Tómács Tibor

Matematikai statisztika gyakorlatok

ÚJ SZÉCHENYI TERV

ESZTERHÁZY KÁROLY KATOLIKUS EGYETEM
Matematikai és Informatikai Intézet

Tómács Tibor

Matematikai statisztika gyakorlatok

Átdolgozott kiadás

Utolsó módosítás:
2023. december 1.

A jegyzet szabadon letölthető az alábbi linkről:
https://tomacstibor.uni-eszterhazy.hu/tananyagok/Matematikai_statisztika_gyakorlatok.pdf

Eger, 2023

Szerző:
Dr. Tómács Tibor
egyetemi docens
Eszterházy Károly Katolikus Egyetem

Bíráló:
Dr. Sztrik János
egyetemi tanár
Debreceni Egyetem

Készült a TÁMOP-4.1.2-08/1/A-2009-0038 támogatásával.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Tartalomjegyzék

Előszó

Ez a tananyag az Eszterházy Károly Katolikus Egyetem matematikai statisztika gyakorlataiból készült. Alapvetően Tómács Tibor [13] tananyagára építünk, amelyben az elméleti alapok találhatóak meg. Természetesen a két műben a jelölések és a szóhasználat is megegyezik, így itt alkalmazásukkor már nem ismertetjük az elméleti részben bevezetett jelöléseket, csak összefoglaljuk a *Jelölések* című részben.

Ez a tananyag inkább számítógéppel megoldható gyakorlatokat, míg az előbb említett mű, a szükséges definíciókon és tételeken túl, elméleti számításokat igénylő feladatokat tartalmaz.

A matematikai statisztika elméletének gyakorlatba való átültetésére mindenekelőtt mintarealizációkra lesz szükségünk. Ezeket néhány esetben mi fogjuk generálni számítógéppel, de lesznek olyan esetek is, amikor adott mintát kell vizsgálnunk. A feladatoknál megadott minták a PDF-hez vannak csatolva, így olyan PDF-nézőt használjon, amely a csatolt állományok kezelését támogatja. Ilyen például a Firefox böngésző beépített PDF-nézője, az Adobe Reader vagy a SumatraPDF.

A mintagenerálást és annak statisztikai elemzését is a *Microsoft Office Excel* program magyar nyelvű változatával végezzük. Az Excel alapfokú használatát ismertnek tételezzük fel, ennek ellenére a példák megoldását olyan részletesen mutatjuk meg, amennyire csak lehet. Itt jegyezzük meg, hogy további számos programcsomag készült statisztikai adatok feldolgozására (SPSS, SAS, MatLab, Maple, R-nyelvű statisztikai rutinok, stb.).

Minden fejezet tartalmaz mintapéldákat részletesen megoldva. A fejezetek végén gyakorlatokat találhatunk, melyhez szükség szerint útmutatót is adunk.

A statisztikában szokásos táblázatokat ebben a tananyagban nem mellékeljük, mert az ezekben található értékeket számítógép segítségével fogjuk kiszámolni.

A tananyag vége egy összefoglalót tartalmaz, melyben megtalálható minden olyan információ, amely a példák és gyakorlatok megoldásához szükséges.

Jelölések

Általános

$\mathbb{N}$	a pozitív egész számok halmaza
$\mathbb{R}$	a valós számok halmaza
$\mathbb{R}^n$	$\mathbb{R}$-nek önmagával vett n-szeres Descartes-szorzata
$\mathbb{R}_+$	a pozitív valós számok halmaza
(a, b)	rendezett elempár vagy nyílt intervallum
$\simeq$	közelítőleg egyenlő
$[x]$	az x valós szám egész része
f^{-1}	az f függvény inverze
$A^\top$	az A mátrix transzponáltja
A^{-1}	az A mátrix inverze

Valószínűségszámítás

$\mathrm{P}(A)$	az A esemény valószínűsége
$\mathrm{E}\,\xi$	ξ várható értéke
$\mathrm{D}\,\xi,\ \mathrm{D}^2\,\xi$	ξ szórása illetve szórásnégyzete
$\mathrm{cov}(\xi, \eta)$	kovariancia
$\mathrm{corr}(\xi, \eta)$	korrelációs együttható
φ	a standard normális eloszlás sűrűségfüggvénye
Φ	a standard normális eloszlás eloszlásfüggvénye
Γ	Gamma-függvény
I_A	az A esemény indikátorváltozója
$\mathrm{Bin}(r; p)$	az r-edrendű p paraméterű binomiális eloszlású valószínűségi változók halmaza
$\mathrm{Exp}(\lambda)$	a λ paraméterű exponenciális eloszlású valószínűségi változók halmaza

$\mathrm{Norm}(m;\sigma)$	az m várható értékű és σ szórású normális eloszlású valószínűségi változók halmaza
$\mathrm{Gamma}(r;\lambda)$	az r-edrendű λ paraméterű gamma-eloszlású valószínűségi változók halmaza
$\mathrm{Khi}(s)$	az s szabadsági fokú khi-négyzet eloszlású valószínűségi változók halmaza
$\mathrm{T}(s)$	az s szabadsági fokú t-eloszlású valószínűségi változók halmaza
$\mathrm{F}(s_1;s_2)$	az s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változók halmaza
$F[\mathrm{V}]$	Ha ξ valószínűségi változó és V a ξ-vel azonos eloszlású valószínűségi változók halmaza, akkor $F[\mathrm{V}]$ a V-beli valószínűségi változók közös eloszlásfüggvényét jelenti. Például $\Phi = F[\mathrm{Norm}(0;1)]$.

Matematikai statisztika

F_n^*	tapasztalati eloszlásfüggvény
$\overline{\xi}$	a ξ-re vonatkozó minta átlaga (mintaátlag)
$S_n,\ S_n^2$	tapasztalati szórás illetve szórásnégyzet
$S_{\xi,n},\ S_{\xi,n}^2$	ξ-re vonatkozó tapasztalati szórás illetve szórásnégyzet
$S_n^*,\ S_n^{*2}$	korrigált tapasztalati szórás illetve szórásnégyzet
$S_{\xi,n}^*,\ S_{\xi,n}^{*2}$	ξ-re vonatkozó korrigált tapasztalati szórás illetve szórásnégyzet
$\xi_1^*,\dots,\xi_n^*$	rendezett minta
$\mathrm{Cov}_n(\xi,\eta)$	tapasztalati kovariancia
$\mathrm{Corr}_n(\xi,\eta)$	tapasztalati korrelációs együttható
$\widehat{\vartheta}$	a ϑ paraméter becslése
$H_0,\ H_1$	nullhipotézis, ellenhipotézis

1. fejezet

Mintagenerálás

A statisztikai elemzések során ismeretlen eloszlású valószínűségi változókat vizsgálunk oly módon, hogy a valószínűségi változóra több mérést is elvégzünk. A kapott számokat *mintarealizációnak* nevezzük. Ha ξ a vizsgált valószínűségi változó, akkor a rávonatkozó mintarealizáció elemeit $\xi_1(\omega), \ldots, \xi_n(\omega)$ módon jelöljük, ahol $\xi_1, \ldots, \xi_n$ a ξ-vel azonos eloszlású független valószínűségi változók, ω pedig a kísérletsorozatban bekövetkező elemi esemény. A $\xi_1, \ldots, \xi_n$ valószínűségi változókat a ξ-re vonatkozó *mintának* nevezzük.

A gyakorlati óráink során mérések helyett számítógéppel állítjuk elő a mintarealizációt. Számítógépes algoritmussal generált véletlen számot *pszeudo-* vagy *álvéletlennek* nevezzük. Például az úgynevezett kongruens módszeren alapuló algoritmust n-szer lefuttatva, a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó n elemű mintarealizációt állíthatunk elő. Ennek az elméletével itt nem foglalkozunk. (Részletesebben lásd például [1, 5, 12].)

Megjegyezzük, hogy valódi véletlent is használhatunk minta generálására a következő címen található internetes szolgáltatással: `http://www.random.org`.

Ebben a fejezetben azt fogjuk részletezni, hogy a $[0,1]$ intervallumon egyenletes eloszlásból hogyan lehet más eloszlást generálni. Emlékeztetőül, egy valószínűségi változó a $[0,1]$ intervallumon egyenletes eloszlású, ha a $[0,1]$ intervallum egy tetszőleges h hosszúságú részintervallumába h valószínűséggel eshet. A következő állítás szerint bármely eloszlású valószínűségi változó előáll $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó valamely transzformáltjaként.

1.1. Tétel. *Legyen $F\colon \mathbb{R} \to \mathbb{R}$ egy eloszlásfüggvény és*

$$G\colon \mathbb{R} \to \mathbb{R}, \quad G(x) := \begin{cases} \sup\{y \in \mathbb{R} : F(y) < x\}, & \text{ha } 0 < x < 1, \\ 0, & \text{különben.} \end{cases}$$

Ha ξ egyenletes eloszlású valószínűségi változó a $[0, 1]$ intervallumon, akkor $G(\xi)$ olyan valószínűségi változó, melynek eloszlásfüggvénye F.

1.2. *Megjegyzés.* Az 1.1. tételben, ha F invertálható eloszlásfüggvény, azaz szigorúan monoton növekvő, akkor $0 < x < 1$ esetén

$$G(x) = \sup\{y \in \mathbb{R} : F(y) < x\} = \sup\{y \in \mathbb{R} : y < F^{-1}(x)\} = F^{-1}(x).$$

Ezért a G függvény $(0, 1)$-re vett leszűkítettjét az F *általánosított inverzének* is nevezik.

1.3. *Megjegyzés.* Az 1.1. tételben, ha F egy olyan valószínűségi változó eloszlásfüggvénye, amely az $x_1 < x_2 < \cdots < x_r$ értékeket veheti fel rendre $p_1, p_2, \ldots, p_r$ valószínűségekkel, akkor

$$G(x) = \begin{cases} x_1, & \text{ha } 0 < x \le p_1, \\ x_2, & \text{ha } p_1 < x \le p_1 + p_2, \\ x_3, & \text{ha } p_1 + p_2 < x \le p_1 + p_2 + p_3, \\ \vdots & \\ x_{r-1}, & \text{ha } p_1 + \cdots + p_{r-2} < x \le p_1 + \cdots + p_{r-1}, \\ x_r, & \text{ha } p_1 + \cdots + p_{r-1} < x < 1. \end{cases}$$

Könnyen látható, hogy a $G(x)$ felírásában a $<$ és $\le$ relációs jelek tetszőlegesen felcserélhetőek, hiszen ez nem változtat a $G(\xi)$ eloszlásán.

Hasonló állítás fogalmazható meg akkor is, ha megszámlálhatóan végtelen sok értéket felvevő valószínűségi változót akarunk transzformálni egyenletes eloszlásból.

1.1. Egyenletes eloszlás

Excel-ben a `VÉL()` függvénnyel tudunk $[0, 1]$ intervallumon egyenletes eloszlású (pszeudo)véletlen számot generálni.

1.4. Példa. Generáljon $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó 20 elemű mintarealizációt.

Megoldás (1). Az `A1` cellába írja be, hogy `=VÉL()` majd *Enter*. Ezután a kitöltőjelet húzza le a 20. sorig. A *kitöltőjel* a kijelölés jobb alsó sarkában lévő négyzet.

Az így generált számok minden újraszámolásnál megváltoznak, ami nem kívánatos, hiszen a mintarealizációt a feladatokban rögzítettnek tekintjük. (Próbálja ezt ki az

F9 funkcióbillentyű megnyomásával, melynek hatására az Excel minden képletet újraszámol.) A mintarealizáció elemeinek rögzítéséhez tegye a következőket:

1. Lépjen az `A` oszlop fejlécére, nyomja meg a jobb egérgombot, majd válassza a *Másolás* pontot.
2. Lépjen a `B` oszlop fejlécére, nyomja meg a jobb egérgombot, válassza az *Irányított beillesztés* pontot, jelölje be az *Értéket*, majd nyomja meg az *OK* gombot.

Megoldás (2). Az előző megoldás abban az esetben nem kényelmes, ha 20 helyett például 10 000 elemű mintarealizációt kell generálni, hiszen ekkor a kitöltőjelet az `A10000` celláig kellene lehúzni, ami sokáig tart. Ebben az esetben a következő módszer praktikusabb.

Lépjen az `A1` cellára, a Név mezőbe írja be, hogy `A10000`, majd *Shift* + *Enter*. Ez kijelöli az `A1:A10000` cellatartományt. Írja be, hogy `=VÉL()` majd *Ctrl* + *Enter*, ami a kijelölt cellatartomány minden cellájába `=VÉL()` függvényt ír. Ezután az előző módszerrel a `B` oszlopba rögzítheti a mintát.

Megoldás (3). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. Először aktiválja az *Analysis ToolPak* bővítményt: *Fájl/Beállítások/Bővítmények* majd *Ugrás* gomb. Pipálja ki az *Analysis ToolPak* sort majd *OK* (lásd a 10.7.1. alszakaszt).

Ezután *Adatok/Adatelemzés*, a felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza az *Egyenletes* sort, változók számát állítsa `1`-re (mert csak egy mintát generálunk), a véletlen számok számát állítsa `20`-ra (mert a mintaelemek száma 20). A paraméterek legyenek `0` valamint `1`. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `A1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció az `A` oszlopban. Ennek rögzítésére nincs szükség, mert a cellákban csak számokat ír be a program, nem függvényeket.

Az (1) és (3) megoldást megnézheti a következő videón:

A példák megoldásánál az Adatok/Adatelemzés *menüpontot a teljesség kedvéért mutatjuk meg. Dolgozatírásnál kérjük ne használja, mert a javítás során nem lehet visszakövetni a mintarealizáció készítésének menetét.*

A következő tétel azt mutatja meg, hogy egy $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változóból hogyan transzformálhatunk tetszőleges $[a, b]$ intervallumon egyenletes eloszlású valószínűségi változót.

1.5. Tétel. *Legyen* ξ *a* $[0, 1]$ *intervallumon egyenletes eloszlású valószínűségi változó és* $a, b \in \mathbb{R},\ a < b$*. Ekkor* $a + (b - a)\xi$ *az* $[a, b]$ *intervallumon egyenletes eloszlású.*

Bizonyítás. A tétel az 1.1. tételből következik, hiszen $G(x) = F^{-1}(x) = a + (b - a)x$, ha $0 < x < 1$. □

1.6. Példa. Generáljon $[-2, 5]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó 100 elemű mintarealizációt.

Megoldás (1)*.* Az előző tétel alapján, ha ξ a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változó, akkor $-2 + (5 - (-2))\xi = -2 + 7\xi$ a $[-2, 5]$ intervallumon egyenletes eloszlású. Tehát az `A1` cellába írja be, hogy `=-2+7*VÉL()`, *Enter*, az `A1` cella kitöltőjelét húzza le a 100. sorig, majd rögzítse a mintarealizáció elemeit.

Megoldás (2)*.* Hasonlóan az előző feladathoz ez is megoldható az *Adatok/Adatelemzés* menüponttal, csak itt a paraméterek `-2` valamint `5` lesznek, illetve a véletlen számok száma `100`.

1.2. Exponenciális eloszlás

1.7. Tétel. *Legyen* ξ *a* $[0, 1]$ *intervallumon egyenletes eloszlású valószínűségi változó és* $\lambda > 0$*. Ekkor* $\frac{-\ln\xi}{\lambda}$ *exponenciális eloszlású* λ *paraméterrel.*

Bizonyítás. Ha $x > 0$, akkor $\mathrm{P}\left(\frac{-\ln\xi}{\lambda} < x\right) = \mathrm{P}(\xi > e^{-\lambda x}) = 1 - e^{-\lambda x}$, illetve ha $x \leq 0$, akkor $\mathrm{P}\left(\frac{-\ln\xi}{\lambda} < x\right) = 0$. □

1.8. *Megjegyzés.* Az 1.1. tételben $G(x) = F^{-1}(x) = \frac{-\ln(1-x)}{\lambda}$, ha $0 < x < 1$, vagyis $\frac{-\ln(1-\xi)}{\lambda}$ is exponenciális eloszlású λ paraméterrel.

1.9. Példa. Generáljon $\lambda = 5{,}6$ paraméterű exponenciális eloszlású valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Megoldás. Az `A1` cellába írja be, hogy `=-LN(VÉL())/5,6`, a kitöltőjelet húzza le a 10. sorig, majd rögzítse a mintarealizáció elemeit.

1.3. Normális eloszlás

A következő tétel az 1.1. tétel következménye.

1.10. Tétel. *Ha* ξ *egyenletes eloszlású valószínűségi változó a* $[0, 1]$ *intervallumon,* $m \in \mathbb{R}$*,* $\sigma > 0$ *és* $F = F[\mathrm{Norm}(m; \sigma)]$*. Ekkor* $F^{-1}(\xi)$ *illetve* $m + \sigma\Phi^{-1}(\xi)$ *is normális eloszlású valószínűségi változók* m *és* σ *paraméterekkel.*

Ebben a tételben F^{-1} illetve Φ^{-1} nem elemi függvények. A következő tétel azt mutatja, hogy elemi függvénnyel is megkaphatjuk a normális eloszlást az egyenletesből.

1.11. Tétel (Box–Muller-transzformáció)**.** *Legyenek ξ, η a $[0,1]$ intervallumon egyenletes eloszlású független valószínűségi változók, $m \in \mathbb{R}$ és $\sigma > 0$. Ekkor*

$$m + \sigma\sqrt{-2\ln\xi}\cos(2\pi\eta)$$

normális eloszlású m várható értékkel és σ szórással.

1.12. Példa. Generáljon $m = 4$ várható értékű és $\sigma = 1{,}2$ szórású normális eloszlású valószínűségi változóra vonatkozó 20 elemű mintarealizációt.

Megoldás (1). Az előző két tétel alapján az `A1` cellába írja be az alábbiak egyikét:

```
=NORM.INVERZ(VÉL();4;1,2)
=4+1,2*NORM.S.INVERZ(VÉL())
=4+1,2*GYÖK(-2*LN(VÉL()))*COS(2*PI()*VÉL())
```

Nyomjon *Enter*-t, lépjen vissza `A1`-re, ezután a kitöltőjelet húzza le a 20. sorig, majd rögzítse a mintarealizáció elemeit.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Normális* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `20`-ra. A várható érték `4`, a szórás `1,2`. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `A1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció az `A` oszlopban.

1.4. Diszkrét egyenletes eloszlás

1.13. Tétel. *Legyen ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó, $m \in \mathbb{N}$ és $x_1, \dots, x_m$ különböző valós számok. Ekkor $x_{[m\xi]+1}$ diszkrét egyenletes eloszlású az $\{x_1, \dots, x_m\}$ halmazon.*

Bizonyítás. A tétel az 1.3. megjegyzés következménye, de közvetlenül is bizonyítható, hiszen $\mathrm{P}(x_{[m\xi]+1} = x_i) = \mathrm{P}([m\xi] + 1 = i) = \mathrm{P}\left(\frac{i-1}{m} \leq \xi < \frac{i}{m}\right) = \frac{1}{m}$. □

1.14. Példa. Modellezzen 10 dobást egy szabályos kockával. Másképpen fogalmazva, generáljon az $\{1,2,3,4,5,6\}$ halmazon diszkrét egyenletes eloszlású valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Megoldás (1). Az előző tétel alapján, ha ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó, akkor $[6\xi]+1$ diszkrét egyenletes eloszlású az $\{1,2,3,4,5,6\}$ halmazon. Így `A1`-be írja be, hogy `=INT(6*VÉL())+1` vagy az ezzel egyenértékű `VÉLETLEN.KÖZÖTT(1;6)` függvényt, *Enter*, az `A1` cella kitöltőjelét húzza le a 10. sorig, majd rögzítse a mintarealizáció elemeit.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. Az `A` oszlopba írja be `1`-től `6`-ig az egész számokat, a `B1`,…,`B6` cellákba pedig az `=1/6`-t. Ezután *Adatok/Adatelemzés*, a felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Diszkrét* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `10`-re. Kattintson az *Érték és valószínűség bemeneti tartománya* felirat mezőjére, majd jelölje ki az `A1:B6` cellatartományt. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `C1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció a `C` oszlopban. Ez utóbbi megoldás videón:

1.5. Karakterisztikus eloszlás

1.15. Tétel. *Legyen ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó és $0<p<1$. Ekkor $\mathrm{I}_{\xi<p}$ karakterisztikus eloszlású p paraméterrel, ahol I az indikátorváltozót jelenti.*

Bizonyítás. A tétel az 1.3. megjegyzés következménye, de közvetlenül is bizonyítható, hiszen $\mathrm{P}(\mathrm{I}_{\xi<p}=1)=\mathrm{P}(\xi<p)=p$ és $\mathrm{P}(\mathrm{I}_{\xi<p}=0)=\mathrm{P}(\xi\geq p)=1-p$. □

1.16. Példa. Figyeljen meg 30 független kísérletben egy 0,4 valószínűségű eseményt oly módon, hogy ha bekövetkezik, akkor leírja az 1 számot, míg ha nem, akkor a 0 számot. Másképpen fogalmazva, generáljon $p=0{,}4$ paraméterű karakterisztikus eloszlású (vagy más néven *Bernoulli-eloszlású*) valószínűségi változóra vonatkozó 30 elemű mintarealizációt.

Megoldás (1). Az `A1` cellába írja be, hogy `=HA(VÉL()<0,4;1;0)`. Nyomjon *Enter*-t, melynek hatására, ha `VÉL()<0,4` teljesül, akkor az eredmény 1, különben 0. Lépjen vissza `A1`-re, ezután a kitöltőjelet húzza le a 30. sorig, majd rögzítse a mintarealizáció elemeit.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Bernoulli* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `30`-ra, a p értékét pedig `0,4`-re. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `A1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció az `A` oszlopban.

1.6. Binomiális eloszlás

Ismert, hogy r darab független p paraméterű karakterisztikus eloszlású valószínűségi változó összege r-edrendű p paraméterű binomiális eloszlású. Ebből következően teljesül a következő tétel.

1.17. Tétel. *Legyenek $\xi_1, \ldots, \xi_r$ a $[0,1]$ intervallumon egyenletes eloszlású független valószínűségi változók és $0 < p < 1$. Ekkor $\sum_{i=1}^{r} \mathrm{I}_{\xi_i < p}$ r-edrendű p paraméterű binomiális eloszlású.*

1.18. *Megjegyzés.* Az 1.3. megjegyzés alapján $G(\xi)$ is r-edrendű p paraméterű binomiális eloszlású, ahol ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó és $p_k := \binom{r}{k} p^k (1-p)^{r-k}$ jelöléssel

$$G(x) = \begin{cases} 0, & \text{ha } 0 < x \le p_0, \\ 1, & \text{ha } p_0 < x \le p_0 + p_1, \\ 2, & \text{ha } p_0 + p_1 < x \le p_0 + p_1 + p_2, \\ \vdots & \\ r-1, & \text{ha } p_0 + \cdots + p_{r-2} < x \le p_0 + \cdots + p_{r-1}, \\ r, & \text{ha } p_0 + \cdots + p_{r-1} < x < 1. \end{cases} \tag{1.1}$$

A $p_0 + \cdots + p_k$ értéke $\boxed{\texttt{BINOM.ELOSZL(}k\texttt{;}r\texttt{;}p\texttt{;IGAZ)}}$ módon számolható ki Excelben.

1.19. Példa. Generáljon egy 0,8 valószínűségű esemény 5 kísérlet utáni gyakoriságára vonatkozó 20 elemű mintarealizációt. Másképpen fogalmazva, generáljon $r = 5$ rendű $p = 0{,}8$ paraméterű binomiális eloszlású valószínűségi változóra vonatkozó 20 elemű mintarealizációt.

Megoldás (1). Az előző tétel és a karakterisztikus eloszlás generálásánál leírtak alapján az `A1` cellába írja be, hogy `=HA(VÉL()<0,8;1;0)`. A kitöltőjelet húzza jobbra az `E` oszlopig. Az `F1` cellába írja be, hogy `=SZUM(A1:E1)`, vagy nyomja meg az *Alt+Shift+7* gombokat, majd nyomjon *Enter*-t. Jelölje ki az `A1:F1` cellatartományt, majd a kitöltőjelet húzza le a 20. sorig. Ekkor a mintarealizáció az `F` oszlopban lesz. Végül rögzítse a mintarealizáció elemeit. Mindez videón:

Megoldás (2). Másik megoldásként használjuk fel az 1.18. megjegyzést. A `B1` cellába írja be a `0` számot, a `C1`-be az `1`-et, és így tovább, a `G1`-be az `5`-öt. Az `A2` cellába írjon `0`-t, míg a `B2`-be a következőt: `=BINOM.ELOSZL(B1;5;0,8;IGAZ)`. A `B2` kitöltőjelét húzza `G2`-ig. Írja az `A3`-ba, hogy `=VÉL()`, a `B3`-ba pedig `=HA(ÉS($A3>A$2;$A3<=B$2);B$1;0)`. A `B3` kitöltőjelét húzza `G3`-ig. A `H3`-ba írja `=SZUM(B3:G3)`. Ezután az `A3:H3` cellatartomány kitöltőjelét húzza le a 22. sorig. Ekkor a mintarealizáció a `H` oszlopban lesz. Végül rögzítse a mintarealizáció elemeit.

Megoldás (3). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Binomiális* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `20`-ra, a p értékét `0,8`-re, a kísérletek számát pedig `5`-re. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `A1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció az `A` oszlopban.

1.7. Hipergeometrikus eloszlás

Legyen egy dobozban N darab golyó, melyből M darab piros. Visszatevés nélkül kiveszünk véletlenszerűen r darab golyót a dobozból. Legyen ξ a kivett piros golyók száma. Ekkor a ξ valószínűségi változó hipergeometrikus eloszlású N, M, r paraméterekkel, ahol feltételezzük, hogy $M < N$ és $r \leq \min\{M, N-M\}$.

A következő állítás ennek a kísérletnek a modellezése alapján bizonyítható.

1.20. Tétel. *Legyen $y_1, \dots, y_r$ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó mintarealizáció, és*

$$x_0 := 0, \quad x_i := \begin{cases} x_{i-1}+1, & \textit{ha } y_i < \dfrac{M-x_{i-1}}{N-i+1}, \\ x_{i-1}, & \textit{különben}, \end{cases} \quad (i = 1, \dots, r).$$

Ekkor x_r a ξ-re vonatkozó 1 elemű mintarealizáció.

1.21. *Megjegyzés.* Az 1.3. megjegyzés alapján $G(\xi)$ is N, M, r paraméterű hipergeometrikus eloszlású, ahol $M < N$, $r \leq \min\{M, N-M\}$, ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó és $p_k := \frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}}$ jelöléssel (1.1) teljesül. A $p_0 + \cdots + p_k$ értéke $\boxed{\texttt{HIPGEOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}M\texttt{;}N\texttt{;IGAZ)}}$ módon számolható ki Excelben.

1.22. Példa. Generáljon $N = 10$, $M = 5$, $r = 4$ paraméterű hipergeometrikus eloszlású valószínűségi változóra vonatkozó 20 elemű mintarealizációt.

Megoldás (1). Először az 1.20. tétel alapján oldjuk meg a feladatot. Az `A1`, `B1`, `C1`, `D1` cellákba írja be rendre az `1`, `2`, `3`, `4` számokat. Az `A2`-be írja be, hogy `=0`, a `B2`-be pedig, hogy `=HA(VÉL()<(5-A2)/(10-A$1+1);A2+1;A2)`. A `B2` kitöltőjelét húzza `E2`-ig. Ezután az `A2:E2` cellatartomány kitöltőjelét húzza le a 21. sorig. Ekkor a mintarealizáció az `E` oszlopban lesz. Végül rögzítse a mintarealizáció elemeit.

Megoldás (2). Második megoldásként használjuk fel az 1.21. megjegyzést. A `B1` cellába írja be a `0` számot, a `C1`-be az `1`-et, és így tovább, az `F1`-be a `4`-et. Az `A2` cellába írjon `0`-t, míg a `B2`-be a következőt: `=HIPGEOM.ELOSZLÁS(B1;4;5;10;IGAZ)`. A `B2` kitöltőjelét húzza `F2`-ig. Írja az `A3`-ba `=VÉL()`, a `B3`-ba pedig `=HA(ÉS($A3>A$2;$A3<=B$2);B$1;0)`. A `B3` kitöltőjelét húzza `F3`-ig. A `G3`-ba írja `=SZUM(B3:F3)`. Ezután az `A3:G3` cellatartomány kitöltőjelét húzza le a 22. sorig. Ekkor a mintarealizáció a `G` oszlopban lesz. Végül rögzítse a mintarealizáció elemeit.

Megoldás (3). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. Az `A` oszlopba írja `0`-tól `4`-ig az egész számokat, a `B1`-be pedig

`=HIPGEOM.ELOSZLÁS(A1;4;5;10;HAMIS)`.

A `B1` cella kitöltőjelére kattintson kétszer. Ezután *Adatok/Adatelemzés*, a felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Diszkrét* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `20`-ra. Kattintson az *Érték és valószínűség bemeneti tartománya* felirat mezőjére, majd jelölje ki az `A1:B5` cellatartományt. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `C1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció a `C` oszlopban.

1.8. Gyakorlatok

1.1. gyakorlat. Generáljon $\mu = 4$ és $\sigma = 0{,}5$ paraméterű Cauchy-eloszlású valószínűségi változóra vonatkozó 15 elemű mintarealizációt.

Útmutatás. Ha ξ és η független standard normális eloszlású valószínűségi változók, akkor $\mu + \sigma\frac{\xi}{\eta}$ Cauchy-eloszlású μ és σ paraméterekkel. Azt is felhasználhatja, hogy a μ és σ paraméterű Cauchy-eloszlás eloszlásfüggvényének inverze $F^{-1}(x) = \mu +$ $+ \sigma \operatorname{tg} \frac{\pi}{2}(2x - 1)$.

1.2. gyakorlat. Generáljon $s = 4$ szabadsági fokú khi-négyzet eloszlású valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Útmutatás. Ha $\xi_1, \ldots, \xi_s$ standard normális eloszlású független valószínűségi változók, akkor a $\xi_1^2 + \cdots + \xi_s^2$ valószínűségi változó s szabadsági fokú khi-négyzet eloszlású. Azt is felhasználhatja, hogy az s szabadsági fokú khi-négyzet eloszlás eloszlásfüggvényének inverzét a $\boxed{\texttt{KHINÉGYZET.INVERZ(}x\texttt{;}s\texttt{)}}$ függvénnyel számolhatja ki a $0 < x < 1$ helyen.

1.3. gyakorlat. Generáljon $s = 4$ szabadsági fokú t-eloszlású valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Útmutatás. Ha ξ standard normális eloszlású és η s szabadsági fokú khi-négyzet eloszlású független valószínűségi változók, akkor a $\xi\sqrt{\frac{s}{\eta}}$ valószínűségi változó s szabadsági fokú t-eloszlású. Azt is felhasználhatja, hogy az s szabadsági fokú t-eloszlás eloszlásfüggvényének inverzét a $\boxed{\texttt{T.INVERZ(}x\texttt{;}s\texttt{)}}$ függvénnyel számolhatja ki a $0 < x < 1$ helyen.

1.4. gyakorlat. Generáljon $s_1 = 2$ és $s_2 = 3$ szabadsági fokú F-eloszlású valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Útmutatás. Ha ξ s_1 szabadsági fokú és η s_2 szabadsági fokú khi-négyzet eloszlású független valószínűségi változók, akkor az $\frac{s_2\xi}{s_1\eta}$ valószínűségi változó s_1 és s_2 szabadsági fokú F-eloszlású. Azt is felhasználhatja, hogy az s_1 és s_2 szabadsági fokú F-eloszlás eloszlásfüggvényének inverzét az $\boxed{\texttt{F.INVERZ(}x\texttt{;}s_1\texttt{;}s_2\texttt{)}}$ függvénnyel számolhatja ki a $0 < x < 1$ helyen.

1.5. gyakorlat. Generáljon $r = 3$ rendű $\lambda = 2{,}1$ paraméterű gamma-eloszlású valószínűségi változóra vonatkozó 15 elemű mintarealizációt.

Útmutatás. Legyenek a $\xi_1, \ldots, \xi_r$ azonos λ paraméterű exponenciális eloszlású független valószínűségi változók. Ekkor a $\xi_1 + \cdots + \xi_r$ valószínűségi változó r-edrendű λ paraméterű gamma-eloszlású. Azt is felhasználhatja, hogy az r rendű λ paraméterű gamma-eloszlás eloszlásfüggvényének inverzét a $\boxed{\texttt{GAMMA.INVERZ(}x\texttt{;}r\texttt{;}1/\lambda\texttt{)}}$ függvénnyel számolhatja ki a $0 < x < 1$ helyen.

1.6. gyakorlat. Egy valószínűségi változó az $x_1 = 1{,}1$, $x_2 = 2{,}2$ és $x_3 = 3{,}3$ értékeket veheti fel, rendre $p_1 = 0{,}2$, $p_2 = 0{,}3$ és $p_3 = 0{,}5$ valószínűségekkel. Generáljon erre a valószínűségi változóra vonatkozó 10 elemű mintarealizációt.

Megoldás (1). Először az 1.3. megjegyzés segítségével oldjuk meg a feladatot. A `B1`, `C1`, `D1` cellákba rendre írja be az `1,1`, `2,2` és `3,3` értékeket, a `B2`, `C2`, `D2` cellákba pedig a `0,2`, `0,3` és `0,5` értékeket. Az `A3`-ba írjon `0`-t, majd `B3`-ba `=A3+B2`. A `B3` kitöltőjelét húzza `D3`-ig. Az `A4`-be `=VÉL()`, majd `B4`-be `=HA(ÉS($A4>A$3;$A4<=B$3);B$1;0)`. A `B4` kitöltőjelét húzza `D4`-ig. Az `E4`-be `=SZUM(B4:D4)`. Ezután az `A4:E4` cellatartomány kitöltőjelét húzza le a 14. sorig. A mintarealizáció ekkor az `E` oszlopban van. Végül rögzítse a mintarealizáció elemeit.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. Az `A` oszlopba írja be az x_i értékeit, a `B`-be pedig a p_i értékeit. Ezután *Adatok/Adatelemzés*, a felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Diszkrét* sort, változók számát állítsa `1`-re, a véletlen számok számát állítsa `10`-re. Kattintson az *Érték és valószínűség bemeneti tartománya* felirat mezőjére, majd jelölje ki az `A1:B3` cellatartományt. Kattintson a *Kimeneti tartomány* feliratra, majd a mellette lévő mezőre, végül az `C1` cellára. Az *OK* gomb megnyomásával elkészül a mintarealizáció a `C` oszlopban.

1.7. gyakorlat. Egy kísérletet ismételjünk egymástól függetlenül, amíg egy rögzített A esemény be nem következik. Legyen ξ a végrehajtott kísérletek száma. A ξ valószínűségi változót *geometriai eloszlásúnak* nevezzük. Írjon programot, mely ξ-re vonatkozó mintarealizációt generál. Oldja meg a feladatot Excelben is.

Útmutatás. Tegyük fel, hogy a vizsgált A esemény valószínűsége p. Legyen $y_1, \dots, y_r$ a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó olyan mintarealizáció, melyre teljesül, hogy

$$y_1 \geq p,\ y_2 \geq p,\ \dots,\ y_{r-1} \geq p \quad \text{és} \quad y_r < p.$$

Ha $y_1 < p$, akkor legyen $r = 1$. Könnyű belátni, hogy az így definiált r a ξ-re vonatkozó 1 elemű mintarealizáció.

Az 1.3. megjegyzéssel Excelben is megoldható a feladat, ahol $x_k = k$ és $p_k = = p(1-p)^{k-1}$ $(k = 0, 1, 2, \dots)$.

1.8. gyakorlat. Írjon programot, mely Poisson-eloszlású valószínűségi változóra vonatkozó mintarealizációt generál. Oldja meg a feladatot Excelben is.

Útmutatás. Legyen $y_0, y_1, \dots, y_r$ a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó olyan mintarealizáció, melyre teljesül, hogy

$$y_0 y_1 \cdots y_{r-1} \geq e^{-\lambda} \quad \text{és} \quad y_0 y_1 \cdots y_r < e^{-\lambda}.$$

Ekkor r a ξ-re vonatkozó 1 elemű mintarealizáció.

A feladat Excelben az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Véletlenszám-generálás* sort, majd *OK*. Az eloszlásnál válassza a *Poisson* sort.

2. fejezet

Tapasztalati eloszlás

Ebben a fejezetben a generált mintarealizáció alapján fogjuk grafikusan megbecsülni a valószínűségi változó eloszlásfüggvényét, diszkrét esetben az eloszlását, illetve abszolút folytonos esetben a sűrűségfüggvényét.

2.1. Tapasztalati eloszlásfüggvény

Egy valószínűségi változó eloszlásfüggvényének az értéke adott $x \in \mathbb{R}$ helyen annak a valószínűsége, hogy a valószínűségi változó x-nél kisebb értéket vesz fel. Ezt a gyakorlatban nem ismerjük, így a relatív gyakorisággal fogjuk becsülni, amit *tapasztalati eloszlásfüggvénynek* nevezünk az x helyen és $F_n^*(x)$ módon jelölünk. Tehát az F_n^* tapasztalati eloszlásfüggvény értéke adott $x \in \mathbb{R}$ helyen az x-nél kisebb elemek száma a mintarealizációban, osztva a mintarealizáció elemeinek a számával. Ez egy olyan lépcsős függvény, melyben a szakadási pontok a mintarealizáció értékeinél vannak. Pontosabban, ha a mintarealizáció $x_1 = \xi_1(\omega), \ldots, x_n = \xi_n(\omega)$, akkor az $(x_i, F_n^*(x_i))$ koordinátájú pontok az F_n^* „lépcsőfokainak” a jobb oldali végpontjai. A legmagasabb lépcsőfok kezdőpontja a $(\max\{x_1, \ldots, x_n\}, 1)$ koordinátájú pont.

2.1. Tétel (A matematikai statisztika alaptörvénye)**.** *A tapasztalati eloszlásfüggvény* 1 *valószínűséggel egyenletesen konvergál* $\mathbb{R}$*-en a valódi eloszlásfüggvényhez, azaz nagy elemszámú mintarealizáció esetén a tapasztalati eloszlásfüggvény jól közelíti a valódit.*

Ezt a törvényt többféle eloszlással is bemutatjuk a következő videóban.

Az itt használt program letölthető a következő helyről:

https://tomacstibor.uni-eszterhazy.hu/tananyagok/valdem/valdem.zip

A gyakorlatok során azzal a könnyítéssel oldjuk meg a feladatokat, hogy a lépcsőfokok jobb végpontjait összekötjük a következő lépcsőfok bal végpontjával. Így egy folytonos vonalat kapunk. Sajnos az Excelben nincs lépcsős diagramtípus (szemben a LibreOffice programmal), ezért a lépcsőfokok mindkét végének koordinátáit meg kell adni és azokat összekötni szakaszokkal.

2.2. Példa. Modellezzen 100 dobást egy szabályos kockával, azaz generáljon az $\{1, 2, 3, 4, 5, 6\}$ halmazon diszkrét egyenletes eloszlású valószínűségi változóra vonatkozó 100 elemű mintarealizációt. Ábrázolja a kapott mintarealizációhoz tartozó tapasztalati eloszlásfüggvényt.

Megoldás. Az `A1` cellába írja be, hogy `=VÉLETLEN.KÖZÖTT(1;6)`, nyomjon *Enter*-t, lépjen vissza `A1`-re, ezután a kitöltőjelet húzza le a 100. sorig, majd rögzítse a mintarealizáció elemeit a `B` oszlopba. A `C1:C6` cellatartományba írja be rendre az 1, 2, 3, 4, 5, 6 számokat, azaz a lehetséges dobásértékeket. Ezután a `D` és `E` oszlopok első 4 sorába írja a következőket:

	D	E
1	`=C1-1`	`0`
2	`=C1`	`0`
3	`=D2`	`=E4`
4	`=INDEX(C:C;SOR(D4)/2)`	`=DARABTELI(B:B;"<"&D4)/DARAB(B:B)`

A `SOR(D2)` értéke a `D2` cella sorának száma, azaz 2, továbbá `INDEX(C:C;2)` a `C2` cella értékével fog megegyezni. A `DARABTELI(B:B;"<"&D4)` a `B` oszlopban a `D4` cella értékénél kisebb értékek száma, továbbá `DARAB(B:B)` a `B` oszlopban található értékek száma. *Figyelem! A* ***`DARABTELI`*** *függvénynév egy téves fordítás eredménye, amit 2023. márciusa utáni Excel-verziókban kijavítottak* ***`DARABHA`*** *névre. Az eredeti angol elnevezésben nincs változás, ezért a* ***`DARATELI`*** *nem szerepel az elavult függvénynevek között.*

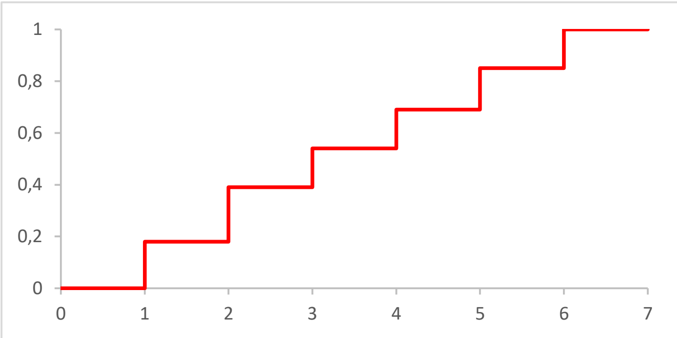
A $(\texttt{D1}, \texttt{E1})$ illetve $(\texttt{D2}, \texttt{E2})$ koordinátájú pontok az első lépcsőfok bal illetve jobb végpontjai, míg a $(\texttt{D3}, \texttt{E3})$ illetve $(\texttt{D4}, \texttt{E4})$ koordinátájú pontok a második lépcsőfok bal illetve jobb végpontjai. A `D1`-ben csak arra kell figyelni, hogy `C1`-nél kisebb érték szerepeljen.

A többi lépcsőfokot úgy kapjuk meg, hogy a `D3:E4` cellatartomány kitöltőjelét lehúzzuk a 14. sorig. Azért 14, mert minden lépcsőhöz 2 sor tartozik, továbbá a 6 lehetséges különböző értékhez 1-gyel több, azaz 7 lépcsőfok tartozik. Általánosságban tehát, ha r darab különböző lehetséges érték van, akkor $2(r+1)$-edik sorig kell lehúzni.

Az utolsó lépcsőfok még nem jó, hiszen `D14` üres cellára hivatkozik. Ezért javítsa ki azt `=D13+1`-re. Itt csak arra kell ügyelni, hogy `D13`-nál nagyobb érték legyen.

Az ábra elkészítéséhez jelölje ki a `D` és `E` oszlopokat, majd

Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont vonalakkal*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

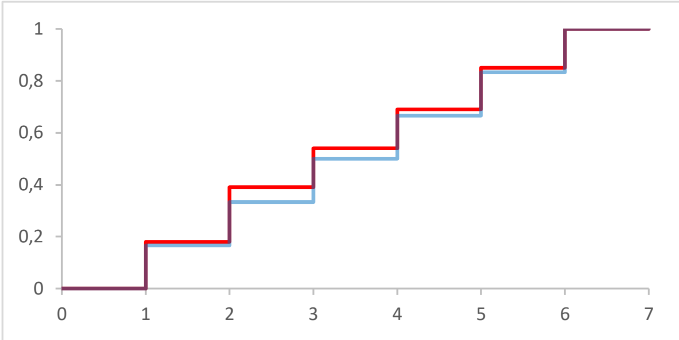
2.3. Példa. Az előző példában kapott grafikonon rajzolja fel a valódi eloszlásfüggvényt, azaz a $\{1,2,3,4,5,6\}$ halmazon diszkrét egyenletes eloszlású valószínűségi változó eloszlásfüggvényét is.

Megoldás. Az előző munkalapon dolgozzon. Az `F` oszlop első négy sorába írja a következőket.

	F
1	0
2	0
3	=F4
4	=F2+1/6

Ezek megadják a valódi eloszlásfüggvény első két lépcsőfokának magasságát. Az `F3:F4` cellatartomány kitöltőjelét húzza le a 14. sorig, amivel megkapjuk a többi lépcsőfok magasságát is. Ezután kattintson a grafikonra, majd

Tervezés → *Adatok kijelölése* → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$D$1:$D$14` →
Adatsor Y értékei: `=Munka1!$F$1:$F$14` → *OK* → *OK*



A színt és a vonalvastagságot igény szerint beállíthatja. Érdemes megnézni a két függvény viszonyát nagyobb (pl. 1000) mintaelemszám esetén is. A megoldás menetét végigkövetheti a következő videón.

2.4. Példa. Generáljon $\lambda = 3$ paraméterű exponenciális eloszlású valószínűségi változóra vonatkozó 100 elemű mintarealizációt. Ábrázolja a kapott mintarealizációhoz tartozó tapasztalati eloszlásfüggvényt.

Megoldás. A mintarealizációt korábban láttuk hogyan kell generálni. Az `A1` cellába írja be, hogy `=-LN(VÉL())/3`, a kitöltőjelet húzza le a 100. sorig, majd rögzítse a mintarealizáció elemeit a `B` oszlopba.

Mivel folytonos az eloszlás, ezért itt nem tudjuk felsorolni a lehetséges értékeket, mint diszkrét esetben. Ehelyett a mintarealizáció elemeit sorba rendezzük, hiszen itt lesznek a tapasztalati eloszlásfüggvény töréspontjai. A rendezéshez kattintson a `B` oszlop jelölőjére, majd

Kezdőlap $\to$ *Rendezés és szűrés* $\to$ *Rendezés méret szerint (növekvő)* $\to$
Folytatja az aktuális kijelöléssel $\to$ *Rendezés*

Ezután hasonlóan járunk el, mint a diszkrét egyenletes eloszlásnál. Töltse ki a `C1:D4` cellatartományt a következő módon:

	C	D
1	0	0
2	=B1	0
3	=C2	=D4
4	=INDEX(B:B;SOR(C4)/2)	=DARABTELI(B:B;"<"&C4)/DARAB(B:B)

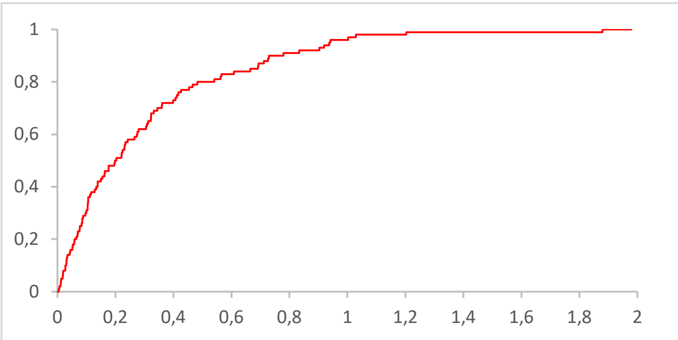
A (`C1`, `D1`) illetve (`C2`, `D2`) koordinátájú pontok az első lépcsőfok bal illetve jobb végpontjai, míg a (`C3`, `D3`) illetve (`C4`, `D4`) koordinátájú pontok a második lépcsőfok bal illetve jobb végpontjai. A `C1`-be azért került `0`, mert az exponenciális eloszlás esetén negatív értékeket 1 valószínűséggel nem vehet fel a valószínűségi változó értéke. Általánosságban arra kell ügyelni, hogy `C1`-be a `B1` értékénél kisebb szám kerüljön.

A többi lépcsőfokot úgy kapjuk meg, hogy a `C3:D4` cellatartomány kitöltőjelét lehúzzuk a 202. sorig. Azért 202, mert minden lépcsőhöz 2 sor tartozik, továbbá a 100 mintarealizáció értékhez 1-gyel több, azaz 101 lépcsőfok tartozik. Általánosságban tehát, ha n elemű a mintarealizáció, akkor $2(n+1)$-edik sorig kell lehúzni.

Az utolsó lépcsőfok még nem jó, hiszen `C202` üres cellára hivatkozik. Ezért javítsa ki azt `=C201+0,1`-re. Itt csak arra kell figyelni, hogy a `C201` cellaértéknél nagyobb szám szerepeljen.

Az ábra elkészítéséhez jelölje ki a `C` és `D` oszlopokat, majd

Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont vonalakkal*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

2.5. Példa. Az előző grafikonon ábrázolja a valódi eloszlásfüggvényt, azaz a $\lambda = 3$ paraméterű exponenciális eloszlású valószínűségi változó eloszlásfüggvényét, és hasonlítsa össze a tapasztalati eloszlásfüggvénnyel.

Megoldás. Az eloszlásfüggvényt jelen esetben 0-tól 2-ig kell megrajzolni. Ehhez ezen az intervallumon 0,1 lépésközökkel kiszámoljuk a függvényértékeket, majd egy sima (differenciálható) görbével összekötve ábrázoljuk.

A megoldást az előző munkalapon végezze el. Az `E1` cellába írjon `0`-t, az `E2`-be pedig `0,1`-et. Az `E1:E2` cellatartomány kitöltőjelét húzza le addig, amíg 2-t nem kap (21. sor). Az `F1` cellába írja a következőt:

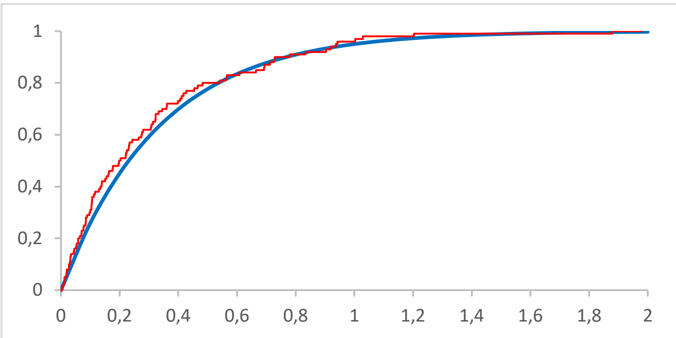
`=EXP.ELOSZL(E1;3;IGAZ)`

Ez a $\lambda = 3$ paraméterű exponenciális eloszlású valószínűségi változó eloszlásfüggvényének értékét adja az `E1` értékének a helyén. Ezután az `F1` kitöltőjelét húzza le addig, amíg az `E` oszlopban van mellette adat (most a 21. sorig). Kattintson a grafikonra, majd

Tervezés → *Adatok kijelölése* → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$E$1:$E$21` →
Adatsor Y értékei: `=Munka1!$F$1:$F$21` → *OK* → *OK*

A `21` helyére értelemszerűen az a sorszám kerül, ameddig az `E` oszlopban vannak számok. Ezt automatikusan is elvégezheti a következő módon: Kattintson az *Adatsor X értékei* alatti mezőre, utána az `E1` cellára, majd nyomjon *Ctrl+Shift+Le* billentyűkombinációt. Hasonlóan, kattintson az *Adatsor Y értékei* alatti mezőre, törölje a benne található `={1}` tartalmat, majd kattintson az `F1` cellára, végül nyomjon *Ctrl+Shift+Le* billentyűkombinációt.

Formátum → bal oldali legördülő listában: *Adatsor2* → *Kijelölés formázása* →
Adatsor formázása panelen:
Kitöltés és vonal ikon → *Vonal* → *Görbített vonal*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

2.2. Vonaldiagram

Diszkrét ξ valószínűségi változóra vonatkozó $x_1, \dots, x_n$ mintarealizáció esetén a *tapasztalati eloszlás* x_i-hez ($i = 1, \dots, n$) hozzárendeli az x_i-vel egyenlő elemek számát a mintarealizációban, elosztva n-nel, amely a $\{\xi = x_i\}$ esemény relatív gyakorisága. A statisztika alaptörvénye szerint ez a $\mathrm{P}(\xi = x_i)$ eloszlásérték annál pontosabb becslése, minél nagyobb a minta elemszáma. Ezt a függvényt célszerű *vonaldiagrammal* ábrázolni, amely azt jelenti, hogy az $(x_i, 0)$ pontot összekötjük az (x_i, p_i) ponttal $(i = 1, \ldots, n)$, ahol p_i a tapasztalati eloszlás értéke az x_i helyen.

Diszkrét eloszlás esetén célszerűbb tapasztalati eloszlásfüggvény helyett tapasztalati eloszlást ábrázolni, mert az jóval karakterisztikusabb.

2.6. Példa. Generáljon $r = 5$ rendű és $p = 0{,}3$ paraméterű binomiális eloszlású valószínűségi változóra vonatkozó 100 elemű mintarealizációt. Ábrázolja a kapott mintarealizációhoz tartozó tapasztalati eloszlást vonaldiagrammal.

Megoldás. A korábban ismertetett módon generálja le a mintarealizációt, majd rögzítse azt a `G` oszlopban. A `H1:H6` cellatartományba írja be a valószínűségi változó lehetséges 0, 1, 2, 3, 4, 5 értékeit. Ezután ezekhez az értékekhez kiszámoljuk a tapasztalati eloszlás értékeket. Ez a korábbi módszer logikájával

`=DARABTELI(G:G;"="&H1)/DARAB(G:G)`

módon történhet, de ez ekvivalens a következő `I1` cellába írásával:

`=DARABTELI(G:G;H1)/DARAB(G:G)`.

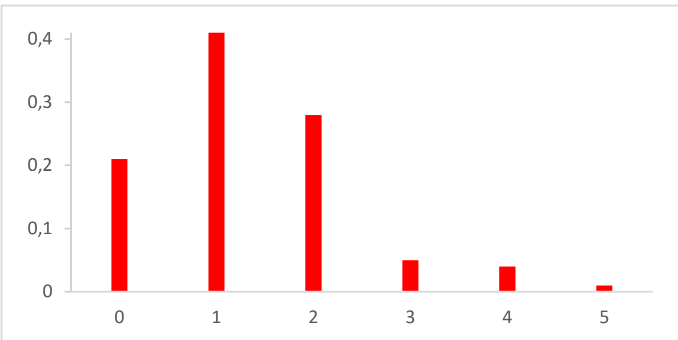
Nyomjon *Enter*-t, majd az `I1` cella kitöltőjelét húzza le a 6. sorig.

A következőkben ábrázoljuk a (`H1`, `I1`), ..., (`H6`, `I6`) koordinátájú pontokhoz tartozó vonaldiagramot, ami a tapasztalati eloszlás vonaldiagramja lesz. Ehhez jelölje ki az `I1:I6` cellatartományt, majd

Beszúrás → *Diagramok* → *Oszlop- és sávdiagram beszúrása* → *Csoportosított oszlop*

Kattintson az ábrára, majd

Tervezés → *Adatok kijelölése* →
Vízszintes (kategória-) tengely feliratai: Szerkesztés →
Tengely felirattartománya: `=Munka1!$H$1:$H$6` → *OK*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

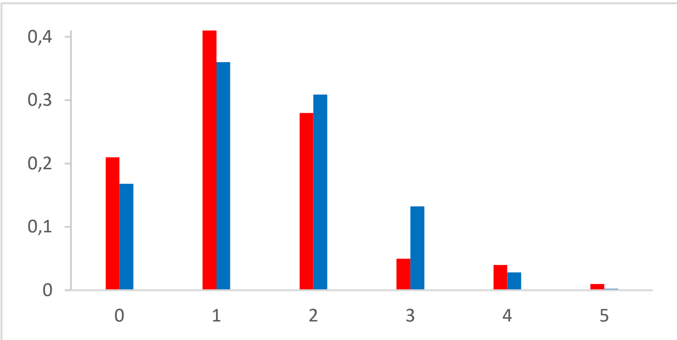
2.7. Példa. Az előző grafikonon ábrázolja a valódi eloszlást is vonaldiagrammal.

Megoldás. A megoldást az előző munkalapon végezze el. A `J1` cellába írja a következőt:

`=BINOM.ELOSZL(H1;5;0,3;HAMIS)`

Ez kiszámolja az eloszlás értékét a `H1` cella értékének a helyén. A `J1` cella kitöltőjelét húzza le a 6. sorig. Kattintson az ábrára, majd

Tervezés → *Adatok kijelölése* → *Hozzáadás* →
Adatsor értékei: `=Munka1!$J$1:$J$6` → *OK* → *OK*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

2.3. Sűrűséghisztogram

Legyen $x_0 < x_1 < \cdots < x_r$. Tegyük fel, hogy az abszolút folytonos ξ-re vonatkozó mintarealizáció minden eleme benne van az (x_0, x_r) intervallumban. Minden $[x_{j-1}, x_j)$ intervallum fölé rajzoljunk egy y_j magasságú téglalapot úgy, hogy a téglalap területe a ξ valódi f sűrűségfüggvényének görbéje alatti területet becsülje az $[x_{j-1}, x_j)$ intervallumon. Hasonlóan az eddigiekhez, egy esemény valószínűségét itt is az esemény relatív gyakoriságával becsüljük. Így tehát

$$\int_{x_{j-1}}^{x_j} f(x)\,\mathrm{d}x = \mathrm{P}(x_{j-1} \leq \xi < x_j) \simeq$$

$$\simeq \frac{[x_{j-1}, x_j) \text{ intervallumba eső mintaelemek száma}}{n} = y_j(x_j - x_{j-1}),$$

Ebből

$$y_j = \frac{[x_{j-1}, x_j) \text{ intervallumba eső mintaelemek száma}}{n(x_j - x_{j-1})} \quad (j = 1, \ldots, r).$$

A kapott oszlopdiagramot *sűrűséghisztogramnak* nevezzük, amely tehát a valódi f sűrűségfüggvényt a j-edik részintervallumon az y_j konstanssal közelíti. A statisztika alaptörvénye szerint ez a közelítés a minta elemszámának növelésével egyre pontosabb.

Abszolút folytonos eloszlás esetén érdemes a tapasztalati eloszlásfüggvény mellett ezt is ábrázolni, mert karakterisztikusabb az alakja.

2.8. Példa. Generáljon standard normális eloszlású valószínűségi változóra vonatkozó 200 elemű mintát. Rajzolja meg a sűrűséghisztogramot a $(-4, 4)$ intervallumon 10 darab egyenlő hosszúságú részintervallum esetén.

Megoldás. Generálja le a mintarealizációt az `A` oszlopban a korábban tanult módszerek valamelyikével, például a következő függvénnyel: `=NORM.S.INVERZ(VÉL())`. Rögzítse a `B` oszlopba. Töltse ki a `C1:D4` cellatartományt a következő módon:

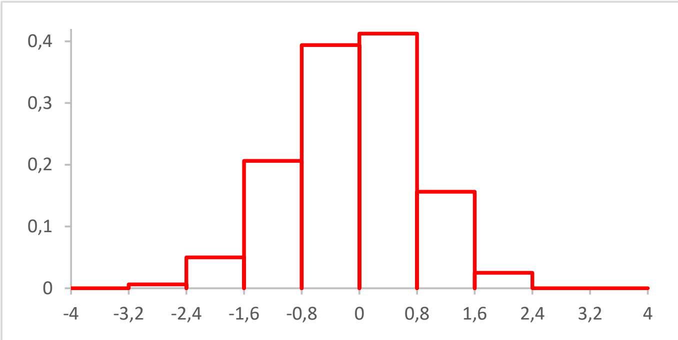
	C	D
1	-4	0
2	=C1	=DARABHATÖBB(B:B;">="&C2;B:B;"<"&C3)/(0,8*DARAB(B:B))
3	=C2+0,8	=D2
4	=C3	=0

A (`C1`, `D1`), (`C2`, `D2`), (`C3`, `D3`), (`C4`, `D4`) koordinátájú pontok rendre az első téglalap bal alsó, bal felső, jobb felső és jobb alsó csúcspontjainak koordinátái. A `D2` és `C3` cellákban azért szerepel 0,8, mert egy részintervallum hossza $\frac{4-(-4)}{10} = 0{,}8$.

Mivel a második téglalap bal alsó csúcspontja egybeesik az első téglalap jobb alsó csúcspontjával, ezért a további téglalapok meghatározásához elég a `C2:D4` cellatartomány kitöltőjelét lehúzni a 31. sorig lehúzni. Azért 31, mert összesen 10 részintervallum van és mindegyikhez 3 sor tartozik, kivéve az első intervallumot, amelyhez 4. Általánosságban, ha r darab részintervallum van, akkor a $(3r + 1)$-edik sorig kell lehúzni.

Az ábra elkészítéséhez jelölje ki a `C` és `D` oszlopokat, majd

Beszúrás → Diagramok →
Pont- (xy) vagy buborékdiagram beszúrása → Pont vonalakkal



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

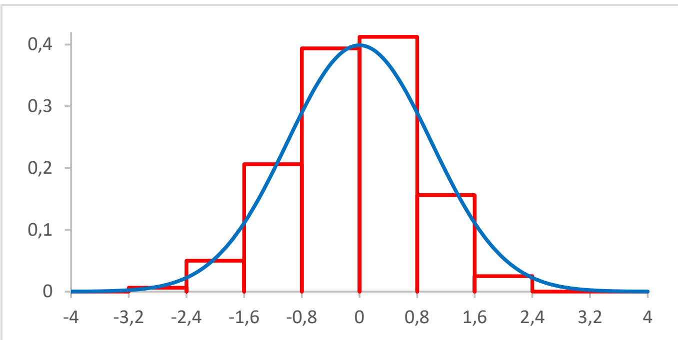
2.9. Példa. Az előző grafikonban ábrázolja a standard normális eloszlás sűrűségfüggvényét is, majd hasonlítsa össze a kapott sűrűséghisztogrammal.

Megoldás. A megoldást az előző munkalapon végezze el. Először a valódi sűrűségfüggvény értékeit a $[-4, 4]$ intervallumon fogjuk kiszámolni 0,2 lépésközzel. Írja be az `E1` cellába, hogy `-4` illetve az `E2` cellába, hogy `-3,8`. Az `E1:E2` cellatartomány kitöltőjelét húzza le a `4` értékig (41. sorig). Ezután az `F1` cellában számolja ki a standard normális eloszlás sűrűségfüggvényének értékét az `E1` cella értékénél:

`=NORM.S.ELOSZLÁS(E1;HAMIS)`

Az `F1` cella kitöltőjelét húzza le a 41. sorig. A következőkben megrajzoljuk a valódi sűrűségfüggvényt. Lépjen a diagram területére, majd

Tervezés → *Adatok kijelölése* → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$E$1:$E$41` →
Adatsor Y értékei: `=Munka1!$F$1:$F$41` → *OK* → *OK* →
Formátum → bal oldali legördülő listában: *Adatsor2* → *Kijelölés formázása* →
Adatsor formázása panelen:
Kitöltés és vonal ikon → *Vonal* → *Görbített vonal*



A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

2.4. Gyakorlatok

2.1. gyakorlat. Generáljon az $\{1, 2, 3, 4, 5\}$ halmazon diszkrét egyenletes eloszlású valószínűségi változóra vonatkozó 200 elemű mintarealizációt. Ábrázolja a kapott mintarealizációhoz tartozó tapasztalati eloszlást, majd a valódi eloszlást vonaldiagrammal.

2.2. gyakorlat. Generáljon $r = 6$ rendű $p = 0{,}6$ paraméterű binomiális eloszlású valószínűségi változóra vonatkozó 200 elemű mintarealizációt. Ebből rajzolja meg a tapasztalati eloszlásfüggvényt. Ábrázolja a valódi eloszlásfüggvényt is, majd hasonlítsa őket össze.

Útmutatás. A vizsgált valószínűségi változót jelölje ξ. Értékkészlete $\{0, 1, \ldots ,6\}$, így a valódi eloszlásfüggvénynek ezekben a pontokban kell kiszámolni az értékét. Ismert, hogy ξ eloszlásfüggvénye a $k = 1, 2, 3, 4, 5, 6$ értékeknél

$$\mathrm{P}(\xi < k) = \sum_{i=0}^{k-1} \binom{r}{i} p^i (1-p)^{r-i}.$$

Az ábrázolásnál használja fel, hogy Excel-ben

$$\boxed{\texttt{BINOM.ELOSZL(}k\texttt{;}r\texttt{;}p\texttt{;IGAZ)}} = \sum_{i=0}^{k} \binom{r}{i} p^i (1-p)^{r-i},$$

így

$$\boxed{\texttt{BINOM.ELOSZL(}k-1\texttt{;}r\texttt{;}p\texttt{;IGAZ)}} = \mathrm{P}(\xi < k) \ \ (k = 1, 2, 3, 4, 5, 6).$$

2.3. gyakorlat. Legyen egy dobozban $N = 10$ darab golyó, melyből $M = 5$ darab piros. Visszatevés nélkül kiveszünk véletlenszerűen $r = 4$ darab golyót a dobozból. Legyen ξ a kivett piros golyók száma. (Tehát ξ hipergeometrikus eloszlású.) Generáljon ξ-re vonatkozó 250 elemű mintarealizációt. Ebből rajzolja meg a tapasztalati eloszlást vonaldiagrammal. Ábrázolja a valódi eloszlást is vonaldiagrammal, majd hasonlítsa őket össze.

Útmutatás. Ismert, hogy

$$\mathrm{P}(\xi = k) = \frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}} \quad (k = 0, \ldots, r),$$

mely Excel-ben $\boxed{\texttt{HIPGEOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}M\texttt{;}N\texttt{;HAMIS)}}$ függvénnyel számolható.

2.4. gyakorlat. Generáljon $\lambda = 3{,}2$ paraméterű Poisson-eloszlású valószínűségi változóra vonatkozó 900 elemű mintarealizációt. Ebből rajzolja meg a tapasztalati eloszlást vonaldiagrammal. Ábrázolja a valódi eloszlást is vonaldiagrammal, majd hasonlítsa őket össze. Ezután ábrázolja a tapasztalati eloszlásfüggvényt azon intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, majd a valódi eloszlásfüggvényt ugyanezen az intervallumon.

Útmutatás. A vizsgált valószínűségi változót jelölje ξ. Ismert, hogy

$$\mathrm{P}(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0, 1, \ldots),$$

mely Excel-ben $\boxed{\texttt{POISSON.ELOSZLÁS(}k\texttt{;}\lambda\texttt{;HAMIS)}}$ függvénnyel számolható. Ha a `HAMIS` szó helyett `IGAZ` szerepel a függvényben, akkor az a

$$\mathrm{P}(\xi \leq k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!}e^{-\lambda}$$

értékét számolja ki.

2.5. gyakorlat. Egy kísérletet ismételjünk egymástól függetlenül, amíg egy rögzített $p = 0{,}3$ valószínűségű esemény be nem következik. Legyen ξ a végrehajtott kísérletek száma. (Tehát ξ geometriai eloszlású valószínűségi változó.) Generáljon ξ-re vonatkozó 700 elemű mintarealizációt. Ebből rajzolja meg a tapasztalati eloszlást vonaldiagrammal. Ábrázolja a valódi eloszlást is vonaldiagrammal, majd hasonlítsa őket össze. Ezután ábrázolja a tapasztalati eloszlásfüggvényt azon intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, majd a valódi eloszlásfüggvényt ugyanezen az intervallumon.

Útmutatás. Ismert, hogy

$$\mathrm{P}(\xi = k) = p(1-p)^{k-1} \quad (k = 1, 2, \ldots).$$

Excel-ben a hatványozás ^ jellel vagy a $\boxed{\texttt{HATVÁNY}}$ függvénnyel történik. Például $0{,}7^3$

$\boxed{\texttt{0,7^3}}$ vagy $\boxed{\texttt{HATVÁNY(0,7;3)}}$ módon számolható ki. Másrészt

$$\mathrm{P}(\xi \le k) = \sum_{i=1}^{k-1} p(1-p)^{i-1} = 1 - (1-p)^{k-1} \quad (k = 2, 3, \dots).$$

2.6. gyakorlat. Generáljon a ξ valószínűségi változóra vonatkozó 500 elemű mintarealizációt, ahol ξ eloszlása

(1) 23 várható értékű és 2 szórású normális;

(2) 5 szabadsági fokú khi-négyzet;

(3) 3 szabadsági fokú t;

(4) 2 és 3 szabadsági fokú F.

Ábrázolja a tapasztalati eloszlásfüggvényt azon az intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, majd a valódi eloszlásfüggvényt ugyanezen az intervallumon.

Útmutatás. (1) m várható értékű és σ szórású normális eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \in \mathbb{R}$ helyen $\boxed{\texttt{NORM.ELOSZLÁS(}x\texttt{;}m\texttt{;}\sigma\texttt{;IGAZ)}}$. Itt jegyezzük meg, hogy ha speciálisan $m = 0$ és $\sigma = 1$, azaz standard normális az eloszlás, akkor $\boxed{\texttt{NORM.ELOSZLÁS(}x\texttt{;0;1;IGAZ)}}$ helyett használható a következő is: $\boxed{\texttt{NORM.S.ELOSZLÁS(}x\texttt{;IGAZ)}}$.

(2) Az s szabadsági fokú khi-négyzet eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \ge 0$ helyen $\boxed{\texttt{KHINÉGYZET.ELOSZLÁS(}x\texttt{;}s\texttt{;IGAZ)}}$.

(3) Az s szabadsági fokú t-eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \in \mathbb{R}$ helyen $\boxed{\texttt{T.ELOSZL(}x\texttt{;}s\texttt{;IGAZ)}}$.

(4) Az s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \ge 0$ helyen $\boxed{\texttt{F.ELOSZL(}x\texttt{;}s_1\texttt{;}s_2\texttt{;IGAZ)}}$.

2.7. gyakorlat. Generáljon $\lambda = 4$ paraméterű exponenciális eloszlású valószínűségi változóra vonatkozó 500 elemű mintarealizációt. Rajzolja meg a sűrűséghisztogramot azon az intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, 10 darab egyenlő hosszúságú részintervallum esetén. Ugyanezen az intervallumon ábrázolja a valódi sűrűségfüggvényt is.

Útmutatás. A λ paraméterű exponenciális eloszlású valószínűségi változó sűrűségfüggvényének értéke az $x \ge 0$ helyen $\boxed{\texttt{EXP.ELOSZL(}x\texttt{;}\lambda\texttt{;HAMIS)}}$.

2.8. gyakorlat. Generáljon a ξ valószínűségi változóra vonatkozó 500 elemű mintarealizációt, ahol ξ eloszlása

(1) $[-5, 4]$ intervallumon egyenletes;

(2) $r = 2$ rendű $\lambda = 1$ paraméterű gamma;

(3) standard Cauchy;

(4) $s = 6$ szabadsági fokú khi-négyzet.

Ábrázolja a tapasztalati eloszlásfüggvényt, illetve a sűrűséghisztogramot 10 darab egyenlő hosszúságú részintervallum esetén, azon az intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, majd a valódi eloszlásfüggvényt illetve a sűrűségfüggvényt ugyanezen az intervallumon.

Útmutatás. (2) Az r-edrendű λ paraméterű gamma-eloszlású valószínűségi változó eloszlásfüggvénye $x \geq 0$ esetén $\boxed{\texttt{GAMMA.ELOSZL(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;IGAZ)}}$ illetve a sűrűségfüggvénye $\boxed{\texttt{GAMMA.ELOSZL(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;HAMIS)}}$ függvényekkel számolható ki.

(3) A standard Cauchy-eloszlású valószínűségi változó sűrűségfüggvénye

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \frac{1}{\pi(1 + x^2)},$$

illetve eloszlásfüggvénye

$$F : \mathbb{R} \to \mathbb{R}, \quad F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x.$$

Az $\operatorname{arctg}(x)$ az $\boxed{\texttt{ARCTAN(}x\texttt{)}}$ függvénnyel számolható. De azt is felhasználhatjuk, hogy a standard Cauchy-eloszlás megegyezik az 1 szabadsági fokú t-eloszlással.

(4) Az s szabadsági fokú khi-négyzet eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \geq 0$ helyen $\boxed{\texttt{KHINÉGYZET.ELOSZLÁS(}x\texttt{;}s\texttt{;IGAZ)}}$, a sűrűségfüggvénye pedig $\boxed{\texttt{KHINÉGYZET.ELOSZLÁS(}x\texttt{;}s\texttt{;HAMIS)}}$.

2.9. gyakorlat. Generáljon a ξ valószínűségi változóra vonatkozó 500 elemű mintarealizációt, ahol ξ eloszlása

(1) 3 szabadsági fokú t;

(2) 2 és 3 szabadsági fokú F.

Ábrázolja a sűrűséghisztogramot 10 darab egyenlő hosszúságú részintervallum esetén, azon az intervallumon, amelyen a mintarealizáció elemei elhelyezkednek, majd a valódi sűrűségfüggvényt ugyanezen az intervallumon.

Útmutatás. Az s szabadsági fokú t-eloszlású valószínűségi változó sűrűségfüggvényének az értéke $x \in \mathbb{R}$ helyen $\boxed{\texttt{T.ELOSZL(}x\texttt{;}s\texttt{;HAMIS)}}$. illetve az s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \geq 0$ helyen $\boxed{\texttt{F.ELOSZL(}x\texttt{;}s_1\texttt{;}s_2\texttt{;HAMIS)}}$.

3. fejezet

Grafikus illeszkedésvizsgálat

Tegyük fel, hogy a vizsgált valószínűségi változó abszolút folytonos. Ha a tapasztalati eloszlásfüggvény vagy a sűrűséghisztogram segítségével megsejtjük, hogy mi a vizsgált valószínűségi változó eloszlása, akkor bizonyos esetekben az úgynevezett grafikus illeszkedésvizsgálattal győződhetünk meg sejtésünk igazáról.

3.1. Paraméter nélküli eloszlásra vonatkozó illeszkedésvizsgálat

Legyen $x_1 < x_2 < \cdots < x_r$, továbbá tegyük fel, hogy a mintarealizáció legkisebb eleme kisebb x_1-nél, a mintarealizáció legnagyobb eleme pedig nagyobb x_r-nél.

A matematikai statisztika alaptétele szerint a tapasztalati eloszlásfüggvény nagy elemszámú mintarealizáció esetén jól közelíti a valódi eloszlásfüggvényt, azaz ha n a mintarealizáció elemszáma, akkor

$$F_n^*(x_i) \simeq F(x_i), \quad i = 1, \ldots, r,$$

ahol F a vizsgált valószínűségi változó valódi eloszlásfüggvénye. Ebből F-nek az $[x_1, x_r]$ intervallumon vett invertálhatóságát feltételezve azt kapjuk, hogy

$$F^{-1}\Big(F_n^*(x_i)\Big) \simeq x_i, \quad i = 1, \ldots, r,$$

azaz $y_i := F^{-1}\Big(F_n^*(x_i)\Big)$ jelöléssel az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy origón átmenő 1 meredekségű egyenesre esnek.

3.1. Példa. A minta-04.txt fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e standard Cauchy-eloszlású.

Megoldás. Ha a vizsgált valószínűségi változó standard Cauchy-eloszlású, akkor

$$F_n^*(x_i) \simeq \frac{1}{\pi} \operatorname{arctg} x_i + \frac{1}{2}, \quad i = 1, \ldots, r,$$

azaz

$$y_i := \operatorname{tg}\left(F_n^*(x_i) - \frac{1}{2}\right)\pi \simeq x_i, \quad i = 1, \ldots, r.$$

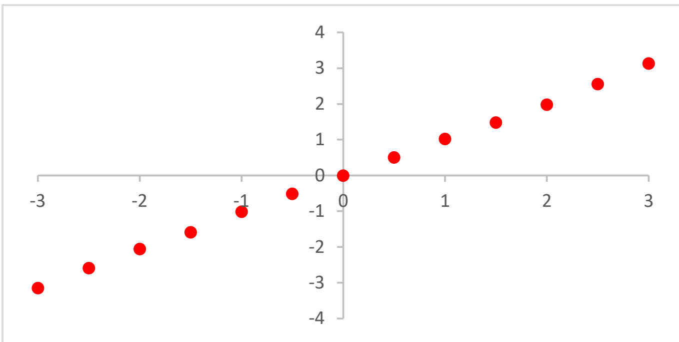
Így az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy origón átmenő 1 meredekségű egyenesre esnek.

A mintarealizációt másolja egy Excel-munkalap `A` oszlopába. Használjuk a -3-tól 3-ig terjedő egyenletes beosztást 0,5 hosszúságú részintervallumokkal, azaz legyen $x_1 = -3, x_2 = -2{,}5, x_3 = -2, \ldots, x_{13} = 3$. Ezeket rendre írja be a `B1:B13` cellatartományba. A `C` oszlopban számoljuk ki az y_i értékeket. Ehhez írja `C1`-be, hogy

```
=TAN((DARABTELI(A:A;"<"&B1)/DARAB(A:A)-1/2)*PI())
```

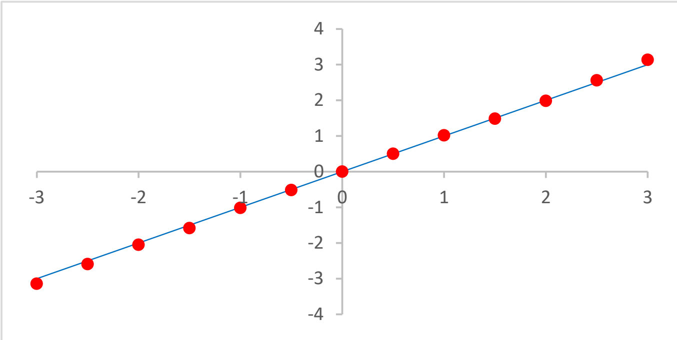
majd a `C1` kitöltőjelét húzza le a 13. sorig. Ábrázoljuk a kapott pontokat. Ehhez jelölje ki a `B1:C13` cellatartományt, majd

Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont*



Az origón átmenő 1 meredekségű egyenessel való összehasonlítás miatt húzzuk meg a diagramon ezt az egyenest. Ehhez a `D1` és `E1` cellákba írja a `-3` értéket, illetve a `D2` és `C2` cellákba a `3` értéket, kattintson a diagramra, majd

Tervezés → *Adatok kijelölése* → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$D$1:$D$2` →
Adatsor Y értékei: `=Munka1!$E$1:$E$2` → *OK* → *OK* →
Formátum → bal oldali legördülő listában: *Adatsor2* →*Kijelölés formázása* →
Adatsor formázása panelen:
Kitöltés és vonal ikon → *Vonal* → *Folytonos vonal* →
Jelölő → *Jelölő beállításai* → *Nincs*



A kapott pontok nagyon jól illeszkednek az egyenesre, így nagy valószínűséggel állítható, hogy a vizsgált valószínűségi változó standard Cauchy-eloszlású.

A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

3.2. Egyparaméteres eloszlásra vonatkozó illeszkedésvizsgálat

Ebben az esetben azt vizsgáljuk, hogy a valószínűségi változó származhat-e olyan eloszlástípusból, amelynek egy ismeretlen paramétere van. Például egy valószínűségi változó lehet-e exponenciális eloszlású, amelynek a λ paramétere ismeretlen?

Általánosságban jelölje ϑ az ismeretlen paramétert. Tegyük fel, hogy G és g olyan függvények, melyekre az eloszlástípus eloszlásfüggvénye előáll az x helyen $G\big(g(\vartheta)x\big)$ alakban, továbbá G invertálható. Az exponenciális eloszlás ilyen, hiszen $G(t) = 1 - e^t$, $g(\vartheta) = -\vartheta$ és $\vartheta = \lambda$ választással $G\big(g(\vartheta)x\big) = 1 - e^{-\lambda x}$, továbbá G invertálható és $G^{-1}(t) = \ln(1-t)$.

Legyen $x_1 < x_2 < \cdots < x_r$, továbbá tegyük fel, hogy a mintarealizáció legkisebb eleme kisebb x_1-nél, a mintarealizáció legnagyobb eleme pedig nagyobb x_r-nél. Ha n a mintarealizáció elemszáma és a valószínűségi változó valóban a sejtett eloszlástípusba tartozik, akkor

$$F_n^*(x_i) \simeq G\big(g(\vartheta)x_i\big), \quad i = 1, \ldots, r,$$

Ebből

$$G^{-1}\big(F_n^*(x_i)\big) \simeq g(\vartheta)x_i, \quad i = 1, \ldots, r,$$

azaz $y_i := G^{-1}\big(F_n^*(x_i)\big)$ jelöléssel az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy origón átmenő $g(\vartheta)$ meredekségű egyenesre esnek.

A *legkisebb négyzetek elve* alapján az origón átmenő $y = mx$ egyenletű egyenesek közül az illeszkedik legjobban az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontokra, melyre $\sum_{i=1}^{r}(mx_i - y_i)^2$ értéke a legkisebb. Ezt az egyenest az adott pontokhoz tartozó *fixpontos regressziós egyenesnek* (vagy *lineáris trendvonalnak*) nevezzük. Jelölje ennek meredekségét m^*. Ekkor megoldva a

$$g(\vartheta) = m^*$$

egyenletet, a $\hat{\vartheta}$ megoldása a ϑ becslése. Tehát ez a módszer nem csak arra ad feleletet, hogy tényleg ebbe az eloszlástípusba tartozik-e a vizsgált valószínűségi változó, hanem az ismeretlen paraméter értékét is megbecsülhetjük vele.

3.2. Példa. A `minta-03.txt` fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e exponenciális eloszlású. Ha igen, akkor lineáris regresszióval becsülje meg a paramétert.

Megoldás. Ha teljesül, hogy a vizsgált valószínűségi változó exponenciális eloszlású λ paraméterrel, akkor

$$F_n^*(x_i) \simeq 1 - e^{-\lambda x_i}, \quad i = 1, \ldots, r,$$

azaz

$$y_i := \ln\Big(1 - F_n^*(x_i)\Big) \simeq -\lambda x_i, \quad i = 1, \ldots, r.$$

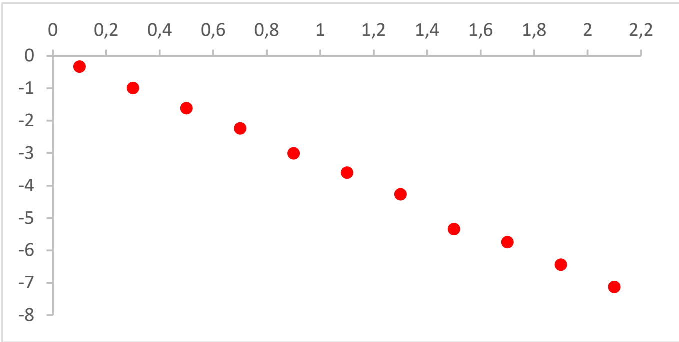
Így az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $-\lambda$ a meredeksége és átmegy az origón. Ha m^* az origón átmenő regressziós egyenes meredeksége, akkor $-\lambda = m^*$ egyenletet megoldva kapjuk λ becslését, azaz $\hat{\lambda} = -m^*$.

A mintarealizációt másolja egy Excel-munkalap `A` oszlopába. Vizsgálja meg a legkisebb és legnagyobb értékét a mintarealizációnak a `=MIN(A:A)` és `=MAX(A:A)` függvényekkel. Azt kapjuk, hogy 0,002 a legkisebb és 2,1932 a legnagyobb érték. Ennek alapján lehet $x_1 = 0{,}1$ és $x_r = 2{,}1$. Célszerű az $r = 11$ választás, mert ekkor egyenletes beosztást választva 10 egyenlő hosszúságú részintervallumot kapunk, melyek hossza $\frac{2{,}1-0{,}1}{10} = 0{,}2$. Így $x_1 = 0{,}1$; $x_2 = 0{,}3$; $x_3 = 0{,}5$; $x_4 = 0{,}7$; $x_5 = 0{,}9$; $x_6 = 1{,}1$; $x_7 = 1{,}3$; $x_8 = 1{,}5$; $x_9 = 1{,}7$; $x_{10} = 1{,}9$; $x_{11} = 2{,}1$. Ezeket írja be a `B1:B11` cellatartományba. Ezután a `C1` cellában számolja ki az y_1 értékét. Ehhez írja be a következőt:

```
=LN(1-DARABTELI(A:A;"<"&B1)/DARAB(A:A))
```

A `C1` cella kitöltőjelét húzza le a 11. sorig. Ábrázoljuk a kapott pontokat. Ehhez jelölje ki a `B1:C11` cellatartományt, majd

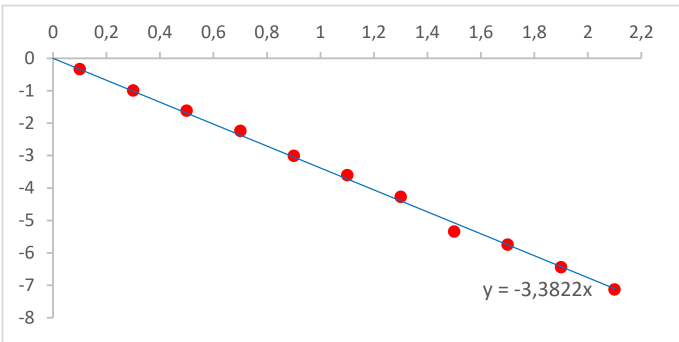
Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont*



A kapott pontok jól illeszkednek egy olyan egyenesre, amely átmegy az origón. Így nagy biztonsággal állíthatjuk, hogy a mintarealizáció exponenciális eloszlásból származik.

A λ paraméter becsléséhez ábrázoljuk az origón átmenő lineáris trendvonalat. Kattintson az ábrára, majd

Tervezés → *Diagram-összetevő hozzáadása* →
Trendvonal → *További trendvonal-beállítások* →
Trendvonal formázása panelen:
Trendvonal beállításai ikon → *Lineáris* → *Metszéspont:* 0 →
Egyenlet látszik a diagramon



A meredekségből látható, hogy λ becslése 3,3822. Összehasonlításképpen közöljük, hogy a vizsgált valószínűségi változó exponenciális eloszlású volt $\lambda = 3{,}2$ paraméterrel.

A színt és a vonalvastagságot igény szerint beállíthatja. A megoldás menetét végigkövetheti a következő videón.

3.3. Kétparaméteres eloszlásra vonatkozó illeszkedésvizsgálat

Most azt vizsgáljuk, hogy a valószínűségi változó származhat-e olyan eloszlástípusból, amelynek két ismeretlen paramétere van. Például egy valószínűségi változó lehet-e normális eloszlású, amelynek m és σ paraméterei ismeretlenek?

Általánosságban jelölje ϑ_1 és ϑ_2 az ismeretlen paramétereket. Tegyük fel, hogy G, g_1 és g_2 olyan függvények, melyekre az eloszlástípus eloszlásfüggvénye előáll az x helyen $G\Big(g_1(\vartheta_1, \vartheta_2)x + g_2(\vartheta_1, \vartheta_2)\Big)$ alakban, továbbá G invertálható. A normális eloszlás ilyen, hiszen $G(t) = \Phi(t)$, $g_1(\vartheta_1, \vartheta_2) = \frac{1}{\vartheta_2}$, $g_2(\vartheta_1, \vartheta_2) = -\frac{\vartheta_1}{\vartheta_2}$, $\vartheta_1 = m$ és $\vartheta_2 = \sigma$ választással $G\Big(g_1(\vartheta_1, \vartheta_2)x + g_2(\vartheta_1, \vartheta_2)\Big) = \Phi\left(\frac{x-m}{\sigma}\right)$, továbbá G invertálható.

Legyen $x_1 < x_2 < \cdots < x_r$, továbbá tegyük fel, hogy a mintarealizáció legkisebb eleme kisebb x_1-nél, a mintarealizáció legnagyobb eleme pedig nagyobb x_r-nél. Ha n a mintarealizáció elemszáma és a valószínűségi változó valóban a sejtett eloszlástípusba tartozik, akkor

$$F_n^*(x_i) \simeq G\Big(g_1(\vartheta_1, \vartheta_2)x_i + g_2(\vartheta_1, \vartheta_2)\Big), \quad i = 1, \ldots, r,$$

Ebből

$$G^{-1}\Big(F_n^*(x_i)\Big) \simeq g_1(\vartheta_1, \vartheta_2)x_i + g_2(\vartheta_1, \vartheta_2), \quad i = 1, \ldots, r,$$

azaz $y_i := G^{-1}\Big(F_n^*(x_i)\Big)$ jelöléssel az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek meredeksége $g_1(\vartheta_1, \vartheta_2)$ és a függőleges tengelymetszete $g_2(\vartheta_1, \vartheta_2)$.

A *legkisebb négyzetek elve* alapján az origón átmenő $y = mx + b$ egyenletű egyenesek közül az illeszkedik legjobban az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontokra, melyre $\sum\limits_{i=1}^{r}(mx_i + b - y_i)^2$ értéke a legkisebb. Ezt az egyenest az adott pontokhoz tartozó *regressziós egyenesnek* (vagy *lineáris trendvonalnak*) nevezzük. Jelölje ennek meredekségét m^* és a függőleges tengelymetszetét b^*. Ekkor megoldva a

$$\begin{aligned} g_1(\vartheta_1, \vartheta_2) &= m^* \\ g_2(\vartheta_1, \vartheta_2) &= b^* \end{aligned}$$

egyenletrendszert, a $(\hat{\vartheta}_1, \hat{\vartheta}_2)$ megoldása a $(\vartheta_1, \vartheta_2)$ becslése. Tehát ez a módszer nem csak arra ad feleletet, hogy tényleg ebbe az eloszlástípusba tartozik-e a vizsgált valószínűségi változó, hanem az ismeretlen paraméterek értékeit is megbecsülhetjük vele.

3.3. Példa. A `minta-01.txt` fájlban található mintarealizáció alapján nézze meg, hogy a vizsgált valószínűségi változó lehet-e normális eloszlású. Ha igen, akkor becsülje meg lineáris regresszióval a paramétereket.

Megoldás. Ha a vizsgált valószínűségi változó normális eloszlású m várható értékkel és σ szórással, akkor

$$F_n^*(x_i) \simeq \Phi\left(\frac{x_i - m}{\sigma}\right), \quad i = 1, \dots, r,$$

azaz

$$y_i := \Phi^{-1}\Big(F_n^*(x_i)\Big) \simeq \frac{1}{\sigma}x_i - \frac{m}{\sigma}, \quad i = 1, \dots, r.$$

Így az $(x_1, y_1), \dots, (x_r, y_r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $\frac{1}{\sigma}$ a meredeksége és $-\frac{m}{\sigma}$ értéknél metszi a függőleges tengelyt. Ha m^* illetve b^* a regressziós egyenes meredeksége illetve függőleges tengelymetszete, akkor az

$$\begin{aligned} \frac{1}{\sigma} &= m^* \\ -\frac{m}{\sigma} &= b^* \end{aligned}$$

egyenletrendszer megoldása $\hat{m} = -\frac{b^*}{m^*}$ és $\hat{\sigma} = \frac{1}{m^*}$, melyek az m és σ becslései.

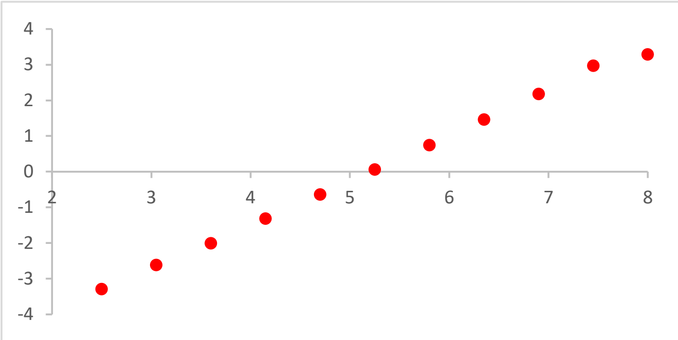
A mintarealizációt másolja egy Excel-munkalap `A` oszlopába. Vizsgálja meg a legkisebb és legnagyobb értékét a mintarealizációnak a `=MIN(A:A)` és `=MAX(A:A)` függvényekkel. Azt kapjuk, hogy 2,495 a legkisebb és 8,0063 a legnagyobb érték. Ennek alapján lehet $x_1 = 2{,}5$ és $x_r = 8$. Célszerű az $r = 11$ választás, mert ekkor egyenletes beosztást választva 10 egyenlő hosszúságú részintervallumot kapunk, melyek hossza $\frac{8-2{,}5}{10} = 0{,}55$. Így $x_1 = 2{,}5; x_2 = 3{,}05; \dots ; x_{11} = 8$. Ezeket az értékeket írja be a `B1:B11` cellatartományba. Ezután a `C1` cellában számolja ki az y_1 értékeket a

```
=NORM.S.INVERZ(DARABTELI(A:A;"<"&B1)/DARAB(A:A))
```

függvénnyel, majd a kitöltőjelét húzza le a 11. sorig. Ábrázoljuk a kapott pontokat. Ehhez jelölje ki a `B1:C11` cellatartományt, majd

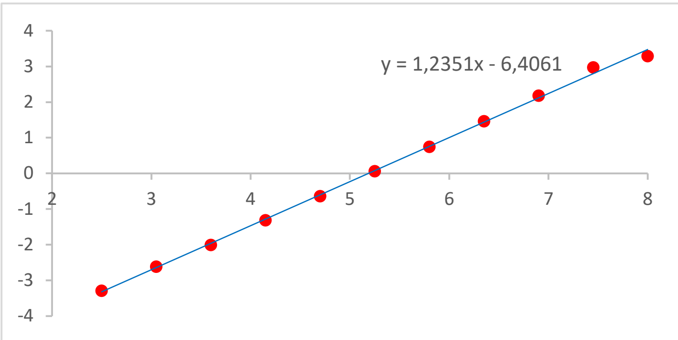
Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont*

A 11 darab pont nagyon jó közelítéssel egy egyenesen helyezkedik el, így normális eloszlásúnak tekinthetjük a vizsgált valószínűségi változót.



A paraméterek becsléséhez ábrázoljuk a lineáris trendvonalat. Kattintson az ábrára, majd

Tervezés → *Diagram-összetevő hozzáadása* →
Trendvonal → *További trendvonal-beállítások* →
Trendvonal formázása panelen:
Trendvonal beállításai ikon → *Lineáris* → *Egyenlet látszik a diagramon*



A lineáris trendvonal meredekségét a `MEREDEKSÉG` függvénnyel, illetve a függőleges tengelymetszet értékét a `METSZ` függvénnyel számolhatjuk ki. Ennek alapján m becslése

```
=-METSZ(C1:C11;B1:B11)/MEREDEKSÉG(C1:C11;B1:B11)
```

és σ becslése

```
=1/MEREDEKSÉG(C1:C11;B1:B11)
```

Az eredmény $m \simeq 5{,}1866$ és $\sigma \simeq 0{,}8096$. Ellenőrzésképpen közöljük, hogy a felhasznált mintarealizáció $m = 5{,}2$ és $\sigma = 0{,}8$ paraméterű normális eloszlásból származik.

Az egész megoldást végigkövetheti a következő videón.

3.4. Gyakorlatok

3.1. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e egyenletes eloszlású. Ha igen, akkor a kapott ábra alapján becsülje meg a paramétereket.

Útmutatás. Ha teljesül, hogy a vizsgált valószínűségi változó egyenletes eloszlású az $[a, b]$ intervallumon, akkor

$$y_i := F_n^*(x_i) \simeq \frac{x_i - a}{b - a} = \frac{1}{b - a} x_i - \frac{a}{b - a}, \quad i = 1, \ldots, r.$$

Ebből $\hat{a} = -\frac{b^*}{m^*}$ és $\hat{b} = -\frac{1-b^*}{m^*}$. Az összehasonlításhoz közöljük, hogy a vizsgált valószínűségi változó egyenletes eloszlású volt a $[2{,}5; 8]$ intervallumon.

3.2. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e exponenciális eloszlású.

Útmutatás. A 3.2. példa alapján eljárva azt állíthatjuk, hogy a minta nem exponenciális eloszlásból származik.

3.3. gyakorlat. A `minta-04.txt` fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e normális eloszlású. Ha igen, akkor a kapott ábra alapján becsülje meg a paramétereket. Használjon -3-tól 3-ig terjedő beosztást 0,5 hosszúságú részintervallumokkal.

Útmutatás. A 3.3. példa alapján eljárva azt állíthatjuk, hogy a minta nem normális eloszlásból származik.

3.4. gyakorlat. Generáljon Excel segítségével 3000 elemű mintarealizációt egyenletes, exponenciális, normális illetve Cauchy-eloszlású valószínűségi változóra vonatkozóan a korábban ismertetett módszerekkel. Grafikus illeszkedésvizsgálattal igazolja, hogy az így generált mintarealizációk valóban olyan eloszlásúak, mint aminek az elmélet szerint kell lennie. A paramétereket is becsüljük meg.

4. fejezet

Statisztikák

Tegyük fel, hogy egy ismeretlen eloszlású ξ valószínűségi változó várható értékét kell meghatározni. Mivel az eloszlást nem ismerjük, ezért a mintarealizáció alapján kell becslést adni. Bizonyítható, hogy a ξ-re vonatkozó $\xi_1, \ldots, \xi_n$ minta elemeinek a számtani közepe, azaz $\frac{1}{n}(\xi_1 + \cdots + \xi_n)$, jó becslése a várható értéknek. Általánosan fogalmazva itt egy olyan függvényt definiáltunk, amely valószínűségi változókból álló rendezett n-eshez egy valószínűségi változót rendel. Az ilyen függvényeket *statisztikának* nevezzük. A következőkben felsorolunk néhány nevezetes statisztikát, ahol a ξ-re vonatkozó minta $\xi_1, \ldots, \xi_n$.

mintaátlag	$\overline{\xi} := \dfrac{1}{n}\sum_{i=1}^{n}\xi_i$
tapasztalati szórásnégyzet	$S_n^2 := \dfrac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2$
tapasztalati szórás	$S_n := \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2}$
korrigált tapasztalati szórásnégyzet	$S_n^{*2} := \dfrac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2$
korrigált tapasztalati szórás	$S_n^{*} := \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2}$
k-adik tapasztalati momentum ($k \in \mathbb{N}$)	$\dfrac{1}{n}\sum_{i=1}^{n}\xi_i^k$
k-adik tapasztalati centrált momentum ($k \in \mathbb{N}$)	$\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^k$
tapasztalati ferdeség	$\dfrac{1}{nS_n^3}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^3$

tapasztalati lapultság $$\frac{1}{nS_n^4}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^4-3$$

Legyen ω egy rögzített elemi esemény. A $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)$ számokat állítsuk növekvő sorrendbe. Az így kapott n tagú sorozatban az i-edik tagot jelölje $\xi_i^*(\omega)$. Ekkor a $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ valószínűségi változókat *rendezett mintának* nevezzük.

A $\xi_n^* - \xi_1^*$ statisztikát *mintaterjedelemnek* nevezzük. A $\frac{\xi_1^*+\xi_n^*}{2}$ az úgynevezett *terjedelemközép*.

A *tapasztalati medián* legyen $\xi^*_{\frac{n+1}{2}}$, ha n páratlan, illetve $\frac{1}{2}\left(\xi^*_{\frac{n}{2}}+\xi^*_{\frac{n}{2}+1}\right)$, ha n páros.

A $100t$*%-os tapasztalati kvantilis* ($0 \le t \le 1$) legyen $\xi^*_{[nt]+1}$, ha $nt \notin \mathbb{N}$, illetve $t\xi^*_{nt} + (1-t)\xi^*_{nt+1}$, ha $nt \in \mathbb{N}$. (Vegyük észre, hogy az 50%-os tapasztalati kvantilis a tapasztalati mediánnal egyenlő.) A 25%-os tapasztalati kvantilist *tapasztalati alsó kvartilisnek*, illetve a 75%-os tapasztalati kvantilist *tapasztalati felső kvartilisnek* nevezzük.

A *tapasztalati módusz* a mintaelemek között a leggyakrabban előforduló. Ha több ilyen is van, akkor azok között a legkisebb.

Legyen a ξ-re vonatkozó minta $\xi_1, \dots, \xi_n$ és az η-ra vonatkozó minta $\eta_1, \dots, \eta_n$. Ezek *tapasztalati kovarianciája*

$$\mathrm{Cov}_n(\xi,\eta) := \frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_i - \frac{1}{n}\sum_{i=1}^{n}\xi_i \cdot \frac{1}{n}\sum_{i=1}^{n}\eta_i,$$

illetve *tapasztalati korrelációs együtthatója*

$$\mathrm{Corr}_n(\xi,\eta) := \frac{\mathrm{Cov}_n(\xi,\eta)}{S_{\xi,n}\cdot S_{\eta,n}}.$$

4.1. Példa. A `minta-02.txt` fájlban található mintarealizáció esetén adja meg a harmadik tapasztalati momentumot, harmadik tapasztalati centrált momentumot, továbbá a rendezett mintát.

Megoldás. A mintarealizációt másolja az `A` oszlopba. A `B1` cellába írja a következőt: `=A1^3`. A `B1` cella kitöltőjelére kattintson kétszer.

A `C1` cellába írja a következőt: `=(A1-ÁTLAG(A:A))^3`. A `C1` cella kitöltőjelére kattintson kétszer.

Ezután a `D1` cellába írja a következőt: `=ÁTLAG(B:B)`. A kapott érték négy tizedesjegyre kerekítve 186,4503, mely a harmadik tapasztalati momentum.

Az `E1` cellába írja a következőt: `=ÁTLAG(C:C)`. A kapott érték négy tizedesjegyre kerekítve 0,0787, mely a harmadik tapasztalati centrált momentum.

A rendezett minta megadásához az `A` oszlopot másolja át az `F` oszlopba, kattintson az `F` oszlop jelölőjére, majd

Kezdőlap → *Rendezés és szűrés* → *Rendezés méret szerint (növekvő)* → *Folytatja az aktuális kijelöléssel* → *Rendezés*

Ezután a rendezett mintát az `F` oszlop tartalmazza.

4.2. Példa. A `minta-02.txt` fájlban található mintarealizáció esetén számolja ki a következő statisztikákat: minta elemszáma; mintaterjedelem; terjedelemközép; mintaátlag; tapasztalati szórás; tapasztalati szórásnégyzet; korrigált tapasztalati szórás; korrigált tapasztalati szórásnégyzet; tapasztalati medián; tapasztalati módusz.

Megoldás. A mintarealizációt másolja az `A` oszlopba. Az előző statisztikákat a következő módon számolhatja ki:

minta elemszáma	`=DARAB(A:A)`
mintaterjedelem	`=MAX(A:A)-MIN(A:A)`
terjedelemközép	`=(MIN(A:A)+MAX(A:A))/2`
mintaátlag	`=ÁTLAG(A:A)`
tapasztalati szórás	`=SZÓR.S(A:A)`
tapasztalati szórásnégyzet	`=VAR.S(A:A)`
korrigált tapasztalati szórás	`=SZÓR.M(A:A)`
korrigált tapasztalati szórásnégyzet	`=VAR.M(A:A)`
tapasztalati medián	`=MEDIÁN(A:A)`
tapasztalati módusz	`=MÓDUSZ.EGY(A:A)`

4.3. Példa. Tekintsük a ξ-re vonatkozó 0,12; 0,63; 0,44; 0,50; 0,66; 0,81; 0,91; 0,96; 0,41; 0,67 mintarealizációt és az η-ra vonatkozó 1,03; 0,13; 1,50; 1,21; 1,73; 1,13; 1,73; 0,65; 1,10; 0,01 mintarealizációt. Számolja ki a tapasztalati kovarianciát és korrelációs együtthatót.

Megoldás. Az első mintarealizációt tegye az `A` oszlopba, a másodikat pedig a `B` oszlopba. A tapasztalati kovarianciát a `=KOVARIANCIA.S(A:A;B:B)`, míg a tapasztalati korrelációs együtthatót a `=KORREL(A:A;B:B)` függvénnyel számolhatja ki.

4.1. Gyakorlatok

4.1. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció esetén adja meg a

$$\sum_{i=1}^{n} x_i,\ \sum_{i=1}^{n} x_i^2,\ \sum_{i=1}^{n} (x_i - \overline{x})^2,\ \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|,\ \prod_{i=1}^{n} x_i,$$

értékeket, ahol $x_1, \ldots, x_n$ a mintarealizáció elemei és $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Útmutatás. Ha a mintarealizáció az `A` oszlopban van, akkor használja rendre a következő függvényeket:

`=SZUM(A:A)`

`=NÉGYZETÖSSZEG(A:A)`

`=SQ(A:A)`

`=ÁTL.ELTÉRÉS(A:A)`

`=SZORZAT(A:A)`.

4.2. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció első 100 elemének számolja ki a mértani illetve harmonikus közepét.

Útmutatás. Használjuk a `MÉRTANI.KÖZÉP` és `HARM.KÖZÉP` függvényeket.

4.3. gyakorlat. Legyenek az $x_1, \ldots, x_{20}$ számok a `minta-01.txt` első 20 eleme, továbbá az $y_1, \ldots, y_{20}$ számok a `minta-02.txt` első 20 eleme. Számolja ki a következő értékeket:

$$\sum_{i=1}^{20} x_i y_i, \ \sum_{i=1}^{20} (x_i - y_i)^2, \ \sum_{i=1}^{20} (x_i^2 - y_i^2), \ \sum_{i=1}^{20} (x_i^2 + y_i^2).$$

Útmutatás. Használja rendre a következő függvényeket:

`SZORZATÖSSZEG`, `SZUMXBŐLY2`, `SZUMX2BŐLY2`, `SZUMX2MEGY2`.

4.4. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció esetén adja meg

(1) a 3-nál kisebb elemek összegét;

(2) a 3-nál nagyobb de 4-nél kisebb vagy egyenlő elemek összegét;

(3) a 3-nál kisebb elemek számát;

(4) a 3-nál nagyobb de 4-nél kisebb vagy egyenlő elemek számát;

(5) a 3-nál kisebb elemek átlagát;

(6) a 3-nál nagyobb de 4-nél kisebb vagy egyenlő elemek átlagát.

Útmutatás. Ha a mintarealizáció az `A` oszlopban van, akkor használja rendre a következő függvényeket:

`=SZUMHA(A:A;"<3")`

`=SZUMHATÖBB(A:A;A:A;">3";A:A;"<=4")`

`=DARABTELI(A:A;"<3")`

`=DARABHATÖBB(A:A;">3";A:A;"<=4")`

`=ÁTLAGHA(A:A;"<3")`

`=ÁTLAGHATÖBB(A:A;A:A;">3";A:A;"<=4")`.

4.5. gyakorlat. Adja meg az $\mathrm{I}_{3<\xi\leq 4}$ indikátorváltozóra vonatkozó mintarealizációt, ha a `minta-02.txt` fájlban található a ξ-re vonatkozó mintarealizáció.

Útmutatás. A mintarealizációt másolja az `A` oszlopba. A `B1` cellába írja a következőt: `=HA(ÉS(A1>3;A1<=4);1;0)`. Ezután a `B1` cella kitöltőjelére kattintson kétszer.

4.6. gyakorlat. Adja meg az x_{10}^* és x_{n-5}^* értékeit, ahol a ξ-re vonatkozó mintarealizáció a `minta-02.txt` fájlban található, továbbá $x_1^*,\dots,x_n^*$ a rendezett mintarealizációt jelöli. A rendezett mintarealizáció hányadik eleme a mintarealizáció 5. eleme? A mintarealizációnak hányadik legnagyobb eleme 7,963?

Útmutatás. A mintarealizációt másolja az `A` oszlopba. Ekkor
$x_k^* = \boxed{\texttt{KICSI(A:A;}k\texttt{)}}$
$x_{n-k}^* = \boxed{\texttt{NAGY(A:A;}k+1\texttt{)}}$
$\min\{k : x_k^* = x_i\} = \boxed{\texttt{RANG.EGY(}x_i\texttt{;A:A;1)}}$
$\min\{k : x_{n-k}^* = x_i\} + 1 = \boxed{\texttt{RANG.EGY(}x_i\texttt{;A:A;0)}}$.

4.7. gyakorlat. A `minta-02.txt` fájlban található mintarealizáció esetén adja meg a 30%-os tapasztalati kvantilist, továbbá a tapasztalati alsó illetve felső kvartilist.

Útmutatás. Ha a mintarealizáció az `A` oszlopban van, akkor a $100t$%-os tapasztalati kvantilis, a tapasztalati alsó illetve felső kvartilis rendre a következő módon számolható ki:
$\boxed{\texttt{PERCENTILIS.TARTALMAZ(A:A;}t\texttt{)}}$
`KVARTILIS.TARTALMAZ(A:A;1)`
`KVARTILIS.TARTALMAZ(A:A;3)`.

4.8. gyakorlat. A `minta-01.txt` illetve `minta-04.txt` fájlban található mintarealizációk esetén adja meg a tapasztalati ferdeséget és tapasztalati lapultságot. Ennek alapján melyik minta származhat normális eloszlásból? Az eredményt vesse össze a grafikus illeszkedésvizsgálatnál tapasztaltakkal.

Útmutatás. Az eloszlás ferdeségének illetve lapultságának természetes becsléseként definiáltuk a tapasztalati ferdeséget illetve lapultságot:

$$S_n^{-3}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^3 \quad \text{illetve} \quad S_n^{-4}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^4-3.$$

Ezeket a harmadik tapasztalati centrált momentum kiszámolásához hasonlóan határozhatjuk meg. A kapott értékek 5 tizedesjegyre kerekítve `minta-01` esetén $-0{,}03636$ illetve 0,00135 továbbá `minta-04` esetén 54,17325 illetve 2952,01016. Mivel ezek az értékek `minta-01` esetén 0-hoz közeliek, míg `minta-04` esetén távoliak, ezért az előbbi minta származhat normális eloszlásból, de az utóbbi nem. Ez a grafikus illeszkedésvizsgálatnál tapasztaltakkal is összhangban van.

A tapasztalati ferdeség kiszámolására az Excelben használhatjuk a `FERDESÉG.P` függvényt is. Van egy `FERDESÉG` függvény is, de ez más becslést használ az eloszlás ferdeségére:

$$\frac{n\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})^3}{(n-1)(n-2)S_n^{*3}}.$$

Excelben a lapultságot csúcsosságnak nevezik, pontosabban a tapasztalati lapultságot a `CSÚCSOSSÁG` függvénnyel számolhatjuk, de ez is más becslést használ az eloszlás lapultságára a korábban ismertetetthez képest:

$$\frac{n(n+1)\sum\limits_{i=1}^{n}(\xi_i-\overline{\xi})^4}{(n-1)(n-2)(n-3)S_n^{*4}}-\frac{3(n-1)^2}{(n-2)(n-3)}.$$

Ezek a statisztikák nagy n esetén körülbelül megegyeznek az előbbi statisztikákkal. A `FERDESÉG` és `CSÚCSOSSÁG` értékei 5 tizedesjegyre kerekítve `minta-01` esetén $-0{,}03638$ illetve 0,00436 továbbá `minta-04` esetén 54,20035 illetve 2956,93809.

5. fejezet

Intervallumbecslések

Legyen ξ a vizsgált valószínűségi változó és ξ eloszlásának legyen ϑ egy ismeretlen paramétere. Intervallumbecslésnél egy olyan intervallumot adunk meg, amelybe a ϑ valódi értéke nagy valószínűséggel beleesik. Ezen intervallum alsó és felső végpontját egy-egy statisztikával adjuk meg. A becslő intervallumot *konfidenciaintervallumnak* nevezzük. Ennek az $1-\alpha$ *biztonsági szintje* az az érték, amelytől nagyobb vagy egyenlő valószínűséggel esik ϑ a konfidenciaintervallumba.

5.1. Normális eloszlás paramétereinek becslése

5.1. Példa. A `minta-06.txt` fájlban található mintarealizáció normális eloszlásból származik, melynek szórása 0,7. Adjon a várható értékre 0,99 biztonsági szintű konfidenciaintervallumot.

Megoldás. Legyen $\xi_1, \ldots, \xi_n$ a minta. Ismert, hogy $1-\alpha = 0{,}99$, $\sigma = 0{,}7$

$$\begin{aligned}
\tau_1 &= \overline{\xi} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\\
\tau_2 &= \overline{\xi} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)
\end{aligned}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,99 biztonsági szintű konfidenciaintervallum a várható értékre.

A mintarealizációt másolja az `A` oszlopba. Mivel

$$\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = \boxed{\texttt{MEGBÍZHATÓSÁG.NORM}(\alpha;\sigma;n)},$$

ezért τ_1 értéke

```
=ÁTLAG(A:A)-MEGBÍZHATÓSÁG.NORM(0,01;0,7;DARAB(A:A))
```

módon, míg τ_2 értéke

`=ÁTLAG(A:A)+MEGBÍZHATÓSÁG.NORM(0,01;0,7;DARAB(A:A))`

módon számolható. A kapott eredmények négy tizedesjegyre kerekítve: $\tau_1 \simeq 15{,}1966$ és $\tau_2 \simeq 15{,}3655$. Ellenőrzésképpen közöljük, hogy a várható érték valódi értéke 15,3.

5.2. Példa. A `minta-07.txt` fájlban található mintarealizáció normális eloszlásból származik, melynek várható értéke 1251. Adjon a szórásra 0,9 biztonsági szintű konfidenciaintervallumot.

Megoldás. Legyen $\xi_1, \dots, \xi_n$ a minta. Ismert, hogy $1 - \alpha = 0{,}9$, $m = 1251$

$$F := F[\mathrm{Khi}(n)]$$
$$\tau_1 := \sqrt{\frac{1}{F^{-1}\left(1 - \frac{\alpha}{2}\right)} \sum_{i=1}^{n} (\xi_i - m)^2}$$
$$\tau_2 := \sqrt{\frac{1}{F^{-1}\left(\frac{\alpha}{2}\right)} \sum_{i=1}^{n} (\xi_i - m)^2}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,9 biztonsági szintű konfidenciaintervallum a szórásra.

A mintarealizációt másolja az `A` oszlopba. A `B1`-be írja be, hogy `=A1-1251`, majd a `B1` kitöltőjelére kattintson kétszer. Ezután τ_1 értéke

`=GYÖK(NÉGYZETÖSSZEG(B:B)/KHINÉGYZET.INVERZ(1-0,1/2;DARAB(A:A)))`

módon, míg τ_2 értéke

`=GYÖK(NÉGYZETÖSSZEG(B:B)/KHINÉGYZET.INVERZ(0,1/2;DARAB(A:A)))`

módon számolható. A kapott eredmények négy tizedesjegyre kerekítve: $\tau_1 \simeq 3{,}1223$ és $\tau_2 \simeq 3{,}3754$. Ellenőrzésképpen közöljük, hogy a szórás valódi értéke 3,2.

5.3. Példa. A `minta-08.txt` fájlban található mintarealizáció normális eloszlásból származik. Adjon a szórásra 0,95 biztonsági szintű konfidenciaintervallumot.

Megoldás. Legyen n a mintarealizáció elemeinek a száma. Ekkor $1 - \alpha = 0{,}95$

$$F := F[\mathrm{Khi}(n-1)]$$
$$\tau_1 := S_n \sqrt{\frac{n}{F^{-1}\left(1 - \frac{\alpha}{2}\right)}}$$
$$\tau_2 := S_n \sqrt{\frac{n}{F^{-1}\left(\frac{\alpha}{2}\right)}}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,95 biztonsági szintű konfidenciaintervallum a szórásra.

A mintarealizációt másolja az `A` oszlopba. Ezután τ_1 értéke

```
=SZÓR.S(A:A)*GYÖK(DARAB(A:A)/KHINÉGYZET.INVERZ(1-0,05/2;DARAB(A:A)-1))
```

módon, míg τ_2 értéke

```
=SZÓR.S(A:A)*GYÖK(DARAB(A:A)/KHINÉGYZET.INVERZ(0,05/2;DARAB(A:A)-1))
```

módon számolható. A kapott eredmények négy tizedesjegyre kerekítve: $\tau_1 \simeq 1{,}9081$ és $\tau_2 \simeq 2{,}8465$. Ellenőrzésképpen közöljük, hogy a valódi szórás 2,4.

5.4. Példa. A `minta-09.txt` fájlban található mintarealizáció normális eloszlásból származik. Adjon a várható értékre 0,94 biztonsági szintű konfidenciaintervallumot.

Megoldás. Ha $\xi_1, \ldots, \xi_n$ a minta, akkor $1 - \alpha = 0{,}94$

$$
\begin{aligned}
F &:= F[\mathrm{T}(n-1)] \\
\tau_1 &:= \overline{\xi} - \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right) \\
\tau_2 &:= \overline{\xi} + \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right)
\end{aligned}
$$

jelölésekkel $[\tau_1, \tau_2]$ 0,94 biztonsági szintű konfidenciaintervallum a várható értékre.

A mintarealizációt másolja az `A` oszlopba. Mivel

$$\frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right) = \boxed{\texttt{MEGBÍZHATÓSÁG.T}(\alpha;S_n^*;n)},$$

ezért τ_1 értéke

```
=ÁTLAG(A:A)-MEGBÍZHATÓSÁG.T(0,06;SZÓR.M(A:A);DARAB(A:A))
```

módon, míg τ_2 értéke

```
=ÁTLAG(A:A)+MEGBÍZHATÓSÁG.T(0,06;SZÓR.M(A:A);DARAB(A:A))
```

módon számolható. A kapott eredmények négy tizedesjegyre kerekítve: $\tau_1 \simeq 4{,}2918$ és $\tau_2 \simeq 4{,}8270$. Ellenőrzésképpen közöljük, hogy a valódi várható érték 4,6.

5.2. Valószínűség becslése

5.5. Példa. Egy ismeretlen p valószínűségű esemény $n = 55$ kísérletből $k = 39$ alkalommal következett be. Adjon p-re 0,98 biztonsági szintű konfidenciaintervallumot.

Megoldás. Legyen ξ a figyelt esemény indikátorváltozója. Ekkor $\overline{\xi}$ az esemény relatív gyakoriságát, azaz $\frac{k}{n}$-et jelenti. Az $1-\alpha = 0{,}98$

$$\tau_1 := \frac{1}{n} \max \left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} \left(1-\overline{\xi}\right)^{n-i} < \frac{\alpha}{2} \right\}$$
$$\tau_2 := \frac{1}{n} \min \left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} \left(1-\overline{\xi}\right)^{n-i} \geq 1-\frac{\alpha}{2} \right\}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,98 biztonsági szintű konfidenciaintervallum p-re. Mivel

$$\min \left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} \left(1-\overline{\xi}\right)^{n-i} \geq x \right\} = \boxed{\texttt{BINOM.INVERZ}(n;\overline{\xi};x)}$$

módon számolható, ezért

$$\max \left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} \left(1-\overline{\xi}\right)^{n-i} < x \right\} = \boxed{\texttt{BINOM.INVERZ}(n;\overline{\xi};x)\texttt{-1}}.$$

Így τ_1

`=(BINOM.INVERZ(55;39/55;0,02/2)-1)/55`

módon, míg τ_2

`=BINOM.INVERZ(55;39/55;1-0,02/2)/55`

módon számolható. A kapott eredmények négy tizedesjegyre kerekítve: $\tau_1 \simeq 0{,}5455$ és $\tau_2 \simeq 0{,}8364$. Ellenőrzésképpen közöljük, hogy a feladatban megadott gyakoriság $p = 0{,}72$ paraméterű karakterisztikus eloszlásból származó minta generálása révén adódott.

5.3. Gyakorlatok

5.1. gyakorlat. Oldjuk meg az 5.3. példát annak ismeretében, hogy a várható érték 14. Melyik módszer ad jobb becslést?

Útmutatás. Az 5.2. példa megoldását használjuk. A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 1,9012 illetve 2,8245. Ennek az intervallumnak a hossza 0,9233, míg az előbb kapott intervallum hossza 0,9384, azaz 0,0151-del hosszabb. Tehát a várható érték ismeretében egy kicsit jobb becslést kaptunk.

5.2. gyakorlat. Oldjuk meg az 5.4. példát annak ismeretében, hogy a szórás 0,8. Melyik módszer ad jobb becslést?

Útmutatás. Az 5.1. példa megoldását használjuk. A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 4,2585 illetve 4,8604. Ennek az intervallumnak a hossza 0,6019, míg az előbb kapott intervallum hossza 0,5352, azaz 0,0667-del rövidebb. Tehát a szórás ismeretében rosszabb becslést kaptunk.

5.3. gyakorlat. A `minta-03.txt` fájlban található mintarealizációról a grafikus illeszkedésvizsgálatnál láttuk, hogy exponenciális eloszlásból származik. Ennek a mintarealizációnak az első 100 elemét a `minta-10.txt` fájl tartalmazza. Ebből adjunk az eloszlás λ paraméterére 0,9 biztonsági szintű konfidenciaintervallumot.

Útmutatás. Ha $\xi_1, \ldots, \xi_n$ a minta, akkor $1 - \alpha = 0{,}9$

$$\begin{aligned} F &:= F[\mathrm{Gamma}(n; 1)] \\ \tau_1 &:= \frac{1}{n\overline{\xi}} F^{-1}\left(\frac{\alpha}{2}\right) \\ \tau_2 &:= \frac{1}{n\overline{\xi}} F^{-1}\left(1 - \frac{\alpha}{2}\right) \end{aligned}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,9 biztonsági szintű konfidenciaintervallum λ-ra. A számoláshoz tudnunk kell, hogy $F = F[\mathrm{Gamma}(n; 1)]$ esetén

$$F^{-1}(x) = \boxed{\texttt{GAMMA.INVERZ}(x\texttt{;}n\texttt{;1})}.$$

A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 2,6750 illetve 3,7197. Ellenőrzésképpen közöljük, hogy λ valódi értéke 3,2.

5.4. gyakorlat. Egy esemény 10 000 kísérletből 2562 alkalommal következett be. Adjon az esemény valószínűségére 0,99 biztonsági szintű konfidenciaintervallumot.

Útmutatás. Az 5.5. példához hasonló módon is megoldható a feladat, de mivel a kísérletek száma most nagy, ezért a számolásnál a Moivre–Laplace-tételt is alkalmazhatjuk. Eszerint, ha n a kísérletek száma és $\overline{\xi}$ az ismeretlen p valószínűségű esemény relatív gyakorisága, akkor $1 - \alpha = 0{,}99$

$$\begin{aligned} c &:= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\ \tau_1 &:= \frac{\overline{\xi} + \frac{c^2}{2n} - \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \\ \tau_2 &:= \frac{\overline{\xi} + \frac{c^2}{2n} + \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \end{aligned}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,99 biztonsági szintű konfidenciaintervallum p-re. Az így kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 0,2451 illetve 0,2676. Ennek és az előző intervallumnak a hossza gyakorlatilag megegyezik, így hasonlóan jó mindkét becslés.

A számolás tovább egyszerűsíthető, ha figyelembe vesszük, hogy most $\frac{1}{n} = 0{,}0001$ elhanyagolhatóan kicsi $\frac{1}{\sqrt{n}} = 0{,}01$-hoz képest. Ekkor

$$\tau_1 \simeq \overline{\xi} - \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}$$
$$\tau_2 \simeq \overline{\xi} + \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}.$$

Az így kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 0,2450 illetve 0,2674.

5.5. gyakorlat. A `minta-11.txt` fájlban található mintarealizáció $[7, b]$ intervallumon egyenletes eloszlásból származik. Adjon b-re 0,95 biztonsági szintű konfidenciaintervallumot.

Útmutatás. Ha $\xi_1, \ldots, \xi_n$ a minta, akkor $a = 7$, $1 - \alpha = 0{,}95$

$$F := F[\mathrm{Gamma}(n; 1)]$$
$$c_1 := F^{-1}\left(\frac{\alpha}{2}\right)$$
$$c_2 := F^{-1}\left(1 - \frac{\alpha}{2}\right)$$
$$\tau_1 := a + \left(e^{c_1}\prod_{i=1}^{n}(\xi_i - a)\right)^{\frac{1}{n}}$$
$$\tau_2 := a + \left(e^{c_2}\prod_{i=1}^{n}(\xi_i - a)\right)^{\frac{1}{n}}$$

jelölésekkel $[\tau_1, \tau_2]$ 0,95 biztonsági szintű konfidenciaintervallum b-re. A számoláshoz használja a `KITEVŐ` és `SZORZAT` függvényeket. A kapott intervallum alsó illetve felső végpontja két tizedesjegyre kerekítve 13,25 illetve 17,86. Ellenőrzésképpen közöljük, hogy b valódi értéke 15.

6. fejezet

Paraméteres hipotézisvizsgálatok

Grafikus illeszkedésvizsgálatnál azt néztük meg, hogy lehet-e például normális eloszlású a vizsgált valószínűségi változó. Tehát egy feltételezésről, hipotézisről döntöttünk. A *hipotézisvizsgálatokban* vagy más néven *statisztikai próbákban* szintén a statisztikai mezőre vonatkozó hipotézisekről döntjük el a mintarealizáció alapján, hogy igaz vagy sem, de ennek nem kell feltétlenül az eloszlásra vonatkoznia. Lehet például az a hipotézis, hogy egy valószínűségi változó várható értéke megfelel az előírásnak, vagy két valószínűségi változó független, vagy a várható értékeik megegyeznek stb. Ha a hipotézis ismert eloszlástípusból származó valószínűségi változók paramétereire vonatkozik, akkor *paraméteres hipotézisvizsgálatról* beszélünk.

Azt a feltételezést, amelyről döntést akarunk hozni, *nullhipotézisnek* nevezzük és H_0-val jelöljük. Ha H_0-t elutasítjuk, akkor egy azzal ellentétes állítást fogadunk el, melyet *ellenhipotézisnek* nevezünk és H_1-gyel jelölünk. Döntésünk lehet helyes vagy hibás a következő táblázatnak megfelelően:

	H_0-t elfogadjuk	H_0-t elutasítjuk
H_0 igaz	helyes döntés	*elsőfajú hiba*
H_1 igaz	*másodfajú hiba*	helyes döntés

Ha H_0 teljesülése esetén az elsőfajú hiba valószínűsége maximum α lehet, akkor ezt a számot a *próba terjedelmének*, az $1-\alpha$ számot pedig a *próba szintjének* nevezzük. A statisztikai próba menete a következő:

1. Megadunk egy H_0 teljesülése esetén ismert eloszlású τ statisztikát, mely lényegesen másképpen viselkedik H_0 illetve H_1 teljesülése esetén. Az ilyen statisztikát *próbastatisztikának* nevezzük.
2. Rögzített α ismeretében megadunk egy I intervallumot úgy, hogy H_0 teljesülése esetén a $\tau \in I$ esemény valószínűsége maximum (vagy ha lehet, pontosan)

α legyen. Az I intervallumot *kritikus tartománynak*, míg a komplementerét *elfogadási tartománynak* nevezzük.

3. Ha a mintarealizáció alapján τ a kritikus tartományba esik, akkor H_0-t elutasítjuk (azaz H_1-gyet fogadjuk el), ellenkező esetben pedig H_0-t elfogadjuk a H_1 ellenhipotézissel szemben.

Ekkor α terjedelmű próbát kapunk. Az I megválasztása τ-hoz nem egyértelmű. A lehetséges esetekből úgy kell választani, hogy a másodfajú hiba valószínűsége minél kisebb legyen. Ezért ugyanazon nullhipotézis esetén a különböző ellenhipotézisekkel szemben más és más kritikus tartomány a megfelelő.

A gyakorlatban a próbastatisztikát nem nekünk kell kitalálni, hanem már ismert statisztikai próbák közül választunk a feladat feltételeinek és a célnak megfelelően. A következőkben tárgyalt statisztikai próbákra teljesülnek a következők:

- *Torzítatlan*, azaz H_0-t nagyobb valószínűséggel utasítjuk el, ha H_1 igaz, mint amikor H_0 igaz.
- *Konzisztens*, azaz a minta elemszámának növelésével a másodfajú hiba valószínűsége 0-hoz tart.

Előfordulhat, hogy különböző szinteken különböző döntéseket hozunk ugyanazzal a próbával. Ennek a kellemetlen tulajdonságnak az az oka, hogy α csökkentésével a másodfajú hiba valószínűsége nő. Ilyenkor a konzisztenciát kihasználva, növeljük meg a minta elemszámát úgy, hogy a másodfajú hiba valószínűsége kellően lecsökkenjen.

6.1. Egymintás u-próba

$\xi \in \mathrm{Norm}(m;\sigma)$, m ismeretlen, σ ismert, a ξ-re vonatkozó minta n elemű, $m_0 \in \mathbb{R}$.

$H_0\colon m = m_0$	kritikus tartomány
$H_1\colon m \neq m_0$	$2 - 2\Phi(\lvert u\rvert) < \alpha$
$H_1\colon m < m_0$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u < 0$
$H_1\colon m > m_0$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u > 0$

ahol

$$u = \frac{\overline{\xi} - m_0}{\sigma}\sqrt{n}.$$

Az egymintás u-próba a gazdasági statisztikában *egymintás z-próba* néven ismert.

6.1. Példa. A `minta-12.txt` fájlban található mintarealizáció normális eloszlásból származik. Tudjuk, hogy a szórás 2. Teljesülhet-e, hogy a várható érték nagyobb 13,8-nél? Döntsön 0,99 szinten.

Megoldás. A mintarealizációt másoljuk az A oszlopba, majd számoljuk ki a mintaátlagot $\boxed{\texttt{=ÁTLAG(A:A)}}$ módon. Ennek értéke négy tizedesjegyre kerekítve 13,9824. A kérdés, hogy ez 13,8-től véletlenül nagyobb vagy van valami oka?

Egymintás u-próba alkalmazható, ahol $H_0\colon m = 13{,}8$ és $H_1\colon m > 13{,}8$. Akkor fogunk H_1 mellett dönteni, ha $1 - \Phi(|u|) < \alpha$ és $u > 0$.

A szint 0,99, azaz $\alpha = 0{,}01$. Mivel $\overline{\xi} - m_0 = 13{,}9824 - 13{,}8 > 0$, ezért $u > 0$, másrészt emiatt $1 - \Phi(|u|) = 1 - \Phi(u)$. Mivel

$$1 - \Phi(u) = \boxed{\texttt{Z.PRÓB(A:A;}m_0\texttt{;}\sigma\texttt{)}},$$

ezért egy cellába írjuk a következőt: $\boxed{\texttt{=Z.PRÓB(A:A;13,8;2)}}$. Ennek értéke négy tizedesjegyre kerekítve 0,0047, amitől nagyobb az α. Tehát a kritikus tartományban van a próbastatisztika, azaz H_0-t elutasítjuk és ezzel a H_1-gyet elfogadjuk. Tehát 0,99 szinten a várható érték nagyobb 13,8-nél.

6.2. Kétmintás u-próba

$\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, m_1, m_2 ismeretlenek, σ_1, σ_2 ismertek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2\Phi(\lvert u\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u < 0$
$H_1\colon m_1 > m_2$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u > 0$

ahol

$$u = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

A kétmintás u-próba a gazdasági statisztikában *kétmintás z-próba* néven ismert.

6.2. Példa. A `minta-14.txt` fájlban található mintarealizáció normális eloszlásból származik, melynek szórása 2. A `minta-15.txt` fájlban található mintarealizáció szintén normális eloszlásból származik, melynek szórása 3. Teljesülhet-e, hogy a két minta várható értéke megegyezik? Döntsön 0,98 szinten.

Megoldás (1). Kétmintás u-próba alkalmazható $H_0\colon m_1 = m_2$ és $H_1\colon m_1 \neq m_2$ hipotézisekkel. Az u próbastatisztika értéke körülbelül 4,34, melyből $2 - 2\Phi(|u|)$ értéke öt tizedesjegy pontossággal 0,00001. Mivel $\alpha = 0{,}02$, ezért az ellenhipotézist fogadjuk el, azaz a két várható érték nem egyezik meg.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Kétmintás z-próba a várható értékre* sort, majd *OK*. Kattintson az *1. változótartomány* felirat mezőjére, majd jelölje ki az `A` oszlopot. Kattintson a *2. változótartomány* felirat mezőjére, majd jelölje ki a `B` oszlopot. Az *1. (ismert) változó variancia* értéke `4`, hiszen a szórás 2, így a szórásnégyzet, más néven variancia értéke $2^2 = 4$. Hasonlóan, a *2. (ismert) változó variancia* értéke `9`. Kattintson a *Kimeneti tartomány* felirat mezőjére, majd a `C1` cellára. Végül *OK*. A kapott táblázatban `z` jelenti az u próbastatisztika értékét, `P(Z<=z) egyszélű` jelenti az $1 - \Phi(|u|)$ értékét és `P(Z<=z) kétszélű` jelenti a $2 - 2\Phi(|u|)$ értékét.

6.3. Egymintás t-próba

$\xi \in \mathrm{Norm}(m; \sigma)$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n elemű, $m_0 \in \mathbb{R}$.

$H_0\colon m = m_0$	kritikus tartomány
$H_1\colon m \neq m_0$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m < m_0$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m > m_0$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol

$$t = \frac{\overline{\xi} - m_0}{S_n^*}\sqrt{n} \quad \text{és} \quad F = F[\mathrm{T}(n-1)].$$

6.3. Példa. A `minta-12.txt` fájlban található mintarealizáció normális eloszlásból származik. Teljesülhet-e, hogy a várható érték egyenlő 14-gyel? Döntsön 0,99 szinten.

Megoldás. Egymintás t-próba alkalmazható, ahol $H_0\colon m = 14$ és $H_1\colon m \neq 14$. A szint 0,99, azaz $\alpha = 0{,}01$. Ezt kell összehasonlítani $2 - 2F(|t|)$ értékével. Excelben

$$2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;2;1)}},$$

ahol az `A` oszlopban van a mintarealizáció, és a B oszlop minden olyan cellájában az m_0 értéke van, amely mellett található a mintarealizációnak egy eleme. Eszerint tehát másoljuk a mintarealizációt az `A` oszlopba, majd a `B1` cellába írjuk be, hogy 14 (most ez az m_0). Ezután a `B1` cella kitöltőjelére kattintsunk kétszer. Ezzel a mintarealizáció minden tagja mellé 14 kerül. Már csak egy üres cellába az előbb említett módon ki kell számolni a $2 - 2F(|t|)$ értékét. Ez négy tizedesjegyre kerekítve 0,7998. Ettől kisebb α, így a nullhipotézist fogadjuk el. Tehát a várható érték 14.

6.4. F-próba

$\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon \sigma_1 = \sigma_2$	kritikus tartomány
$H_1\colon \sigma_1 \neq \sigma_2$	$2\min\{F(\mathsf{F}), 1-F(\mathsf{F})\} < \alpha$
$H_1\colon \sigma_1 < \sigma_2$	$\min\{F(\mathsf{F}), 1-F(\mathsf{F})\} < \alpha$ és $\mathsf{F} < 1$
$H_1\colon \sigma_1 > \sigma_2$	$\min\{F(\mathsf{F}), 1-F(\mathsf{F})\} < \alpha$ és $\mathsf{F} > 1$

ahol

$$\mathsf{F} = \frac{{S^*}^2_{\xi,n_1}}{{S^*}^2_{\eta,n_2}} \quad \text{és} \quad F := F[\mathrm{F}(n_1-1;n_2-1)].$$

6.4. Példa. A minta-12.txt és minta-14.txt fájlban található mintarealizációk normális eloszlásból származnak. Teljesülhet-e, hogy a két minta szórása megegyezik? Döntsön 0,99 szinten.

Megoldás (1). F-próba alkalmazható $H_0\colon \sigma_1 = \sigma_2$ és $H_1\colon \sigma_1 \neq \sigma_2$ hipotézisekkel. Ha az `A` illetve a `B` oszlopokba rakjuk a két mintát, akkor

$$2\min\{F(\mathsf{F}),1-F(\mathsf{F})\} = \boxed{\texttt{F.PRÓB(A:A;B:B)}},$$

amely most négy tizedesjegyre kerekítve 0,6588. Az $\alpha = 0{,}01$ ettől kisebb, azaz a nullhipotézist fogadjuk el. Tehát a szórások megegyeznek.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Kétmintás F-próba a szórásnégyzetre* sort, majd *OK*. Kattintson az *1. változótartomány* felirat mezőjére, majd jelölje ki az `A` oszlopot. Kattintson a *2. változótartomány* felirat mezőjére, majd jelölje ki a `B` oszlopot. Kattintson a *Kimeneti tartomány* felirat mezőjére, majd a `C1` cellára. Végül *OK*. A kapott táblázatban `F` jelenti az F próbastatisztika értékét, `P(F<=f) egyszélű` jelenti a $\min\{F(\mathsf{F}),1-F(\mathsf{F})\}$ értékét.

6.5. Kétmintás t-próba

$\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, $\sigma_1 = \sigma_2$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2-2F(\lvert t\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1-F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m_1 > m_2$	$1-F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol

$$t = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{n_1 S^2_{\xi,n_1} + n_2 S^2_{\eta,n_2}}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \quad \text{és} \quad F = F[\mathrm{T}(n_1 + n_2 - 2)].$$

A feltételben szerepel, hogy a szórásoknak meg kell egyezni, melyről F-próbával dönthetünk.

6.5. Példa. A `minta-12.txt` és `minta-14.txt` fájlban található mintarealizációk származhatnak-e azonos normális eloszlásból? Döntsön 0,99 szinten.

Megoldás (1). Az előző példában láttuk, hogy a szórások megegyeznek. Ezért a várható értékek egyezésére alkalmazhatunk kétmintás t-próbát $H_0\colon m_1 = m_2$ és $H_1\colon m_1 \neq m_2$ hipotézisekkel. Ha az `A` illetve a `B` oszlopokba rakjuk a két mintát, akkor

$$2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;2;2)}},$$

amely most kb. $7{,}4972 \cdot 10^{-44}$, azaz kisebb az $\alpha = 0{,}01$ értéktől. Így az ellenhipotézist fogadjuk el. Tehát a várható értékek különböznek, vagyis nem azonos normális eloszlásból származik a két minta.

Kiszámolva a mintaátlagokat kiderül, hogy az `A` oszlopban nagyobb a mintaátlag, mint a `B`-ben, azaz $\overline{\xi} - \overline{\eta} > 0$. Ebből $t > 0$, továbbá

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;2)}}$$

értéke kb. $3{,}7486 \cdot 10^{-44}$, azaz kisebb az $\alpha = 0{,}01$ értéktől. Így $H_1\colon m_1 > m_2$ esetén a kritikus tartományba esik a próbastatisztika, azaz amellett döntünk, hogy az A oszlopban található mintarealizációhoz tartozó valószínűségi változó várható értéke a nagyobb.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Kétmintás t-próba egyenlő szórásnégyzeteknél* sort, majd *OK*. Kattintson az *1. változótartomány* felirat mezőjére, majd jelölje ki az `A` oszlopot. Kattintson a *2. változótartomány* felirat mezőjére, majd jelölje ki a `B` oszlopot. Kattintson a *Kimeneti tartomány* felirat mezőjére, majd a `C1` cellára. Végül *OK*. A kapott táblázatban `t érték` jelenti a t próbastatisztika értékét, `P(T<=t) egyszélű` jelenti az $1 - F(|t|)$ értékét és `P(T<=t) kétszélű` jelenti a $2 - 2F(|t|)$ értékét.

6.6. Scheffé-módszer

$\xi \in \text{Norm}(m_1;\sigma_1)$, $\eta \in \text{Norm}(m_2;\sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű, $n_1 \leq n_2$.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m_1 > m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol

$$\zeta_i := \xi_i - \sqrt{\frac{n_1}{n_2}}\eta_i + \frac{1}{\sqrt{n_1 n_2}}\sum_{k=1}^{n_1}\eta_k - \overline{\eta} \quad (i = 1, \ldots, n_1)$$

jelöléssel

$$t = \frac{\overline{\zeta}}{S^*_{\zeta,n_1}}\sqrt{n_1} \quad \text{és} \quad F = F[\mathrm{T}(n_1 - 1)].$$

Vagyis a $\zeta_1, \ldots, \zeta_{n_1}$ mintára alkalmazzuk az egymintás t-próbát $m_0 = 0$ választással.

Speciálisan $n_1 = n_2$ esetén $\zeta_i = \xi_i - \eta_i$ teljesül. A módszert ekkor *párosított t-próbának* is nevezik. Ebben az esetben a módszer akkor is alkalmazható, ha a minták nem függetlenek, de csak akkor, ha $\xi - \eta$ normális eloszlású.

6.6. Példa. Egy iskolában új módszert akarnak kipróbálni a gyerekek problémamegoldó képességeinek javítására. A kísérlet elején kitöltetnek 50 diákkal egy ilyen képességet mérő tesztet. A kapott eredményeket %-ban, a diákok névsorával megegyező sorrendben rögzítették a `minta-20.txt` fájlban. Egy éven keresztül alkalmazzák a módszert ezeken a diákokon. A tesztet az egy év leteltével megismétlik. A kapott eredményeket ismét a diákok névsorával megegyező sorrendben leírták a `minta-21.txt` fájlban. Ennek alapján döntsön 0,99 szinten arról, hogy a módszer sikeresnek mondható-e.

Megoldás (1). Az első illetve második teszt eredményeit másolja az `A` illetve `B` oszlopba. Mivel a két minta nem tekinthető függetlennek, ezért a különbség mintát kell megvizsgálni, hogy normális eloszlású-e. A `C1` cellába írja be, hogy `=A1-B1`, majd a kitöltőjelre kattintson kétszer. Így a `C` oszlopban megjelenik a különbségminta. Erre grafikus normalitásvizsgálatot végzünk a korábban ismertetett módon, melyből kiderül, hogy a különbség minta normális eloszlásúnak tekinthető. Így alkalmazhatjuk a várható értékek összehasonlítására a párosított t-próbát.

Az első mintarealizáció átlaga 65,94, a másodiké pedig 69,78. A módszernek köszönhető az átlagnövekedés, vagy a véletlen műve? Az ellenhipotézis legyen az, hogy a módszer sikeres, azaz a második minta várható értéke nagyobb mint az elsőé. Mivel $\overline{\zeta} = \overline{\xi} - \overline{\eta} = 65{,}94 - 69{,}78 < 0$, ezért $t < 0$ továbbá

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;1)}}$$

értéke négy tizedesjegyre kerekítve 0,0002. Ettől $\alpha = 0{,}01$ nagyobb, tehát az ellenhipotézist fogadjuk el, azaz a módszer sikeresnek tekinthető.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Kétmintás párosított t-próba a várható értékre* sort, majd *OK*. Kattintson az *1. változótartomány* felirat mezőjére, majd jelölje ki az `A` oszlopot. Kattintson a *2. változótartomány* felirat mezőjére, majd jelölje ki a `B` oszlopot. Kattintson a *Kimeneti tartomány* felirat mezőjére, majd a `C1` cellára. Végül *OK*. A kapott táblázatban **`t érték`** jelenti a t próbastatisztika értékét, **`P(T<=t) egyszélű`** jelenti az $1 - F(|t|)$ értékét és **`P(T<=t) kétszélű`** jelenti a $2 - 2F(|t|)$ értékét.

6.7. Példa. A `minta-06.txt` és `minta-08.txt` fájlokban normális eloszlású független minták vannak. Döntsön a várható értékeik egyezéséről 0,98 szinten.

Megoldás. A **`minta-06`**-ot illetve **`minta-08`**-at másolja az `A` illetve `B` oszlopba. Először végezzen F-próbát. Ennek az lesz az eredménye, hogy a szórások nem egyeznek meg. Ezért a kétmintás t-próba nem alkalmazható, így a Scheffé-módszert használjuk. A $\zeta_1, \ldots, \zeta_{n_1}$ minta elkészítéséhez tegye a következőket.

Egyelőre a `C` és `D` oszlopokat hagyja üresen. Számolja ki az első illetve második minta elemszámát. Kapjuk, hogy ez 456 illetve 50. Mivel a második minta elemszáma kisebb, ezért ez lesz a ξ-re vonatkozó mintarealizáció. Először számolja ki az

$$\zeta_1 := \xi_1 - \sqrt{\frac{n_1}{n_2}}\eta_1 + \frac{1}{\sqrt{n_1 n_2}} \sum_{k=1}^{n_1} \eta_k - \overline{\eta}$$

értékét a `C1` cellába

$$\boxed{\texttt{=B1-GYÖK(50/456)*A1+SZUM(A1:A50)/GYÖK(50*456)-ÁTLAG(A:A)}}$$

módon. Húzza le a `C1` cella kitöltőjelét az 50. sorig. Ezzel a `C` oszlopban elkészült a $\zeta_1, \ldots, \zeta_{n_1}$ minta. A Scheffé-módszer szerint erre kell alkalmazni az egymintás t-próbát $m_0 = 0$ választással. Tehát a `D1` cellába írja be, hogy `0`, majd a kitöltőjelet húzza le az 50. sorig. Ekkor

$$2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(C:C;D:D;2;1)}}.$$

Ennek értéke kb. $3{,}5514 \cdot 10^{-5}$ ami kisebb $\alpha = 0{,}02$-nál, így a várható értékek nem egyeznek meg.

A `C` oszlop átlaga negatív, amely $\overline{\zeta}$ értékét adja. Így $t < 0$, továbbá

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(C:C;D:D;1;1)}}$$

értéke kb. $1{,}7757 \cdot 10^{-5}$ ami kisebb $\alpha = 0{,}02$-nál. Így ξ-nek, azaz a második mintához tartozó valószínűségi változónak a várható értéke kisebb.

6.7. Welch-próba

$\xi \in \mathrm{Norm}(m_1; \sigma_1)$, $\eta \in \mathrm{Norm}(m_2; \sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0 \colon m_1 = m_2$	kritikus tartomány
$H_1 \colon m_1 \neq m_2$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1 \colon m_1 < m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1 \colon m_1 > m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol

$$t := \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\dfrac{S^{*2}_{\xi,n_1}}{n_1} + \dfrac{S^{*2}_{\eta,n_2}}{n_2}}} \qquad \text{és} \qquad F \simeq F[\mathrm{T}(s)].$$

Az s szabadsági fok c értékének kerekítése a legközelebbi egészre, ahol

$$a := \frac{S^{*2}_{\xi,n_1}}{n_1}, \quad b := \frac{S^{*2}_{\eta,n_2}}{n_2}, \quad c := \frac{(a+b)^2}{\dfrac{a^2}{n_1 - 1} + \dfrac{b^2}{n_2 - 1}}.$$

Ezt a próbát akkor szokták alkalmazni, ha az F-próba a szórások különbözőségét mutatja ki, ugyanis a szórások egyezése esetén a kétmintás t-próba megbízhatóbb.

Legyen $F_k := F[\mathrm{T}(k)]$, ahol $k \in \mathbb{N}$. Az előbb az $F(|t|) \simeq F_s(|t|)$ közelítést alkalmaztuk, de ezt lehet finomítani *polinomiális interpolációval.* Ehhez tekintsük azt a g polinomot, melyre $g(k) = F_k(|t|)$ minden $k \in \mathbb{N}$ esetén. Ezután az $F(|t|) \simeq g(c)$ közelítéssel számolhatunk tovább.

6.8. Példa. A `minta-06.txt` és `minta-08.txt` fájlokban normális eloszlású független minták vannak. Döntsön a várható értékeik egyezéséről 0,99 szinten.

Megoldás (1). A `minta-06`-ot illetve `minta-08`-at másolja az `A` illetve `B` oszlopba. Először végezzen F-próbát. Ennek az lesz az eredménye, hogy a szórások nem egyeznek meg. Ezért a kétmintás t-próba nem alkalmazható, így a Welch-próbát használjuk. Ekkor

$$2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;2;3)}}.$$

Ennek értéke kb. $2{,}7336 \cdot 10^{-5}$ ami kisebb $\alpha = 0{,}01$-nál, így a várható értékek nem egyeznek meg.

Az első minta átlaga nagyobb a másodiknál, ezért $t > 0$, továbbá

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;3)}}$$

értéke kb. $1{,}3668 \cdot 10^{-5}$ ami kisebb $\alpha = 0{,}01$-nál. Így az első mintához tartozó valószínűségi változónak a várható értéke nagyobb.

Megoldás (2). A feladat az *Adatok/Adatelemzés* menüponttal is megoldható. A felugró ablakban a legördülő listában válassza a *Kétmintás t-próba nem-egyenlő szórásnégyzeteknél* sort, majd *OK*. Kattintson az *1. változótartomány* felirat mezőjére, majd jelölje ki az `A` oszlopot. Kattintson a *2. változótartomány* felirat mezőjére, majd jelölje ki a `B` oszlopot. Kattintson a *Kimeneti tartomány* felirat mezőjére, majd a `C1` cellára. Végül *OK*. A kapott táblázatban `t érték` jelenti a t próbastatisztika értékét, `P(T<=t) egyszélű` jelenti az $1-F(|t|)$ értékét és `P(T<=t) kétszélű` jelenti a $2 - 2F(|t|)$ értékét.

Azt fogja tapasztalni, hogy ebben az esetben egy kicsi eltérés lesz a `T.PRÓB` függvénnyel számolt értékektől. Ennek az az oka, hogy a `T.PRÓB` függvény pontos szabadsági fokkal számol polinomiális interpolációt használva, míg az utóbbi módszer egészre kerekített szabadsági fokot vesz alapul.

6.8. Khi-négyzet próba normális eloszlás szórására

$\xi \in \mathrm{Norm}(m;\sigma)$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n elemű, $\sigma_0 > 0$.

$H_0\colon \sigma = \sigma_0$	kritikus tartomány
$H_1\colon \sigma \neq \sigma_0$	$2\min\{F(\chi^2), 1-F(\chi^2)\} < \alpha$
$H_1\colon \sigma < \sigma_0$	$\min\{F(\chi^2), 1-F(\chi^2)\} < \alpha$ és $\chi^2 < n-1$
$H_1\colon \sigma > \sigma_0$	$\min\{F(\chi^2), 1-F(\chi^2)\} < \alpha$ és $\chi^2 > n-1$

ahol

$$\chi^2 = \frac{{S_n^*}^2}{\sigma_0^2}(n-1) \quad \text{és} \quad F = F[\mathrm{Khi}(n-1)].$$

6.9. Példa. Egy alkatrész valamelyik paraméterére vonatkozó normális eloszlású minta a `minta-07.txt` fájlban található. Előzetes vizsgálat kimutatta, hogy a várható érték megfelel az előírásnak. A selejtarány alacsonyan tartása miatt a szórás nem lehet nagyobb 3-nál. Eleget tesznek-e a legyártott alkatrészek ennek a feltételnek? Döntsön 0,98 szinten.

Megoldás. A mintarealizációt másolja az `A` oszlopba. Ennek korrigált tapasztalati szórását `=SZÓR.M(A:A)` módon számolhatja ki, amely négy tizedesjegyre kerekítve 3,2451, ami nagyobb $\sigma_0 = 3$-tól. Ebből következően $\chi^2 > n - 1$.

Legyen $H_1 \colon \sigma > 3$, azaz, hogy a szórás nem felel meg a feltételnek. Ekkor $F(\chi^2)$ értéke

`=KHINÉGYZET.ELOSZLÁS(VAR.M(A:A)*(DARAB(A:A)-1)/9;DARAB(A:A)-1;IGAZ)`,

vagy

`=KHINÉGYZET.ELOSZLÁS(VAR.S(A:A)*DARAB(A:A)/9;DARAB(A:A)-1;IGAZ)`,

módon számolható, amely 0,9997 négy tizedesjegyre kerekítve. Így $\min\{F(\chi^2), 1 - F(\chi^2)\} = 0{,}0003$ négy tizedesjegyre kerekítve. Ettől nagyobb $\alpha = 0{,}02$, tehát az ellenhipotézist fogadjuk el, azaz a gyártmány nem felel meg a szórásra vonatkozó feltételnek.

6.9. Statisztikai próba az exponenciális eloszlás paraméterére

$\xi \in \mathrm{Exp}(\lambda)$, ahol λ ismeretlen, a ξ-re vonatkozó minta n elemű, $\lambda_0 > 0$.

$H_0 \colon \lambda = \lambda_0$	kritikus tartomány
$H_1 \colon \lambda \neq \lambda_0$	$2\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$
$H_1 \colon \lambda < \lambda_0$	$\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$ és $\gamma > n$
$H_1 \colon \lambda > \lambda_0$	$\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$ és $\gamma < n$

ahol

$$\gamma = \lambda_0 n \overline{\xi} \quad \text{és} \quad F = F[\mathrm{Gamma}(n; 1)].$$

6.10. Példa. Egy rendelet szerint a háziorvosi rendelőkben az orvosoknak minden beteggel átlagosan 5 percnél többet kell foglalkoznia. Egy adott rendelőben feljegyeznek néhány beteggel való konzultáció idejét percben. A mintát a `minta-22.txt` fájl tartalmazza, mely exponenciális eloszlású. Ez alapján betartja-e az itteni orvos a rendeletet? Hozzon döntést 0,99 szinten.

Megoldás. Másolja a mintát az `A` oszlopba. A mintaátlag 7,0669 perc, tehát úgy tűnik, hogy az orvos betartja a rendeletet. Kérdés, hogy ez csak véletlen, vagy szignifikánsan nagyobb az átlagos konzultáció idő 5 percnél? Legyen az az ellenhipotézis, hogy az orvos betartja a rendeletet, azaz $\mathrm{E}\,\xi = \frac{1}{\lambda} > 5$. Így tehát $H_1\colon \lambda < 0{,}2$. Mivel $\lambda_0 = 0{,}2$ és $\overline{\xi} > \frac{1}{\lambda_0} = 5$, ezért $\gamma > n$, másrészt

$$F(\gamma) = \boxed{\texttt{GAMMA.ELOSZL(0,2*SZUM(A:A);DARAB(A:A);1;IGAZ)}},$$

amely most 0,9989 négy tizedesjegyre kerekítve. Ebből $\min\{F(\gamma), 1 - F(\gamma)\} = 0{,}0011$. Ettől $\alpha = 0{,}01$ nagyobb, ezért az ellenhipotézist fogadjuk el, miszerint az orvos betartja a rendeletet.

6.10. Statisztikai próba valószínűségre

$\xi \in \mathrm{Bin}(1; p)$, p ismeretlen, a ξ-re vonatkozó minta n elemű, $0 < p_0 < 1$.

$H_0\colon p = p_0$	kritikus tartomány
$H_1\colon p \neq p_0$	$n\overline{\xi} < F^{-1}\left(\frac{\alpha}{2}\right)$ vagy $n\overline{\xi} > F^{-1}\left(1 - \frac{\alpha}{2}\right)$
$H_1\colon p < p_0$	$n\overline{\xi} < F^{-1}(\alpha)$
$H_1\colon p > p_0$	$n\overline{\xi} > F^{-1}(1 - \alpha)$

ahol

$$F^{-1}(x) = \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \geq x \right\}.$$

6.11. Példa. A tejiparban hasznos lehetne egy olyan eljárás, melynek révén nagyobb arányban születne üszőborjú, mint bikaborjú, hiszen ekkor több fejőstehenet nevelhetnének fel azonos születésszám mellett. Egy kutató javasol egy módszert erre. Az állításának ellenőrzésére elvégeznek 100 ilyen eljárást, melynek révén 61 darab üszőborjú született. Ennek alapján döntsön 0,99 szinten arról, hogy hatásos-e a módszer.

Megoldás. Az új eljárás révén az üszőborjú születésének relatív gyakorisága (mely a valószínűség hatásos becslése) 0,61. Kérdés, hogy a relatív gyakoriság csak a véletlennek köszönhetően nagyobb $\frac{1}{2}$-nél, vagy azért mert a valószínűség is nagyobb $\frac{1}{2}$-nél? Legyen ξ az indikátorváltozója annak az eseménynek, hogy az eljárás révén üszőborjú születik és p ennek az eseménynek a valószínűsége. Ekkor $n = 100$ és $n\overline{\xi} = 61$. A módszer akkor hatékony, ha $p > \frac{1}{2}$. Alkalmazzuk az előbb ismertetett statisztikai próbát $p_0 = \frac{1}{2}$ választással. Tehát $H_0\colon p = \frac{1}{2}$ és $H_1\colon p > \frac{1}{2}$. Mivel

$$\min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \geq x \right\} = \boxed{\texttt{BINOM.INVERZ}(n;p;x)},$$

így most

$$F^{-1}(1-\alpha) = F^{-1}(0{,}99) = \boxed{\texttt{BINOM.INVERZ(100;1/2;0,99)}} = 62.$$

Mivel $n\overline{\xi} = 61 < 62$ így a nullhipotézist fogadjuk el, azaz a módszer nem hatékony.

Azonban, ha 0,95 szinten döntenénk, akkor $F^{-1}(0{,}95) = 58$ miatt már az ellenhipotézist fogadnánk el, hiszen $n\overline{\xi} = 61 > 58$. Így a válaszunk a szint függvénye. Ezért tanácsos további kísérleteket végezni.

6.12. Példa. Tegyük fel, hogy az előző feladatban további 100 esetet megvizsgálnak, és azt kapják, hogy a most már összesen 200 esetből 125 alkalommal született üszőborjú. Így is döntsön 0,99 szinten arról, hogy hatásos-e a módszer. Kell-e újabb kísérleteket végezni?

Megoldás. Ekkor $F^{-1}(0{,}99) = 116$ (a feltétel az, hogy ez az érték 101 és 199 közé essen, ami teljesül), így a módszert hatékonynak mondhatjuk, hiszen $125 > 116$. Az F^{-1} monoton növekvő, így ennél kisebb szinten is ugyanígy döntenénk. Ezért további kísérletekre nincs szükség.

6.11. Gyakorlatok

6.1. gyakorlat. Egy gépsoron csavarokat készítenek. Az előírás az, hogy a csavarok hossza 14 mm legyen. Néhány hosszát lemérik. A `minta-12.txt` fájlban található a mintarealizáció, mely normális eloszlásból származik és tudjuk, hogy a szórás 2. Eleget tesznek-e a csavarok a hosszúságra vonatkozó előírásnak, vagy állítani kell a gép pontosságán? Döntsön 0,99 szinten.

Útmutatás. Használjon egymintás u-próbát $H_0\colon m = 14$ és $H_1\colon m \neq 14$ hipotézisekkel. Könnyen látható, hogy $1 - \Phi(|u|) = \min\{1 - \Phi(u), \Phi(u)\}$, így

$$2 - 2\Phi(|u|) = \boxed{\texttt{2*MIN(Z.PRÓB(A:A;14;2);1-Z.PRÓB(A:A;14;2))}}.$$

6.2. gyakorlat. Egy kereskedő egy malomtól nagy tételben lisztet rendel 1 kg-os kiszerelésben. A megvásárolt tételből 100 zacskót lemérnek grammban. A mintarealizáció a `minta-13.txt` fájlban található. Tudjuk, hogy a minta normális eloszlásból származik, melynek 10 gramm a szórása. Döntsön 0,99 szinten, hogy a kereskedő elfogadja-e a szállítmányt.

Útmutatás. Használjon egymintás u-próbát $H_0\colon m = 1000$ nullhipotézissel. A kereskedő csak akkor nem fogadja el a szállítmányt, ha a $H_1\colon m < 1000$ ellenhipotézis teljesül.

6.3. gyakorlat. Oldja meg az előző gyakorlatot úgy is, ha nem ismerjük a szórást.

Útmutatás. Használjon egymintás t-próbát az előző hipotézisekkel. Excelben

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;1)}},$$

ahol az `A` oszlopban van a mintarealizáció, és a B oszlop minden olyan cellájában az $m_0 = 1000$ értéke van, amely mellett található a mintarealizációnak egy eleme.

6.4. gyakorlat. Egy kórháznak olyan fájdalomcsillapítóra van szüksége, amely 12 percen belül hat. Egy bizonyos fajtát kipróbálnak néhány betegen. A hatás elérését percben mérik. A `minta-14.txt` fájlban található a mintarealizáció, mely normális eloszlásból származik. Döntsön 0,99 szinten, hogy ez a fájdalomcsillapító megfelel-e az igényeknek.

Útmutatás. Használjon egymintás t-próbát $H_0\colon m = 12$ és $H_1\colon m > 12$ hipotézisekkel.

6.5. gyakorlat. Két fájdalomcsillapító injekció hatásosságát mérik. Mindkettőt kipróbálják több betegen. Percben mérik a hatásának elérését. Az első fájdalomcsillapítóra vonatkozó mintarealizáció a `minta-14.txt` fájlban található. Ez normális eloszlásból származik, melynek szórása 2. A második fájdalomcsillapítóra vonatkozó mintarealizáció a `minta-15.txt` fájlban található. Ez szintén normális eloszlásból származik, melynek szórása 3. Melyik szer tekinthető hatásosabbnak? Döntsön 0,99 szinten.

Útmutatás. Használjon kétmintás u-próbát.

6.6. gyakorlat. Az előző gyakorlatot oldja meg a szórások ismerete nélkül is. Változott-e a döntése?

Útmutatás. Használjon F-próbát majd Scheffé-módszert vagy Welch-próbát.

6.7. gyakorlat. Két gépsor által gyártott csavarokat ellenőrzik. A csavarokból mintát vesznek és ezeket lemérik mm-ben. Az első illetve második gépre vonatkozó minta a `minta-06.txt` illetve `minta-08.txt` fájlokban található, melyek normális eloszlásúak. Ezekből egymintás t-próbákkal megállapították, hogy mindkét gép eleget tesz a hosszúságra vonatkozó előírásoknak. A gépek pontosságát így már csak a szórásuk határozza meg. Döntsön 0,98 szinten arról, hogy melyik gépsor tekinthető pontosabbnak.

Útmutatás. Használjon F-próbát.

6.8. gyakorlat. Két különböző márkájú golflabdát tesztelnek. Egy golfozó ugyanazzal az ütővel mindkét márkájú labdából elüt néhányat. Az ütéstávolságokat lemérik méterben. Az első ill. második márkára vonatkozó minta a `minta-16.txt` illetve `minta-17.txt` fájlokban található, melyek normális eloszlásúak. Melyik labdamárka tekinthető jobbnak 0,99 szinten, ha csak az ütőtávolság várható értékét vesszük alapul?

Útmutatás. Először F-próbát alkalmazzon, melynek az lesz az eredménye, hogy a szórások egyformának tekinthetők 0,99 szinten. Így a várható értékre kétmintás t-próba alkalmazható.

Először legyen az az ellenhipotézis, hogy a labdamárkák különböző minőségűek. Ekkor a $2 - 2F(|t|)$ értékét kell kiszámolni, amely Excelben

$$\boxed{\texttt{T.PRÓB(A:A;B:B;2;2)}},$$

ahol az `A` oszlopban az első márkára, míg a `B` oszlopban a második márkára vonatkozó adatok vannak. Ebből azt fogjuk kapni, hogy különbözőek a labdák. Ezután már fölösleges az egyoldali ellenhipotézisre is megcsinálni a próbát, elég csak a mintaátlagok viszonyát megvizsgálni. Mivel az első minta átlaga nagyobb, így azt kapjuk, hogy az első labdamárka jobb.

6.9. gyakorlat. Egy lőszergyártó cég azt állítja, hogy sikerült kifejlesztenie egy mesterlövő puskához egy olyan új lőszert, amellyel nagyobb a találati pontosság, mint a hagyományossal. Ennek ellenőrzésére ugyanakkora távolságból lőnek egy célra a hagyományos és az új lőszerrel is. A találat távolságát a célponttól mm-ben mérik. A hagyományos illetve új lőszerre kapott minták a `minta-18.txt` illetve `minta-19.txt` fájlokban vannak, melyek normális eloszlásúak. Döntsön 0,99 szinten arról, hogy igaz-e a gyár állítása.

Útmutatás. A `minta-18`-at illetve a `minta-19`-et másolja az `A` illetve `B` oszlopba. Először F-próbát alkalmazzon, melynek az lesz az eredménye, hogy a szórások egyformának tekinthetők 0,99 szinten. Így a várható értékre kétmintás t-próba alkalmazható. Legyen az az ellenhipotézis, hogy az új lőszernek nagyobb a találati pontossága. Mivel az első minta átlaga nagyobb, ezért $t > 0$, továbbá

$$1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;2)}}.$$

Ettől nagyobb az α, ezért az ellenhipotézist fogadjuk el. Tehát a gyár állítása igaz.

6.10. gyakorlat. A `minta-13.txt` fájlban egy normális eloszlásból származó mintarealizáció található. Döntsön 0,99 szinten arról, hogy a szórás értéke 10.

Útmutatás. Használjon khi-négyzet próbát.

6.11. gyakorlat. A `minta-10.txt` fájlban egy exponenciális eloszlásból származó mintarealizáció található. Döntsön 0,99 szinten arról, hogy a paraméter értéke 2,3.

6.12. gyakorlat. Egy dobókockával 1000 dobásból 186 alkalommal dobtunk hatost. Döntsön 0,99 szinten arról, hogy a hatos dobásának a valószínűsége $\frac{1}{6}$.

6.13. gyakorlat. Az ötös lottó 3000 sorsolásából 190 alkalommal húzták ki az 1 számot. Valaki azt állítja, hogy ez gyanúsan sok, valami csalás van a dologban. Döntsön 0,99 szinten arról, hogy igaza van-e az illetőnek.

Útmutatás. Rendes esetben annak a valószínűsége, hogy egy lottó ötösben szerepeljen az 1 szám $\frac{\binom{89}{4}}{\binom{90}{5}} = \frac{5}{90}$. Legyen p a valódi valószínűsége ennek az eseménynek. Ekkor az illető állítása $H_1 \colon p > \frac{5}{90}$. A kritikus érték 196, melytől a gyakoriság kisebb, azaz nem esik a kritikus tartományba. Így nincs igaza az illetőnek, ez a gyakoriság még nem gyanúsan sok. Másrészt például 0,9 szinten már azt kapnánk, hogy H_1 igaz, így biztosabb válaszhoz további sorsolásokra lesz szükség.

7. fejezet

Nemparaméteres hipotézisvizsgálatok

Ebben a fejezetben $1-\alpha$ ismét a próba szintjét jelenti.

7.1. Tiszta illeszkedésvizsgálat valószínűségre

$A_1, \dots, A_r$ teljes eseményrendszer, $p_1, \dots, p_r \in \mathbb{R}_+$, $p_1 + \dots + p_r = 1$.

$$\boxed{H_0\colon A_i \text{ valószínűsége } p_i\ \forall i},$$

Legyen ϱ_i az A_i gyakorisága n kísérlet után ($\varrho_i \geq 10\ \forall i$),

$$\nu_i = np_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\mathrm{Khi}(r-1)].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.1. Példa. Egy dobókockával 400-szor dobtunk. Az 1, 2, 3, 4, 5, 6 dobások gyakoriságai rendre 69, 50, 57, 64, 63, 97. Döntsön 0,99 szinten arról, hogy szabályos-e a kocka.

Megoldás. Legyen az a nullhipotézis, hogy a dobókocka szabályos, azaz $p_1 = \dots = p_6 = \frac{1}{6}$. Az `A1:A6` cellákba írja be rendre a 69, 50, 57, 64, 63, 97 gyakoriságokat, melyek mindegyike nagyobb 10-nél, így alkalmazható a próba. A `B1` cellába írja be $\nu_1 = np_1$ értékét, azaz `=400/6`. Ezután a `B1` cella kitöltőjelére kattintson kétszer. Tehát most `A1:A6` tartalmazza a ϱ_i, míg `B1:B6` a ν_i értékeket. Ekkor

$$1 - F(\chi^2) = \boxed{\texttt{KHINÉGYZET.PRÓBA(A1:A6;B1:B6)}}.$$

A kijött érték 0,0014 négy tizedesjegyre kerekítve, ami kisebb $\alpha = 0{,}01$-nál, így elutasítjuk a nullhipotézist, azaz a dobókocka cinkelt.

7.2. Tiszta illeszkedésvizsgálat eloszlásfüggvényre

Legyen ξ a vizsgált valószínűségi változó és F_0 egy eloszlásfüggvény.

$$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F_0}$$

Legyen $a_1 < a_2 < \cdots < a_{r-1}$, $I_1 = (-\infty, a_1)$, $I_2 = [a_1, a_2)$, $I_3 = [a_2, a_3)$, $\ldots$, $I_{r-1} = [a_{r-2}, a_{r-1})$, $I_r = [a_{r-1}, \infty)$. Jelölje ϱ_i a ξ-re vonatkozó n elemű mintában az I_i intervallumba eső mintaelemek számát ($\varrho_i \geq 10\ \forall i$), továbbá legyen $p_1 = F_0(a_1)$, $p_2 = F_0(a_2) - F_0(a_1)$, $p_3 = F_0(a_3) - F_0(a_2)$, $\ldots$, $p_{r-1} = F_0(a_{r-1}) - F_0(a_{r-2})$, $p_r = 1 - F_0(a_{r-1})$,

$$\nu_i = np_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\mathrm{Khi}(r-1)].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.2. Példa. A `minta-23.txt` fájlban található mintarealizáció alapján döntse el 0,99 szinten, hogy a vizsgált valószínűségi változó lehet-e $\lambda = 1$ paraméterű Poisson-eloszlású.

Megoldás. Legyen az a nullhipotézis, hogy a vizsgált valószínűségi változó $\lambda = 1$ paraméterű Poisson-eloszlású. A mintarealizáció elemeit másolja az `A` oszlopba. A `B1:B3` cellatartományba írja be rendre a 0, 1, 2 számokat, majd a `C1` cellába, hogy

`=POISSON.ELOSZLÁS(B1;1;HAMIS)`.

A `C1` kitöltőjelére kattintson kétszer. Végül a `C4` cellába írja be, hogy `=1-SZUM(C1:C3)`. Ezzel a `C` oszlopban megjelennek a p_1, p_2, p_3, p_4 értékek, ahol $a_1 = 1$, $a_2 = 2$, $a_3 = 3$ és F_0 a $\lambda = 1$ paraméterű Poisson-eloszlás eloszlásfüggvénye. Mivel $p_4 = 0{,}0803$ négy tizedesjegyre kerekítve, ezért az utolsó intervallum további felbontására nincs szükség.

A `D1` cellába írja be, hogy `=DARABTELI(A:A;B1)`, majd a kitöltőjelet húzza le a 3. sorig. A `D4` cellába írja ezt: `=DARABTELI(A:A;">=3")`. Ezzel megkapja a ϱ_i gyakoriságokat, melyek mindegyike nagyobb 10-nél, így alkalmazható a próba. Az `E` oszlopban számolja ki a $\nu_i = np_i$ értékeket. Írja az `E1` cellába, hogy `=DARAB(A:A)*C1`, majd a kitöltőjelet húzza le a 4. sorig. Ekkor

$1 - F(\chi^2) =$ `KHINÉGYZET.PRÓBA(D1:D4;E1:E4)`.

A kijött érték 0,3017 négy tizedesjegyre kerekítve, ami nagyobb $\alpha = 0{,}01$-nál, így elfogadjuk a nullhipotézist.

7.3. Példa. A `minta-27.txt` fájlban található mintarealizációról döntse el 0,99 szinten, hogy származhat-e standard normális eloszlásból.

Megoldás. Legyen az a nullhipotézis, hogy standard normális eloszlásból származik a mintarealizáció. A mintarealizáció elemeit másolja az `A` oszlopba. Az osztópontok legyenek $a_1 = -2$, $a_2 = -1$, $a_3 = 0$, $a_4 = 1$, $a_5 = 2$. A `B1:B5` cellatartományba írja be ezeket a számokat, majd a `C1` cellába, hogy

`=NORM.S.ELOSZLÁS(B1;IGAZ)`,

a `C2`-be pedig

`=NORM.S.ELOSZLÁS(B2;IGAZ)-NORM.S.ELOSZLÁS(B1;IGAZ)`.

A `C2` kitöltőjelét húzza le az 5. sorig. Végül a `C6` cellába írja be, hogy `=1-SZUM(C1:C5)`. Ezzel a `C` oszlopban megjelennek a p_i értékek. Mivel $p_1 = p_6 = 0{,}0228$ négy tizedesjegyre kerekítve, ezért az első és utolsó intervallumok további felbontására nincs szükség.

A `D1` cellába írja be, hogy `=DARABTELI(A:A;"<"&B1)`, a `D2`-be pedig

`=DARABHATÖBB(A:A;">="&B1;A:A;"<"&B2)`.

A `D2` kitöltőjelét húzza le az 5. sorig. A `D6` cellába írja ezt: `=DARABTELI(A:A;">="&B5)`. Ezzel megkapja a ϱ_i gyakoriságokat, melyek mindegyike nagyobb 10-nél, így alkalmazható a próba. Az `E` oszlopban számolja ki a $\nu_i = np_i$ értékeket. Írja az `E1` cellába, hogy `=DARAB(A:A)*C1`, majd a kitöltőjelet húzza le a 6. sorig. Ekkor

$1 - F(\chi^2) =$ `KHINÉGYZET.PRÓBA(D1:D6;E1:E6)`.

A kijött érték 0,8337 négy tizedesjegyre kerekítve, ami nagyobb $\alpha = 0{,}01$-nál, így elfogadjuk a nullhipotézist.

7.3. Becsléses illeszkedésvizsgálat

Legyen ξ a vizsgált valószínűségi változó és F_ϑ eloszlásfüggvény minden $\vartheta \in \Theta \subset \mathbb{R}^v$ esetén.

$$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F_\vartheta \text{ valamely } \vartheta \in \Theta \text{ esetén}}$$

Legyen $a_1 < a_2 < \cdots < a_{r-1}$, $I_1 = (-\infty, a_1)$, $I_2 = [a_1, a_2)$, $I_3 = [a_2, a_3)$, $\ldots$, $I_{r-1} = [a_{r-2}, a_{r-1})$, $I_r = [a_{r-1}, \infty)$. Jelölje ϱ_i a ξ-re vonatkozó n elemű mintában az I_i intervallumba eső mintaelemek számát ($\varrho_i \geq 10\ \forall i$). Legyen $\widehat{\vartheta}$ a ϑ maximum likelihood becslése H_0 feltételezésével, továbbá $\widehat{p}_1 = F_{\widehat{\vartheta}}(a_1)$, $\widehat{p}_2 = F_{\widehat{\vartheta}}(a_2) - F_{\widehat{\vartheta}}(a_1)$, $\widehat{p}_3 = F_{\widehat{\vartheta}}(a_3) - F_{\widehat{\vartheta}}(a_2)$, $\ldots$, $\widehat{p}_{r-1} = F_{\widehat{\vartheta}}(a_{r-1}) - F_{\widehat{\vartheta}}(a_{r-2})$, $\widehat{p}_r = 1 - F_{\widehat{\vartheta}}(a_{r-1})$,

$$\nu_i = n\widehat{p}_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\mathrm{Khi}(r-1-v)].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.4. Példa. A `minta-24.txt` fájlban található mintarealizáció alapján döntse el 0,99 szinten, hogy a vizsgált valószínűségi változó lehet-e normális eloszlású.

Megoldás. Itt nem tudjuk használni a `KHINÉGYZET.PRÓBA` függvényt, mert az $r-1$ szabadsági fokkal számol $r-1-v$ helyett. Így „direkt" módon fogjuk meghatározni χ^2 és $1 - F(\chi^2)$ értékekeit.

A nullhipotézis legyen az, hogy a vizsgált valószínűségi változó normális eloszlású. Ennek $v = 2$ darab paramétere van, a várható érték és a szórás. Ezek maximum likelihood becslései H_0 esetén a mintaátlag illetve a tapasztalati szórás.

Az $a_1, a_2, \ldots, a_9$ osztópontok legyenek rendre a $6, 7, \ldots, 14$ számok, melyeket írjon be a `B1:B9` cellatartományba. A `C1` cellába írja be, hogy

```
=NORM.ELOSZLÁS(B1;ÁTLAG(A:A);SZÓR.S(A:A);IGAZ)
```

a `C2`-be pedig

```
=NORM.ELOSZLÁS(B2;ÁTLAG(A:A);SZÓR.S(A:A);IGAZ)-
 NORM.ELOSZLÁS(B1;ÁTLAG(A:A);SZÓR.S(A:A);IGAZ)
```

A `C2` kitöltőjelét húzza le a 9. sorig. Végül a `C10` cellába írja be, hogy `=1-SZUM(C1:C9)`. Ezzel a `C` oszlopban megjelennek a $\widehat{p}_i$ értékek. Mivel $\widehat{p}_1 = 0{,}0223$ és $\widehat{p}_{10} = 0{,}0231$ négy tizedesjegyre kerekítve, ezért az első és utolsó intervallumok további felbontására nincs szükség.

A `D1` cellába írja be, hogy `=DARABTELI(A:A;"<"&B1)`, a `D2`-be pedig

```
=DARABHATÖBB(A:A;">="&B1;A:A;"<"&B2)
```

.

A `D2` kitöltőjelét húzza le a 9. sorig. A `D10` cellába írja ezt: `=DARABTELI(A:A;">="&B9)`. Ezzel megkapja a ϱ_i gyakoriságokat, melyek mindegyike nagyobb 10-nél, így alkalmazható a próba.

Az `E` oszlopban számoljuk ki a $\frac{(\varrho_i - \nu_i)^2}{\nu_i}$ értékeket. Ehhez az `E1` cellába írja be, hogy

$\boxed{\texttt{=(D1-DARAB(A:A)*C1)\^{}2/(DARAB(A:A)*C1)}}$,

majd a kitöltőjelét húzza le a 10. sorig. Végül kiszámoljuk az $1 - F(\chi^2)$ értékét, melyben $r - 1 - v = 10 - 1 - 2 = 7$ a szabadsági fok:

$\boxed{\texttt{=1-KHINÉGYZET.ELOSZLÁS(SZUM(E1:E10);7;IGAZ)}}$.

A kapott érték 0,6064 négy tizedesjegyre kerekítve, ami nagyobb $\alpha = 0{,}01$-nál, így elfogadjuk a nullhipotézist. Tehát a mintarealizáció normális eloszlásból származik.

7.4. Függetlenségvizsgálat eseményrendszerekre

$A_1, \ldots, A_r$ és $B_1, \ldots, B_s$ két teljes eseményrendszer. A nullhipotézisben azt feltételezzük, hogy a két eseményrendszer független egymástól, azaz

$$\boxed{H_0\colon\ \mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\,\mathrm{P}(B_j)\ \forall i, j},$$

ahol P a valódi valószínűség. Végezzünk n darab kísérletet. Legyen ϱ_{ij} az $A_i \cap B_j$ gyakorisága ($\varrho_{ij} \geq 10$), k_i az A_i gyakorisága, l_j az B_j gyakorisága,

$$\nu_{ij} = \frac{k_i l_j}{n}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F \simeq F[\mathrm{Khi}((r-1)(s-1))].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.5. Példa. A következő táblázat 200 ember haj- és szemszínét tartalmazza:

	szőke haj	barna haj	fekete haj
kék szem	42	28	3
barna szem	17	89	21

Ebből a mintából döntse el 0,99 szinten, hogy független-e az embereknél a hajnak és a szemnek a színe, vagy van valamilyen genetikai kapcsolat a kettő között.

Megoldás. A nullhipotézis az lesz, hogy független a szem- és hajszín. Mivel a ϱ_{ij} gyakoriságok nagyobbak 10-nél, ezért alkalmazható a próba. Az előző táblázat értékeit gépelje be az `A1:C2` cellatartományba. A `D1` cellába írja be, hogy $\boxed{\texttt{=SZUM(A1:C1)}}$, majd a kitöltőjelet húzza le a 2. sorig. Ezután az `A3` cellába írja be, hogy $\boxed{\texttt{=SZUM(A1:A2)}}$, majd a kitöltőjelet húzza jobbra a `D` oszlopig. A `D1` és `D2` tartalmazza a k_1 és k_2 értékeket, az `A3`, `B3`, `C3` pedig az l_1, l_2, l_3 értékeket, végül `D3`-ban van az n értéke.

Ennek alapján elkészítjük a $\nu_{ij} = \frac{1}{n} k_i l_j$ táblázatát. Az `F1` cellába írja a következőt:

$$\boxed{\texttt{=\$D1*A\$3/\$D\$3}}.$$

A kitöltőjelet húzza jobbra a `H` oszlopig, majd le a 2. sorig. Tehát `A1:C2` tartalmazza a ϱ_{ij}, `F1:H2` pedig a ν_{ij} értékeket. Ekkor

$$1 - F(\chi^2) = \boxed{\texttt{KHINÉGYZET.PRÓBA(A1:C2;F1:H2)}},$$

mely kb. $2 \cdot 10^{-10}$, ami kisebb $\alpha = 0{,}01$-nál. Így elutasítjuk a nullhipotézist, azaz a szem- és hajszín között van genetikai kapcsolat.

7.5. Függetlenségvizsgálat két valószínűségi változóra

A vizsgált valószínűségi változók ξ és η.

$$\boxed{H_0 \colon \xi \text{ és } \eta \text{ függetlenek}}$$

A (ξ, η)-ra vonatkozó minta $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$.

Legyen $a_0 < a_1 < \cdots < a_r$, tegyük fel, hogy $\xi_1, \ldots, \xi_n$ minden eleme benne van az $[a_0, a_r)$ intervallumban. Jelölje k_i a $\xi_1, \ldots, \xi_n$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát.

Legyen $b_0 < b_1 < \cdots < b_s$, tegyük fel, hogy $\eta_1, \ldots, \eta_n$ minden eleme benne van a $[b_0, b_s)$ intervallumban. Jelölje l_j az $\eta_1, \ldots, \eta_n$ mintában a $[b_{j-1}, b_j)$ intervallumba eső elemek számát.

Jelölje ϱ_{ij} a $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ mintában az $[a_{i-1}, a_i) \times [b_{j-1}, b_j)$ tartományba eső elemek számát ($\varrho_{ij} \geq 10$),

$$\nu_{ij} = \frac{k_i l_j}{n}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F \simeq F[\mathrm{Khi}((r-1)(s-1))].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.6. Példa. A `minta-25.txt` fájlban található (ξ, η)-ra vonatkozó mintarealizáció alapján döntse el 0,99 szinten, hogy ξ és η függetlenek-e.

Megoldás. Legyen az a nullhipotézis, hogy ξ és η független. Az `A` oszlopba kerül a ξ-re vonatkozó mintarealizáció, míg a `B` oszlopba kerül az η-ra vonatkozó mintarealizáció.

Az A és B oszlopok minimumának és maximumának vizsgálatával kiderül, hogy minden mintaelem pozitív és kisebb 4-nél. Legyenek $a_0, \ldots, a_5$ rendre 0, 0,13, 0,3, 0,5 és 4. Ezeket írja a `D2:D6` cellákba. Legyenek $b_0, \ldots, b_6$ rendre 0, 0,1, 0,2, 0,3, 0,5, 1 és 4. Ezeket írja a `E1:K1` cellákba.

Számolja ki a ϱ_{ij} gyakoriságokat. Az `E2` cellába gépelje a következőt:

`=DARABHATÖBB($A:$A;">="&$D2;$A:$A;"<"&$D3;$B:$B;">="&E$1;$B:$B;"<"&F$1)`.

A kitöltőjelet húzza jobbra a `J` oszlopig, majd lefelé az 5. sorig. Látjuk, hogy minden gyakoriság nagyobb 10-nél, ezért alkalmazhatjuk a próbát. Ha ez nem teljesülne, akkor az osztópontokon kellene változtatni.

A `K2`-be írja be, hogy `=SZUM(E2:J2)`, majd a kitöltőjelet húzza le az 5. sorig. Ezzel megkapjuk a k_i értékeket. Ezután az `E6` cellába gépelje be, hogy `=SZUM(E2:E5)`, majd a kitöltőjelet húzza jobbra a `K` oszlopig. Az `E6:J6` cellatartományban lesznek az l_j értékek, a `K6`-ban pedig n.

Most következhet a $\nu_{ij} = \frac{1}{n} k_i l_j$ táblázata. Az `E8` cellába gépelje be, hogy

`=$K2*E$6/$K$6`,

a kitöltőjelet húzza jobbra a `J` oszlopig, majd lefelé az 11. sorig. Tehát `E2:J5` tartalmazza a ϱ_{ij}, `E8:J11` pedig a ν_{ij} értékeket. Ekkor

$$1 - F(\chi^2) = \boxed{\texttt{KHINÉGYZET.PRÓBA(E2:J5;E8:J11)}},$$

amely négy tizedesjegyre kerekítve 0,3675, ami nagyobb $\alpha = 0{,}01$-nál. Ebből következően elfogadjuk a nullhipotézist, azaz ξ és η függetlenek.

7.6. Homogenitásvizsgálat

Legyenek a ξ és η független valószínűségi változókra vonatkozó minták $\xi_1, \ldots, \xi_{n_1}$ illetve $\eta_1, \ldots, \eta_{n_2}$.

$$\boxed{H_0\colon \xi \text{ és } \eta \text{ azonos eloszlású}}$$

Legyen $a_0 < a_1 < \cdots < a_r$, tegyük fel, hogy mindkét minta minden eleme benne van az $[a_0, a_r)$ intervallumban. Jelölje ϱ_{i1} a $\xi_1, \ldots, \xi_{n_1}$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát ($\varrho_{i1} \geq 10$), illetve ϱ_{i2} az $\eta_1, \ldots, \eta_{n_2}$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát ($\varrho_{i2} \geq 10$), továbbá

$$\nu_{ij} = \frac{(\varrho_{i1} + \varrho_{i2}) n_j}{n_1 + n_2}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{2} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F = F[\mathrm{Khi}(r-1)].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

7.7. Példa. A minta-26.txt illetve minta-27.txt fájlban található mintarealizációkról döntse el 0,99 szinten, hogy származhatnak-e azonos eloszlásból.

Megoldás. A nullhipotézis jelentse azt, hogy a két mintarealizáció azonos eloszlásból származik. A `minta-26`-ot illetve `minta-27`-et másolja az `A` illetve `B` oszlopba.

Mindkét mintarealizáció benne van a $[-50, 1200)$ intervallumban. Legyenek az $a_0, \ldots, a_7$ értékek rendre $-50, -2, -1, -0{,}5, 0{,}5, 1, 2, 1200$. Ezeket írja a `C1:C8` cellákba. A `D1` cellába gépelje a következőt:

`=DARABHATÖBB(A:A;">="&$C1;A:A;"<"&$C2)`.

A kitöltőjelet húzza jobbra eggyel, majd le a 7. sorig. A `D1:E7` tartalmazza a ϱ_{ij} gyakoriságokat, melyek nagyobbak 10-nél, ezért a próba alkalmazható.

Az `F1` cellába írja be, hogy `=D1+E1`, majd a kitöltőjelet húzza le a 7. sorig. A `D8` cellába írja be, hogy `=SZUM(D1:D7)`, majd a kitöltőjelet húzza jobbra az `F` oszlopig.

Az `F1:F7` tartalmazza a $\varrho_{i1} + \varrho_{i2}$, míg `D8:E8` az n_j értékeket és `F8`-ban van $n_1 + n_2$. Ezekkel elkészítjük a $\nu_{ij} = \frac{(\varrho_{i1}+\varrho_{i2})n_j}{n_1+n_2}$ táblázatát. A `G1` cellába írja be, hogy

`=$F1*D$8/$F$8`,

a kitöltőjelet húzza jobbra eggyel, majd le a 7. sorig. Ekkor `D1:E7` tartalmazza a ϱ_{ij}, míg `G1:H7` a ν_{ij} értékeket. Így

$$1 - F(\chi^2) = \boxed{\texttt{KHINÉGYZET.PRÓBA(D1:E7;G1:H7)}},$$

ami kb. $4 \cdot 10^{-31}$. Ez kisebb $\alpha = 0{,}01$-nál, melyből következően elutasítjuk a nullhipotézist, azaz a két mintarealizáció különböző eloszlásból származik.

7.7. Kétmintás előjelpróba

(ξ, η)-ra vonatkozó minta $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$.

$H_0\colon\ \mathrm{P}(\xi > \eta) = \frac{1}{2}$	kritikus tartomány
$H_1\colon\ \mathrm{P}(\xi > \eta) \neq \frac{1}{2}$	$B < F^{-1}(\frac{\alpha}{2})$ vagy $B > F^{-1}(1 - \frac{\alpha}{2})$
$H_1\colon\ \mathrm{P}(\xi > \eta) < \frac{1}{2}$	$B < F^{-1}(\alpha)$
$H_1\colon\ \mathrm{P}(\xi > \eta) > \frac{1}{2}$	$B > F^{-1}(1 - \alpha)$

ahol B azon (ξ_i, η_i) mintaelemek száma, melyekre $\xi_i - \eta_i$ pozitív, továbbá

$$F^{-1}(x) = \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \left(\frac{1}{2}\right)^n \geq x \right\}.$$

7.8. Példa. Migrénes fejfájásra kifejlesztenek egy új fájdalomcsillapítót. Tesztelésnél 50 páciensből 35-nél bizonyult az új gyógyszer tartósabb hatásúnak, mint a régi gyógyszere. Ennek alapján döntsön 0,99 szinten arról, hogy jobb-e az új gyógyszer.

Megoldás. Legyen ξ illetve η egy adott páciensnél az új illetve a régi gyógyszer hatásának az ideje. Vizsgáljuk a

$$H_0\colon\ \mathrm{P}(\xi > \eta) = \tfrac{1}{2}$$
$$H_1\colon\ \mathrm{P}(\xi > \eta) > \tfrac{1}{2}$$

hipotéziseket. A feladat szerint $n = 50$, $B = 35$ és $1 - \alpha = 0{,}99$. Így a már korábban megismert `BINOM.INVERZ` függvénnyel

$$F^{-1}(1-\alpha) = \boxed{\texttt{BINOM.INVERZ(50;1/2;0,99)}},$$

melynek 33 az értéke. Ettől B nagyobb, ezért a H_1 ellenhipotézist fogadjuk el, miszerint az emberek több mint felénél az új gyógyszer huzamosabb ideig hat, mint a régi.

7.8. Kolmogorov – Szmirnov-féle kétmintás próba

ξ és η folytonos eloszlásfüggvényű független valószínűségi változók, az ezekre vonatkozó minták $\xi_1, \dots, \xi_n$ illetve $\eta_1, \dots, \eta_n$ $(n > 30)$.

$$\boxed{H_0\colon \xi \text{ és } \eta \text{ azonos eloszlású}}$$

ξ-re illetve η-ra vonatkozó mintákhoz tartozó tapasztalati eloszlásfüggvények F_n^* illetve G_n^*,

$$D = \sqrt{\frac{n}{2}} \sup_{x\in\mathbb{R}} |F_n^*(x) - G_n^*(x)| \quad \text{és} \quad K(z) = 1 + 2\sum_{i=1}^{\infty} (-1)^i e^{-2i^2z^2}.$$

Kritikus tartomány: $K(D) \geq 1 - \alpha$.

A D kiszámolásához az F monoton növekedése miatt csak a két tapasztalati eloszlásfüggvény szakadási pontjaiban kell megvizsgálni a differenciákat. Így kapjuk, hogy

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - G_n^*(x)| = \max_{i=1,\dots,n} \max\Big\{|F_n^*(\xi_i) - G_n^*(\xi_i)|, |F_n^*(\eta_i) - G_n^*(\eta_i)|\Big\}.$$

Számolásnál még azt is vegyük figyelembe, hogy K határeloszlást jelent, így $\mathbb{R}_+$-on monoton növekvő.

7.9. Példa. Számolja ki adott $z \geq 1$ esetén $K(z)$ értékét.

Megoldás. A `B1` cellába írja be, hogy `z=`, majd a `C1` cellába egy konkrét z értéket, például most legyen 1. A `C1` cellát nevezze el `z`-nek. Ezután az A oszlopba számolja ki a $(-1)^i e^{-2i^2z^2}$ értékeket, ahol i a sor száma. Egy cella sorának a számát a `SOR` függvénnyel, míg az exponenciális függvényt a `KITEVŐ` függvénnyel kapja meg. Tehát az `A1` cellába írja a következőt:

`=(-1)^SOR(A1)*KITEVŐ(-2*SOR(A1)^2*z^2)`.

Az `A1` cella kitöltőjelét húzza le a 19. sorig. Az `A19` cella értéke már 0 lesz, pontosabban olyan kicsi szám, amit az Excel már nem tud ábrázolni. Mivel $e^{-2i^2z^2}$ monoton csökkenő i-ben, ezért biztos, hogy $i \geq 19$ esetén az Excel mindig 0-t írna ki. Ezért a szummázásban ezek a tagok már nem jelentenek számottevő értéket.

Most számolja ki $K(z)$ értékét. A `B2` cellába írja be, hogy `K(z)=`, majd a `C2` cellába, hogy `=1+2*SZUM(A:A)`. Mivel most a z értékéhez 1 van írva, ezért a $K(1)$-et kapjuk meg, ami négy tizedesjegyre kerekítve 0,7300.

Mivel $e^{-2i^2z^2}$ monoton csökkenő z-ben is, ezért z növelésével a szummázásban számottevő tagok száma nem nőhet. Így az `A` oszlopban tetszőleges $z \geq 1$ esetben sem kell újabb szumma tagokat számolni.

7.10. Példa. A `minta-25.txt` fájlban található ξ-re (első oszlopban) illetve η-ra (második oszlopban) vonatkozó mintarealizációk alapján döntse el 0,99 szinten, hogy ξ és η azonos eloszlású-e, ha tudjuk, hogy mindkét valószínűségi változónak folytonos az eloszlásfüggvénye.

Megoldás. Legyen az a nullhipotézis, hogy ξ és η azonos eloszlású. A ξ-re illetve η-ra vonatkozó mintarealizációt másolja az A illetve B oszlopba.

A `DARAB` függvénnyel ellenőrizheti, hogy a közös mintaelemszám $n = 500 > 30$. Tehát alkalmazhatjuk a Kolmogorov – Szmirnov-féle kétmintás próbát.

Az `A` oszlopot nevezze el `xi`-nek, a `B` oszlopot pedig `eta`-nak. A `C1` cellába írja be, hogy

`=ABS(DARABTELI(xi;"<"&A1)/500-DARABTELI(eta;"<"&A1)/500)`.

Ez $|F_n^*(\xi_1) - G_n^*(\xi_1)|$ értéke. A kitöltőjelet húzza jobbra eggyel, majd a kitöltőjelre kattintson kétszer. Ezzel $|F_n^*(\xi_i) - G_n^*(\xi_i)|$ értékeit kapta meg a `C` oszlopban, míg $|F_n^*(\eta_i) - G_n^*(\eta_i)|$ értékeit a `D` oszlopban. Ezért

$$D=\sqrt{\frac{n}{2}}\max_{i=1,\dots,n}\max\Big\{|F_n^*(\xi_i)-G_n^*(\xi_i)|,|F_n^*(\eta_i)-G_n^*(\eta_i)|\Big\}=\boxed{\texttt{GYÖK(500/2)*MAX(C:D)}}.$$

Ennek értéke 0,7906 négy tizedesjegyre kerekítve. Tehát K monoton növekedéséből és az előző feladat megoldásából kapjuk, hogy $K(D)\le K(1)\simeq 0{,}73<0{,}99=1-\alpha$. Így tehát elfogadjuk a nullhipotézist, azaz ξ és η azonos eloszlású.

7.9. Kolmogorov – Szmirnov-féle egymintás próba

Legyen ξ folytonos eloszlásfüggvényű valószínűségi változó, az erre vonatkozó minta $\xi_1,\dots,\xi_n$ $(n>30)$.

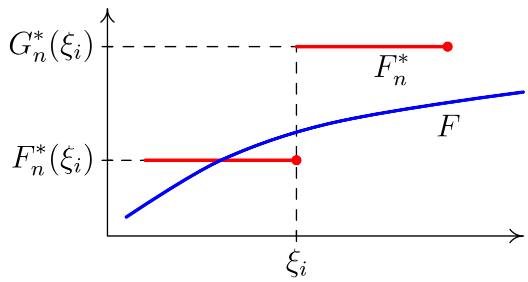
$$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F}$$

F_n^* a tapasztalati eloszlásfüggvény,

$$D=\sqrt{n}\sup_{x\in\mathbb{R}}|F_n^*(x)-F(x)| \quad \text{és} \quad K(z)=1+2\sum_{i=1}^{\infty}(-1)^i e^{-2i^2z^2}.$$

Kritikus tartomány: $K(D)\ge 1-\alpha$.

A D kiszámolásához vegyük figyelembe, hogy most egy lépcsős és egy folytonos függvény értékeinek abszolút eltérését vizsgáljuk. Így nem elég csak a lépcsők jobb végpontjaiban megnézni ezeket az értékeket, ezt meg kell tenni a bal végpontokban is. Máshol viszont nem kell, mert F monoton növekedő.



Egy lépcsőfok bal végpontjának magasságát a ξ_i mintaelem helyén megkapjuk a

$$G_n^*(\xi_i)=\frac{k_i}{n}=\lim_{x\to\xi_i+0}F_n^*(x)$$

képlettel, ahol k_i azon mintaelemek száma, melyek nem nagyobbak ξ_i-nél. Így könnyen ellenőrizhető, hogy

$$\sup_{x\in\mathbb{R}}|F_n^*(x)-F(x)|=\max_{i=1,\dots,n}\max\Big\{|F_n^*(\xi_i)-F(\xi_i)|,|G_n^*(\xi_i)-F(\xi_i)|\Big\}.$$

7.11. Példa. Tudjuk, hogy a `minta-27.txt` fájlban található mintarealizáció egy folytonos eloszlásfüggvényű valószínűségi változóra vonatkozik. Döntse el 0,99 szinten, hogy származhat-e standard Cauchy-eloszlásból.

Megoldás. Legyen az a nullhipotézis, hogy a mintarealizáció standard Cauchy-eloszlásból származik. Másoljuk a mintarealizációt az `A` oszlopba. A `DARAB` függvénnyel meggyőződhet róla, hogy $n = 730 > 30$, tehát a Kolmogorov – Szmirnov-féle egymintás próba alkalmazható. Tekintve, hogy a standard Cauchy-eloszlás eloszlásfüggvénye $F(x) = \frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2}$, így a `B1` cellába írja be, hogy

`=ABS(DARABTELI(A:A;"<"&A1)/730-ARCTAN(A1)/PI()-1/2)`,

míg a `C1` cellába, hogy

`=ABS(DARABTELI(A:A;"<="&A1)/730-ARCTAN(A1)/PI()-1/2)`.

Jelölje ki a `B1:C1` cellatartományt, és kattintson kétszer a kitöltőjelre. Ezzel a `B` oszlopban megkapta az $|F_n^*(\xi_i) - F(\xi_i)|$, míg a `C` oszlopban a $|G_n^*(\xi_i) - F(\xi_i)|$ értékeit. Így D `=GYÖK(730)*MAX(B:C)`, módon számolható, amelynek 3,4344 az értéke négy tizedesjegyre kerekítve. A $K(z)$ értékeire vonatkozó feladat megoldásából ellenőrizheti, hogy $K(3{,}4344) \simeq 1 > 1 - \alpha = 0{,}99$. Így elutasítjuk a nullhipotézist.

7.10. Gyakorlatok

7.1. gyakorlat. Egy genetikai törvény szerint, ha az egyik szülő A, a másik B vércsoportú, akkor a gyerekeik A, AB vagy B vércsoportú lehet 1 : 2 : 1 arányban. 300 ilyen vizsgált gyerek 30%-a volt A, 45%-a AB és a többi B vércsoportú. Alátámasztják ezek az adatok ezt a genetikai törvényt 0,99 szinten?

Útmutatás. A feladatot tiszta illeszkedésvizsgálattal oldja meg. Legyen az a nullhipotézis, hogy az adatok alátámasztják a genetikai törvényt. Ekkor $1 - F(\chi^2) \simeq$ $\simeq 0{,}1054 > 0{,}01 = \alpha$, azaz a nullhipotézist elfogadjuk.

7.2. gyakorlat. A `minta-22.txt` fájlban található mintarealizációról egy korábbi példa kapcsán azt állítottuk, hogy az exponenciális eloszlású. Igazoljuk ezt az állítást becsléses illeszkedésvizsgálattal 0,99 szinten.

Útmutatás. Legyen az a nullhipotézis, hogy exponenciális eloszlásból származik a mintarealizáció. Az osztópontok legyenek rendre 2, 4, 6, 10. A λ paraméter maximum likelihood becslése $1/\overline{\xi}$. A szabadsági fok $5 - 1 - 1 = 3$ lesz. Mindezek figyelembevételével kapjuk, hogy $1 - F(\chi^2) \simeq 0{,}8597 > 0{,}01 = \alpha$, azaz a nullhipotézist elfogadjuk.

7.3. gyakorlat. Televízióban az „A” márkájú fogkrémet hetente 1 órát, a „B” márkájú fogkrémet 25 percet illetve a „C” márkájú fogkrémet egyáltalán nem reklámozzák. Arra vagyunk kíváncsiak, hogy hatással vannak-e a fogkrémfogyasztási szokásokra a reklámok. Ennek érdekében megkérdeztek 610 embert arról, hogy a három márka közül melyiket használja, és hogy hetente hány órát tölt tévénézéssel. A kapott eredményeket a következő táblázat tartalmazza.

	„A”	„B”	„C”
5 óránál kevesebb	80	64	60
5–15 óra között	75	70	60
15 óra felett	90	65	46

Ebből a mintából döntsön 0,99 szinten a feltett kérdésre vonatkozóan.

Útmutatás. Végezzen függetlenségvizsgálatot az adott kontingencia táblázat alapján. Ekkor $1 - F(\chi^2) \simeq 0{,}3958 > 0{,}01 = \alpha$, azaz elfogadhatjuk a nullhipotézist, miszerint a vizsgált szempontok függetlenek egymástól. Tehát ezen adatok alapján a fogkrémfogyasztási szokásokra nincsenek hatással a reklámok.

7.4. gyakorlat. A `minta-28.txt` fájlban található (ξ, η)-ra vonatkozó mintarealizáció alapján döntse el 0,99 szinten, hogy ξ és η függetlenek-e.

Útmutatás. Végezzen függetlenségvizsgálatot. Legyen pl. $a_0 = -19,\ a_1 = -1,\ a_2 = 0,\ a_3 = 2,\ a_4 = 21$ és $b_0 = -19,\ b_1 = -1,\ b_2 = 1,\ b_3 = 2,\ b_4 = 21$. Ekkor $1 - F(\chi^2) < \alpha$, melyből következően ξ és η nem függetlenek.

7.5. gyakorlat. Két cinkelt kockával dobunk. Az egyikre illetve másikra vonatkozó minta a `minta-29.txt` illetve `minta-30.txt` fájlokban található. Döntse el 0,99 szinten, hogy azonosan vannak-e „cinkelve” a kockák.

Útmutatás. Végezzen homogenitásvizsgálatot. Legyen $a_0 = 1,\ a_1 = 2,\ a_2 = 3,\ a_3 = 4,\ a_4 = 5,\ a_5 = 6,\ c_a = 7$. Ekkor $1 - F(\chi^2) \simeq 0{,}2322 > 0{,}01 = \alpha$, melyből következően a két mintarealizáció azonos eloszlásból származik, azaz a két kocka azonos módon van „cinkelve”.

7.6. gyakorlat. Egy sportszergyár a legújabb gerelyt teszteli. 22 gerelyhajító dobott a régivel és az újjal is, akik közül 15 dobott nagyobbat az újjal. Döntse el 0,99 szinten, hogy $\frac{1}{2}$-nél nagyobb valószínűséggel jobb-e az új gerely a réginél.

Útmutatás. Végezzen kétmintás előjelpróbát.

7.7. gyakorlat. A `minta-26.txt` illetve `minta-27.txt` fájlokban található folytonos eloszlású valószínűségi változókra vonatkozó mintarealizációkról homogenitásvizsgálattal már korábban megállapítottuk, hogy nem azonos eloszlásból származnak. Mutassa ki ugyanezt Kolmogorov – Szmirnov-féle kétmintás próbával is.

Útmutatás. A `minta-27`-ben több elem található, ezért először a végéből töröljön annyit, hogy a mintaelemek száma megegyezzen. A D statisztika értékére négy tizedesjegyre kerekítve 2,2326-ot kapunk. Felhasználva a K függvény értékeire korábban gyártott Excel-táblázatot, $K(2{,}2326) \simeq 0{,}9999 > 0{,}99 = 1 - \alpha$ adódik, azaz a két eloszlás valóban nem egyezik meg.

7.8. gyakorlat. A `minta-27.txt` fájlban található mintarealizációról korábban tiszta illeszkedésvizsgálattal beláttuk, hogy standard normális eloszlásból származik. Mutassa ki ugyanezt Kolmogorov – Szmirnov-féle egymintás próbával is.

Útmutatás. A D statisztika értékére négy tizedesjegyre kerekítve 0,7618-at kapunk, továbbá $K(0{,}7618) \simeq 0{,}3927 < 0{,}99 = 1 - \alpha$ adódik, így elfogadjuk a nullhipotézist, azaz a minta standard normális eloszlásból származik.

8. fejezet

Szórásanalízis

A szórásanalízis is hipotézisvizsgálat, itt három típusával foglalkozunk.

8.1. Egyszeres osztályozás

Vizsgáljuk a ξ valószínűségi változót egyetlen tényező r darab különböző szintjén. Az i. szinthez tartozó valószínűségi változót jelölje ξ_i. Feltesszük, hogy $\xi_i \in$ $\in \mathrm{Norm}(m_i; \sigma)$ $(i = 1, 2, \ldots, r)$ függetlenek, ahol az $m_1, m_2, \ldots, m_r, \sigma$ paraméterek ismeretlenek. A ξ_i-re $(i = 1, 2, \ldots, r)$ vonatkozó minta legyen

$$\xi_{i1}, \xi_{i2}, \ldots, \xi_{in_i}.$$

A nullhipotézis az lesz, hogy az egyetlen tényező különböző szintjei nincsenek hatással a mért értékekre, azaz

$$H_0 \colon m_1 = m_2 = \cdots = m_r.$$

Vezessük be a következő jelöléseket:

$$\begin{aligned}
n &:= n_1 + n_2 + \cdots + n_r, \\
\overline{\xi}_{..} &:= \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \xi_{ij}, \\
\overline{\xi}_{i.} &:= \frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{ij}, \\
Q_1 &:= \sum_{i=1}^{r} n_i \left(\overline{\xi}_{i.} - \overline{\xi}_{..} \right)^2, \\
Q_2 &:= \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(\xi_{ij} - \overline{\xi}_{i.} \right)^2,
\end{aligned}$$

$$\mathsf{F} := \frac{n-r}{r-1} \cdot \frac{Q_1}{Q_2},$$
$$F = F[\mathrm{F}(r-1; n-r)].$$

Ekkor $1 - F(\mathsf{F}) \geq \alpha$ elfogadási tartománnyal α terjedelmű próbát kapunk H_0-ra.

8.1. Példa. Egy gazdaságban a búza terméshozamára vagyunk kíváncsiak. Azt vizsgáljuk, hogy egyetlen tényező, a búza fajtája milyen hatással van a terméshozamra. A gazdaság $r = 3$ különböző búzafajtát termeszt, azaz a vizsgált tényezőnek 3 különböző szintje van. Az 1. fajtát $n_1 = 4$, a 2. fajtát $n_2 = 3$, végül a 3. fajtát $n_3 = 5$ különböző parcellán termesztik. A ξ_{ij} jelentse az i. fajta j. parcellán kapott terméshozamát tonna/hektár-ban mérve. A kapott mintarealizációk a következők:

$$\xi_{11}(\omega) = 5{,}24 \quad \xi_{12}(\omega) = 4{,}17 \quad \xi_{13}(\omega) = 4{,}35 \quad \xi_{14}(\omega) = 4{,}77$$
$$\xi_{21}(\omega) = 5{,}09 \quad \xi_{22}(\omega) = 6{,}05 \quad \xi_{23}(\omega) = 5{,}89$$
$$\xi_{31}(\omega) = 4{,}18 \quad \xi_{32}(\omega) = 4{,}10 \quad \xi_{33}(\omega) = 4{,}17 \quad \xi_{34}(\omega) = 3{,}98 \quad \xi_{35}(\omega) = 3{,}60$$

Döntsön $1 - \alpha = 0{,}95$ szinten arról, hogy a különböző búzafajták nincsenek hatással a terméshozamra.

Megoldás. Gépelje be az adatokat a következő ábra szerint:

	A	B	C	D	E
1	5,24	4,17	4,35	4,77	
2	5,09	6,05	5,89		
3	4,18	4,1	4,17	3,98	3,6

A feladatot az *Adatok/Adatelemzés* menüponttal fogjuk megoldani. A legördülő listában válassza az *Egytényezős varianciaanalízis* sort, majd *OK*.

Bemeneti tartomány: `$A$1:$E$3`
Csoportosítási alap: Sorok
☐ *Feliratok az első oszlopban*
Alfa: `0,05`
Kimeneti tartomány: `$A$4`
OK

Ekkor a következő táblázatot kapjuk:

Tényezők	*SS*	*df*	*MS*	*F*	*p-érték*	*F krit.*
Csoportok között	5,2334	2	2,6167	16,3286	0,00101	4,256495
Csoporton belül	1,4423	9	0,1603			

A táblázatban $Q_1 = 5{,}2334$, $Q_2 = 1{,}4423$, $\mathsf{F} = 16{,}3286$ és $1 - F(\mathsf{F}) = 0{,}0010$. Mivel $1 - F(\mathsf{F}) = 0{,}0010 < 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elutasítjuk azt a hipotézist, hogy a búza fajtája nincs hatással a terméshozamra.

8.2. Kétszeres osztályozás interakció nélkül

Vizsgáljuk két tényező hatását egy ξ valószínűségi változóra. Legyen az 1. tényezőnek r_1, míg a 2. tényezőnek r_2 különböző szintje. Jelölje ξ_{ij} az 1. tényező i. szintjéhez és a 2. tényező j. szintjéhez tartozó valószínűségi változót. Feltesszük, hogy $\xi_{ij} \in$ $\in \text{Norm}(m_{ij}; \sigma)$ $(i = 1, 2, \ldots, r_1;\ j = 1, 2, \ldots, r_2)$ függetlenek, ahol minden paraméter ismeretlen.

Két nullhipotézist fogunk vizsgálni. A $H_0^{(1)}$ szerint az 1. tényező különböző szintjei nincsenek hatással ξ-re. A $H_0^{(2)}$ szerint a 2. tényező különböző szintjei nincsenek hatással ξ-re.

Vezessük be a következő jelöléseket:

$$\begin{aligned}
\overline{\xi}_{..} &:= \frac{1}{r_1 r_2} \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \xi_{ij}, \\
\overline{\xi}_{i.} &:= \frac{1}{r_2} \sum_{j=1}^{r_2} \xi_{ij} \quad (i = 1, 2, \ldots, r_1), \\
\overline{\xi}_{.j} &:= \frac{1}{r_1} \sum_{i=1}^{r_1} \xi_{ij} \quad (j = 1, 2, \ldots, r_2), \\
Q_1 &:= r_2 \sum_{i=1}^{r_1} \left(\overline{\xi}_{i.} - \overline{\xi}_{..} \right)^2, \\
Q_2 &:= r_1 \sum_{j=1}^{r_2} \left(\overline{\xi}_{.j} - \overline{\xi}_{..} \right)^2, \\
Q_3 &:= \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \left(\xi_{ij} - \overline{\xi}_{i.} - \overline{\xi}_{.j} + \overline{\xi}_{..} \right)^2, \\
\mathsf{F}_1 &:= (r_2 - 1) \cdot \frac{Q_1}{Q_3}, \\
\mathsf{F}_2 &:= (r_1 - 1) \cdot \frac{Q_2}{Q_3}, \\
F_1 &:= F\Big[\mathrm{F}\Big(r_1 - 1; (r_1 - 1)(r_2 - 1)\Big)\Big], \\
F_2 &:= F\Big[\mathrm{F}\Big(r_2 - 1; (r_1 - 1)(r_2 - 1)\Big)\Big].
\end{aligned}$$

Ekkor $1 - F_1(\mathsf{F}_1) \geq \alpha$ illetve $1 - F_2(\mathsf{F}_2) \geq \alpha$ elfogadási tartományokkal α terjedelmű próbát kapunk $H_0^{(1)}$ illetve $H_0^{(2)}$ esetén.

8.2. Példa. A 8.1. példát tovább gondolva, tegyük fel, hogy nem csak a búza fajtáját, hanem a parcella talajtípusát is vizsgálni szeretnénk a terméshozamot illetően, vagyis nem egy, hanem két tényező hatását figyeljük. Tegyük fel, hogy $r_1 = 3$ fajta búzát $r_2 = 4$ típusú talajba vetették. Azaz az 1. tényezőnek 3, a 2. tényezőnek pedig 4 szintje van. A ξ_{ij} jelentse az i. búzafajta j. talajtípuson vett terméshozamát. A kapott mintarealizációk a következők:

$$
\begin{array}{llll}
\xi_{11}(\omega) = 7{,}51 & \xi_{12}(\omega) = 6{,}34 & \xi_{13}(\omega) = 5{,}07 & \xi_{14}(\omega) = 6{,}17 \\
\xi_{21}(\omega) = 5{,}43 & \xi_{22}(\omega) = 4{,}81 & \xi_{23}(\omega) = 3{,}42 & \xi_{14}(\omega) = 4{,}00 \\
\xi_{31}(\omega) = 5{,}76 & \xi_{32}(\omega) = 4{,}71 & \xi_{33}(\omega) = 4{,}45 & \xi_{34}(\omega) = 4{,}33
\end{array}
$$

Döntsünk 0,95 szinten a következő nullhipotézisekről:

$H_0^{(1)}$: a különböző búzafajták nincsenek hatással a terméshozamra,
$H_0^{(2)}$: a különböző talajtípusok nincsenek hatással a terméshozamra.

Megoldás. Gépelje be az adatokat a következő ábra szerint:

	A	B	C	D
1	7,51	6,34	5,07	6,17
2	5,43	4,81	3,42	4
3	5,76	4,71	4,45	4,33

A feladatot az *Adatok/Adatelemzés* menüponttal fogjuk megoldani. A legördülő listában válassza az *Kéttényezős varianciaanalízis ismétlések nélkül* sort, majd *OK*.

Bemeneti tartomány: `$A$1:$D$3`
☐ *Feliratok*
Alfa: `0,05`
Kimeneti tartomány: `$A$4`
OK

Ekkor a következő táblázatot kapjuk:

Tényezők	*SS*	*df*	*MS*	*F*	*p-érték*	*F krit.*
Sorok	7,6532	2	3,8266	35,9278	0,0005	5,1433
Oszlopok	5,9744	3	1,9915	18,6978	0,0019	4,7571
Hiba	0,6391	6	0,1065			

A táblázatban $Q_1 = 7{,}6532$, $Q_2 = 5{,}9744$, $Q_3 = 0{,}6391$, $\mathsf{F}_1 = 35{,}9278$, $\mathsf{F}_2 = 18{,}6978$, $1 - F_1(\mathsf{F}_1) = 0{,}0005$ és $1 - F_2(\mathsf{F}_2) = 0{,}0019$.

Mivel $1 - F_1(\mathsf{F}_1) = 0{,}0005 < 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elutasítjuk a $H_0^{(1)}$ hipotézist, azaz a különböző búzafajták hatással vannak a terméshozamra.

Másrészt $1 - F_2(\mathsf{F}_2) = 0{,}0019 < 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elutasítjuk a $H_0^{(2)}$ hipotézist, azaz a különböző talajtípusok hatással vannak a terméshozamra.

Fontos, hogy ebben az esetben azt nem tudjuk vizsgálni, hogy a két tényező milyen hatással van egymásra, azaz, hogy egy konkrét búzafajta különbözőképpen terem-e a különböző talajtípusokon, vagy hogy egy konkrét talajtípuson különbözőképpen teremnek-e a különböző búzafajták, mert minden búzafajta–talajtípus kombinációból csak egy mintaelemünk van.

8.3. Kétszeres osztályozás interakcióval

Vizsgáljuk két tényező hatását egy ξ valószínűségi változóra. Legyen az 1. tényezőnek r_1, míg a 2. tényezőnek r_2 különböző szintje. Jelölje ξ_{ij} az 1. tényező i. szintjéhez és a 2. tényező j. szintjéhez tartozó valószínűségi változót. Feltesszük, hogy $\xi_{ij} \in$ $\in \mathrm{Norm}(m_{ij}; \sigma)$ $(i = 1, 2, \ldots, r_1;\ j = 1, 2, \ldots, r_2)$ függetlenek, ahol minden paraméter ismeretlen. Minden ξ_{ij}-hez készítsünk egy s elemű mintát:

$$\xi_{ij1}, \xi_{ij2}, \ldots, \xi_{ijs}.$$

Három nullhipotézist fogunk vizsgálni:

$H_0^{(1)}$: az 1. tényező különböző szintjei nincsenek hatással ξ-re,
$H_0^{(2)}$: a 2. tényező különböző szintjei nincsenek hatással ξ-re,
$H_0^{(3)}$: a két tényező együttes hatása nem befolyásolja a ξ értékét.

Vezessük be a következő jelöléseket:

$$\begin{aligned}
\overline{\xi}_{...} &:= \frac{1}{r_1 r_2 s} \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sum_{k=1}^{s} \xi_{ijk}, \\
\overline{\xi}_{i..} &:= \frac{1}{r_2 s} \sum_{j=1}^{r_2} \sum_{k=1}^{s} \xi_{ijk} \quad (i = 1, 2, \ldots, r_1), \\
\overline{\xi}_{.j.} &:= \frac{1}{r_1 s} \sum_{i=1}^{r_1} \sum_{k=1}^{s} \xi_{ijk} \quad (j = 1, 2, \ldots, r_2), \\
\overline{\xi}_{ij.} &:= \frac{1}{s} \sum_{k=1}^{s} \xi_{ijk} \quad (i = 1, 2, \ldots, r_1;\ j = 1, 2, \ldots, r_2), \\
Q_1 &:= r_2 s \sum_{i=1}^{r_1} \left(\overline{\xi}_{i..} - \overline{\xi}_{...} \right)^2, \\
Q_2 &:= r_1 s \sum_{j=1}^{r_2} \left(\overline{\xi}_{.j.} - \overline{\xi}_{...} \right)^2,
\end{aligned}$$

$$
\begin{aligned}
Q_3 &:= s\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}\left(\overline{\xi}_{ij.}-\overline{\xi}_{i..}-\overline{\xi}_{.j.}+\overline{\xi}_{...}\right)^2,\\
Q_4 &:= \sum_{i=1}^{r_1}\sum_{j=1}^{r_2}\sum_{k=1}^{s}\left(\xi_{ijk}-\overline{\xi}_{ij.}\right)^2,\\
\mathsf{F}_1 &:= \frac{r_1r_2(s-1)}{r_1-1}\cdot\frac{Q_1}{Q_4},\\
\mathsf{F}_2 &:= \frac{r_1r_2(s-1)}{r_2-1}\cdot\frac{Q_2}{Q_4},\\
\mathsf{F}_3 &:= \frac{r_1r_2(s-1)}{(r_1-1)(r_2-1)}\cdot\frac{Q_3}{Q_4},\\
F_1 &:= F\Big[\mathrm{F}\Big(r_1-1;r_1r_2(s-1)\Big)\Big],\\
F_2 &:= F\Big[\mathrm{F}\Big(r_2-1;r_1r_2(s-1)\Big)\Big],\\
F_3 &:= F\Big[\mathrm{F}\Big((r_1-1)(r_2-1);r_1r_2(s-1)\Big)\Big].
\end{aligned}
$$

Ekkor $1-F_1(\mathsf{F}_1)\geq\alpha$, $1-F_2(\mathsf{F}_2)\geq\alpha$ illetve $1-F_3(\mathsf{F}_3)\geq\alpha$ elfogadási tartományokkal α terjedelmű próbát kapunk rendre $H_0^{(1)}$, $H_0^{(2)}$ és $H_0^{(3)}$ esetén.

8.3. Példa. Folytatva a 8.2. példát, ha a két tényező közötti kapcsolatot is vizsgálni szeretnénk, akkor minden búzafajta–talajtípus kombinációból több mérést kell végeznünk. Legyen az előző példában minden kombinációra $s=3$ mérésünk. Jelentse ξ_{ijk} az i. búzafajta j. talajtípuson vett terméshozamára vonatkozó, azaz az (i,j) *cellában* végzett k. mérési eredményt. A kapott mintarealizációk a következők:

$$
\begin{array}{llll}
\xi_{111}(\omega)=7{,}51 & \xi_{121}(\omega)=6{,}34 & \xi_{131}(\omega)=5{,}07 & \xi_{141}(\omega)=6{,}17\\
\xi_{112}(\omega)=7{,}03 & \xi_{122}(\omega)=5{,}81 & \xi_{132}(\omega)=4{,}19 & \xi_{142}(\omega)=5{,}90\\
\xi_{113}(\omega)=6{,}91 & \xi_{123}(\omega)=6{,}61 & \xi_{133}(\omega)=5{,}27 & \xi_{143}(\omega)=6{,}28\\[1ex]
\xi_{211}(\omega)=5{,}43 & \xi_{221}(\omega)=4{,}81 & \xi_{231}(\omega)=3{,}42 & \xi_{141}(\omega)=4{,}00\\
\xi_{212}(\omega)=4{,}95 & \xi_{222}(\omega)=3{,}82 & \xi_{232}(\omega)=3{,}19 & \xi_{142}(\omega)=3{,}80\\
\xi_{213}(\omega)=5{,}48 & \xi_{223}(\omega)=4{,}18 & \xi_{233}(\omega)=2{,}02 & \xi_{143}(\omega)=3{,}94\\[1ex]
\xi_{311}(\omega)=5{,}76 & \xi_{321}(\omega)=4{,}71 & \xi_{331}(\omega)=4{,}45 & \xi_{341}(\omega)=4{,}33\\
\xi_{312}(\omega)=5{,}90 & \xi_{322}(\omega)=5{,}24 & \xi_{332}(\omega)=4{,}65 & \xi_{342}(\omega)=5{,}41\\
\xi_{313}(\omega)=6{,}01 & \xi_{323}(\omega)=4{,}07 & \xi_{333}(\omega)=4{,}59 & \xi_{343}(\omega)=5{,}70
\end{array}
$$

Döntsön a következő nullhipotézisekről:

$H_0^{(1)}$: a különböző búzafajták nincsenek hatással a terméshozamra,
$H_0^{(2)}$: a különböző talajtípusok nincsenek hatással a terméshozamra,
$H_0^{(3)}$: a búzafajta és a talajtípus között nincs kapcsolat a terméshozamot illetően.

Megoldás. Gépelje be az adatokat a következő ábra szerint:

	A	B	C	D	E
1					
2		7,51	6,34	5,07	6,17
3		7,03	5,81	4,19	5,9
4		6,91	6,61	5,27	6,28
5		5,43	4,81	3,42	4
6		4,95	3,82	3,19	3,8
7		5,48	4,18	2,02	3,94
8		5,76	4,71	4,45	4,33
9		5,9	5,24	4,65	5,41
10		6,01	4,07	4,59	5,7

Fontos, hogy az első sorba illetve oszlopba ne vigyünk adatokat. Oda csak feliratok kerülhetnek.

A feladatot az *Adatok/Adatelemzés* menüponttal fogjuk megoldani. A legördülő listában válassza az *Kéttényezős varianciaanalízis ismétlésekkel* sort, majd *OK*.

Bemeneti tartomány: `$A$1:$E$10`
Mintánként hány sor: `3`
Alfa: `0,05`
Kimeneti tartomány: `$A$11`
OK

Ekkor a következő táblázatot kapjuk:

Tényezők	*SS*	*df*	*MS*	*F*	*p-érték*	*F krit.*
Minta	24,1034	2	12,0517	59,3583	5E-10	3,4028
Oszlopok	18,2751	3	6,0917	30,0035	3E-08	3,0088
Kölcsönhatás	2,0206	6	0,3368	1,6587	0,1746	2,5082
Belül	4,8728	24	0,2030			

A táblázatban $Q_1 = 24{,}1034$, $Q_2 = 18{,}2751$, $Q_3 = 2{,}0206$, $Q_4 = 4{,}8728$, $\mathsf{F}_1 = 59{,}3583$, $\mathsf{F}_2 = 30{,}0035$, $\mathsf{F}_3 = 1{,}6587$, $1 - F_1(\mathsf{F}_1) = 5 \cdot 10^{-10}$, $1 - F_2(\mathsf{F}_2) = 3 \cdot 10^{-8}$ és $1 - F_3(\mathsf{F}_3) = 0{,}1746$.

Mivel $1 - F_1(\mathsf{F}_1) = 5 \cdot 10^{-10} < 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elutasítjuk a $H_0^{(1)}$ hipotézist, azaz a különböző búzafajták hatással vannak a terméshozamra.

Másrészt $1 - F_2(\mathsf{F}_2) = 3 \cdot 10^{-8} < 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elutasítjuk a $H_0^{(2)}$ hipotézist, azaz a különböző talajtípusok hatással vannak a terméshozamra.

Harmadrészt $1 - F_3(\mathsf{F}_3) = 0{,}1746 > 0{,}05 = \alpha$, ezért $1 - \alpha = 0{,}95$ szinten elfogadjuk a $H_0^{(3)}$ hipotézist, azaz a búzafajta és a talajtípus között nincs kapcsolat a terméshozamot illetően.

9. fejezet

Regressziószámítás

Az $\eta, \xi_1, \ldots, \xi_k$ valószínűségi változók esetén adjuk meg a legjobb

$$\eta \simeq g(\xi_1, \ldots, \xi_k) \tag{9.1}$$

közelítést adó g függvényt. Ezt úgy értjük, hogy az

$$\mathrm{E}\Big(\eta - g(\xi_1, \ldots, \xi_k)\Big)^2$$

értékét kell minimalizálni. Ez az úgynevezett *legkisebb négyzetek elve*. Az így kapott g függvényt *regressziós felületnek* nevezzük. Ha g lineáris, akkor $k = 1$ illetve $k = 2$ esetén a g függvényt *elsőfajú regressziós egyenesnek* illetve *elsőfajú regressziós síknak* nevezzük. A regressziós felület továbbá $\xi_1, \ldots, \xi_k$ ismeretében η megbecsülhető (9.1) alapján.

9.1. Lineáris regresszió

Sok esetben a regressziós felület meghatározása bonyolult feladat. Ilyenkor azzal egyszerűsíthetjük a problémát, hogy $\mathrm{E}\Big(\eta - g(\xi_1, \ldots, \xi_k)\Big)^2$ minimumát csak a

$$g(x_1, \ldots, x_k) = a_0 + a_1 x_1 + \cdots + a_k x_k \quad (a_0, a_1, \ldots, a_k \in \mathbb{R})$$

alakú – azaz lineáris – függvények között keressük. Természetesen ennek csak akkor van értelme, ha a lineáris kapcsolatnak nem mond ellent valamilyen elvi vagy tapasztalati megfontolás. Ezt a típusú regressziószámítást *lineáris regressziónak* nevezzük. A feladat megoldásában szereplő $a_0, \ldots, a_k$ konstansokat a *lineáris regresszió együtthatóinak* nevezzük. Az így kapott g függvényt $k = 1$ illetve $k = 2$ esetén *másodfajú* *regressziós egyenesnek* illetve *másodfajú regressziós síknak* nevezzük.

A lineáris regresszió együtthatóinak értékét a gyakorlatban kellő információ hiányában nem tudjuk kiszámolni. Így ekkor az $(\eta, \xi_1, \ldots, \xi_k)$-ra vonatkozó minta alapján kell ezeket megbecsülni. Legyen ez a minta

$$(\eta_i, \xi_{i1}, \ldots, \xi_{ik}) \quad i = 1, \ldots, n.$$

Bevezetjük a következő jelöléseket:

$$\begin{aligned} a &:= (a_0, \ldots, a_k)^\top \\ Y &:= (\eta_1, \ldots, \eta_n)^\top \\ X &:= \begin{pmatrix} 1 & \xi_{11} & \ldots & \xi_{1k} \\ 1 & \xi_{21} & \ldots & \xi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_{n1} & \ldots & \xi_{nk} \end{pmatrix}. \end{aligned}$$

Az $\mathrm{E}(\eta - a_0 - a_1\xi_1 - \cdots - a_k\xi_k)^2$ várható értéket az

$$\frac{1}{n}\sum_{i=1}^{n}(\eta_i - a_0 - a_1\xi_{i1} - \cdots - a_k\xi_{ik})^2$$

átlaggal becsüljük, így az a becslése azon vektor, amely mellett ez az átlag minimális. Bizonyítható, hogy az a-ra vonatkozó

$$X^\top Y = X^\top X a$$

úgynevezett *normálegyenlet* $\widehat{a} = (\widehat{a}_0, \ldots, \widehat{a}_k)^\top$-val jelölt megoldása szolgáltatja a lineáris regresszió együtthatóinak becslését. Ebből, ha $X^\top X$ invertálható mátrix, akkor

$$\widehat{a} = (\widehat{a}_0, \ldots, \widehat{a}_k)^\top = (X^\top X)^{-1} X^\top Y.$$

Ezután az $\eta \simeq \widehat{a}_0 + \widehat{a}_1\xi_1 + \cdots + \widehat{a}_k\xi_k$ közelítést fogjuk használni. Speciálisan $k = 1$ esetén a lineáris regresszió együtthatóinak becslése ($\xi := \xi_1$ jelöléssel)

$$\widehat{a}_1 = \frac{\mathrm{Cov}_n(\eta, \xi)}{S^2_{\xi,n}}, \quad \widehat{a}_0 = \overline{\eta} - \widehat{a}_1\overline{\xi},$$

ahol $\mathrm{Cov}_n(\eta, \xi)$ a tapasztalati kovarianciája az (η, ξ)-re vonatkozó mintának. Ennek alapján a továbbiakban az $\eta \simeq \widehat{a}_0 + \widehat{a}_1\xi$ közelítést fogjuk használni. (Megjegyezzük, hogy a grafikus illeszkedésvizsgálatnál is ezt alkalmaztuk.) Az η és ξ közötti lineáris kapcsolat feltételezésének jogosságát az ún. *determinációs együtthatóval* szokták mérni, amely a tapasztalati korrelációs együttható négyzete, azaz

$$R^2 := \mathrm{Corr}_n(\eta, \xi).$$

Ennek értéke minél közelebb van 1-hez annál jobban feltételezhető a lineáris kapcsolat. Ha 0-hoz van közel, akkor a két valószínűségi változó között nem célszerű függvény-kapcsolatot keresni, mert feltehetőleg függetlenek (pontosabban korrelálatlanok).

9.1. Példa. Jelentse η a talajvízszintet mm-ben és ξ az őszi csapadék mennyiségét cm-ben. Az (η, ξ)-re vonatkozó elmúlt 18 évi mérésből származó mintarealizációt a `minta-31.txt` fájl tartalmazza. Ez alapján becsülje meg a lineáris regresszió együtthatóit. A becsült másodfajú regressziós egyenest ábrázolja a mintarealizációval együtt. Számolja ki a determinációs együtthatót, majd becsülje meg a talajvízszintet, ha az őszi csapadék 29,6 cm.

Megoldás. Töltse le a `minta-31.txt` fájlt, nyissa meg például Notepad programmal. *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a minta tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára és *Ctrl+V* segítségével illessze be a mintarealizációt.

Az $\widehat{a}_1, \widehat{a}_0$ együtthatók kiszámolásához jelölje ki a `C1:D1` cellatartományt, majd gépelje be a következő tömbképletet:

`=LIN.ILL(A1:A18;B1:B18)`.

Végül nyomjon *Ctrl+Shift+Enter*-t. Ennek hatására `C1` fogja $\widehat{a}_1$ értékét, illetve `D1` fogja $\widehat{a}_0$ értékét tartalmazni.

A determinációs együtthatót a következő módon számolhatja ki:

`=RNÉGYZET(A1:A18;B1:B18)`.

Ennek értéke négy tizedesjegyre kerekítve 0,7931. Ez azt jelenti, hogy a lineáris közelítés jónak mondható.

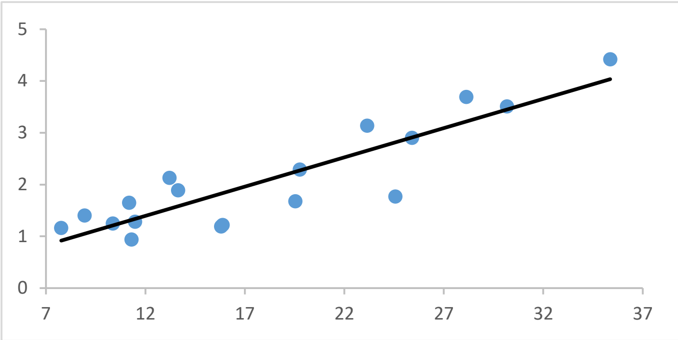
Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A1:A18`), majd

Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés → *Adatsor X értékei:* `=Munka1!$B$1:$B$18` → *OK* → *OK*

Ezzel megjelennek a mintarealizáció pontjai. Következzen a másodfajú regressziós egyenes becslésének a meghúzása (az Excel ezt *trendvonalnak* nevezi). Lépjen rá valamelyik kék jelölő pontra. Helyi menüből válassza a *Trendvonal felvétele* pontot és válassza ki a *Lineáris* típust.



A talajvízszint becslése, ha az őszi csapadék 29,6 cm, ezek alapján $\widehat{a}_0 + 29{,}6\widehat{a}_1$, amit a következő módon is kiszámolhatunk:

`=TREND(A1:A18;B1:B18;29,6)`.

A kapott érték 3,38 két tizedesjegyre kerekítve. Tehát 29,6 cm csapadék lehullása után az adatok alapján 3,38 mm-re becsüljük a talajvízszintet.

9.2. Példa. Jelentse η a Duna egy árhullámának tetőző vízállását Budapesten cm-ben, ξ_1 az árhullámot kiváltó csapadék mennyiségét mm-ben és ξ_2 a Duna vízállását Budapestnél az esőzés kezdetekor cm-ben. Az (η, ξ_1, ξ_2)-re vonatkozó elmúlt 26 évi mérésből származó mintarealizációt a `minta-32.txt` fájl tartalmazza. Ez alapján becsülje meg a lineáris regresszió együtthatóit. Az idén az árhullámot kiváltó csapadék 102 mm volt, illetve a Duna vízállása Budapestnél az esőzés kezdetekor 648 cm volt. Ezekből az adatokból becsülje meg, hogy a Duna árhullámának tetőző vízállása Budapesten hány cm lesz.

Megoldás. Nyissa meg a letöltött `minta-32.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára, és *Ctrl+V* segítségével illessze be a mintarealizációt. Az $\widehat{a}_2, \widehat{a}_1, \widehat{a}_0$ együtthatók kiszámolásához jelölje ki a `D1:F1` cellatartományt, majd gépelje be a következőt:

`=LIN.ILL(A1:A26;B1:C26)`.

Végül nyomja meg a *Ctrl+Shift+Enter* billentyűket. Ennek hatására az $\widehat{a}_2, \widehat{a}_1, \widehat{a}_0$ értékek rendre megjelennek a `D1`, `E1`, `F1` cellákban. A `D2` cellába gépelje be a 102 értéket, az `E2` cellába a 648 értéket. Ekkor a tetőző vízállás becslése a következő függvénnyel számolható ki:

$\boxed{\texttt{=TREND(A1:A26;B1:C26;D2:E2)}}$.

A kapott érték egészre kerekítve 800 cm.

9.2. Fixpontos lineáris regresszió

A lineáris regresszió feladata tovább szűkíthető, ha tudjuk, hogy a keresett lineáris függvény áthalad egy rögzített ponton.

Legyenek $t_0, \dots, t_k \in \mathbb{R}$ rögzített konstansok. Az $\mathrm{E}\Big(\eta - g(\xi_1, \dots, \xi_k)\Big)^2$ minimumát keressük azon

$$g(x_1, \dots, x_k) = a_0 + a_1 x_1 + \dots + a_k x_k \quad (a_0, a_1, \dots, a_k \in \mathbb{R})$$

függvények között, melyekre teljesül, hogy $g(t_1, \dots, t_k) = t_0$, azaz a g függvény áthalad a $(t_1, \dots, t_k, t_0)$ úgynevezett *fixponton*. Ez az úgynevezett *fixpontos lineáris regresszió*. A megoldást adó g függvényt $k = 1$ illetve $k = 2$ esetén *fixpontos regressziós egyenesnek* illetve *fixpontos regressziós síknak* nevezzük.

A fixpontos lineáris regresszió együtthatóira az $(\eta, \xi_1, \dots, \xi_k)$ valószínűségi vektorváltozóra vonatkozó $(\eta_i, \xi_{i1}, \dots, \xi_{ik})$, $i = 1, \dots, n$ minta alapján adhatunk becslést. Ha $t_0 = \dots = t_k = 0$ (azaz $y = g(x_1, \dots, x_k)$ átmegy az origón), akkor

$$Y := (\eta_1, \dots, \eta_n)^\top$$
$$X := \begin{pmatrix} \xi_{11} & \dots & \xi_{1k} \\ \xi_{21} & \dots & \xi_{2k} \\ \vdots & \ddots & \vdots \\ \xi_{n1} & \dots & \xi_{nk} \end{pmatrix}$$

jelölésekkel

$$(\widehat{a}_1, \dots, \widehat{a}_k)^\top = (X^\top X)^{-1} X^\top Y.$$

Ezután az $\eta \simeq \widehat{a}_1 \xi_1 + \dots + \widehat{a}_k \xi_k$ közelítést fogjuk használni. Tetszőleges $t_0, \dots, t_k \in \mathbb{R}$ esetén az előző becslési eljárást hajtsuk végre az $(\eta - t_0, \xi_1 - t_1, \dots, \xi_k - t_k)$-ra vonatkozó mintára. Az így kapott $\widehat{a}_1, \dots, \widehat{a}_k$ értékekkel az $\eta - t_0 \simeq \widehat{a}_1(\xi_1 - t_1) + \dots + \widehat{a}_k(\xi_k - t_k)$, azaz

$$\eta \simeq t_0 + \widehat{a}_1(\xi_1 - t_1) + \dots + \widehat{a}_k(\xi_k - t_k)$$

közelítést fogjuk használni.

9.3. Példa. Jelentse η egy vizsgált ellenálláson átfolyó áram erősségét Amperben, illetve ξ az ellenállásra adott feszültséget Voltban. Az (η, ξ)-re vonatkozó 10 mérésből származó mintarealizációt a `minta-33.txt` fájl tartalmazza. Természetesen $\xi = 0$ esetén $\eta = 0$. Ez alapján becsülje meg a fixpontos lineáris regresszió a_1 együtthatóját. A kapott egyenest ábrázolja a mintarealizációval együtt. Adjon becslést arra, hogy mekkora lesz az áramerősség 12 V ráadott feszültség esetén.

Megoldás. Nyissa meg a letöltött `minta-33.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára és *Ctrl+V* segítségével illessze be a mintarealizációt. Az $\widehat{a}_1$ kiszámolásához használja a következőt:

`=LIN.ILL(A1:A10;B1:B10;HAMIS)`.

A kapott érték 0,0254 négy tizedesjegyre kerekítve, így a továbbiakban az $\eta \simeq 0{,}0254\xi$ közelítést lehet használni. (Vagyis az ellenállás becslése $\frac{1}{0{,}0254} \simeq 39{,}4$ Ohm.)

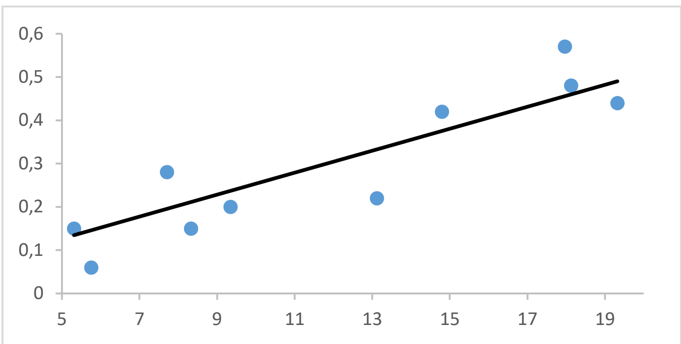
Most következik a grafikon. Jelölje ki az η-ra vonatkozó `A1:A10` mintát, majd

Beszúrás $\to$ *Diagramok* $\to$
Pont- (xy) vagy buborékdiagram beszúrása $\to$ *Pont*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés $\to$ *Adatsor X értékei:* `=Munka1!$B$1:$B$10` $\to$ *OK* $\to$ *OK*

Ezzel megjelennek a mintarealizáció pontjai. Következzen a fixpontos regressziós egyenes becslésének a meghúzása. Tudjuk, hogy $t_0 = t_1 = 0$, azaz most az origó a fixpont. Lépjen rá valamelyik kék jelölő pontra. Helyi menüből válassza a *Trendvonal felvétele* pontot. Válassza ki a *Lineáris* típust, a *Metszéspontot* pipálja ki és állítsa 0-ra (ez a t_0 értéke).



Az ábra csak akkor helyes, ha $t_1 = 0$, mert annak értékét nem lehet állítani Excelben. Az áramerősség becslése 12 V esetén

$\boxed{\texttt{=TREND(A1:A10;B1:B10;12;HAMIS)}}$

módon számolható. A kapott érték 0,30 két tizedesjegyre kerekítve. Tehát 12 V feszültség esetén az átfolyó áram erősségét 0,3 A-ra becsüljük.

9.4. Példda. Az (η, ξ_1, ξ_2)-re vonatkozó mintarealizációt a `minta-34.txt` fájl tartalmazza. Ez alapján becsülje meg a fixpontos lineáris regresszió a_1, a_2 együtthatóit $(t_1, t_2, t_0) = (0{,}4; 0{,}75; 1)$ fixpont esetén. Ebből adjon becslést η-ra, ha $\xi_1 = 1{,}3$ és $\xi_2 = 7{,}5$.

Megoldás. Nyissa meg a letöltött `minta-34.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára, és *Ctrl+V* segítségével illessze be a mintarealizációt. A $t_0 = 1$, $t_1 = 0{,}4$, $t_2 = 0{,}75$ értékeket írja be rendre a `G1`, `H1`, `I1` cellákba. A `D2` cellába írja be, hogy $\boxed{\texttt{=A1-G\$1}}$. A kitöltőjelet húzza `F2`-ig, majd a kitöltőjelre kattintson kétszer.

Az $\widehat{a}_2, \widehat{a}_1$ kiszámolásához jelölje ki a `G2:H2` cellatartományt, gépelje be a következőt:

$\boxed{\texttt{=LIN.ILL(D1:D15;E1:F15;HAMIS)}}$,

majd nyomja meg a *Ctrl+Shift+Enter* billentyűket. A kapott értékek 1,9983 és 0,3312 négy tizedesjegyre kerekítve, így a továbbiakban $\xi_1 = x_1$ és $\xi_2 = x_2$ esetén

$$\eta \simeq 1 + 0{,}3312(x_1 - 0{,}4) + 1{,}9983(x_2 - 0{,}75)$$

közelítést lehet használni. Ennek értékét kell kiszámolni $x_1 = 1{,}3$ és $x_2 = 7{,}5$ esetén. Ehhez az $x_1 - t_1 = 1{,}3 - 0{,}4$ és $x_2 - t_2 = 7{,}5 - 0{,}75$ értékeket gépelje a `G3` és `H3` cellákba, majd

$\boxed{\texttt{=G1+TREND(D1:D15;E1:F15;G3:H3;HAMIS)}}$.

A kapott érték 14,79 két tizedesjegyre kerekítve. Tehát $\eta \simeq 14{,}79$, ha $\xi_1 = 1{,}3$ és $\xi_2 = 7{,}5$.

9.3. Nemlineáris regresszió

A lineáris regressziós közelítés sok esetben nem célszerű, mert valamilyen elvi vagy tapasztalati tény a lineáris kapcsolatnak ellentmond. Ilyenkor meg kell tippelni, hogy milyen típusú függvény közelíti jobban a kapcsolatot a lineárisnál (hatvány, exponenciális, logaritmus, stb.), majd a regressziós függvény keresését le kell szűkíteni erre a csoportra. Néhány esetben valamilyen transzformációval ez a keresés visszavezethető a lineáris esetre. Most csak ilyen esetekkel foglalkozunk $k = 1$ esetén, azaz amikor az η valószínűségi változót akarjuk becsülni ξ valamilyen függvényével.

9.3.1. Polinomos regresszió

Ebben az esetben a regressziós függvényt

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_r x^r \quad (a_0, \ldots, a_r \in \mathbb{R}_+)$$

alakban keressük. Ekkor az $a_0, \ldots, a_r$ együtthatókat az $\eta, \xi, \xi^2, \ldots, \xi^r$ között végrehajtott lineáris regresszió adja.

9.5. Példa. Az (η, ξ)-re vonatkozó mintarealizációt a `minta-35.txt` fájl tartalmazza. Ez alapján becsülje meg a másodfokú polinomos regressziós függvényt. A kapott parabolát ábrázolja a mintarealizációval együtt.

Megoldás. Nyissa meg a letöltött `minta-35.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára és *Ctrl+V* segítségével illessze be a mintarealizációt. A `C1` cellába írja be, hogy `=B1^2`, majd a kitöltőjelre kattintson kétszer. Az $\widehat{a}_2, \widehat{a}_1, \widehat{a}_0$ kiszámolásához jelölje ki az `D1:F1` tartományt, gépelje be a következőt:

`=LIN.ILL(A1:A10;B1:C10)`,

majd nyomja meg a *Ctrl+Shift+Enter* billentyűket. A kapott értékek $-4{,}3773$, $15{,}6534$ és $-7{,}2032$ négy tizedesjegyre kerekítve, így a másodfokú polinomos regressziós függvény becslése:

$$y = -4{,}3773x^2 + 15{,}6534x - 7{,}2032.$$

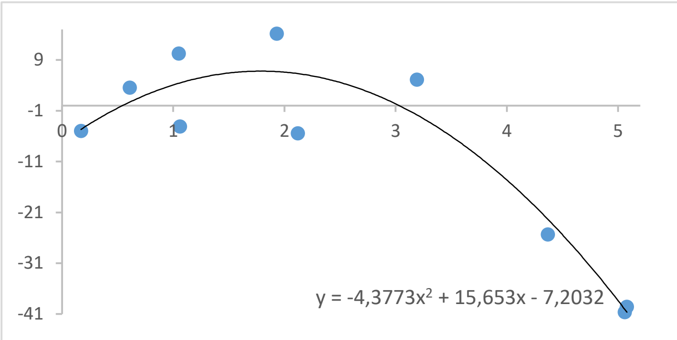
Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A1:A10`), majd

Beszúrás → *Diagramok* →
Pont- (xy) vagy buborékdiagram beszúrása → *Pont*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés → *Adatsor X értékei:* `=Munka1!$B$1:$B$10` → *OK* → *OK*

Ezzel megjelentek a mintarealizáció pontjai. Következzen a másodfokú polinomos regressziós függvény becslésének a megrajzolása. Lépjen rá valamelyik kék jelölő pontra. Helyi menüből válassza a *Trendvonal felvétele* pontot és válassza ki a *Polinomiális Fokszám 2)* típust.



9.3.2. Hatványkitevős regresszió

Ebben az esetben a regressziós függvényt

$$y = ax^b \quad (a \in \mathbb{R}_+,\ b \in \mathbb{R})$$

alakban keressük. Ez azzal ekvivalens, hogy $\ln y = \ln a + b \ln x$, így ekkor $\ln \eta$ és $\ln \xi$ között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy $a_0 = \ln a,\ a_1 = b$, azaz $a = e^{a_0},\ b = a_1$.

9.6. Példa. Az (η, ξ)-re vonatkozó mintarealizációt a `minta-36.txt` fájl tartalmazza. Ez alapján becsülje meg a hatványkitevős regressziós függvényt. A kapott függvényt ábrázolja a mintarealizációval együtt.

Megoldás. Nyissa meg a letöltött `minta-36.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára és *Ctrl+V* segítségével illessze be a mintarealizációt. A `C1` cellába írja be, hogy `=LN(A1)`, a kitöltőjelet húzza a `D1` celláig, majd a kitöltőjelre kattintson kétszer. Az $\widehat{a}_1, \widehat{a}_0$ kiszámolásához jelölje ki az `E1:F1` tartományt, gépelje be a következőt:

`=LIN.ILL(C1:C20;D1:D20)`,

majd nyomja meg a *Ctrl+Shift+Enter* billentyűket. Ekkor $\widehat{a} =$ `KITEVŐ(F1)` $= 3{,}0982$ és $\widehat{b} = \widehat{a}_1 = 3{,}4833$ négy tizedesjegyre kerekítve, így a hatványkitevős regressziós függvény becslése:

$$y = 3{,}0982x^{3{,}4833}.$$

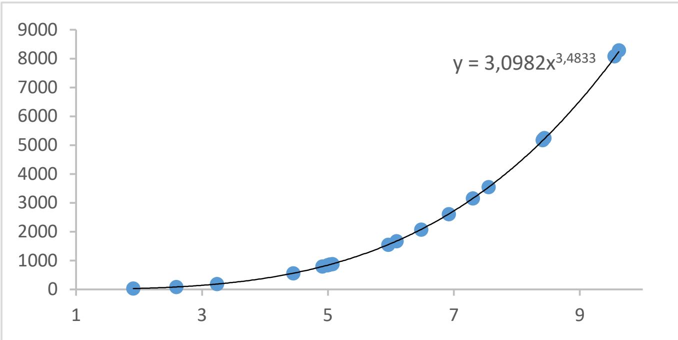
Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A1:A20`), majd

Beszúrás $\to$ *Diagramok* $\to$
Pont- (xy) vagy buborékdiagram beszúrása $\to$ *Pont*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés $\to$ *Adatsor X értékei:* `=Munka1!$B$1:$B$20` $\to$ *OK* $\to$ *OK*

Ezzel megjelentek a mintarealizáció pontjai. Következzen a hatványkitevős regressziós függvény becslésének a megrajzolása. Lépjen rá valamelyik kék jelölő pontra. Helyi menüből válassza a *Trendvonal felvétele* pontot, majd válassza ki a *Hatványos* típust.



9.3.3. Exponenciális regresszió

Ebben az esetben a regressziós függvényt

$$y = ab^x \quad (a, b \in \mathbb{R}_+)$$

alakban keressük. Ez azzal ekvivalens, hogy $\ln y = \ln a + (\ln b)x$, így ekkor $\ln \eta$ és ξ között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy $a_0 = \ln a,\ a_1 = \ln b$, azaz $a = e^{a_0},\ b = e^{a_1}$.

9.7. Példa. Az (η, ξ)-re vonatkozó mintarealizációt a `minta-37.txt` fájl tartalmazza. Ez alapján becsülje meg az exponenciális regressziós függvényt. A kapott függvényt ábrázolja a mintarealizációval együtt. Becsülje meg ebből η értékét, ha $\xi = 5$.

Megoldás. Nyissa meg a letöltött `minta-37.txt` fájlt például Notepad programmal, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A1` cellára és *Ctrl+V* segítségével illessze be a mintarealizációt. Az előző megoldás logikáját is lehet követni, de Excelben erre az esetre van külön függvény. A $\widehat{b}, \widehat{a}$ kiszámolásához jelölje ki a `C1:D1` tartományt, gépelje be a következőt:

$\boxed{\texttt{=LOG.ILL(A1:A10;B1:B10)}}$,

majd nyomja meg a *Ctrl+Shift+Enter* billentyűket. Ekkor $\widehat{b} = 3{,}0495$ és $\widehat{a} = 4{,}8127$ négy tizedesjegyre kerekítve, így az exponenciális regressziós függvény becslése:

$$y = 4{,}8127 \cdot 3{,}0495^x.$$

Ezután

$\boxed{\texttt{=NÖV(A1:A10;B1:B10;5)}}$.

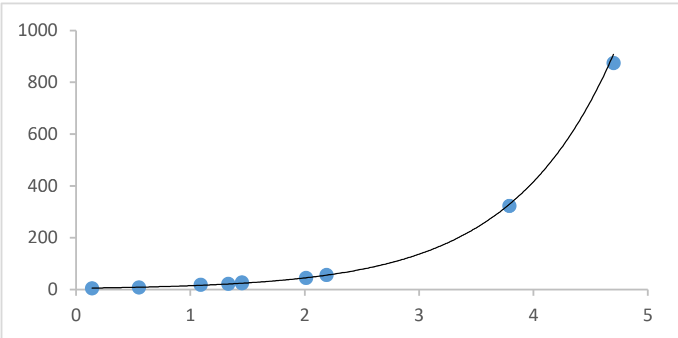
értéke 1269,14, amely az η becslése $\xi_1 = 5$ esetén. Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A1:A10`), majd

Beszúrás $\to$ *Diagramok* $\to$
Pont- (xy) vagy buborékdiagram beszúrása $\to$ *Pont*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés $\to$ *Adatsor X értékei:* `=Munka1!$B$1:$B$10` $\to$ *OK* $\to$ *OK*

Ezzel megjelentek a mintarealizáció pontjai. Következzen az exponenciális regressziós függvény becslésének a megrajzolása. Lépjen rá valamelyik kék jelölő pontra. Helyi menüből válassza a *Trendvonal felvétele* pontot, majd válassza ki a *Exponenciális* típust.



9.3.4. Logaritmikus regresszió

Ebben az esetben a regressziós függvényt

$$y = a + b \ln x \quad (a, b \in \mathbb{R})$$

alakban keressük. Így ekkor η és $\ln \xi$ között lineáris regressziót végrehajtva, $a = a_0$, $b = a_1$. (Lásd a 9.5. gyakorlatot.)

9.3.5. Hiperbolikus regresszió

Ebben az esetben a regressziós függvényt

$$y = \frac{1}{a+bx} \quad (a, b \in \mathbb{R})$$

alakban keressük. Ez azzal ekvivalens, hogy $y^{-1} = a + bx$, így ekkor η^{-1} és ξ között lineáris regressziót végrehajtva, $a = a_0,\ b = a_1$. (Lásd a 9.6. gyakorlatot.)

9.4. Gyakorlatok

9.1. gyakorlat. Az $(\eta, \xi_1, \xi_2, \xi_3)$ valószínűségi vektorváltozóra vonatkozó mintarealizációt a `minta-38.txt` fájl tartalmazza. Ez alapján becsülje meg a lineáris regresszió együtthatóit, majd ebből η értékét, ha $\xi_1 = 6{,}3,\ \xi_2 = 0{,}7,\ \xi_3 = 0{,}9$.

9.2. gyakorlat. Az (η, ξ_1)-re vonatkozó mintarealizációt a `minta-39.txt` fájl tartalmazza. Ez alapján becsülje meg a $(t_1, t_0) = (0, 3)$ fixpontos lineáris regresszió a_1 együtthatóját. A kapott egyenest ábrázolja a mintarealizációval együtt. Adjon becslést arra, hogy mekkora lesz η, ha $\xi_1 = 1{,}6$.

Útmutatás. Nézze át a fixpontos lineáris regressziónál található példákat. Az ábrázolásnál a trendvonal felvételénél a metszéspontot állítsa 3-ra.

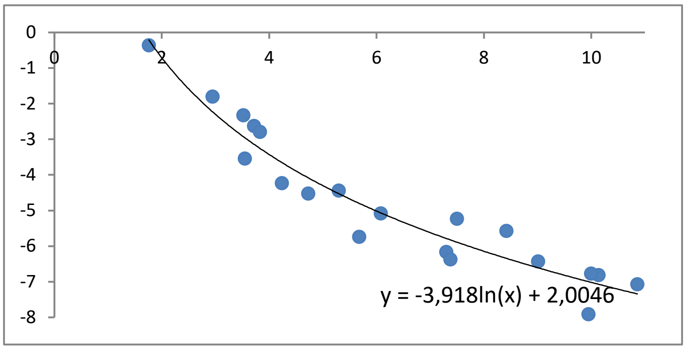
9.3. gyakorlat. Az $(\eta, \xi_1, \xi_2, \xi_3)$ valószínűségi vektorváltozóra vonatkozó mintarealizációt a `minta-38.txt` fájl tartalmazza. Ez alapján becsülje meg a $(t_1, t_2, t_3, t_0) = (1, 1, 1, 1)$ fixpontos lineáris regresszió együtthatóit, majd ebből η értékét, ha $\xi_1 = 6{,}3,\ \xi_2 = 0{,}7,\ \xi_3 = 0{,}9$.

9.4. gyakorlat. Oldja meg az exponenciális regresszióra vonatkozó példát `LOG.ILL` és `NÖV` függvények nélkül.

Útmutatás. Használja fel az exponenciális regresszió és a lineáris regresszió kapcsolatát.

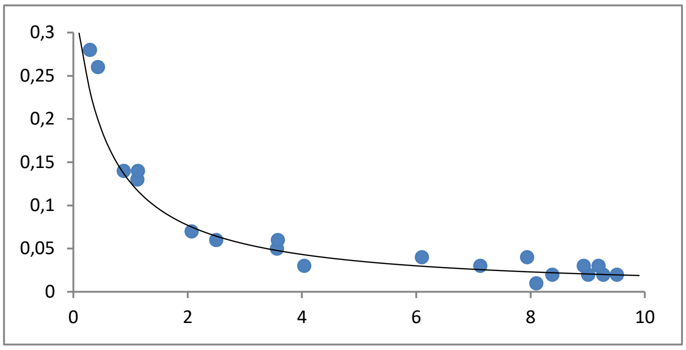
9.5. gyakorlat. Az (η, ξ)-re vonatkozó mintarealizációt a `minta-40.txt` fájl tartalmazza. Ez alapján becsülje meg a logaritmikus regressziós függvényt. A kapott függvényt ábrázolja a mintarealizációval együtt. Becsülje meg ebből η értékét, ha $\xi = 5{,}3$.

Útmutatás. Használja fel a logaritmikus regresszió és a lineáris regresszió kapcsolatát. A trendvonal felvételénél a *logaritmikus* pontot jelölje ki. Az eredményt a következő ábra mutatja.



9.6. gyakorlat. Az (η, ξ)-re vonatkozó mintarealizációt a `minta-41.txt` fájl tartalmazza. Ez alapján becsülje meg a hiperbolikus regressziós függvényt. A kapott függvényt ábrázolja a mintarealizációval együtt. Becsülje meg ebből η értékét, ha $\xi = 4{,}2$.

Útmutatás. Használja fel a hiperbolikus regresszió és a lineáris regresszió kapcsolatát. Az eredményt a következő ábra mutatja.



A becsült görbe egyenlete $y = \frac{1}{2{,}8347+5{,}0766x}$. A trendvonal ábrázolásánál vegyen fel sűrűn pontokat a görbén és folytonos vonallal húzza azokat össze, úgy, ahogy azt a tapasztalati és valódi eloszlásfüggvény egy diagramon való ábrázolásánál tettük.

10. fejezet

Összefoglaló

10.1. Eloszlások generálása

10.1.1. Egyenletes eloszlásból származtatott eloszlások

Itt az $\eta, \eta_0, \eta_1, \eta_2, \ldots$ független, a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változókat jelent.

- **Diszkrét egyenletes eloszlás**
  Ha $m \in \mathbb{N}$, akkor $[m\eta]+1$ diszkrét egyenletes eloszlású az $\{1,\ldots,m\}$ halmazon.

- **Karakterisztikus eloszlás**
  Ha $0 < p < 1$, akkor $\mathrm{I}_{\eta<p}$ karakterisztikus eloszlású p paraméterrel.

- **Binomiális eloszlás**
  Ha $r \in \mathbb{N}$ és $0 < p < 1$, akkor $\sum\limits_{i=1}^{r} \mathrm{I}_{\eta_i<p}$ r-edrendű p paraméterű binomiális eloszlású.

- **Hipergeometrikus eloszlás**
  Legyen $r, M, N \in \mathbb{N}$, $M < N$, továbbá $r \leq \min\{M, N-M\}$. Ekkor

$$\xi_0 \equiv 0, \quad \xi_i := \begin{cases} \xi_{i-1}+1, & \text{ha } \eta_i < \frac{M-\xi_{i-1}}{N-i+1}, \\ \xi_{i-1}, & \text{különben}, \end{cases} \quad (i = 1,\ldots,r)$$

  jelöléssel ξ_r hipergeometrikus eloszlású N, M, r paraméterekkel.

- **Poisson-eloszlás**
  Ha $\lambda > 0$, akkor $\min\left\{s \in \mathbb{N}\cup\{0\} : \eta_0\eta_1\cdots\eta_s < e^{-\lambda}\right\}$ Poisson-eloszlású λ paraméterrel.

- **Geometriai eloszlás**
  Ha $0 < p < 1$, akkor $\min\{s \in \mathbb{N} : \eta_s < p\}$ geometriai eloszlású p paraméterrel.

- **Folytonos egyenletes eloszlás**
  Ha $a, b \in \mathbb{R}$, $a < b$, akkor $a + (b-a)\eta$ az $[a,b]$ intervallumon egyenletes eloszlású.

- **Exponenciális eloszlás**
  Ha $\lambda > 0$, akkor $-\frac{\ln \eta}{\lambda}$ exponenciális eloszlású λ paraméterrel.

- **Gamma-eloszlás**
  Ha $\lambda > 0$ és $r \in \mathbb{N}$, akkor $-\frac{1}{\lambda}\sum_{i=1}^{r} \ln \eta_i$ r-edrendű λ paraméterű gamma-eloszlású. Tetszőleges $r, \lambda > 0$ esetén $F^{-1}(\eta)$ r-edrendű λ paraméterű gamma-eloszlású, ahol $F = F[\mathrm{Gamma}(r;\lambda)]$.

- **Normális eloszlás**
  Ha $m \in \mathbb{R}$ és $\sigma > 0$, akkor $m + \sigma\sqrt{-2\ln \eta_1}\cos(2\pi\eta_2)$ illetve $F^{-1}(\eta)$ normális eloszlású m várható értékkel és σ szórással, ahol $F = F[\mathrm{Norm}(m;\sigma)]$. (Standard normális eloszlás esetén $m = 0$, $\sigma = 1$ és $F = \Phi$.)

10.1.2. Normális eloszlásból származtatott eloszlások

Itt az η, η_i $(i \in \mathbb{N})$ független standard normális eloszlású valószínűségi változókat, míg ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változót jelent.

- **Khi-négyzet eloszlás**
  Ha $s \in \mathbb{N}$, akkor $\sum_{i=1}^{s} \eta_i^2$ illetve $F^{-1}(\xi)$ khi-négyzet eloszlású s szabadsági fokkal, ahol $F = F[\mathrm{Khi}(s)]$.

- **t-eloszlás**
  Ha $s \in \mathbb{N}$, akkor $\eta\sqrt{s/\sum_{i=1}^{s}\eta_i^2}$ illetve $F^{-1}(\xi)$ t-eloszlású s szabadsági fokkal, ahol $F = F[\mathrm{T}(s)]$.

- **Cauchy-eloszlás**
  Ha $\mu \in \mathbb{R}$ és $\sigma > 0$, akkor $\mu + \sigma\frac{\eta_1}{\eta_2}$ illetve $\mu + \sigma \,\mathrm{tg}\, \frac{\pi}{2}(2\xi - 1)$ Cauchy-eloszlású μ és σ paraméterekkel. (Standard Cauchy-eloszlás esetén $\mu = 0$ és $\sigma = 1$.)

- **F-eloszlás**
  Ha $s_1, s_2 \in \mathbb{N}$, akkor $\frac{s_2}{s_1}\sum_{i=1}^{s_1}\eta_i^2 / \sum_{i=s_1+1}^{s_1+s_2}\eta_i^2$ illetve $F^{-1}(\xi)$ F-eloszlású s_1 és s_2 szabadsági fokkal, ahol $F = F[\mathrm{F}(s_1;s_2)]$.

10.2. Grafikus illeszkedésvizsgálat

Legyen $x_1 < x_2 < \cdots < x_r$, továbbá tegyük fel, hogy a mintarealizáció legkisebb eleme nagyobb x_1-nél, a mintarealizáció legnagyobb eleme pedig kisebb x_r-nél.

- **Exponencialitásvizsgálat**
  Ha a vizsgált valószínűségi változó exponenciális eloszlású λ paraméterrel, akkor $y_i := \ln\Big(1 - F_n^*(x_i)\Big)$ jelöléssel az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $-\lambda$ a meredeksége és átmegy az origón.

- **Normalitásvizsgálat**
  Ha a vizsgált valószínűségi változó normális eloszlású m várható értékkel és σ szórással, akkor $y_i := \Phi^{-1}\Big(F_n^*(x_i)\Big)$ jelöléssel az $(x_1, y_1), \ldots, (x_r, y_r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $\frac{1}{\sigma}$ a meredeksége és $-\frac{m}{\sigma}$ értéknél metszi a függőleges tengelyt.

10.3. Intervallumbecslések

Legyen a ξ valószínűségi változóra vonatkozó minta $\xi_1, \ldots, \xi_n$, és $1 - \alpha$ a becsülendő paraméterre vonatkozó $[\tau_1, \tau_2]$ konfidenciaintervallum biztonsági szintje.

- $\xi \in \text{Norm}(m; \sigma)$
  m az ismeretlen becsülendő paraméter, σ ismert
  $\tau_1 = \overline{\xi} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \tau_2 = \overline{\xi} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \text{Norm}(m; \sigma)$
  m ismert, σ az ismeretlen becsülendő paraméter
  $F = F[\text{Khi}(n)]$
  $\tau_1 = \sqrt{\frac{1}{F^{-1}\left(1-\frac{\alpha}{2}\right)} \sum_{i=1}^{n} (\xi_i - m)^2} \quad \tau_2 = \sqrt{\frac{1}{F^{-1}\left(\frac{\alpha}{2}\right)} \sum_{i=1}^{n} (\xi_i - m)^2}$

- $\xi \in \text{Norm}(m; \sigma)$
  m ismeretlen, σ az ismeretlen becsülendő paraméter
  $n \geq 2, \ F = F[\text{Khi}(n-1)]$
  $\tau_1 = S_n\sqrt{\frac{n}{F^{-1}\left(1-\frac{\alpha}{2}\right)}} \quad \tau_2 = S_n\sqrt{\frac{n}{F^{-1}\left(\frac{\alpha}{2}\right)}}$

- $\xi \in \text{Norm}(m; \sigma)$
  m az ismeretlen becsülendő paraméter, σ ismeretlen

$n \geq 2,\ F = F[\mathrm{T}(n-1)]$
$\tau_1 = \overline{\xi} - \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \tau_2 = \overline{\xi} + \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \mathrm{Exp}(\lambda)$
  λ az ismeretlen becsülendő paraméter
  $F = F[\mathrm{Gamma}(n;1)]$
  $\tau_1 = (n\overline{\xi})^{-1} F^{-1}\left(\frac{\alpha}{2}\right) \quad \tau_2 = (n\overline{\xi})^{-1} F^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \mathrm{Bin}(1;p)$
  p az ismeretlen becsülendő paraméter
  $\tau_1 = \frac{1}{n} \max\left\{ z \in \mathbb{N} : \sum\limits_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} (1-\overline{\xi})^{n-i} < \frac{\alpha}{2} \right\}$
  $\tau_2 = \frac{1}{n} \min\left\{ z \in \mathbb{N} : \sum\limits_{i=0}^{z} \binom{n}{i} \overline{\xi}^{\,i} (1-\overline{\xi})^{n-i} \geq 1 - \frac{\alpha}{2} \right\}$
  Nagy n-re:
  $c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
  $$\tau_1 = \frac{\overline{\xi} + \frac{c^2}{2n} - \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \simeq \overline{\xi} - \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}$$
  $$\tau_2 = \frac{\overline{\xi} + \frac{c^2}{2n} + \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \simeq \overline{\xi} + \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}$$

- ξ az $[a,b]$ intervallumon egyenletes eloszlású
  a ismert, b az ismeretlen becsülendő paraméter
  $F = F[\mathrm{Gamma}(n;1)],\ c_1 = F^{-1}\left(\frac{\alpha}{2}\right),\ c_2 = F^{-1}\left(1 - \frac{\alpha}{2}\right)$
  $\tau_1 = a + \left(e^{c_1} \prod\limits_{i=1}^{n} (\xi_i - a)\right)^{\frac{1}{n}}$
  $\tau_2 = a + \left(e^{c_2} \prod\limits_{i=1}^{n} (\xi_i - a)\right)^{\frac{1}{n}}$

10.4. Paraméteres hipotézisvizsgálatok

A következőkben $1-\alpha$ a próba szintjét jelenti.

- **Egymintás u-próba**
  $\xi \in \mathrm{Norm}(m;\sigma)$, m ismeretlen, σ ismert, a ξ-re vonatkozó minta n elemű, $m_0 \in \mathbb{R}$.

$H_0\colon m = m_0$	kritikus tartomány
$H_1\colon m \neq m_0$	$2 - 2\Phi(\lvert u\rvert) < \alpha$
$H_1\colon m < m_0$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u < 0$
$H_1\colon m > m_0$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u > 0$

ahol
$$u = \frac{\overline{\xi} - m_0}{\sigma}\sqrt{n}.$$

- **Kétmintás u-próba**

  $\xi \in \mathrm{Norm}(m_1; \sigma_1)$, $\eta \in \mathrm{Norm}(m_2; \sigma_2)$ függetlenek, m_1, m_2 ismeretlenek, σ_1, σ_2 ismertek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2\Phi(\lvert u\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u < 0$
$H_1\colon m_1 > m_2$	$1 - \Phi(\lvert u\rvert) < \alpha$ és $u > 0$

ahol
$$u = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

- **Egymintás t-próba**

  $\xi \in \mathrm{Norm}(m; \sigma)$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n elemű, $m_0 \in \mathbb{R}$.

$H_0\colon m = m_0$	kritikus tartomány
$H_1\colon m \neq m_0$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m < m_0$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m > m_0$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol
$$t = \frac{\overline{\xi} - m_0}{S_n^*}\sqrt{n} \quad \text{és} \quad F = F[\mathrm{T}(n-1)].$$

- **Kétmintás t-próba**

  $\xi \in \mathrm{Norm}(m_1; \sigma_1)$, $\eta \in \mathrm{Norm}(m_2; \sigma_2)$ függetlenek, $\sigma_1 = \sigma_2$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m_1 > m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

ahol
$$t = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{n_1 S_{\xi,n_1}^2 + n_2 S_{\eta,n_2}^2}}\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \quad \text{és} \quad F = F[\mathrm{T}(n_1 + n_2 - 2)].$$

- **Scheffé-módszer**

  $\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű, $n_1 \le n_2$.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m_1 > m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

  ahol
  $$\zeta_i := \xi_i - \sqrt{\frac{n_1}{n_2}}\eta_i + \frac{1}{\sqrt{n_1 n_2}}\sum_{k=1}^{n_1}\eta_k - \overline{\eta} \quad (i = 1,\ldots,n_1)$$
  jelöléssel
  $$t = \frac{\overline{\zeta}}{S^*_{\zeta,n_1}}\sqrt{n_1} \quad \text{és} \quad F = F[\mathrm{T}(n_1-1)].$$
  Speciálisan $n_1 = n_2$ esetén $\zeta_i = \xi_i - \eta_i$ teljesül. A módszert ekkor *párosított t-próbának* is nevezik. Ebben az esetben a módszer akkor is alkalmazható, ha a minták nem függetlenek, de csak akkor, ha $\xi - \eta$ normális eloszlású.

- **Welch-próba**

  $\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon m_1 = m_2$	kritikus tartomány
$H_1\colon m_1 \neq m_2$	$2 - 2F(\lvert t\rvert) < \alpha$
$H_1\colon m_1 < m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t < 0$
$H_1\colon m_1 > m_2$	$1 - F(\lvert t\rvert) < \alpha$ és $t > 0$

  ahol
  $$t := \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\dfrac{{S^*_{\xi,n_1}}^2}{n_1} + \dfrac{{S^*_{\eta,n_2}}^2}{n_2}}} \quad \text{és} \quad F \simeq F[\mathrm{T}(s)].$$
  Az s szabadsági fok a c értékének kerekítése a legközelebbi egészre, ahol
  $$a := \frac{{S^*_{\xi,n_1}}^2}{n_1}, \quad b := \frac{{S^*_{\eta,n_2}}^2}{n_2}, \quad c := \frac{(a+b)^2}{\dfrac{a^2}{n_1-1} + \dfrac{b^2}{n_2-1}}.$$

- **F-próba**

  $\xi \in \mathrm{Norm}(m_1;\sigma_1)$, $\eta \in \mathrm{Norm}(m_2;\sigma_2)$ függetlenek, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n_1 elemű, az η-ra vonatkozó minta n_2 elemű.

$H_0\colon \sigma_1 = \sigma_2$	kritikus tartomány
$H_1\colon \sigma_1 \neq \sigma_2$	$2\min\{F(\mathsf{F}), 1 - F(\mathsf{F})\} < \alpha$
$H_1\colon \sigma_1 < \sigma_2$	$\min\{F(\mathsf{F}), 1 - F(\mathsf{F})\} < \alpha$ és $\mathsf{F} < 1$
$H_1\colon \sigma_1 > \sigma_2$	$\min\{F(\mathsf{F}), 1 - F(\mathsf{F})\} < \alpha$ és $\mathsf{F} > 1$

ahol

$$\mathsf{F} = \frac{S^{*2}_{\xi,n_1}}{S^{*2}_{\eta,n_2}} \quad \text{és} \quad F := F[\mathrm{F}(n_1 - 1; n_2 - 1)].$$

- **Khi-négyzet próba normális eloszlás szórására**

  $\xi \in \mathrm{Norm}(m;\sigma)$, a paraméterek ismeretlenek, a ξ-re vonatkozó minta n elemű, $\sigma_0 > 0$.

$H_0\colon \sigma = \sigma_0$	kritikus tartomány
$H_1\colon \sigma \neq \sigma_0$	$2\min\{F(\chi^2), 1 - F(\chi^2)\} < \alpha$
$H_1\colon \sigma < \sigma_0$	$\min\{F(\chi^2), 1 - F(\chi^2)\} < \alpha$ és $\chi^2 < n-1$
$H_1\colon \sigma > \sigma_0$	$\min\{F(\chi^2), 1 - F(\chi^2)\} < \alpha$ és $\chi^2 > n-1$

ahol

$$\chi^2 = \frac{S^{*2}_n}{\sigma_0^2}(n-1) \quad \text{és} \quad F = F[\mathrm{Khi}(n-1)].$$

- **Statisztikai próba az exponenciális eloszlás paraméterére**

  $\xi \in \mathrm{Exp}(\lambda)$, ahol λ ismeretlen, a ξ-re vonatkozó minta n elemű, $\lambda_0 > 0$.

$H_0\colon \lambda = \lambda_0$	kritikus tartomány
$H_1\colon \lambda \neq \lambda_0$	$2\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$
$H_1\colon \lambda < \lambda_0$	$\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$ és $\gamma > n$
$H_1\colon \lambda > \lambda_0$	$\min\{F(\gamma), 1 - F(\gamma)\} < \alpha$ és $\gamma < n$

ahol

$$\gamma = \lambda_0 n\overline{\xi} \quad \text{és} \quad F = F[\mathrm{Gamma}(n;1)].$$

- **Statisztikai próba valószínűségre**

  $\xi \in \mathrm{Bin}(1;p)$, p ismeretlen, a ξ-re vonatkozó minta n elemű, $0 < p_0 < 1$.

$H_0\colon p = p_0$	kritikus tartomány
$H_1\colon p \neq p_0$	$n\overline{\xi} < F^{-1}\left(\frac{\alpha}{2}\right)$ vagy $n\overline{\xi} > F^{-1}\left(1 - \frac{\alpha}{2}\right)$
$H_1\colon p < p_0$	$n\overline{\xi} < F^{-1}(\alpha)$
$H_1\colon p > p_0$	$n\overline{\xi} > F^{-1}(1-\alpha)$

ahol

$$F^{-1}(x) = \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} p_0^i (1-p_0)^{n-i} \geq x \right\}.$$

10.5. Nemparaméteres hipotézisvizsgálatok

A következőkben $1-\alpha$ a próba szintjét jelenti.

- **Tiszta illeszkedésvizsgálat valószínűségre**
  $A_1, \dots, A_r$ teljes eseményrendszer, $p_1, \dots, p_r \in \mathbb{R}_+$, $p_1 + \cdots + p_r = 1$.

  $$\boxed{H_0\colon A_i \text{ valószínűsége } p_i \ \forall i},$$

  Legyen ϱ_i az A_i gyakorisága n kísérlet után ($\varrho_i \geq 10\ \forall i$),

  $$\nu_i = np_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\mathrm{Khi}(r-1)].$$

  Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Tiszta illeszkedésvizsgálat eloszlásfüggvényre**
  Legyen ξ a vizsgált valószínűségi változó és F_0 egy eloszlásfüggvény.

  $$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F_0}$$

  Legyen $a_1 < a_2 < \cdots < a_{r-1}$, $I_1 = (-\infty, a_1)$, $I_2 = [a_1, a_2)$, $I_3 = [a_2, a_3)$, $\dots$, $I_{r-1} = [a_{r-2}, a_{r-1})$, $I_r = [a_{r-1}, \infty)$. Jelölje ϱ_i a ξ-re vonatkozó n elemű mintában az I_i intervallumba eső mintaelemek számát ($\varrho_i \geq 10\ \forall i$), továbbá legyen $p_1 = F_0(a_1)$, $p_2 = F_0(a_2) - F_0(a_1)$, $p_3 = F_0(a_3) - F_0(a_2)$, $\dots$, $p_{r-1} = F_0(a_{r-1}) - F_0(a_{r-2})$, $p_r = 1 - F_0(a_{r-1})$,

  $$\nu_i = np_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\mathrm{Khi}(r-1)].$$

  Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Becsléses illeszkedésvizsgálat**
  Legyen ξ a vizsgált valószínűségi változó és F_ϑ eloszlásfüggvény minden $\vartheta \in \Theta \subset \mathbb{R}^v$ esetén.

  $$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F_\vartheta \text{ valamely } \vartheta \in \Theta \text{ esetén}}$$

  Legyen $a_1 < a_2 < \cdots < a_{r-1}$, $I_1 = (-\infty, a_1)$, $I_2 = [a_1, a_2)$, $I_3 = [a_2, a_3)$, $\dots$, $I_{r-1} = [a_{r-2}, a_{r-1})$, $I_r = [a_{r-1}, \infty)$. Jelölje ϱ_i a ξ-re vonatkozó n elemű mintában az I_i intervallumba eső mintaelemek számát ($\varrho_i \geq 10\ \forall i$). Legyen

$\widehat{\vartheta}$ a ϑ maximum likelihood becslése H_0 feltételezésével, továbbá $\widehat{p}_1 = F_{\widehat{\vartheta}}(a_1)$, $\widehat{p}_2 = F_{\widehat{\vartheta}}(a_2) - F_{\widehat{\vartheta}}(a_1)$, $\widehat{p}_3 = F_{\widehat{\vartheta}}(a_3) - F_{\widehat{\vartheta}}(a_2)$, $\ldots$, $\widehat{p}_{r-1} = F_{\widehat{\vartheta}}(a_{r-1}) - F_{\widehat{\vartheta}}(a_{r-2})$, $\widehat{p}_r = 1 - F_{\widehat{\vartheta}}(a_{r-1})$,

$$\nu_i = n\widehat{p}_i, \quad \chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - \nu_i)^2}{\nu_i} \quad \text{és} \quad F \simeq F[\text{Khi}(r-1-v)].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Függetlenségvizsgálat eseményrendszerekre**
  $A_1, \ldots, A_r$ és $B_1, \ldots, B_s$ két teljes eseményrendszer. A nullhipotézisben azt feltételezzük, hogy a két eseményrendszer független egymástól, azaz

  $$\boxed{H_0\colon\ \mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\,\mathrm{P}(B_j)\ \forall i, j},$$

  ahol P a valódi valószínűség. Végezzünk n darab kísérletet. Legyen ϱ_{ij} az $A_i \cap B_j$ gyakorisága ($\varrho_{ij} \geq 10$), k_i az A_i gyakorisága, l_j az B_j gyakorisága,

  $$\nu_{ij} = \frac{k_i l_j}{n}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F \simeq F[\text{Khi}((r-1)(s-1))].$$

  Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Függetlenségvizsgálat két valószínűségi változóra**
  A vizsgált valószínűségi változók ξ és η.

  $$\boxed{H_0\colon \xi \text{ és } \eta \text{ függetlenek}}$$

  A (ξ, η)-ra vonatkozó minta $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$.

  Legyen $a_0 < a_1 < \cdots < a_r$, tegyük fel, hogy $\xi_1, \ldots, \xi_n$ minden eleme benne van az $[a_0, a_r)$ intervallumban. Jelölje k_i a $\xi_1, \ldots, \xi_n$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát.

  Legyen $b_0 < b_1 < \cdots < b_s$, tegyük fel, hogy $\eta_1, \ldots, \eta_n$ minden eleme benne van a $[b_0, b_s)$ intervallumban. Jelölje l_j az $\eta_1, \ldots, \eta_n$ mintában a $[b_{j-1}, b_j)$ intervallumba eső elemek számát.

  Jelölje ϱ_{ij} a $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ mintában az $[a_{i-1}, a_i) \times [b_{j-1}, b_j)$ tartományba eső elemek számát ($\varrho_{ij} \geq 10$),

  $$\nu_{ij} = \frac{k_i l_j}{n}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F \simeq F[\text{Khi}((r-1)(s-1))].$$

Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Homogenitásvizsgálat**

  Legyenek $\xi_1, \dots, \xi_{n_1}$ és $\eta_1, \dots, \eta_{n_2}$ a ξ illetve η független valószínűségi változókra vonatkozó minták.

  $$\boxed{H_0\colon \xi \text{ és } \eta \text{ azonos eloszlású}}$$

  Legyen $a_0 < a_1 < \cdots < a_r$, tegyük fel, hogy mindkét minta minden eleme benne van az $[a_0, a_r)$ intervallumban. Jelölje ϱ_{i1} a $\xi_1, \dots, \xi_{n_1}$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát ($\varrho_{i1} \geq 10$), illetve ϱ_{i2} az $\eta_1, \dots, \eta_{n_2}$ mintában az $[a_{i-1}, a_i)$ intervallumba eső elemek számát ($\varrho_{i2} \geq 10$), továbbá

  $$\nu_{ij} = \frac{(\varrho_{i1} + \varrho_{i2})n_j}{n_1 + n_2}, \quad \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{2} \frac{(\varrho_{ij} - \nu_{ij})^2}{\nu_{ij}} \quad \text{és} \quad F = F[\text{Khi}(r-1)].$$

  Kritikus tartomány: $1 - F(\chi^2) < \alpha$.

- **Kétmintás előjelpróba**

  (ξ, η)-ra vonatkozó minta $(\xi_1, \eta_1), \dots, (\xi_n, \eta_n)$.

$H_0\colon\ \mathrm{P}(\xi > \eta) = \frac{1}{2}$	kritikus tartomány
$H_1\colon\ \mathrm{P}(\xi > \eta) \neq \frac{1}{2}$	$B < F^{-1}(\frac{\alpha}{2})$ vagy $B > F^{-1}(1 - \frac{\alpha}{2})$
$H_1\colon\ \mathrm{P}(\xi > \eta) < \frac{1}{2}$	$B < F^{-1}(\alpha)$
$H_1\colon\ \mathrm{P}(\xi > \eta) > \frac{1}{2}$	$B > F^{-1}(1 - \alpha)$

  ahol B azon (ξ_i, η_i) mintaelemek száma, melyekre $\xi_i - \eta_i$ pozitív, továbbá

  $$F^{-1}(x) = \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \left(\frac{1}{2}\right)^n \geq x \right\}.$$

- **Kolmogorov – Szmirnov-féle kétmintás próba**

  ξ és η folytonos eloszlásfüggvényű független valószínűségi változók, az ezekre vonatkozó minták $\xi_1, \dots, \xi_n$ illetve $\eta_1, \dots, \eta_n$ ($n > 30$).

  $$\boxed{H_0\colon \xi \text{ és } \eta \text{ azonos eloszlású}}$$

  ξ-re illetve η-ra vonatkozó mintákhoz tartozó tapasztalati eloszlásfüggvények F_n^* illetve G_n^*,

  $$D = \sqrt{\frac{n}{2}} \max_{i=1,\dots,n} \max\Big\{ |F_n^*(\xi_i) - G_n^*(\xi_i)|, |F_n^*(\eta_i) - G_n^*(\eta_i)| \Big\},$$

$$K(z) = 1 + 2\sum_{i=1}^{\infty}(-1)^i e^{-2i^2z^2}.$$

Kritikus tartomány: $K(D) \geq 1 - \alpha$.

- **Kolmogorov – Szmirnov-féle egymintás próba**

  ξ folytonos eloszlásfüggvényű valószínűségi változó, az erre vonatkozó minta $\xi_1, \dots, \xi_n$ ($n > 30$).

  $$\boxed{H_0\colon \xi \text{ eloszlásfüggvénye } F}$$

  $$D = \sqrt{n} \max_{i=1,\dots,n} \max\Big\{|F_n^*(\xi_i) - F(\xi_i)|, |G_n^*(\xi_i) - F(\xi_i)|\Big\},$$

  ahol F_n^* a tapasztalati eloszlásfüggvény, k_i azon mintaelemek száma, melyek nem nagyobbak ξ_i-nél és $G_n^*(\xi_i) = \frac{k_i}{n}$.

  $$K(z) = 1 + 2\sum_{i=1}^{\infty}(-1)^i e^{-2i^2z^2}.$$

  Kritikus tartomány: $K(D) \geq 1 - \alpha$.

10.6. Regressziószámítás

Az $\eta, \xi_1, \dots, \xi_k$ valószínűségi változókra adjuk meg azt az $\eta \simeq g(\xi_1, \dots, \xi_k)$ közelítést adó g függvényt, melyre $\mathrm{E}\Big(\eta - g(\xi_1, \dots, \xi_k)\Big)^2$ minimális. Az ilyen tulajdonságú g függvényt (*regressziós függvény*) a gyakorlatban csak becsülni tudjuk az $(\eta, \xi_1, \dots, \xi_k)$ valószínűségi vektorváltozóra vonatkozó

$$(\eta_i, \xi_{i1}, \dots, \xi_{ik}), \quad i = 1, \dots, n$$

minta alapján. Legyen ez a becslés $\widehat{g}$. Ezután az $\boxed{\eta \simeq \widehat{g}(\xi_1, \dots, \xi_k)}$ közelítést fogjuk használni.

- **Lineáris regresszió**

  A regressziós függvényt csak a

  $$\boxed{g(x_1, \dots, x_k) = a_0 + a_1x_1 + \dots + a_kx_k \quad (a_0, \dots, a_k \in \mathbb{R})}$$

  alakú függvények között keressük. Ekkor az

  $$\boxed{\eta \simeq \widehat{a}_0 + \widehat{a}_1\xi_1 + \dots + \widehat{a}_k\xi_k}$$

közelítést fogjuk használni, ahol $\widehat{a}_0, \dots, \widehat{a}_k$ rendre $a_0, \dots, a_k$ becslései.

- **Fixpontos lineáris regresszió**
  Legyenek $t_0, \dots, t_k \in \mathbb{R}$ rögzített konstansok. A regressziós függvényt

$$\boxed{g(x_1, \dots, x_k) = t_0 + a_1(x_1 - t_1) + \cdots + a_k(x_k - t_k) \quad (a_1, \dots, a_k \in \mathbb{R})}$$

  alakban keressük. Ekkor az

$$\boxed{\eta \simeq t_0 + \widehat{a}_1(\xi_1 - t_1) + \cdots + \widehat{a}_k(\xi_k - t_k)}$$

  közelítést fogjuk használni, ahol $\widehat{a}_1, \dots, \widehat{a}_k$ rendre $a_1, \dots, a_k$ becslései.

- **Polinomos regresszió**
  $k = 1$ és a regressziós függvényt

$$\boxed{y = a_0 + a_1 x + a_2 x^2 + \cdots + a_r x^r \quad (a_0, \dots, a_r \in \mathbb{R}_+)}$$

  alakban keressük. Az $a_0, \dots, a_r$ együtthatókat az $\boxed{\eta, \xi_1, \xi_1^2, \dots, \xi_1^r}$ között végrehajtott lineáris regresszió adja.

- **Hatványkitevős regresszió**
  $k = 1$ és a regressziós függvényt

$$\boxed{y = ax^b \quad (a \in \mathbb{R}_+,\ b \in \mathbb{R})}$$

  alakban keressük. Ez azzal ekvivalens, hogy

$$\ln y = \ln a + b \ln x,$$

  így ekkor $\boxed{\ln \eta \text{ és } \ln \xi_1}$ között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy

$$\boxed{a = e^{a_0}, \quad b = a_1}.$$

- **Exponenciális regresszió**
  $k = 1$ és a regressziós függvényt

$$\boxed{y = ab^x \quad (a, b \in \mathbb{R}_+)}$$

alakban keressük. Ez azzal ekvivalens, hogy

$$\ln y = \ln a + (\ln b)x,$$

így ekkor $\boxed{\ln \eta \text{ és } \xi_1}$ között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy

$$\boxed{a = e^{a_0}, \quad b = e^{a_1}}.$$

- **Logaritmikus regresszió**
  $k = 1$ és a regressziós függvényt

  $$\boxed{y = a + b \ln x \quad (a, b \in \mathbb{R})}$$

  alakban keressük. Így ekkor $\boxed{\eta \text{ és } \ln \xi_1}$ között lineáris regressziót végrehajtva,

  $$\boxed{a = a_0, \ b = a_1}.$$

- **Hiperbolikus regresszió**
  $k = 1$ és a regressziós függvényt

  $$\boxed{y = \frac{1}{a + bx} \quad (a, b \in \mathbb{R})}$$

  alakban keressük. Ez azzal ekvivalens, hogy

  $$y^{-1} = a + bx,$$

  így ekkor $\boxed{\eta^{-1} \text{ és } \xi_1}$ között lineáris regressziót végrehajtva,

  $$\boxed{a = a_0, \ b = a_1}.$$

10.7. Excel függvények

10.7.1. Analysis ToolPak aktiválása

Az *Adatok/Adatelemzés* menüpont használatához aktiválja az *Analysis ToolPak* bővítményt: *Fájl/Beállítások/Bővítmények* majd *Ugrás* gomb. Pipálja ki az *Analysis ToolPak* sort majd *OK*.

10.7.2. Képlet bevitele

Minden képletet `=` jellel kell kezdeni. Ha a képlet egyértékű eredményt ad, akkor nyomjon *Enter*-t.

10.7.3. Tömbképlet bevitele

Ha a képlet eredménye tömb (például egy mátrix inverze), akkor először jelölje ki a megfelelő méretű tömböt, gépelje be a képletet (előtte `=`), majd nyomjon *Ctrl+Shift+Enter*-t.

10.7.4. Tömbképlet javítása

Ha egy tömbképletet javítani akar, akkor jelölje ki a tömbképletre vonatkozó tömböt, *F2*, javítás, majd *Ctrl+Shift+Enter*.

10.7.5. Műveletek

`+` összeadás

`-` kivonás

`*` szorzás

`/` osztás

`^` hatványozás

10.7.6. Relációk

`=` egyenlő

`<` kisebb

`>` nagyobb

`<=` kisebb vagy egyenlő

`>=` nagyobb vagy egyenlő

`<>` nem egyenlő

10.7.7. Konstansok

$e =$ `KITEVŐ(1)`

$\pi =$ `PI()`

10.7.8. Logikai függvények

`HA(feltétel;ha igaz;ha hamis)`

`ÉS(feltétel1;feltétel2;...)`

`VAGY(feltétel1;feltétel2;...)`

10.7.9. Elemi függvények

$|x| =$ `ABS(x)` $x \in \mathbb{R}$
$[x] =$ `INT(x)` $x \in \mathbb{R}$
$\operatorname{sign} x =$ `ELŐJEL(x)` $x \in \mathbb{R}$
$\ln x =$ `LN(x)` $x > 0$
$\log_a x =$ `LOG(x;a)` $x > 0,\ a > 0,\ a \neq 1$
$\sqrt{x} =$ `GYÖK(x)` $x \geq 0$
$x^a =$ `HATVÁNY(x;a)` $=$ `x^a`
$e^x =$ `KITEVŐ(x)` $x \in \mathbb{R}$
$\sin x =$ `SIN(x)` $x \in \mathbb{R}$
$\cos x =$ `COS(x)` $x \in \mathbb{R}$
$\operatorname{tg} x =$ `TAN(x)` $x \in \mathbb{R},\ x \neq k\frac{\pi}{2}$, ahol k páratlan egész
$\arcsin x =$ `ARCSIN(x)` $x \in [-1, 1]$
$\arccos x =$ `ARCCOS(x)` $x \in [-1, 1]$
$\operatorname{arctg} x =$ `ARCTAN(x)` $x \in \mathbb{R}$
$\Gamma(x) = \int\limits_0^\infty u^{x-1}e^{-u}\,\mathrm{d}u =$ `GAMMA(x))` $x > 0$

10.7.10. Mátrixok

`MDETERM(tömb)` A `tömb`-ben található $n \times n$ típusú mátrix determinánsa

`TRANSZPONÁLÁS(tömb)` A `tömb`-ben található $m \times n$ típusú mátrix transzponáltja, mely egy $n \times m$ méretű tömbben helyezkedik el (tömbképlet!).

`INVERZ.MÁTRIX(tömb)` A `tömb`-ben található $n \times n$ típusú mátrix inverze, mely egy $n \times n$ méretű tömbben helyezkedik el (tömbképlet!).

`MSZORZAT(tömb1;tömb2)` A `tömb1`-ben található $m \times n$ típusú mátrix és a `tömb2`-ben található $n \times k$ típusú mátrix szorzata, mely egy $m \times k$ méretű tömbben helyezkedik el (tömbképlet!).

10.7.11. Kombinatorika

$m! = \boxed{\texttt{FAKT}(m)}$ $m \in \mathbb{N}$

$m!! = \boxed{\texttt{FAKTDUPLA}(m)}$ $m \in \mathbb{N}$ ($m!!$ az ún. *szemifaktoriális*, amely $1 \cdot 3 \cdot \ldots \cdot m$, ha m páratlan, illetve $2 \cdot 4 \cdot \ldots \cdot m$, ha m páros.)

$\binom{m}{k} = \boxed{\texttt{KOMBINÁCIÓK}(m;k)}$ $m \in \mathbb{N},\ k = 0, \ldots, m$

$\frac{m!}{(m-k)!} = \boxed{\texttt{VARIÁCIÓK}(m;k)}$ $m \in \mathbb{N},\ k = 0, \ldots, m$

$\frac{(k_1+k_2+\cdots+k_r)!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!} = \boxed{\texttt{SZORHÁNYFAKT}(k_1;k_2;\ldots;k_r)}$ $k_1, k_2, \ldots, k_r \in \mathbb{N}$

10.7.12. Pszeudo-véletlen szám generálása

$\boxed{\texttt{VÉL()}}$ $[0, 1]$ intervallumon egyenletes eloszlású pszeudo-véletlen szám

$\boxed{\texttt{VÉLETLEN.KÖZÖTT}(a;b)} = \boxed{\texttt{INT((}b\texttt{-}a\texttt{+1)*VÉL())+}a}$ $(a, b \in \mathbb{N},\ a < b)$ diszkrét egyenletes eloszlású pszeudo-véletlen szám az $\{a, a+1, \ldots, b\}$ halmazon

10.7.13. Statisztikák

Legyen a ξ valószínűségi változóra vonatkozó $x_1, \ldots, x_n$ mintarealizáció az `A` oszlopban. Jelölje $x_1^*, \ldots, x_n^*$ a rendezett mintarealizációt. Ekkor

$x_1^* = \boxed{\texttt{MIN(A:A)}}$

$x_n^* = \boxed{\texttt{MAX(A:A)}}$

$x_k^* = \boxed{\texttt{KICSI(A:A;}k\texttt{)}}$ $k = 1, \ldots, n$

$x_{n-k}^* = \boxed{\texttt{NAGY(A:A;}k+1\texttt{)}}$ $k = 0, \ldots, n-1$

$\min\{k : x_k^* = x_i\} = \boxed{\texttt{RANG.EGY(}x_i\texttt{;A:A;1)}}$ $i = 1, \ldots, n$

$\min\{k : x_{n-k}^* = x_i\} + 1 = \boxed{\texttt{RANG.EGY(}x_i\texttt{;A:A;0)}}$ $i = 1, \ldots, n$

$n = \boxed{\texttt{DARAB(A:A)}}$

$\overline{\xi} = \boxed{\texttt{ÁTLAG(A:A)}}$

$S_n = \boxed{\texttt{SZÓR.S(A:A)}}$

$S_n^2 = \boxed{\texttt{VAR.S(A:A)}}$

$S_n^* = \boxed{\texttt{SZÓR.M(A:A)}}$

$S_n^{*2} = \boxed{\texttt{VAR.M(A:A)}}$

tapasztalati medián = $\boxed{\texttt{MEDIÁN(A:A)}}$

tapasztalati módusz = $\boxed{\texttt{MÓDUSZ.EGY(A:A)}}$

$100t$%-os tapasztalati kvantilis = $\boxed{\texttt{PERCENTILIS.TARTALMAZ(A:A;}t\texttt{)}}$ $0 \le t \le 1$

tapasztalati alsó kvartilis = $\boxed{\texttt{KVARTILIS.TARTALMAZ(A:A;1)}}$

tapasztalati felső kvartilis = $\boxed{\texttt{KVARTILIS.TARTALMAZ(A:A;3)}}$

tapasztalati ferdeség = $\boxed{\texttt{FERDESÉG.P(A:A)}}$

tapasztalati lapultság (csúcsosság) = $\boxed{\texttt{CSÚCSOSSÁG(A:A)}}$

$\sum_{i=1}^{n} x_i = \boxed{\texttt{SZUM(A:A)}}$
$\sum_{i=1}^{n} x_i^2 = \boxed{\texttt{NÉGYZETÖSSZEG(A:A)}}$
$\sum_{i=1}^{n} (x_i - \overline{\xi})^2 = \boxed{\texttt{SQ(A:A)}}$
$\frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{\xi}| = \boxed{\texttt{ÁTL.ELTÉRÉS(A:A)}}$
$\prod_{i=1}^{n} x_i = \boxed{\texttt{SZORZAT(A:A)}}$
$\sqrt[n]{\prod_{i=1}^{n} x_i} = \boxed{\texttt{MÉRTANI.KÖZÉP(A:A)}}\ x_i > 0\ (i = 1, \ldots, n)$
$\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1} = \boxed{\texttt{HARM.KÖZÉP(A:A)}}\ x_i > 0\ (i = 1, \ldots, n)$
a-nál kisebb elemek száma $=$ $\boxed{\texttt{DARABTELI(A:A;"<"\&}a\texttt{)}}$ $a \in \mathbb{R}$ (2023. márciusa utáni Excel-verziókban ez a függvény `DARABHA` néven érhető el.)
$(a, b]$-beli elemek száma $=$ $\boxed{\texttt{DARABHATÖBB(A:A;">"\&}a\texttt{;A:A;"<="\&}b\texttt{)}}$ $a, b \in \mathbb{R}$
a-nál kisebb elemek összege $=$ $\boxed{\texttt{SZUMHA(A:A;"<"\&}a\texttt{)}}$ $a \in \mathbb{R}$
$(a, b]$-beli elemek összege $=$ $\boxed{\texttt{SZUMHATÖBB(A:A;A:A;">"\&}a\texttt{;A:A;"<="\&}b\texttt{)}}$ $a, b \in \mathbb{R}$
a-nál kisebb elemek átlaga $=$ $\boxed{\texttt{ÁTLAGHA(A:A;"<"\&}a\texttt{)}}$ $a \in \mathbb{R}$
$(a, b]$-beli elemek átlaga $=$ $\boxed{\texttt{ÁTLAGHATÖBB(A:A;A:A;">"\&}a\texttt{;A:A;"<="\&}b\texttt{)}}$ $a, b \in \mathbb{R}$

Legyen a ξ-re vonatkozó mintarealizáció $x_1, \ldots, x_n$ és az η-ra vonatkozó mintarealizáció $y_1, \ldots, y_n$. Az `A` oszlop i-edik sorában legyen x_i, illetve a `B` oszlop i-edik sorában legyen y_i. Ekkor

$\mathrm{Cov}_n(\xi, \eta) = \boxed{\texttt{KOVARIANCIA.S(A:A;B:B)}}$
$\mathrm{Corr}_n(\xi, \eta) = \boxed{\texttt{KORREL(A:A;B:B)}}$
$R^2 = \mathrm{Corr}_n^2(\xi, \eta) = \boxed{\texttt{RNÉGYZET(A:A;B:B)}}$
$\sum_{i=1}^{n} x_i y_i = \boxed{\texttt{SZORZATÖSSZEG(A:A;B:B)}}$
$\sum_{i=1}^{n} (x_i - y_i)^2 = \boxed{\texttt{SZUMXBŐLY2(A:A;B:B)}}$
$\sum_{i=1}^{n} (x_i^2 - y_i^2) = \boxed{\texttt{SZUMX2BŐLY2(A:A;B:B)}}$
$\sum_{i=1}^{n} (x_i^2 + y_i^2) = \boxed{\texttt{SZUMX2MEGY2(A:A;B:B)}}$

10.7.14. Eloszlásfüggvények

- **Binomiális eloszlás** (r-edrendű p paraméterű)
  $\sum_{i=0}^{k} \binom{r}{i} p^i (1-p)^{r-i} = \boxed{\texttt{BINOM.ELOSZL(}k\texttt{;}r\texttt{;}p\texttt{;IGAZ)}}$
  $r \in \mathbb{N},\ k = 0, \ldots, r,\ 0 < p < 1$

- **Hipergeometrikus eloszlás**

$\sum\limits_{i=0}^{k} \frac{\binom{M}{i}\binom{N-M}{r-i}}{\binom{N}{r}} = \boxed{\texttt{HIPGEOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}M\texttt{;}N\texttt{;IGAZ)}}$
$r, M, N \in \mathbb{N},\ M < N,\ r \leq \min\{M, N-M\},\ k = 0, \ldots, r$

- **Poisson-eloszlás** (λ paraméterű)
  $\sum\limits_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} = \boxed{\texttt{POISSON.ELOSZLÁS(}k\texttt{;}\lambda\texttt{;IGAZ)}}\ \lambda > 0,\ k = 0, 1, \ldots$

- **Exponenciális eloszlás** (λ paraméterű)
  $F(x) = 1 - e^{-\lambda x} = \boxed{\texttt{EXP.ELOSZL(}x\texttt{;}\lambda\texttt{;IGAZ)}}\ \lambda > 0,\ x \geq 0$

- **F-eloszlás** (s_1 és s_2 szabadsági fokú)
  $F(x) = \boxed{\texttt{F.ELOSZL(}x\texttt{;}s_1\texttt{;}s_2\texttt{;IGAZ)}}\ s_1, s_2 \in \mathbb{N},\ x \geq 0$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $F(x) = \boxed{\texttt{GAMMA.ELOSZL(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;IGAZ)}}\ r, \lambda > 0,\ x \geq 0$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $F(x) = \boxed{\texttt{KHINÉGYZET.ELOSZLÁS(}x\texttt{;}s\texttt{;IGAZ)}}\ s \in \mathbb{N},\ x \geq 0$

- **Normális eloszlás** (m és σ paraméterű)
  $F(x) = \Phi\left(\frac{x-m}{\sigma}\right) = \boxed{\texttt{NORM.ELOSZLÁS(}x\texttt{;}m\texttt{;}\sigma\texttt{;IGAZ)}}$
  $m, x \in \mathbb{R},\ \sigma > 0$

- **Standard normális eloszlás**
  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t = \boxed{\texttt{NORM.S.ELOSZLÁS(}x\texttt{;IGAZ)}}\ x \in \mathbb{R}$

- **t-eloszlás** (s szabadsági fokú)
  $F(x) = \boxed{\texttt{T.ELOSZL(}x\texttt{;}s\texttt{;IGAZ)}}\ s \in \mathbb{N},\ x \in \mathbb{R}$

10.7.15. Inverz eloszlásfüggvények

- **Exponenciális eloszlás** (λ paraméterű)
  $F^{-1}(x) = \boxed{\texttt{-LN(1-}x\texttt{)/}\lambda}\ \lambda > 0,\ 0 < x < 1$

- **F-eloszlás** (s_1 és s_2 szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{F.INVERZ(}x\texttt{;}s_1\texttt{;}s_2\texttt{)}}\ s_1, s_2 \in \mathbb{N},\ 0 < x < 1$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $F^{-1}(x) = \boxed{\texttt{GAMMA.INVERZ(}x\texttt{;}r\texttt{;1/}\lambda\texttt{)}}\ r, \lambda > 0,\ 0 < x < 1$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{KHINÉGYZET.INVERZ(}x\texttt{;}s\texttt{)}}\ s \in \mathbb{N},\ 0 < x < 1$

- **Normális eloszlás** (m és σ paraméterű)
  $F^{-1}(x) = \boxed{\texttt{NORM.INVERZ(}x\texttt{;}m\texttt{;}\sigma\texttt{)}}\ m \in \mathbb{R},\ \sigma > 0,\ 0 < x < 1$

- **Standard normális eloszlás**
  $\Phi^{-1}(x) = \boxed{\texttt{NORM.S.INVERZ(}x\texttt{)}}\ 0 < x < 1$

- **t-eloszlás** (s szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{T.INVERZ(}x\texttt{;}s\texttt{)}}\ s \in \mathbb{N},\ 0 < x < 1$

10.7.16. Eloszlások

- **Binomiális eloszlás** (r-edrendű p paraméterű)
  $\binom{r}{k}p^k(1-p)^{r-k} = \boxed{\texttt{BINOM.ELOSZL(}k\texttt{;}r\texttt{;}p\texttt{;HAMIS)}}$
  $r \in \mathbb{N},\ k = 0, \ldots, r,\ 0 < p < 1$

- **Hipergeometrikus eloszlás**
  $\frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}} = \boxed{\texttt{HIPGEOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}M\texttt{;}N\texttt{;HAMIS)}}$
  $r, M, N \in \mathbb{N},\ M < N,\ r \leq \min\{M, N-M\},\ k = 0, \ldots, r$

- **Poisson-eloszlás** (λ paraméterű)
  $\frac{\lambda^k}{k!}e^{-\lambda} = \boxed{\texttt{POISSON.ELOSZLÁS(}k\texttt{;}\lambda\texttt{;HAMIS)}}\ \lambda > 0,\ k = 0, 1, \ldots$

10.7.17. Sűrűségfüggvények

- **Exponenciális eloszlás** (λ paraméterű)
  $f(x) = \lambda e^{-\lambda x} = \boxed{\texttt{EXP.ELOSZL(}x\texttt{;}\lambda\texttt{;HAMIS)}}\ \lambda > 0,\ x \geq 0$

- **F-eloszlás** (s_1 és s_2 szabadsági fokú)
  $F(x) = \boxed{\texttt{F.ELOSZL(}x\texttt{;}s_1\texttt{;}s_2\texttt{;HAMIS)}}\ s_1, s_2 \in \mathbb{N},\ x \geq 0$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $f(x) = \boxed{\texttt{GAMMA.ELOSZL(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;HAMIS)}}\ r, \lambda > 0,\ x \geq 0$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $f(x) = \boxed{\texttt{KHINÉGYZET.ELOSZLÁS(}x\texttt{;}s\texttt{;HAMIS)}}\ x \geq 0$

- **Normális eloszlás** (m és σ paraméterű)
  $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right) = \boxed{\texttt{NORM.ELOSZLÁS(}x\texttt{;}m\texttt{;}\sigma\texttt{;HAMIS)}}$
  $m, x \in \mathbb{R},\ \sigma > 0$

- **Standard normális eloszlás**
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \boxed{\texttt{NORM.S.ELOSZLÁS(}x\texttt{;HAMIS)}}\ x \in \mathbb{R}$

- **t-eloszlás** (s szabadsági fokú)
  $F(x) = \boxed{\texttt{T.ELOSZL(}x\texttt{;}s\texttt{;HAMIS)}}\ s \in \mathbb{N},\ x \in \mathbb{R}$

10.7.18. Grafikus illeszkedésvizsgálat

$\boxed{\texttt{MEREDEKSÉG(tömb\_}y_i\texttt{;tömb\_}x_i\texttt{)}}$ Az $(x_i, y_i),\ i = 1, \ldots, r$ pontokra illesztett lineáris trendvonal meredeksége.
$\boxed{\texttt{METSZ(tömb\_}y_i\texttt{;tömb\_}x_i\texttt{)}}$ Az $(x_i, y_i),\ i = 1, \ldots, r$ pontokra illesztett lineáris trendvonal függőleges tengelymetszete.

10.7.19. Intervallumbecslés

$\frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \boxed{\texttt{MEGBÍZHATÓSÁG.NORM(}\alpha\texttt{;}\sigma\texttt{;}n\texttt{)}}\ 0 < \alpha < 1,\ \sigma > 0,\ n \in \mathbb{N}$
$\frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right) = \boxed{\texttt{MEGBÍZHATÓSÁG.T(}\alpha\texttt{;}S_n^*\texttt{;}n\texttt{)}}\ F = F[\mathrm{T}(n-1)],\ 0 < \alpha < 1,\ n \in \mathbb{N}$
$\min\left\{c \in \mathbb{N} : \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \geq x\right\} = \boxed{\texttt{BINOM.INVERZ(}n\texttt{;}p\texttt{;}x\texttt{)}}\ n \in \mathbb{N},\ 0 < p < 1,\ 0 < x < 1$

10.7.20. Paraméteres hipotézisvizsgálatok

A ξ-re illetve η-ra vonatkozó mintarealizációk az `A` illetve `B` oszlopokban vannak.

- **Egymintás u-próba**
  $1 - \Phi(u) = \boxed{\texttt{Z.PRÓB(A:A;}m_0\texttt{;}\sigma\texttt{)}}$
  $2 - 2\Phi(|u|) = \boxed{\texttt{2*MIN(Z.PRÓB(A:A;}m_0\texttt{;}\sigma\texttt{);1-Z.PRÓB(A:A;}m_0\texttt{;}\sigma\texttt{))}}$

- **Egymintás t-próba**
  A ξ-re vonatkozó mintarealizáció minden tagja mellett szerepeljen m_0 értéke a `B` oszlopban.
  $2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;2;1)}}$
  $1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;1)}}$

- **F-próba**
  $2\min\{F(\mathsf{F}), 1 - F(\mathsf{F})\} = \boxed{\texttt{F.PRÓB(A:A;B:B)}}$

- **Kétmintás t-próba**
  $2 - 2F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;2;2)}}$
  $1 - F(|t|) = \boxed{\texttt{T.PRÓB(A:A;B:B;1;2)}}$

- **Scheffé-módszer azonos mintaelemszámra (párosított t-próba)**
  $2 - 2F(|t|) =$ `T.PRÓB(A:A;B:B;2;1)`
  $1 - F(|t|) =$ `T.PRÓB(A:A;B:B;1;1)`

- **Scheffé-módszer különböző mintaelemszámra**
  Az ζ-ra vonatkozó mintarealizáció a `C` oszlopban van és minden tagja mellett szerepeljen 0 a `D` oszlopban.
  $2 - 2F(|t|) =$ `T.PRÓB(C:C;D:D;2;1)`
  $1 - F(|t|) =$ `T.PRÓB(C:C;D:D;1;1)`

- **Welch-próba**
  $2 - 2F(|t|) =$ `T.PRÓB(A:A;B:B;2;3)`
  $1 - F(|t|) =$ `T.PRÓB(A:A;B:B;1;3)`

- **Statisztikai próba valószínűségre**
  $F^{-1}(x) =$ $\boxed{\texttt{BINOM.INVERZ(}n\texttt{;}p_0\texttt{;}x\texttt{)}}$

10.7.21. Nemparaméteres hipotézisvizsgálatok

- **Tiszta illeszkedésvizsgálat**
  $1 - F(\chi^2) =$ $\boxed{\texttt{KHINÉGYZET.PRÓBA(}\varrho_i\ \texttt{tartománya;}\nu_i\ \texttt{tartománya)}}$

- **Függetlenségvizsgálat**
  $1 - F(\chi^2) =$ $\boxed{\texttt{KHINÉGYZET.PRÓBA(}\varrho_{ij}\ \texttt{tartománya;}\nu_{ij}\ \texttt{tartománya)}}$

- **Homogenitásvizsgálat**
  $1 - F(\chi^2) =$ $\boxed{\texttt{KHINÉGYZET.PRÓBA(}\varrho_{ij}\ \texttt{tartománya;}\nu_{ij}\ \texttt{tartománya)}}$

- **Kétmintás előjelpróba**
  $F^{-1}(x) =$ $\boxed{\texttt{BINOM.INVERZ(}n\texttt{;1/2;}x\texttt{)}}$

10.7.22. Regressziószámítás

- **Lineáris regresszió**
  `eta`: η-ra vonatkozó mintarealizációt tartalmazó $n \times 1$ méretű tömb.
  `xi`: $(\xi_1, \ldots, \xi_k)$-ra vonatkozó mintarealizációt tartalmazó $n \times k$ méretű tömb.
  `x`: $x_1, \ldots, x_k$ számokat tartalmazó $1 \times k$ méretű tömb.
  $(\widehat{a}_k, \widehat{a}_{k-1}, \ldots, \widehat{a}_0) =$ `LIN.ILL(eta;xi)` $(1 \times (k+1)$ méretű tömbképlet!)
  $\widehat{a}_0 + \widehat{a}_1 x_1 + \cdots + \widehat{a}_k x_k =$ `TREND(eta;xi;x)`

- **Fixpontos lineáris regresszió**

  `eta-t`: $(\eta - t_0)$-ra vonatkozó mintarealizációt tartalmazó $n \times 1$ méretű tömb.

  `xi-t`: $(\xi_1 - t_1, \ldots, \xi_k - t_k)$-ra vonatkozó mintarealizációt tartalmazó $n \times k$ méretű tömb.

  `x-t`: $x_1 - t_1, \ldots, x_k - t_k$ számokat tartalmazó $1 \times k$ méretű tömb.

  $(\widehat{a}_k, \widehat{a}_{k-1}, \ldots, \widehat{a}_1) = \boxed{\texttt{LIN.ILL(eta-t;xi-t;HAMIS)}}$ ($1 \times k$ méretű tömbképlet!)

  $\widehat{a}_1(x_1 - t_1) + \cdots + \widehat{a}_k(x_k - t_k) = \boxed{\texttt{TREND(eta-t;xi-t;x-t;HAMIS)}}$

- **Exponenciális regresszió**

  `eta`: η-ra vonatkozó mintarealizációt tartalmazó $n \times 1$ méretű tömb.

  `xi`: ξ_1-re vonatkozó mintarealizációt tartalmazó $n \times 1$ méretű tömb.

  $(\widehat{b}, \widehat{a}) = \boxed{\texttt{LOG.ILL(eta;xi)}}$ (1×2 méretű tömbképlet!)

  $\widehat{a} \cdot \widehat{b}^x = \boxed{\texttt{NÖV(eta;xi;}x\texttt{)}}$

Irodalomjegyzék

[1] DEÁK I.: *Véletlenszám-generátorok és alkalmazásuk*, Akadémiai Kiadó, Budapest, 1986.

[2] FAZEKAS I. (szerk.): *Bevezetés a matematikai statisztikába*, Kossuth Egyetemi Kiadó, Debrecen, 2000.

[3] HUNYADI L., MUNDRUCZÓ GY., VITA L.: *Statisztika*, Aula Kiadó, Budapesti Közgazdaságtudományi Egyetem, 1996.

[4] KOVALCSIKNÉ PINTÉR O.: *Az Excel függvényei A-tól Z-ig*, ComputerBooks, Budapest, 2008.

[5] LOVÁSZ L.: *Véletlen és álvéletlen*, Természet világa, 2000. II. különszám.

[6] LUKÁCS O.: *Matematikai statisztika példatár*, Műszaki Könyvkiadó, Budapest, 1987.

[7] MESZÉNA GY., ZIERMANN M.: *Valószínűségelmélet és matematikai statisztika*, Közgazdasági és Jogi Könyvkiadó, Budapest, 1981.

[8] MOGYORÓDI J., MICHALETZKY GY. (szerk.): *Matematikai statisztika*, Nemzeti Tankönyvkiadó, Budapest, 1995.

[9] PÉTERFY K.: *Microsoft Office Excel 2007 – Függvények* (magyar változat), Mercator Stúdió, 2007.

[10] PRÉKOPA A.: *Valószínűségelmélet műszaki alkalmazásokkal*, Műszaki Könyvkiadó, Budapest, 1962.

[11] RÉNYI A.: *Valószínűségszámítás*, Tankönyvkiadó, Budapest, 1966.

[12] RÉVÉSZ P.: *Mennyire véletlen a véletlen?* Akadémiai Kiadó, Budapest, 1984.

[13] TÓMÁCS TIBOR: *Matematikai statisztika*