloszlás sűrűségfüggvényét és eloszlásfüggvényét az

$$f_{\chi^2(n-1,1)}(x) = \begin{cases} \dfrac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} x^{\frac{n-3}{2}} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

illetve az

$$\begin{aligned} F_{\chi^2(n-1,1)}(x) &= \int_{-\infty}^{x} f_{\chi^2(n-1,1)}(t)\,\mathrm{d}t \\ &= \begin{cases} \dfrac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \displaystyle\int_0^x t^{\frac{n-3}{2}} e^{-\frac{t}{2}}\,\mathrm{d}t, & x > 0, \\ 0, & x \leq 0 \end{cases} \end{aligned}$$

alakra hozhatjuk.

Megoldás – folytatás

- Ezért a további számításainkat a $(0, +\infty)$ tartományra kell korlátoznunk.
- Ennek megfelelően, amikor a $H_0 : \sigma^2 = \sigma_0^2$ nullhipotézishez a $H_k : \sigma^2 \neq \sigma_0^2$ kétoldali alternatív hipotézist társítjuk, akkor a V_n változóra szerkesztett megbízhatósági intervallumot már nem feltételezhetjük az origóra nézve szimmetrikusnak, ahogy azt eddig tettük az egy- és kétmintás U-, illetve T-próbák esetén.
- Hasonlóan értelmetlen lenne a $H_0 : \sigma^2 = \sigma_0^2$ és $H_j : \sigma^2 > \sigma_0^2$ hipotézispárosítás (azaz jobb oldali próba) esetén azt feltételeznünk, hogy a V_n változót magába foglaló konfidenciaintervallum alsó végpontja a $-\infty$.

Megoldás – folytatás

- Mindez viszont könnyen orvosolható, ha az adott $\alpha \in (0,1)$ szignifikanciaszint által eredményezett

$$\begin{aligned}
\chi^2_{\frac{\alpha}{2}} &= F^{-1}_{\chi^2(n-1,1)}\left(\tfrac{\alpha}{2}\right),\\
\chi^2_{1-\frac{\alpha}{2}} &= F^{-1}_{\chi^2(n-1,1)}\left(1-\tfrac{\alpha}{2}\right),\\
\chi^2_{1-\alpha} &= F^{-1}_{\chi^2(n-1,1)}(1-\alpha),\\
\chi^2_{\alpha} &= F^{-1}_{\chi^2(n-1,1)}(\alpha)
\end{aligned} \tag{2}$$

kvantilisek segítségével a két-, a jobb és a bal odali próbák alkalmazásakor a V_n valószínűségi változóra rendre az alábbi megbízhatósági intervallumokat fogalmazzuk meg:

  - $V_n \in \left(v^k_{\min}, v^k_{\max}\right) = \left(\chi^2_{\frac{\alpha}{2}}, \chi^2_{1-\frac{\alpha}{2}}\right)$;
  - $V_n \in \left(v^j_{\min}, v^j_{\max}\right) = \left(0, \chi^2_{1-\alpha}\right)$;
  - $V_n \in \left(v^b_{\min}, v^b_{\max}\right) = \left(\chi^2_{\alpha}, +\infty\right)$.

Megoldás – folytatás

- Vegyük észre, hogy a kétoldali próba esetén az $\frac{\alpha}{2}$ és $1-\frac{\alpha}{2}$ valószínűségek a $(0,1)$ intervallumban az $\frac{1}{2}$ értékre nézve szimmetrikusak, továbbá a lehetetlen eseményt a 0-val és nem a $-\infty$ szimbólummal reprezentáltuk!
- Ugyanakkor a későbbi számítások és implementálás alkalmával ügyeljünk arra, hogy a (2)-es kvantilisek 2-es felső indexe nem hatványkitevőt jelöl, hanem egyszerűen csak a χ^2-eloszlásnévre utal!
- Felhasználva a V_n változó alakját és a fenti konfidenciaintervallumokat, egyszerű átalakításokkal az olvasó könnyen beláthatja, hogy a két-, a jobb, és a bal oldali χ^2-próbák esetén az ismeretlen elméleti σ szórásra rendre az alábbi intervallumbecsléseket kapjuk:
  - $\left(\sigma_{\text{min}}^k, \sigma_{\text{max}}^k\right) = \left(\overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}}}, \overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}}}}\right);$
  - $\left(\sigma_{\text{min}}^j, \sigma_{\text{max}}^j\right) = \left(\overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{1-\alpha}}}, +\infty\right);$
  - $\left(\sigma_{\text{min}}^b, \sigma_{\text{max}}^b\right) = \left(0, \overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{\alpha}}}\right).$

Megoldás – folytatás

- Amennyiben a $H_0 : \sigma^2 = \sigma_0^2$ nullhipotézist igaznak feltételezzük, akkor mindhárom próba esetén a
$$V_n^0 = \frac{(n-1)\,\overline{\sigma}^2}{\sigma_0^2}$$
érték közös, amit a próbák χ^2-értékének is nevezünk.
- A kétoldali szignifikanciapróbához tartozó p-érték kiszámításánál is ügyelnünk kell, hiszen az $f_{\chi^2(n-1,1)}$ sűrűségfüggvény alakja már nem szimmetrikus sem az oordináta-tengelyre, sem bármilyen más függőleges tengelyre nézve.
- Ezért az eddigi szignifikanciapróbákhoz képest, a jelen esetben, a kétoldali szignifikanciapróba p-értékét a
$$\begin{aligned} p_k &= 2\min\left\{P\left(V_n < V_n^0 \mid H_0\right), P\left(V_n > V_n^0 \mid H_0\right)\right\} \\ &= 2\min\left\{F_{\chi^2(n-1,1)}\left(V_n^0\right), 1 - F_{\chi^2(n-1,1)}\left(V_n^0\right)\right\} \end{aligned}$$
alakban határozhatjuk meg, a jobb és bal odali szignifikanciapróbák p-értékét pedig a már ismerős módon:
  - $p_j = P\left(V_n > V_n^0 \mid H_0\right) = 1 - F_{\chi^2(n-1,1)}\left(V_n^0\right)$;
  - $p_b = P\left(V_n < V_n^0 \mid H_0\right) = F_{\chi^2(n-1,1)}\left(V_n^0\right)$.

Megoldás – folytatás

- Minden $a \in \{k, j, b\}$ oldalú χ^2-próba esetén a már megszokott minta alapján dönthetünk: amennyiben $V_n^0 \in \left(v_{\min}^a, v_{\max}^a\right)$, vagy $\sigma_0 \in \left(\sigma_{\min}^a, \sigma_{\max}^a\right)$, akkor a $H_0 : \sigma^2 = \sigma_0^2$ nullhipotézist, máskülönben a $H_a : \sigma^2 \operatorname{op}_a \sigma_0^2$ alternatív hipotézist fogadjuk el.
- Hasonlóképpen, az $a \in \{k, j, b\}$ odalú szignifikanciapróba esetén, ha $p_a > \alpha$, akkor a $H_0 : \sigma^2 = \sigma_0^2$ nullhipotézist, ellenben a $H_a : \sigma^2 \operatorname{op}_a \sigma_0^2$ alternatív hipotézist fogadjuk el.
- A most vázolt, ismeretlen elméleti szórásra vonatkozó, egymintás χ^2-próba kétoldali esetét az 1. kódrészletben listáztuk.
- A forráskód befejzését kitűzött feladatként az olvasóra bízzuk. Vegyük észre, hogy a kódrészlet a χ^2-eloszlás eloszlásfüggvényét és annak inverzét közelítő chi2cdf, illetve chi2inv MATLAB®-beli függvényekre épül!

1. Kódrészlet. Az egymintás χ^2-négyzet próba részleges MATLAB®-beli implementálása

```
% ——————————
% Description
% ——————————
% The function performs a one—sample χ^2—test of the null hypothesis H_0 : σ^2 = σ_0^2 that the
% data in the vector X = {X_j}_{j=1}^n comes from an N(μ, σ) distribution, where the
% theoretical standard deviation σ > 0 is an unknown parameter.
%
% The test is performed against the alternative hypothesis specified by the input
% parameter tail.
%
% ————
% Input
% ————
% X = {X_j}_{j=1}^n — an independent and identically distributed sample of the normal
%                    distribution N(μ, σ), where the theoretical standard deviation σ > 0
%                    is unknown
% sigma_0          — denotes the guess σ_0 of the user for the unknown theoretical standard
%                    deviation σ
% alpha            — represents the significance level α ∈ (0, 1)
% tail             — a parameter that can be set either by one of the strings 'both',
%                    'right', 'left', or by using one of the numbers 0, 1, −1 (it determines
%                    the type of the alternative hypothesis)
%

```

Implementálás – II

Egymintás χ^2-próba az ismeretlen elméleti szórásnégyzetre

```
% ______
% Output
% ______

% ci_chi2     - confidence interval for the random variable $V_n = \frac{(n-1)\,\overline{\sigma}^2}{\sigma^2}$
% ci_std      - confidence interval for standard deviation $\sigma$
% chi2_value - value of the test, assuming that the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ is true
% p_value     - probability of observing the given result, or one more extreme, by
%               chance if the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ is true (small $p$-values cast
%               doubt on the validity of the null hypothesis)
% H           - the code of the accepted hypothesis (0 = null hypothesis,
%               1 = alternative hypothesis defined by the input parameter tail)
%
function [ci_chi2, ci_std, chi2_value, p_value, H] = Chi2Test(X, sigma_0, alpha, tail)

% check the validity of input parameters!
...

% get the size of the sample $\{X_j\}_{j=1}^n$
n = length(X);

% calculate the sample variance $\overline{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$
sigma_2_ = var(X);

% calculate the chi2-value $V_n^0 = \frac{(n-1)\,\overline{\sigma}^2}{\sigma_0^2}$
```

Implementálás – III

Egymintás χ^2-próba az ismeretlen elméleti szórásnégyzetre

```
chi2_value = (n-1) * sigma_2_ / sigma_0^2;

% calculate the confidence intervals
switch (tail)
    case {'both', 0}
      % determine the confidence interval $\left(\chi^2_{\frac{\alpha}{2}} = F^{-1}_{\chi^2(n-1,1)}\left(\frac{\alpha}{2}\right), \chi^2_{1-\frac{\alpha}{2}} = F^{-1}_{\chi^2(n-1,1)}\left(1-\frac{\alpha}{2}\right)\right)$
      ci_chi2(1) = chi2inv(alpha/2, n-1);
      ci_chi2(2) = chi2inv(1 - alpha/2, n-1);

      % store the end points of the confidence interval $\left(\overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}}}, \overline{\sigma}\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}}}}\right)$
      ci_std(1) = sqrt((n - 1) * sigma_2_ / ci_chi2(2));
      ci_std(2) = sqrt((n - 1) * sigma_2_ / ci_chi2(1));

        % calculate the p-value $p = 2\min\left\{F_{\chi^2(n-1,1)}\left(V_n^0\right), 1 - F_{\chi^2(n-1,1)}\left(V_n^0\right)\right\}$
      probability = chi2cdf(chi2_value, n-1);
      p_value = 2.0 * min(probability, 1 - probability);

    case {'right', 1}
      ...

    case {'left', -1}
      ...
end
```

Implementálás – IV

Egymintás χ^2-próba az ismeretlen elméleti szórásnégyzetre

```
% make your decision based on confidence intervals, and note that, the
% condition (sigma_0 > ci_std(1) && sigma_0 < ci_std(2)) would also be correct!
H = ~(chi2_value > ci_chi2(1) && chi2_value < ci_chi2(2));

% do you have any doubt? -- if yes, then apply the corresponding significance test!
if (p_value < alpha)
    disp('Warning: small p-value cast doubt on the validity of the null-hypothesis!');
end
```

- Emlékezzünk arra, hogy a kétmintás T-próba alkalmazása során két ismeretlen paraméterezésű, független normális eloszlás elméleti várható értékét hasonlítottuk össze független mintavételi adatok alapján úgy, hogy két esetet különböztettünk meg aszerint, hogy az eloszlások ismeretlen elméleti szórását megegyezőnek, vagy különbözőnek feltételeztük!
- Mi több, az esetszétválasztás a kétmintás T-próbát implementáló TTest2D függvény bemenő paraméterlistájára is kihatott egy felhasználó által önkényesen állítható equal_std nevű logikai változó formájában (lásd a 10. labor 2. kódrészletét).
- A függvényt felhasználó számára egyrészt sokkal kényelmesebb, másrészt adott szignifikanciaszint mellett biztonságosabb is lenne, ha a két ismeretlen elméleti szórás megegyezőségét találgatások helyett valamilyen statisztikai próbával döntenénk el.
- Amennyiben találnánk ilyen módszert, akkor megszabadulhatnánk a hibalehetőséget rejtő equal_std nevű logikai változótól.
- Amint hamarosan látni fogjuk, az alábbi tételben bevezetett valószínűségi változót a fenti ismeretlen elméleti szórások összehasonlítására használhatjuk.

2. Tétel (Az F-próba valószínűségi változója)

Tekintsük az $X \sim \mathcal{N}(\mu_1, \sigma_1)$ és $Y \sim \mathcal{N}(\mu_2, \sigma_2)$ független valószínűségi változókat, valamint ezeknek egy-egy független $\{X_i\}_{i=1}^m$ és $\{Y_j\}_{j=1}^n$ mintavételét ($m \geq 2$, $n \geq 2$), ahol $\mu_1, \mu_2 \in \mathbb{R}$ és $\sigma_1, \sigma_2 > 0$! Ekkor az

$$F_{m,n} = \frac{\overline{\sigma}_1^2}{\sigma_1^2} \cdot \frac{\sigma_2^2}{\overline{\sigma}_2^2} \tag{3}$$

valószínűségi változó $(m-1, n-1)$-szabadságfokú Snedecor–Fisher-eloszlást követ.

2. Feladat (Kétmintás F-próba)

Tekintsük az $X \sim \mathcal{N}(\mu_1, \sigma_1)$ és $Y \sim \mathcal{N}(\mu_2, \sigma_2)$ független valószínűségi változókat, valamint ezeknek egy-egy független $\{X_i\}_{i=1}^{m}$ és $\{Y_j\}_{j=1}^{n}$ mintavételét ($m \geq 2$, $n \geq 2$), ahol a $\mu_1, \mu_2 \in \mathbb{R}$ várható értékek tetszőlegesek, továbbá a $\sigma_1, \sigma_2 > 0$ szórásokat nem ismerjük! Legyen $\lambda = \dfrac{\sigma_1}{\sigma_2}$! Az $\alpha \in (0,1)$ szignifikanciaszint mellett döntsük el a

$$H_0 : \lambda = 1$$

nullhipotézis helyességét a

$$H_k : \lambda^2 \neq 1 \text{ (kétoldali próba)},$$

$$H_j : \lambda^2 > 1 \text{ (jobb oldali próba)},$$

$$H_b : \lambda^2 < 1 \text{ (bal oldali próba)}$$

lehetséges alternatív hipotézisekkel szemben! Ugyanakkor szerkesszünk mindhárom esetben megbízhatósági intervallumot az ismeretlen elméleti szórások λ arányára, határozzuk meg az egyes esetekhez tartozó szignifikanciapróbák p-értékét, továbbá implementáljuk is ezt a statisztikai módszert!

Megoldás

- Mindenekelőtt vegyük észre, hogy a null- és a lehetséges alternatív hipotézisek pontosan azt fejezik ki, hogy az ismeretlen σ_1 és σ_2 elméleti szórások megegyeznek (H_0), különböznek (H_k), $\sigma_1 > \sigma_2$ (H_j), illetve $\sigma_1 < \sigma_2$ (H_b)! Tehát az alábbiakban vázolt kétmintás F-próba esetén valóban az ismeretlen elméleti szórások összehasonlításáról beszélhetünk.
- A feladat megoldásához a (3)-as $\mathcal{F}(m-1, n-1)$-eloszlású valószínűségi változót használjuk fel.
- Vegyük észre, hogy a $\lambda = \dfrac{\sigma_1}{\sigma_2}$ arány segítségével a változót az

$$F_{m,n} = \frac{1}{\lambda^2} \cdot \frac{\overline{\sigma}_1^2}{\overline{\sigma}_2^2}$$

  alakra hozhatjuk!
- A 2. labor 39. oldalán bevezetett Snedecor–Fisher-eloszlás sűrűség- és eloszlásfüggvényét tanulmányozva, könnyen beláthatjuk, hogy csak a $(0, +\infty)$ intervallumon dolgozhatunk, azaz egy teljesen hasonló jelenségbe ütközünk, mint az 1. feladatban leírt egymintás χ^2-próba esetén.

Megoldás – folytatás

- Ezek szerint az

$$\begin{aligned} f_{\frac{\alpha}{2}} &= F^{-1}_{\mathcal{F}(m-1,n-1)}\left(\tfrac{\alpha}{2}\right), \\ f_{1-\frac{\alpha}{2}} &= F^{-1}_{\mathcal{F}(m-1,n-1)}\left(1-\tfrac{\alpha}{2}\right), \\ f_{1-\alpha} &= F^{-1}_{\mathcal{F}(m-1,n-1)}\left(1-\alpha\right), \\ f_{\alpha} &= F^{-1}_{\mathcal{F}(m-1,n-1)}\left(\alpha\right) \end{aligned} \tag{4}$$

kvantilisek segítségével, a két-, a jobb és bal oldali F-próbák esetén az $F_{m,n}$ valószínűségi változóra rendre az

- $F_{m,n} \in \left(f^k_{\min}, f^k_{\max}\right) = \left(f_{\frac{\alpha}{2}}, f_{1-\frac{\alpha}{2}}\right)$;
- $F_{m,n} \in \left(f^j_{\min}, f^j_{\max}\right) = \left(0, f_{1-\alpha}\right)$;
- $F_{m,n} \in \left(f^b_{\min}, f^b_{\max}\right) = \left(f_{\alpha}, +\infty\right)$

megbízhatósági intervallumokat kapjuk, amelyek alapján könnyen adódnak az ismeretlen elméleti szórások λ arányára is az alábbi intervallumbecslések:

Megoldás – folytatás

- $\left(\lambda^k_{min}, \lambda^k_{max}\right) = \left(\frac{1}{\sqrt{f_{1-\frac{\alpha}{2}}}} \cdot \frac{\overline{\sigma}_1}{\overline{\sigma}_2}, \frac{1}{\sqrt{f_{\frac{\alpha}{2}}}} \cdot \frac{\overline{\sigma}_1}{\overline{\sigma}_2}\right)$;
- $\left(\lambda^j_{min}, \lambda^j_{max}\right) = \left(\frac{1}{\sqrt{f_{1-\alpha}}} \cdot \frac{\overline{\sigma}_1}{\overline{\sigma}_2}, +\infty\right)$;
- $\left(\lambda^b_{min}, \lambda^b_{max}\right) = \left(0, \frac{1}{\sqrt{f_{\alpha}}} \cdot \frac{\overline{\sigma}_1}{\overline{\sigma}_2}\right)$.

- Ha a $H_0 : \lambda^2 = 1$ nullhipotézist igaznak feltételezzük, akkor megkaphatjuk az F-próba

$$F^0_{m,n} = \frac{1}{1} \cdot \frac{\overline{\sigma}^2_1}{\overline{\sigma}^2_2} = \frac{\overline{\sigma}^2_1}{\overline{\sigma}^2_2}$$

  f-értékét.

- Amennyiben a két-, a jobb, vagy a bal odali szignifikanciapróbát alkalmazzuk, akkor rendre az alábbi p-értékeket kell figyelembe vennünk:
  - $p_k = 2\min\left\{F_{\mathcal{F}(m-1,n-1)}\left(F^0_{m,n}\right), 1 - F_{\mathcal{F}(m-1,n-1)}\left(F^0_{m,n}\right)\right\}$
  - $p_j = 1 - F_{\mathcal{F}(m-1,n-1)}\left(F^0_{m,n}\right)$;
  - $p_b = F_{\mathcal{F}(m-1,n-1)}\left(F^0_{m,n}\right)$.

Megoldás – folytatás

- Bármely $a \in \{k, j, b\}$ oldalú F-próba esetén a már megszokott módon dönthetünk: amennyiben $F_{m,n}^0 \in \left(f_{\min}^a, f_{\max}^a\right)$, vagy $1 \in \left(\lambda_{\min}^a, \lambda_{\max}^a\right)$, akkor a $H_0 : \lambda^2 = 1$ nullhipotézist, máskülönben a $H_a : \lambda^2 \operatorname{op}_a 1$ alternatív hipotézist fogadjuk el.
- Az $a \in \{k, j, b\}$ odalú szignifikanciapróba esetén, ha $p_a > \alpha$, akkor a $H_0 : \lambda^2 = 1$ nullhipotézist, ellenben a $H_a : \lambda^2 \operatorname{op}_a 1$ alternatív hipotézist fogadjuk el.
- A fent vázolt statisztikai módszer kódolását kitűzött feladatként az olvasóra bízzuk! Segítségképpen a 2. kódrészletben megadtuk az implementálandó függvény leírását és fejlécét.

2. Kódrészlet. A kétmintás F-próba függvényének fejléce

```
% ---------------------------------------------------------------------
% Description
% ---------------------------------------------------------------------
% The function performs a two-sample F-test of the null hypothesis $H_0 : \lambda^2 = \frac{\sigma_1^2}{\sigma_2^2} = 1$
% that the data in the two independent samples $\mathbf{X} = \{X_i\}_{i=1}^m$ and $\mathbf{Y} = \{Y_j\}_{j=1}^n$, of
% random variables $X \sim \mathcal{N}(\mu_1, \sigma_1)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2)$, come from distributions
% with equal theoretical standard deviations.
%
% The test is performed against the alternative hypothesis specified by the input
% parameter tail.
%
% ---------------------------------------------------------------------
% Input
% ---------------------------------------------------------------------
% $\mathbf{X} = \{X_i\}_{i=1}^m$ - an independent and identically distributed sample of the normal
%               distribution $\mathcal{N}(\mu_1, \sigma_1)$, where $\sigma_1 > 0$ is unknown
% $\mathbf{Y} = \{Y_j\}_{j=1}^n$ - an independent and identically distributed sample of the normal
%               distribution $\mathcal{N}(\mu_2, \sigma_2)$, where $\sigma_2 > 0$ is unknown
% alpha       - represents the significance level $\alpha \in (0,1)$
% tail        - a parameter that can be set either by one of the strings 'both',
%               'right', 'left', or by using one of the numbers 0, 1, -1 (it determines
%               the type of the alternative hypothesis)
%
```

Fejléc – II

Kétmintás F-próba az ismeretlen elméleti szórásnégyzetek összehasonlítására

```
% ----------
% Output
% ----------
% ci_f        - confidence interval for the random variable F_{m,n} = 1/lambda^2 * overline{sigma}_1^2 / overline{sigma}_2^2
%
% ci_lambda   - confidence interval for the ratio lambda = sigma_1 / sigma_2 of the theoretical standard
%               deviations
% f_value     - value of the test, assuming that the null hypothesis H_0 : lambda^2 = 1 is true
% p_value     - probability of observing the given result, or one more extreme,
%               by chance if the null hypothesis H_0 : lambda^2 = 1 is true (small p-values
%               cast doubt on the validity of the null hypothesis)
% H           - the code of the accepted hypothesis (0 = null hypothesis,
%               1 = alternative hypothesis defined by the input parameter tail)
%
function [ci_f, ci_lambda, f_value, p_value, H] = FTest2D(X, Y, alpha, tail)
```

1. feladat

- Egy prémium minőségű szarvasgombás olajokat gyártó vállalat termékeinek vizsgálata során az ellenőrök véletlenszerűen kiválasztottak néhány 100 milliliteres üveget, majd lemérték azok olajtartalmát. A milliliterben kifejezett értékeket az

$$X = [100.8, 102.5, 98.2, 97.5, 99.1, 99.4, 100.9, 95.6, 99.3, 99.1, 98.3, 99.6, 96.2, 99.0, 100.8, 97.5, 99.3, 97.2, 98.7, 98.2, 99.0, 98.6, 98.8, 97.3, 100.6, 99.3, 96.5]$$

  tömb tartalmazza.
- Tudva, hogy az üvegekben levő szarvasgombás olaj mennyisége normális eloszlást követ, 96%-os valószínűség mellett döntsétek el, hogy megvizsgált üvegekben levő szarvasgombás olaj mennyiségének elméleti hibája meghaladja-e az előírt maximális 1.2 milliliteres értéket!

2. feladat

- A Fogyasztóvédelmi Hivatal emberei a karácsony előtti razzia alkalmával ellenőrzést tartanak két cukrászdában. A vizsgálat során találomra kiválasztanak néhány 750 grammos karácsonyi süteményválogatást tartalmazó dobozt és lemérik azok súlyát. A mérések grammban kifejezett eredményét az

$$\begin{aligned} X = [&750, 753, 747, 748, 751, 748, 751, 753, 752, 749, 750, \\ &753, 748, 752, 748, 752, 749, 751, 750, 749], \\ Y = [&746, 753, 747, 751, 751, 747, 748, 745, 754, 747, 744, \\ &748, 750, 747, 748, 746, 754, 752, 745] \end{aligned}$$

  tömbök tartalmazzák.
- Tudva, hogy a süteményes dobozok súlya normális eloszlást követ, döntsétek el az $\alpha = 0.05$ szignifikanciaszint mellett, hogy a két cukrászda esetén a süteményes dobozok súlyának ismeretlen elméleti szórása különbözik-e, vagy sem!
- Az előző alpont eredményét felhasználva, döntsétek el 91%-os biztonsággal, hogy az első cukrászda esetén a süteményes dobozok átlagos súlya nagyobb-e, mint a második cukrászda esetén!