

KHI-NÉGYZET PRÓBA EXCEL VBA ŰRLAPJÁNAK KIALAKÍTÁSA

Hampel György

Abstract: Normális eloszlású sokaság szórására vonatkozó hipotézis tesztelésére a khi-négyszet próbát kell alkalmazni. Ennek végrehajtását teszi egyszerűvé és jól áttekinthetővé a Microsoft Excel Visual Basic for Applications szolgáltatásával elkészült alkalmazás, melynek fejlesztése kerül bemutatásra. Egyetlen űrlap biztosítja az adatok megadását, valamint az eredmények megjelenítését az űrlapon elhelyezett vezérlőkkel. A program eseményvezérelt programozási technikával készült el, emiatt felhasználóbarát módon reagál a használat során bekövetkező eseményekre. Matematikai statisztikai ismeretek nélkül is egyszerűen kezelhető a program. A tesztelendő hipotézis olyan listából választható ki, melyben minden lehetséges változat elérhető. Az adatok megadásához kiválasztható, hogy a statisztikai minta teljes adatsorát adnánk meg, vagy csak a minta azon jellemzőit, melyek a kiértékeléshez szükségesek. Az eredmények könnyen értelmezhetőek, a számított eredményeken túl érzékeltetett válaszként is megjelenik a tesztelendő hipotézisre adott válasz.

Abstract: The chi-square test should be used to test the hypothesis about the variance of a normally distributed population. The implementation of this is made simple and clear by the application created with the Microsoft Excel Visual Basic for Applications service, the development of which is presented. A single form ensures the entry of data and the display of results with the controls placed on the form. The program was created using an event-driven programming technique, so it reacts to events that occur during use in a user-friendly way. The program is easy to use even without knowledge of mathematical statistics. The hypothesis to be tested can be selected from a list in which all possible variants are available. To enter the data, it is possible to choose whether to enter the entire data set of the statistical sample or only those characteristics of the sample that are necessary for the evaluation. The results are easy to interpret, in addition to the calculated results, the answer to the hypothesis to be tested is also displayed as a full-sentence answer.

Kulcsszavak: matematikai statisztika, khi-négyszet próba, Excel VBA, programozás

Keywords: mathematical statistics, chi-square test, Excel VBA, programming

1. Bevezetés

Korunk egyik nagy kihívása, hogy hatalmas adatmennyiséget képzünk nem csak a tudományos tevékenységek során, de szinte minden területen. Már a sportesemények is elképzelhetetlenek statisztikai kimutatásokkal jellemzésük nélkül. Egyre több kutatás foglalkozik az adatok tárolásához kialakított adatbázisokkal, a hozzá tartozó szolgáltatások fejlesztésével (Fabulya, 2018a; Fabulya, 2018b). Az adatok feldolgozásához matematikai statisztikai technikákat alkalmazva lesz lehetőségünk hasznosítani a tárolt adatban megjelenő információt (Móri, 2011; Fabulya, 2017).

Jelen kutatás egy nagyobb projekt része, melynek célja matematikai statisztikai eljárások fejlesztése a Microsoft Excel táblázatkezelő program eseményvezérelt programozási technikát biztosító Visual Basic for Applications (VBA) szolgáltatásával. Egyszerűen kialakítható az alkalmazás megjelenését és kezelését nyújtó felület párbeszédablaka, melyen vezérlőket elhelyezve, viselkedésüket programozással tehetjük könnyen használhatóvá. Az Excelt számos kutatásban alkalmazzák nem csak az adatok kiértékelését és grafikonon megjelenítését támogató képességei, de a programozási lehetőségei miatt is (Fabulya, 2019). Elkészíthető

pénzügyi tervek kiértékelése (Zsótér, 2017), beruházások gazdaságossági elemzése (Zsótér–Túri, 2017), piackutatási eredmények kiértékelése (Zsótér–Kaliczka, 2014, Zsótér, 2007), de még dokumentumok generálását is automatizálhatjuk (Fabulya, 2020).

2. Anyag és módszer

2.1. A khi-négyzet próba

A matematikai statisztika a khi-négyzet próba elnevezést több hipotézisvizsgálati eljárásra is alkalmazza:

- paraméteres próba sokasági szórásra,
- függetlenségvizsgálat,
- illeszkedésvizsgálat,
- homogenitásvizsgálat.

Ezek közül a sokasági szórásra alkalmazott próbához alakítottunk ki alkalmazást, így ennek matematikai statisztikai hátterét tekintjük át. A próbát akkor alkalmazhatjuk, ha az adataink numerikus típusúak és a statisztikai sokaság normális eloszlású (Obádovics, 2020). A próba végrehajtásának menete megegyezik a többi paraméteres próbáéval (Hampel, 2022):

- Elsőfajú hibavalószínűség (ε) választása.
- Null (H_0) és alternatív hipotézis (H_1) felállítása.
- A próba statisztikai függvény értékének (χ^2) kiszámítása.
- A kritikus tartomány határának (χ_p^2) vagy a próba szignifikanciaszintjének (p) meghatározása.
- A döntés meghozatala a hipotézis elfogadásáról vagy elvetéséről.

Az elsőfajú hibavalószínűség értéknek leggyakrabban 0,05 vagy 0,01 értéket szoktak alkalmazni, mert ekkor nagyon kis valószínűséggel követünk el elsőfajú hibát a döntéskor. A null és alternatív hipotézis felállítása az alapján történik, hogy a vizsgálandó hipotézisben milyen reláció szerepel a sokasági szórás (σ) feltételezett értékére (σ_0). A különböző esetekben alkalmazandó null és alternatív hipotézist az 1. táblázat mutatja.

1. táblázat: Null és alternatív hipotézis

Vizsgálandó hipotézis	H_0	H_1
$\sigma = \sigma_0$	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$
$\sigma \neq \sigma_0$	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$
$\sigma > \sigma_0$	$\sigma = \sigma_0$	$\sigma > \sigma_0$
$\sigma < \sigma_0$	$\sigma = \sigma_0$	$\sigma < \sigma_0$
$\sigma \geq \sigma_0$	$\sigma = \sigma_0$	$\sigma < \sigma_0$
$\sigma \leq \sigma_0$	$\sigma = \sigma_0$	$\sigma > \sigma_0$

Forrás: a szerző saját szerkesztése.

A khi-négyzet próba statisztikai függvénye az (1) képletben látható.

$$\chi^2 = \frac{(n - 1) \cdot s^2}{\sigma_0^2} \quad (1)$$

ahol:

χ^2 a statisztikai függvény értéke

s a minta szórása

σ_0 a statisztikai sokaság feltételezett szórása

n a minta elemszáma

A képlet szerint a számításhoz elegendő a minta két jellemzője, de ha a minta szórását nem ismernénk, akkor ezt a minta adataiból kell meghatározni a (2) képlettel. Ekkor viszont a minta átlagának kiszámítására is szükségünk lehet a (3) képlet alapján.

$$s = \sqrt{\frac{1}{n - 1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (3)$$

ahol:

s a minta szórása

\bar{x} a minta átlaga

x_i az i -edik mintaelem értéke

n a minta elemszáma

A döntés a kritikus tartománnyal vagy a szignifikanciaszinttel hozható meg. Az ellenhipotézistől függ, hogy ezek hogyan határozhatók meg. A kiszámítás a 2. táblázatban látható, ahol az F a khi-négyzet eloszlásnak az eloszlásfüggvénye a megfelelő szabadságfok mellett, míg χ^2_p a kritikus tartomány határa egy p valószínűség esetén. Az ellenhipotézis a kritikus tartományba eső χ^2 érték esetén, vagy a $p < \varepsilon$ feltétel teljesülésekor fogadható el, amikor a szignifikanciaszint kisebb, mint az elsőfajú hibavalószínűség (Michaletzky–Mogyoródi, 1995).

2. táblázat: A döntés feltétele és a szignifikanciaszint kiszámítása

H_1	H_1 elfogadási feltétele kritikus tartománnyal	Szignifikanciaszint (p)
$\sigma \neq \sigma_0$	$\chi^2 > \chi^2_{1-\frac{\varepsilon}{2}}$ vagy $\chi^2 < \chi^2_{\frac{\varepsilon}{2}}$	$p = \begin{cases} 2 \cdot (1 - F(\chi^2)) & \text{ha } s > \sigma_0 \\ 2 \cdot F(\chi^2) & \text{ha } s < \sigma_0 \end{cases}$
$\sigma < \sigma_0$	$\chi^2 < \chi^2_{\varepsilon}$	$p = F(\chi^2)$
$\sigma > \sigma_0$	$\chi^2 > \chi^2_{1-\varepsilon}$	$p = 1 - F(\chi^2)$

Forrás: a szerző saját szerkesztése.

A khi-négyzet eloszlás eloszlásfüggvénye (F) egy valószínűséget eredményez a (4) képlet összefüggése alapján.

$$P(\chi^2 < \chi^2_p) = F(\chi^2_p) = p \quad (4)$$

ahol:

χ^2 egy khi-négyzet eloszlású valószínűségi változó

χ^2_p a khi-négyzet eloszlás p valószínűséghez tartozó kvantilise

F a khi-négyzet eloszlás eloszlásfüggvénye

2.2. A Microsoft Excel programozhatósági lehetősége

Eseményvezérelt programozási technikát alkalmazhatunk a Visual Basic for Applications (VBA) szolgáltatással. A program kezelőfelülete vezérlőkkel alakítható ki (Matteson, 1995). A VBA programozási nyelven a szekvenciális, szelekciós és iterációs programszerkezet egyszerűen megvalósítható (Zimmerman, 1996). A kezelőfelület elkészítése grafikusán történik a vezérlők elhelyezésével a program párbeszédablakán (Kovalcsik, 2005).

3. Eredmények és értékelésük

A khi-négyzet próbát kiértékelő alkalmazás elkészítésekor a következő szempontok valósultak meg:

- Az alkalmazás kezelése egyetlen párbeszédablakon történik. A programot tartalmazó Excel fájl megnyitásakor ez a kezelő felület automatikusan megjelenik.
- Az adatbevitel a párbeszédablak egyik lapján történik, míg az eredmények a másik lapon látható.
- Lehetőség van a minta jellemzőinek (elemszám, szórás) megadására a számításokhoz, de jelölőnégyzettel kiválasztható a minta adatsorának bevitele a jellemzők hiányában.
- A kiértékelendő hipotézis megadását legördülő lista biztosítja, melyben minden reláció elérhető.

- Az eredmények a felhasználótól rejtett Excel munkalapon kialakított formulákkal képződnek, de a párbeszédablakon jelennek meg.

3.1. Az adatbevitel párbeszédablaka

A programot tartalmazó Excel fájl megnyitásakor automatikusan megjelenik a kezelőfelület párbeszédablaka, mely az 1. ábrán látható. Ehhez eseményvezérelt VBA programot kellett készíteni, mely a beállított esemény (Workbook_Open) bekövetkezésekor fut le, és megjeleníti a párbeszédablakot (UserForm1) az alábbi eljárásban szereplő VBA kóddal:

```
Private Sub Workbook_Open()
    UserForm1.Show
End Sub
```

1. ábra: Az alkalmazás felhasználói felülete

Forrás: a szerző saját szerkesztése

Az Adatok lapon a minta két jellemző értékét, elemszámát és szórását elegendő megadni a számításokhoz, ha ezek ismertek lennének. Legördülő listából választható

ki a hipotézis típusa, míg beviteli mezőben adható meg az elsőfajú hibaválósínúség értéke (ε) és a sokasági szórás feltételezett értéke (σ_0).

Ha a minta jellemzőit nem ismerjük, akkor az adatsort szükséges megadnunk. Ezt a jelölőnégyzet értékének megváltozásakor (CheckBox1_Change) lefutó program teszi lehetővé, megjelenítve a szükséges vezérlőket az alábbi VBA programmal:

```
Private Sub CheckBox1_Change()
    Call kinezet
End Sub
Public Sub kinezet()
    If CheckBox1 Then
        Frame1.Visible = True
        Frame2.Visible = False
    Else
        Frame1.Visible = False
        Frame2.Visible = True
    End If
End Sub
```

A fenti programrész végrehajtásakor a jelölőnégyzet állapotának megfelelő keret (Frame) lesz látható, a keretben kialakított vezérlőkkel. Így lecserélhető a mintajellemzők megadásának kerete az adatsor keretére, mely a 2. ábrán látható.

2. ábra: A minta adatsorának megadása

Forrás: a szerző saját szerkesztése

Amikor befejeztük az adatok megadását, akkor válthatunk egy Multipage vezérlővel az *Adatok* és az *Eredmények* lapok között.

3.2. Az eredmények megjelenítése

A számítások az Excel táblázatkezelő programnak egy munkalapján kialakított formulákkal valósul meg, a felhasználótól rejtett módon. Ehhez a program párbeszédablakán lévő adatokat át kell helyezni a munkalapra, ahol automatikusan képződnek az eredmények. A munkalap egy cellája értékadó utasítással veszi fel a párbeszédablak egy vezérlőjében tárolt értéket, az alábbi típusú VBA utasítás hatására:

```
Cells(4, 31) = TextBox3.Text
```

Ekkor a munkalap 4. sorának 31. cellájába kerül a `TextBox3` beviteli mező értéke. A munkalapon kiszámított eredmények a párbeszédablakon lesznek láthatók olyan VBA utasítások hatására, melyek egy cella értékét adják át egy felirat (`Label`) vezérlőnek:

```
Label7.Caption = Cells(8, 13).Text
```

A kiszámított eredmények a legfontosabb adatokkal és hipotézisekkel együtt jelennek meg az Eredmények lapon. A jobb értelmezhetőség érdekében egészmondatos válaszok is olvashatók, ahogy ez a 3. ábrán látható.

3. ábra: Az eredmények lap

Khi-négyzet próba szórásra

Adatok | Eredmények

Vizsgálandó hipotézis: $\sigma \leq 0,5$
A próba hipotézisei:
 $H_0: \sigma = 0,5$
 $H_1: \sigma > 0,5$

$$\chi^2 = \frac{(n - 1) \cdot s^2}{\sigma_0^2}$$

Adatok	Számítások
$\varepsilon = 0,05$	$\chi^2 = 15,3673$
$\sigma_0 = 0,5$	$\chi^2(1-\varepsilon) = 18,307$
$n = 11$	$p = 0,1192$
$s = 0,6198$	

H_1 elutasítva. ($\sigma > 0,5$ hamis)
 $\varepsilon = 0,05$ esetén az $\sigma \leq 0,5$ elfogadható.
A próba szignifikancia szintje: $p = 0,1192$

Kilépés

4. Következtetések

A Microsoft Excel táblázatkezelő program Visual Basic for Applications szolgáltatásával felhasználóbarát módon kezelhető alkalmazás fejleszthető sokasági szórásra vonatkozó khi-négyzet próba kiértékelésére. Így matematikai statisztikai ismeretek nélkül végezhető el a számítás, míg az eredmények könnyen értelmezhető formában jelennek meg.

Irodalomjegyzék

- Fabulya Z. (2017): Hőkezelési folyamatok összehangolása Excel VBA szolgáltatásokkal. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 12(4), 19–25. <https://doi.org/10.14232/jtgf.2017.4.19-25>
- Fabulya Z. (2018a): Access alkalmazás kialakítása ügyfélközpontú szolgáltatások nyilvántartására. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 13(1-2), 67–76. <https://doi.org/10.14232/jtgf.2018.1-2.67-76>
- Fabulya Z. (2018b): Access alkalmazás kialakítása dolgozói jelenlét nyilvántartására. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 13(3-4), 151–160. <https://doi.org/10.14232/jtgf.2018.3-4.151-160>
- Fabulya Z. (2019): Excel VBA függvények kialakítása háromdimenziós vektorok matematikai alkalmazására. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 14(1), 29–34. <https://doi.org/10.14232/jtgf.2019.1.29-34>
- Fabulya Z. (2020): VBA program fejlesztése feladatsorok dokumentumainak generálására. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 15(3-4), 139–143. <https://doi.org/10.14232/jtgf.2020.3-4.139-143>
- Hampel Gy. (2022): Alkalmazás fejlesztése egymintás u-próba kiértékelésére VBA vezérlőkkel Excel környezetben. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 17(1-2), 119–126. <https://doi.org/10.14232/jtgf.2022.1-2.119-126>
- Kovalcsik G. (2005): *Az Excel programozása*. Computerbooks, Budapest.
- Matteson, B. L. (1995): *Microsoft Excel Visual Basic Programmer's Guide*. MicrosoftPress, Washington.
- Michaletzky Gy., Mogyoródi J. (1995): *Matematikai statisztika*. Nemzeti Tankönyvkiadó, Budapest.
- Móri T. (2011): *Statisztikai hipotézisvizsgálat*. Typotex Kft., Budapest.
- Obádovics J. Gy. (2020): *Valószínűségszámítás és matematikai statisztika*, Scolar Kiadó Kft., Budapest. ISBN: 9789632440675
- Zimmerman M. W. (1996): *Microsoft Office 97 Visual Basic Programmer's Guide*, MicrosoftPress, Washington.
- Zsótér B. (2007): A Hotel Nonius szolgáltatásait igénybe vevők földrajzi megoszlása. *Agrár- és Vidékfejlesztési Szemle*, 2 (2): 201–206.
- Zsótér B. (2017): Financial planning in connection with accomodation development in a sport centre. *Quaestus Multidisciplinary Research Journal*, 4 (11): 172–177.
- Zsótér B., Kaliczka Renáta (2014): Examinations carried out in relation to the shopping habits and satisfaction of costumers in the shops of Coop Szeged Ltd. *Review of Faculty of Engineering Analecta Technica Szegedinensia*, 8 (1): 38–41.
- Zsótér B., Túri I. (2017): Economical calculations related to a smoking technology investment of a pork processing plant. *Annals of Faculty of Engineering Hunedoara*, 15 (4): 57–61.