Tómács Tibor

Matematikai statisztika gyakorlatok

Eszterházy Károly Főiskola
Matematikai és Informatikai Intézet

Tómács Tibor

Matematikai statisztika gyakorlatok

Eger, 2012

Szerző:
Dr. Tómács Tibor
főiskolai docens
Eszterházy Károly Főiskola

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 6

Jelölések 7

1. Mintagenerálás 9

1.1. Egyenletes eloszlás . . . 9
1.2. Diszkrét egyenletes eloszlás . . . 11
1.3. Karakterisztikus eloszlás . . . 11
1.4. Binomiális eloszlás . . . 12
1.5. Exponenciális eloszlás . . . 12
1.6. Normális eloszlás . . . 13
1.7. Gyakorlatok . . . 14

2. Tapasztalati eloszlás 18

2.1. Tapasztalati eloszlásfüggvény . . . 18
2.2. Vonaldiagram . . . 25
2.3. Sűrűséghisztogram . . . 28
2.4. Gyakorlatok . . . 31

3. Grafikus illeszkedésvizsgálat 37

3.1. Általános vizsgálat . . . 37
3.2. Grafikus normalitásvizsgálat . . . 37
3.3. Grafikus exponencialitásvizsgálat . . . 39
3.4. Gyakorlatok . . . 40

4. Statisztikák 44

4.1. Gyakorlatok . . . 46

5. Intervallumbecslések 49

5.1. Normális eloszlás paramétereinek becslése . . . 49
5.2. Valószínűség becslése . . . 52
5.3. Gyakorlatok . . . 53

6. Paraméteres hipotézisvizsgálatok 57

6.1. Egymintás u-próba . . . 58
6.2. Kétmintás u-próba . . . 59
6.3. Egymintás t-próba . . . 60

6.4. F-próba . . . 60
6.5. Kétmintás t-próba . . . 61
6.6. Scheffé-módszer . . . 62
6.7. Khi-négyzet próba . . . 64
6.8. Statisztikai próba az exponenciális eloszlás paraméterére . . . 65
6.9. Statisztikai próba valószínűségre . . . 66
6.10. Gyakorlatok . . . 67

7. Nemparaméteres hipotézisvizsgálatok 72
7.1. Tiszta illeszkedésvizsgálat . . . 72
7.2. Becsléses illeszkedésvizsgálat . . . 74
7.3. Függetlenségvizsgálat . . . 75
7.4. Homogenitásvizsgálat . . . 79
7.5. Kétmintás előjelpróba . . . 80
7.6. Kolmogorov – Szmirnov-féle kétmintás próba . . . 81
7.7. Kolmogorov – Szmirnov-féle egymintás próba . . . 83
7.8. Gyakorlatok . . . 85

8. Regressziószámítás 88
8.1. Lineáris regresszió . . . 88
8.2. Fixpontos lineáris regresszió . . . 92
8.3. Nemlineáris regresszió . . . 95
8.3.1. Polinomos regresszió . . . 96
8.3.2. Hatványkitevős regresszió . . . 97
8.3.3. Exponenciális regresszió . . . 99
8.3.4. Logaritmikus regresszió . . . 100
8.3.5. Hiperbolikus regresszió . . . 100
8.4. Gyakorlatok . . . 100

9. Összefoglaló 103
9.1. Eloszlások generálása . . . 103
9.1.1. Egyenletes eloszlásból származtatott eloszlások . . . 103
9.1.2. Normális eloszlásból származtatott eloszlások . . . 104
9.2. Grafikus illeszkedésvizsgálat . . . 104
9.3. Intervallumbecslések . . . 105
9.4. Paraméteres hipotézisvizsgálatok . . . 106
9.5. Nemparaméteres hipotézisvizsgálatok . . . 109

9.6. Regressziószámítás 112
9.7. Excel függvények 115
9.7.1. Logikai függvények 116
9.7.2. Elemi függvények 116
9.7.3. Mátrixok 117
9.7.4. Kombinatorika 117
9.7.5. Pszeudo-véletlen szám generálása 117
9.7.6. Statisztikák 117
9.7.7. Eloszlások 119
9.7.8. Eloszlásfüggvények 119
9.7.9. Sűrűségfüggvények 120
9.7.10. Inverz eloszlásfüggvények 120
9.7.11. Grafikus illeszkedésvizsgálat 121
9.7.12. Intervallumbecslés 121
9.7.13. Paraméteres hipotézisvizsgálatok 121
9.7.14. Nemparaméteres hipotézisvizsgálatok 122
9.7.15. Regressziószámítás 122

Irodalomjegyzék **124**

Előszó

Ez a tananyag az egri Eszterházy Károly Főiskola matematikai statisztika gyakorlataiból készült, melyet matematika tanárszakos és programtervező informatikus hallgatóknak szánunk.

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 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, sok esetben videóval is bemutatva. 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
I_A	az A esemény indikátorváltozója
$\mathbf{Bin}(r; p)$	az r-edrendű p paraméterű binomiális eloszlású valószínűségi változók halmaza
$\mathbf{Exp}(\lambda)$	a λ paraméterű exponenciális eloszlású valószínűségi változók halmaza
$\mathbf{Norm}(m; \sigma)$	az m várható értékű és σ szórású normális eloszlású valószínűségi változók halmaza
$\mathbf{Norm}_d(m; A)$	az m és A paraméterű d-dimenziós normális eloszlású valószínűségi változók halmaza
$\mathbf{Gamma}(r; \lambda)$	az r-edrendű λ paraméterű gamma-eloszlású valószínűségi változók halmaza
$\mathbf{Khi}(s)$	az s szabadsági fokú khi-négyzet eloszlású valószínűségi változók halmaza

$\mathbf{t}(s)$	az s szabadsági fokú t-eloszlású valószínűségi változók halmaza
$\mathbf{F}(s_1; s_2)$	az s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változók halmaza
$F \sim \mathbf{V}$	Ha ξ valószínűségi változó, és $\mathbf{V}$ a ξ-vel azonos eloszlású valószínűségi változók halmaza, akkor ez azt jelöli, hogy F a $\mathbf{V}$-beli valószínűségi változók közös eloszlásfüggvénye. Például $\Phi \sim \mathbf{Norm}(0; 1)$.

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
$\xi_1^*, \ldots, \xi_n^*$	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. Mintagenerálás

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 elmélete igen terjedelmes és túlmutat ezen mű keretein. (Részletesebben lásd például [1, 5, 12].) Itt csak azt fogjuk részletezni, hogy egyenletes eloszlásból hogyan lehet más eloszlást generálni.

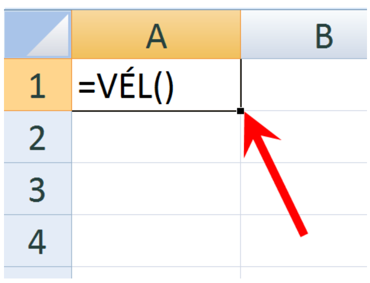
Megjegyezzük, hogy valódi véletlent is használhatunk minta generálására, a következő címen található internetes szolgáltatással: http://www.random.org.

1.1. Egyenletes eloszlás

Excel-ben a `VÉL()` függvénnyel tudunk $[0, 1]$ intervallumon egyenletes eloszlású (pszeudo)véletlen számot generálni. Ennek a függvénynek az értéke minden esetben $[0, 1)$-beli.

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

Megoldás. Az `A1` cellába írja be, hogy `=VÉL()`. (Egy képletet mindig `=` jellel kell kezdeni.) Ezután a kitöltőjelet húzza le a 20. sorig. (A *kitöltőjel* a kijelölés jobb alsó sarkában lévő fekete négyzet, amire a következő ábrán egy piros nyíl mutat.)



A következő videón mindezt megnézheti a gyakorlatban.

VIDEÓ

A mintarealizáció elemeinek rögzítése. 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.

3. Lépjen az `A` oszlop fejlécére, nyomja meg a jobb egérgombot, majd válassza a *Törlés* pontot.

Mindezeket a következő videón is megnézheti:

VIDEÓ

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.

1.2. 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. Legyen ξ eloszlásfüggvénye F, illetve $a+(b-a)\xi$ eloszlásfüggvénye G. Ekkor

$$G(x)=\mathrm{P}\big(a+(b-a)\xi<x\big)=\mathrm{P}\left(\xi<\frac{x-a}{b-a}\right)=F\left(\frac{x-a}{b-a}\right)=$$
$$=\begin{cases}0, & \text{ha } \frac{x-a}{b-a}<0\\ \frac{x-a}{b-a}, & \text{ha } 0\leqslant\frac{x-a}{b-a}\leqslant 1\\ 1, & \text{ha } \frac{x-a}{b-a}>1\end{cases}=\begin{cases}0, & \text{ha } x<a\\ \frac{x-a}{b-a}, & \text{ha } a\leqslant x\leqslant b\\ 1, & \text{ha } x>b\end{cases}$$

melyből adódik az állítás.

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

Megoldás. 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()`, a kitöltőjelet húzza le a 100. sorig, majd rögzítse a mintarealizáció elemeit. Mindez videón:

VIDEÓ

1.2. Diszkrét egyenletes eloszlás

1.4. Tétel. *Legyen ξ a $[0,1]$ intervallumon egyenletes eloszlású valószínűségi változó, $m \in \mathbb{N}$ és $x_1, \dots, x_m \in \mathbb{R}$. Feltesszük, hogy az x_i-k mindegyike különbözik a többitől. Ekkor $x_{[m\xi]+1}$ diszkrét egyenletes eloszlású az $\{ x_1, \dots, x_m \}$ halmazon.*

Bizonyítás. $\mathrm{P}(x_{[m\xi]+1} = x_i) = \mathrm{P}([m\xi]+1 = i) = \mathrm{P}\left(\frac{i-1}{m} \leqslant \xi < \frac{i}{m}\right) = \frac{1}{m}$.

1.5. 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. 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. Az egészrész-függvény az Excelben `INT()`.

Így `A1`-be írja be, hogy `=INT(6*VÉL())+1`. Ezután a kitöltőjelet húzza le a 10. sorig, majd rögzítse a mintarealizáció elemeit.

Az Excel erre a feladatra egy más megoldást is kínál. Az `A1` cellába az előbbi helyett írja be, hogy `=RANDBETWEEN(1;6)`. Mindez videón:

V I D E Ó

1.3. Karakterisztikus eloszlás

1.6. 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. $\mathrm{P}(\mathrm{I}_{\xi<p} = 1) = \mathrm{P}(\xi < p) = p$ és $\mathrm{P}(\mathrm{I}_{\xi<p} = 0) = \mathrm{P}(\xi \geqslant p) = 1-p$, melyből következik az állítás.

1.7. 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 valószínűségi változóra vonatkozó 30 elemű mintarealizációt.

Megoldás. 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. (A `HA` függvény leírását olvassa el az Excel súgójából.) Lépjen vissza `A1`-re, ezután a kitöltőjelet húzza le a 30. sorig, majd rögzítse a mintarealizáció elemeit. Mindez videón:

VIDEÓ

1.4. 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.

1.8. 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.*

1.9. Példa. Generáljon egy 0,8 valószínűségű esemény 5 kísérlet utáni gyakoriságára vonatkozó 20 elemű mintarealizációt.

Megoldás. A gyakoriság binomiális eloszlású, így a feladat egy $r = 5$ rendű $p = 0{,}8$ paraméterű binomiális eloszlású valószínűségi változóra vonatkozó 20 elemű mintarealizáció generálása.

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. (A `SZUM` függvény leírását olvassa el az Excel súgójából.)

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:

VIDEÓ

1.5. Exponenciális eloszlás

1.10. 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 $x \leqslant 0$, akkor $\mathrm{P}\left(\frac{-\ln\xi}{\lambda} < x\right) = 0$.

1.11. 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.

1.6. Normális eloszlás

1.12. Tétel (Box–Muller-transzformáció, 1958)**.** *Legyenek ξ, η a $[0,1]$ intervallumon egyenletes eloszlású független valószínűségi változók. Ekkor $\sqrt{-2\ln\xi}\cos(2\pi\eta)$ standard normális eloszlású.*

Bizonyítás. Ha $x > 0$, akkor $\mathrm{P}\left(\sqrt{-2\ln\xi} < x\right) = \mathrm{P}\left(\xi > e^{-\frac{x^2}{2}}\right) = 1 - e^{-\frac{x^2}{2}}$, illetve ha $x \leqslant 0$, akkor $\mathrm{P}\left(\sqrt{-2\ln\xi} < x\right) = 0$. Így $\sqrt{-2\ln\xi}$ sűrűségfüggvénye

$$f(x) = \begin{cases} 0, & \text{ha } x \leqslant 0, \\ \left(1 - e^{-\frac{x^2}{2}}\right)' = xe^{-\frac{x^2}{2}}, & \text{ha } x > 0. \end{cases}$$

Ha $-1 < x \leqslant 1$, akkor

$$\begin{aligned} &\mathrm{P}\big(\cos(2\pi\eta) < x\big) = \\ &= \mathrm{P}\left(\frac{1}{2\pi}\arccos x < \xi < 1 - \frac{1}{2\pi}\arccos x\right) = 1 - \frac{1}{\pi}\arccos x. \end{aligned}$$

$\mathrm{P}\big(\cos(2\pi\eta) < x\big) = 1$, ha $x > 1$, illetve $\mathrm{P}\big(\cos(2\pi\eta) < x\big) = 0$, ha $x \leqslant -1$. Így $\cos(2\pi\eta)$ sűrűségfüggvénye

$$g(x) = \begin{cases} \left(1 - \frac{1}{\pi}\arccos x\right)' = \frac{1}{\pi\sqrt{1-x^2}}, & \text{ha } -1 < x \leqslant 1, \\ 0 & \text{különben.} \end{cases}$$

Ismert, hogy f és g sűrűségfüggvényű független valószínűségi változók szorzatának sűrűségfüggvénye $h(z) = \int\limits_{-\infty}^{\infty} f(y)g\left(\frac{z}{y}\right)\frac{1}{|y|}\,\mathrm{d}y$. (Lásd például Rényi A. [11, 189. oldal].) Így $\sqrt{-2\ln\xi}\cos(2\pi\eta)$ sűrűségfüggvénye

$$\begin{aligned} h(z) &= \int\limits_{-\infty}^{\infty} f(y)g\left(\frac{z}{y}\right)\frac{1}{|y|}\,\mathrm{d}y = \int\limits_{0}^{\infty} e^{-\frac{y^2}{2}} g\left(\frac{z}{y}\right)\mathrm{d}y = \\ &= \int\limits_{|z|}^{\infty} e^{-\frac{y^2}{2}} \frac{1}{\pi\sqrt{1-\left(\frac{z}{y}\right)^2}}\,\mathrm{d}y = \frac{1}{\pi}\int\limits_{|z|}^{\infty} ye^{-\frac{y^2}{2}}\frac{1}{\sqrt{y^2-z^2}}\,\mathrm{d}y = \\ &= \frac{1}{\pi}e^{-\frac{z^2}{2}}\int\limits_{|z|}^{\infty} ye^{-\frac{y^2-z^2}{2}}\frac{1}{\sqrt{y^2-z^2}}\,\mathrm{d}y = \frac{1}{\pi}e^{-\frac{z^2}{2}}\int\limits_{0}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \\ &= \frac{1}{2\pi}e^{-\frac{z^2}{2}}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}. \end{aligned}$$

Az integrálásban $x = \sqrt{y^2 - z^2}$ helyettesítést alkalmaztunk.

1.13. Következmény. *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.

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

Megoldás. Az előző következmény alapján az `A1` cellába írja be, hogy

`=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.

1.7. Gyakorlatok

1.1. gyakorlat. Generáljon Cauchy-eloszlású valószínűségi változóra vonatkozó 15 elemű mintarealizációt.

Útmutatás. Ha ξ és η független standard normális eloszlású valószínűségi változók, akkor $\frac{\xi}{\eta}$ Cauchy-eloszlású.

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 $\sum_{i=1}^{s} \xi_i^2$ valószínűségi változó s szabadsági fokú khi-négyzet eloszlású. Így hasonlóan járhat el, mint a binomiális eloszlás generálásánál, de itt a standard normális eloszlásból induljon ki, ne a karakterisztikusból, továbbá `SZUM` helyett `NÉGYZETÖSSZEG` függvényt használjon.

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. Így felhasználhatja az előző gyakorlatot, továbbá a négyzetgyök számolásához alkalmazza a `GYÖK` függvényt.

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.

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 $\sum_{i=1}^{r} \xi_i$ valószínűségi változó r-edrendű λ paraméterű gamma-eloszlású. Így hasonlóan járhat el, mint a binomiális eloszlás generálásánál, de itt az exponenciális eloszlásból induljon ki, ne a karakterisztikusból.

1.6. gyakorlat. 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. Írjon programot, mely ξ-re vonatkozó mintarealizációt generál. (A ξ valószínűségi változót *hipergeometrikus eloszlásúnak* nevezzük.)

Útmutatás. Legyen $y_1, \ldots, y_r$ a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változóra vonatkozó mintarealizáció, és

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

Ekkor x_r a ξ-re vonatkozó 1 elemű mintarealizáció. (Ennek belátását az Olvasóra bízzuk.)

1.7. gyakorlat. Excel segítségével is generáljon 10 elemű mintarealizációt az előző feladatban szereplő ξ-re, $N = 20$, $M = 7$, $r = 5$ választással.

Útmutatás. Ha a *Munka1* munkalap `A1` cellája a dobozban lévő piros golyók számát, illetve a *Munka2* munkalap `A1` cellája a dobozban lévő golyók számát tartalmazza, akkor a *Munka1* munkalap `B1` cellájába

```
=HA(VÉL()<A1/Munka2!A1;A1-1;A1)
```

írva, az első húzás utáni dobozban maradt piros golyók számát kapjuk. A részletes megoldást végigkövetheti a következő videón.

VIDEÓ

1.8. gyakorlat. Egy kísérletet ismételjünk egymástól függetlenül, amíg egy rögzített A esemény be nem következik. Legyen ξ a végrehajtott kísérletek száma. Írjon programot, mely ξ-re vonatkozó mintarealizációt generál. (A ξ valószínűségi változót geometriai eloszlásúnak nevezzük.)

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

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

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

1.9. gyakorlat. Írjon programot, mely Poisson-eloszlású valószínűségi változóra vonatkozó mintarealizációt generál. (A ξ Poisson-eloszlású $\lambda > 0$ paraméterrel, ha az értékkészlete $\{0, 1, 2, \ldots\}$ és $\mathrm{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ minden $k = 0, 1, 2, \ldots$ esetén.)

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

$$\prod_{i=1}^{r-1} y_i \geqslant e^{-\lambda} \quad \text{és} \quad \prod_{i=1}^{r} y_i < e^{-\lambda}.$$

Ha $y_1 < e^{-\lambda}$, akkor legyen $r := 1$. Az így definiált r esetén $r - 1$ a ξ-re vonatkozó 1 elemű mintarealizáció. Könnyen látható, hogy ez az állítás ekvivalens a következő tétellel:

1.15. Tétel. *Legyenek $\eta_0, \eta_1, \ldots$ a $[0,1]$ intervallumon egyenletes eloszlású független valószínűségi változók és $\lambda > 0$. Ekkor*

$$\eta := \min\left\{ s : \prod_{i=0}^{s} \eta_i < e^{-\lambda} \right\}$$

Poisson-eloszlású λ paraméterrel.

Bizonyítás. $\mathrm{P}(\eta = 0) = \mathrm{P}\left(\eta_0 < e^{-\lambda}\right) = e^{-\lambda} = \frac{\lambda^0}{0!} e^{-\lambda}$, illetve az egyenletes eloszlás és a geometriai valószínűségi mező kapcsolata alapján

$$\mathrm{P}(\eta = 1) = \mathrm{P}\left(\eta_0 \geqslant e^{-\lambda}, \eta_0\eta_1 < e^{-\lambda}\right) = \int\limits_{e^{-\lambda}}^{1} \frac{e^{-\lambda}}{x_0}\,\mathrm{d}x_0 = \frac{\lambda^1}{1!}e^{-\lambda},$$

továbbá ha $k = 2, 3, \ldots$, akkor

$$\mathrm{P}(\eta = k) = \mathrm{P}\left(\eta_0 \cdots \eta_{k-1} \geqslant e^{-\lambda}, \eta_0 \cdots \eta_k < e^{-\lambda}\right) =$$
$$= \int\limits_{e^{-\lambda}}^{1} \int\limits_{\frac{e^{-\lambda}}{x_0}}^{1} \int\limits_{\frac{e^{-\lambda}}{x_0 x_1}}^{1} \cdots \int\limits_{\frac{e^{-\lambda}}{x_0 \cdots x_{k-2}}}^{1} \frac{e^{-\lambda}}{x_0 \cdots x_{k-1}}\,\mathrm{d}x_{k-1} \cdots \mathrm{d}x_0 = \frac{\lambda^k}{k!}e^{-\lambda}.$$

1.10. gyakorlat. Írjon programot, mely a következő eloszlású valószínűségi változókra vonatkozó mintarealizációkat generál: egyenletes, diszkrét egyenletes, karakterisztikus, binomiális, exponenciális, normális. Hasonló program letölthető a következő helyről:

PROGRAM

A program indítása után nyomja meg a *Mintagenerálás* gombot. A paraméterek beállítása után nyomja meg a megfelelő eloszlás gombját. Ekkor a mintarealizáció a vágólapra kerül. Ezután ezt bemásolhatjuk például egy Excel-munkalapra. Próbáljon ki néhány konkrét esetet.

2. Tapasztalati eloszlás

Ebben a fejezetben generált mintarealizáció alapján ábrázolunk tapasztalati eloszlásfüggvényt, vonaldiagramot és sűrűséghisztogramot.

2.1. Tapasztalati eloszlásfüggvény

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 $x_1 = \xi_1(\omega), \dots, x_n = \xi_n(\omega)$, akkor az $(x_i, F_n^*(x_i))$ koordinátájú pontok az F_n^* „lépcsőfokainak” a jobb oldali végpontjai. A legmagasabb lépcsőfok kezdőpontja a $(\max\{\, x_1, \dots, x_n \,\}$,1) koordinátájú pont. A következő feladatok megoldásában ezt a tényt fogjuk felhasználni.

A matematikai statisztika alaptétele szerint a tapasztalati eloszlásfüggvény 1 valószínűséggel egyenletesen konvergál $\mathbb{R}$-en a valódi eloszlásfüggvényhez. Vagyis, ha elég nagy a mintarealizáció elemeinek a száma, akkor a tapasztalati eloszlásfüggvény elég jól közelíti a valódit. Ezt is megvizsgáljuk néhány konkrét esetben.

2.1. 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. A mintarealizációt korábban láttuk hogyan kell generálni: Az `A1` cellába írja be, hogy

`=INT(6*VÉL())+1` vagy `=RANDBETWEEN(1;6)`.

Nyomjon *Enter*-t, lépjen vissza `A1`-re, ezután a kitöltőjelet húzza le a 100. sorig, majd rögzítse a mintarealizáció elemeit. A `B1` cellába írja be, hogy

`=DARABTELI(A:A;"<"&A1)/DARAB(A:A)`.

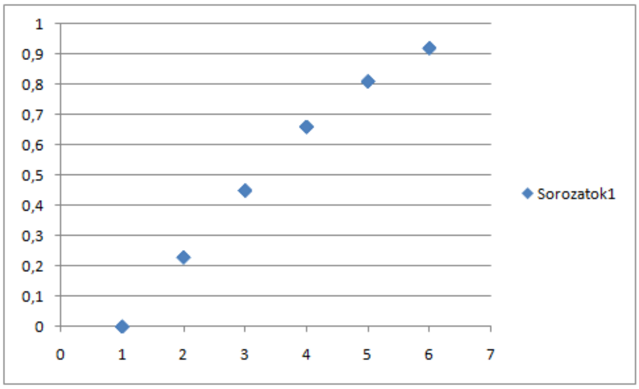
Nyomjon *Enter*-t. Ennek hatására kiszámolja, hogy az `A` oszlopban hány olyan elem van, mely kisebb az `A1` cella értékénél, majd elosztja az `A` oszlopban található számot tartalmazó cellák számával (azaz a mintarealizáció elemeinek a számával). Ez nem más, mint az `A1` cella értékénél felvett tapasztalati eloszlásfüggvény értéke.

Lépjen vissza `B1`-re, a kitöltőjelet húzza le a 100. sorig, majd menjen vissza `A1`-re. Ugyanezt a hatást úgy is elérhetjük, ha a `B1` cella kitöltőjelére kétszer klikkelünk.

A következőkben ábrázoljuk az $(\texttt{A1}, \texttt{B1}), (\texttt{A2}, \texttt{B2}), \ldots$ koordinátájú pontokat. Ehhez jelölje ki az `A` és `B` oszlopokat, majd

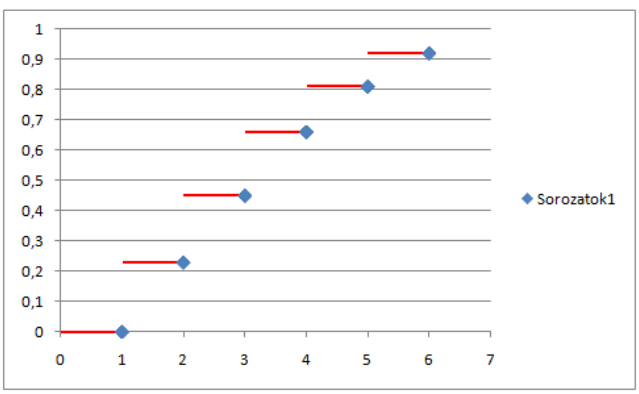
Beszúrás → *Diagramok/Pont* → *Pont csak jelölőkkel*

Ekkor megjelenik egy olyan függvény, amely a keresett lépcsős függvény lépcsőinek a jobb oldali végpontjait ábrázolja.



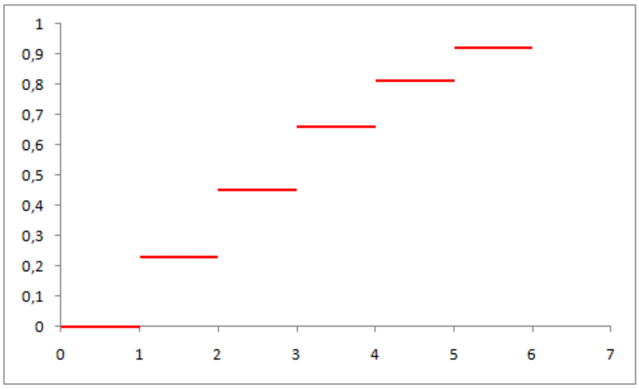
Ezután rajzolja meg a lépcsőfokokat is, felhasználva, hogy ebben az esetben minden lépcsőfok hossza 1.

Elrendezés → *Elemzés/Hibasávok* → *Elemzések standard hibával* →
Aktuális kijelölés/Diagramelemek: Sorozatok1 Y hibasávok →
Delete gomb →
Aktuális kijelölés/Diagramelemek: Sorozatok1 X hibasávok →
Aktuális kijelölés/Kijelölés formázása → *Irány/Mínusz* →
Végpont stílusa/Nyílt→ *A hiba mértéke/Abszolút értékben: 1* →
Vonal színe → *Folytonos vonal* → *Szín:* piros → *Vonalstílus* →
Szélesség: 1,5 pt → *Bezárás*



Ezzel gyakorlatilag kész a feladat, de még érdemes néhány finomítást elvégezni. Törölje a kék pontokat, a vezető rácsokat és a „Sorozatok1” feliratot.

Aktuális kijelölés/Diagramelemek: Sorozatok1 →
Aktuális kijelölés/Kijelölés formázása → Jelölő beállításai →
Jelölő típusa/Nincs → Bezárás →
Tengelyek/Rácsvonalak → Elsődleges vízszintes rácsvonalak → Nincs
Címkék/Jelmagyarázat → Nincs



A korábban leírtak szerint a legmagasabban lévő lépcsőfok itt még nem jelenik meg. Ezt pótolhatja például úgy, hogy az ábrázolt pontok közé szúrja a $(7, 1)$ koordinátájú pontot. Jelölje ki az 1. sort. Nyomja meg a jobb egérgombot, majd *Beszúrás*. Az `A1` cellába írja be, hogy 7, a B1-be pedig hogy 1, majd klikkeljünk a diagramterületre.

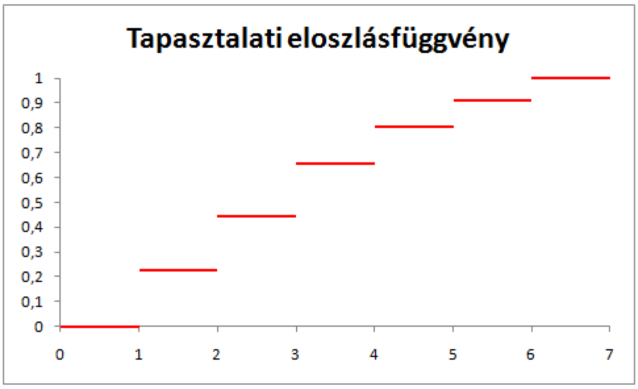
	A	B
1	7	1
2	5	0,80198
3	1	0
4	2	0,22772
5	2	0,22772
6	1	0
7	5	0,80198
8	1	0

A piros nyíllal jelölt pontot húzza fel az 1. sorba. Ezzel megjelenik a hiányzó lépcsőfok is. Végső simításként a vízszintes tengelyen megjelenő maximális értéket rögzítse 7-nek, a függőleges tengelyen megjelenő maximális értéket rögzítse 1-nek, végül adja a diagramnak a „Tapasztalati eloszlásfüggvény” címet.

Tengelyek/Tengelyek → Elsődleges vízszintes tengely →
Elsődleges vízszintes tengely további beállításai →
Maximum: Rögzített 7 → Bezárás →
Tengelyek/Tengelyek → Elsődleges függőleges tengely →

Elsődleges függőleges tengely további beállításai →
Maximum: Rögzített 1 → *Bezárás* →
Címkék/Diagramcím → *A diagram felett* →
A szerkesztőlécbe írja be: Tapasztalati eloszlásfüggvény → *Enter*

Ezzel megkapja a végeredményt:



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

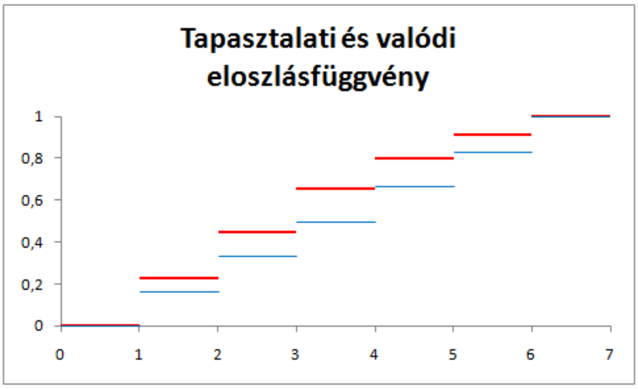
2.2. Példa. Az előző példában kapott grafikonon rajzolja fel a valódi eloszlásfüggvényt, azaz a $\{1, 2, 3, 4, 5, 6\}$ halmazon diszkrét egyenletes eloszlású valószínűségi változó eloszlásfüggvényét is.

Megoldás. Az előző munkalapon dolgozzon. A `C1:C7` cellatartományba írja rendre az 1, 2, 3, 4, 5, 6, 7 számokat (lépcsőfokok végeinek első koordinátái illetve az utolsó lépcsőfok egy pontjának első koordinátája). Ezután a `D1`-be írjon 0 értéket (első lépcsőfok magassága), `D2`-be `=D1+1/6`, majd a `D2` cella kitöltőjelére klikkeljen kétszer. Ezzel megkapja az összes lépcsőfok magasságát. Ezután klikkeljen a grafikonra, majd

Jobb egérgomb → *Helyi menü/Adatok kijelölése* → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$C$1:$C$7` →
Adatsor Y értékei: `=Munka1!$D$1:$D$7` → *OK* → *OK* →
Elrendezés → *Aktuális kijelölés/Diagramelemek: Sorozatok2* →
Elemzés/Hibasávok → *Elemzések standard hibával* →
Aktuális kijelölés/Diagramelemek: Sorozatok2 Y hibasávok →
Delete gomb → *Aktuális kijelölés/Diagramelemek: Sorozatok2 X hibasávok* →
Aktuális kijelölés/Kijelölés formázása → *Irány/Mínusz* → *Végpont stílusa/Nyílt*→
A hiba mértéke/Abszolút értékben: 1 → *Vonal színe* → *Folytonos vonal* →

Szín: kék → *Vonalstílus* → *Szélesség: 1 pt* → *Bezárás*

Klikkeljen a feliratra, és javítsa ki „Tapasztalati és valódi eloszlásfüggvény”-re.



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

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

Megoldás. A feladatot azzal a könnyítéssel oldjuk meg, hogy csak a lépcsőfokok jobb oldali végpontjait ábrázoljuk. Ez abszolút folytonos eloszlás esetén nem zavaró, mert a legtöbb lépcsőfok hossza nagyon rövid lesz az ábra felbontásához képest (legalábbis ha a mintarealizáció elemeinek a száma nagy). Mivel a lépcsőfokok száma 1 valószínűséggel 101 lesz, ezért az sem lesz zavaró, hogy az utolsó 1 magasságban levő lépcsőfokot nem rajzoljuk ki.

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 `B1` cellába írja be, hogy

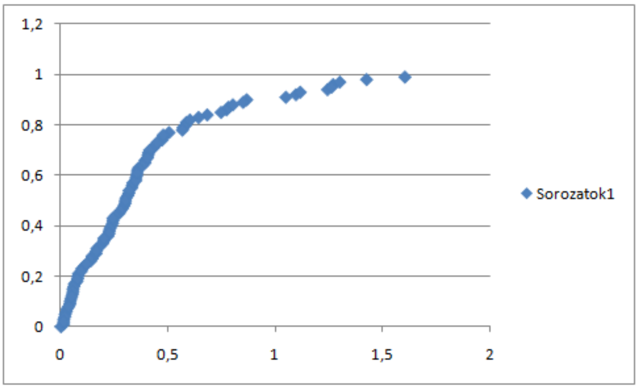
`=DARABTELI(A:A;"<"&A1)/DARAB(A:A)`.

Nyomjon *Enter*-t, majd a `B1` cella kitöltőjelére klikkeljen kétszer.

A következőkben ábrázolja az $(\mathtt{A1}, \mathtt{B1}), (\mathtt{A2}, \mathtt{B2}), \ldots$ koordinátájú pontokat. Ehhez jelölje ki az `A` és `B` oszlopokat, majd

Beszúrás → Diagramok/Pont → Pont csak jelölőkkel

Ekkor megjelenik az előbb ismertetett függvény.

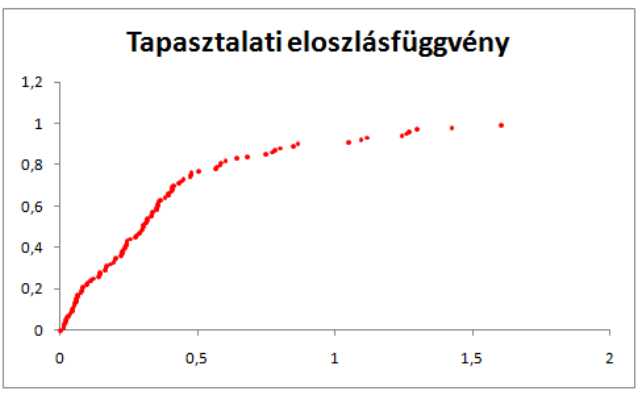


Még néhány finomítást érdemes elvégezni. Törölje a vezető rácsokat és a „Sorozatok1” feliratot. Ezt elvégezheti a korábban leírtak szerint is, de ráklikkelve az adott objektumra, majd a jobb egérgombot lenyomva, a helyi menüből is végrehajthatja.

Ezután a ◆ adatjelölőket változtassa 2 pt méretű piros ponttá. Ehhez klikkeljen valamelyik jelölő pontra, majd a jobb egérgombot megnyomva, a helyi menüből válassza ki az *Adatsorok formázása* pontot.

Jelölő beállításai → Beépített → Típus: pont → Méret: 2 →
Jelölőkitöltés → Egyszínű kitöltés → Szín: piros →
Jelölővonal színe → Folytonos vonal → Szín: piros → Bezárás

Végül adja a diagramnak a „Tapasztalati eloszlásfüggvény” címet az előző feladat megoldásában leírtak szerint.



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

2.4. 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. A megoldást az előző munkalapon végezze el. A `C1` cellába írja a vízszintes tengely minimális értékét (most ez 0). A `C2`-be írja be, hogy `=C1+0,1`. Itt 0,1 az a lépésköz, amellyel a függvény pontjait ábrázoljuk. Ezután a kitöltőjelet húzza le addig, amíg a vízszintes tengely maximális értékéig nem ér (jelen esetben 2-ig). A `D1` cellába írja a következőt:

`=EXP.ELOSZLÁS(C1;3;IGAZ)`

Ez a $\lambda = 3$ paraméterű exponenciális eloszlású valószínűségi változó eloszlásfüggvényének értékét adja a `C1` értékének a helyén. Ha `IGAZ` helyett `HAMIS` szerepelne a képletben, akkor eloszlásfüggvény helyett sűrűségfüggvényt számolna. Ezután a `D1` cella kitöltőjelére klikkeljen kétszer.

Klikkeljen a grafikonra, majd helyi menüben válassza az *Adatok kijelölése* pontot.

Hozzáadás →
Adatsor X értékei: `=Munka1!$C$1:$C$21` →
Adatsor Y értékei: `=Munka1!$D$1:$D$21` → *OK* → *OK*

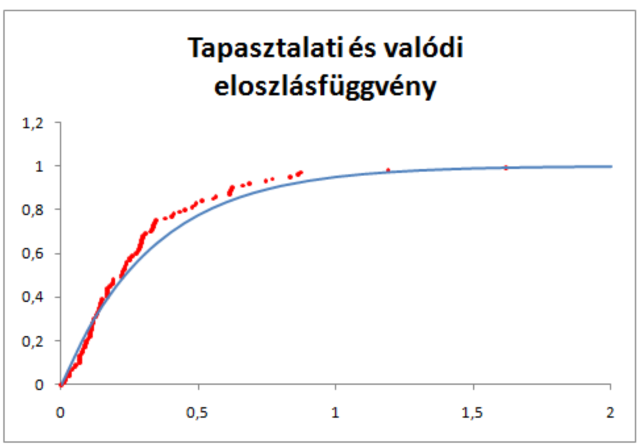
A 21 helyére értelemszerűen az a sorszám kerül, ameddig a `C` oszlopban vannak számok. Lépjen valamelyik *Sorozatok2* pontra, majd helyi menüben *Sorozat-diagramtípus módosítása*. Ezután

Pont (X, Y)/Pont görbített vonalakkal → *OK*

Lépjen a *Sorozatok2* vonalra, majd helyi menüben *Adatsorok formázása*. Ezután

Vonal színe/Folytonos vonal/Szín: kék → *Vonalstílus/Szélesség: 1,5 pt* → *Bezárás*

Klikkeljen a feliratra, és javítsa ki „Tapasztalati és valódi eloszlásfüggvény”-re.



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

2.2. Vonaldiagram

Diszkrét valószínűségi változóra vonatkozó $x_1, \ldots, x_n$ mintarealizáció esetén a tapasztalati eloszlás x_i-hez $(i = 1, \ldots, n)$ hozzárendeli az x_i-vel egyenlő elemek számát a mintarealizációban, elosztva n-nel. 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.

2.5. 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 az `A` oszlopba. Ezután minden mintarealizáció elemhez kiszámoljuk a tapasztalati eloszlás értéket. Ez a korábbi módszer logikájával

`=DARABTELI(A:A;"="&A1)/DARAB(A:A)`

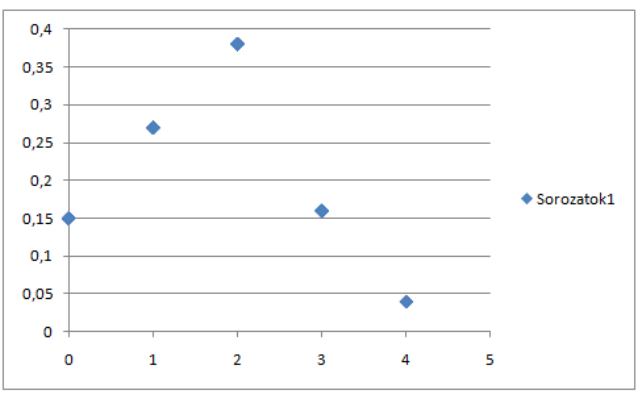
módon történhet. De ez ekvivalens a következő `B1` cellába írásával:

`=DARABTELI(A:A;A1)/DARAB(A:A)`.

Nyomjon *Enter*-t, majd a `B1` cella kitöltőjelére klikkeljen kétszer.

A következőkben ábrázolja az `(A1, B1), (A2, B2), ...` koordinátájú pontokat. Ehhez jelölje ki az `A` és `B` oszlopokat, majd

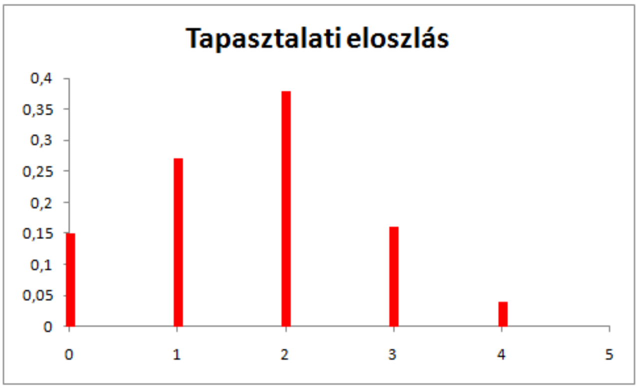
Beszúrás → Diagramok/Pont → Pont csak jelölőkkel



Ezután elkészítjük a vonaldiagramot. Jelenítsen meg hibasávokat százalékkal, törölje az X hibasávokat, majd az Y hibasávok formázásánál

Irány: Mínusz $\rightarrow$ *Végpont stílusa: Nyílt* $\rightarrow$
A hiba mértéke: Százalék: 100% $\rightarrow$ *Vonal színe* $\rightarrow$
Folytonos vonal $\rightarrow$ *Szín: piros* $\rightarrow$ *Vonalstílus* $\rightarrow$ *Szélesség: 5 pt* $\rightarrow$
Bezárás

Végül törölje a kék pontokat, a vezető rácsokat és a „Sorozatok1” feliratot, továbbá adja a diagramnak a „Tapasztalati eloszlás” címet.



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

Az Excelben járatosabb Olvasónak feltűnhet, hogy miért nem az oszlopdiagram típust választottuk az ábrázolásnál pontdiagram helyett, hiszen ekkor nem lenne szükség a hibasávokra. Ennek az az oka, hogy az Excel oszlopdiagram esetén a vízszintes tengelyen nem értékeket, hanem úgynevezett kategóriákat jelenít meg egymástól azonos távolságokra. Így ha a vizsgált diszkrét valószínűségi változó egymást követő lehetséges értékei nem azonos távolságokra vannak egymástól, akkor az oszlopdiagramos ábrázolás rossz megoldást adna, míg az előbb ismertetett megoldás akkor is helyes lenne.

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

Megoldás. Azt fogjuk felhasználni, hogy Excel-ben

$$\boxed{\texttt{BINOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}p\texttt{;HAMIS)}} = \binom{r}{k} p^k (1-p)^{r-k}.$$

Ha `HAMIS` helyett `IGAZ` kerül a képletbe, akkor ezzel a

$$\sum_{i=0}^{k} \binom{r}{i} p^i (1-p)^{r-i}$$

képlet számolható ki. Itt $r \in \mathbb{N},\ 0 \leqslant p \leqslant 1,\ k = 0, \ldots, r$.

A megoldást az előző munkalapon végezze el. Mivel a valódi eloszlás értelmezési tartománya $\{0, 1, 2, 3, 4, 5\}$, ezért a `C1`, `C2`, `C3`, `C4`, `C5`, `C6` cellákba rendre írja be a 0, 1, 2, 3, 4, 5 számokat. A `D1` cellába írja a következőt:

`=BINOM.ELOSZLÁS(C1;5;0,3;HAMIS)`

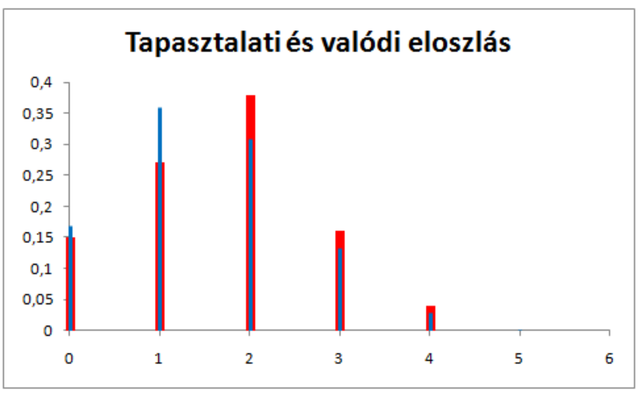
A `D1` cella kitöltőjelére klikkeljen kétszer. Most rátérünk a függvény ábrázolására. Helyi menüben válassza az *Adatok kijelölése* pontot.

Hozzáadás →
Adatsor X értékei: `=Munka1!$C$1:$C$6` →
Adatsor Y értékei: `=Munka1!$D$1:$D$6` → *OK* → *OK*

Jelenítsen meg hibasávokat százalékkal a Sorozatok2-höz, törölje az X hibasávokat, majd az Y hibasávok formázásánál

Irány: Mínusz → *Végpont stílusa: Nyílt* →
A hiba mértéke: Százalék: 100% → *Vonal színe* →
Folytonos vonal → *Szín: kék* → *Vonalstílus* → *Szélesség: 2 pt* →
Bezárás

Legvégül a feliratot változtassa „Tapasztalati és valódi eloszlás”-ra.



Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

2.3. Sűrűséghisztogram

Legyen $r \in \mathbb{N}$, $x_0, x_1, \ldots, x_r \in \mathbb{R}$ és $x_0 < x_1 < \cdots < x_r$. Tegyük fel, hogy a ξ-re vonatkozó $\xi_1(\omega), \ldots, \xi_n(\omega)$ 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 valódi f sűrűségfüggvény 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

$$\int\limits_{x_{j-1}}^{x_j} f(x)\,\mathrm{d}x = \mathrm{P}(x_{j-1} \leqslant \xi < x_j) \simeq \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{x_{j-1} \leqslant \xi_i(\omega) < x_j} = y_j(x_j - x_{j-1}),$$

melyből

$$y_j = \frac{\sum_{i=1}^{n} \mathrm{I}_{x_{j-1} \leqslant \xi_i(\omega) < x_j}}{n(x_j - x_{j-1})} \quad (j = 1, \ldots, r).$$

A kapott oszlopdiagramot *sűrűséghisztogramnak* nevezzük, amely tehát a valódi f sűrűségfüggvényt a j-edik részintervallumon az y_j konstanssal közelíti.

2.7. 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 és rögzítse az `A` oszlopba a korábban tanult képlettel:

`=GYÖK(-2*LN(VÉL()))*COS(2*PI()*VÉL())`.

A `B` oszlopba írja be az osztópontokat (egy részintervallum hossza $\frac{4-(-4)}{10} = 0{,}8$). `B1`-be írjon -4-et, `B2`-be pedig `=B1+0,8`-at, majd a kitöltőjelet húzza le a 11. sorig (mert itt lesz az értéke 4).

Ezután `C1`-be számolja ki a sűrűséghisztogram $[-4; -3{,}2)$ fölötti téglalapjának magasságát a következő képlettel:

`=DARABHATÖBB(A:A;">="&B1;A:A;"<"&B2)/(0,8*DARAB(A:A))`.

(A `DARABHATÖBB` függvény leírását olvassa el a súgóban.) A `C1` cella kitöltőjelére klikkeljen kétszer. Jelölje ki a `B1:C10` cellatartományt.

Beszúrás/Diagramok/Pont/Pont csak jelölőkkel

Ennek hatására megjelennek a sűrűséghisztogram téglalapjainak a bal felső pontjai. Törölje a „Sorozatok1” feliratot és a vezető rácsokat. Húzza meg a téglalapok bal oldalát és a tetejét.

Elrendezés/Elemzés/Hibasávok/Hibasávok standard hibával →
Aktuális kijelölés/Sorozatok1 X hibasávok → Kijelölés formázása →
Irány/Plusz → Végpont stílusa/Nyílt → A hiba mértéke/Abszolút értékben: 0,8 →
Vonal színe/Folytonos vonal/Szín: piros → Bezárás

Elrendezés/Aktuális kijelölés/Sorozatok1 Y hibasávok → Kijelölés formázása →
Irány/Mínusz → Végpont stílusa/Nyílt → A hiba mértéke/Százalék: 100% →
Vonal színe/Folytonos vonal → Bezárás

A következő lépésben a téglalapok jobb felső pontjait ábrázolja.

Helyi menü/Adatok kijelölése → Hozzáadás →
Adatsor X értékei: `=Munka1!$B$2:$B$11` →
Adatsor Y értékei: `=Munka1!$C$1:$C$10` → *OK* → *OK*

Húzza meg a téglalapok jobb oldalait.

Elrendezés/Aktuális kijelölés/Sorozatok2 →
Elemzés/Hibasávok/Hibasávok standard hibával →
Aktuális kijelölés/Sorozatok2 X hibasávok → Delete gomb →
Aktuális kijelölés/Sorozatok2 Y hibasávok → Kijelölés formázása →
Irány/Mínusz → Végpont stílusa/Nyílt → A hiba mértéke/Százalék: 100% →
Vonal színe/Folytonos vonal → Bezárás

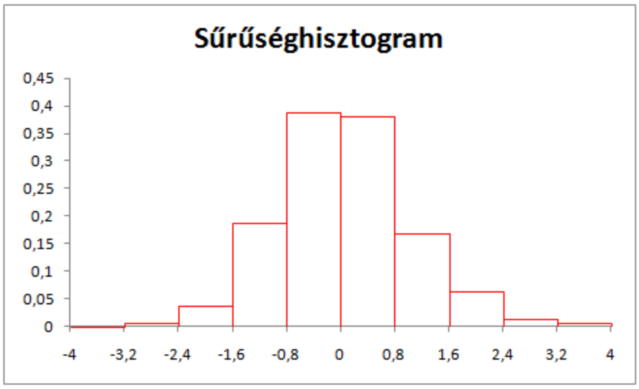
Rejtse el a Sorozatok1 jelölőit.

Elrendezés/Aktuális kijelölés/Sorozatok1 → Kijelölés formázása →
Jelölő beállításai → Jelölő típusa/Nincs → Bezárás

Hasonlóan rejtse el a Sorozatok2 jelölőit is. Ezután még a vízszintes tengelyen állítson be néhány dolgot. Klikkeljen a vízszintes tengely valamely értékére, majd

Helyi menü/Tengely formázása → Minimum/Rögzített: -4 →
Maximum/Rögzített: 4 → Fő lépték/Rögzített: 0,8 →
Függőleges tengely metszéspontja/Ezen értéknél: -4 → Bezárás

Végül adja a diagramnak „Sűrűséghisztogram” címet. A következő eredményt kapjuk.



A megoldást végigkövetheti a következő videón.

VIDEÓ

2.8. 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 `=E1+0,2`. Az `E2` cella kitöltőjelét húzza le a 4 értékig (41. sorig). Ezután az `F1` cellában számolja ki a standard normális eloszlás sűrűségfüggvényének értékét az `E1` cella értékénél. Ennek érdekében írja `F1`-be:

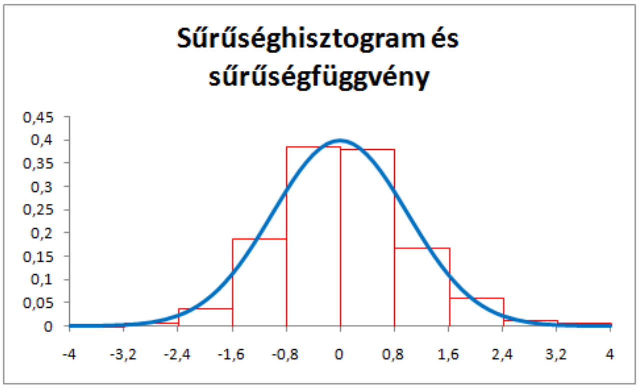
```
=NORM.ELOSZL(E1;0;1;HAMIS)
```

Itt 0 a várható értéket, míg 1 a szórást jelenti. Ha HAMIS helyett az IGAZ logikai értéket írjuk be, akkor az eloszlásfüggvényt számoljuk. Az `F1` cella kitöltőjelére klikkeljen kétszer.

A következőkben megrajzoljuk a valódi sűrűségfüggvényt. Lépjen a diagram területére, majd helyi menüben

Adatok kijelölése → *Hozzáadás* →
Adatsor X értékei: `=Munka1!$E$1:$E$41` →
Adatsor Y értékei: `=Munka1!$F$1:$F$41` → *OK* → *OK* →
Elrendezés/Sorozatok3 → *Tervezés/Más diagramtípus* →
Pont (X, Y)/Pont görbített vonalakkal → *OK*

Végül a kapott függvény színét állítsa kékre és adja a diagramnak a „Sűrűséghisztogram és sűrűségfüggvény” címet.



A megoldást végigkövetheti a következő videón.

VIDEÓ

2.4. Gyakorlatok

2.1. gyakorlat. A matematikai statisztika alaptörvényét többféle eloszlással is bemutatjuk a következő videóban.

VIDEÓ

Az itt használt program letölthető a következő helyről:

PROGRAM

Vizsgálja meg Ön is ezzel a programmal néhány esetben a tapasztalati eloszlásfüggvény konvergenciáját.

2.2. 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.

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

Útmutatás. A vizsgált valószínűségi változót jelölje ξ. Értékkészlete $\{0, 1, \dots, 6\}$, 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

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

így

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

2.4. 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

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

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

2.5. gyakorlat. Generáljon az Ön által készített (vagy a letöltött) programmal $\lambda = 3{,}2$ 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

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

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

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

értékét számolja ki.

2.6. 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 az Ön által készített (vagy a letöltött) programmal 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

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

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

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

2.7. 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

$$\boxed{\texttt{NORM.ELOSZL(}x\texttt{;}m\texttt{;}\sigma\texttt{;IGAZ)}}$$

Itt jegyezzük meg, hogy ha speciálisan $m = 0$ és $\sigma = 1$, azaz standard normális az eloszlás, akkor $\boxed{\texttt{NORM.ELOSZL(}x\texttt{;0;1;IGAZ)}}$ helyett használható a következő is:

$$\boxed{\texttt{STNORMELOSZL(}x\texttt{)}}.$$

(2) Az s szabadsági fokú khi-négyzet eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \geqslant 0$ helyen

$$\boxed{\texttt{1-KHI.ELOSZLÁS(}x\texttt{;}s\texttt{)}}.$$

(3) Az s szabadsági fokú t-eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \geqslant 0$ helyen

$$\boxed{\texttt{1-T.ELOSZLÁS(}x\texttt{;}s\texttt{;1)}}$$

illetve $x < 0$ esetén

$$\boxed{\texttt{T.ELOSZLÁS(}-x\texttt{;}s\texttt{;1)}}.$$

Itt jegyezzük meg, hogy ilyen eloszlású ξ esetén, ha $x \geqslant 0$, akkor

$$P(\xi > x) = \boxed{\texttt{T.ELOSZLÁS(}x\texttt{;}s\texttt{;1)}} \text{ (egyszélű eloszlás)}$$
$$P(|\xi| > x) = \boxed{\texttt{T.ELOSZLÁS(}x\texttt{;}s\texttt{;2)}} \text{ (kétszélű eloszlás).}$$

(4) Az s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változó eloszlásfüggvényének az értéke $x \geqslant 0$ helyen

$$\boxed{\texttt{1-F.ELOSZLÁS(}x\texttt{;}s_1\texttt{;}s_2\texttt{)}}.$$

2.8. 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 eloszlásfüggvényének értéke az $x \geqslant 0$ helyen

$$\boxed{\texttt{EXP.ELOSZLÁS(}x\texttt{;}\lambda\texttt{;HAMIS)}}.$$

2.9. 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) 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

$$\boxed{\texttt{GAMMA.ELOSZLÁS(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;IGAZ)}}$$

illetve sűrűségfüggvénye

$$\boxed{\texttt{GAMMA.ELOSZLÁS(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;HAMIS)}}$$

függvényekkel számolható, ha $x \geqslant 0$.
(3) Cauchy-eloszlású valószínűségi változó sűrűségfüggvénye

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

illetve eloszlásfüggvénye

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

Itt az $\operatorname{arctg}(x)$ az $\boxed{\texttt{ARCTAN(}x\texttt{)}}$ függvénnyel számolható. De azt is felhasználhatjuk, hogy a Cauchy-eloszlás megegyezik az 1 szabadsági fokú t-eloszlással.
(4) Használja fel, hogy az s szabadsági fokú khi-négyzet eloszlás megegyezik az $r = \frac{s}{2}$ rendű $\lambda = \frac{1}{2}$ paraméterű gamma-eloszlással.

2.10. 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. s szabadsági fokú t-eloszlású valószínűségi változó sűrűségfüggvénye

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

illetve s_1 és s_2 szabadsági fokú F-eloszlású valószínűségi változó sűrűségfüggvénye

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} 0, & \text{ha } x \leqslant 0, \\ \frac{\Gamma\left(\frac{s_1+s_2}{2}\right)}{\Gamma\left(\frac{s_1}{2}\right)\Gamma\left(\frac{s_2}{2}\right)} \sqrt{\frac{s_1^{s_1} s_2^{s_2} x^{s_1-2}}{(s_1 x+s_2)^{s_1+s_2}}}, & \text{ha } x > 0, \end{cases}$$

ahol $\Gamma(x) = \int\limits_0^\infty u^{x-1} e^{-u} \, \mathrm{d}u = \boxed{\texttt{KITEVŐ(GAMMALN(}x\texttt{))}}$ $(x > 0)$ az úgynevezett gamma-függvény.

3. Grafikus illeszkedésvizsgálat

Ebben a fejezetben azt vizsgáljuk, hogyan lehet grafikus úton eldönteni a vizsgált valószínűségi változóról, hogy milyen eloszláscsaládba tartozik.

3.1. Általános vizsgálat

Legyen $r \in \mathbb{N}$, $x_1, \ldots, x_r \in \mathbb{R}$ és $x_1 < x_2 < \cdots < x_r$. A matematikai statisztika alaptétele szerint a tapasztalati eloszlásfüggvény nagy elemszámú minta esetén jól közelíti a valódi eloszlásfüggvényt, azaz ha n a minta elemszáma, akkor

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

ahol F a vizsgált valószínűségi változó valódi eloszlásfüggvénye és n nagy. Ebből F 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_i, y_i) $(i = 1, \ldots, r)$ koordinátájú pontok körülbelül egy egyenesre esnek.

3.2. Grafikus normalitásvizsgálat

Az előző módszert most speciálisan a normális eloszlásra alkalmazzuk.

3.1. Példa. A `minta-01.txt` fájlban található mintarealizáció alapján nézze meg, hogy a vizsgált valószínűségi változó lehet-e normális eloszlású.

Megoldás. Legyen $r \in \mathbb{N}$, $x_1, \ldots, x_r \in \mathbb{R}$ és $x_1 < x_2 < \cdots < x_r$. Jelölje n a mintarealizáció elemeinek a számát és k_i az x_i-nél kisebb elemek számát a mintarealizációban. Ekkor $F_n^*(x_i) = \frac{k_i}{n}$. Ha teljesül, hogy a vizsgált valószínűségi változó normális eloszlású m várható értékkel és σ szórással, akkor

$$\frac{k_i}{n} \simeq \Phi\left(\frac{x_i - m}{\sigma}\right), \quad i = 1, \ldots, r,$$

azaz

$$\Phi^{-1}\left(\frac{k_i}{n}\right) \simeq \frac{1}{\sigma}x_i - \frac{m}{\sigma}, \quad i = 1, \ldots, r.$$

Így $y_i := \Phi^{-1}\left(\frac{k_i}{n}\right)$ jelöléssel az (x_i, y_i) $(i = 1, \ldots, r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $\frac{1}{\sigma}$ a meredeksége és $-\frac{m}{\sigma}$ értéknél metszi a függőleges tengelyt.

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 tekinthetjük például az $\mathbb{R}$ következő beosztását: $x_1 = 3{,}5$; $x_2 = 4$; $x_3 = 4{,}5$; $x_4 = 5$; $x_5 = 5{,}5$; $x_6 = 6$; $x_7 = 6{,}5$; $x_8 = 7$; $x_9 = 7{,}5$. Ezeket az értékeket írja be a `B1-B9` cellákba.

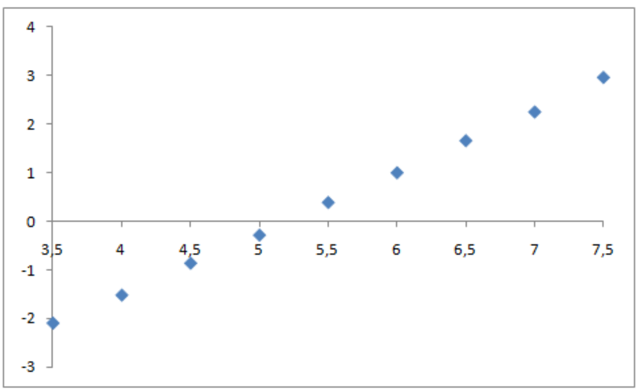
Ezután a `C1-C9` cellákban számolja ki az $y_i = \Phi^{-1}\left(\frac{k_i}{n}\right)$ értékeket. Excelben a $\Phi(x)$ értékét `INVERZ.STNORM(`x`)` függvénnyel számolhatjuk ki minden $x \in \mathbb{R}$ esetén. Ennek alapján a `C1` cellába írja be, hogy

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

majd a kitöltőjelre klikkeljen kétszer. Következhet az ábrázolás. Jelölje ki `B1-C9` cellatartományt, majd

Beszúrás → Diagramok/Pont → Pont csak jelölőkkel

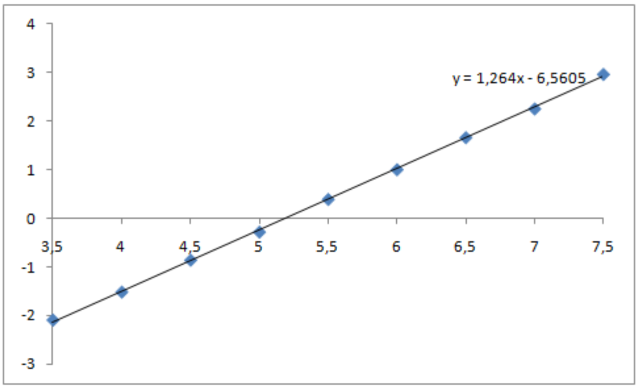
törölje a vezető rácsokat és a „Sorozatok1” feliratot, majd a vízszintes tengelyen rögzítse a minimális értéket 3,5-nek, a maximális értéket pedig 7,5-nek.



Amint látható a 9 darab pont nagyon jó közelítéssel egy egyenesen helyezkedik el, így normális eloszlásúnak tekinthetjük a vizsgált valószínűségi változót.

A továbbiakban ebből lehetőségünk van megbecsülni a normális eloszlás paramétereit, hiszen az egyenes meredeksége körülbelül $\frac{1}{\sigma}$ illetve körülbelül $-\frac{m}{\sigma}$ értéknél metszi a függőleges tengelyt. Ehhez először azt kell eldönteni, hogy a 9 darab pontra melyik egyenes illeszkedik a legjobban. Az elfogadott kritérium az úgynevezett *legkisebb négyzetek módszere*, mely szerint azt az egyenest tekintjük, melytől a pontok távolságainak négyzetösszege minimális. Ezt *lineáris trendvonalnak* vagy *lineáris regressziónak* is nevezik. Az Excelben ez egyszerűen ábrázolható. Klikkeljen valamelyik kék pontra, majd a helyi menüben válassza a *Trendvonal felvétele* pontot.

Pipálja ki az *Egyenlet látszik a diagramon* lehetőséget, majd nyomja meg a *Bezárás* gombot.



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

```
=1/MEREDEKSÉG(C1:C9;B1:B9)
```

és m becslése

```
=-METSZ(C1:C9;B1:B9)/MEREDEKSÉG(C1:C9;B1:B9)
```

Az eredmény $\sigma \simeq 0{,}7911$ és $m \simeq 5{,}1902$. Ellenőrzésképpen közöljük, hogy a felhasznált mintarealizáció $m = 5{,}2$ és $\sigma = 0{,}8$ paraméterű normális eloszlásból származik. Az egész megoldást végigkövetheti a következő videón.

VIDEÓ

3.3. Grafikus exponencialitásvizsgálat

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

Megoldás. Az előző megoldás jelöléseit használva, ha teljesül, hogy a vizsgált valószínűségi változó exponenciális eloszlású λ paraméterrel, akkor

$$\frac{k_i}{n} \simeq 1 - e^{-\lambda x_i}, \quad i = 1, \ldots, r,$$

azaz

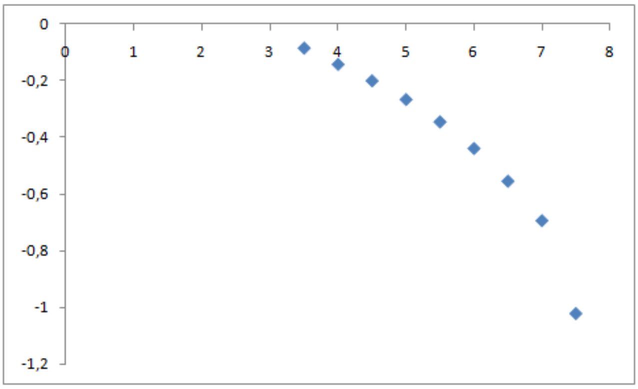
$$\ln\left(1 - \frac{k_i}{n}\right) \simeq -\lambda x_i, \quad i = 1, \ldots, r.$$

Így $y_i := \ln\left(1 - \frac{k_i}{n}\right)$ jelöléssel az (x_i, y_i) $(i = 1, \ldots, r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $-\lambda$ a meredeksége és átmegy az origón.

A mintarealizációt másolja egy Excel-munkalap `A` oszlopába. Vizsgálja meg a legkisebb és legnagyobb értékét a mintarealizációnak. Azt kapjuk, hogy 2,5002 a legkisebb és 7,9942 a legnagyobb érték. Így használhatjuk az előző megoldásbeli beosztást, melyet a `B` oszlopba írjunk. Ezután a `C1` cellába írja be, hogy

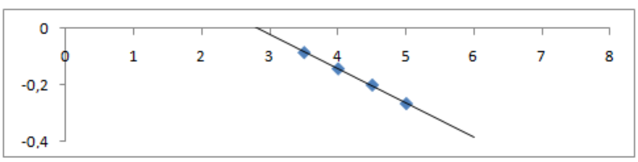
```
=LN(1-DARABTELI(A:A;"<"&B1)/DARAB(A:A))
```

majd a kitöltőjelre klikkeljen kétszer. Az ábrázolást az előző megoldáshoz hasonlóan végezheti el.



A kapott ábra azt mutatja, hogy a pontok inkább valamilyen ívelt görbén helyezkednek el, mintsem egy egyenesen, ezért nagy biztonsággal állíthatjuk, hogy a vizsgált valószínűségi változó az adott mintarealizáció alapján nem exponenciális eloszlású.

Ha csak az első négy pontot hagyjuk meg, akkor az már jó közelítéssel elhelyezhető egy egyenesen, de ez az egyenes messze halad el az origótól, így pusztán ezen pontok figyelembevételével is azt állíthatjuk, hogy a minta nem exponenciális eloszlásból származik.



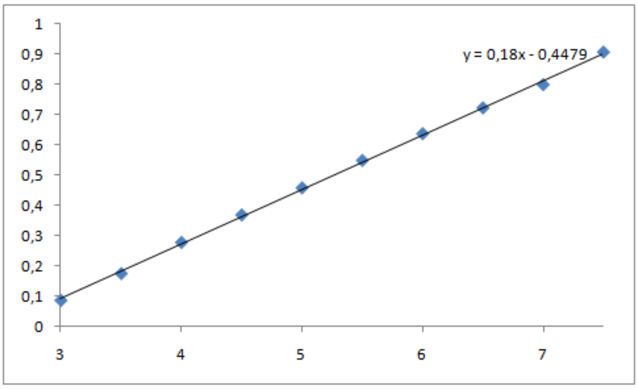
3.4. Gyakorlatok

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

$$\frac{k_i}{n} \simeq \frac{x_i - a}{b - a} = \frac{1}{b-a} x_i - \frac{a}{b-a}, \quad i = 1, \dots, r.$$

Ha például az $x_1 = 3$; $x_2 = 3{,}5$; $x_3 = 4$; $x_4 = 4{,}5$; $x_5 = 5$; $x_6 = 5{,}5$; $x_7 = 6$; $x_8 = 6{,}5$; $x_9 = 7$; $x_{10} = 7{,}5$ beosztást használjuk, akkor a következő ábrát kapjuk:

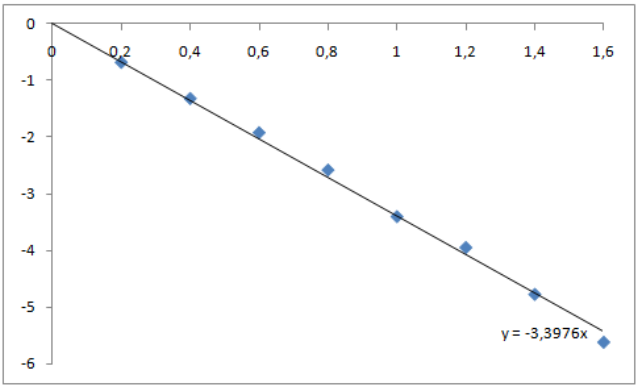


Így nagy valószínűséggel mondhatjuk, hogy a minta egyenletes eloszlásból származik. A `MEREDEKSÉG` és `METSZ` függvények segítségével az a becslése 2,4881 illetve a b becslése 8,0430. Összehasonlításként közöljük, hogy a vizsgált valószínűségi változó egyenletes eloszlású volt a $[2{,}5; 8]$ intervallumon.

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

Útmutatás. A grafikus exponencialitásvizsgálatban leírtak szerint járjon el. Használhatja például az $x_1 = 0{,}2$; $x_2 = 0{,}4$; $x_3 = 0{,}6$; $x_4 = 0{,}8$; $x_5 = 1$; $x_6 = 1{,}2$; $x_7 = 1{,}4$; $x_8 = 1{,}6$ beosztást. Ekkor a kapott pontok nagyon jól illeszkednek egy olyan egyenesre, amely átmegy az origón. Így nagy biztonsággal állíthatjuk, hogy a mintarealizáció exponenciális eloszlásból származik.

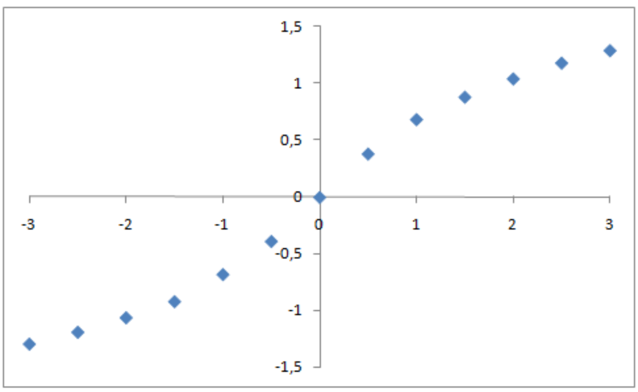
A λ paraméter becslésénél továbbra is azt az egyenest keressük, amely a legkisebb négyzetek módszerével adódik, de most a vizsgálandó egyenesek körét leszűkíthetjük azokra, amelyek átmennek az origón. Ezt úgy tehetjük meg, ha a trendvonal beállításánál kipipáljuk a *Metszéspont: 0* opciót.



A meredekségből tehát látható, hogy λ becslése 3,3976. Összehasonlításképpen közöljük, hogy a vizsgált valószínűségi változó exponenciális eloszlású volt $\lambda = 3{,}2$ paraméterrel.

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.

Útmutatás. A grafikus normalitásvizsgálatban leírtak szerint járjon el. Használhatja például a -3-tól 3-ig terjedő egyenletes beosztást 0,5 hosszúságú részintervallumokkal. Ekkor a következőt kapjuk:



Ebből egyértelműen látható, hogy a minta nem normális eloszlásból származik.

3.4. gyakorlat. A `minta-04.txt` fájlban található mintarealizáció alapján vizsgálja meg, hogy a vizsgált valószínűségi változó lehet-e Cauchy-eloszlású.

Útmutatás. Ha teljesül, hogy a vizsgált valószínűségi változó Cauchy-eloszlású, akkor

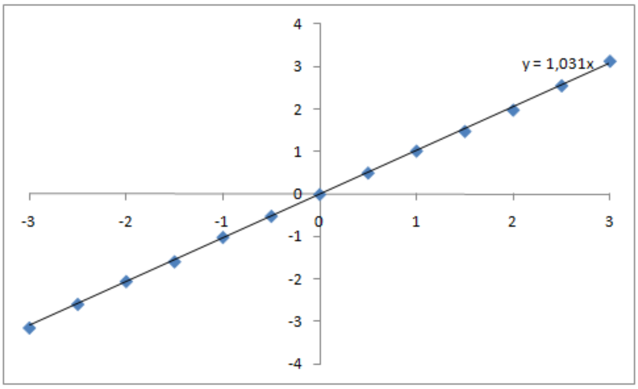
$$\frac{k_i}{n} \simeq \frac{1}{\pi} \operatorname{arctg} x_i + \frac{1}{2}, \quad i = 1, \ldots, r,$$

azaz

$$\operatorname{tg}\left(\left(\frac{k_i}{n}-\frac{1}{2}\right)\pi\right) \simeq x_i, \quad i=1,\ldots,r.$$

Így $y_i := \operatorname{tg}\left(\left(\frac{k_i}{n}-\frac{1}{2}\right)\pi\right)$ jelöléssel az (x_i, y_i) $(i=1,\ldots,r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek 1 a meredeksége és átmegy az origón. Excelben a $\operatorname{tg} x$ az $\boxed{\texttt{TAN}(x)}$ függvénnyel számolható, ahol x radiánban van megadva.

Ismét használjuk a -3-tól 3-ig terjedő egyenletes beosztást 0,5 hosszúságú részintervallumokkal. Ekkor a következőt kapjuk:



A kapott egyenes 1,031 meredekségű, ami jó közelítéssel 1, így nagy valószínűséggel állítható, hogy a vizsgált valószínűségi változó Cauchy-eloszlású.

3.5. 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.

4. Statisztikák

Ebben a fejezetben azt vizsgáljuk, hogyan lehet a különböző statisztikákat kiszámolni Excelben.

4.1. 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:

	A	B	C
1	7,46	minta elemszáma	=DARAB(A:A)
2	4,1276	mintaterjedelem	=MAX(A:A)-MIN(A:A)
3	7,4835	terjedelemközép	=(MIN(A:A)+MAX(A:A))/2
4	5,3306	mintaátlag	=ÁTLAG(A:A)
5	4,6688	tapasztalati szórás	=SZÓRÁSP(A:A)
6	5,0706	tapasztalati szórásnégyzet	=VARP(A:A)
7	7,496	korrigált tapasztalati szórás	=SZÓRÁS(A:A)
8	3,8952	korrigált tapasztalati szórásnégyzet	=VAR(A:A)
9	6,5811	tapasztalati medián	=MEDIÁN(A:A)

	A	B	C	D	E
1	7,46	minta elemszáma	1500		
2	4,1276	mintaterjedelem	5,494		
3	7,4835	terjedelemközép	5,2472		
4	5,3306	mintaátlag	5,263688		
5	4,6688	tapasztalati szórás	1,60214677		
6	5,0706	tapasztalati szórásnégyzet	2,56687427		
7	7,496	korrigált tapasztalati szórás	1,60268108		
8	3,8952	korrigált tapasztalati szórásnégyzet	2,56858666		
9	6,5811	tapasztalati medián	5,25955		

Természetesen

$$\boxed{\texttt{VARP(A:A)}} = \boxed{\texttt{SZÓRÁSP(A:A)\^{}2}}$$
$$\boxed{\texttt{VAR(A:A)}} = \boxed{\texttt{SZÓRÁS(A:A)\^{}2}}.$$

4.2. Példa. A `minta-05.txt` fájlban található mintarealizáció esetén számolja ki a tapasztalati móduszt.

Megoldás. A mintarealizációt másolja az `A` oszlopba. Ekkor a tapasztalati módusz értéke a `=MÓDUSZ(A:A)` függvénnyel számolható ki, amely most 2-vel egyenlő.

4.3. 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 klikkeljen kétszer.

A `C1` cellába írja a következőt: `=(A1-ÁTLAG(A:A))^3`. A `C1` cella kitöltőjelére klikkeljen 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, majd

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

Ezután a rendezett mintát az `F` oszlop tartalmazza.

4.4. Példa. Tekintsük a következő kétdimenziós valószínűségi vektorváltozóra vonatkozó mintarealizációt:

$$(0{,}12; 1{,}03)\ (0{,}63; 0{,}13)\ (0{,}44; 1{,}50)\ (0{,}50; 1{,}21)\ (0{,}66; 1{,}73)$$
$$(0{,}81; 1{,}13)\ (0{,}91; 1{,}73)\ (0{,}96; 0{,}65)\ (0{,}41; 1{,}10)\ (0{,}67; 0{,}01).$$

Számolja ki a tapasztalati kovarianciát és korrelációs együtthatót.

Megoldás. A rendezett számpárok első elemeit tegye az `A` oszlopba, a második elemeket pedig a `B` oszlopba. A tapasztalati kovarianciát a `KOVAR`, míg a tapasztalati korrelációs együtthatót a `KORREL` függvénnyel számolhatja ki az alábbiak szerint:

	A	B	C	D
1	0,12	1,03	tapasztalati kovariancia	=KOVAR(A:A;B:B)
2	0,63	0,13	tapasztalati korrelációs együttható	=KORREL(A:A;B:B)
3	0,44	1,5		
4	0,5	1,21		
5	0,66	1,73		
6	0,81	1,13		
7	0,91	1,73		
8	0,96	0,65		
9	0,41	1,1		
10	0,67	0,01		

	A	B	C	D
1	0,12	1,03	tapasztalati kovariancia	-0,006082
2	0,63	0,13	tapasztalati korrelációs együttható	-0,044347395
3	0,44	1,50		
4	0,50	1,21		
5	0,66	1,73		
6	0,81	1,13		
7	0,91	1,73		
8	0,96	0,65		
9	0,41	1,10		
10	0,67	0,01		

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 elemeit jelenti, és $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Útmutatás. Használjuk rendre a következő függvényeket:

`SZUM`

`NÉGYZETÖSSZEG`

`SQ`

`ÁTL.ELTÉRÉS`

`SZORZAT`.

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. 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")`.

4.4. gyakorlat. Adja meg az $\mathrm{I}_{3<\xi\leqslant 4}$ 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 klikkeljen kétszer.

4.5. gyakorlat. Adja meg a ξ_{10}^* és ξ_{n-5}^* értékeit, ahol a ξ-re vonatkozó mintarealizáció a `minta-02.txt` fájlban található, továbbá $\xi_1^*, \ldots, \xi_n^*$ a rendezett mintát jelöli. A rendezett mintának hányadik eleme a mintarealizáció 5. eleme? A mintarealizációnak hányadik legnagyobb eleme 7,963?

Útmutatás. A mintarealizációt másolja az `A` oszlopba. Ekkor
$\xi_k^* =$ `KICSI(A:A;k)`
$\xi_{n-k}^* =$ `NAGY(A:A;k+1)`
$\min\{\, k : \xi_k^* = \xi_i \,\} =$ `SORSZÁM(ξ_i;A:A;1)`
$\min\{\, k : \xi_{n-k}^* = \xi_i \,\} + 1 =$ `SORSZÁM(ξ_i;A:A;0)`.

4.6. 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:
`PERCENTILIS(A:A;t)`, `KVARTILIS(A:A;1)`, `KVARTILIS(A:A;3)`.

4.7. 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:

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

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 van egy `FERDESÉG` függvény, de ez más becslést használ az eloszlás ferdeségére:

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

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:

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

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 $2956{,}93809$.

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

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

Útmutatás. Használja rendre a következő függvényeket:
`SZORZATÖSSZEG`, `SZUMXBŐLY2`, `SZUMX2BŐLY2`, `SZUMX2MEGY2`.

5. Intervallumbecslések

Legyen ξ a vizsgált valószínűségi változó, amelyre a mintát vonatkoztatjuk. A ξ eloszlásának legyen ϑ egy ismeretlen becsülendő paramétere. Intervallumbecslésnél egy olyan intervallumot adunk meg, amelybe a ϑ valódi értéke nagy valószínűséggel beleesik. Ezen intervallum alsó és felső végpontját egy-egy statisztika realizációjával adjuk meg. A becslő intervallumot *konfidenciaintervallumnak* nevezzük. Ennek a biztonsági szintje az az érték, amelynél nagyobb vagy egyenlő valószínűséggel teljesül, hogy ϑ a konfidenciaintervallumba esik.

5.1. Normális eloszlás paramétereinek becslése

5.1. Példa. A `minta-06.txt` fájlban található mintarealizációról grafikus illeszkedésvizsgálattal győződjön meg, hogy normális eloszlásból származik. Tudjuk, hogy a szórás 0,7. Adjon a várható értékre 0,99 biztonsági szintű konfidenciaintervallumot.

Megoldás. A grafikus illeszkedést hasonlóan csináljuk mint korábban, így itt már nem részletezzük. Legyen ξ az a valószínűségi változó, amelyre a mintarealizáció vonatkozik és n a mintaelemeszám. Ismert, hogy $1-\alpha = 0{,}99$, $\sigma = 0{,}7$

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

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

A mintarealizációt másolja az `A` oszlopba. Ezután vegye fel az alapadatokat, számolja ki α-t, a mintaelemszámot és a mintaátlagot a következő ábrának megfelelően:

	A	B	C
1	17,2095	biztonsági szint =	0,99
2	16,087	szórás =	0,7
3	14,6349	alfa =	=1-C1
4	15,7631	mintaelemszám =	=DARAB(A:A)
5	15,1532	mintaátlag =	=ÁTLAG(A:A)
6	13,1924		

A szórás, α, mintaelemszám és mintaátlag értékeit tartalmazó cellákat nevezze el rendre `szórás`, `alfa`, `n`, `átlag` módon. Ehhez lépjen az adott cellára, majd a szerkesztőléc mellett balra található *név mezőbe* írja a megfelelő nevet. Végül üssön *Enter*-t.

Ezután következhet a várható értékre vonatkozó konfidenciaintervallum alsó és felső végpontjának a kiszámítása. Ehhez tudnunk kell, hogy

$$\Phi^{-1}(x) = \boxed{\texttt{INVERZ.STNORM}(x)} \quad (0 < x < 1).$$

Így az alsó végpont

```
=átlag-INVERZ.STNORM(1-alfa/2)*szórás/GYÖK(n)
```

módon, míg a felső végpont

```
=átlag+INVERZ.STNORM(1-alfa/2)*szórás/GYÖK(n)
```

módon számolható. A számolás kicsit egyszerűbben is elvégezhető, ha tudjuk, hogy Excelben

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

Ebben az esetben az alsó végpont

```
=átlag-MEGBÍZHATÓSÁG(alfa;szórás;n)
```

módon, míg a felső végpont

```
=átlag+MEGBÍZHATÓSÁG(alfa;szórás;n)
```

módon számolható. A kapott eredményeket kerekítsük négy tizedesjegyre. A végeredmény a következő ábrán látható:

	A	B	C
1	17,2095	biztonsági szint =	0,99
2	16,087	szórás =	0,7
3	14,6349	alfa =	0,01
4	15,7631	mintaelemszám =	456
5	15,1532	mintaátlag =	15,28107785
6	13,1924	**A várható értékre vonatkozó konfidenciaintervallum**	
7	17,9192	**alsó végpontja =**	15,1966
8	16,7218	**felső végpontja =**	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. (Erről grafikus illeszkedésvizsgálattal meggyőződhet.) Tudjuk, hogy a várható érték 1251. Adjon a szórásra 0,9 biztonsági szintű konfidenciaintervallumot.

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

$$
\begin{aligned}
&F \sim \mathbf{Khi}(n) \\
&c_1 := F^{-1}\left(\frac{\alpha}{2}\right) \\
&c_2 := F^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&\tau_1 := \sqrt{\frac{\sum_{i=1}^n (\xi_i - m)^2}{c_2}} \\
&\tau_2 := \sqrt{\frac{\sum_{i=1}^n (\xi_i - m)^2}{c_1}}
\end{aligned}
$$

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

A mintarealizációt másolja az `A` oszlopba. A `B` oszlopot hagyja üresen. Írja be az alapadatokat (biztonsági szint, várható érték), majd számolja ki a mintaelemszámot és az α-t. A várható érték, mintaelemszám és α értékeit tartalmazó cellákat nevezze el rendre `m`, `n`, `alfa` módon.

Ezután a `B1` cellába írja be, hogy `=m`, majd a kitöltőjelre klikkeljen kétszer. Erre azért van szükség, mert így a $\sum_{i=1}^n (\xi_i - m)^2$ realizációja

`SZUMXBŐLY2(A:A;B:B)`

módon kiszámolható. Most számolja ki c_1 és c_2 értékeit. Ehhez tudnunk kell, hogy $F \sim \mathbf{Khi}(n)$ esetén

$$F^{-1}(x) = \boxed{\texttt{INVERZ.KHI}(1 - x;n)} \quad (0 < x < 1).$$

Így $c_1 =$ `INVERZ.KHI(1-alfa/2;n)` és $c_2 =$ `INVERZ.KHI(alfa/2;n)`. A c_1 és c_2 értékeket tartalmazó cellákat nevezze el az egyszerűség kedvéért rendre `c_1` és `c_2` módon. Ezután következhet a szórásra vonatkozó konfidenciaintervallum alsó és felső végpontjának a kiszámítása. Az alsó végpont

`=GYÖK(SZUMXBŐLY2(A:A;B:B)/c_2)`

módon, míg a felső végpont

`=GYÖK(SZUMXBŐLY2(A:A;B:B)/c_1)`

módon számolható. A kapott eredményeket kerekítsük négy tizedesjegyre. A végeredmény a következő ábrán látható:

	A	B	C	D	E	F
1	1250,6	1251	biztonsági szint =	0,9		
2	1246,8	1251	várható érték =	1251		
3	1247,8	1251	mintaelemszám =	891		
4	1250,1	1251	alfa =	0,1		
5	1256	1251	c_1 =	822,7201		
6	1249,1	1251	c_2 =	961,5537		
7	1249,2	1251	**A szórásra vonatkozó konfidenciaintervallum**			
8	1249,3	1251	**alsó végpontja =**	3,1223		
9	1246,8	1251	**felső végpontja =**	3,3754		

Ellenőrzésképpen közöljük, hogy a szórás valódi értéke 3,2.

5.2. Valószínűség becslése

5.3. 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 $1 - \alpha = 0{,}98$

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

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

Írja be az alapadatokat (n, k, biztonsági szint), majd számolja ki az α-t. Az n, k és α értékeit tartalmazó cellákat nevezze el rendre `n`, `k`, `alfa` módon.

Ezután következhet a p-re vonatkozó konfidenciaintervallum alsó és felső végpontjának a kiszámítása. Ehhez szükség van a `KRITBINOM` függvényre, melynek jelentése

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

Ebből következően

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

Így az alsó végpont

```
=(KRITBINOM(n;k/n;alfa/2)-1)/n
```

módon, míg a felső végpont

```
=KRITBINOM(n;k/n;1-alfa/2)/n
```

módon számolható. A kapott eredményeket kerekítsük négy tizedesjegyre. A végeredmény a következő ábrán látható:

	A	B	C
1	n =	55	
2	k =	39	
3	biztonsági szint =	0,98	
4	alfa =	0,02	
5	**A p-re vonatkozó konfidenciaintervallum**		
6	**alsó végpontja =**	0,5455	
7	**felső végpontja =**	0,8364	

Ellenőrzésképpen közöljük, hogy a feladatban megadott $k = 39$ gyakoriság, egy $p = 0{,}72$ paraméterű karakterisztikus eloszlásból származó minta generálása révén adódott.

5.3. Gyakorlatok

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

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

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

jelölésekkel $[\tau_1, \tau_2]$ 0,95 biztonsági szintű konfidenciaintervallum a szórásra. A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 1,9081 illetve 2,8465. Ellenőrzésképpen közöljük, hogy a valódi szórás 2,4.

5.2. gyakorlat. Oldjuk meg az előző feladatot annak ismeretében, hogy a várható érték 14. Melyik módszer ad jobb becslést?

Útmutatás. A normális eloszlás szórásának becslésére vonatkozó 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.

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

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

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

jelölésekkel $[\tau_1, \tau_2]$ 0,94 biztonsági szintű konfidenciaintervallum a várható értékre. A számoláshoz tudnunk kell, hogy $F \sim \mathbf{t}(s)$ esetén

$$
\begin{aligned}
F^{-1}(x) &= \boxed{\texttt{-INVERZ.T(}2x\texttt{;}s\texttt{)}} \text{ ha } 0 < x < \tfrac{1}{2}, \\
F^{-1}(x) &= \boxed{\texttt{INVERZ.T(}2-2x\texttt{;}s\texttt{)}} \text{ ha } \tfrac{1}{2} \leqslant x < 1.
\end{aligned}
$$

Most $x = 1 - \frac{\alpha}{2} \geqslant \frac{1}{2}$ és $2 - 2x = \alpha$, ezért $F^{-1}\left(1 - \frac{\alpha}{2}\right) = \boxed{\texttt{INVERZ.T(}\alpha\texttt{;}n-1\texttt{)}}$. A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 4,2918 illetve 4,8270. Ellenőrzésképpen közöljük, hogy a valódi várható érték 4,6.

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

Útmutatás. A normális eloszlás várható értékének becslésére vonatkozó 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.

5.5. 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 $1 - \alpha = 0{,}9$

$$F \sim \mathbf{Gamma}(n; 1)$$
$$\tau_1 := \frac{1}{n\overline{\xi}} F^{-1}\left(\frac{\alpha}{2}\right)$$
$$\tau_2 := \frac{1}{n\overline{\xi}} F^{-1}\left(1 - \frac{\alpha}{2}\right)$$

jelölésekkel $[\tau_1, \tau_2]$ 0,9 biztonsági szintű konfidenciaintervallum λ-ra. A számoláshoz tudnunk kell, hogy $F \sim \mathbf{Gamma}(r; \lambda')$ esetén

$$F^{-1}(x) = \boxed{\texttt{INVERZ.GAMMA(}x\texttt{;}r\texttt{;1/}\lambda'\texttt{)}} \quad (0 < x < 1).$$

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

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

Útmutatás. Először oldjuk meg a gyakorlatot úgy, ahogy azt egy korábbi hasonló példában tettük. A kapott intervallum alsó illetve felső végpontja négy tizedesjegyre kerekítve 0,2449 illetve 0,2675. Ellenőrzésképpen közöljük, hogy a feladatban megadott gyakoriság, egy $p = 0{,}26$ paraméterű karakterisztikus eloszlásból származó minta generálása révén adódott.

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 $1 - \alpha = 0{,}99$

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

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

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

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

5.7. 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 $a = 7$, $1 - \alpha = 0{,}95$

$$\begin{aligned}
&F \sim \mathbf{Gamma}(n; 1) \\
&c_1 := F^{-1}\left(\frac{\alpha}{2}\right) \\
&c_2 := F^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&\tau_1 := a + \left(e^{c_1}\prod_{i=1}^{n}(\xi_i - a)\right)^{\frac{1}{n}} \\
&\tau_2 := a + \left(e^{c_2}\prod_{i=1}^{n}(\xi_i - a)\right)^{\frac{1}{n}}
\end{aligned}$$

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. 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 eloszláscsaládból 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. Általában H_0 és H_1 közül az egyik mindig bekövetkezik, de ez nem mindig van így (lásd például az úgynevezett egyoldali ellenhipotéziseket). 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. (Ha nincs ilyen, akkor a sejtésünk legyen H_1 és ezután H_0-t úgy választjuk meg, hogy már legyen hozzá próbastatisztika.)

2. Rögzített α ismeretében megadunk egy $K \subset \mathbb{R}$ halmazt úgy, hogy H_0 teljesülése esetén a $\tau \in K$ esemény valószínűsége maximum (vagy ha lehet, pontosan) α legyen. A $\tau \in K$ eseményt *kritikus tartománynak*, míg az ellenkezőjét *elfogadási tartománynak* nevezzük.

3. Ha a mintarealizáció alapján teljesül $\tau \in K$, akkor H_0-t elutasítjuk, azaz H_1-gyet fogadjuk el, míg $\tau \notin K$ esetben H_0-t elfogadjuk a H_1 ellenhipotézissel szemben.

Ekkor α terjedelmű próbát kapunk. A K 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

$\xi \in \mathbf{Norm}(m; \sigma)$, m ismeretlen, σ ismert, $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta, $m_0 \in \mathbb{R}$ rögzített.

$$\boxed{H_0\colon m = m_0}$$

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

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

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 mintarealizáció átlaga három tizedesjegyre kerekítve 13,982. A kérdés az, hogy 13,8-től csak véletlenül nagyobb, vagy szignifikánsan, azaz van valami oka.

Egymintás u-próba alkalmazható, ahol $H_0\colon m = 13{,}8$ és $H_1\colon m > 13{,}8$. A szint 0,99, azaz $\alpha = 0{,}01$. Ezt kell összehasonlítani $1 - \Phi(u)$ értékével. Excelben

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

ahol a mintarealizáció az `A` oszlopban van. Tehát másoljuk a mintarealizációt az `A` oszlopba, majd egy cellába írjuk a következőt:

$$\boxed{\texttt{=Z.PRÓBA(A:A;13,8;2)}}.$$

Ennek az értéke hat tizedesjegyre kerekítve 0,004685, 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

$\xi \in \mathbf{Norm}(m_1;\sigma_1)$, $\eta \in \mathbf{Norm}(m_2;\sigma_2)$ függetlenek, m_1, m_2 ismeretlenek, σ_1, σ_2 ismertek, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta.

$$\boxed{H_0\colon m_1 = m_2}$$

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

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

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. 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 $2 - 2\Phi(\lvert u\rvert)$ értéke öt tizedesjegy pontossággal 0,00001. Mivel $\alpha = 0{,}02$, ezért az ellenhipotézist fogadjuk el, azaz a két várható érték nem egyezik meg.

6.3. Egymintás t-próba

$\xi \in \mathbf{Norm}(m;\sigma)$, ahol m,σ ismeretlenek, $\xi_1,\dots,\xi_n$ a ξ-re vonatkozó minta, $n \geqslant 2$, $m_0 \in \mathbb{R}$ rögzített.

$$\boxed{H_0\colon m = m_0}$$

$$t = \frac{\overline{\xi} - m_0}{S_n^*}\sqrt{n}$$

$$F \sim \mathbf{t}(n-1)$$

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

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, ha a szórást nem ismerjük.

Megoldás. A mintarealizáció átlaga eltér 14-től. Kérdés, hogy ez véletlen vagy szignifikáns eltérés.

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

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

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 klikkeljünk 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 most négy tizedesjegyre kerekítve 0,7998 lesz. Ettől kisebb α, így a nullhipotézist fogadjuk el. Tehát a várható érték 14.

6.4. F-próba

$\xi \in \mathbf{Norm}(m_1;\sigma_1)$, $\eta \in \mathbf{Norm}(m_2;\sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\xi_1,\dots,\xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1,\dots,\eta_{n_2}$ az η-ra vonatkozó minta, $n_1 \geqslant 2$, $n_2 \geqslant$ $\geqslant 2$.

$$\boxed{H_0\colon \sigma_1 = \sigma_2}$$

$$\mathsf{F} = \frac{{S^*}^2_{\xi,n_1}}{{S^*}^2_{\eta,n_2}}$$

$$F \sim \mathbf{F}(n_1 - 1; n_2 - 1)$$

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

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. 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

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

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

6.5. Kétmintás t-próba

$\xi \in \mathbf{Norm}(m_1; \sigma_1)$, $\eta \in \mathbf{Norm}(m_2; \sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\sigma_1 = \sigma_2$, $\xi_1, \dots, \xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1, \dots, \eta_{n_2}$ az η-ra vonatkozó minta, $n_1 \geqslant 2$, $n_2 \geqslant 2$.

$$\boxed{H_0\colon m_1 = m_2}$$

$$t = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{n_1 S^2_{\xi,n_1} + n_2 S^2_{\eta,n_2}}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

$$F \sim \mathbf{t}(n_1 + n_2 - 2)$$

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

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

Megoldás. 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

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

amely most gyakorlatilag 0. Így $\alpha = 0{,}01$ ettől nagyobb, azaz az ellenhipotézist fogadjuk el. Tehát a várható értékek különböznek, vagyis nem azonos normális eloszlásból származik a két minta.

6.6. Scheffé-módszer

$\xi \in \mathbf{Norm}(m_1;\sigma_1)$, $\eta \in \mathbf{Norm}(m_2;\sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta, $2 \leqslant n_1 \leqslant n_2$.

$$\boxed{H_0\colon m_1 = m_2}$$

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

Speciálisan $n_1 = n_2$ esetén $\zeta_i = \xi_i - \eta_i$ teljesül.

$$t = \frac{\overline{\zeta}}{S^*_{\zeta,n_1}} \sqrt{n_1}$$

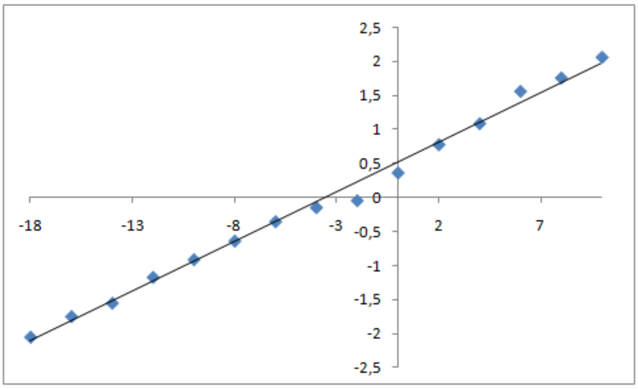
$$F \sim \mathbf{t}(n_1 - 1)$$

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

$n_1 = n_2$ esetén 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 gyerekkel 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. 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 klikkeljen kétszer. Így a `C` oszlopban megjelenik a különbség minta. Erre grafikus normalitásvizsgálatot végzünk a korábban ismertetett módon.



Ebből kapjuk, hogy a különbség minta normális eloszlásúnak tekinthető. Így alkalmazhatjuk a várható értékek összehasonlítására a Scheffé-módszert. Az ellenhipotézis legyen az, hogy a módszer sikeres, azaz a második minta várható értéke nagyobb mint az elsőé. Így az $F(t)$ értékét kell kiszámolni, melyet `=T.PRÓBA(A:A;B:B;1;1)` módon lehet megtenni, ha az első minta átlaga kisebb. Ez most teljesül, így kapjuk, hogy $F(t) = 0{,}000161$ hat tizedesjegyre kerekítve. Ettől α nagyobb, tehát az ellenhipotézist fogadjuk el, azaz a módszer sikeresnek tekinthető.

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 F-próbát csináljon. Ennek az lesz az eredménye, hogy a szórások nem egyeznek meg. Ezért a kétmintás t-próba nem alkalmazható, így marad a Scheffé-módszer. A $\zeta_1, \dots, \zeta_{n_1}$ 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. Külön számolja ki az $\frac{1}{\sqrt{n_1 n_2}} \sum_{k=1}^{n_1} \eta_k - \overline{\eta}$ értékét

`=SZUM(A1:A50)/GYÖK(50*456)-ÁTLAG(A:A)`

módon. A cella neve legyen például $\boxed{\texttt{konst}}$. Ezután a ζ_1 értékét számolja ki a `C1` cellába:

$$\boxed{\texttt{=B1-GYÖK(50/456)*A1+konst}}.$$

Klikkeljen kétszer a `C1` cella kitöltőjelére. Ezzel a `C` oszlopban elkészült a $\zeta_1, \dots, \zeta_{n_1}$ minta. A Scheffé-módszer szerint erre kell alkalmazni az egymintás t-próbát $m_0 = 0$ választással. Tehát a `D1` cellába írja, hogy 0, majd klikkeljen kétszer a `D1` cella kitöltőjelére. Ekkor

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

Ennek értéke most gyakorlatilag 0, tehát α nagyobb. Így a várható értékek nem egyeznek meg. Kiszámolva az átlagokat, az elsőé nagyobb, tehát az egyoldali ellenhipotézisek esetén azt a döntést hoznánk, hogy az első minta várható értéke nagyobb.

6.7. Khi-négyzet próba

$\xi \in \mathbf{Norm}(m; \sigma)$, ahol m, σ ismeretlenek, $\xi_1, \dots, \xi_n$ a ξ-re vonatkozó minta, $n \geqslant 2$.

$$\boxed{H_0\colon \sigma = \sigma_0}$$

$$\chi^2 = \frac{S_n^2}{\sigma_0^2} n$$

$$F \sim \mathbf{Khi}(n-1)$$

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

6.8. 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. Legyen az ellenhipotézis az, hogy nem felel meg a feltételnek, azaz $H_1\colon \sigma > 3$. Az $1 - F(\chi^2)$ értékét kell kiszámolni. A mintát másolja az `A` oszlopba. Ekkor $\chi^2 = \boxed{\texttt{VARP(A:A)*DARAB(A:A)/9}}$. Ezt a cellát nevezze el $\boxed{\texttt{khi}}$-nek. Ezután

$$1 - F(\chi^2) = \boxed{\texttt{KHI.ELOSZLÁS(khi;DARAB(A:A)-1)}}.$$

Ennek értéke 0,0003 négy tizedesjegyre kerekítve. Ettől nagyobb az α, tehát az ellenhipotézist fogadjuk el. Így a döntésünk az, hogy a gyártmány nem felel meg a szórásra vonatkozó feltételnek.

6.8. Statisztikai próba az exponenciális eloszlás paraméterére

$\xi \in \mathbf{Exp}(\lambda)$, ahol λ ismeretlen, $\xi_1, \dots, \xi_n$ a ξ-re vonatkozó minta, $\lambda_0 \in \mathbb{R}_+$ rögzített.

$$\boxed{H_0 \colon \lambda = \lambda_0}$$

$$\gamma = \lambda_0 n \overline{\xi}$$

$$F \sim \mathbf{Gamma}(n; 1)$$

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

6.9. Példa. Az egészségügyi miniszter hoz egy rendeletet, miszerint 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. A mintaátlag 7,07 perc, tehát úgy tűnik, hogy az orvos betartja a rendeletet. Kérdés, hogy ez csak véletlen, vagy szignifikánsan nagyobb a konzultáció idő 5 percnél. Legyen az az ellenhipotézis, hogy az orvos betartja a rendeletet, azaz $\mathrm{E}\,\xi = \frac{1}{\lambda} > 5$. Így tehát $H_1 \colon \lambda < 0{,}2$. Az $1 - F(\gamma)$ értékét kell összehasonlítani α-val. Másolja a mintát az `A` oszlopba. Ekkor

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

amely most 0,001 három tizedesjegyre kerekítve. Ettől α nagyobb, ezért az ellenhipotézist fogadjuk el, miszerint az orvos betartja a rendeletet.

6.9. Statisztikai próba valószínűségre

$\xi \in \mathbf{Bin}(1;p)$, p ismeretlen, $0 < p_0 < 1$ rögzített és $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta.

$$\boxed{H_0 \colon p = p_0}$$

$$F^{-1}(x) = \min\Big\{ z \in \mathbb{N} : \sum_{i=0}^{z} \tbinom{n}{i} p_0^i (1-p_0)^{n-i} \geqslant x \Big\}.$$

Ha $\min\{ np_0, n(1-p_0) \} \geqslant 10$, akkor

$$F^{-1}(x) \simeq np_0 - \frac{1}{2} + \sqrt{np_0(1-p_0)}\,\Phi^{-1}(x).$$

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

Az előbbi kritikus tartományok végpontjaira (*kritikus értékek*) további feltételek kellenek:

H_1	feltétel
$p \neq p_0$	$1 \leqslant F^{-1}\left(\frac{\alpha}{2}\right) < np_0 < F^{-1}\left(1-\frac{\alpha}{2}\right) \leqslant n-1$
$p < p_0$	$1 \leqslant F^{-1}(\alpha) < np_0$
$p > p_0$	$np_0 < F^{-1}(1-\alpha) \leqslant n-1$

Ezen feltételek mindig teljesíthetők α és n alkalmas megválasztásával.

6.10. 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. Jelentse ξ az eljárás révén üszőborjú születésének indikátorváltozóját és p annak valószínűségét, hogy egy ilyen módszer alkalmazásával üszőborjú születik. Ekkor tehát $n = 100$ 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

$$H_0 \colon p = \frac{1}{2}$$

$$H_1\colon p > \frac{1}{2}$$

hipotézisekről döntünk. A szint 0,99, így a kritikus érték

$$F^{-1}(1-\alpha) = F^{-1}(0{,}99) = \min\Big\{\, z \in \mathbb{N} : \sum_{i=0}^{z} \tbinom{100}{i}\big(\tfrac{1}{2}\big)^{100} \geqslant 0{,}99 \,\Big\}.$$

A feltétel az, hogy ez az érték 51 és 99 közé essen. Excelben

$$\min\Big\{\, z \in \mathbb{N} : \sum_{i=0}^{z} \tbinom{n}{i} p^i (1-p)^{n-i} \geqslant x \,\Big\} = \boxed{\texttt{KRITBINOM}(n\texttt{;}p\texttt{;}x\texttt{)}},$$

azaz $F^{-1}(0{,}99)$ $\boxed{\texttt{=KRITBINOM(100;1/2;0,99)}}$ módon számolható ki. Így kapjuk, hogy $F^{-1}(0{,}99) = 62$. Erre teljesülnek a feltételek, és $n\bar{\xi} = 61 < 62$ 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 $n\bar{\xi} = 61 > 58$. Így a válaszunk a szint függvénye. Ezért tanácsos további kísérleteket végezni.

6.11. 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.

Megjegyezzük, hogy a számoláshoz $\min\{\, np_0, n(1-p_0) \,\} \geqslant 10$ miatt a közelítő formulát is használhatjuk:

$$F^{-1}(0{,}99) \simeq 200 \cdot \tfrac{1}{2} - \tfrac{1}{2} + \sqrt{200 \cdot \tfrac{1}{2} \cdot (1-\tfrac{1}{2})}\,\Phi^{-1}(0{,}99) \simeq 115{,}9$$

egy tizedesjegyre kerekítve. Mivel ez az elfogadási tartomány felső határa, ezért célszerű fölfelé kerekíteni, azaz 116-ot kapunk, mint az előbb.

6.10. 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

$$2 - 2\Phi(|u|) = 2\min\{\, 1 - \Phi(u), \Phi(u)\,\},$$

így

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

módon számolható.

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. A $\Phi(u)$ értéke négy tizedesjegyre kerekítve 0,1442, melytől kisebb α, így elfogadjuk a nullhipotézist. Tehát a szállítmányt átveheti a kereskedő.

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

Útmutatás. Használjon egymintás t-próbát az előző hipotézisekkel. Excelben, ha $\overline{\xi} \leqslant$ $\leqslant m_0$, akkor

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

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.

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 megvegye-e a szert a kórház.

Útmutatás. Használjon egymintás t-próbát $H_0\colon m = 12$ és $H_1\colon m > 12$ hipotézisekkel. Excelben, ha $\overline{\xi} \geqslant m_0$, akkor

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

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.

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.

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. A `minta-06` korrigált tapasztalati szórása kisebb a `minta-08` korrigált tapasztalati szórásánál, ezért a „második gép pontosabb az elsőnél” hipotézist biztosan elutasítjuk. Most vizsgáljuk az „első gép pontosabb a másodiknál” hipotézist, mint ellenhipotézist. Ekkor az $F(\mathsf{F})$ értékét kell összehasonlítani α-val. A korrigált tapasztalati szórások előbbi relációja mellett – ha az első gép adatai az `A` oszlopban, míg a második gép adatai a `B` oszlopban vannak –, Excelben $F(\mathsf{F}) = \boxed{\texttt{F.PRÓBA(A:A;B:B)/2}}$. Ennek értéke gyakorlatilag 0, így tehát azt állíthatjuk, hogy az első gép pontosabb a másodiknál.

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

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

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. Ekkor az $1 - F(t)$ értékét kell $\alpha = 0{,}01$-dal összehasonlítani. Mivel az első minta átlaga nagyobb, ezért Excelben

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

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. Nemparaméteres hipotézisvizsgálatok

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

7.1. Tiszta illeszkedésvizsgálat

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

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

ahol P a valódi valószínűség. $\varrho_i(\geqslant 10)$ az A_i gyakorisága n kísérlet után.

$$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - np_i)^2}{np_i}, \quad F \sim \mathbf{Khi}(r-1)$$

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

7.1. Példa. Egy dobókockával 400-szor dobtunk, és a következő gyakoriságok jöttek ki:

1	2	3	4	5	6
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 $np_i = \frac{400}{6}$ minden $i \in \{1,2,3,4,5,6\}$ esetén. 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, hogy `=400/6`. Ezután a `B1` cella kitöltőjelére klikkeljen kétszer. Ekkor

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

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

A tiszta illeszkedésvizsgálat alkalmazható valószínűségi változó eloszlásának tesztelésére is.

ξ-re vonatkozó minta $\xi_1, \dots, \xi_n$.

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

$$a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < a_r = \infty$$

$$\varrho_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \geqslant 10$$

$$p_i = \mathrm{P}(a_{i-1} \leqslant \xi < a_i) = F_0(a_i) - F_0(a_{i-1}) \text{ (azaz } \mathrm{P} \in \mathcal{P}_{H_0})$$

$$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - np_i)^2}{np_i}, \quad F \sim \mathbf{Khi}(r-1)$$

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` cellákba írja be rendre a 0, 1, 2 számokat, majd a `C1` cellába, hogy

`=POISSON(B1;1;HAMIS)`.

A `C1` kitöltőjelére klikkeljen kétszer. Végül a `C4` cellába írja be, hogy

`=1-SZUM(C1:C3)`.

Ezzel a `C` oszlopban megjelentek a

$$p_i = \mathrm{P}(a_{i-1} \leqslant \xi < a_i) = F_0(a_i) - F_0(a_{i-1})$$

értékek, ahol $a_0 = -\infty,\ a_1 = 1,\ a_2 = 2,\ a_3 = 3,\ a_4 = \infty$ 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 kitöltőjelre klikkeljen kétszer. A D4 cellát javítsa ki `=DARABTELI(A:A;">=3")` módon. A kapott gyakoriságok mindegyike nagyobb 10-nél, így alkalmazható a próba. Az `E` oszlopban számolja ki az np_i értékeket. Írja az `E1` cellába, hogy `=DARAB(A:A)*C1`, majd a kitöltőjelre klikkeljen kétszer. Ekkor

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

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

7.2. Becsléses illeszkedésvizsgálat

ξ-re vonatkozó minta $\xi_1, \ldots, \xi_n$ és minden $\vartheta = (\vartheta_1, \ldots, \vartheta_v) \in \Theta \subset \mathbb{R}^v$ esetén F_ϑ eloszlásfüggvény.

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

$$a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < a_r = \infty$$

$$\varrho_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \geqslant 10$$

$\widehat{\vartheta}_i$ a ϑ_i maximum likelihood becslése H_0 feltételezésével.

$$\widehat{p}_i = \mathrm{P}_{(\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_v)}(a_{i-1} \leqslant \xi < a_i) = F_{(\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_v)}(a_i) - F_{(\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_v)}(a_{i-1})$$

$$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - n\widehat{p}_i)^2}{n\widehat{p}_i}, \qquad F \sim \mathbf{Khi}(r-1-v)$$

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

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

Megoldás. A nullhipotézis legyen az, hogy a vizsgált valószínűségi változó normális eloszlású. Tudjuk, hogy ennek teljesülése esetén a várható érték illetve a szórás maximum likelihood becslése a mintaátlag illetve a tapasztalati szórás. A következő táblázatot fogjuk elkészíteni:

	A	B	C	D	E	F	G
1	10,83	$\widehat{m} =$	10,015025	a_i	ϱ_i	$n\widehat{p}_i$	$\frac{(\varrho_i - n\widehat{p}_i)^2}{n\widehat{p}_i}$
2	7,75	$\widehat{\sigma} =$	1,99908529	-1000	17	17,83900	0,039459833
3	6,52	$n =$	800	6	36	34,76210	0,04408214
4	10,38	$1 - F(\chi^2) =$	0,6064	7	67	72,78577	0,459912899
5	9,44			8	135	119,26608	2,075663545
6	6,77			9	145	152,94833	0,413054141
7	11,25			10	147	153,51252	0,276282891
8	6,75			11	113	120,59084	0,477821047
9	8,03			12	85	74,13833	1,591295014
10	8,03			13	37	35,66993	0,049595752
11	11,48			14	18	18,48710	0,012834212
12	8,73			1000			

A mintarealizációt másolja az `A` oszlopba. A mintaátlagot, tapasztalati szórást és a mintarealizáció elemeinek a számát számolja ki a `C1`, `C2` illetve `C3` cellákba az

`=ÁTLAG(A:A)`, `=SZÓRÁSP(A:A)` illetve `=DARAB(A:A)` függvényekkel. Rendre nevezze el a cellákat `m`, `szigma` illetve `n` módon.

A `D` oszlopba kerülnek az a_i osztópontok. Figyelembe véve $\widehat{m}$ és $\widehat{\sigma}$ értékét, gyakorlati szempontból $-\infty$ helyett -1000 illetve ∞ helyett 1000 írható. Az $a_0, \ldots, a_{10}$ helyére írja a $-1000, 6, 7, \ldots$,14, 1000 számokat a `D2:D12` cellatartományba. Az `E2` cellába számolja ki a ϱ_1 gyakoriságot:

`=DARABHATÖBB(A:A;">="&D2;A:A;"<"&D3)`.

Az `F` oszlopba kerülnek az $n\widehat{p}_i$ értékei. Ehhez vegyük figyelembe, hogy most

$$\widehat{p}_i = F_{(\widehat{m},\widehat{\sigma})}(a_i) - F_{(\widehat{m},\widehat{\sigma})}(a_{i-1}),$$

ahol $F_{(\widehat{m},\widehat{\sigma})} \sim \mathbf{Norm}(\widehat{m}; \widehat{\sigma})$. Így tehát az `F2` cellába írjuk be, hogy

`=n*(NORM.ELOSZL(D3;m;szigma;IGAZ)-NORM.ELOSZL(D2;m;szigma;IGAZ))`.

Ha ezekből kitöltenénk az `E` és `F` oszlopokat, majd erre alkalmaznánk a `KHI.PRÓBA` függvényt úgy, mint az előző példában, akkor a kapott érték $r - 1 = 9$ szabadsági fokkal lenne kiszámolva. De ez most nem jó, mert a két becsléssel ($v = 2$) kettővel csökkent a szabadsági fok. Így ki kell számolni a χ^2 statisztikát, majd az $1 - F(\chi^2)$ értékét $r - 1 - v = 7$ szabadsági fokkal. Ehhez a G oszlopban számolja ki a $\frac{(\varrho_i - n\widehat{p}_i)^2}{n\widehat{p}_i}$ értékeket. A `G2` cellába írja be, hogy

`=(E2-F2)^2/F2`.

A táblázat első sora alapján töltse ki a hiányzó cellákat. Jelölje ki az `E2:G2` cellatartományt, majd a kitöltőjelet húzza le a 11. sorig. Látható, hogy a ϱ_i gyakoriságok mindegyike nagyobb 10-nél, így alkalmazható a próba. Számolja ki az $1 - F(\chi^2)$ értékét a `C4` cellába:

`=KHI.ELOSZLÁS(SZUM(G:G);7)`.

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

7.3. Függetlenségvizsgálat

$A_1, \ldots, A_r$ és $B_1, \ldots, B_s$ két teljes eseményrendszer.

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

ahol P a valódi valószínűség. A kontingencia táblázat

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_s & \\
\hline
A_1 & \varrho_{11} & \varrho_{12} & \dots & \varrho_{1s} & k_1 \\
A_2 & \varrho_{21} & \varrho_{22} & \dots & \varrho_{2s} & k_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_r & \varrho_{r1} & \varrho_{r2} & \dots & \varrho_{rs} & k_r \\
\hline
 & l_1 & l_2 & \dots & l_s & n
\end{array}
$$

melyben $\varrho_{ij} \geqslant 10$ minden i, j esetén.

$$
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \frac{1}{n} k_i l_j)^2}{\frac{1}{n} k_i l_j}, \qquad F \sim \mathbf{Khi}((r-1)(s-1))
$$

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

7.4. 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 nullhipotézisünk az lesz, hogy független a szem- és hajszín. Mivel a ϱ_{ij} gyakoriságok nagyobbak 10-nél, ezért alkalmazható a próba. Először készítse el a kontingencia táblázatot. Az előző táblázat értékeit gépelje be az `A1:C2` cellatartományba, majd számolja ki a perem-gyakoriságokat. A `D1` cellába írja be, hogy `=SZUM(A1:C1)`, majd a kitöltőjelet húzza le a 2. sorig. Ezután az `A3` cellába írja be, hogy `=SZUM(A1:A2)`, majd a kitöltőjelet húzza jobbra a `D` oszlopig.

Ennek alapján elkészítjük az $\frac{1}{n} k_i l_j$ táblázatát. Ehhez szükség lesz az Excel *relatív és abszolút hivatkozásának* a fogalmára. Amikor egy cellába például azt írja, hogy `=A1`, majd a kitöltőjelet lehúzza a következő celláig, akkor abban `=A2` jelenik meg. Ez az úgynevezett relatív hivatkozás. Ha ezt a hatást nem akarja, akkor a sor számát abszolúttá kell tenni úgy, hogy a sorszám elé \$ jelet kell gépelni: `=A$1`. Ekkor már hiába húzza a kitöltőjelet le vagy fel, a sorszám nem változik. De ha jobbra vagy balra húzza, akkor az oszlop azonosítója változni fog, mert az még relatív. Értelemszerűen ezt is abszolúttá lehet tenni egy elé írt \$ jellel: `=$A$1`. Ha mindkét azonosítót egyszerre akarja abszolútra változtatni, akkor a kurzorral lépjen az `A1` szövegre és nyomja meg az *F4* funkcióbillentyűt. Ekkor automatikusan megjelennek a \$ jelek mindkét azonosító előtt. A cellára való hivatkozást abszolúttá lehet tenni úgy is, hogy nevet adunk a cellának és ezzel hivatkozunk rá. Ezt már eddig is csináltuk. Ez a megoldás azért is szerencsés, mert ezzel átláthatóbbak lesznek a képletek.

Mindezek alapján 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. Ezután

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

	A	B	C	D	E	F	G	H
1	42	28	3	73		21,535	42,705	8,76
2	17	89	21	127		37,465	74,295	15,24
3	59	117	24	200				
4								
5								
6			$1 - F(\chi^2) =$	2,102E-10				
7								

Most $1 - F(\chi^2) \simeq 2{,}102 \cdot 10^{-10} < \alpha = 0{,}01$, melyből következően elutasítjuk a nullhipotézist, azaz a szem- és hajszín között van genetikai kapcsolat.

Függetlenségvizsgálat alkalmazható valószínűségi változók függetlenségének tesztelésére is.

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

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

$$a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < a_r = \infty$$

$$b_0 = -\infty < b_1 < b_2 < \cdots < b_{s-1} < a_s = \infty$$

$$k_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \qquad l_j = \sum_{z=1}^{n} \mathrm{I}_{\eta_z \in [b_{j-1}, b_j)}$$

$$\varrho_{ij} = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \, \mathrm{I}_{\eta_z \in [b_{j-1}, b_j)} \geqslant 10$$

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \frac{1}{n} k_i l_j)^2}{\frac{1}{n} k_i l_j}, \qquad F \sim \mathbf{Khi}((r-1)(s-1))$$

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

7.5. 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.

Ctrl+A-val jelölje ki a teljes mintát, majd *Ctrl+C*-vel tegye a vágólapra. Nyisson egy Excel munkalapot és álljon az `A1` cellára. Ezután *Ctrl+V*-vel az `A` oszlopba kerül a ξ-re vonatkozó mintarealizáció, míg a `B` oszlopba kerül az η-ra vonatkozó mintarealizáció. Lépjen az `A` oszlop azonosítójára, majd nevezze el `xi`-nek. Ezután lépjen a `B` oszlop azonosítójára, majd nevezze el `eta`-nak.

Mindkét mintában tekinthető $-\infty$ 0-nak és ∞ 1000-nek. Legyen $a_1 = 0{,}13$, $a_2 = 0{,}3$, $a_3 = 0{,}5$ és $b_1 = 0{,}1$, $b_2 = 0{,}2$, $b_3 = 0{,}3$, $b_4 = 0{,}5$, $b_5 = 1$. Az a_i osztópontokat írja a `D2-D6` cellákba, míg a b_i osztópontokat az `E1-K1` cellákba.

	A	B	C	D	E	F	G	H	I	J	K
1	1,426	0,325			**0**	**0,1**	**0,2**	**0,3**	**0,5**	**1**	**1000**
2	1,042	1,468		**0**	18	22	10	20	22	12	104
3	0,326	0,705		**0,13**	17	12	16	21	32	14	112
4	1,190	0,078		**0,3**	23	14	14	18	20	18	107
5	0,331	0,598		**0,5**	40	25	13	41	38	20	177
6	0,112	0,696		**1000**	98	73	53	100	112	64	500
7	2,176	0,206									
8	1,141	0,006			20,384	15,184	11,024	20,8	23,296	13,312	
9	0,884	0,519			21,952	16,352	11,872	22,4	25,088	14,336	
10	0,459	0,442			20,972	15,622	11,342	21,4	23,968	13,696	
11	1,134	0,064			34,692	25,842	18,762	35,4	39,648	22,656	
12	0,577	0,552									
13	0,314	0,088			$1-F(\chi^2) =$ 0,3317						

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

`=DARABHATÖBB(xi;">="&$D2;xi;"<"&$D3;eta;">="&E$1;eta;"<"&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.

Határozza meg a perem-gyakoriságokat. A `K2`-be írja be, hogy `=SZUM(E2:J2)`, majd a kitöltőjelet húzza le az 5. sorig. Ezután az `E6` cellába gépelje be, hogy `=SZUM(E2:E5)`, majd a kitöltőjelet húzza jobbra a `K` oszlopig.

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

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

a kitöltőjelet húzza jobbra a `J` oszlopig, majd lefelé az 11. sorig. Ezután

$1 - F(\chi^2) =$ `KHI.PRÓBA(E2:J5;E8:J11)`.

Most $1 - F(\chi^2) \simeq 0{,}3317 > \alpha = 0{,}01$, melyből következően elfogadjuk a nullhipotézist, azaz ξ és η független.

7.4. Homogenitásvizsgálat

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

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

$$c_0 = -\infty < c_1 < c_2 < \dots < c_{r-1} < c_r = \infty$$

$$k_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [c_{i-1}, c_i)} \geqslant 10, \quad l_j = \sum_{z=1}^{n} \mathrm{I}_{\eta_z \in [c_{j-1}, c_j)} \geqslant 10$$

$$\chi^2 = n_1 n_2 \sum_{i=1}^{r} \frac{\left(\frac{k_i}{n_1} - \frac{l_i}{n_2}\right)^2}{k_i + l_i}, \qquad F \sim \mathbf{Khi}(r-1)$$

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

A feladatok megoldásánál érdemes felhasználni, hogy a homogenitásvizsgálat megegyezik az alábbi kontingencia táblázatra vonatkozó függetlenségvizsgálattal:

$$\begin{array}{|cc|c}
\hline
k_1 & l_1 & k_1 + l_1 \\
k_2 & l_2 & k_2 + l_2 \\
\vdots & \vdots & \vdots \\
k_r & l_r & k_r + l_r \\
\hline
n_1 & n_2 & n_1 + n_2
\end{array}$$

7.6. 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 mintában tekinthető $-\infty$ -2000-nek illetve a ∞ 2000-nek. Legyen $c_1 = -2$, $c_2 = -1$, $c_3 = -0{,}5$, $c_4 = 0{,}5$, $c_5 = 1$, $c_6 = 2$. A c_i osztópontokat írja a `D2-D9` cellákba. Az `E2` cellába gépelje a következőt:

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

	A	B	C	D	E	F	G	H	I	J
1	0,0429	0,1162		c_i	k_i	l_i				
2	-0,6741	0,2242		-2000	85	18	103		42,848	60,152
3	0,3395	0,1977		-2	39	97	136		56,576	79,424
4	12,6127	-0,3287		-1	48	113	161		66,976	94,024
5	0,1109	0,9731		-0,5	154	267	421		175,136	245,864
6	2,8904	0,4905		0,5	60	111	171		71,136	99,864
7	1,0527	1,624		1	60	106	166		69,056	96,944
8	-0,4969	0,0148		2	74	18	92		38,272	53,728
9	3,828	0,063		2000	520	730	1250			
10	2,4266	1,6285								
11	0,5655	-0,8709			$1 - F(\chi^2) =$ 4E-31					

A kitöltőjelet húzza jobbra eggyel, majd le a 8. sorig. A kijött gyakoriságok mindegyike nagyobb 10-nél, ezért a próba alkalmazható. Most következnek a perem-gyakoriságok. A `G2` cellába írja be, hogy `=E2+F2`, majd a kitöltőjelet húzza le a 8. sorig. Az `E9` cellába írja be, hogy `=SZUM(E2:E8)`, majd a kitöltőjelet húzza jobbra a `G` oszlopig.

A perem-gyakoriságokból pontosan úgy készítjük el a nullhipotézis teljesülése esetén várható gyakoriságokat, mint a függetlenségvizsgálatban. Ez most megfelel a $\frac{(k_i+l_i)n_j}{n_1+n_2}$ táblázatnak. Az `I2` cellába írja be, hogy

`=$G2*E$9/$G$9`,

a kitöltőjelet húzza jobbra eggyel, majd le a 8. sorig. Ezután

$$1 - F(\chi^2) = \boxed{\texttt{KHI.PRÓBA(E2:F8;I2:J8)}}.$$

Most $1 - F(\chi^2) \simeq 4 \cdot 10^{-31} < \alpha = 0{,}01$, melyből következően elutasítjuk a nullhipotézist, azaz a két mintarealizáció különböző eloszlásból származik.

7.5. Kétmintás előjelpróba

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

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

$$B = \sum_{i=1}^{n} \mathrm{I}_{\xi_i > \eta_i}, \quad F^{-1}(x) = \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \left(\frac{1}{2}\right)^n \geqslant x \right\}$$

$n \geqslant 20$ esetén $F^{-1}(x) \simeq \frac{1}{2}\left(n - 1 + \sqrt{n}\Phi^{-1}(x)\right)$.

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

7.7. 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 $\boxed{\texttt{KRITBINOM}}$ függvénnyel

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

melynek most 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.

7.6. 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^*.

$$D = \sqrt{\tfrac{n}{2}} \sup_{x\in\mathbb{R}} |F_n^*(x) - G_n^*(x)|$$

$$K(z) = 1 + 2\sum_{i=1}^{\infty} (-1)^i e^{-2i^2z^2}$$

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

A gyakorlatban D kiszámolásához elég csak a két tapasztalati eloszlásfüggvény összes szakadási pontjában megvizsgálni a differenciákat. Tehát, ha

$$x_1 := \xi_1, \dots, x_n := \xi_n, x_{n+1} := \eta_1, \dots, x_{2n} := \eta_n,$$

akkor

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - G_n^*(x)| = \max_{i=1,\dots,2n} |F_n^*(x_i) - G_n^*(x_i)|.$$

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

7.8. Példa. Excelben számolja ki adott $z \geqslant 1$ esetén $K(z)$ értékét.

Megoldás. A `B1` cellába írja be, hogy $\boxed{\texttt{z =}}$, majd a `C1` cellába egy konkrét z értéket, például most legyen 1. A `C1` cellát nevezze el $\boxed{\texttt{z}}$-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. Amint látni fogja, 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 $i \geqslant 19$ 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 $z \geqslant 1$ esetben sem kell újabb szumma tagokat számolni.

	A	B	C
1	-0,13534	z =	1
2	0,000335	K(z) =	0,7300
3	-1,5E-08		
4	1,27E-14		
5	-1,9E-22		
6	5,38E-32		
7	-2,7E-43		
8	2,57E-56		
9	-4,4E-71		
10	1,38E-87		
11	-8E-106		
12	8,4E-126		
13	-2E-147		
14	5,7E-171		
15	-4E-196		
16	4,4E-223		
17	-1E-251		
18	3,8E-282		
19	0		

7.9. 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 $\boxed{\texttt{DARAB}}$ függvénnyel ellenőrizheti, hogy a közös mintaelemszám $n = 500 > 30$. Tehát alkalmazhatjuk a Kolmogorov – Szmirnov-féle kétmintás próbát.

Az `A` oszlopot nevezze el $\boxed{\texttt{xi}}$-nek, a `B` oszlopot pedig $\boxed{\texttt{eta}}$-nak. A `C1` cellába írja be, hogy

$$\boxed{\texttt{=ABS(DARABTELI(xi;"<"\&A1)/500-DARABTELI(eta;"<"\&A1)/500)}}.$$

Ez $|F_n^*(x_1) - G_n^*(x_1)|$, ahol $x_1 = \xi_1$. A kitöltőjelet húzza jobbra eggyel, majd a kitöltőjelre klikkeljen kétszer. Ezzel $|F_n^*(x_i) - G_n^*(x_i)|$ értékeit kapta meg a `C` oszlopban $i = 1, \ldots, n$ esetekre, míg a `D` oszlopban $i = n+1, \ldots, 2n$ esetekre, ahol $x_1 = \xi_1, \ldots, x_n = \xi_n, x_{n+1} = \eta_1, \ldots, x_{2n} = \eta_n$. Mivel

$$D = \sqrt{\tfrac{n}{2}} \max_{i=1,\ldots,2n} |F_n^*(x_i) - G_n^*(x_i)|,$$

ezért ez Excelben

$$D = \boxed{\texttt{GYÖK(500/2)*MAX(C:D)}}$$

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

7.7. Kolmogorov – Szmirnov-féle egymintás próba

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

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

F_n^* a tapasztalati eloszlásfüggvény.

$$D = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|$$

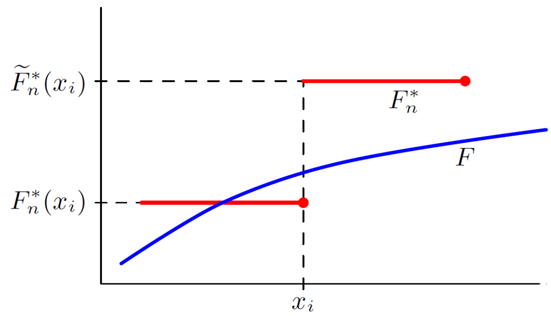
$$K(z) = 1 + 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2 z^2}$$

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

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, úgy mint a kétmintás esetben. Ezt meg kell tenni a bal végpontokban is. Máshol viszont nem kell, mert F monoton növekedő. Az összes lépcsőfok bal végpontját megkapjuk az

$$\widetilde{F}_n^*(x_i) := \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}_{\xi_k \leqslant x_i}$$

képlettel, ahol x_i befutja a mintarealizáció összes elemét.



Így kapjuk, hogy

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| = \max_{i=1,\dots,n} \max\{\, |F_n^*(x_i) - F(x_i)|, |\widetilde{F}_n^*(x_i) - F(x_i)|\,\}$$

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

Megoldás. Legyen az a nullhipotézis, hogy a mintarealizáció 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 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 klikkeljen kétszer a kitöltőjelre. Ezzel a `B` oszlopban megkapta az $|F_n^*(x_i)-F(x_i)|$, míg a `C` oszlopban a $|\widetilde{F}_n^*(x_i)-F(x_i)|$ értékeit. Így kapjuk, hogy

$D =$ `GYÖK(730)*MAX(B:C)`,

amelynek most 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, azaz a mintarealizáció nem Cauchy-eloszlásból származik.

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

Útmutatás. A feladatot tiszta illeszkedésvizsgálattal oldja meg. Legyen az a nullhipotézis, hogy standard normális eloszlásból származik a mintarealizáció. Az osztópontok legyenek $a_0 = -\infty,\ a_1 = -2,\ a_2 = -1,\ a_3 = 0,\ a_4 = 1,\ a_5 = 2,\ a_6 = \infty$. A számolásnál $-\infty$ helyére írható pl. -1000, illetve ∞ helyére 1000. Ekkor $1-F(\chi^2) \simeq$ $\simeq 0{,}8337 > 0{,}01 = \alpha$, azaz a nullhipotézist elfogadjuk.

7.3. 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 $a_0 = -\infty,\ a_1 = 2,\ a_2 = 4,\ a_3 = 6,\ a_4 =$ $= 10,\ a_5 = \infty$. A számolásnál $-\infty$ helyére írható pl. 0, illetve ∞ helyére 1000. 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.

7.4. 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.

7.5. 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. Mindkét mintában tekinthető $-\infty$ -1000-nek és ∞ 1000-nek. Legyen például $a_1 = -1,\ a_2 = 0,\ a_3 = 2$ és $b_1 = -1,\ b_2 = 1,\ b_3 = 2$. Ekkor $1 - F(\chi^2) \simeq 0 < 0{,}01 = \alpha$, melyből következően ξ és η nem függetlenek.

7.6. 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 $c_0 = 1,\ c_1 = 2,\ c_2 = 3,\ c_3 = 4,\ c_4 = 5,\ c_5 = 6,\ c_6 = 7$. 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”.

7.7. 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. A számolásnál alkalmazzon folytonossági korrekciót.

Útmutatás. Végezzen kétmintás előjelpróbát. Azt kapjuk, hogy

$$F^{-1}(0{,}99) \simeq \frac{1}{2}\left(21 + \sqrt{22}\Phi^{-1}(0{,}99)\right) \simeq 15{,}9558 > 15 = B,$$

tehát az ellenhipotézist nem fogadjuk el, azaz $\frac{1}{2}$-nél nem nagyobb valószínűséggel jobb az új gerely a réginél.

7.8. 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.

7.9. 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. Regressziószámítás

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

$$\eta \simeq g(\xi_1, \ldots, \xi_k)$$

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 lesz.

8.1. Lineáris regresszió

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

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

alakú – azaz lineáris – függvények között keressük. Ezt a típusú regressziószámítást *lineáris regressziónak* nevezzük. A feladat megoldásában szereplő $a_0, \ldots, a_k$ konstansokat a *lineáris regresszió együtthatóinak* nevezzük. $k = 1$ illetve $k = 2$ esetén a lineáris regresszióval kapott g függvényt *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:

$$a := (a_0, \ldots, a_k)^\top$$
$$Y := (\eta_1, \ldots, \eta_n)^\top$$

$$X := \begin{pmatrix} 1 & \xi_{11} & \dots & \xi_{1k} \\ 1 & \xi_{21} & \dots & \xi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_{n1} & \dots & \xi_{nk} \end{pmatrix}.$$

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

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

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

$$\widehat{a} = (\widehat{a}_0, \dots, \widehat{a}_k)^\top = (X^\top X)^{-1} X^\top Y.$$

Ezután az $\eta \simeq \widehat{a}_0 + \widehat{a}_1\xi_1 + \dots + \widehat{a}_k\xi_k$ közelítést fogjuk használni. Speciálisan $k = 1$ esetén a lineáris regresszió együtthatóinak becslése

$$\begin{aligned} \widehat{a}_0 &= \overline{\eta} - \frac{\mathrm{Cov}_n(\eta, \xi_1)}{S^2_{\xi_1,n}}\overline{\xi_1}, \\ \widehat{a}_1 &= \frac{\mathrm{Cov}_n(\eta, \xi_1)}{S^2_{\xi_1,n}}, \end{aligned}$$

ahol $\mathrm{Cov}_n(\eta, \xi_1)$ a tapasztalati kovarianciája az (η, ξ_1)-re vonatkozó mintának. Ennek alapján a továbbiakban az $\eta \simeq \widehat{a}_0 + \widehat{a}_1\xi_1$ közelítést fogjuk használni. A grafikus illeszkedésnél pontosan ezt a közelítést alkalmaztuk.

8.1. Példa. Jelentse η a talajvízszintet mm-ben és ξ_1 az őszi csapadék mennyiségét cm-ben. Az (η, ξ_1)-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. Becsülje meg a talajvízszintet, ha az őszi csapadék 29,6 cm.

Megoldás. *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a minta tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A2` cellára, és *Ctrl+V* segítségével illessze be a mintarealizációt.

Az $\widehat{a}_1, \widehat{a}_0$ együtthatók kiszámolásához jelölje ki a D2:E2 cellatartományt, majd gépelje be a következőt:

$$\boxed{\texttt{=LIN.ILL(A2:A19;B2:B19)}}.$$

Végül nyomja meg a *Shift+Ctrl+Enter* billentyűket. Ezt az úgynevezett tömbképleteknél, mint ez is, mindig így kell csinálni. Ennek hatására D2 fogja $\widehat{a}_1$ értékét, illetve E2 fogja $\widehat{a}_0$ értékét tartalmazni. Megjegyezzük, hogy ebben az esetben a következőképpen is számolhatott volna:

$$\widehat{a}_1 = \boxed{\texttt{KOVAR(A2:A19;B2:B19)/VARP(B2:B19)}}$$
$$\widehat{a}_0 = \boxed{\texttt{ÁTLAG(A2:A19)-a\_1*ÁTLAG(B2:B19)}}$$

ahol az $\widehat{a}_1$ értékét tartalmazó cella neve a_1. Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (A2:A19), majd

Beszúrás → *Diagramok/Pont/Pont csak jelölőkkel*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

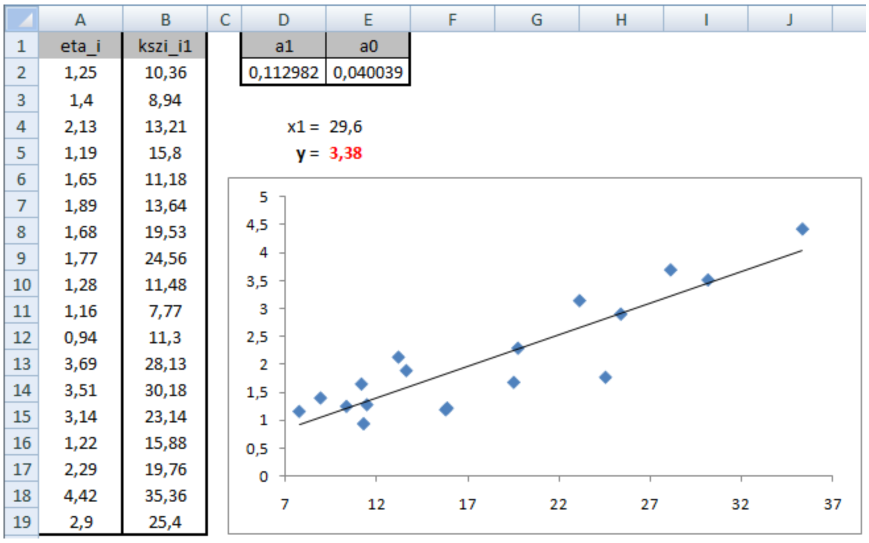
Szerkesztés → *Adatsor X értékei:* =Munka1!$B$2:$B$19 → *OK* → *OK*

Ezzel megjelentek 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. Válassza ki a *lineáris* típust, majd *Bezárás*.

Az E4 cellába írja be a 29,6 értéket, majd az E5 cellába, hogy

$$\boxed{\texttt{=TREND(A2:A19;B2:B19;E4)}}.$$

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. Megjegyezzük, hogy a 3,38 értéket $\boxed{\texttt{=E2+D2*E4}}$ módon is megkaphatjuk.



8.2. 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 `minta-32.txt` fájlt, 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 `E2:G2` cellatartományt, majd gépelje be a következőt:

$$\boxed{\texttt{=LIN.ILL(A1:A26;B1:C26)}}.$$

Végül nyomja meg a *Shift+Ctrl+Enter* billentyűket. Ennek hatására az $\widehat{a}_2, \widehat{a}_1, \widehat{a}_0$ értékek rendre megjelennek az `E2`, `F2`, `G2` cellákban. Az `E5` cellába gépelje be a 102 értéket, az `F5` cellába a 648 értéket, majd az `E7` cellába, hogy

$$\boxed{\texttt{=TREND(A1:A26;B1:C26;E5:F5)}}.$$

A kapott érték 800 cm kerekítve. Tehát az adatok alapján a Duna árhullámának becsült tetőző vízállása Budapesten 800 cm lesz.

	A	B	C	D	E	F	G
1	590	58	405		a2	a1	a0
2	660	52	450		0,43882371	2,345935	276,366
3	780	133	350				
4	770	179	285		x1	x2	
5	710	96	330		102	648	
6	640	72	400				
7	670	72	550	y =	800		
8	520	43	480				

8.3. Példa. Az előző példát oldja meg a `LIN.ILL` illetve `TREND` függvények használata nélkül is, csak a normálegyenlet segítségével.

Megoldás. Használja az előző munkalapot. Jelölje ki a `B` oszlopot, majd helyi menüből válassza a *Beszúrás* pontot. A `B1` cellába írja be, hogy 1, majd a kitöltőjelre klikkeljen kétszer. Nevezze el a `B1:D26` tömböt `X`-nek, illetve az `A1:A26` tömböt `Y`-nak. Ezután jelölje ki a `K1:K3` tömböt, írja a szerkesztőlécbe, hogy

`=MSZORZAT(INVERZ.MÁTRIX(MSZORZAT(TRANSZPONÁLÁS(X);X));MSZORZAT(TRANSZPONÁLÁS(X);Y))`,

majd nyomja meg a *Shift+Ctrl+Enter* billentyűket. Ennek hatására az $\widehat{a}_0, \widehat{a}_1, \widehat{a}_2$ értékek rendre megjelennek az `K1`, `K2`, `K3` cellákban.

	A	B	C	D	E	F	G	H	I	J	K
1	590	1	58	405		a2	a1	a0		a0	276,366
2	660	1	52	450		0,43882371	2,345935	276,366		a1	2,345935
3	780	1	133	350						a2	0,438824
4	770	1	179	285		x1	x2				
5	710	1	96	330		102	648			y =	800
6	640	1	72	400							
7	670	1	72	550	y =	800					

Végül a `K5` cellába írja be, hogy `=K1+K2*F5+K3*G5`. Az ábrán láthatjuk, hogy pontosan azokat az eredményeket kaptuk, mint az előbb.

8.2. Fixpontos lineáris regresszió

A lineáris regresszió feladata tovább szűkíthető, ha tudjuk, hogy a keresett lineáris függvény áthalad egy rögzített ponton.

Legyenek $t_0, \ldots, t_k \in \mathbb{R}$ rögzített konstansok. Az $\mathrm{E}\big(\eta - g(\xi_1, \ldots, \xi_k)\big)^2$ minimumát keressük azon

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

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 azzal ekvivalens, hogy

$$g(x_1, \dots, x_k) = t_0 + a_1(x_1 - t_1) + \dots + a_k(x_k - t_k) \ (a_1, \dots, a_k \in \mathbb{R}).$$

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 becslést adhatunk.

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.

8.4. Példa. Jelentse η egy vizsgált ellenálláson átfolyó áram erősségét Amperben, illetve ξ_1 az ellenállásra adott feszültséget Voltban. Az (η, ξ_1)-re vonatkozó 10 mérésből származó mintarealizációt a `minta-33.txt` fájl tartalmazza. Természetesen $\xi_1 = 0$ 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 `minta-33.txt` fájlt, 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 a `D1` cellába gépelje be a következőt:

$$\boxed{\texttt{=LIN.ILL(A1:A10;B1:B10;HAMIS)}}.$$

A kapott érték 0,0254 négy tizedesjegyre kerekítve, így a továbbiakban az $\eta \simeq \simeq 0{,}0254\xi_1$ közelítést lehet használni. (Vagyis az ellenállás becslése $R \simeq \frac{1}{0{,}0254} \simeq \simeq 39{,}4\,\mathrm{Ohm}$.) Most következik a grafikon. Jelölje ki az η-ra vonatkozó `A1:A10` mintát, majd

Beszúrás → Diagramok/Pont/Pont csak jelölőkkel

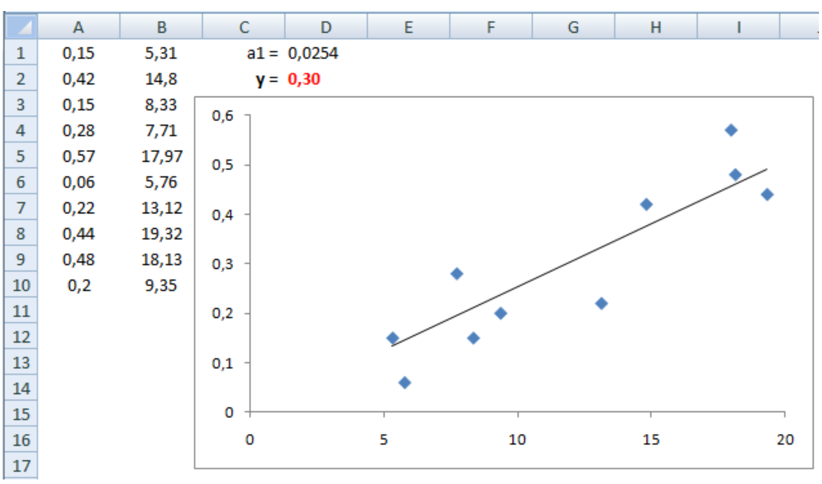
Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés → Adatsor X értékei: `=Munka1!$B$1:$B$10` → *OK* → *OK*

Ezzel megjelentek a mintarealizáció pontjai. Következzen a 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, állítsa 0-ra (ez a t_0 értéke), majd *Bezárás*. (Az ábra csak akkor helyes, ha $t_1 = 0$, mert annak értékét nem lehet állítani.) A `D2` cellába írja be, hogy

$$\boxed{\texttt{=TREND(A1:A10;B1:B10;12;HAMIS)}}.$$

A kapott érték 0,30 két tizedesjegyre kerekítve. Tehát 12 V feszültség esetén az átfolyó áram erősségét 0,3 A-ra becsüljük.



8.5. Példa. 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 $(t_1, t_2, t_0) = (\frac{2}{5}, \frac{3}{4}, 1)$ fixpont esetén. Ebből adjon becslést η-ra, ha $\xi_1 = 1{,}3$ és $\xi_2 = 7{,}5$.

Megoldás. Nyissa meg a `minta-34.txt` fájlt, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A2` 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 `G2`, `H2`, `I2` cellákba. A `D2` cellába írja be, hogy `=A2-G$2`. A kitöltőjelet húzza `F2`-ig, majd a kitöltőjelre klikkeljen kétszer.

Az $\widehat{a}_2, \widehat{a}_1$ kiszámolásához jelölje ki a `H5:I5` tartományt, gépelje be a következőt:

`=LIN.ILL(D2:D16;E2:F16;HAMIS)`,

majd nyomja meg a *Shift+Ctrl+Enter* billentyűket. A kapott értékek 1,9983 és 0,3312 négy tizedesjegyre kerekítve, így a továbbiakban az

$$\eta \simeq 1 + 0{,}3312(\xi_1 - 0{,}4) + 1{,}9983(\xi_2 - 0{,}75)$$

közelítést lehet használni. Az $x_1 - t1 = 1{,}3 - 0{,}4$ és $x_2 - t_2 = 7{,}5 - 0{,}75$ értékeket gépelje a `H8` és `I8` cellákba, majd az `I10`-be, hogy

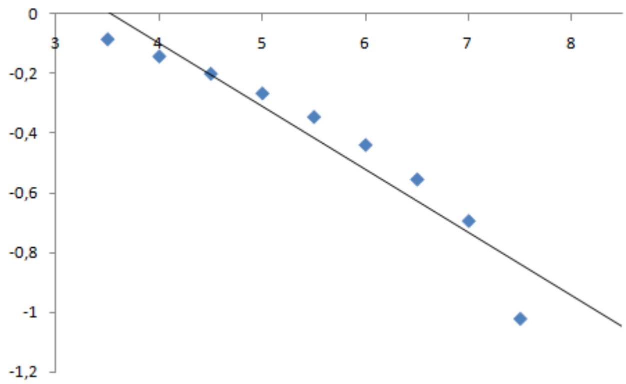
`=G2+TREND(D2:D16;E2:F16;H8:I8;HAMIS)`.

A kapott érték 14,79 két tizedesjegyre kerekítve. Tehát $\eta \simeq 14{,}79$, ha $\xi_1 = 1{,}3$ és $\xi_2 = 7{,}5$.

	A	B	C	D	E	F	G	H	I
1	eta	xi1	xi2	eta-t0	xi1-t1	xi2-t2	t0	t1	t2
2	12,76	0,57	6,62	11,76	0,17	5,87	1	0,4	0,75
3	19,3	1,2	9,75	18,3	0,8	9			
4	5,03	1,84	2,29	4,03	1,44	1,54		a2	a1
5	18	3,87	8,81	17	3,47	8,06		1,9983	0,3312
6	3,63	1,9	2,07	2,63	1,5	1,32			
7	7,86	2,18	3,62	6,86	1,78	2,87		x1-t1	x2-t2
8	8,41	3,07	4,14	7,41	2,67	3,39		0,9	6,75
9	17,27	3,63	8,12	16,27	3,23	7,37			
10	15,51	0,13	8,24	14,51	-0,27	7,49		y =	14,79
11	16,04	2,72	8,18	15,04	2,32	7,43			
12	5,97	0,8	2,98	4,97	0,4	2,23			
13	7,72	1,32	3,86	6,72	0,92	3,11			
14	17,17	0,01	8,8	16,17	-0,39	8,05			
15	9,51	2,47	4,46	8,51	2,07	3,71			
16	3,17	3,43	1,46	2,17	3,03	0,71			

8.3. Nemlineáris regresszió

A lineáris regressziós közelítés sokszor nagyon durva becslést adhat. A mintarealizációt jelentő pontok ábrázolásával $k = 1$ esetén, jól szemléltethető ez a probléma.



Látszik, hogy ebben az esetben „hiba” lenne lineáris regressziót alkalmazni. Ilyenkor érdemes megtippelni, 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.

8.3.1. 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 $\eta, \xi_1, \xi_1^2, \ldots, \xi_1^r$ között végrehajtott lineáris regresszió adja.

8.6. Példa. Az (η, ξ_1)-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 `minta-35.txt` fájlt, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A2` cellára, és *Ctrl+V* segítségével illessze be a mintarealizációt. A `C2` cellába írja be, hogy `=B2^2`, majd a kitöltőjelre klikkeljen kétszer. Az $\widehat{a}_2, \widehat{a}_1, \widehat{a}_0$ kiszámolásához jelölje ki az `E2:G2` tartományt, gépelje be a következőt:

`=LIN.ILL(A2:A11;B2:C11)`,

majd nyomja meg a *Shift+Ctrl+Enter* billentyűket. A kapott értékek −4,3773, 15,6534 és −7,2032 négy tizedesjegyre kerekítve, így a másodfokú polinomos regressziós függvény becslése:

$$y = -4{,}3773x^2 + 15{,}6534x - 7{,}2032.$$

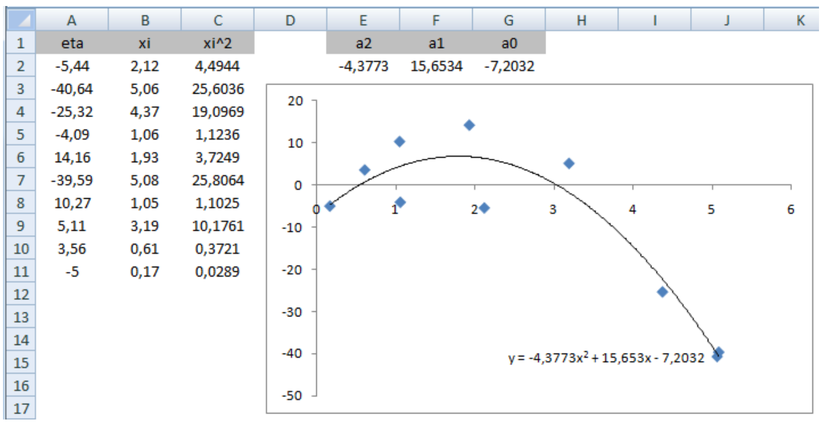
Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A2:A11`), majd

Beszúrás $\rightarrow$ *Diagramok/Pont/Pont csak jelölőkkel*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés $\rightarrow$ *Adatsor X értékei:* `=Munka1!$B$2:$B$11` $\rightarrow$ *OK* $\rightarrow$ *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. Válassza ki a *Polinomiális (Sorrend:2)* típust, majd *Bezárás*.



	A	B	C	D	E	F	G
1	eta	xi	xi^2		a2	a1	a0
2	-5,44	2,12	4,4944		-4,3773	15,6534	-7,2032
3	-40,64	5,06	25,6036				
4	-25,32	4,37	19,0969				
5	-4,09	1,06	1,1236				
6	14,16	1,93	3,7249				
7	-39,59	5,08	25,8064				
8	10,27	1,05	1,1025				
9	5,11	3,19	10,1761				
10	3,56	0,61	0,3721				
11	-5	0,17	0,0289				

8.3.2. 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 $\ln \xi_1$ között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy $a_0 = \ln a,\ a_1 = b$, azaz $a = e^{a_0},\ b = a_1$.

8.7. Példa. Az (η, ξ_1)-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 `minta-36.txt` fájlt, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A2` cellára, és *Ctrl+V* segítségével illessze be a mintarealizációt. A `C2` cellába írja be, hogy `=LN(A2)`, a kitöltőjelet húzza a `D2` celláig, majd a kitöltőjelre klikkeljen kétszer. Az $\widehat{a}_1, \widehat{a}_0$ kiszámolásához jelölje ki az `F2:G2` tartományt, gépelje be a következőt:

`=LIN.ILL(C2:C21;D2:D21)`,

majd nyomja meg a *Shift+Ctrl+Enter* billentyűket. Ekkor $a =$ `KITEVŐ(G2)` $=$ 3,0982 és $b = a_1 = 3{,}4833$ négy tizedesjegyre kerekítve, így a hatványkitevős regressziós függvény becslése:

$$y = 3{,}0982x^{3{,}4833}.$$

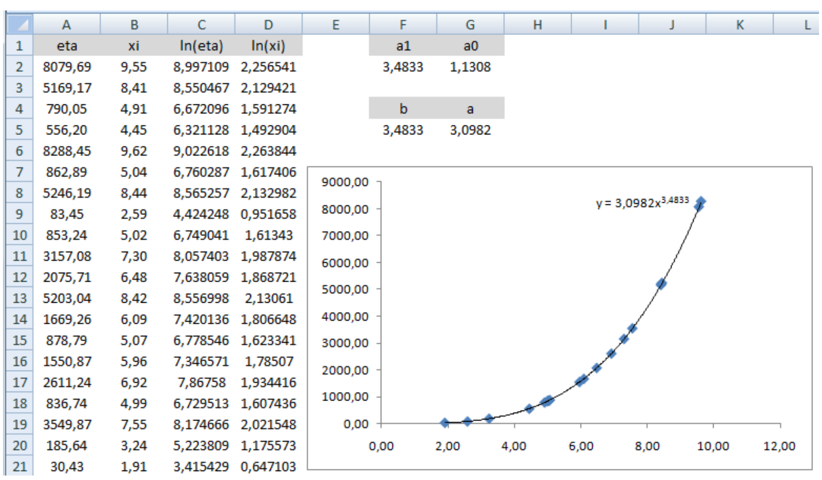
Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A2:A21`), majd

Beszúrás → Diagramok/Pont/Pont csak jelölőkkel

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés → Adatsor X értékei: `=Munka1!$B$2:$B$21` *→ OK → 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. Válassza ki a *Hatványos* típust, majd *Bezárás*.



	A	B	C	D	E	F	G
1	eta	xi	ln(eta)	ln(xi)		a1	a0
2	8079,69	9,55	8,997109	2,256541		3,4833	1,1308
3	5169,17	8,41	8,550467	2,129421			
4	790,05	4,91	6,672096	1,591274		b	a
5	556,20	4,45	6,321128	1,492904		3,4833	3,0982
6	8288,45	9,62	9,022618	2,263844			
7	862,89	5,04	6,760287	1,617406			
8	5246,19	8,44	8,565257	2,132982			
9	83,45	2,59	4,424248	0,951658			
10	853,24	5,02	6,749041	1,61343			
11	3157,08	7,30	8,057403	1,987874			
12	2075,71	6,48	7,638059	1,868721			
13	5203,04	8,42	8,556998	2,13061			
14	1669,26	6,09	7,420136	1,806648			
15	878,79	5,07	6,778546	1,623341			
16	1550,87	5,96	7,346571	1,78507			
17	2611,24	6,92	7,86758	1,934416			
18	836,74	4,99	6,729513	1,607436			
19	3549,87	7,55	8,174666	2,021548			
20	185,64	3,24	5,223809	1,175573			
21	30,43	1,91	3,415429	0,647103			

8.3.3. 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_1 között lineáris regressziót végrehajtva, a kapott a_0, a_1 együtthatókra teljesül, hogy $a_0 = \ln a,\ a_1 = \ln b$, azaz $a = e^{a_0},\ b = e^{a_1}$.

8.8. Példa. Az (η, ξ_1)-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 $\xi_1 = 5$.

Megoldás. Nyissa meg a `minta-37.txt` fájlt, majd *Ctrl+A* és *Ctrl+C* segítségével tegye a vágólapra a tartalmát. Nyisson meg egy üres munkalapot Excelben, lépjen az `A2` 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 `D2:E2` tartományt, gépelje be a következőt:

`=LOG.ILL(A2:A11;B2:B11)`,

majd nyomja meg a *Shift+Ctrl+Enter* billentyűket. Ekkor $b = 3{,}0495$ és $a = 4{,}8127$ négy tizedesjegyre kerekítve, így az exponenciális regressziós függvény becslése:

$$y = 4{,}8127 \cdot 3{,}0495^x.$$

Ezután az `E4` cellába írja be, hogy

`=NÖV(A2:A11;B2:B11;5)`.

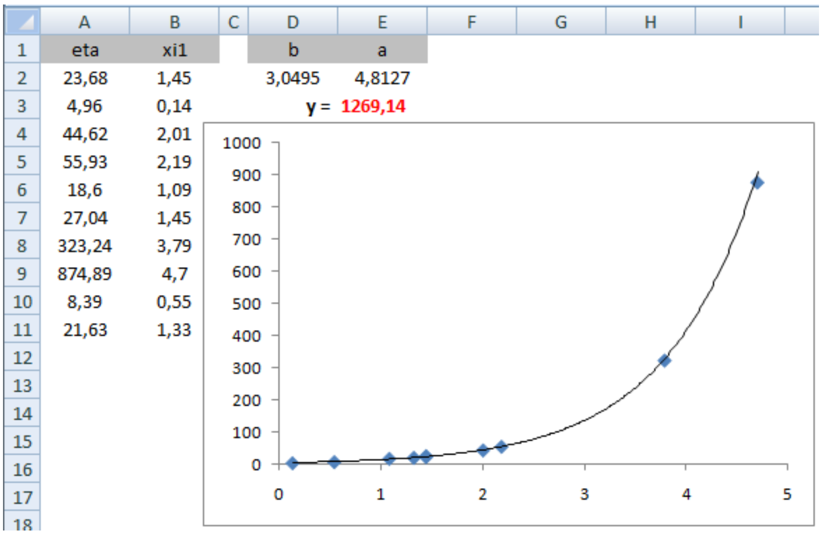
A kapott 1269,14 érték az η becslése $\xi_1 = 5$ esetén. Most következik a grafikon. Jelölje ki az η-ra vonatkozó mintát (`A2:A11`), majd

Beszúrás $\rightarrow$ *Diagramok/Pont/Pont csak jelölőkkel*

Lépjen a diagramterületre, majd helyi menüből (jobb egérgomb) válassza az *Adatok kijelölése* pontot.

Szerkesztés $\rightarrow$ *Adatsor X értékei:* `=Munka1!$B$2:$B$11` $\rightarrow$ *OK* $\rightarrow$ *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. Válassza ki a *Exponenciális* típust, majd *Bezárás*.



8.3.4. 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 $\ln \xi_1$ között lineáris regressziót végrehajtva, $a = a_0, \; b = a_1$.

8.3.5. 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_1 között lineáris regressziót végrehajtva, $a = a_0, \; b = a_1$.

8.4. Gyakorlatok

8.1. 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 $\xi_1 = 6{,}3,\ \ \xi_2 = 0{,}7,\ \ \xi_3 = 0{,}9$.

8.2. 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.

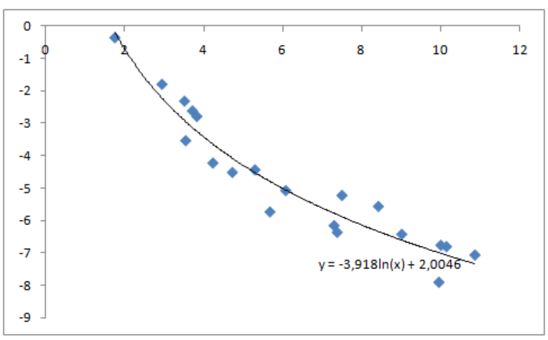
8.3. gyakorlat. Az $(\eta, \xi_1, \xi_2, \xi_3)$ valószínűségi vektorváltozóra vonatkozó mintarealizációt a `minta-38.txt` fájl tartalmazza. Ez alapján becsülje meg a $(t_1, t_2, t_3, t_0) = (1, 1, 1, 1)$ fixpontos lineáris regresszió együtthatóit, majd ebből η értékét, ha $\xi_1 = 6{,}3,\ \ \xi_2 = 0{,}7,\ \ \xi_3 = 0{,}9$.

8.4. 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.

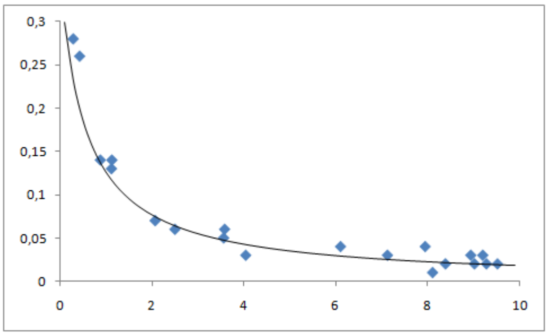
8.5. gyakorlat. Az (η, ξ_1)-re vonatkozó mintarealizációt a `minta-40.txt` fájl tartalmazza. Ez alapján becsülje meg a logaritmikus regressziós függvényt. A kapott függvényt ábrázolja a mintarealizációval együtt. Becsülje meg ebből η értékét, ha $\xi_1 = 5{,}3$.

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.



8.6. gyakorlat. Az (η, ξ_1)-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 $\xi_1 = 4{,}2$.

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.

9. Összefoglaló

9.1. Eloszlások generálása

9.1.1. 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 $\sum_{i=1}^{r} \mathrm{I}_{\eta_i < p}$ r-edrendű p paraméterű binomiális eloszlású.

- **Hipergeometrikus eloszlás**
  Legyen $r, M, N \in \mathbb{N}$, $M < N$, továbbá $r \leqslant \min\{ M, N - M \}$. 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
  $$\mathrm{P}(\xi_r = k) = \frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}} \quad (k = 0, \ldots, r),$$
  azaz ξ_r hipergeometrikus eloszlású N, M, r paraméterekkel.

- **Poisson-eloszlás**
  Ha $\lambda > 0$, akkor
  $$\min\Big\{ s : \prod_{i=0}^{s} \eta_i < e^{-\lambda} \Big\}$$
  Poisson-eloszlású λ paraméterrel.

- **Geometriai eloszlás**
  Ha $0 < p < 1$, akkor $\min\{ s : \eta_s < p \}$ 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.

- **Normális eloszlás**
  Ha $m \in \mathbb{R}$ és $\sigma > 0$, akkor
  $$m + \sigma\sqrt{-2\ln \eta_1}\cos(2\pi\eta_2)$$
  normális eloszlású m várható értékkel és σ szórással.

9.1.2. Normális eloszlásból származtatott eloszlások

Itt az η, η_i $(i \in \mathbb{N})$ független standard normális eloszlású valószínűségi változókat jelent.

- **Khi-négyzet eloszlás**
  Ha $s \in \mathbb{N}$, akkor $\sum_{i=1}^{s} \eta_i^2$ khi-négyzet eloszlású s szabadsági fokkal.

- **t-eloszlás**
  Ha $s \in \mathbb{N}$, akkor $\eta\sqrt{\frac{s}{\sum_{i=1}^{s}\eta_i^2}}$ t-eloszlású s szabadsági fokkal.

- **Cauchy-eloszlás**
  $\frac{\eta_1}{\eta_2}$ Cauchy-eloszlású.

- **F-eloszlás**
  Ha $s_1, s_2 \in \mathbb{N}$, akkor $\frac{s_2\sum_{i=1}^{s_1}\eta_i^2}{s_1\sum_{i=s_1+1}^{s_1+s_2}\eta_i^2}$ F-eloszlású s_1 és s_2 szabadsági fokkal.

9.2. Grafikus illeszkedésvizsgálat

Legyen $x_1, \ldots, x_r \in \mathbb{R}$ és $x_1 < x_2 < \cdots < x_r$. Jelölje n a mintarealizáció elemeinek a számát és k_i az x_i-nél kisebb elemek számát a mintarealizációban.

- **Normalitásvizsgálat**
  Ha teljesül, hogy a vizsgált valószínűségi változó normális eloszlású m várható értékkel és σ szórással, akkor $\boxed{y_i := \Phi^{-1}\left(\frac{k_i}{n}\right)}$ jelöléssel az (x_i, y_i) $(i = 1, \dots, r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $\frac{1}{\sigma}$ a meredeksége és $-\frac{m}{\sigma}$ értéknél metszi a függőleges tengelyt.

- **Exponencialitásvizsgálat**
  Ha teljesül, hogy a vizsgált valószínűségi változó exponenciális eloszlású λ paraméterrel, akkor $\boxed{y_i := \ln\left(1 - \frac{k_i}{n}\right)}$ jelöléssel az (x_i, y_i) $(i = 1, \dots, r)$ koordinátájú pontok körülbelül egy olyan egyenesre esnek, melynek $-\lambda$ a meredeksége és átmegy az origón.

9.3. Intervallumbecslések

Legyen a ξ valószínűségi változóra vonatkozó minta $\xi_1, \dots, \xi_n$, és $1-\alpha$ a becsülendő paraméterre vonatkozó $[\tau_1, \tau_2]$ konfidenciaintervallum biztonsági szintje.

- $\xi \in \mathbf{Norm}(m; \sigma)$
  m az ismeretlen becsülendő paraméter, σ ismert
  $\tau_1 = \overline{\xi} - \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
  $\tau_2 = \overline{\xi} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \mathbf{Norm}(m; \sigma)$
  m ismert, σ az ismeretlen becsülendő paraméter
  $F \sim \mathbf{Khi}(n)$
  $\tau_1 = \sqrt{\frac{\sum_{i=1}^{n}(\xi_i - m)^2}{F^{-1}\left(1-\frac{\alpha}{2}\right)}}$
  $\tau_2 = \sqrt{\frac{\sum_{i=1}^{n}(\xi_i - m)^2}{F^{-1}\left(\frac{\alpha}{2}\right)}}$

- $\xi \in \mathbf{Norm}(m; \sigma)$
  m ismeretlen, σ az ismeretlen becsülendő paraméter
  $n \geqslant 2,\ F \sim \mathbf{Khi}(n-1)$
  $\tau_1 = S_n \sqrt{\frac{n}{F^{-1}\left(1-\frac{\alpha}{2}\right)}}$
  $\tau_2 = S_n \sqrt{\frac{n}{F^{-1}\left(\frac{\alpha}{2}\right)}}$

- $\xi \in \mathbf{Norm}(m; \sigma)$
  m az ismeretlen becsülendő paraméter, σ ismeretlen

$n \geqslant 2,\ F \sim \mathbf{t}(n-1)$
$\tau_1 = \overline{\xi} - \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right)$
$\tau_2 = \overline{\xi} + \frac{S_n^*}{\sqrt{n}} F^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \mathbf{Exp}(\lambda)$
  λ az ismeretlen becsülendő paraméter
  $F \sim \mathbf{Gamma}(n;1)$
  $\tau_1 = \dfrac{1}{n\overline{\xi}} F^{-1}\left(\frac{\alpha}{2}\right)$
  $\tau_2 = \dfrac{1}{n\overline{\xi}} F^{-1}\left(1 - \frac{\alpha}{2}\right)$

- $\xi \in \mathbf{Bin}(1;p)$
  p az ismeretlen becsülendő paraméter
  $\tau_1 = \frac{1}{n} \max\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^i (1-\overline{\xi})^{n-i} < \frac{\alpha}{2} \right\}$
  $\tau_2 = \frac{1}{n} \min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} \overline{\xi}^i (1-\overline{\xi})^{n-i} \geqslant 1 - \frac{\alpha}{2} \right\}$
  Nagy n-re:
  $c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
  $\tau_1 = \dfrac{\overline{\xi} + \frac{c^2}{2n} - \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \simeq \overline{\xi} - \dfrac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}$
  $\tau_2 = \dfrac{\overline{\xi} + \frac{c^2}{2n} + \frac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi}) + \frac{c^2}{4n}}}{1 + \frac{c^2}{n}} \simeq \overline{\xi} + \dfrac{c}{\sqrt{n}}\sqrt{\overline{\xi}(1-\overline{\xi})}$

- ξ az $[a,b]$ intervallumon egyenletes eloszlású
  a ismert, b az ismeretlen becsülendő paraméter
  $F \sim \mathbf{Gamma}(n;1),\ c_1 = F^{-1}\left(\frac{\alpha}{2}\right),\ c_2 = F^{-1}\left(1 - \frac{\alpha}{2}\right)$
  $\tau_1 = a + \left(e^{c_1} \prod_{i=1}^{n} (\xi_i - a)\right)^{\frac{1}{n}}$
  $\tau_2 = a + \left(e^{c_2} \prod_{i=1}^{n} (\xi_i - a)\right)^{\frac{1}{n}}$

9.4. Paraméteres hipotézisvizsgálatok

A következőkben $1-\alpha$ a próba szintjét jelenti.

- **Egymintás u-próba**
  $\xi \in \mathbf{Norm}(m;\sigma)$, m ismeretlen, σ ismert, $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta, $m_0 \in \mathbb{R}$ rögzített.
  $\boxed{H_0\colon m = m_0}$
  $u = \dfrac{\overline{\xi} - m_0}{\sigma}\sqrt{n}$

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

- **Kétmintás u-próba**

  $\xi \in \mathbf{Norm}(m_1; \sigma_1)$, $\eta \in \mathbf{Norm}(m_2; \sigma_2)$ függetlenek, m_1, m_2 ismeretlenek, σ_1, σ_2 ismertek, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta.

  $\boxed{H_0\colon m_1 = m_2}$

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

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

- **Egymintás t-próba**

  $\xi \in \mathbf{Norm}(m; \sigma)$, m, σ ismeretlenek, $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta, $n \geqslant 2$, $m_0 \in \mathbb{R}$ rögzített.

  $\boxed{H_0\colon m = m_0}$

  $t = \dfrac{\overline{\xi} - m_0}{S_n^*}\sqrt{n} \quad$ és $\quad F \sim \mathbf{t}(n-1)$

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

- **Kétmintás t-próba**

  $\xi \in \mathbf{Norm}(m_1; \sigma_1)$, $\eta \in \mathbf{Norm}(m_2; \sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\sigma_1 = \sigma_2$, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta, $n_1 \geqslant 2$, $n_2 \geqslant 2$.

  $\boxed{H_0\colon m_1 = m_2}$

  $t = \frac{\overline{\xi} - \overline{\eta}}{\sqrt{n_1 S_{\xi,n_1}^2 + n_2 S_{\eta,n_2}^2}}\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \quad$ és $\quad F \sim \mathbf{t}(n_1 + n_2 - 2)$

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

- **Scheffé-módszer**

$\xi \in \mathbf{Norm}(m_1;\sigma_1)$, $\eta \in \mathbf{Norm}(m_2;\sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta, $2 \leqslant n_1 \leqslant n_2$.

$\boxed{H_0\colon m_1 = m_2}$

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

($n_1 = n_2$ esetén $\zeta_i = \xi_i - \eta_i$)

$t = \frac{\overline{\zeta}}{S^*_{\zeta,n_1}}\sqrt{n_1}$ és $F \sim \mathbf{t}(n_1 - 1)$

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

$n_1 = n_2$ esetén a módszer akkor is alkalmazható, ha a minták nem függetlenek, de csak akkor, ha $\xi - \eta$ normális eloszlású.

- **F-próba**

$\xi \in \mathbf{Norm}(m_1;\sigma_1)$, $\eta \in \mathbf{Norm}(m_2;\sigma_2)$ függetlenek, $m_1, m_2, \sigma_1, \sigma_2$ ismeretlenek, $\xi_1, \ldots, \xi_{n_1}$ a ξ-re vonatkozó, illetve $\eta_1, \ldots, \eta_{n_2}$ az η-ra vonatkozó minta ($n_1 \geqslant 2$, $n_2 \geqslant 2$).

$\boxed{H_0\colon \sigma_1 = \sigma_2}$

$\mathsf{F} = \dfrac{{S^*_{\xi,n_1}}^2}{{S^*_{\eta,n_2}}^2}$ és $F \sim \mathbf{F}(n_1 - 1; n_2 - 1)$

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

- **Khi-négyzet próba**

$\xi \in \mathbf{Norm}(m;\sigma)$, m, σ ismeretlenek, $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta ($n \geqslant 2$).

$\boxed{H_0\colon \sigma = \sigma_0}$

$\chi^2 = \dfrac{S_n^2}{\sigma_0^2} n$ és $F \sim \mathbf{Khi}(n-1)$

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

- **Statisztikai próba az exponenciális eloszlás paraméterére**

  $\xi \in \mathbf{Exp}(\lambda)$, λ ismeretlen, $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta, $\lambda_0 \in \mathbb{R}_+$ rögzített.

  $\boxed{H_0 \colon \lambda = \lambda_0}$

  $\gamma = \lambda_0 n \overline{\xi}$ és $F \sim \mathbf{Gamma}(n; 1)$

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

- **Statisztikai próba valószínűségre**

  $\xi \in \mathbf{Bin}(1; p)$, p ismeretlen, $0 < p_0 < 1$ rögzített és $\xi_1, \ldots, \xi_n$ a ξ-re vonatkozó minta.

  $\boxed{H_0 \colon p = p_0}$

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

  $\min\{\, np_0, n(1-p_0) \,\} \geqslant 10$ esetén

  $F^{-1}(x) \simeq np_0 - \frac{1}{2} + \sqrt{np_0(1-p_0)}\,\Phi^{-1}(x)$

H_1	kritikus tartomány
$p \neq p_0$	$n\overline{\xi} < F^{-1}\left(\frac{\alpha}{2}\right)$ vagy $n\overline{\xi} > F^{-1}\left(1 - \frac{\alpha}{2}\right)$
$p < p_0$	$n\overline{\xi} < F^{-1}(\alpha)$
$p > p_0$	$n\overline{\xi} > F^{-1}(1-\alpha)$
	feltétel
$p \neq p_0$	$1 \leqslant F^{-1}\left(\frac{\alpha}{2}\right) < np_0 < F^{-1}\left(1 - \frac{\alpha}{2}\right) \leqslant n-1$
$p < p_0$	$1 \leqslant F^{-1}(\alpha) < np_0$
$p > p_0$	$np_0 < F^{-1}(1-\alpha) \leqslant n-1$

9.5. Nemparaméteres hipotézisvizsgálatok

A következőkben $1 - \alpha$ a próba szintjét jelenti.

- **Tiszta illeszkedésvizsgálat valószínűségre**

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

  $\boxed{H_0\colon\ \mathrm{P}(A_i) = p_i\ \forall i}$, ahol P a valódi valószínűség

  $\varrho_i (\geqslant 10)$ az A_i gyakorisága n kísérlet után

  $$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - np_i)^2}{np_i} \quad \text{és} \quad F \sim \mathbf{Khi}(r-1)$$

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

- **Tiszta illeszkedésvizsgálat eloszlásfüggvényre**

  ξ-re vonatkozó minta $\xi_1, \dots, \xi_n$

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

  $a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < a_r = \infty$

  $$\varrho_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \geqslant 10$$

  $p_i = \mathrm{P}(a_{i-1} \leqslant \xi < a_i) = F_0(a_i) - F_0(a_{i-1})$, azaz $\mathrm{P} \in \mathcal{P}_{H_0}$

  $$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - np_i)^2}{np_i} \quad \text{és} \quad F \sim \mathbf{Khi}(r-1)$$

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

- **Becsléses illeszkedésvizsgálat**

  ξ-re vonatkozó minta $\xi_1, \dots, \xi_n$

  F_ϑ eloszlásfüggvény minden $\vartheta = (\vartheta_1, \dots, \vartheta_v) \in \Theta \subset \mathbb{R}^v$ esetén.

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

  $a_0 = -\infty < a_1 < a_2 < \cdots < a_{r-1} < a_r = \infty$

  $$\varrho_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \geqslant 10$$

  $\widehat{\vartheta}_i$ a ϑ_i maximum likelihood becslése H_0 feltételezésével

  $\widehat{p}_i = \mathrm{P}_{(\widehat{\vartheta}_1, \dots, \widehat{\vartheta}_v)}(a_{i-1} \leqslant \xi < a_i) = F_{(\widehat{\vartheta}_1, \dots, \widehat{\vartheta}_v)}(a_i) - F_{(\widehat{\vartheta}_1, \dots, \widehat{\vartheta}_v)}(a_{i-1})$

  $$\chi^2 = \sum_{i=1}^{r} \frac{(\varrho_i - n\widehat{p}_i)^2}{n\widehat{p}_i} \quad \text{és} \quad F \sim \mathbf{Khi}(r-1-v)$$

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

- **Függetlenségvizsgálat eseményrendszerekre**

  $A_1, \dots, A_r$ és $B_1, \dots, B_s$ két teljes eseményrendszer.

  $\boxed{H_0\colon\ \mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\,\mathrm{P}(B_j)\ \forall i, j}$, ahol P a valódi valószínűség.

  A kontingencia táblázat

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_s & \\
\hline
A_1 & \varrho_{11} & \varrho_{12} & \dots & \varrho_{1s} & k_1 \\
A_2 & \varrho_{21} & \varrho_{22} & \dots & \varrho_{2s} & k_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_r & \varrho_{r1} & \varrho_{r2} & \dots & \varrho_{rs} & k_r \\
\hline
 & l_1 & l_2 & \dots & l_s & n
\end{array}
$$

$\varrho_{ij} \geqslant 10$ minden i, j esetén

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \frac{1}{n} k_i l_j)^2}{\frac{1}{n} k_i l_j} \qquad \text{és} \quad F \sim \mathbf{Khi}\big(_{(r-1)(s-1)}\big)$$

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

- **Függetlenségvizsgálat valószínűségi változókra**

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

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

  $a_0 = -\infty < a_1 < a_2 < \dots < a_{r-1} < a_r = \infty$

  $b_0 = -\infty < b_1 < b_2 < \dots < b_{s-1} < a_s = \infty$

  $$k_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \qquad l_j = \sum_{z=1}^{n} \mathrm{I}_{\eta_z \in [b_{j-1}, b_j)}$$

  $$\varrho_{ij} = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [a_{i-1}, a_i)} \, \mathrm{I}_{\eta_z \in [b_{j-1}, b_j)} \geqslant 10$$

  $$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\varrho_{ij} - \frac{1}{n} k_i l_j)^2}{\frac{1}{n} k_i l_j} \qquad \text{és} \quad F \sim \mathbf{Khi}\big(_{(r-1)(s-1)}\big)$$

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

- **Homogenitásvizsgálat**

  ξ és η független valószínűségi változók, az ezekre vonatkozó minták $\xi_1, \dots, \xi_{n_1}$ illetve $\eta_1, \dots, \eta_{n_2}$.

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

  $c_0 = -\infty < c_1 < c_2 < \dots < c_{r-1} < c_r = \infty$

  $$k_i = \sum_{z=1}^{n} \mathrm{I}_{\xi_z \in [c_{i-1}, c_i)} \geqslant 10 \quad l_j = \sum_{z=1}^{n} \mathrm{I}_{\eta_z \in [c_{j-1}, c_j)} \geqslant 10$$

  $$\chi^2 = n_1 n_2 \sum_{i=1}^{r} \frac{\left(\frac{k_i}{n_1} - \frac{l_i}{n_2}\right)^2}{k_i + l_i} \qquad \text{és} \quad F \sim \mathbf{Khi}(r-1)$$

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

- **Kétmintás előjelpróba**

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

$\boxed{H_0:\ \mathrm{P}(\xi > \eta) = \frac{1}{2}}$

$B = \displaystyle\sum_{i=1}^{n} \mathrm{I}_{\xi_i > \eta_i}$

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

$n \geqslant 20$ esetén $F^{-1}(x) \simeq \frac{1}{2}\left(n - 1 + \sqrt{n}\Phi^{-1}(x)\right)$

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

- **Kolmogorov – Szmirnov-féle kétmintás próba**
  ξ és η folytonos eloszlásfüggvényű független valószínűségi változók, az ezekre vonatkozó minták $\xi_1, \ldots, \xi_n$ illetve $\eta_1, \ldots, \eta_n$ $(n > 30)$
  $\boxed{H_0: \xi \text{ és } \eta \text{ azonos eloszlású}}$
  ξ-re illetve η-ra vonatkozó mintákhoz tartozó tapasztalati eloszlásfüggvények F_n^* illetve G_n^*
  $D = \sqrt{\frac{n}{2}} \max_{i=1,\ldots,2n} |F_n^*(x_i) - G_n^*(x_i)|,$
  ahol $x_1 = \xi_1, \ldots, x_n = \xi_n, x_{n+1} = \eta_1, \ldots, x_{2n} = \eta_n$
  $K(z) = 1 + 2\sum_{i=1}^{\infty} (-1)^i e^{-2i^2 z^2}$
  Kritikus tartomány: $K(D) \geqslant 1 - \alpha$

- **Kolmogorov – Szmirnov-féle egymintás próba**
  ξ folytonos eloszlásfüggvényű valószínűségi változó, az erre vonatkozó minta $\xi_1, \ldots, \xi_n$ $(n > 30)$
  $\boxed{H_0: \xi \text{ eloszlásfüggvénye } F}$
  F_n^* a tapasztalati eloszlásfüggvény
  $\widetilde{F}_n^*(x) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}_{\xi_k \leqslant x}$
  $D = \sqrt{n} \max_{i=1,\ldots,n} \max\{\, |F_n^*(x_i) - F(x_i)|, |\widetilde{F}_n^*(x_i) - F(x_i)| \,\}$
  $K(z) = 1 + 2\sum_{i=1}^{\infty} (-1)^i e^{-2i^2 z^2}$
  Kritikus tartomány: $K(D) \geqslant 1 - \alpha$

9.6. Regressziószámítás

Az $\eta, \xi_1, \ldots, \xi_k$ valószínűségi változókra adjuk meg azt az $\eta \simeq g(\xi_1, \ldots, \xi_k)$ közelítést adó g függvényt, melyre $\mathrm{E}\big(\eta - g(\xi_1, \ldots, \xi_k)\big)^2$ minimális. Az ilyen tulajdonságú g függvényt (*regressziós függvény*) a gyakorlatban csak becsülni tudjuk az

$(\eta, \xi_1, \ldots, \xi_k)$ valószínűségi vektorváltozóra vonatkozó

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

minta alapján. Legyen ez a becslés $\widehat{g}$. Ezután az $\boxed{\eta \simeq \widehat{g}(\xi_1, \ldots, \xi_k)}$ közelítést fogjuk használni.

- **Lineáris regresszió**
  A regressziós függvényt csak a

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

  alakú függvények között keressük. Ekkor az

  $$\boxed{\eta \simeq \widehat{a}_0 + \widehat{a}_1 \xi_1 + \cdots + \widehat{a}_k \xi_k}$$

  közelítést fogjuk használni, ahol $\widehat{a}_0, \ldots, \widehat{a}_k$ rendre $a_0, \ldots, a_k$ becslései.

- **Fixpontos lineáris regresszió**
  Legyenek $t_0, \ldots, t_k \in \mathbb{R}$ rögzített konstansok. A regressziós függvényt

  $$\boxed{g(x_1, \ldots, x_k) = t_0 + a_1(x_1 - t_1) + \cdots + a_k(x_k - t_k) \quad (a_1, \ldots, a_k \in \mathbb{R})}$$

  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 $\widehat{a}_1, \ldots, \widehat{a}_k$ rendre $a_1, \ldots, a_k$ becslései.

  A $(t_1, \ldots, t_k, t_0)$ pontot *fixpontnak* nevezzük, mert a kapott g biztosan ráilleszkedik.

- **Polinomos regresszió**
  $k = 1$ és a regressziós függvényt

  $$\boxed{y = a_0 + a_1 x + a_2 x^2 + \cdots + a_r x^r \quad (a_0, \ldots, a_r \in \mathbb{R}_+)}$$

  alakban keressük. Az $a_0, \ldots, a_r$ együtthatókat az $\boxed{\eta, \xi_1, \xi_1^2, \ldots, \xi_1^r}$ 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,

$$\boxed{a = a_0, \ b = a_1}.$$

9.7. Excel függvények

Képlet bevitele

Minden képletet = jellel kell kezdeni. Ha a képlet egyértékű eredményt ad, akkor nyomjon *Enter*-t.

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.

Tömbképlet javítása

Ha egy tömbképletet javítani akar, akkor jelölje ki a tömbképletre vonatkozó tömböt, *F2*, javítás, majd *Ctrl+Shift+Enter*.

Műveletek

$\boxed{+}$ összeadás
$\boxed{-}$ kivonás
$\boxed{*}$ szorzás
$\boxed{/}$ osztás
$\boxed{\hat{}}$ hatványozás

Relációk

$\boxed{=}$ egyenlő
$\boxed{<}$ kisebb
$\boxed{>}$ nagyobb
$\boxed{<=}$ kisebb vagy egyenlő

$\boxed{\texttt{>=}}$ nagyobb vagy egyenlő

$\boxed{\texttt{<>}}$ nem egyenlő

Konstansok

$e = \boxed{\texttt{KITEVŐ(1)}}$

$\pi = \boxed{\texttt{PI()}}$

9.7.1. Logikai függvények

$\boxed{\texttt{HA(feltétel;ha igaz;ha hamis)}}$

$\boxed{\texttt{ÉS(feltétel1;feltétel2;...)}}$

$\boxed{\texttt{VAGY(feltétel1;feltétel2;...)}}$

9.7.2. Elemi függvények

$|x| = \boxed{\texttt{ABS}(x)}\ x \in \mathbb{R}$

$[x] = \boxed{\texttt{INT}(x)}\ x \in \mathbb{R}$

$\operatorname{sign} x = \boxed{\texttt{ELŐJEL}(x)}\ x \in \mathbb{R}$

$\ln x = \boxed{\texttt{LN}(x)}\ x > 0$

$\log_a x = \boxed{\texttt{LOG}(x;a)}\ x > 0,\ a > 0,\ a \neq 1$

$\sqrt{x} = \boxed{\texttt{GYÖK}(x)}\ x \geqslant 0$

$x^a = \boxed{\texttt{HATVÁNY}(x;a)} = \boxed{x\texttt{\^{}}a}$

$e^x = \boxed{\texttt{KITEVŐ}(x)}\ x \in \mathbb{R}$

$\sin x = \boxed{\texttt{SIN}(x)}\ x \in \mathbb{R}$

$\cos x = \boxed{\texttt{COS}(x)}\ x \in \mathbb{R}$

$\operatorname{tg} x = \boxed{\texttt{TAN}(x)}\ x \in \mathbb{R},\ x \neq k\frac{\pi}{2}$, ahol k páratlan egész

$\arcsin x = \boxed{\texttt{ARCSIN}(x)}\ x \in [-1, 1]$

$\arccos x = \boxed{\texttt{ARCCOS}(x)}\ x \in [-1, 1]$

$\operatorname{arctg} x = \boxed{\texttt{ARCTAN}(x)}\ x \in \mathbb{R}$

$\sum_{i=0}^{r} a_i x^{k+im} = \boxed{\texttt{SERIESSUM}(x;k;m;\texttt{A:A})}$, ahol az A oszlopban vannak az $a_0, \ldots, a_r$ valós számok ($x \in \mathbb{R},\ k, m \in \mathbb{N}$)

$\frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) = \boxed{\texttt{FISHER}(x)}\ -1 < x < 1$

$\frac{e^{2x}-1}{e^{2x}+1} = \boxed{\texttt{INVERZ.FISHER}(x)}\ x \in \mathbb{R}$

$\ln\Gamma(x) = \ln\int\limits_0^\infty u^{x-1}e^{-u}\,\mathrm{d}u = \boxed{\texttt{GAMMALN}(x)}\ x > 0$

$\Gamma(x) = \int\limits_0^\infty u^{x-1}e^{-u}\,\mathrm{d}u = \boxed{\texttt{KITEVŐ(GAMMALN}(x)\texttt{)}}\ x > 0$

9.7.3. Mátrixok

$\boxed{\texttt{MDETERM(tömb)}}$ A `tömb`-ben található $n \times n$ típusú mátrix determinánsa
$\boxed{\texttt{TRANSZPONÁLÁS(tömb)}}$ A `tömb`-ben található $m \times n$ típusú mátrix transzponáltja, mely egy $n \times m$ méretű tömbben helyezkedik el (tömbképlet!).
$\boxed{\texttt{INVERZ.MÁTRIX(tömb)}}$ A `tömb`-ben található $n \times n$ típusú mátrix inverze, mely egy $n \times n$ méretű tömbben helyezkedik el (tömbképlet!).
$\boxed{\texttt{MSZORZAT(tömb1;tömb2)}}$ 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!).

9.7.4. Kombinatorika

$m! = \boxed{\texttt{FACT}(m)}$ $m \in \mathbb{N}$
$m!! = \boxed{\texttt{FACTDOUBLE}(m)}$ $m \in \mathbb{N}$ ($m!!$ 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.)
$\binom{m}{k} = \boxed{\texttt{KOMBINÁCIÓK}(m;k)}$ $m \in \mathbb{N},\ k = 0, \ldots, m$
$\frac{m!}{(m-k)!} = \boxed{\texttt{VARIÁCIÓK}(m;k)}$ $m \in \mathbb{N},\ k = 0, \ldots, m$
$\frac{(k_1+k_2+\cdots+k_r)!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!} = \boxed{\texttt{MULTINOMIAL}(k_1;k_2;\ldots;k_r)}$ $k_1, k_2, \ldots, k_r \in \mathbb{N}$

9.7.5. Pszeudo-véletlen szám generálása

$\boxed{\texttt{VÉL()}}$ $[0,1)$ intervallumon egyenletes eloszlású pszeudo-véletlen szám
$\boxed{\texttt{RANDBETWEEN}(a;b)}$ ($a, b \in \mathbb{N},\ a < b$) diszkrét egyenletes eloszlású pszeudo-véletlen szám az $\{\, a, a+1, \ldots, b \,\}$ halmazon

9.7.6. 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

$x_1^* = \boxed{\texttt{MIN(A:A)}}$
$x_n^* = \boxed{\texttt{MAX(A:A)}}$
$x_k^* = \boxed{\texttt{KICSI}(\texttt{A:A};k)}$ $k = 1, \ldots, n$
$x_{n-k}^* = \boxed{\texttt{NAGY}(\texttt{A:A};k+1)}$ $k = 0, \ldots, n-1$
$\min\{\, k : x_k^* = x_i \,\} = \boxed{\texttt{SORSZÁM}(x_i;\texttt{A:A;1})}$ $i = 1, \ldots, n$
$\min\{\, k : x_{n-k}^* = x_i \,\} + 1 = \boxed{\texttt{SORSZÁM}(x_i;\texttt{A:A;0})}$ $i = 1, \ldots, n$
$n = \boxed{\texttt{DARAB(A:A)}}$
$\overline{\xi} = \boxed{\texttt{ÁTLAG(A:A)}}$

$S_n = \boxed{\texttt{SZÓRÁSP(A:A)}}$
$S_n^2 = \boxed{\texttt{VARP(A:A)}}$
$S_n^* = \boxed{\texttt{SZÓRÁS(A:A)}}$
$S_n^{*2} = \boxed{\texttt{VAR(A:A)}}$
tapasztalati medián $= \boxed{\texttt{MEDIÁN(A:A)}}$
tapasztalati módusz $= \boxed{\texttt{MÓDUSZ(A:A)}}$
$100t$%-os tapasztalati kvantilis $= \boxed{\texttt{PERCENTILIS(A:A;}t\texttt{)}}$ $0 \leqslant t \leqslant 1$
tapasztalati alsó kvartilis $= \boxed{\texttt{KVARTILIS(A:A;1)}}$
tapasztalati felső kvartilis $= \boxed{\texttt{KVARTILIS(A:A;3)}}$
tapasztalati ferdeség $= \boxed{\texttt{FERDESÉG(A:A)}}$
tapasztalati lapultság (csúcsosság) $= \boxed{\texttt{CSÚCSOSSÁG(A:A)}}$
$\sum_{i=1}^n x_i = \boxed{\texttt{SZUM(A:A)}}$
$\sum_{i=1}^n x_i^2 = \boxed{\texttt{NÉGYZETÖSSZEG(A:A)}}$
$\sum_{i=1}^n (x_i - \overline{\xi})^2 = \boxed{\texttt{SQ(A:A)}}$
$\frac{1}{n} \sum_{i=1}^n |x_i - \overline{\xi}| = \boxed{\texttt{ÁTL.ELTÉRÉS(A:A)}}$
$\prod_{i=1}^n x_i = \boxed{\texttt{SZORZAT(A:A)}}$
$\sqrt[n]{\prod_{i=1}^n x_i} = \boxed{\texttt{MÉRTANI.KÖZÉP(A:A)}}$ $x_i > 0$ $(i = 1, \ldots, n)$
$\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}\right)^{-1} = \boxed{\texttt{HARM.KÖZÉP(A:A)}}$ $x_i > 0$ $(i = 1, \ldots, n)$
$\sum_{x_i < a} x_i = \boxed{\texttt{SZUMHA(A:A;"<}a\texttt{")}}$ $a \in \mathbb{R}$
$\sum_{a < x_i \leqslant b} x_i = \boxed{\texttt{SZUMHATÖBB(A:A;A:A;">}a\texttt{";A:A;"<=}b\texttt{")}}$ $a, b \in \mathbb{R}$
$\frac{1}{n} \sum_{x_i < a} x_i = \boxed{\texttt{ÁTLAGHA(A:A;"<}a\texttt{")}}$ $a \in \mathbb{R}$
$\frac{1}{n} \sum_{a < x_i \leqslant b} x_i = \boxed{\texttt{ÁTLAGHATÖBB(A:A;A:A;">}a\texttt{";A:A;"<=}b\texttt{")}}$ $a, b \in \mathbb{R}$
$\sum_{i=1}^n \mathrm{I}_{x_i < a} = \boxed{\texttt{DARABTELI(A:A;"<}a\texttt{")}}$ $a \in \mathbb{R}$
$\sum_{i=1}^n \mathrm{I}_{a < x_i \leqslant b} = \boxed{\texttt{DARABHATÖBB(A:A;">}a\texttt{";A:A;"<=}b\texttt{")}}$ $a, b \in \mathbb{R}$

Legyen a (ξ, η) kétdimenziós valószínűségi vektorváltozóra vonatkozó mintarealizáció $(x_1, y_1), \ldots, (x_n, y_n)$. Az `A` oszlop i-edik sorában legyen x_i, illetve a `B` oszlop i-edik sorában legyen y_i. Ekkor

$\mathrm{Cov}_n(\xi, \eta) = \boxed{\texttt{KOVAR(A:A;B:B)}}$
$\mathrm{Corr}_n(\xi, \eta) = \boxed{\texttt{KORREL(A:A;B:B)}}$
$\sum_{i=1}^n x_i y_i = \boxed{\texttt{SZORZATÖSSZEG(A:A;B:B)}}$
$\sum_{i=1}^n (x_i - y_i)^2 = \boxed{\texttt{SZUMXBŐLY2(A:A;B:B)}}$
$\sum_{i=1}^n (x_i^2 - y_i^2) = \boxed{\texttt{SZUMX2BŐLY2(A:A;B:B)}}$
$\sum_{i=1}^n (x_i^2 + y_i^2) = \boxed{\texttt{SZUMX2MEGY2(A:A;B:B)}}$

9.7.7. Eloszlások

- **Binomiális eloszlás** (r-edrendű p paraméterű)
  $\binom{r}{k}p^k(1-p)^{r-k} = \boxed{\texttt{BINOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}p\texttt{;HAMIS)}}$
  $r \in \mathbb{N},\ k = 0,\ldots,r,\ 0 < p < 1$

- **Hipergeometrikus eloszlás**
  $\frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}} = \boxed{\texttt{HIPERGEOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}M\texttt{;}N\texttt{)}}$
  $r, M, N \in \mathbb{N},\ M < N,\ r \leqslant \min\{M, N-M\},\ k = 0,\ldots,r$

- **Poisson-eloszlás** (λ paraméterű)
  $\frac{\lambda^k}{k!}e^{-\lambda} = \boxed{\texttt{POISSON(}k\texttt{;}\lambda\texttt{;HAMIS)}}$ $\lambda > 0,\ k = 0,1,\ldots$

9.7.8. Eloszlásfüggvények

- **Binomiális eloszlás** (r-edrendű p paraméterű)
  $\sum_{i=0}^{k}\binom{r}{i}p^i(1-p)^{r-i} = \boxed{\texttt{BINOM.ELOSZLÁS(}k\texttt{;}r\texttt{;}p\texttt{;IGAZ)}}$
  $r \in \mathbb{N},\ k = 0,\ldots,r,\ 0 < p < 1$

- **Poisson-eloszlás** (λ paraméterű)
  $\sum_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda} = \boxed{\texttt{POISSON(}k\texttt{;}\lambda\texttt{;IGAZ)}}$ $\lambda > 0,\ k = 0,1,\ldots$

- **Exponenciális eloszlás** (λ paraméterű)
  $F(x) = 1 - e^{-\lambda x} = \boxed{\texttt{EXP.ELOSZLÁS(}x\texttt{;}\lambda\texttt{;IGAZ)}}$ $\lambda > 0,\ x \geqslant 0$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $F(x) = \boxed{\texttt{GAMMA.ELOSZLÁS(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;IGAZ)}}$ $r, \lambda > 0,\ x \geqslant 0$

- **Standard normális eloszlás**
  $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t = \boxed{\texttt{STNORMELOSZL(}x\texttt{)}}$ $x \in \mathbb{R}$

- **Normális eloszlás** (m és σ paraméterű)
  $F(x) = \Phi\left(\frac{x-m}{\sigma}\right) = \boxed{\texttt{NORM.ELOSZL(}x\texttt{;}m\texttt{;}\sigma\texttt{;IGAZ)}}$
  $m, x \in \mathbb{R},\ \sigma > 0$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $F(x) = \boxed{\texttt{1-KHI.ELOSZLÁS(}x\texttt{;}s\texttt{)}}$ $s \in \mathbb{N},\ x \geqslant 0$

- **t-eloszlás** (s szabadsági fokú)
  $F(x) = \boxed{\texttt{1-T.ELOSZLÁS(}x\texttt{;}s\texttt{;1)}}$ $s \in \mathbb{N},\ x \geqslant 0$
  $F(x) = \boxed{\texttt{T.ELOSZLÁS(}-x\texttt{;}s\texttt{;1)}}$ $s \in \mathbb{N},\ x < 0$
  $\mathrm{P}(|\xi| > x) = 2 - 2F(x) = \boxed{\texttt{T.ELOSZLÁS(}x\texttt{;}s\texttt{;2)}}$ $s \in \mathbb{N},\ x \geqslant 0$

- **F-eloszlás** (s_1 és s_2 szabadsági fokú)
  $F(x) = \boxed{\texttt{1-F.ELOSZLÁS(}x\texttt{;}s_1\texttt{;}s_2\texttt{)}}$ $s_1, s_2 \in \mathbb{N},\ x \geqslant 0$

9.7.9. Sűrűségfüggvények

- **Exponenciális eloszlás** (λ paraméterű)
  $f(x) = \lambda e^{-\lambda x} = \boxed{\texttt{EXP.ELOSZLÁS(}x\texttt{;}\lambda\texttt{;HAMIS)}}$ $\lambda > 0,\ x \geqslant 0$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $f(x) = \boxed{\texttt{GAMMA.ELOSZLÁS(}x\texttt{;}r\texttt{;1/}\lambda\texttt{;HAMIS)}}$ $r, \lambda > 0,\ x \geqslant 0$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $f(x) = \boxed{\texttt{GAMMA.ELOSZLÁS(}x\texttt{;}s\texttt{/2;2;HAMIS)}}$ $x \geqslant 0$

- **Standard normális eloszlás**
  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \boxed{\texttt{NORM.ELOSZL(}x\texttt{;0;1;HAMIS)}}$ $x \in \mathbb{R}$

- **Normális eloszlás** (m és σ paraméterű)
  $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right) = \boxed{\texttt{NORM.ELOSZL(}x\texttt{;}m\texttt{;}\sigma\texttt{;HAMIS)}}$
  $m, x \in \mathbb{R},\ \sigma > 0$

9.7.10. Inverz eloszlásfüggvények

- **Normális eloszlás** (m és σ paraméterű)
  $F^{-1}(x) = \boxed{\texttt{INVERZ.NORM(}x\texttt{;}m\texttt{;}\sigma\texttt{)}}$ $m \in \mathbb{R},\ \sigma > 0,\ 0 < x < 1$

- **Standard normális eloszlás**
  $\Phi^{-1}(x) = \boxed{\texttt{INVERZ.STNORM(}x\texttt{)}}$ $0 < x < 1$

- **Khi-négyzet eloszlás** (s szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{INVERZ.KHI(}1-x\texttt{;}s\texttt{)}}$ $s \in \mathbb{N},\ 0 < x < 1$

- **t-eloszlás** (s szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{-INVERZ.T(}2x\texttt{;}s\texttt{)}}$ $s \in \mathbb{N},\ 0 < x < 0{,}5$
  $F^{-1}(x) = \boxed{\texttt{INVERZ.T(}2-2x\texttt{;}s\texttt{)}}$ $s \in \mathbb{N},\ 0{,}5 \leqslant x < 1$

- **Gamma-eloszlás** (r-edrendű λ paraméterű)
  $F^{-1}(x) = \boxed{\texttt{INVERZ.GAMMA(}x\texttt{;}r\texttt{;1/}\lambda\texttt{)}}$ $r, \lambda > 0,\ 0 < x < 1$

- **F-eloszlás** (s_1 és s_2 szabadsági fokú)
  $F^{-1}(x) = \boxed{\texttt{INVERZ.F(}1-x\texttt{;}s_1\texttt{;}s_2\texttt{)}}$ $s_1, s_2 \in \mathbb{N},\ 0 < x < 1$

9.7.11. 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.

9.7.12. Intervallumbecslés

$\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = \boxed{\texttt{MEGBÍZHATÓSÁG(}\alpha\texttt{;}\sigma\texttt{;}n\texttt{)}}$ $0 < \alpha < 1$, $\sigma > 0$, $n \in \mathbb{N}$

$\frac{S_n^*}{\sqrt{n}}F^{-1}\left(1-\frac{\alpha}{2}\right) \simeq \boxed{\texttt{MEGBÍZHATÓSÁG(}\alpha\texttt{;}S_n^*\texttt{;}n\texttt{)}}$ $F \sim \mathbf{t}(n-1)$, $0 < \alpha < 1$, $\sigma > 0$, $n \in \mathbb{N}$. A becslés annál pontosabb, minél nagyobb az n.

$\min\left\{ c \in \mathbb{N} : \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \geqslant x \right\} = \boxed{\texttt{KRITBINOM(}n\texttt{;}p\texttt{;}x\texttt{)}}$ $n \in \mathbb{N}$, $0 < p < 1$, $0 < x < 1$

9.7.13. Paraméteres hipotézisvizsgálatok

A ξ-re illetve η-ra vonatkozó mintarealizációk az `A` illetve `B` oszlopokban vannak.

- **Egymintás u-próba**
  $1-\Phi(u) = \boxed{\texttt{Z.PRÓBA(A:A;}m_0\texttt{;}\sigma\texttt{)}}$
  $2-2\Phi(|u|) = \boxed{\texttt{2*MIN(Z.PRÓBA(A:A;}m_0\texttt{;}\sigma\texttt{);1-Z.PRÓBA(A:A;}m_0\texttt{;}\sigma\texttt{))}}$

- **Egymintás t-próba**
  A ξ-re vonatkozó mintarealizáció minden tagja mellett szerepeljen m_0 értéke a `B` oszlopban.
  $2-2F(|t|) = \boxed{\texttt{T.PRÓBA(A:A;B:B;2;1)}}$
  $F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;1)}}$ ha $\overline{\xi} \leqslant m_0$
  $1-F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;1)}}$ ha $\overline{\xi} \geqslant m_0$

- **F-próba**
  $2\min\{ F(\mathsf{F}), 1-F(\mathsf{F}) \} = \boxed{\texttt{F.PRÓBA(A:A;B:B)}}$
  $F(\mathsf{F}) = \boxed{\texttt{F.PRÓBA(A:A;B:B)/2}}$, ha ${S^*_{\xi,n_1}}^2 \leqslant {S^*_{\eta,n_2}}^2$
  $1-F(\mathsf{F}) = \boxed{\texttt{F.PRÓBA(A:A;B:B)/2}}$, ha ${S^*_{\xi,n_1}}^2 \geqslant {S^*_{\eta,n_2}}^2$

- **Kétmintás t-próba**
  $2-2F(|t|) = \boxed{\texttt{T.PRÓBA(A:A;B:B;2;2)}}$
  $F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;2)}}$ ha $\overline{\xi} \leqslant \overline{\eta}$
  $1-F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;2)}}$ ha $\overline{\xi} \geqslant \overline{\eta}$

- **Scheffé-módszer azonos mintaelemszámra**
  $2-2F(|t|) = \boxed{\texttt{T.PRÓBA(A:A;B:B;2;1)}}$
  $F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;1)}}$ ha $\overline{\xi} \leqslant \overline{\eta}$
  $1-F(t) = \boxed{\texttt{T.PRÓBA(A:A;B:B;1;1)}}$ ha $\overline{\xi} \geqslant \overline{\eta}$

- **Scheffé-módszer különböző mintaelemszámra**
  Az ζ-ra vonatkozó mintarealizáció a `C` oszlopban van, és minden tagja mellett szerepeljen 0 a `D` oszlopban.
  $2-2F(|t|) = \boxed{\texttt{T.PRÓBA(C:C;D:D;2;1)}}$
  $F(t) = \boxed{\texttt{T.PRÓBA(C:C;D:D;1;1)}}$ ha $\overline{\xi} \leqslant \overline{\eta}$
  $1-F(t) = \boxed{\texttt{T.PRÓBA(C:C;D:D;1;1)}}$ ha $\overline{\xi} \geqslant \overline{\eta}$

- **Statisztikai próba az exponenciális eloszlás paraméterére**
  $F(\gamma) = \boxed{\texttt{GAMMA.ELOSZLÁS(}\lambda_0\texttt{*SZUM(A:A);DARAB(A:A);1;IGAZ)}}$

- **Statisztikai próba valószínűségre**
  $\min\left\{ z \in \mathbb{N} : \sum_{i=0}^{z} \binom{n}{i} p_0^i (1-p_0)^{n-i} \geqslant x \right\} = \boxed{\texttt{KRITBINOM(}n\texttt{;}p_0\texttt{;}x\texttt{)}}$

9.7.14. Nemparaméteres hipotézisvizsgálatok

- **Tiszta illeszkedésvizsgálat**
  $1-F(\chi^2) = \boxed{\texttt{KHI.PRÓBA(}\varrho_i\texttt{ tartománya;}np_i\texttt{ tartománya)}}$

- **Becsléses illeszkedésvizsgálat**
  $\Sigma_i := \frac{(\varrho_i - n\widehat{p}_i)^2}{n\widehat{p}_i}$
  $1-F(\chi^2) = \boxed{\texttt{=KHI.ELOSZLÁS(SZUM(}\Sigma_i\texttt{ tartománya);}r-1-v\texttt{)}}$

- **Függetlenségvizsgálat**
  $1-F(\chi^2) = \boxed{\texttt{KHI.PRÓBA(}\varrho_{ij}\texttt{ tartománya;}k_i l_j/n\texttt{ tartománya)}}$

- **Homogenitásvizsgálat**
  $\nu_{ij} := \frac{(k_i+l_i)n_j}{n_1+n_2}$
  $1-F(\chi^2) = \boxed{\texttt{KHI.PRÓBA(}k_i, l_j\texttt{ tartománya;}\nu_{ij}\texttt{ tartománya)}}$

- **Kétmintás előjelpróba**
  $F^{-1}(x) = \boxed{\texttt{KRITBINOM(}n\texttt{;1/2;}x\texttt{)}}$

9.7.15. 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.

  $(\widehat{a}_k, \widehat{a}_{k-1}, \ldots, \widehat{a}_1) =$ `LIN.ILL(eta-t;xi-t;HAMIS)` $(1 \times k$ méretű tömbképlet!)

  $\widehat{a}_1(x_1 - t_1) + \cdots + \widehat{a}_k(x_k - t_k) =$ `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.

  $(\widehat{b}, \widehat{a}) =$ `LOG.ILL(eta;xi)` $(1 \times 2$ méretű tömbképlet!)

  $\widehat{a} \cdot \widehat{b}^x =$ `NÖV(eta;xi;x)`

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 (http://www.sulinet.hu/termeszetvilaga/archiv/2000/0014/02.html).

[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, Eszterházy Károly Főiskola, Eger, 2012.