

# **PÁROS T-PRÓBA PROGRAMOZHATÓ KIALAKÍTÁSA EXCEL VBA KÖRNYEZETBEN**

Hampel György

**Abstract:** Az Excel táblázatkezelő programmal olyan számológépet tudunk kialakítani, melyeket többször is használhatunk. Egy-egy adat megváltozásakor automatikusan újraszámítódik a teljes tábla. Az adatok megváltoztatását programozott módon is elvégezhetjük a Visual Basic for Application (VBA) szolgáltatás segítségével. Összetett számítások esetén célszerű arra törekedni, hogy a kialakított számológép egyszerűen tegye lehetővé a paraméterek értékétől függő eredmények képzését.

Jelen publikáció a páros t-próba automatizált végrehajtása kialakításának egy lehetséges módját tárgyalja. Statisztikai programok nem teszik lehetővé a programozással automatizált kiértékeléseket, ezért az Excel VBA alkalmazása jelentősen felgyorsítja az ismételt elvégzendő számítások előkészítését és végrehajtását. A cikkben bemutatom egy könnyen és kényelmesen használható kezelőfelület elkészítését, valamint a hozzá tartozó VBA programokat.

**Abstract:** We can create spreadsheets in Excel that can be used several times. When the data changes, the entire table is automatically recalculated. The data change can be done in a programmed method using Visual Basic for Application (VBA). In case of complex calculations, it is advisable to develop a spreadsheet which easily allows producing the results depending on the values of parameters.

This publication discusses a possible way of creating an automated execution of the paired t-test. Statistical programs do not allow automatic evaluations with programming; therefore, the use of Excel VBA significantly speeds up the preparation and execution of calculations that have to be repeated. In this article, I present the preparation of an easy-to-use user interface and the associated VBA programs.

*Kulcsszavak:* Microsoft Excel, VBA, statisztikai kiértékelés, páros t-próba

*Keywords:* Microsoft Excel, VBA, statistical evaluation, paired t-test

## **1. Bevezetés**

Statisztikai kiértékelések során ma már természetes, hogy számológépet használunk. Viszont ekkor az ismert statisztikai programcsomagokat alkalmazzuk, míg az Excel táblázatkezelő programra ritkán esik a választásunk. Elkerülhetjük a nagyszámú, ismétlődő kiértékelések hosszadalmas, monoton végrehajtását, ha kihasználjuk az Excel nyújtotta programozási lehetőséget a Visual Basic for Application (VBA) szolgáltatással (Fabulya, 2017). Elegendő egy olyan számológépet kialakítanunk, melyben automatikusan újraszámítódnak a statisztikai eredmények a kiinduló adatok megváltozásakor (Microsoft, 1995). A VBA programunknak csak a kiértékelendő adatsorok és a keletkező eredmények adatainak cellák közötti mozgatását kell automatizálnia.

A páros t-próbát gyakran alkalmazzuk olyan esetekben, amikor két, egymástól nem független, párokba szervezhető adatokkal rendelkező minták összehasonlítását kell elvégeznünk. A próba elvégzéséhez pár formula kiszámítása szükséges (Michaletzky–Mogyoródi, 1995), melyek könnyen kialakíthatók a számológépben.

Az Excel VBA programokat minden olyan területen célszerű alkalmazni, ahol ismétlődő számítási tevékenységek automatizálása megvalósítható (Kovalcsik, 2010). Így pénzügyi befektetések kiértékelésekor (Zsótér–Tóth, 2014), vagy

adatbázisba szervezett adatok esetén is kényelmesebbé, gyorsabbá teheti munkánkat (Fabulya, 2018).

## 2. Anyag és módszer

Munkánkhoz az Excel 2010 verzióját használtuk. Az alábbiakra tekintettel esett erre a választás:

- Programozási lehetőséget biztosít a Visual Basic for Application bővítményével.
- Az adataink Excel fájlokban érhetőek el.
- Egyszerűen kialakíthatjuk a statisztikai kiértékelés számolótábláját munkalapfüggvények alkalmazásával (Wells–Harshberger, 1997).
- A számítógép használók többsége rendelkezik a programmal, ismeri alapvető használatát.

### 2.1. Az Excel és VBA bővítménye

A táblázatkezelő programban egy munkafüzet több lapján alakíthatjuk ki az adatok tárolására és a rajtuk elvégzendő számításokra alkalmas felhasználói felületeket (Zimmerman, 1996). Az automatikus újraszámításokhoz olyan formulákat kell megadnunk, melyek hivatkoznak azokra a cellákra, melyektől az eredmény függ. Az adatok megváltoztathatósága miatt arra is tekintettel kell lennünk a formulák helyes eredménye érdekében, hogy extrém értékek, akár hiányzó adatok mellett is működjenek.

A VBA bővítmény programozási lehetőséget nyújt. Ezt arra használjuk, hogy automatizáljuk a minták adatainak lecserélését a számolótáblán, valamint az eredmények kigyűjtését táblázatba. Értékadó utasítások szükségesek a cellák és a program változói között az adatok mozgatásához. A programjainkban ciklusok biztosítják a tevékenységek ismétlődő végrehajtását, így az összes minta kiértékelését.

### 2.2. A páros t-próba

A páros t-próba segítségével két, egymástól nem független, adatként szervezhető minták alapján tesztelhetjük két statisztikai sokaság várhatóértékei közötti relációra vonatkozó hipotézis helyességét. Az adatoknak intervallum skálán mérhető, numerikus adattípusú kell lennie. Az alkalmazhatóságának feltétele, hogy normális eloszlásúnak legyen tekinthető a két populáció (Móri, 2011).

A próba végrehajtásakor először kiszámítjuk az alkalmazott statisztikai függvény (1) értékét ( $t$ ) a minta adatsorából.

$$d_i = x_i - y_i \quad \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad t = \frac{\bar{d}}{s} \quad s = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}} \quad (1)$$

ahol:

$x_i, y_i$  – a minták  $i$ . elempárja ( $i=1, \dots, n$ ),

$s$  – a becült szórás,

$\bar{d}$  – a mintaelemek különbségének átlaga, (számtani közép),

$n$  – a minta elemszáma.

A döntést két technikával is meghozhatjuk:

- kritikus tartomány alapján,
- szignifikanciaszint ( $p$ ) kiszámításával.

Amikor a döntést a kritikus tartománnyal hozzuk meg, akkor a statisztikai függvény értékét hasonlítjuk össze a tartomány szélét mutató értékkel, mely a Student-féle t-eloszlás eloszlásfüggvényének értéke, mint kritikus érték ( $t_{kr}$ ), mely függ a minta elemszámától ( $n$ ) és az elsőfajú hibavalószínűségtől ( $\varepsilon$ ). Ha a kritikus tartományba esett a statisztikai függvény értéke, akkor a két populáció várhatóértékének szignifikáns, jelentős eltérését jelzi a próba. Szintén ez adódik, ha a szignifikanciaszint kisebb az elsőfajú hibavalószínűségnél, azaz  $p < \varepsilon$  (Móri, 2011).

### 3. Eredmények és értékelésük

Két munkalapot kell kialakítanunk az Excelben ahhoz, hogy a felhasználó csak a saját adatainak szerkesztését tudja elvégezni az egyikben, míg a számításokért felelős munkalap ne legyen szerkeszthető. Viszont így ezek a munkalapok „kommunikálnak” egymással, hiszen a felhasználói felület munkalapjának adatait dolgozza fel a számítások munkalapja, míg a számítások eredménye megjelenik a felhasználói felületen.

#### 3.1. Számítások munkalapja

A számítások munkalap a felhasználó számára rejtett, így nem szükséges nagy hangsúlyt fektetnünk az esztétikus, felhasználóbarát megjelenítésre. Meg kell oldanunk, hogy csak a számítások során értelmezhető, numerikus adatok kerüljenek át a kezelő felület munkalapjáról. Ez megoldható az Excel munkalap függvényeivel a (2) képlet szerint.

$$=HA(SZÁM(Adatok!A3);Adatok!A3;"") \quad (2)$$

A  $HA()$  munkalapfüggvény a  $SZÁM()$  függvénnyel ellenőrzi, hogy az adatkezelő felület *Adatok* munkalapján az *A3* cella tartalma szám vagy sem. Így csak numerikus érték kerül át a számítások munkalapra. E képlet másolatai eredményezik, hogy a rejtett munkalapra helyezzük az adatokat az *A2* cellától kezdődően. A számítások már csak ezeken az adatokon alapulnak. Mivel ezek munkalapfüggvények, ezért automatikusan megjelenik minden adatváltozás a számítások lapján.

Az elsőfajú hibavalószínűség (*D3* cella), valamint a vizsgálandó hipotézis típusa (*D5* cella) szükséges még a számítások elvégzéséhez. Ezek ellenőrzés nélkül vehetők át az *Adatok* munkalapról, mert ott hibás értéket nem fogad el a program a felhasználótól. A számításokor a hipotézis típusa három féle lehet attól függően, hogy milyen irányú eltérést vizsgálunk a két sokaság várhatóértéke között:

- kétszélű ( $m_1 \neq m_2$ ),
- balszélű ( $m_1 < m_2$ ),
- jobbszélű ( $m_1 > m_2$ ).

A felhasználó a vizsgálandó hipotézis megfogalmazásakor tetszőleges relációs jeleket alkalmazhat, de a számítás technikája a fenti típusú ellenhipotézisek (alternatív hipotézisek) egyikére vezet. Az 1. táblázatban látható, hogy a lehetséges felhasználó által vizsgálandó hipotézis (H) típusok esetén a számításokor melyik alternatív hipotézist ( $H_A$ ) kell alkalmazni, és milyen döntést kell hozni H-ra, amikor az alternatív hipotézist kell elfogadni, vagyis amikor a számított szignifikanciaszint kisebb az elsőfajú hibavalószínűsénél ( $p < \varepsilon$ ).

1. táblázat: **Döntés az alkalmazott alternatív hipotézis elfogadásakor**

| H              | $H_A$          | Döntés $H_A$ elfogadásakor |
|----------------|----------------|----------------------------|
| $m_1 = m_2$    | $m_1 \neq m_2$ | H elutasítva               |
| $m_1 \neq m_2$ | $m_1 \neq m_2$ | H elfogadva                |
| $m_1 < m_2$    | $m_1 < m_2$    | H elfogadva                |
| $m_1 \leq m_2$ | $m_1 > m_2$    | H elutasítva               |
| $m_1 > m_2$    | $m_1 > m_2$    | H elfogadva                |
| $m_1 \geq m_2$ | $m_1 < m_2$    | H elutasítva               |

Forrás: A szerző saját szerkesztése.

A számítások során egy-egy cellában a következő értékek alakulnak ki:

- a minta elemszáma ( $n$ ),
- a mintaelemek különbségének átlaga ( $\bar{d}$ ),
- a szórás becslt értéke ( $s$ ),
- a statisztikai függvény értéke ( $t$ ),
- a kritikus tartomány határa ( $t_{kr}$ ),
- a próba szignifikanciaszintje ( $p$ ).

Ezek mindegyike munkalapfüggvényekkel kapható meg:

- DARAB(),
- ÁTLAG(),
- SZÓRÁS(),
- T.INVERZ(),
- T.ELOSZL().

A kritikus tartomány határának, valamint a szignifikanciaszintnek a kiszámításakor azt is figyelembe kell venni, hogy melyik típusú ellenhipotézist alkalmazzuk.

A felhasználói felület felé a legfontosabb eredmény a szignifikanciaszint ( $p$ ), melyből könnyen adódik az eredmény a különböző típusú hipotézisek esetén.

### 3.2. Felhasználói felület kialakítása

Minden adatváltozáskor az *Adatok* munkalapon a számítások automatikusan aktualizálódnak, mivel a számítások munkalapon csak Excel munkalap függvényeket alkalmaztunk. Ez azt jelenti, hogy elegendő csak az *Adatok* munkalapot olyan felhasználói felülettel alakítani, ahol az adatbevitel és az eredmény megjelenítése valósul meg. A kész felület látható az 1. ábrán.

A próba eredményét három módon is megjelenítjük egy-egy bekeretezett részben:

- szövegesen,
- egy + vagy – jellel,
- a szignifikancia szinttel.

1. ábra: A páros t-próba felhasználói felülete

| Adatok |      | Elsőfajú hibavalószínűség ( $\epsilon$ )  | 0,05                |
|--------|------|---|---------------------|
| 48,2   | 51,1 | Hipotézis várhatóértékre  | $m_1 - m_2 = m = 0$ |
| 48,9   | 50,2 |   |                     |
| 49,2   | 50,5 | <b>A próba eredménye:</b><br>Az $m_1 - m_2 = m = 0$ hipotézis elfogadható. → +<br>Szignifikancia szint: 0,13733 |                     |
| 50,2   | 49,2 |   |                     |
| 48,9   | 49,6 |   |                     |
| 51,3   | 49,9 |   |                     |
| 49,7   | 48,8 |   |                     |
| 50,3   | 50,9 |   |                     |
| 48,3   | 51,6 |   |                     |

Forrás: A szerző saját szerkesztése.

Az ábrán látható, hogy a vizsgálandó hipotézis egyszerűbb formájú, ha bevezetjük a két várhatóérték helyett a különbségüket ( $m = m_1 - m_2$ ). A döntéshez csak a szignifikanciaszint értéke szükséges a számítások eredményeként, melyet az elsőfajú hibavalószínűséggel hasonlítunk össze. Az alternatív hipotézist fogadjuk el, ha teljesül a  $p < \epsilon$  feltétel.

A felhasználónak adatok módosítására a világos háttérű cellák esetén van csak lehetősége. Megadható a két minta adatsora, az elsőfajú hibavalószínűség, valamint a hipotézis típusa. A munkalap védelmét bekapcsolva, és a védendő cellákat zároltra állítva érhetjük el, hogy a munkalapon a felhasználó csak a többi cellát módosíthassa.

Érvényesítési beállításokkal védekezhetünk nem értelmezhető adatok ellen. Az elsőfajú hibavalószínűség értéke leggyakrabban 0,01 vagy 0,05 szokott lenni. Itt a 2. ábrán látható érvényesítési beállítás miatt csak 0 és 0,2 közötti érték adható meg.

2. ábra: Adatérvényesítés üzenetei

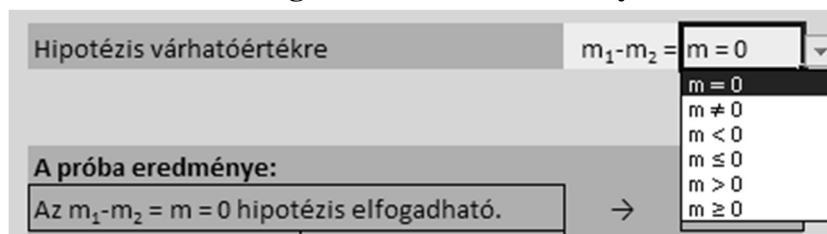
The screenshot shows the 'Páros t-próba' dialog box with the 'Elsőfajú hibavalószínűség ( $\epsilon$ )' field set to 0,25. A 'Javítás kell!' (Correction needed!) error message is displayed, stating 'Nincs 0 és 0,2 között!' (Not between 0 and 0,2!). Below the error message are buttons for 'Ujra' (Again), 'Mégse' (No), and 'Súgó' (Help). To the right, a 'Figyelem!' (Warning!) dialog box is visible, stating 'Az érték 0 és 0,2 között legyen!' (The value must be between 0 and 0,2!). The main dialog box also shows the hypothesis type 'Az  $m_1 - m_2 = m = 0$  hipotézis nem fogadható el.' (The  $m_1 - m_2 = m = 0$  hypothesis is not accepted.) and a '+' button.

Forrás: A szerző saját szerkesztése.

Amikor ráállunk erre a cellára, megjelenik a tájékoztató figyelmeztetés, míg hibás adat esetén hibaüzenet is látható.

A legördülő listás adatérvényesítés (3. ábra) biztosítja, hogy a hipotézis típusaként csak az értelmezhető relációk egyike legyen kiválasztható.

3. ábra: Legördülő listás adatérvényesítés



Forrás: A szerző saját szerkesztése.

### 3.3. Kiértékelések automatizált végrehajtása

A páros t-próbát több minta adatsorán automatikusan végrehajthatjuk az Excel VBA programozási lehetőségével. Amennyiben egy munkalapon (*Adatsorok*) több páros t-próba adatait szerepeltetjük szisztematikus, algoritmizálható elrendezésben, akkor itt a kiértékelés eredményeket is megkaphatjuk egy adatként, például + vagy – jelzés formájában. A programban ciklus szerkezetre van szükségünk. A ciklusban azt kell leprogramoznunk egy minta adatsor hipotézisének kiértékeléséhez, hogy az adatok átkerüljenek a felhasználói felület munkalapjára (*Adatok*), majd az ott adódó eredményt kiolvastva, a döntés eredményét helyezzük az erre a célra fenntartott cellába.

Az *Adatsorok* munkalap a következő szerkezetű:

- két oszlopban helyezkedik el egy kiértékelendő mintapár, mellettük a következő párok,
- az első oszlop első cellája az elsőfajú hibaválószerűség,
- a második oszlop első cellája a hipotézis típusa,
- az első oszlop második cellája az eredmény számára fenntartott cella,
- a két oszlop harmadik cellájától a két minta adatsora következik.

A program a következő:

```
Public Sub kiertekeles()
    s1 = Sheets("Adatsorok")
    s2 = Sheets("Adatok")
    minta = 1
    Do Until s1.Cells(1, 2 * minta - 1) = ""
        'Alapadatok feltöltése
        s2.Cells(2, 7) = s1.Cells(1, 2 * minta - 1)
        s2.Cells(4, 8) = s1.Cells(1, 2 * minta)
        'Törlés
        sor = 3
        Do Until s2.Cells(sor, 1) = ""
            s2.Cells(sor, 1) = ""
```

```

        s2.Cells(sor, 2) = ""
        sor = sor + 1
    Loop
    'Minta feltöltése
    sor = 3
    Do Until s1.Cells(sor, 2 * minta) = ""
        s2.Cells(sor, 1) = s1.Cells(sor, 2 * minta - 1)
        s2.Cells(sor, 2) = s1.Cells(sor, minta)
        sor = sor + 1
    Loop
    'Eredmény kiolvasása
    s1.Cells(2, 2 * minta - 1) = s2.Cells(8, 8)
    minta = minta + 1
End Sub

```

Az adatok áthelyezése az *s1* és *s2* változókkal elérhető *Adatsorok* és *Adatok* munkalapok cellái között történik. Az első mintától (*minta=1*) kezdődik a végrehajtás, mely akkor ér véget, mikor az aktuális oszlop üres cellával nem kezdődik (*do until s1.cells(1,2\*minta-1)=""*). Egy cikluson belül először az alapadatokat helyezzük át, majd a minta adatsorát, de előbb törölni kell, mert bár egy újabb adatsor felülírja az előzőt, de ha az aktuális rövidebb, akkor a maradék adatok hamis eredményt okoznának. A ciklus végén kiolvassuk az eredményt és átállunk a következő minta oszlopára (*minta=minta+1*).

#### 4. Összegzés

Automatizálni tudtuk a páros t-próba számításinak végrehajtását. Kialakítottuk azt a felhasználóbarát felületet, melyen egyszerűen beállítható minden adat, s megjeleníti a próba eredményét. Programozott módon is kialakítottuk az újbóli felhasználást a Visual Basic for Application szolgáltatással. Ehhez rendelkezniünk kell minimális szintű VBA programozói ismeretekkel.

#### Irodalomjegyzék

- Fabulya Z. (2017): Hőkezelési folyamatok összehangolása Excel VBA szolgáltatásokkal. *Jelenkori társadalmi és gazdasági folyamatok*, 12 (4): 19–25.
- Fabulya Z. (2018): Access alkalmazás kialakítása ügyfélközpontú szolgáltatások nyilvántartására. *Jelenkori Társadalmi és Gazdasági Folyamatok*, 13 (1–2): 67–76.
- Zsótér B., Tóth A. (2014): Examination of satisfaction related to investments (2006-2011) accomplished by the local council in Abony. *Analecta Technica Szegedinensia* 8 (1): 33–37.
- Michaletzky Gy., Mogyoródi J. (1995): *Matematikai statisztika*. Nemzeti Tankönyvkiadó, Budapest.
- Kovalcsik G. (2010): *Az Excel programozása*. Computerbooks, Budapest.
- Microsoft Press (1995): *Microsoft Excel/Visual Basic Programmer's Guide*. Microsoft Press, Washington.
- Móri T. (2011): *Statisztikai hipotézisvizsgálat*. Typotex Kft., Budapest.
- Wells E., Harshberger S. (1997): *Microsoft Excel 97 Developer's Handbook*. Microsoft Press, Washington.
- Zimmerman M. W. (1996): *Microsoft Office 97/Visual Basic Programmer's Guide*. Microsoft Press, Washington.