



Közzététel: 2024. január 22.

A tanulmány címe:

Normális eloszlás vizsgálata: melyik tesztet alkalmazzuk a gyakorlatban?

Szerzők:

SZILÁRD PÁLMA BERNADETT

a Budapesti Corvinus Egyetem Heller Farkas Szakkollégiumának tagja

E-mail: szilard.palma@heller.uni-corvinus.hu

BADICS MILÁN CSABA

a Budapesti Corvinus Egyetem Pénzügyi Intézetének tanársegédje

E-mail: milancsaba.badics@uni-corvinus.hu

DOI: <https://doi.org/10.20311/stat2024.01.hu0005>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) *Statisztikai Szemle* c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szjt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szjt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle* c. folyóirat 102. évfolyam 1. számában megjelent, **Szilárd Pálma Bernadett–Badics Milán Csaba** által írt, **Normális eloszlás vizsgálata: melyik tesztet alkalmazzuk a gyakorlatban?** című tanulmány (link csatolása)”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Szilárd Pálma Bernadett – Badics Milán Csaba

Normális eloszlás vizsgálata: melyik tesztet alkalmazzuk a gyakorlatban?

Examining normal distribution: which test to use?

Szilárd Pálma Bernadett, a Budapesti Corvinus Egyetem Heller Farkas Szakkollégiumának tagja

E-mail: szilard.palma@heller.uni-corvinus.hu

Badics Milán Csaba, a Budapesti Corvinus Egyetem Pénzügyi Intézetének tanársegédje

E-mail: milancsaba.badics@uni-corvinus.hu

A statisztikában a normális eloszlásnak központi szerepe van. Kutatásunk az alábbi tizenegy normalitásteszt statisztikai erejét vizsgálja: Shapiro–Wilk- (SW-), Shapiro–Francia- (SF-), Filliben- (F-), Kolmogorov–Smirnov- (KS-), Anderson–Darling- (AD-), Kuiper- (K-), Cramér–von Mises- (CvM-), D’Agostino–Pearson- (DP-), Jarque–Bera- (JB-), Kí-négyzet- (CS-) és Vasicek- (V-) teszt. Az összehasonlításhoz öt, a normálstól eltérő eloszláscsoportból generálunk mintát Monte Carlo-szimulációval: $(-\infty, \infty)$ tartójú szimmetrikus és aszimmetrikus, $(0, \infty)$ tartójú, $(0, 1)$ tartójú és kevert eloszlásokból. A makrogazdasági és pénzügyi adatsorok gyakran 100-nál kisebb elemszámú megfigyelést tartalmaznak, új minta nem szimulálható, ezért kutatásunkban a 10, 20, 30, ..., 100 elemszámú mintaméretre fókuszálunk. Az eredmények alapján a $(-\infty, \infty)$ tartójú eloszlások esetén a regressziós és a korrelációs tesztek – mint például a Shapiro–Wilk- és a Filliben-próba – statisztikai ereje a legmagasabb. A $(0, \infty)$ és $(0, 1)$ tartójú eloszlásokon a Shapiro–Wilk- és a Vasicek-teszt teljesít a legjobban, míg a kevert a mintákon a Jarque–Bera-próba statisztikai ereje a kiemelkedő. A tanulmányunkban kapott eredmények segíthetnek a statisztikusoknak és az ökonómétereknek a megfelelő normalitásteszt kiválasztásában a vizsgált adatgeneráló folyamatról kialakított prior elképzeléseik alapján.

Kulcsszavak: normalitásvizsgálat, Monte Carlo-szimuláció, statisztikai tesztek ereje

In statistics, normal distribution has a fundamental role. This study analyses the statistical power of eleven tests of normality: Shapiro–Wilk (SW), Shapiro–Francia (SF), Filliben (F), Kolmogorov–Smirnov (KS), Anderson–Darling (AD), Kuiper (K), Cramér–von Mises (CvM), D’Agostino–Pearson (DP), Jarque–Bera (JB), Chi-square (CS) and Vasicek (V) test. The power comparison is obtained by sampling data from 5 distribution groups with Monte Carlo simulation: support $(-\infty, \infty)$ symmetric and asymmetric, support $(0, \infty)$, support $(0, 1)$, and contaminated distributions. We focus on the smaller 10, 20, 30, ..., 100 sample sizes given that econometrics often model macroeconomic and financial time-series within this range of samples. Results suggest that the regression and correlation-based normality tests, such as the Shapiro–Wilk and Filliben tests perform well on distributions with support $(-\infty, \infty)$. For distributions with $(0, \infty)$ and $(0, 1)$ support the Shapiro–Wilk and the Vasicek tests have the highest power while on contaminated samples the Jarque–Bera test outperforms them. The results could help statisticians and econometrics to apply the best-performing normality test having a prior assumption or information on the support of the data-generating process.

Keywords: normality test, Monte Carlo simulation, power of statistical tests

A vizsgált adatok eloszlása a statisztikában elengedhetetlen információ, számos tesztnek alapfeltétele. Amennyiben a mintánkról feltételezett eloszlás az adatgeneráló eloszlástól különbözik, a modellezés során kapott eredmények a valós összefüggésektől eltérők lehetnek. Például a konfidencia-intervallum szélessége, a diszkriminanciaelemzés és a parametrikus tesztek eredményei sem megbízhatóak, ha az eloszlást rosszul specifikáljuk.

A statisztikában és az ökonometriában a leggyakrabban a normális eloszlás feltételét teszteljük, grafikus vagy numerikus módszerek segítségével. A grafikus tesztek közül a legelterjedtebb a Q–Q-plot, amely a standardizált empirikus és a standard normális eloszlás kvantiliseit hasonlítja össze. Mindazonáltal a grafikus módszerek hátránya a numerikus módszerekkel szemben az, hogy nem hozható egyértelmű döntés az ábra alapján (*Yap–Sims, 2011; Simić, 2021*).

Numerikus statisztikai tesztek közül több mint negyven létezik a normális eloszlású minta nullhipotézisének tesztelésére (*Dufour et al., 1998*). A próbák négy csoportba sorolhatók: regressziós és korrelációs, empirikus eloszlásfüggvényen alapuló, momentumon alapuló, és egyéb tesztek (*Seier, 2002*). A tesztek a normális eloszlás egy-egy tulajdonságát vizsgálják, például a momentumok értékét, az empirikus és elméleti eloszlás közötti különbséget, vagy a diszjunkt osztályokba tartozó megfigyelések gyakoriságát. Emiatt a tesztek eredményei egymással nem mindig konzisztensek (*Yap–Sims, 2011; Simić, 2021*). Korábbi tanulmányok eredményei alapján a normalitástesztnek ereje a vizsgált minta elemszámának függvényében jelentősen változik, és kis elemszám esetén statisztikai erejük rendkívül alacsony (*Yaizici–Yolcan, 2007; Noughabi–Arghami, 2001*). A közgazdaságtanban ez a kérdés releváns, hiszen a makrogazdasági és pénzügyi adatsorok gyakran 100-nál kisebb elemszámú megfigyelést tartalmaznak, és új minta nem szimulálható. A fenti két ok miatt felmerülhet a kérdés is, hogy kis minta esetén melyik teszt a legalkalmasabb a normális eloszlás vizsgálatára.

Az elmúlt években számos tanulmány vizsgálta a normalitástesztnek elsőfajú hibáját, illetve a tesztek statisztikai erejét Monte Carlo-szimuláció segítségével. A tesztek összehasonlításához a normálistól különböző eloszlásokból szimuláltak mintákat kis, közepes és nagy elemszámmal,¹ amely során néhány kiválasztott teszt

¹ Kis mintaméretnek tekinthető a 10-től 50-ig terjedő elemszám, a közepes minta 50 és 500 elem közötti, a nagy minta 500 elem feletti.

erejét hasonlították össze (*Yaizici–Yolcan, 2007; Yap–Sims, 2011*).² Azonban a tanulmányokban kapott eredmények gyakran inkonzisztensek, aminek az az oka, hogy az aggregált eredmények a vizsgált eloszlások halmazától függenek. Mindemellett a szakirodalomban kevés kutatás hasonlította össze a normalitástesztetek erejét úgy, hogy mind a négy tesztcsoportból több tesztet is kiválasztott.

Kutatásunk célja, hogy átfogó képet adjon a normalitástesztetek széles csoportjának statisztikai erejéről a leggyakrabban előforduló alternatív eloszlások (köztük a kevert eloszlások) esetében. Tanulmányunkban arra a kérdésre próbálunk választ adni, hogy az egyes eloszlásokban, eloszláscsoportokban egy 10-től 100-ig terjedő mintát vizsgálva a négy tesztcsoportból melyik tesztnek a legmagasabb a statisztikai ereje. Kutatásunk az alábbi tizenegy normalitástesztet statisztikai erejét vizsgálja: Shapiro–Wilk- (SW-), Shapiro–Francia- (SF-), Filliben- (F-), Kolmogorov–Smirnov- (KS-), Anderson–Darling- (AD-), Kuiper- (K-), Cramér–von Mises- (CvM-), D’Agostino–Pearson- (DP-), Jarque–Bera- (JB-), Khí-négyzet- (CS-) és Vasicek- (V-) teszt. Az összehasonlításhoz normálistól különböző, öt eloszláscsoportból generálunk mintát Monte Carlo-szimulációval, amelyen a tesztek korrigált kritikus értékei alapján kiszámoljuk az elutasított hipotéziseknek a teljes mintán belüli arányát. Az öt eloszláscsoport: $(-\infty, \infty)$ tartójú szimmetrikus, $(-\infty, \infty)$ tartójú aszimmetrikus, $(0, \infty)$ tartójú, $(0, 1)$ tartójú és kevert eloszlás.

Kutatásunknak három hozzáadott értéke van a szakirodalomhoz. A korábbi tanulmányoktól eltérően nemcsak a legjobban teljesítő tesztet mutatjuk be az eloszlások függvényében, hanem az eloszláscsoportonkénti legjobb tesztet is meghatározzuk. Másrészt legjobb tudomásunk szerint a szakirodalomban nem létezik olyan kutatás, amely a normalitástesztetek ilyen széles körét vizsgálta az általunk bemutatott eloszláscsoportok mindegyikén, kis mintára. A tanulmányunkban vizsgált tizenegy normalitástesztet a korábbi kutatásokban összehasonlított tesztek uniója, így a tesztek halmazának kiválasztása véleményünk szerint kevésbé esetleges.³ Emellett kutatásunk során négy, gyakran alkalmazott eloszláscsoport mellett kevert eloszlásokat is figyelembe veszünk a tesztek összehasonlítása során.

A szakirodalomban csak *Romao és munkatársai (2010)*, *Yap és Sim (2011)*, valamint *Simić (2021)* vizsgálta mind az öt eloszláscsoportot.

A kapott eredmények alapján az ökonometerek kiválaszthatják a legjobban teljesítő normalitástesztet az eloszlásra vonatkozó prior elképzelésük szerint, illetve felmérhetik annak kockázatát, hogy egy másik eloszlás és eloszláscsoport esetén a

² Ezenkívül négy korábbi tanulmány kevert eloszlásokon is vizsgálta a tesztek erejét.

³ Kutatásunkban a szakirodalomban legalább kétszer szereplő összes normalitástesztet bevonjuk. Ez alól kivételt képeznek a tesztek módosított változatai, például a Lilliefors-tesztet a Kolmogorov–Smirnov-teszt módosításának tekintjük, mivel tesztstatisztikájuk megegyezik.

kiválasztott teszt statisztikai ereje mennyivel alacsonyabb. Kutatásunk első felében a releváns szakirodalmat mutatjuk be. Ezután a vizsgált normalitásteszteket és eloszlásokat, illetve a Monte Carlo-szimuláció paramétereit ismertetjük a 2., a 3. és a 4. fejezetben. Végül a Monte Carlo-szimuláció során kapott eredményeket mutatjuk be, és a tesztek alkalmazásához fogalmazunk meg javaslatokat.

1. Irodalmi áttekintés

Egy új normalitásteszt bevezetésekor a szerzők általában összehasonlították próbájuk statisztikai erejét más, gyakran alkalmazott normalitástesztekkel (*Vasicek, 1976; Jarque–Bera, 1987*). Mivel az összehasonlítás egyik cikkben sem volt teljes körű, az elmúlt évtizedben számos tanulmány vizsgálta részletesen a tesztek erejét Monte Carlo-szimuláció segítségével.

Kis, közepes és nagy mintaelemszámot figyelembe véve *Romao és munkatársai (2010)*, *Razali és munkatársai (2011)*, *Yap és Sims (2011)*, *Mbah és Paothong (2015)*, valamint *Wijekularathna és munkatársai (2020)* végeztek szimulációs számításokat. A tanulmányokban kapott eredmények alapján a Shapiro–Wilk- és a Shapiro–Francia-teszt statisztikai ereje a legjobb az aszimmetrikus eloszlások esetén. Nagyobb kurtózisú eloszlásokra azonban *Yap és Sim (2011)* tanulmánya szerint a Jarque–Bera- és az Anderson–Darling-teszt ereje a legmagasabb.

Az 1. táblázatban az irodalmi áttekintésben bemutatott kutatásokból azok a normalitástesztek szerepelnek, amelyek legalább két tanulmányban előfordultak. A táblázat oszlopaiban a cikkekben szereplő próbákat jelöljük: Shapiro–Wilk (SW) (*Shapiro–Wilk, 1965*), Shapiro–Francia (SF) (*Shapiro–Francia, 1972*), Filliben (F) (*Filliben, 1975*), Kolmogorov–Smirnov (KS) (*Kolmogorov, 1933*), Anderson–Darling (AD) (*Anderson–Darling, 1954*), Kuiper (K) (*Kuiper, 1960*), Cramér–von Mises (*Von Mises, 1931*), D’Agostino–Pearson (DP) (*D’agostino et al., 1990*), Jarque–Bera (JB) (*Jarque–Bera, 1987*), Khí-négyzet (CS) (*Pearson, 1894*), Vasicek (V) (*Vasicek, 1976*). A tesztcsoportok sorában a *Sigut és munkatársai (2006)* által alkotott négy kategóriából a tanulmány által vizsgált tesztcsoportok száma szerepel. A mintaméret sorban a minimális és a maximális mintaméret található, zárójelben a tartományban szereplő minták számát jelezzük.

1. táblázat

A normalitás tesztek összehasonlító tanulmányokat összefoglaló táblázat
Summary of the existing literature on the power of normality tests

| Tesztek | Thadewald–Brüning (2007) | Yaizici–Yolacan (2007) | Noughabi–Arghami (2011) | Razali et al. (2011) | Yap–Sim (2011) |
|--------------------|--------------------------|------------------------|-------------------------|----------------------|-------------------|
| Shapiro–Wilk | ✓ | ✓ | ✓ | ✓ | ✓ |
| Shapiro–Francia | | | | | ✓ |
| Filliben | | | | | |
| Kolmogorov–Smirnov | ✓ | ✓ | ✓ | ✓ | ✓ |
| Anderson–Darling | | ✓ | ✓ | ✓ | ✓ |
| Kuiper | ✓ | ✓ | ✓ | | |
| Cramér–von Mises | ✓ | | ✓ | | |
| D’Agostino–Pearson | | | ✓ | | ✓ |
| Jarque–Bera | ✓ | ✓ | ✓ | | ✓ |
| Khí-négyzet | | ✓ | | | ✓ |
| Vasicek | | ✓ | ✓ | | |
| Tesztcsoportok | 3 | 4 | 4 | 2 | 4 |
| Mintaméret | [10, 500] (6) | [20, 50] (4) | [10, 50] (4) | [10, 1000] (15) | [10, 2000] (9) |

| Tesztek | Romao et al. (2010) | Mbah–Paothong (2015) | Ahmad–Khan (2015) | Wijekularathna et al. (2020) | Simić (2021) |
|--------------------|---------------------|----------------------|-------------------|------------------------------|-------------------|
| Shapiro–Wilk | ✓ | ✓ | ✓ | ✓ | ✓ |
| Shapiro–Francia | ✓ | ✓ | ✓ | | ✓ |
| Filliben | ✓ | | | | |
| Kolmogorov–Smirnov | ✓ | ✓ | ✓ | ✓ | ✓ |
| Anderson–Darling | ✓ | ✓ | ✓ | ✓ | ✓ |
| Kuiper | | | | | |
| Cramér–von Mises | | ✓ | ✓ | ✓ | ✓ |
| D’Agostino–Pearson | ✓ | ✓ | ✓ | ✓ | ✓ |
| Jarque–Bera | ✓ | ✓ | ✓ | ✓ | ✓ |
| Khí-négyzet | | ✓ | ✓ | | |
| Vasicek | | | | | |
| Tesztcsoportok | 3 | 4 | 4 | 3 | 4 |
| Mintaméret | [25, 1000] (3) | [10, 5000] (12) | [10, 500] (8) | [5, 2000] (11) | [10, 100] (10) |

A táblázatban az irodalmi áttekintésben bemutatott kutatásokból azok a normalitás tesztek szerepelnek, amelyek legalább két tanulmányban előfordultak. A táblázat oszlopaiban a cikkekben szereplő próbákat jelöljük: Shapiro–Wilk (SW) (*Shapiro–Wilk, 1965*), Shapiro–Francia (SF) (*Shapiro–Francia, 1972*), Filliben (F) (*Filliben, 1975*), Kolmogorov–Smirnov (KS) (*Kolmogorov, 1933*), Anderson–Darling (AD) (*Anderson–Darling, 1954*), Kuiper (K) (*Kuiper, 1960*), Cramér–von Mises (*Von Mises, 1931*), D’Agostino–Pearson (DP) (*D’Agostino et al., 1990*), Jarque–Bera (JB) (Jarque és Bera, 1987), Khí-négyzet (CS) (Pearson, 1894), Vasicek (V) (*Vasicek, 1976*). A tesztcsoportok sorban a *Sigut és munkatársai (2006)* által felállított négy kategóriából a tanulmány által vizsgált tesztcsoportok száma szerepel. A mintaméret sorában a minimális és a maximális mintaméret található, zárójelben a tartományban szereplő minták számát jelezzük.

Thadewald és Brüning (2007), *Ahmad és Khan (2015)* és *Simić (2021)* kis és közepes mintaméretben vizsgálta a normalitástesztetek erejét. Az *Ahmad és Khan (2015)*, valamint a *Simić (2021)* által elvégzett szimulációk alapján a logisztikus és a $(0, 1)$ tartójú eloszlásoktól eltekintve minden más eloszlásra a Shapiro–Wilk- és a Shapiro–Francia-teszt ereje volt a legnagyobb.⁴ *Thadewald és Brüning (2007)* azonban a korlátos tartójú eloszlásoktól eltekintve a Jarque–Bera-próba statisztikai erejét találta a legnagyobb.⁵

Kutatásunk *Yaizici és Yolacan (2007)*, illetve *Noughabi és Arghami (2011)* tanulmányához áll legközelebb. *Yaizici és Yolacan (2007)* tizenkét normalitástesztet hasonlított össze, 20-as, 30-as, 40-es és 50-es mintaméretű eloszlások segítségével. A szimuláció során kapott eredmények alapján a Jarque–Bera-teszt statisztikai ereje a legerősebb aszimmetrikus eloszlások esetén, szimmetrikus eloszlásokon pedig a Kolmogorov–Smirnov-, illetve az Anderson–Darling-próba teljesített a legjobban. *Noughabi és Arghami (2011)* kutatásában hét teszt statisztikai erejét vizsgálta, 10, 20, 30, 40, 50 elemszámú mintán. Tanulmányuk a $(-\infty, \infty)$ tartójú szimmetrikus eloszlásokra a Jarque–Bera-tesztet, az aszimmetrikus esetben pedig a Shapiro–Wilk-próbát találta optimálisnak. A $(0, 1)$ és $(0, \infty)$ tartójú eloszlások esetében a Shapiro–Wilk- és a Vasicek-teszt ereje volt a legmagasabb.

A normalitástesztetek összehasonlító tanulmányok eredményeinek összefoglalása után érdemes kiemelni, hogy a tesztek aggregált statisztikai ereje függhet a szimulációba bevont eloszlásoktól és a mintaelemszámtól. Például a $(-\infty, \infty)$ tartójú, szimmetrikus eloszlásokon a Jarque–Bera-teszt statisztikai ereje a legmagasabb, kivéve a $t(1)$ eloszlást, amelyen az Anderson–Darling-teszt pontosabb (*Noughabi–Arghami, 2001*). Ehhez hasonlóan a szimmetrikus, korlátos tartójú eloszlásokon a Kolmogorov–Smirnov-normalitásteszt statisztikai ereje a legkisebb mintaelemszámokon magas, azonban pontossága a többi tesztől eltérően még a vizsgált 300 elemű mintán sem éri el a 80%-ot (*Yap–Sims, 2011*).

A fenti összefoglalás alapján a szakirodalomban kevés tanulmány hasonlította össze a normalitástesztetek erejét mind a négy tesztcsoporthoz választva, illetve az öt eloszláscsoportra fókuszálva. Tanulmányunkban ezért megvizsgáljuk, hogy a négy csoportból az eloszlások széles körét vizsgálva melyik normalitástesztet teljesít a legjobban. A felsorolt kritikák alapján kutatásunkban a 10-től 100-ig terjedő mintaméretre fókuszálunk. A korábbi kutatásokban vizsgált normalitásteszteteket és mintaelemszámokat az 1. táblázat foglalja össze. Ezenkívül a Függelék A1. táblázata összegzi a tanulmányok által vizsgált eloszlásokat.

⁴ A logisztikus eloszlásra *Ahmad és Khan (2015)* tanulmánya alapján a Jarque–Bera-teszt ereje volt kiemelkedő. A $(0, 1)$ tartójú eloszlások esetében *Simić (2021)* a D’Agostino–Pearson-teszt erejét találta a legmagasabbnak.

⁵ A különbség a vizsgált eloszlásokból adódik. *Thadewald és Brüning (2007)* kevert eloszlások segítségével vizsgálta a tesztek statisztikai erejét, míg a többi tanulmány más eloszlásokból vett mintát.

2. Normális eloszlás tesztelése

Az ökonometriában a normalitás tesztelése az alábbi módon fogalmazható meg. Legyen adott egy $x_1, x_2, x_3, \dots, x_n \sim i. i. d$ minta φ folytonos eloszlásból. Tesztelni szeretnénk a következő nullhipotézist: $H_0: \varphi \sim \mathcal{N}(\mu, \sigma^2)$. Az alternatív hipotézis: $H_1: \varphi \not\sim \mathcal{N}(\mu, \sigma^2)$.

2. táblázat

A kutatásban vizsgált normalitásteszteket összefoglaló táblázat
Tests of normality

| Teszt neve | Tesztstatisztika | Jelölés |
|---------------------------------------|---|---|
| 1. Regressziós és korrelációs tesztek | | |
| Shapiro–Wilk | $\frac{(\sum_{i=1}^n a_i \cdot y_i)^2}{\sum_{i=1}^n y_i - \bar{y}}$ | $y = (y_1, y_2, \dots, y_n)$: a rendezett véletlen minta |
| | | $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$: a mintaátlag |
| | | $a^T = \frac{b^T V^{-1}}{\sqrt{b^T V^{-1} V^{-1} b}}$ |
| | | b : n darab standard normális eloszlású rendezett minta várható értéke. V : n darab standard normális eloszlású rendezett minta kovariancia mátrixa. |
| Shapiro–Francia | $\frac{(\sum_{i=1}^n a_i \cdot y_i)^2}{\sum_{i=1}^n y_i - \bar{y}}$ | $a = \frac{b^T}{\sqrt{b^T b}}$ |
| Filliben | $\frac{\sum_{i=1}^n x_{(i)} \cdot M_{(i)}}{\sqrt{\sum_{i=1}^n M_{(i)}^2} \cdot \sqrt{(n-1) \cdot s^2}}$ | $M_{(i)} = \phi^2(m_{(i)})$ $m_{(1)}, m_{(2)}, \dots, m_{(n)}$: a $[0, 1]$ tartójú, egyenletes eloszlású rendezett minta medián értékei. |
| 2. Empirikus eloszlásfüggvénytesztek | | |
| Kolmogorov–Smirnov | $\max(D^+, D^-)$ | $D^+ = \max\left\{\left(\frac{i}{n}\right) - z_i, i = 1, \dots, n\right\}$ |
| | | $D^- = \max\left\{z_i - \frac{i-1}{n}, i = 1, \dots, n\right\}$ |
| | | $z_i = \Phi\left(\frac{x_i - \bar{x}}{\sigma_x}\right)$ |
| Anderson–Darling | $n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \Psi(F(x)) dF(x)$ | $\Psi(u)$: a nemnegatív súlyfüggvény |
| | | $F_n = \frac{\text{no.of } x_i < x}{n}$: a tapasztalati eloszlásfüggvény F : a tesztelni kívánt eloszlásfüggvény |
| Kuiper | $D^+ + D^-$ | D^+, D^- : a Kolmogorov–Smirnov-tesztben leírt értékek |
| Cramér–von Mises | $\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$ | $F_n = \frac{\text{no.of } x_i < x}{n}$: a tapasztalati eloszlásfüggvény F : a tesztelni kívánt eloszlásfüggvény |

(A táblázat folytatása a következő oldalon)

(folytatás)

| Teszt neve | Tesztstatisztika | Jelölés |
|--------------------|--|---|
| 3. Momentumtesztek | | |
| D'Agostino–Pearson | $Z^2(\sqrt{b_1}) + Z^2(b_2)$ | $\sqrt{b_1}$: ferdeség b_2 : csúcsosság $Z(\sqrt{b_1}) + Z(b_2)$ normális eloszlású, ha a minta normális eloszlásból származik |
| Jarque–Bera | $n \left(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)}{24} \right)$ | $\sqrt{b_1}$: ferdeség b_2 : csúcsosság |
| 4. Egyéb tesztek | | |
| Khí-négyzet | $\sum_{i=1}^n \frac{O_i - E_i}{E_i}$ | O_i : megfigyelt értékek E_i : eloszlás alapján várt értékek |
| Vasicek | $\frac{\exp\{V_{m,n}(x_1, x_2, \dots, x_n)\}}{\sigma(x_1, x_2, \dots, x_n)}$ | $V_{m,n}(x_1, x_2, \dots, x_n)$ $= \frac{1}{n} \sum_{i=1}^n \log \left(\frac{n}{2m} [x_{i+m} - x_{i-m}] \right)$ |

A táblázat soraiban szereplő normalitástesztek: Shapiro–Wilk (SW) (Shapiro és Wilk, 1965), Shapiro–Francia (SF) (Shapiro és Francia, 1972), Filliben (F) (Filliben, 1975), Kolmogorov–Smirnov (KS) (Kolmogorov, 1933), Anderson–Darling (AD) (Anderson és Darling, 1954), Kuiper (K) (Kuiper, 1960), Cramér–von Mises (Von Mises, 1931), D'Agostino–Pearson (DP) (D'agostino et al., 1990), Jarque–Bera (JB) (Jarque és Bera, 1987), Khí-négyzet (CS) (Pearson, 1894), Vasicek (V) (Vasicek, 1976).

A hipotézis megfogalmazása nem (μ, σ) specifikus értékeire értendő, azonban azokra is lehetséges. A frekventista megközelítésben az α , elsőfajú hiba valószínűségének megválasztását követően döntünk a nullhipotézis elutasításáról vagy elfogadásáról. A megközelítés alap gondolata, hogy a nullhipotézis fennállása esetén a tesztstatisztika (S) eloszlása ismert. Ekkor az \mathcal{H} halmaz definíciója:

$$P(S \in \mathcal{H}) = 1 - \alpha$$

A mintából becsült statisztika értéke alapján nem tudjuk elutasítani a nullhipotézist, ha $S \in \mathcal{H}$.

Seier (2002) négy csoportba sorolta az egydimenziós normalitásteszteket.⁶ Kutatásunkban az ezekből a csoportokból leggyakrabban alkalmazott tesztek statisztikai erejét vizsgáljuk: Shapiro–Wilk- (SW-), Shapiro–Francia- (SF-), Filliben- (F-), Kolmogorov–Smirnov- (KS-), Anderson–Darling- (AD-), Kuiper- (K-), Cramér–von Mises- (CvM-), D'Agostino–Pearson- (DP-), Jarque–Bera- (JB-), Khí-négyzet- (CS-) és Vasicek- (V-) teszt. A tanulmányban vizsgált normalitásteszteket a 2. táblázat foglalja össze. A táblázatban megtalálhatóak a próbákhoz

⁶ Romao és munkatársai (2010), Razali és munkatársai (2011), valamint Mbah és Paothong (2015) tanulmánya is ugyanezt a csoportosítást alkalmazta.

tartozó tesztstatisztikák és a legfontosabb jelölések. A tesztek részletesebb bemutatása a Függelék B. fejezetében olvasható.

Az első kategóriába tartoznak a regressziós és a korrelációs tesztek, amelyek alap gondolata az, hogy $y \sim \mathcal{N}(\mu, \sigma^2)$ kifejezhető $y = \mu + \sigma x$, $x \sim \mathcal{N}(0, 1)$ alakban. A regressziós tesztek, mint a Shapiro–Wilk-próba, a rendezett, standardizált empirikus és normális eloszlású minta regressziós egyenesét vizsgálják. A PP-plot alapján a megfigyelt görbe lineáris, ha a minta normális eloszlású (*Shapiro–Wilk, 1965; Shapiro–Francia, 1972*). A korrelációs tesztek a megfigyelt és a standard normális rendezett minta közötti kapcsolat erősségét vizsgálják, például a megfigyelt rendezett minta és a rendezett statisztika mediánja közötti korrelációt kiszámítva (*Filliben, 1975*).

A második csoport az empirikus eloszlásfüggvényt alkalmazó próbákat tartalmazza. A tesztek általában a megfigyelt minta és az elméleti eloszlás megfelelő pontjainak különbségeit súlyozzák. Az Anderson–Darling-próba (*Anderson–Darling, 1954*) az eloszlás szélein felülsúlyozza a különbségeket, míg a Cramér–von Mises-tesztben (*Von Mises, 1931*) egyenlő súllyal szerepelnek az eltérések. A Kolmogorov–Smirnov- (*Kolmogorov, 1933*) és a Kuiper- (*Kuiper, 1960*) tesztben a tesztstatisztika értéke a legnagyobb eltérés.

A harmadik csoport a momentumokon alapuló próbákat tartalmazza, általában a csúcosság- és a ferdeségmintából becsült és normális eloszlás esetén kapott értékeit összehasonlítva. A D’Agostino–Pearson-próba (*D’agostino et al., 1990*) esetében az így képzett tesztstatisztika aszimptotikusan normális eloszlású, míg a Jarque–Bera-tesztnél (*Jarque–Bera, 1987*) Khí-négyzet-eloszlású.

Az utolsó csoport az egyéb tesztek foglalta magában.⁷ A Khí-négyzet-teszt (*Pearson, 1894*) esetében az adott tartományokba esés valószínűsége és a megfigyelések száma alkotja a tesztstatisztika, míg a Vasicek-próba (*Vasicek, 1976*) a folytonos eloszlásfüggvényhez tartozó entrópia értékén alapul.

3. Vizsgált eloszlások

A kutatás során két szimulált adatbázis (eloszláshalmaz) segítségével vizsgáltuk a normalitás tesztek erejét. *Noughabi és Arghami (2011)* húsz eloszlást kategorizált négy csoportba, szimmetria/aszimmetria és az eloszlás tartói szerint,

⁷ A normalitás teszteknek egyéb csoportosítása is létezik. *Yap és Sim (2011)* további megkülönböztetést tett az egyéb próbák között, és a Khí-négyzet-, illetve a spacingteszteket külön kategóriába sorolta.

Esteban és munkatársai (2001) tanulmánya alapján.⁸ Az eloszlásokat a 3. táblázat foglalja össze.

3. táblázat

Normalitásteszték vizsgálatához használt eloszlások *Esteban és munkatársai (2001)*, valamint *Noughabi és Arghami (2011)* tanulmánya alapján
Specification of the non-normal distributions based on the study of
 Esteban et al. (2001), Noughabi and Arghami (2011)

| Eloszláscsoportok | Eloszlások |
|--|---------------------------------------|
| I. $(-\infty, \infty)$ tartójú, szimmetrikus | Student t-eloszlás, 1 szabadságfokkal |
| | Student t-eloszlás, 3 szabadságfokkal |
| | Standard logisztikus |
| | Standard laplace |
| II. $(-\infty, \infty)$ tartójú, aszimmetrikus | Gumbell $\alpha = 0, \beta = 1$ |
| | Gumbell $\alpha = 0, \beta = 2$ |
| | Gumbell $\alpha = 0, \beta = 1/2$ |
| III. $(0, \infty)$ tartójú | Exponenciális $\lambda = 1$ |
| | Gamma $\beta = 1, \alpha = 2$ |
| | Gamma $\beta = 1, \alpha = 1/2$ |
| | Lognormális $(0, 1)$ |
| | Lognormális $(0, 2)$ |
| | Weibull $\beta = 1, \alpha = 1/2$ |
| IV. $(0, 1)$ tartójú | Weibull $\beta = 1, \alpha = 2$ |
| | Egyenletes $[0, 1]$ |
| | Beta $\alpha = 2, \beta = 2$ |
| | Beta $\alpha = 0,5, \beta = 2$ |
| | Beta $\alpha = 3, \beta = 1,5$ |
| | Beta $\alpha = 2, \beta = 1$ |

Normalitásteszték erejének vizsgálatához gyakran alkalmaznak olyan mintát, amely $\mathcal{N}(0, 1)$ standard normális és $\mathcal{N}(\mu, \sigma^2)$ normális eloszlású elemeket együttesen tartalmaz. Elemzésünk robusztusságvizsgálatához kevert eloszlásokat használunk, *Yap és Sim (2011)*, illetve *Wijekularathna és munkatársai (2020)* tanulmányához hasonlóan.

Általánosan felírva a minta az alábbi kevert eloszlásból származik:

$$(1 - \epsilon) \cdot \mathcal{N}(0, 1) + \epsilon \cdot \mathcal{N}(0, 1).$$

Chen (1971) és *Simić (2021)* kutatása alapján az $\epsilon = 5\%$, 10% , a $\mu = (0, 3)$, és a $\sigma^2 = (1,9)$ paramétereket alkalmazzuk a Monte Carlo-szimuláció

⁸ Ezeket az eloszlásokat használta összehasonlító elemzésében *Simić (2021)* is, így kutatásunkban mi is ezt a négy eloszláscsoportot helyezzük a fókuszba.

során. A 4. táblázat mutatja a keveréshez használt normális eloszlásokat és arányukat.

4. táblázat

Normalitástesztvizsgálatához alkalmazott kevert eloszlások
Mixed distributions specifications

| Eloszlás típusa | Eloszlás aránya a teljes mintában (ϵ) | Együttes eloszlás típusa |
|---------------------|--|---------------------------|
| $\mathcal{N}(3, 1)$ | 5% | jobbra ferde |
| $\mathcal{N}(0, 9)$ | 5% | szimmetrikus, vastagszélű |
| $\mathcal{N}(3, 1)$ | 10% | jobbra ferde |
| $\mathcal{N}(0, 9)$ | 10% | szimmetrikus, vastagszélű |

A szimulált minta ϵ hányada tartalmazza a táblázatban specifikált normális eloszlásból szimulált elemeket és $(1 - \epsilon)$ része a standard normális eloszlásból szimulált mintát. A paraméterek *Chen (1971)* és *Simić (2021)* tanulmánya alapján lettek kiválasztva.

4. Monte Carlo-szimuláció

Tanulmányunkban a normalitástesztvizsgálatok statisztikai erejét Monte Carlo-szimuláció segítségével vizsgáljuk. A szimulációt $n = 10, 20, 30, \dots, 100$ elemű mintákra végezzük el, $\alpha = 5\%$ -os szignifikanciaszint megválasztása mellett. Elemzésünkhöz a 3. és a 4. táblázatban szereplő eloszlásokat használjuk, 1000-1000 mintát generálva mindegyik mintaméret-eloszlás párra.

Korábbi tanulmányok alapján elmondható, hogy számos normalitásteszt esetében az empirikus α értéke eltérhet a megválasztott szignifikanciaszinttől (*Yaizici–Yolcan, 2007; Noughabi–Arghami, 2001*). Emiatt a normalitástesztvizsgálatához először standard normális eloszlású mintákat generálunk, amelyek segítségével megtörténik az empirikus α értékek kiszámítása az összes teszt esetén. Minden lehetséges elemszámra 100 000 mintát szimulálunk, *Yap és Sim (2011)* kutatásához hasonlóan. A kapott eredmények az 5. táblázatban láthatók.

Az eredmények jól mutatják, hogy az empirikus α értékek a választott 5%-os szignifikanciaszinttől jelentősen eltérnek a Vasicek- és Khí-négyzet-teszt esetében, továbbá a többi tesztnél is pontatlanok a második, illetve a harmadik tizedesjegyet tekintve. Emiatt a próbák statisztikai erejének vizsgálatához nem az adott próbák hagyományos kritikus értékeit használjuk, hanem a tesztekhez tartozó kritikus értékeket szimuláció segítségével számítjuk ki. Minden elemszámra 100 000 mintát generálunk, és a kapott tesztstatisztikák empirikus eloszlásából számoljuk a kritikus értékeket.

5. táblázat

Normalitástesztek empirikus α értékei
Empirical α values of the normality tests

| Teszt neve | $n = 10$ | $n = 20$ | $n = 30$ | $n = 40$ | $n = 50$ |
|--------------------|----------|----------|----------|----------|-----------|
| Shapiro–Wilk | 0,0506 | 0,0505 | 0,0506 | 0,0508 | 0,0514 |
| Shapiro–Francia | 0,0562 | 0,0532 | 0,0523 | 0,0520 | 0,0525 |
| Filliben | 0,0495 | 0,0507 | 0,0522 | 0,0546 | 0,0566 |
| Kolmogorov–Smirnov | 0,0495 | 0,0510 | 0,0499 | 0,0508 | 0,0507 |
| Andeson–Darling | 0,0520 | 0,0516 | 0,0526 | 0,0507 | 0,0524 |
| Kuiper | 0,0486 | 0,0490 | 0,0492 | 0,0476 | 0,0489 |
| Cramér–von Mises | 0,0503 | 0,0520 | 0,0521 | 0,0511 | 0,0523 |
| D’Agostino–Pearson | 0,0584 | 0,0567 | 0,0573 | 0,0570 | 0,0583 |
| Jarque–Bera | 0,0501 | 0,0502 | 0,0497 | 0,0498 | 0,0502 |
| Khí-négyzet | 0,0000 | 0,0000 | 0,0329 | 0,0730 | 0,0609 |
| Vasicek | 0,1072 | 0,1090 | 0,1002 | 0,0932 | 0,0907 |
| Teszt neve | $n = 60$ | $n = 70$ | $n = 80$ | $n = 90$ | $n = 100$ |
| Shapiro–Wilk | 0,0509 | 0,0506 | 0,0499 | 0,0504 | 0,0505 |
| Shapiro–Francia | 0,0515 | 0,0518 | 0,0513 | 0,0525 | 0,0521 |
| Filliben | 0,0521 | 0,0473 | 0,0473 | 0,0405 | 0,0506 |
| Kolmogorov–Smirnov | 0,0501 | 0,0519 | 0,0513 | 0,0502 | 0,0501 |
| Andeson–Darling | 0,0515 | 0,0514 | 0,0517 | 0,0519 | 0,0526 |
| Kuiper | 0,0496 | 0,0481 | 0,0502 | 0,0499 | 0,0497 |
| Cramér–von Mises | 0,0521 | 0,0522 | 0,0514 | 0,0520 | 0,0526 |
| D’Agostino–Pearson | 0,0559 | 0,0561 | 0,0556 | 0,0550 | 0,0562 |
| Jarque–Bera | 0,0489 | 0,0489 | 0,0486 | 0,0486 | 0,0506 |
| Khí-négyzet | 0,0578 | 0,0583 | 0,0589 | 0,0576 | 0,0572 |
| Vasicek | 0,084/3 | 0,0818 | 0,0812 | 0,0817 | 0,0809 |

A táblázatban a vizsgált tizenegy normalitásteszthez tartozó empirikus α értékek találhatók. A szimulációhoz minden elemszám esetén 100 000-es standard normális eloszlású mintán futtattuk le a tesztek, 5%-os szignifikanciaszint megválasztása mellett. A táblázat soraiban szereplő normalitástesztek: Shapiro–Wilk (SW) (*Shapiro és Wilk, 1965*), Shapiro–Francia (SF) (*Shapiro és Francia, 1972*), Filliben (F) (*Filliben, 1975*), Kolmogorov–Smirnov (KS) (*Kolmogorov, 1933*), Anderson–Darling (AD) (*Anderson és Darling, 1954*), Kuiper (K) (*Kuiper, 1960*), Cramér–von Mises (*Von Mises, 1931*), D’Agostino–Pearson (DP) (*D’Agostino et al., 1990*), Jarque–Bera (JB) (*Jarque és Bera, 1987*), Khí-négyzet (CS) (*Pearson, 1894*), Vasicek (V) (*Vasicek, 1976*).

A Kolmogorov–Smirnov-, az Anderson–Darling-, a Kuiper-, a Cramér–von Mises-, a D’Agostino–Perason-, a Jarque–Bera-, a Vasicek-teszt jobb oldali próba, így a teststatisztikák empirikus eloszlásának az $(1 - \alpha) \cdot 100$ -dik percentilise a megfelelő kritikus érték. A Shapiro–Wilk-, a Shapiro–Francia- és a Filliben-teszt esetében

az empirikus eloszlás $\alpha \cdot 100$ -dik percentilisét vesszük. A Khí-négyzet-teszt esetében *Yap és Sim (2011)* tanulmánya alapján számítjuk ki a kritikus értékeket.⁹

Az így kiszámított kritikus értékek a függelék B.3. alfejezetében található.

5. Eredmények

Ebben a fejezetben a Monte Carlo-szimuláció során kapott eredményeket, a tesztek statisztikai erejét ismertetjük, 5%-os szignifikanciaszint megválasztása mellett, a 3. és a 4. táblázatban bemutatott eloszlásokra és eloszláscsoportokra vonatkozóan. Az alfejezetekben csak az egyes tesztcsoportokban legjobban teljesítő teszteket mutatjuk be, azonban az összes teszt eredményét, a tanulmány ábráit, táblázatait a replikálókóddal együtt a [Githubon](#) közzétesszük. A tesztcsoportokban legjobban teljesítő próbák kiválasztásához az összes, $n = 10, 20, 30, \dots, 100$ elemszámra megvizsgáljuk, hogy melyik teszt statisztikai ereje a legmagasabb. Az így legtöbbször szereplő tesztnek tekintjük a statisztikai erejét a csoporton belül maximálisnak.¹⁰

Amikor a normalitásteszteket a gyakorlatban alkalmazzuk, az esetek legnagyobb részében a mintáról nem tudjuk meghatározni, hogy milyen eloszlású, csak az eloszláscsoportról van prior elképzelésünk. Emiatt mind a négy eloszláscsoportból készítettünk egy-egy olyan mintát, amely egyenlően tartalmazza a csoportban szereplő eloszlásokat. A teljesség kedvéért a normalitástesztek statisztikai erejét a csoportban minden egyes eloszlásra külön-külön is bemutatjuk. Piros színnel jelöljük a regressziós és a korrelációs teszteket, kézzel az eloszlásfüggvényen alapuló próbákat, zölddel a momentumteszteket és lilával az egyéb tesztcsoportot.

Minden egyes eloszláscsoporthoz tartozó ábra több részabrárt tartalmaz. Ezek mindegyikén a mintaméret függvényében mutatjuk be az adott tesztcsoporton belül legjobban teljesítő tesztet. A randomizált mintavételhez tartozó eredmény mindig az ábra jobb alsó sarkában található.

⁹ Minden elemszámra kiszámítottuk a tesztstatisztikát több E_i , $i = 1, 2, \dots, m$ várható csoportra, ahol $0 < m < M$ és $M = \left(\frac{2n^2}{z_\alpha^2}\right)^{\frac{1}{5}}$. *Schorr (1974)* szimulációval szemléltette, hogy a Khí-négyzet-teszt esetén az optimális csoportok száma $M = \left(\frac{2n^2}{z_\alpha^2}\right)^{\frac{1}{5}}$ -nél kisebb kell, hogy legyen. A képletben z_α^2 a standard normális eloszlása $(1 - \alpha) \cdot 100$ -dik percentilise.

¹⁰ A kapott eredmények nem változnak, ha a tesztek statisztikai erejét a vizsgált elemszámokra összegezzük, majd ez alapján rangsorolunk.

5.1. $(-\infty, \infty)$ tartójú, szimmetrikus eloszlások

A szimmetrikus, teljes számegyenesen értelmezett eloszlásokon kapott eredményeket az 1. ábra szemlélteti. A legtöbb esetben a Filliben- és a Jarque–Bera-teszt ereje a legmagasabb.

Az ábra a tesztcsoportonkénti legmagasabb statisztikai erejű normalitástesztet mutatja a $(-\infty, \infty)$ tartójú szimmetrikus eloszlások esetében, a vizsgált minta elemszámának függvényében, 5%-os szignifikanciaszint mellett. A legjobban a Filliben- (F-), a Kuiper- (K-), a Jarque–Bera- (JB-) és a Vasicek-teszt teljesít. A randomizált mintavétel során az első eloszláscsoportban szereplő eloszlásokból szimulált minta egyenlő arányban szerepel.

Hasonló eredményeket kapott *Yap és Sims (2011)*, akik kutatásukban a szimmetrikus, vastagszélű eloszlások esetén a Jarque–Bera-teszt statisztikai erejét találták a legmagasabbnak. Ábráink kiegészítik ezt az eredményt, mivel a Filliben-próbát is figyelembe vesszük összehasonlító elemzésünkben.

A logisztikus és a standard Laplace-eloszlásokra a Khí-négyzet-teszt statisztikai ereje az $n = 40$ -es mintamérettől a legmagasabb, ami megegyezik *Romao és munkatársai (2010)* adott eloszlásokra vonatkozó állításával.¹¹

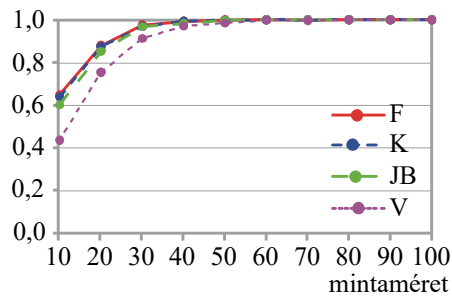
Fontos kiemelni, hogy a szimulált kritikus értékek ellenére is a tesztek statisztikai ereje a $t(1)$ eloszlástól eltekintve egyetlen elemszámra sem éri el a megválasztott 95%-os megbízhatósági szintet. A *Yap és Sim (2011)* tanulmánya által vizsgált eloszlások esetén az elemszámot legalább 300-ra kell emelni ehhez. A kapott eredmények jól illusztrálják továbbá, hogy ha a minta eloszlásának tulajdonságairól van egy helyes prior elképzelésünk, akkor a megfelelő teszt megválasztásával akár 10–30%-kal is növelhető a tesztek pontossága, ahogy ezt a randomizált minta esetén a Filliben- és a Khí-négyzet-próba eredményei is mutatják.

¹¹ *Romao és munkatársai (2010)* kutatásukban a szimmetrikus eloszlásokra a Shapiro-Wilk- és a Khí-négyzet-próba statisztikai erejét találták a legmagasabbnak.

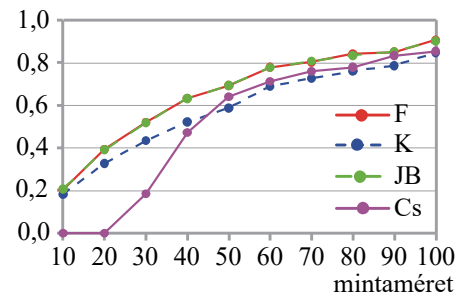
1. ábra

Tesztcsoportonkénti legjobb normalitástesztek I.*Statistical power of the best performing normality tests in each group I.***Student t-eloszlás, 1 szabadságfokkal**

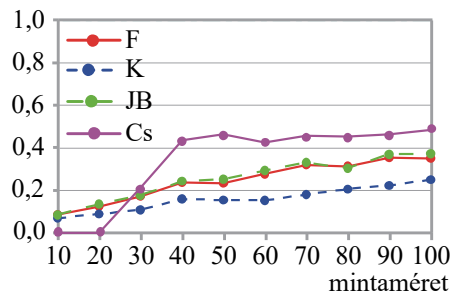
Tesztek statisztikai ereje

**Student t-eloszlás, 3 szabadságfokkal**

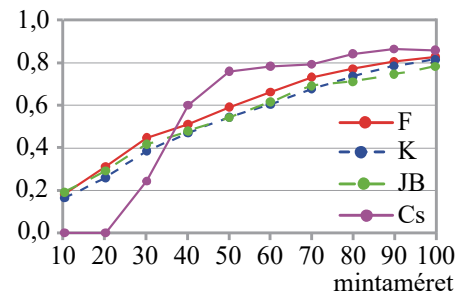
Tesztek statisztikai ereje

**Logisztikus eloszlás**

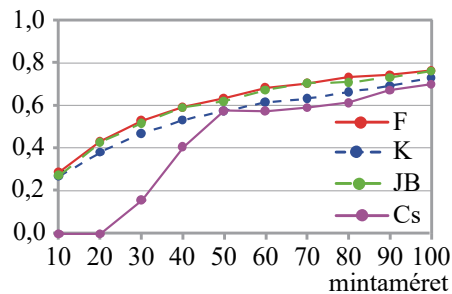
Tesztek statisztikai ereje

**Standard Laplace-eloszlás**

Tesztek statisztikai ereje

**Randomizált mintavétel $(-\infty, \infty)$ tartójú, szimmetrikus eloszlásokból**

Tesztek statisztikai ereje



5.2. $(-\infty, \infty)$ tartójú, aszimmetrikus eloszlások

Az aszimmetrikus eloszlásokat tekintve a négy legjobban teljesítő próba, a Shapiro–Wilk-, a Kuiper-, a Jarque–Bera- és a Khí-négyzet-próba statisztikai ereje hasonló a 30-nál nagyobb méretű mintákra. Az összes mintaméretre és eloszlásra a Shapiro–Wilk-teszt statisztikai ereje a legmagasabb, azonban az egyes eloszlások eredményei közötti különbség minimális.¹²

Kutatásunk megerősíti a korábbi tanulmányokat, ahol ugyanezen próba statisztikai ereje a legmagasabb a kis, a közepes, illetve a nagy mintákon is (*Yaizici–Yolacan, 2007; Noughabi–Arghami, 2011; Yap–Sims, 2011*).

Összefoglalva, a randomizált és az egyedi eloszlásokat tartalmazó mintán is a Shapiro–Wilk-teszt statisztikai ereje a legmagasabb, azonban a vizsgált próbák statisztikai ereje alig különbözik, ellentétben a $(-\infty, \infty)$ tartójú, szimmetrikus eloszlásokkal.

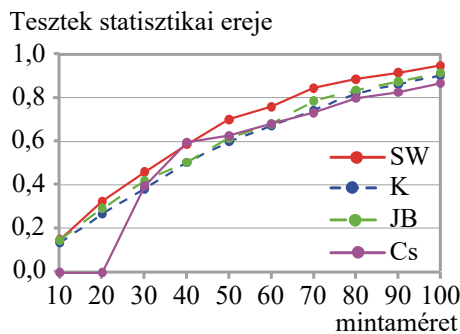
A 2. ábra a tesztcsoportonkénti legmagasabb statisztikai erejű normalitástesztet mutatja a $(-\infty, \infty)$ tartójú aszimmetrikus eloszlások esetén, a vizsgált minta elemszámának függvényében 5%-os szignifikanciaszint mellett. A legjobban a Shapiro–Wilk- (F-), a Kuiper- (K-), a Jarque–Bera- (JB-) és Khí-négyzet- (CS-) teszt teljesít. A randomizált mintavétel során a második eloszláscsoportban szereplő eloszlásokból szimulált minta egyenlő arányban szerepel.

2. ábra

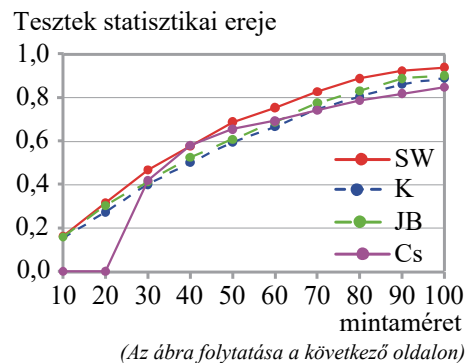
Tesztcsoportonkénti legjobb normalitásteszték II.

Statistical power of the best performing normality tests in each group II.

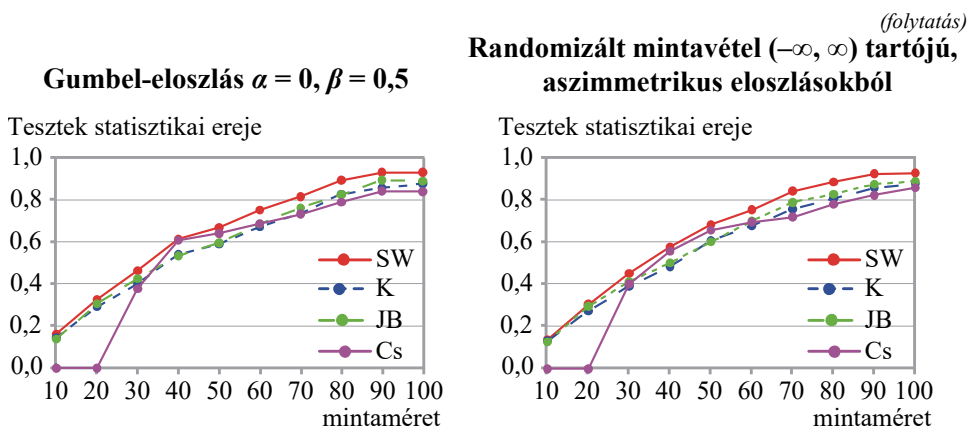
Gumbel-eloszlás $\alpha = 0, \beta = 1$



Gumbel-eloszlás $\alpha = 0, \beta = 2$



¹² A [Githubra](#) feltöltött eredmények és részletesebb ábrák alapján a Shapiro–Francia-teszt statisztikai ereje is hasonlóan magas.



5.3. $(0, \infty)$ tartójú eloszlások

A normalitástesztek statisztikai ereje a $(0, \infty)$ tartójú eloszlásokon a 3. ábrán látható. A négy tesztcsoportból a Shapiro–Wilk-, a Kuiper-, a Jarque–Bera- és a Vasicek-teszt statisztikai ereje a legmagasabb, a Weibull (1,2)-eloszlástól eltekintve, amelyen a Khí-négyzet-próba pontosabb a Vasicek-tesztnél. *Noughabi és Arghami (2011)* kutatásukban ugyancsak a Shapiro–Wilk- és a Vasicek-próba statisztikai erejét találták a legmagasabbnak a $(0, \infty)$ tartójú eloszlások esetén. A randomizált minta alapján a Shapiro–Wilk-próba ereje a legmagasabb, azonban csak 1–10%-kal nagyobb a negyedik (egyéb tesztek) tesztcsoportban szereplő Vasicek-tesztnél.¹³

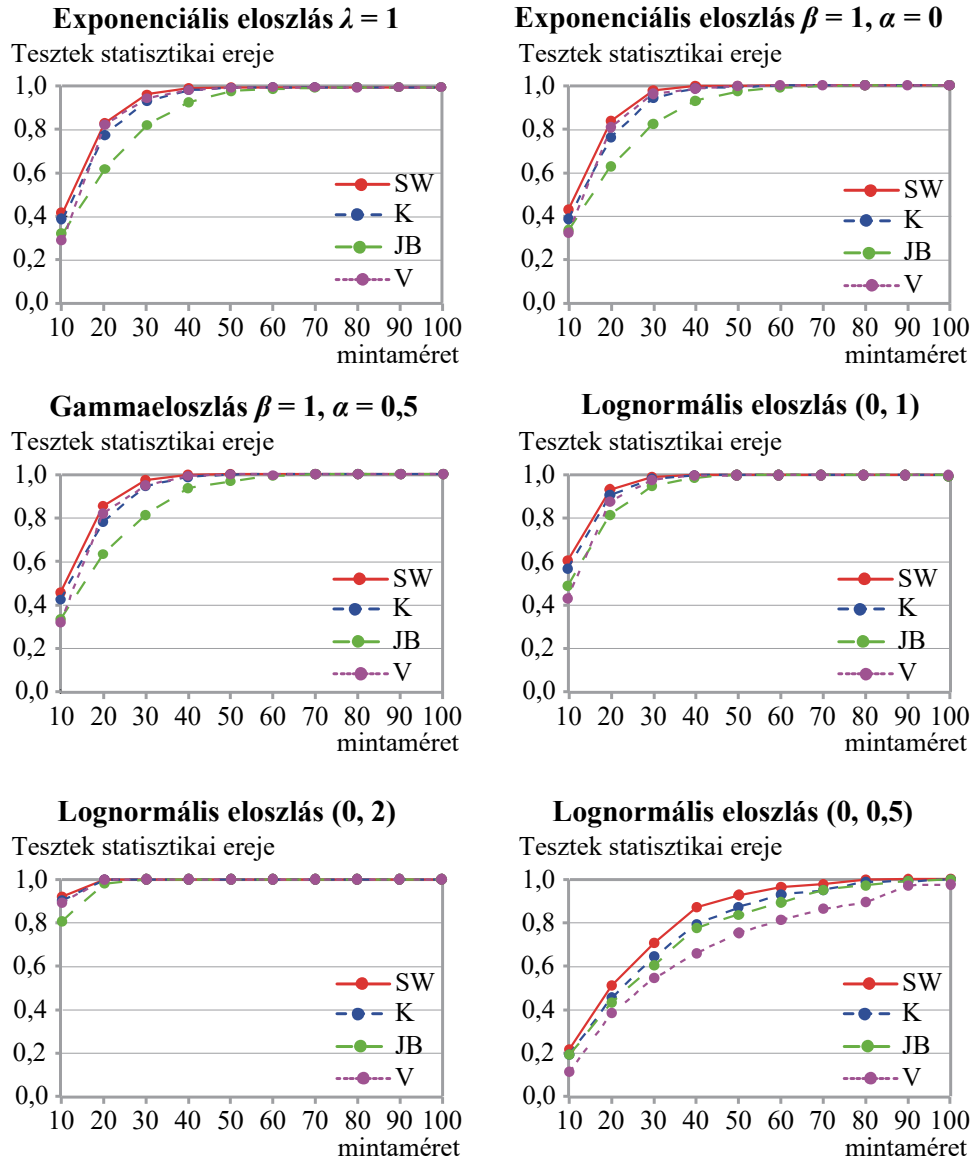
Az egyes eloszlásokat vizsgálva az exponenciális, a gamma és a lognormális eloszlás esetében a tesztek statisztikai ereje 1-hez konvergál a vizsgált elemszámokra. Mindemellett az adott eloszlásokat összehasonlítva a próbák ereje magasabb, ha az eloszlás szórása nagyobb. Például a $\sigma = 2$ paraméterű lognormális eloszlás esetében a legjobban teljesítő tesztek ereje $n = 40$ -es elemszámtól 95%-nál pontosabb. Hasonló megállapítás figyelhető meg a Weibull-eloszlásra.

Összefoglalva: a Shapiro–Wilk-próba a legmegbízhatóbb, azonban a tesztek statisztikai ereje a legtöbb eloszlás esetén alig különbözik. A legjobban teljesítő próbák ereje csak az alacsonyabb szórású lognormális és a Weibull-eloszlás esetében különbözik, átlagosan 10–20%-kal.

¹³ A [Githubon](#) szereplő eredmények alapján a negyedik tesztcsoportban szereplő Khí-négyzet-tesztnél a Shapiro–Wilk-teszt statisztikai ereje 20–60%-kal magasabb.

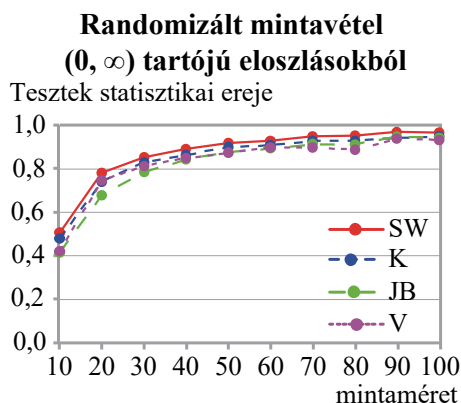
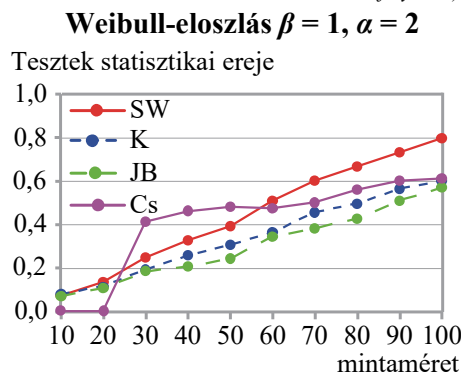
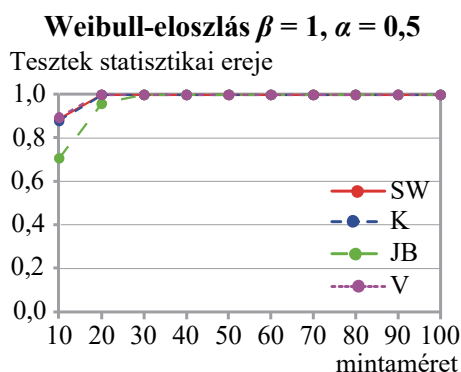
3. ábra

Tesztcsoportonkénti legjobb normalitástesztek III.
Statistical power of the best performing normality tests in each group III.



(Az ábra folytatása a következő oldalon)

(folytatás)



Az ábra a tesztcsoportonkénti legmagasabb statisztikai erejű normalitástesztet mutatja, $(0, \infty)$ tartójú aszimmetrikus eloszlások esetében, a vizsgált minta elemszámának függvényében, 5%-os szignifikanciaszint mellett. A legjobban a Shapiro–Wilk- (SW-), a Kuiper- (K-), a Jarque–Bera- (JB-), a Vasicek- (V-) és Khí-négyzet- (CS-) teszt teljesít. A randomizált mintavétel során a harmadik eloszláscsoportban szereplő eloszlásokból szimulált minta egyenlő arányban szerepel.

5.4. (0, 1) tartójú eloszlások

A $(0, 1)$ tartójú eloszlások esetén a Shapiro–Wilk-, a Kuiper-, a D’Agostino–Pearson- és a Vasicek-próba statisztikai ereje a 4. ábra alapján a legmagasabb, jól-lehet a Shapiro–Francia- és a Khí-négyzet-teszt eredményei is hasonlóak.¹⁴ Az első

¹⁴ A [Githubon](#) szereplő eredmények alapján.

eloszláscsoporttal megegyezően a legjobban teljesítő próbák ereje jelentősen különbözik. A kevert mintát tekintve az 50-es elemszámig legjobban teljesítő Vasicek-próba pontossága 20%-kal is magasabb lehet a D’Agostino–Pearson-momentumokon alapuló tesztnél. Ez a különbség 10%-ra csökken az 50-nél nagyobb mintánál a Shapiro–Wilk-próba esetében.

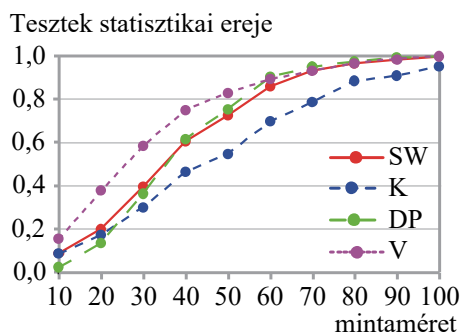
A harmadik eloszláscsoporthoz hasonlóan a nagyobb varianciájú bétaeloszlás esetében magasabb a tesztek ereje. Így a béta (0,5, 0,5) -eloszlásra a próbák ereje 50 elemszám felett 1-hez közelít, míg a legalacsonyabb varianciájú béta (2, 2) -eloszlásra csak a D’Agostino–Pearson-teszt statisztikai ereje magasabb 50%-nál az összes vizsgált elemszámot tekintve. Az egyes eloszlásokat nézve a Vasicek-teszt ereje a legmagasabb az egyenletes béta (2, 1) és a béta (0,5, 0,5) -eloszlásra, míg a béta (3, 1,5) -eloszláson a Shapiro–Wilk-próba statisztikai ereje a legnagyobb.

4. ábra

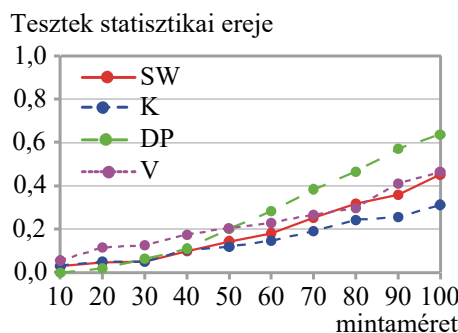
Testcsoportonkénti legjobb normalitástesztek IV.

Statistical power of the best performing normality tests in each group IV.

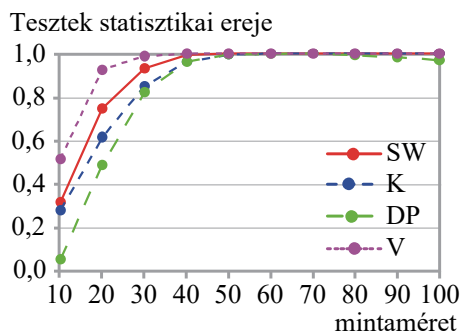
Egyenletes eloszlás (0, 1)



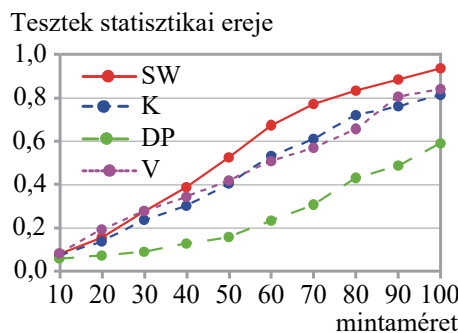
Bétaeloszlás (2, 2)



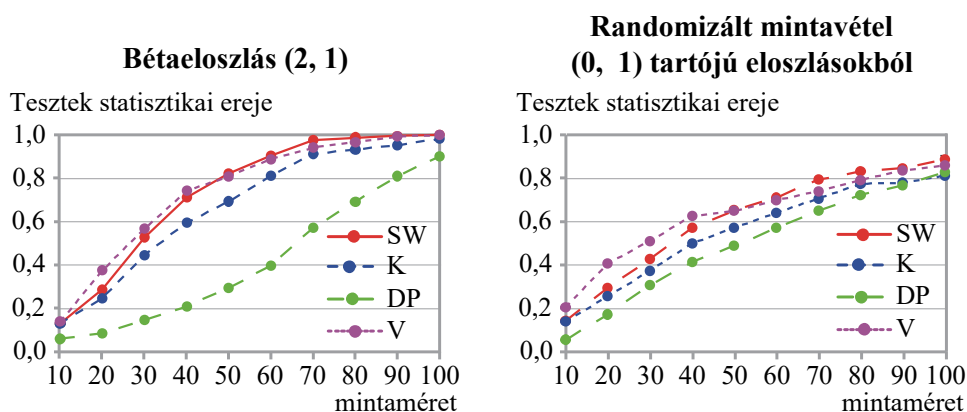
Bétaeloszlás (0,5, 0,5)



Bétaeloszlás (3, 1,5)



(Az ábra folytatása a következő oldalon)
(folytatás)



A 4. ábra a tesztcsoportonkénti legmagasabb statisztikai erejű normalitástesztet mutatja a $(0, 1)$ tartójú eloszlások esetén a vizsgált minta elemszámának függvényében, 5%-os szignifikanciaszint mellett. A legjobban a Shapiro–Wilk- (SW-), a Kuiper- (K-), a D’Agostino–Pearson- (DP-) és Vasicek- (V-) teszt teljesít. A randomizált mintavétel során a negyedik eloszláscsoportban szereplő eloszlásokból szimulált minta egyenlő arányban szerepel.

Összefoglalva: a $(0, 1)$ tartójú eloszlások esetén nem találunk egyetlen legjobban teljesítő normalitástesztet. Általában a Shapiro–Wilk- és a Vasicek-teszt ereje a legmagasabb. *Noughabi és Arghami (2011)* kutatásukban egy 10-től 50-ig terjedő mintán ugyancsak a Vasicek-próba statisztikai erejét emelték ki. *Razali és munkatársai (2011)* pedig kis, közepes és nagy mintaméretet vizsgálva a Shapiro–Wilk-tesztet találták a legjobbnak a bétaeloszlásokra. Elemzésünk megerősíti a két cikk ezekre az eloszláscsoportra vonatkozó eredményét.

5.5. Kevert eloszlások

A kevert eloszlásokon a Filliben-, a Kuiper, a Jarque–Bera- és a Khi-négyzet-próba teljesítenek a legjobban az egyes tesztcsoportokban. Az 5. ábra alapján a vastag szélű eloszlásokra a Jarque–Bera-próba statisztikai ereje a legmagasabb, viszont az aszimmetrikus eloszlásokra a Filliben- és a Khi-négyzet-teszt a hatékonyabb. Azonban fontos kiemelni, hogy a Shapiro–Francia- és a Filliben-teszt statisztikai ereje mind a négy kevert eloszlásra és az összes elemszámra maximum 1%-ban tér el.¹⁵ A D’Agostino–Pearson- és a Jarque–Bera-teszt pontossága is maximum 4,3%-ban más.

¹⁵ A [Githubon](#) szereplő ábrák és táblázatok alapján.

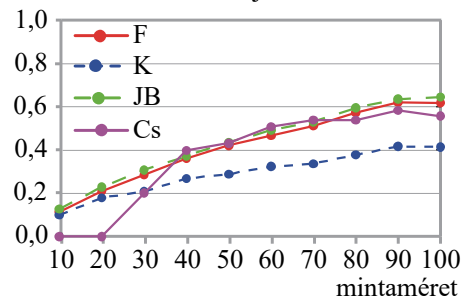
5. ábra

Testcsoportonkénti legjobb normalitástesztek V.

Statistical power of the best performing normality tests in each group V.

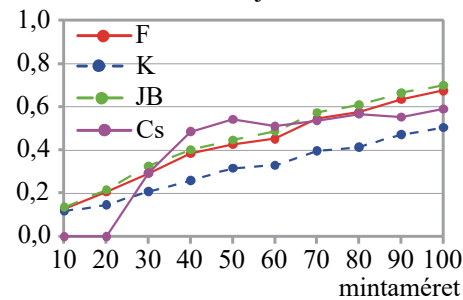
5%-ban kevert, N(0, 9) eloszlással

Tesztek statisztikai ereje



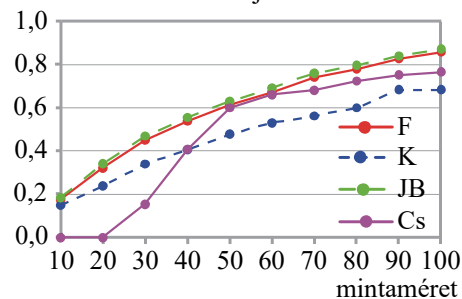
5%-ban kevert, N(3, 1) eloszlással

Tesztek statisztikai ereje



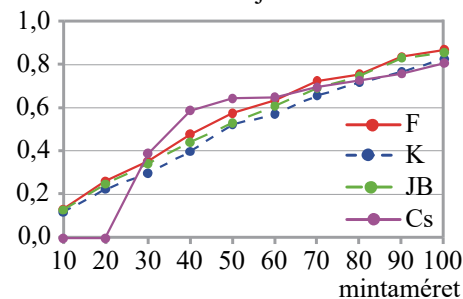
10%-ban kevert, N(0, 9) eloszlással

Tesztek statisztikai ereje



10%-ban kevert, N(3, 1) eloszlással

Tesztek statisztikai ereje



Az 5. ábra a tesztcsoportonkénti legmagasabb statisztikai erejű normalitástesztet mutatja kevert eloszlások esetén, a vizsgált minta elemszámának függvényében, 5%-os szignifikanciaszint mellett. A legjobban a Filliben- (F-), a Kuiper- (K-), a Jarque–Bera- (JB-) és Khí-négyzet- (CS-) teszt teljesít.

Más tanulmányokhoz hasonlóan azt tapasztaljuk, hogy a használt kevert eloszlás arányának növelésével pontosabbak lesznek a tesztek (Chen, 1971). A négy eloszláscsoporttól eltérően a kevert eloszlások esetén nem figyelhető meg a tesztek statisztikai erejének javulása a szórás növekedésével.

5.6. Az eredmények összefoglalása

Kutatásunkban a Shapiro–Wilk- (SW-), a Shapiro–Francia- (SF-), a Filliben- (F-), a Kolmogorov–Smirnov- (KS-), az Anderson–Darling- (AD-), a Kuiper- (K-), a Cramér–von Mises- (CvM-), a D’Agostino–Pearson- (DP-), a Jarque–Bera- (JB), a Khí-négyzet- (CS-) és a Vasicek- (V-) normalitásteszt statisztikai erejét hasonlítottuk össze Monte Carlo-szimuláció segítségével 10-től 100-ig terjedő mintákon, öt eloszláscsoportot figyelembe véve.

Az eredményeket a 6. ábra összegzi, amely a legjobban teljesítő teszteket mutatja az egyes eloszlás-elemszám párokra. A bal oldali ábrán a hőtércép a legjobban és a második legjobban teljesítő próba közötti pontosságot mutatja, míg a jobb oldali ábrán a legjobban teljesítő teszt statisztikai ereje látható.

Az összes eloszlás-elemszám párra a 240 esetből 184-ben találunk olyan, legjobban teljesítő próbát, mint amilyen a Filliben- (7%), a Shapiro–Wilk- (41%) a Kuiper- (1%), a D’Agostino–Pearson- (5%), a Jarque–Bera- (15%), a Khí-négyzet- (19%) és a Vasicek- (12%) teszt volt.

A 6. ábra alapján az első, regressziós és korrelációs tesztcsoport statisztikai ereje a legmagasabb a vizsgált mintának az első két eloszláscsoportján. Jóllehet a teszt kevésbé ismert,¹⁶ a $(-\infty, \infty)$ tartójú szimmetrikus eloszlásokon a Filliben-teszt statisztikai ereje a legmagasabb. Ezen az eloszláscsoporton az egyes tesztcsoportokban legjobban teljesítő tesztek közötti különbség magas, és egyik teszt statisztikai ereje sem éri el a megválasztott 95%-os megbízhatósági szintet. Az aszimmetrikus esetben a Shapiro–Wilk-próba teljesít a legjobban, és statisztikai ereje a legtöbb esetben a mintaelemszám növelésével rövid időn belül 1-hez konvergál.

Amennyiben $(0, 1)$ tartójú eloszlás a prior feltevésünk, a Shapiro–Wilk- és a Vasicek-teszt általában a legpontosabb. A Vasicek-próba statisztikai ereje gyakran 10–20%-kal magasabb a második legjobban teljesítő próbánál. Mindazonáltal a tesztet megfigyelésünk alapján ritkán alkalmazzák a statisztikusok és az ökonometerek az empirikus kutatásokban, és a korábbi tanulmányok közül csak ketten vizsgálták a próba statisztikai erejét (*Noughabi–Arghami, 2001; Yaizici–Yolacan, 2007*). Ennek egyik oka az lehet, hogy a teszt empirikus α értéke (5. táblázat), a megválasztott 5%-os szignifikanciaszinttől 3–5%-kal eltér. Ha a teszt empirikus alkalmazásakor szeretnénk elérni a Monte Carlo-szimulációban kapott magas statisztikai pontosságot, szükségünk van a kritikus értékek korrigálására.

A $(0, \infty)$ tartójú eloszlások esetében a Shapiro–Wilk- és a Vasicek-teszt statisztikai ereje a legmagasabb, és a legjobban teljesítő próbák közötti eltérés alacsony. Az egyenletes eloszlásoknál a D’Agostino–Pearson-próbát is alkalmazhatjuk.

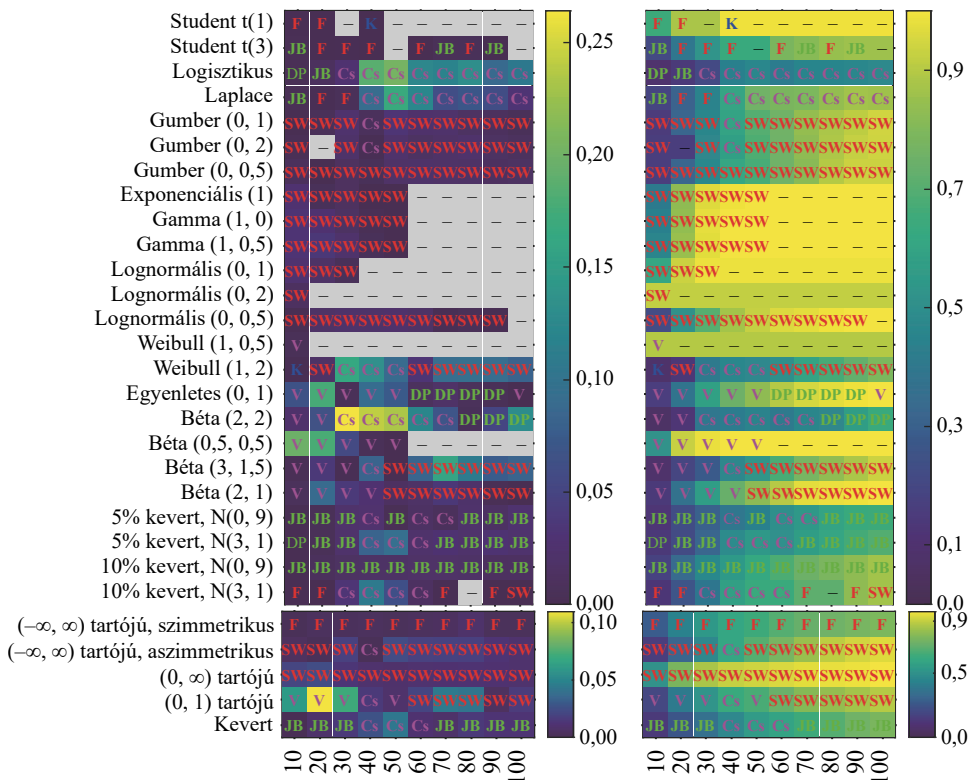
¹⁶ Korábbi kutatásokban csak *Romao és munkatársai (2010)* vizsgálták a Filliben-próba statisztikai erejét kevert eloszlásokra fókuszálva.

A kevert eloszlások esetében a Jarque–Bera-teszt statisztikai ereje a legmagasabb, de a Filliben- és a Khí-négyzet-teszt is alkalmazható.

6. ábra

Az eloszlásonként és eloszláscsoportként legjobban teljesítő normalitásteszt
Best performing normality tests for each distribution and group of distributions

Legjobb teszt **addicionális statisztikai ereje** **Legjobb teszt statisztikai ereje**



Az ábra az eloszlásonként és eloszláscsoportonként legjobban teljesítő tesztet mutatja a minta elemszámának függvényében. Pirossal jelöltük a regressziós és a korrelációs teszteket, kézzel az eloszlásfüggvényen alapuló teszteket, zölddel a momentumpróbákat, és lilával az egyéb tesztcsoportot. A bal oldali ábrán a legjobban teljesítő teszt rövidítése található az egyes (eloszlás, elemszám) kombinációkra, és a hőtérkép megmutatja, hogy a második legjobban teljesítő próba statisztikai ereje átlagosan mennyivel kisebb adott eloszlás és elemszám mellett. A jobb oldali ábrán a legjobban teljesítő teszt statisztikai erejét láthatjuk. Az üresen hagyott cellák esetén nem létezik egyetlen legjobban teljesítő teszt.

Azon eloszlások esetén, amelyek több paraméterkombinációval szerepelnek – például a béta-, a lognormális és a Weibull-eloszlás –, a magasabb varianciájú eloszlásnál a tesztek statisztikai ereje is nagyobb. Kivételt képeznek a kevert eloszlások. A tesztek általános hatékonyságát összefoglalva azt látjuk, hogy több eloszlás

vagy eloszláscsoport esetén még 100-as elemszámú mintára sem 95%-os a legjobban teljesítő próbák megbízhatósági szintje. Véleményünk szerint a Monte Carlo-szimulációnk jól kiegészíti a korábbi kutatásokat, amelyek a Filliben-, a Kuiper- és a Vasicek-próbát nem, vagy csak ritkán vonták be az összehasonlító elemzéseikbe.

6. Konklúzió

Kutatásunkban a korábbi tanulmányok leggyakrabban alkalmazott normalitás-tesztjei közül tizenegyet összehasonlítva vizsgáltuk a próbák statisztikai erejét. Az összehasonlításához normálistól különböző eloszláscsoportokból generáltunk mintát Monte Carlo-szimulációval: $(-\infty, \infty)$ tartójú szimmetrikus, aszimmetrikus, $(0, \infty)$ tartójú, $(0, 1)$ tartójú és kevert eloszlásokból. Az ökonometriai idősorok gyakran 100-nál kisebb elemszámú megfigyelést tartalmaznak, ezért kutatásunkban a 10, 20, 30, ..., 100 elemszámú mintaméreteket vizsgáltuk.

A $(-\infty, \infty)$ tartójú szimmetrikus és aszimmetrikus eloszlásokon a regressziós és korrelációs tesztek statisztikai ereje a legmagasabb. Monte Carlo-szimuláció alapján a $(0, \infty)$ tartójú eloszlások esetén a Shapiro–Wilk- és a Vasicek-próbát érdemes alkalmazni, míg a $(0, 1)$ tartójú eloszlásokon a Shapiro–Wilk- és a Vasicek-próba a leghatékonyabb. A kevert eloszlásokon a Jarque–Bera-próba statisztikai ereje volt a legmagasabb. Emellett a normalitás tesztek jobban teljesítenek, ahogy az eloszlás szórása növekszik, a kevert eloszlásoktól eltekintve.

A korábbi tanulmányoktól és az alkalmazásoktól eltérően a Filliben- és a Vasicek-próba statisztikai ereje kiemelkedően magas volt az első, illetve a negyedik eloszláscsoporton. Ennek az az egyik oka, hogy korábban kevés összehasonlító kutatás vizsgálta ezen teszteket, illetve az empirikus és a megválasztott szignifikanciaszint jelentősen különbözik a két teszt esetében. Emiatt fontosnak tartjuk kiemelni, hogy a kapott eredmények és a próbák rangsorolás szerinti alkalmazása csak korrigált kritikus értékek mellett tükrözi a Monte Carlo-szimulációban kapott eredményeket.

További kutatási célunk kidolgozni a normális eloszlás tesztelésére egy alternatív módszert, amely a kapott eredményeknél magasabb statisztikai erővel rendelkezik a kis minták esetén. Az új teszt megoldást kínálna azon kutatási területeken, ahol a mintaelemszám növelése nem lehetséges, azonban a normalitás tesztek alkalmazása gyakori. Véleményünk szerint a legjobb próbák kombinálása lehetséges alternatíva, korábban az ökonometrián belül kointegrációs (*Bayer–Hanck, 2013*), illetve egységgyöktesztek (*Harvey et al., 2009*) esetében már többször sikeresen alkalmazták. További kutatási terület lehet a normalitás tesztek közötti eltérés matematikai magyarázata az egyes eloszlások és eloszláscsoportok esetén.

Függelék

A. Az irodalmi áttekintést összefoglaló táblázatok

Az irodalmi áttekintés során ismertetjük azon kutatásokat, amelyek a normalitás-tesztek erejét Monte Carlo-szimulációval vizsgálták. Ebben az alfejezetben a szakirodalom bemutatását kiegészítve egy további táblázatban szemléltetjük a korábbi kutatásokban vizsgált eloszlásokat.

Az A1. táblázat összesíti a korábbi kutatásokban vizsgált eloszlásokat a kutatásunkban szereplő öt eloszláscsoport alapján. A sorokban szereplő számok megmutatják, hogy az egyes eloszláscsoportokból az adott kutatást hány eloszlást vizsgált. Fontos kiemelni, hogy a vizsgált eloszlások vagy azok paraméterei eltérhetnek a kutatásunkban szereplő eloszlásoktól.

A1. táblázat

A szakirodalomban vizsgált eloszlások a kutatásban használt eloszláscsoportok alapján
Summary of the distributions applied in the existing existing literature

| Szerzők | I | II | III | IV | Kevert eloszlások |
|------------------------------|---|----|-----|----|-------------------|
| Thadewald és Brünnin (2007) | 0 | 0 | 0 | 30 | 0 |
| Yaizi és Yolcan (2007) | 1 | 0 | 3 | 1 | 0 |
| Noughabi és Arghani (2011) | 3 | 3 | 8 | 6 | 0 |
| Razali et al. (2011) | 5 | 0 | 2 | 4 | 0 |
| Yap és Sim (2011) | 6 | 8 | 1 | 6 | 2 |
| Romao et al. (2010) | 1 | 1 | 13 | 4 | 41 |
| Mbah és Paothong (2015) | 6 | 0 | 0 | 1 | 0 |
| Ahmad és Khan (2015) | 2 | 0 | 2 | 0 | 0 |
| Wijekularathna et al. (2020) | 5 | 0 | 4 | 2 | 6 |
| Simić (2021) | 3 | 3 | 8 | 6 | 4 |

Az eloszláscsoportokat (I, II, III, IV és kevert eloszlások) a 4. és az 5. táblázatban ismertetjük.

B. Normalitás tesztek részletes leírása

A normális eloszlás tesztelése fejezetben bemutatjuk és röviden csoportosítjuk azokat a tesztek, amelyeket a Monte Carlo-szimuláció során összehasonlítottunk.

Ebben az alfejezetben a tesztek összehasonlítását kiegészítve részletesen bemutatjuk az egyes normalitástesztek teststatisztikájának kiszámolási módját.

B.1. Regressziós és korrelációs tesztek

A regressziós és korrelációs próbák felhasználják azt a transzformációt, hogy $X \sim \mathcal{N}(\mu, \sigma^2)$ eloszlású változó felírható $X = \mu + \sigma Z$ alakban, ahol $Z \sim \mathcal{N}(0, 1)$. Az elemzésbe bevont próbák a Shapiro–Wilk- (*Shapiro–Wilk, 1965*) a Shapiro–Francia- (*Shapiro–Francia, 1972*) és a Filliben- (*Filliben, 1975*) teszt.

B.1.1. Shapiro–Wilk-teszt

Legyen $\mathbf{Z}' = (z_1, z_2, z_3, \dots, z_n)$ a megfigyeléstől független, standard normális eloszlású rendezett minta \mathbf{V} kovarianciamátrixszal és $\mathbf{m}' = (m_1, m_2, m_3, \dots, m_n)$ várható értékkel. Ha az x_i normális eloszlású, akkor felírható $x_i = \mu + \sigma z_i$ alakban $\forall i = 1, 2, 3, \dots, n$. A Shapiro–Wilk-tesztstatisztika:

$$W = \frac{(\sum_{i=1}^n a_i \cdot x_i)^2}{\sum_{i=1}^n x_i - \bar{x}}$$

ahol \bar{x} a mintából becsült átlag és az $\mathbf{a}' = (a_1, a_2, a_3, \dots, a_n)$ vektor:

$$\mathbf{a}' = \mathbf{m}'\mathbf{V}^{-1}(\mathbf{m}'\mathbf{V}^{-1}\mathbf{m})^{-(1/2)}$$

B.1.2. Shapiro–Francia-teszt

A Shapiro–Francia-teszt a Shapiro–Wilk-próba egy leegyszerűsítése arra az esetre, amikor a \mathbf{V} kovarianciamátrix nem ismert. Ekkor $\mathbf{V} = \mathbf{I}$ helyettesítéssel a minta függetlenségét feltételezzük és így számítjuk ki a teststatisztikát.

$$W = \frac{(\sum_{i=1}^n a_i \cdot x_i)^2}{\sum_{i=1}^n x_i - \bar{x}}$$

ahol: $\mathbf{a}' = \mathbf{m}'(\mathbf{m}'\mathbf{m})^{-(1/2)}$

B.1.3. Filliben-teszt

Legyen $\mathbf{x}' = (x_1, x_2, x_3, \dots, x_n)$ a rendezett véletlen minta, és legyenek $m_{(1)}, m_{(2)}, m_{(3)}, \dots, m_{(n)}$ a $[0, 1]$ tartójú egyenletes eloszlásból szimulált rendezett minta mediánértékei.

$$m_{(i)} = \begin{cases} 1 - 0,5^{(1/n)} & i = 1 \\ \frac{(i - 0.3175)}{(n + 0.365)} & 1 < i < n \\ 0,5^{(1/n)} & i \geq n \end{cases}$$

Az elméleti eloszlás rendezett mintájának mediánja az $M_{(i)} = \phi^{-1}(m_{(i)})$ a $[0, 1]$ transzformációval kapható meg. A Filliben-tesztstatisztika az elméleti és az empirikus rendezett statisztikák közötti korrelációs együttható:

$$r = \frac{\sum_{i=1}^n x_{(i)} \cdot M_{(i)}}{\sqrt{\sum_{i=1}^n M_{(i)}^2} \cdot \sqrt{(n-1) \cdot s^2}}$$

B.2. Empirikus eloszlásfüggvénytesztek

Az eloszlásfüggvényeken alapuló tesztek a tapasztalati és az elméleti eloszlásfüggvények különbségét hasonlítják össze. Az elemzésbe bevont próbák a Cramér–von Mises- (*Von Mises, 1931*), Kolmogorov–Smirnov- (*Kolmogorov, 1933*), az Anderson–Darling- (*Anderson–Darling, 1954*) és a Kuiper- (*Kuiper, 1960*) teszt.

B.2.1. Kolmogorov–Smirnov-teszt

A Kolmogorov–Smirnov-tesztet nem egy eloszlástípusra, hanem egy eloszlásra lehet felírni. Emiatt csak az tesztelhető, hogy a minta $\mathcal{N}(\bar{x}, \text{std}(x)^2)$ eloszlásból származik, ahol a minta átlaga \bar{x} , szórása, $\text{std}(x)^2$. A teszt a megfigyelt minta illeszkedését az elméleti eloszlásra a legnagyobb vertikális különbséggel méri:

$$D = \sup_x |F_n(x) - F(x)|$$

B.2.2. Anderson–Darling-teszt

Az Anderson–Darling-tesztstatisztika:

$$AD = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \Psi(F(x)) dF(x)$$

ahol $\Psi(F(x)) = [F(x) \cdot (1 - F(x))]^{-1}$, így az empirikus és az elméleti eloszlás közötti különbségek nagyobb súlyt kapnak az eloszlás szélein. A minta elemszámát n , a tapasztalati eloszlásfüggvényt F_n jelöli.

B.2.3. Kuiper-teszt

A Kuiper-teszt a Kolmogorov–Smirnov-teszthez hasonlóan a megfigyelt minta és az elméleti eloszlás közötti maximális különbség alapján teszteli a normális eloszlástól való eltérést.

$$K = D^+ - D^-$$

ahol

$$D^+ = \sup_x (F_n(x) - F(x))$$

és

$$D^- = (F(x) - F_n(x))$$

értékek. A Kuiper-teszt és a Kolmogorov–Smirnov-teszt egymáshoz hasonló. A Kuiper-teszt jelöléseit alkalmazva a Kolmogorov–Smirnov-teszt felírható $D = \max(D^+, D^-)$ alakban.

B.2.4. Cramér–von Mises-teszt

A Cramér–von Mises-tesztstatisztika az Anderson–Darling-tesztstatisztika $\Psi(\cdot) = 1$ érték mellett.

$$\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

Emiatt az empirikus és az elméleti eloszlásfüggvény közötti különbségek azonos súlyt kapnak.

B.3. Momentumtesztek

A momentumokon alapuló normalitástesztetek általában az eloszlás ferdeségét és/vagy csúcosságát használják. A leggyakrabban alkalmazott momentumon alapuló tesztek a Jarque–Bera- (*Jarque és Bera, 1987*) és a D’Agostino–Pearson- (*D’Agostino et al., 1990*) teszt.

B.3.1. D’Agostino–Pearson-teszt

A mintából becsült ferdeség és csúcosság külön-külön alkalmazható a minta normális eloszlásának tesztelésére. *D’Agostino és munkatársai (1990)* a kettőt együtt alkalmazva az alábbi tesztstatisztikát javasolták:

$$Z^2(\sqrt{b_1}) + Z^2(b_2)$$

Ha a minta normális eloszlásból származik, $Z^2(\sqrt{b_1})$ és $Z^2(b_2)$ is normális eloszlású, és a tesztstatisztika Khí-négyzet-eloszlást követ. A ferdeség és a csúcosság is aszimptotikusan normális, azonban a konvergencia sebessége, különösen a csúcosság esetén, lassú. Ezt korrigálja a $Z^2(\cdot)$ transzformáció.

B.3.2. Jarque–Bera-teszt

A Jarque–Bera-teszt a tapasztalati eloszlásból számolt ferdeség és csúcosság értéket hasonlítja a normális eloszláséhoz. A tesztstatisztika:

$$n \left(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)}{24} \right)$$

alakú. $\sqrt{b_1}$ a ferdeséget, b_2 a csúcosságot jelöli. Aszimptotikusan a ferdeség és a csúcosság értéke 0 és 3 normális eloszlás esetén, és a variancia rendre $\frac{6}{n}$ és $\frac{24}{n}$.

A J tesztstatisztika emiatt aszimptotikusan két független standard normális eloszlású változó összege, tehát Khí-négyzet-eloszlású.

B.4. Egyéb tesztek

B.4.1. Kí-négyzet-teszt

Legyen adott $N \in \mathbb{N}$ darab tartomány, és jelölje p_j annak a valószínűségét, hogy az elméleti eloszlásból származó egyelemű minta a j -edik tartományon belüli értéket veszi fel. Jelölje O_j szimulált mintából a j -edik tartományba eső elemek számát és $E_j = p_j \cdot n$ az eloszlás szerinti várható elemek számát. A Kí-négyzet-tesztstatisztika:

$$\sum_{i=1}^n \frac{O_i - E_i}{E_i}$$

alakú.

B.4.1. Vasicek-teszt

A Vasicek-teszt (Vasicek, 1976) az eloszlásfüggvénnyel rendelkező eloszlások entrópiáján alapul:

$$H(f) = - \int_{-\infty}^{\infty} f(x) \cdot \log f(x) dx$$

amely adott $x_1, x_2, x_3, \dots, x_n$ rendezett minta esetén felírható

$$H(f) = - \int_0^1 \log \left(\frac{d}{dp} F^{-1}(p) \right) dp$$

alakban. Az eloszlásfüggvény helyére az empirikus eloszlást írva és a deriváltat a megfigyelések differenciájával közelítve:

$$H_{n,m} = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{n}{2m} \cdot (x_{i+m} - x_{i-m}) \right).$$

Shannon (2001) bizonyította, hogy minden eloszlásfüggvénnyel rendelkező eloszlásra adott variancia mellett a $H(f)$ entrópiát a normális eloszlás maximalizálja, amelynek értéke $\log(\sqrt{2\pi e} \cdot \sigma)$. A $H_{n,m}$ közelítésével van definiálva a tesztstatisztika. Legyen $m < \frac{n}{2}$ és $x_1, x_2, x_3, \dots, x_n$ rendezett minta, a Vasicek-tesztstatisztika:

$$K_{n,m} = \frac{n}{2m \cdot s} \left(\prod_{i=1}^n x_{i+m} - x_{i-m} \right)^{1/n}$$

ahol

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

a minta szórása.

A nullhipotézis fennállása esetén:

$$K_{n,m} \xrightarrow{P} \sqrt{2\pi e}, \text{ ha } n \rightarrow \infty, m \rightarrow \infty, \frac{m}{n} \rightarrow 0.$$

C. Kritikus értékek

Ebben az alfejezetben a normalitástesztek empirikus α értékeit kiegészítő C.1. táblázatként szemléltetjük a Monte Carlo-szimulációba bevont tesztek szimulált kritikus értékeit.

C.1.táblázat

A szakirodalomban vizsgált eloszlások a kutatásban használt eloszláscsoportok

| Teszt neve | $n = 10$ | $n = 20$ | $n = 30$ | $n = 40$ | $n = 50$ |
|--------------------|----------|----------|----------|----------|-----------|
| Shapiro–Wilk | 0,8446 | 0,9043 | 0,9303 | 0,9446 | 0,9539 |
| Shapiro–Francia | 0,8421 | 0,9030 | 0,9296 | 0,9443 | 0,9535 |
| Filliben | 0,9168 | 0,9497 | 0,9637 | 0,9714 | 0,9761 |
| Kolmogorov–Smirnov | 0,4102 | 0,2942 | 0,2420 | 0,2104 | 0,1889 |
| Andeson–Darling | 0,6854 | 0,7234 | 0,7328 | 0,7377 | 0,7426 |
| Kuiper | 1,6490 | 1,7513 | 1,7969 | 1,8241 | 1,8428 |
| Cramér–von Mises | 0,1254 | 0,1262 | 0,1263 | 0,1262 | 0,1267 |
| D’Agostino–Pearson | 6,5343 | 6,3859 | 6,3896 | 6,3765 | 6,3929 |
| Jarque–Bera | 2,5401 | 3,8146 | 4,3984 | 4,7597 | 5,0262 |
| Khí-négyzet | 1,3201 | 1,3841 | 3,6827 | 7,7651 | 9,2430 |
| Vasicek | 0,8575 | 0,4285 | 0,3452 | 0,3002 | 0,2726 |
| Teszt neve | $n = 60$ | $n = 70$ | $n = 80$ | $n = 90$ | $n = 100$ |
| Shapiro–Wilk | 0,9605 | 0,9653 | 0,9691 | 0,9722 | 0,9747 |
| Shapiro–Francia | 0,9603 | 0,9651 | 0,9689 | 0,9721 | 0,9746 |
| Filliben | 0,9797 | 0,9822 | 0,9842 | 0,9858 | 0,9871 |
| Kolmogorov–Smirnov | 0,1728 | 0,1600 | 0,1499 | 0,1414 | 0,1344 |
| Andeson–Darling | 0,7452 | 0,7467 | 0,7444 | 0,7418 | 0,7417 |
| Kuiper | 1,8554 | 1,8661 | 1,8752 | 1,8827 | 1,8884 |
| Cramér–von Mises | 0,1266 | 0,1266 | 0,1261 | 0,1253 | 0,1253 |
| D’Agostino–Pearson | 6,3191 | 6,3119 | 6,2862 | 6,2281 | 6,2502 |
| Jarque–Bera | 5,0969 | 5,2808 | 5,3595 | 5,3879 | 5,4084 |
| Khí-négyzet | 9,8881 | 10,3846 | 10,6044 | 10,7647 | 10,9043 |
| Vasicek | 0,2542 | 0,2409 | 0,2302 | 0,1735 | 0,1664 |

A táblázat a normalitástesztekhez tartozó kritikus értékeket tartalmazza $\alpha = 5\%$ -os szignifikanciaszint megválasztása mellett. A szimulációhoz 100 000 standard normális eloszlású mintát szimuláltunk minden $n = 10, 20, \dots, 100$ értékre.

Irodalom

- Ahmad, F. – Khan, R. A. (2015): A power comparison of various normality tests. *Pakistan Journal of Statistics and Operation Research*, 11(3), 331–345. <https://doi.org/10.18187/pjsor.v11i3.845>
- Anderson, T. W. – Darling, D. A. (1954): A test of goodness of fit. *Journal of the American Statistical Association*, 49(268), 765–769. <https://doi.org/10.1080/01621459.1954.10501232>
- Bayer, C. – Hanck, C. (2013): Combining non-cointegration tests. *Journal of Time Series Analysis*, 34(1), 83–95. <https://doi.org/10.1111/j.1467-9892.2012.00814.x>
- Chen, E. H. (1971): The power of the Shapiro-Wilk test for normality in samples from contaminated normal distributions. *Journal of the American Statistical Association*, 66(336), 760–762. <https://doi.org/10.1080/01621459.1971.10482342>
- D’Agostino, R. B. – Belanger, A. – D’Agostino Jr, R. B. (1990): A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316–321. <https://doi.org/10.1080/00031305.1990.10475751>
- Dufour, J.-M. – Farhat, A. – Gardiol, L. – Khalaf, L. (1998): Simulation-based finite sample normality tests in linear regressions. *The Econometrics Journal*, 1(1), 154–173. <https://doi.org/10.1111/1368-423X.11009>
- Esteban, M. – Castellanos, M. – Morales, D. – Vajda, I. (2001): Monte Carlo comparison of four normality tests using different entropy estimates. *Communications in Statistics-Simulation and computation*, 30(4), 761–785. <https://doi.org/10.1081/SAC-100107780>
- Filliben, J. J. (1975): The probability plot correlation coefficient test for normality. *Technometrics*, 17(1), 111–117. <https://doi.org/10.1080/00401706.1975.10489279>
- Harvey, D. I. – Leybourne, S. J. – Taylor, A. R. (2009): Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric theory*, 25(3), 587–636. <https://doi.org/10.1017/S026646660809018X>
- Jarque, C. M. – Bera, A. K. (1987): A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2), 163–172. <https://doi.org/10.2307/1403192>
- Kolmogorov, A. N. (1933): Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell’inst Ital Degli Att*, 4, 89–91.
- Kuiper, N. H. (1960): Tests concerning random points on a circle. *In Nederl. Akad. Wetensch. Proc.Ser. A*, 63, 38–47.
- Mbah, A. K. – Paothong, A. (2015): Shapiro–Francia test compared to other normality test using expected p-value. *Journal of Statistical Computation and Simulation*, 85(15), 3002–3016. <https://doi.org/10.1080/00949655.2014.947986>
- Noughabi, H. A. – Arghami, N. R. (2011): Monte carlo comparison of seven normality tests. *Journal of Statistical Computation and Simulation*, 81(8), 965–972. <https://doi.org/10.1080/00949650903580047>
- Pearson, K. (1894): Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, A, 85, 71–110.
- Razali, N. M. – Wah, Y. B. – et al. (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics*, 2(1), 21–33.
- Romao, X. – Delgado, R. – Costa, A. (2010): An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, 80(5), 545–591. <https://doi.org/10.1080/00949650902740824>

- Schorr, B. (1974): On the choice of the class intervals in the application of the chi-square test. *Mathematische Operationsforschung und Statistik*, 5(4-5), 357–377. <https://doi.org/10.1080/02331887408801174>
- Seier, E. (2002): Comparison of tests for univariate normality. *Interstat*, 1, 1–17.
- Shannon, C. E. (2001): A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shapiro, S. S. – Francia, R. (1972): An approximate analysis of variance test for normality. *Journal of the American statistical Association*, 67(337), 215–216. <https://doi.org/10.1080/01621459.1972.10481232>
- Shapiro, S. S. – Wilk, M. B. (1965): An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591–611. <https://doi.org/10.2307/2333709>
- Sigut, J. – Pineiro, J. – Estévez, J. – Toledo, P. (2006): A neural network approach to normality testing. *Intelligent Data Analysis*, 10(6), 509–519. <https://doi.org/10.3233/IDA-2006-10603>
- Simić, M. (2021): Testing for normality with neural networks. *Neural Computing and Applications*, 33(23), 16279–16313. <https://doi.org/10.1007/s00521-021-06229-7>
- Thadewald, T. – Büning, H. (2007): Jarque–bera test and its competitors for testing normality—a power comparison. *Journal of Applied Statistics*, 34(1), 87–105. <https://doi.org/10.1080/02664760600994539>
- Vasicek, O. (1976): A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(1), 54–59. <https://doi.org/10.1111/j.2517-6161.1976.tb01566.x>
- Von Mises, R. (1931): Vorlesungen aus dem gebiete der angewandten mathematik: Wahrcheinlichkeitsrechnung und ihre anwendung in der statistik und theoretischen physik. *F. Deuticke*, 1.
- Wijekularathna, D. K. – Manage, A. B. – Scariano, S. M. (2020): Power analysis of several normality tests: A Monte Carlo simulation study. *Communications in Statistics-Simulation and Computation*, 51(3), 757–773. <https://doi.org/10.1080/03610918.2019.1658780>
- Yap, B. W. – Sim, C. H. (2011): Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155. <https://doi.org/10.1080/00949655.2010.520163>
- Yazici, B. – Yolacan, S. (2007): A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, 77(2), 175–183. <https://doi.org/10.1080/10629360600678310>